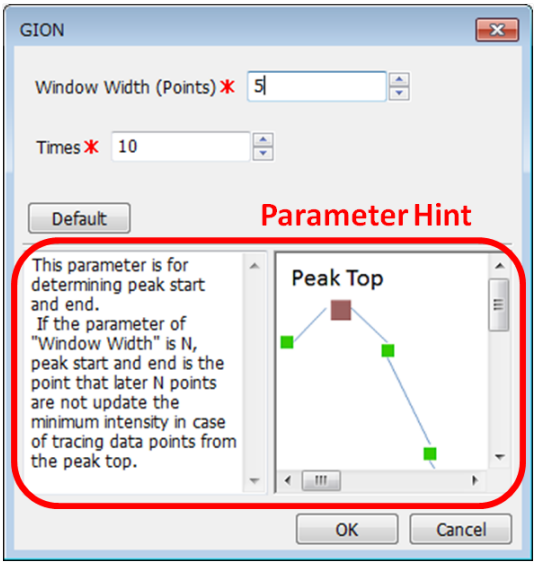
Mass++ Beginners' Guide

Mass++ Beginners' Guide

Table of Contents

Chapter 1. Basic Operation

This chapter describes the basic operations of Mass++. For details regarding each parameter upon operation, refer to [Parameter Hint] displayed at the bottom of each dialog.



1.1. File Open and Processing

This section describes the steps for performing basic operations such as peak detection, comparison of MS spectra, and smoothing of measured spectra in Mass++.

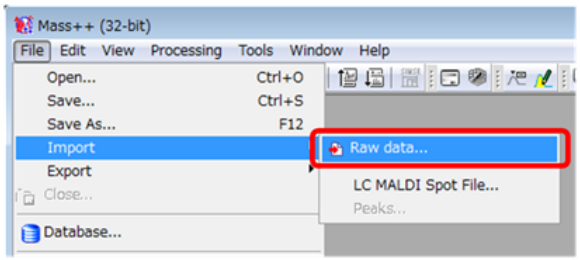
The sample data (and sample Python code: refer "Script Consol" section) that supplied with Mass++ is saved in the following location.

[Folder where Mass++ is installed]/Mass++/data

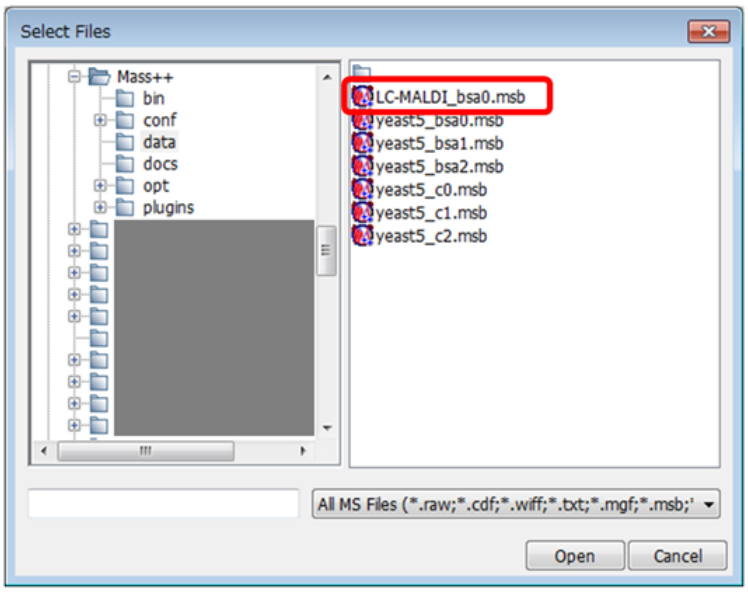
1.1.1. Opening a File

First, open the target data.

After starting Mass++, click [Import] - [Raw Data] from the [File] menu.



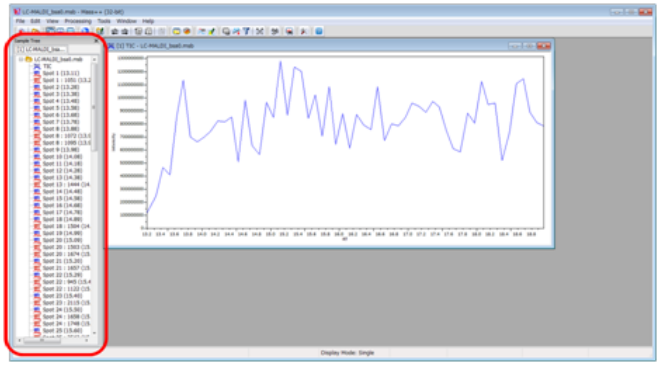
[Select Files] opens.



Here, open the sample data "LC-MALDI_bsa0.msb"that supplied with Mass++.

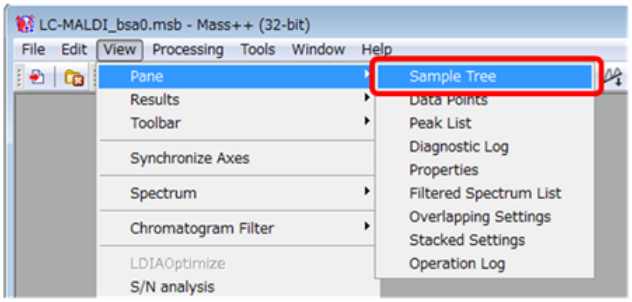
Select the target file and click [Open].

The sample opened in the [Sample Tree] of the main window is displayed. The top spectrum (or chromatogram) of the [Sample Tree] is automatically displayed in the main window.



In this data, a chromatogram opened. This can be closed by clicking the [x] in the upper-right corner of the window.

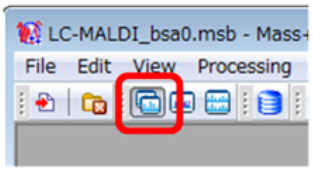
Note: If the [Sample Tree] is not visible, click [Pane] - [Sample Tree] from the [View] menu.



1.1.2. Opening a Spectrum

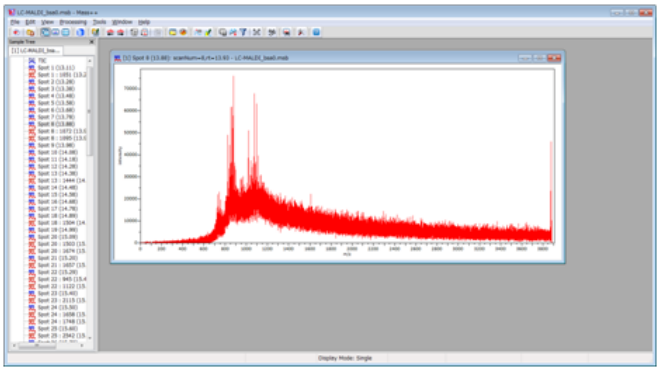
Open a spectrum.

Check that [Display mode] is [Single].

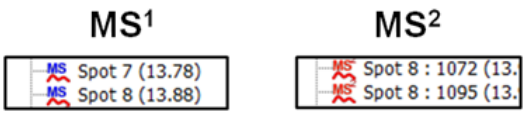


Click on the target spectrum in the [Sample Tree].

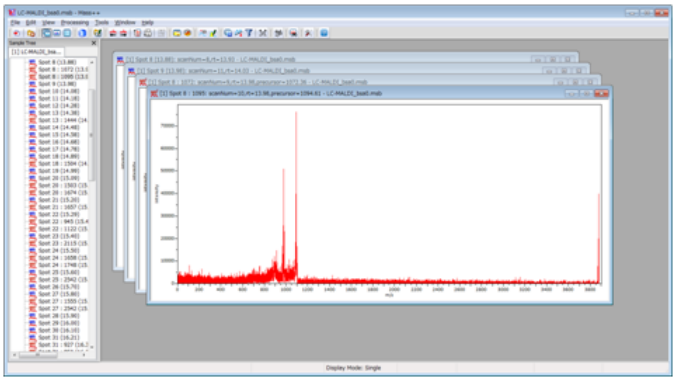
The spectrum is displayed in the main window.



Spectra are often compared by opening in aligned windows. As an example here, 2 MS spectra and 2 MS2 spectra are opened. The MS1 spectra and MS2 spectra are displayed with icons in the [Sample Tree] as shown in the respective figures.



If [Display Mode] is [Single], a new window will be added as shown in the figure.

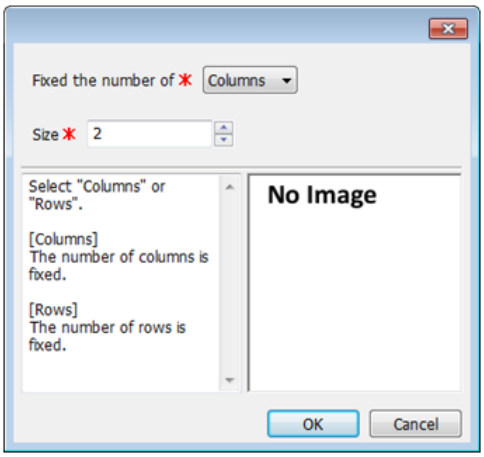


1.1.3. Window Alignment

The four open spectra are aligned 2 windows x 2 windows for easy comparison. Although there are several alignment methods, [Auto Arrange Mode] is used here. In [Auto Arrange Mode], the display position of 2 windows can be switched by dragging and dropping.

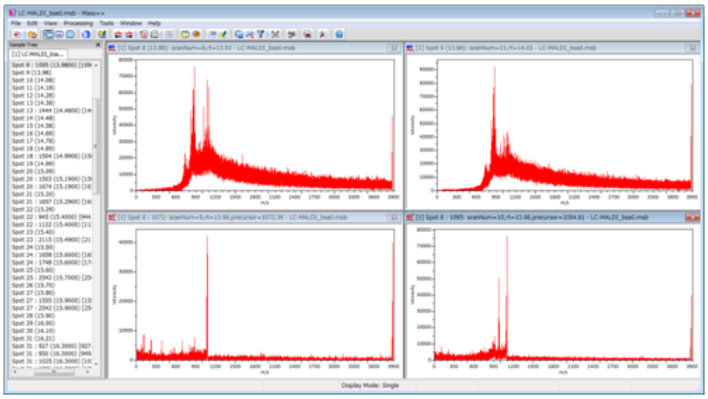
Click [Auto Arrange Mode] from the [Window] menu.

The settings dialog opens.

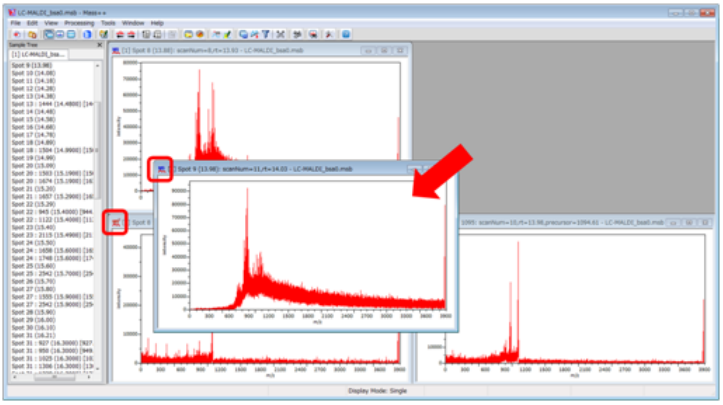


Set both [Columns] and [Rows] under [Fixed the number of] to 2. Click [OK] to close the dialog.

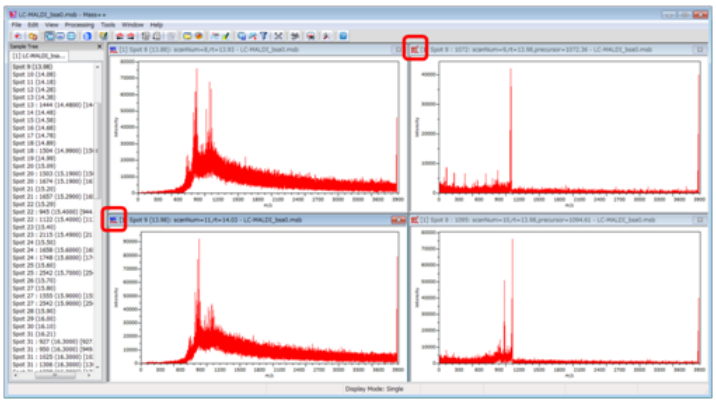
The windows will be aligned.



In [Auto Arrange Mode], to switch a window's display location with another window's display location, click the header of one window and drag onto the window and drop.



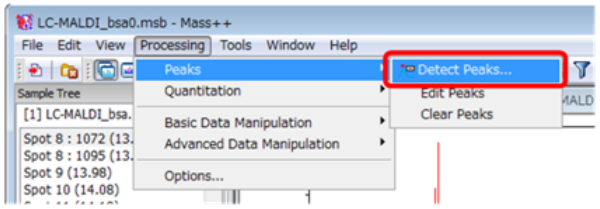
The display position of the window will be changed.



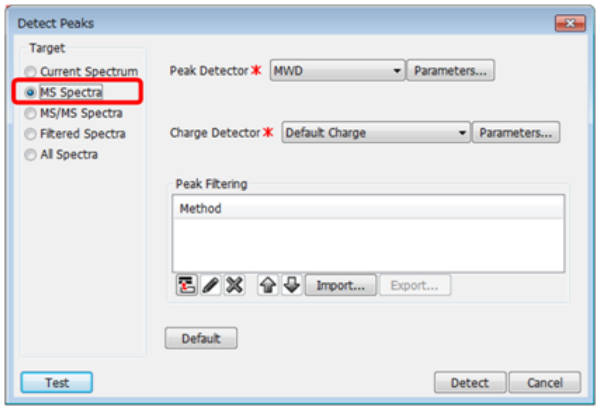
1.1.4. Peak Detection

After the spectrum window is displayed, the spectrum peaks are detected.

Click [Peaks] - [Detect Peaks...] from the [Processing] menu.



[Detect Peaks] opens.



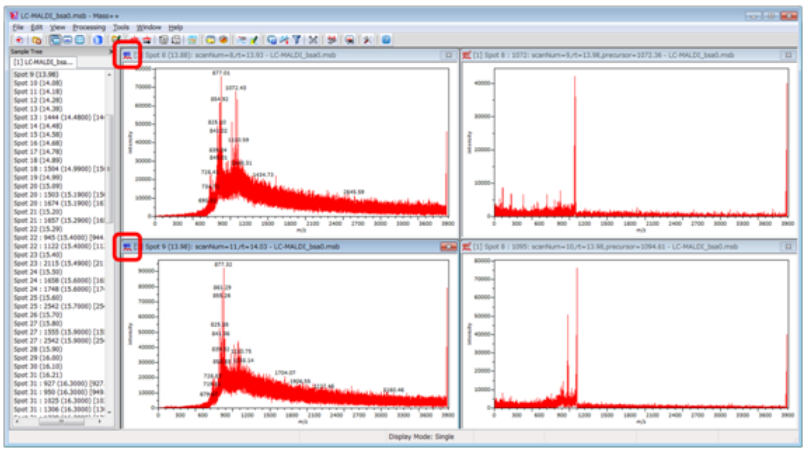
Here, all of the peaks of the MS spectra included in the data are detected.

Select the [MS Spectra] under [Target].

Refer to the individual [Parameter Hint] for other individual settings such as [Peak Detector].

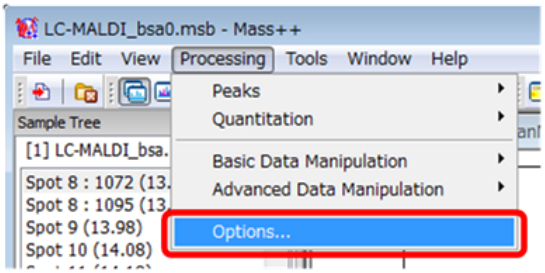
Click [Detect].

Peaks are detected only on the MS spectrum. The m/z value is displayed for the detected peak.

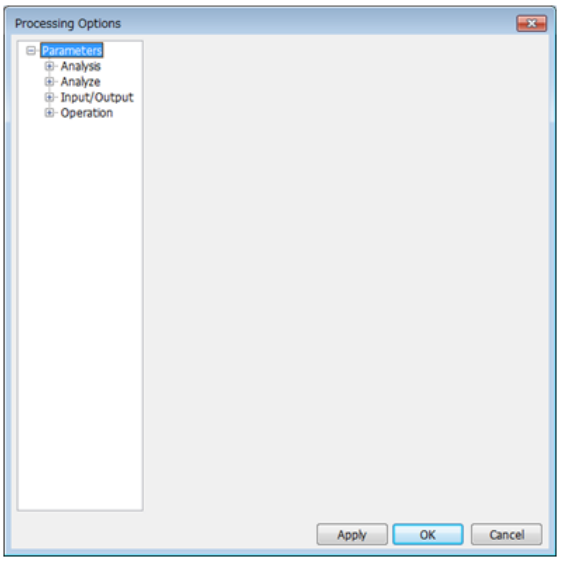


Note: When opening the spectrum window, configure the following settings to automatically perform peak detection on the opened spectrum.

Click [Options] from the [Processing] menu.

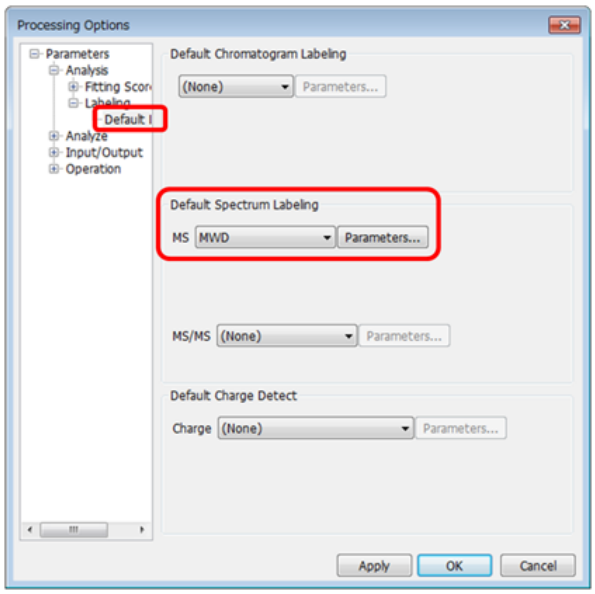


[Processing Options] opens.



Click [Analysis] - [Labeling] - [Default Labeling] from the [Parameters] menu.

The [Default Labeling] settings screen appears.



In this settings screen, the settings for whether to perform peak detection on the open spectrum or chromatogram and also for which peak detection function to use in such cases can be configured.

Click [OK] to close [Processing Options].

Open fresh MS spectra. Depending on the settings, peak detection is automatically performed.

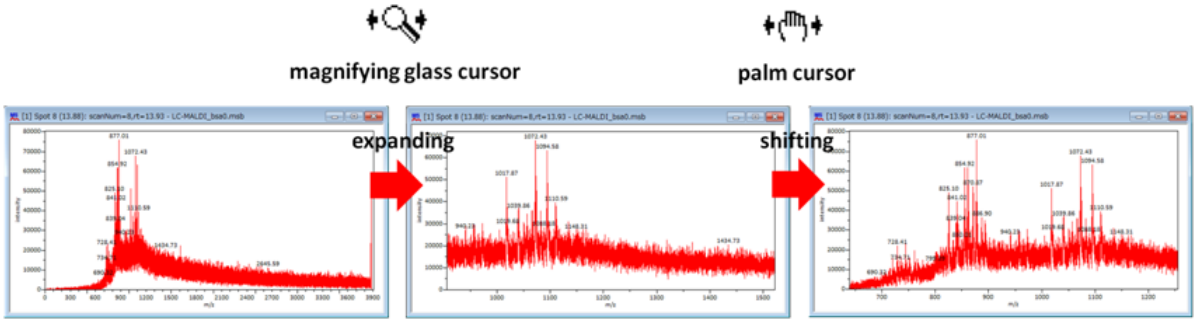
1.1.5. Zooming In, Zooming Out, and Shifting

Next, zoom the horizontal axis (m/z or RT) in or out to check the details of the detected peaks.

Move the mouse pointer to the bottom of the graph axis. Make sure that the mouse pointer becomes a magnifying glass when right-clicked. Dragging right while clicking will zoom in, and dragging left will zoom out.

In addition, make sure that the mouse pointer becomes a palm when clicking.

The display area will be shifted by dragging left or right while clicked.

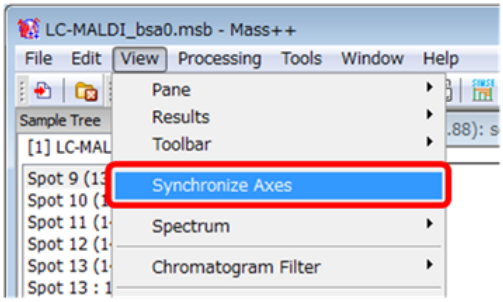


Zooming in, zooming out, and display area shifting can be performed similar to displaying the vertical axis (peak intensity). This zooming in and out can be done regardless of whether peak detection is enabled or disabled.

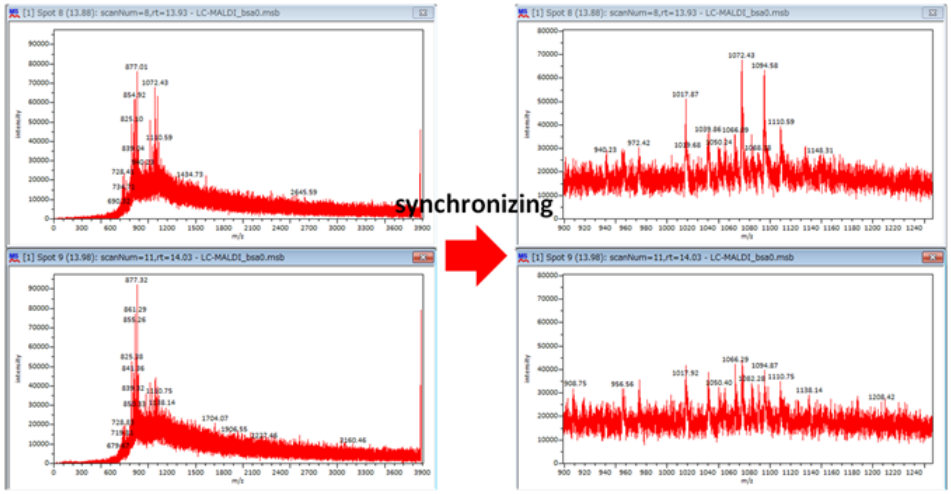
1.1.6. Axis Synchronization

If comparing multiple spectra, when changing the display width of one spectrum, the display width of the other spectra can be conveniently changed automatically in the same way. In this step, the axes display of the spectrum displayed on the main window are synchronized.

Click [Synchronize Axes] from the [View] menu.



When zooming the axis display of a spectrum in/out, make sure that the axis displays of other spectra zoom in/out in the same way.

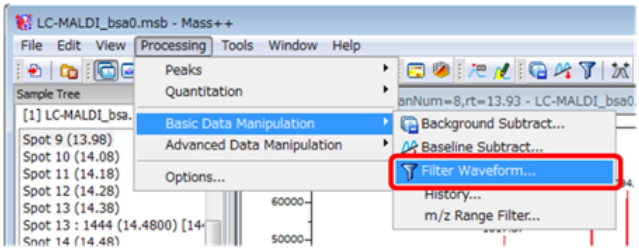


1.1.7. Data Processing (Smoothing)

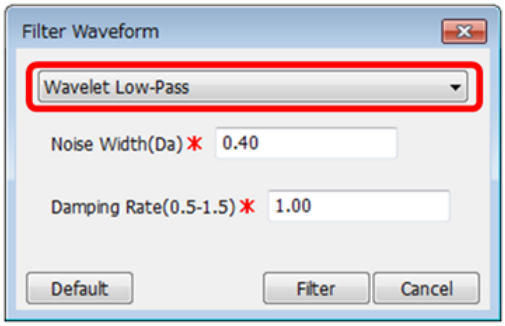
The operations that can be done on a spectrum is not only peak detection. This section describes the noise rejection using smoothing as an example of data processing to a spectrum.

Here, smoothing is applied to a spectrum being displayed.

Click [Basic Data Manipulation] - [Filter Waveform] from the [Processing] menu.



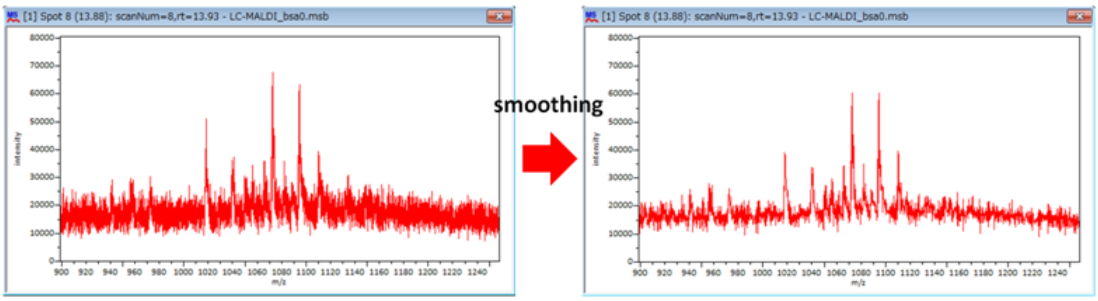
[Filter Waveform] opens.



Here, smoothing is done using Wavelet. Select [Wavelet Low-Pass].

Click [Filter].

Smoothing is applied to the selected spectrum.



If necessary, zoom the display range in and out to check the waveform.

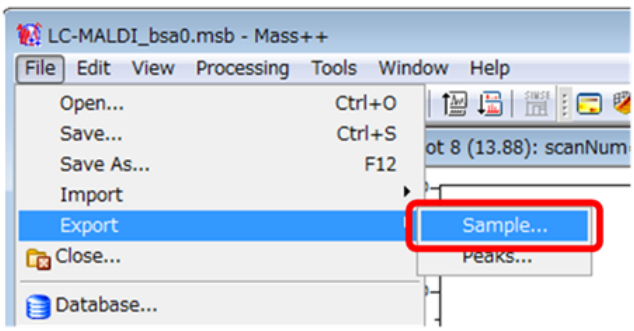
1.1.8. Saving a Spectrum

Save the spectrum after performing the necessary processing.

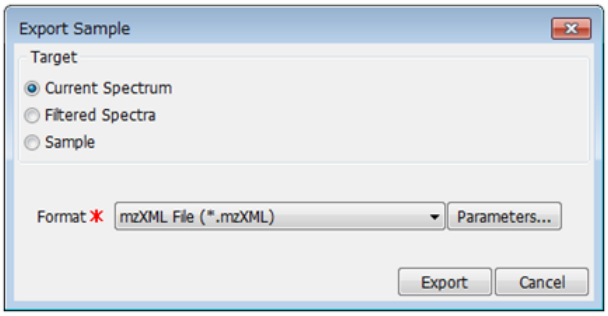
Here, the spectrum after smoothing is saved in the mzXML format.

Select the window where the target spectrum is displayed.

Click [Export] - [Sample...] from the [File] menu.



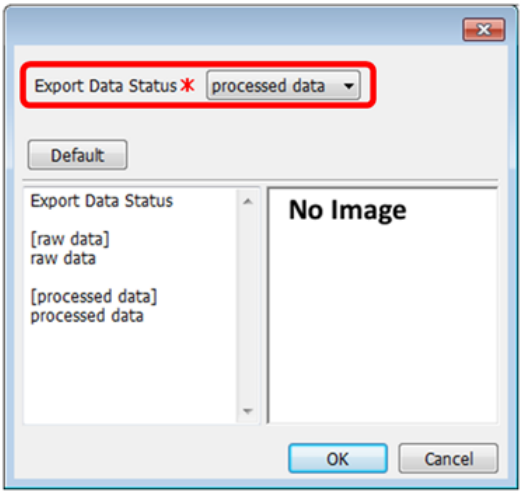
[Export Sample] opens.



Set [Target] to [Current Spectrum] and [Format] to [mzXML File (*.MZXML)].

Click the [Parameters] button.

The parameter settings dialog opens.



In order to output the spectrum for which processing (smoothing, in this case) was performed, set [Export Data Status] to [Processed Data].

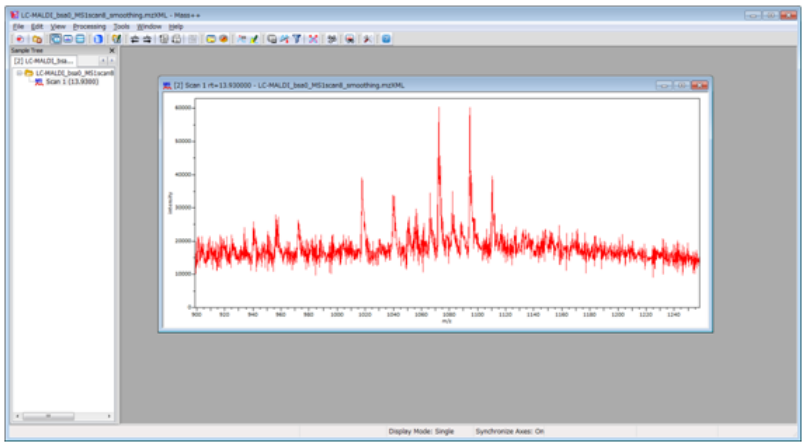
Click [OK] to close the dialog.

Click [Export] under [Export Sample].

[Save As] opens. Input a name to save the file.

For verification purposes, open the output file.

Smoothed data will be saved.

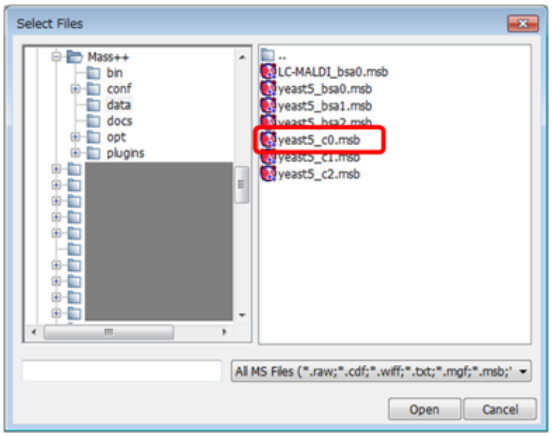


1.2. Heatmap and 3D Display

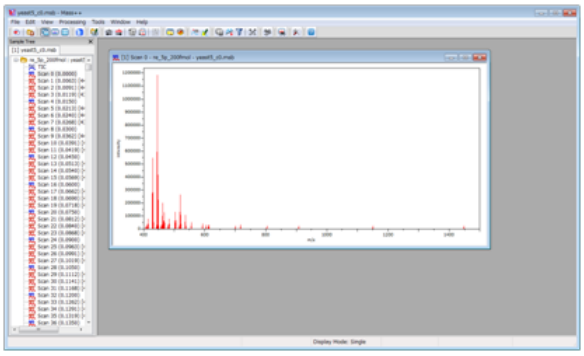
This section describes how to display the RT × m/z two-dimensional data of LC-MS, LC-MALDI, etc., in a heatmap display or in 3D.

1.2.1. Opening a File

First, open the target data. As an example here, from among the LC-MS data that supplied with Mass++, "yeast5_c0.msb" opens.



Then, open a spectrum or a chromatogram in the data. Here, open a spectrum.



Preparation is complete.

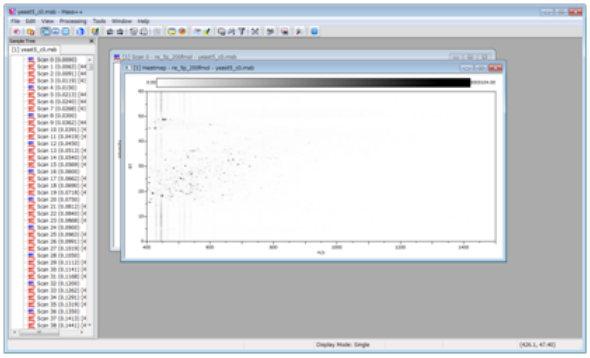
1.2.2. Displaying the Heatmap

Display the heatmap of the opened LC-MS data.

Clicking on the window where the spectrum is displayed will activate it. (If only one window is open in the main window, it has been already active.) The data including the active spectrum will be displayed in heatmap.

Click the following icon.

The heatmap is displayed.

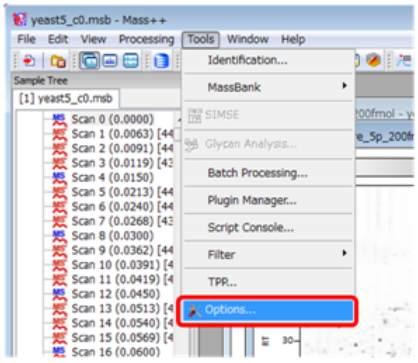


1.2.3. Changing the Display Settings

Under the default settings when installing Mass++, the heatmap will appear as previously shown. That is, the display format becomes "Grayscale" and the peak intensity display scale becomes "Liner".

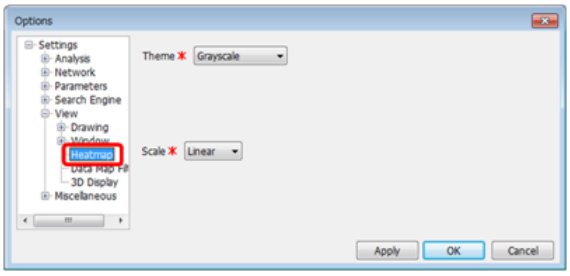
See the following for how to change the display format for increased clarity, how to change Scale to Log in order to highlight small peaks, etc.

Click [Options] from the [Tools] menu.



[Options] is displayed.

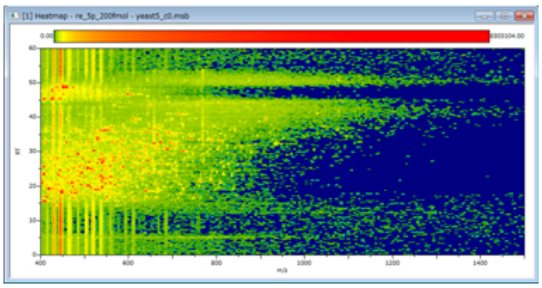
Click [Settings] - [Display] - [Heatmap].



Select the applicable display format from [Theme] and the Scale of the peak intensity from [Scale].

By clicking [Apply], [Options] will not close, and the changed display format will be reflected in the main window heatmap.

For example, if [Apply] is clicked with [Thermography] selected under [Theme] and with [Log] selected under [Scale], the display will change as follows.



Select the appropriate format and click [OK] to close [Options].

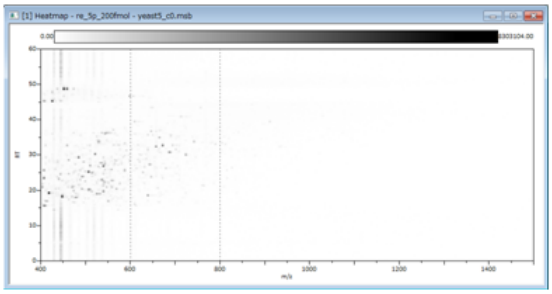
The following explanations will assume the default values for [Theme] and [Scale].

1.2.4. Changing the Display Width (1)

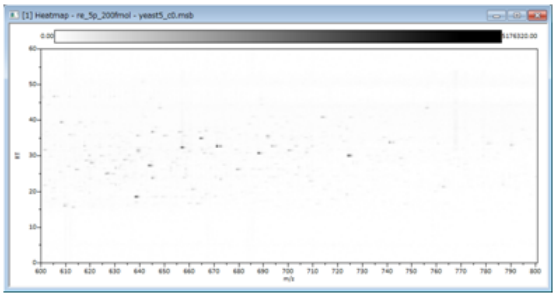
There are three methods for changing the display width for detailed examination of a particular peak on the heatmap.

The first is to move the mouse pointer to the outside of the heat map display axis, and drag-and-drop on the area to zoom in on.

In the following example, dragging is done along the m/z axis from around m/z = 600 to around 800. Dotted lines are drawn at around m/z = 600 and 800.



Dropping will enlarge that area.



The zoom operation can also be done in the RT direction.

To return to the initial display area, double-click on the window where the heatmap is displayed.

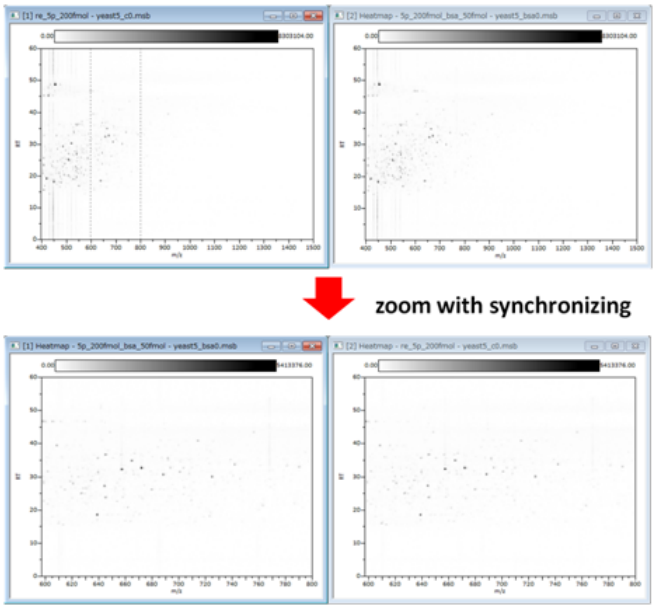
Note: Double-click on the window where the heatmap is displayed after making several times zooming, the display area returns to the initial state.

To "return to the previous display", click [Undo] from the [Edit] menu.

The method for resetting these display areas is similar for all zoom operations.

Note: For multiple heatmaps, to zoom in on a display range at the same time, check [Synchronize Axes] in the [Display] menu.

The zoom operations will be linked.

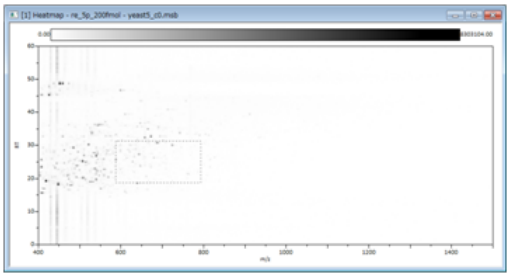


1.2.5. Changing the Display Width (2)

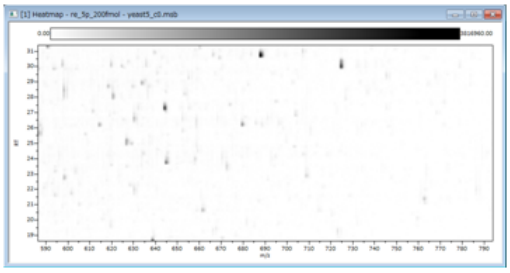
The second method is to move the mouse pointer to within the heat map display axis, and drag-and-drop on the area to zoom in on (m/z × RT).

After moving the mouse pointer to within the heatmap display axis, the icon will become a magnifying glass once clicked.

In the following example, dragging is done along the m/z axis from around (RT, m/z) = (30, 600) to around (20, 800). A dotted line is drawn around that area.



Dropping will enlarge that area.



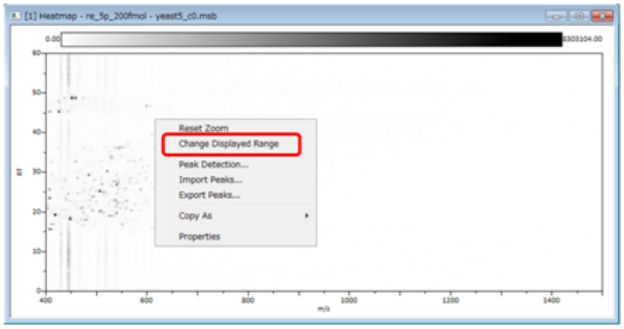
1.2.6. Changing the Display Width (3)

The two methods thus far only allow rough specification of the zoom region.

The third method is to specify the value of the range to be enlarged.

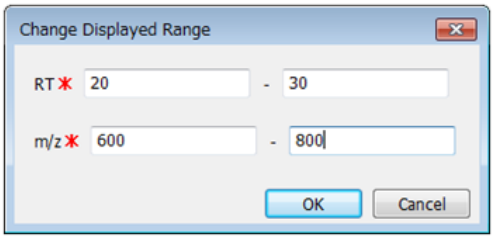
Move the mouse pointer into the display width of the heatmap, and right-click.

The right-click menu is displayed.



Click [Change Displayed Range].

[Change Displayed Range] is displayed. Input the RT width and the m/z width. The figure shows settings for enlarging the display from (RT, m/z) = (30, 600) to (20, 800).



Click [OK] to close the dialog. Check that the display range has been enlarged.

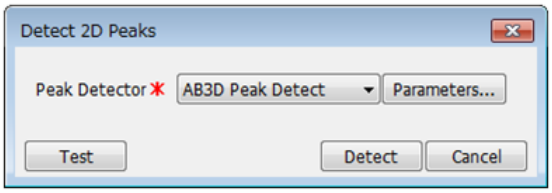
1.2.7. Two-Dimensional Peak Detection

Two-dimensional peak detection is performed to detect peaks on the heatmap.

Two-dimensional peak detection is a form of peak detection that takes isotopic distribution into account.

Click on the window with the heatmap of the two-dimensional peak detection target displayed to activate it.

Click [Peaks] - [Detect Peaks...] from the [Processing] menu.



Select the applicable peak detection function from [Peak Detector].

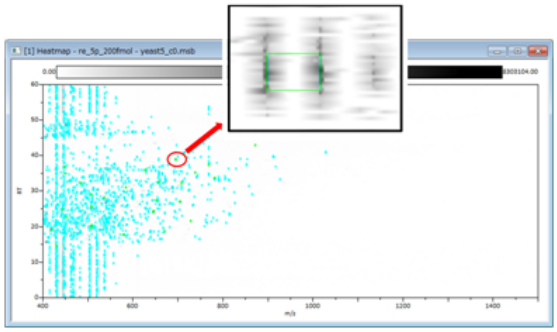
Select "detect MRM3D Peaks" for MRM data.

Because the currently open data is not MRM data, select "AB3D Peak Detect" here.

Click [Detect].

Two-dimensional peak detection is processed over the entire heatmap.

The detected peaks are surrounded by a blue square. Those peaks that are determined to have shared isotopic peaks are surrounded by a green square.



Note: If the data is large, two-dimensional peak detection for the entire data may take a lot of time. The following describes how to perform two-dimensional peak detection for only a specific range.

First, zoom in on the area where two-dimensional peak detection will be performed.

Next, open [Detect 2D Peaks], and after configuring the settings, click [Test].

Two-dimensional peak detection is done only for the display area.

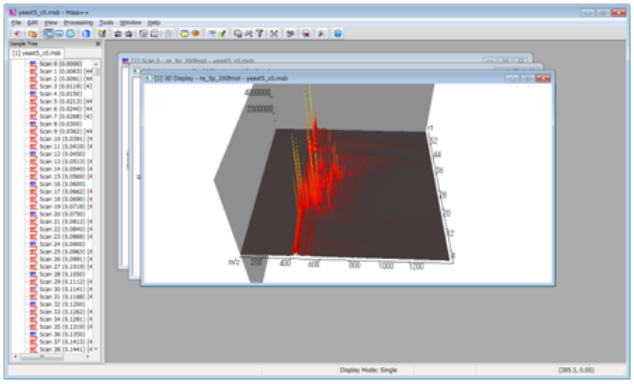
1.2.8. 3D Display

The next section describes the 3D Display.

As with the heatmap, the data from the active spectrum will be subject to the display of the 3D Display.

Click the following icon.

After data processing is performed, the 3D Display is displayed.



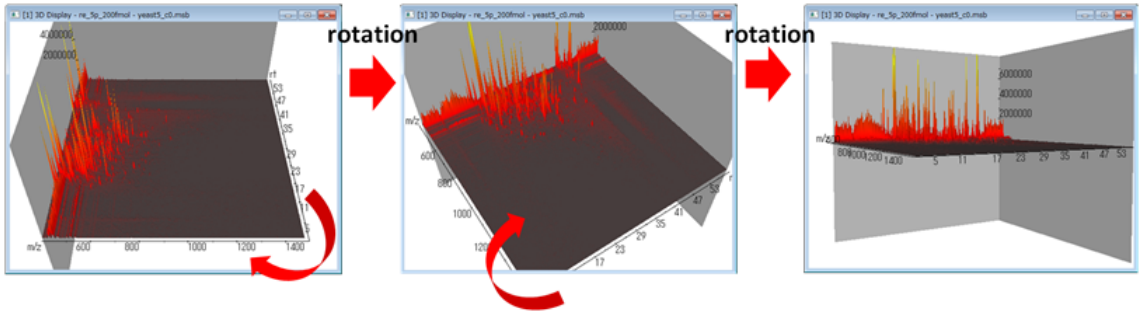
1.2.9. Rotation Display

The following describes how to rotate the 3D Display for detailed observation of a peak.

Move the mouse pointer into the display width of the 3D Display, and left-click.

The mouse pointer will be displayed as an arrow.

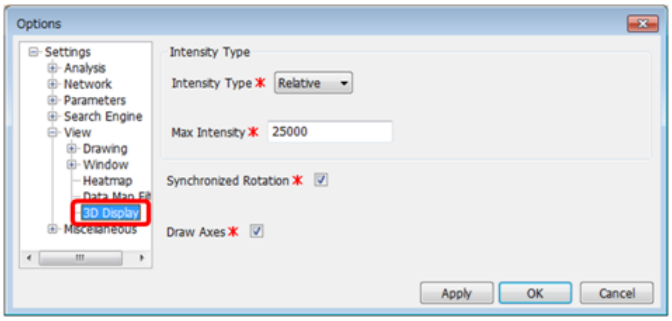
Dragging in that state will rotate the 3D Display. Dropping will stop the rotation.



1.2.10. Changing the Display Settings

To change the display settings, access [Options] as with the heatmap display.

Click [Settings] - [Display] - [3D Display].



Configure the settings for displaying the peak intensity with an absolute value or with a relative value, and the maximum intensity value to display.

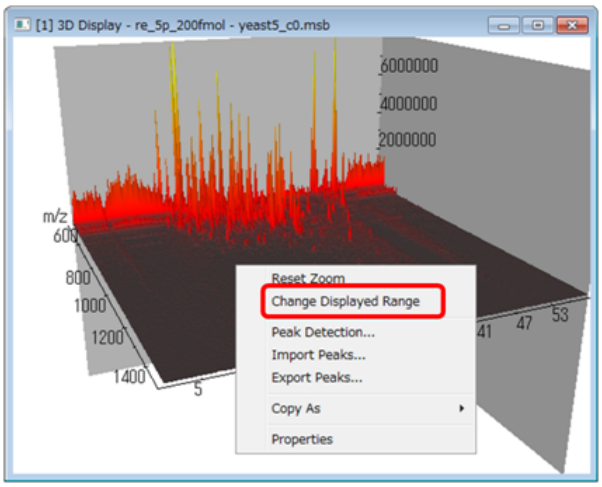
Note: When multiple 3D Displays are open, checking [Synchronized Rotation] will link display rotation.

1.2.11. Changing the Display Width

Change the display width of the 3D Display through the right-click menu.

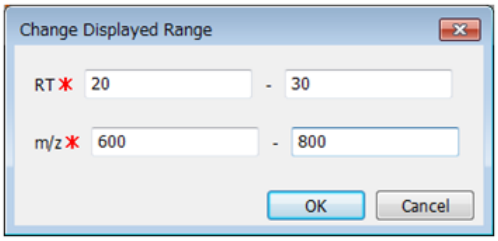
Move the mouse pointer into the display width of the 3D Display, and right-click.

The right-click menu is displayed.

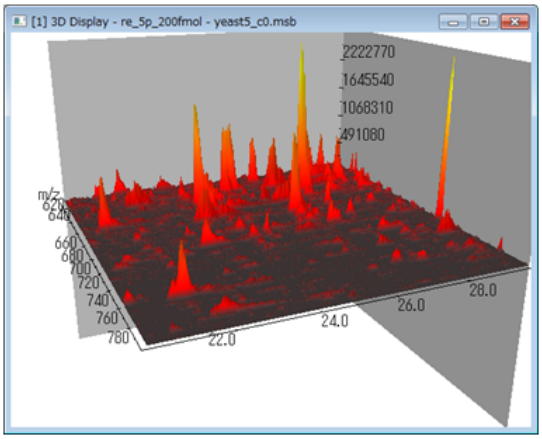


Click [Change Displayed Range].

[Change Displayed Range] is displayed. Input the RT width and the m/z width. The figure shows settings for enlarging the display from (RT, m/z) = (30, 600) to (20, 800).



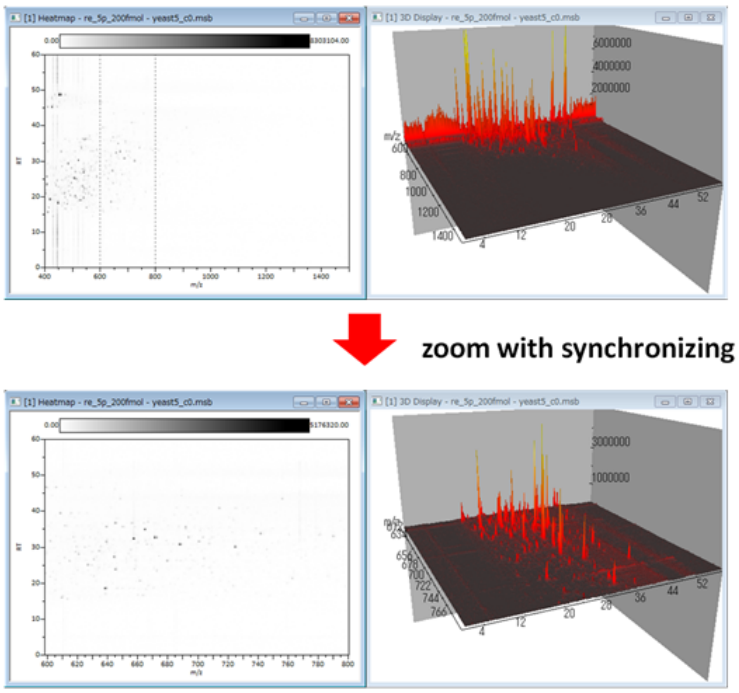
Click [OK] to close the dialog. The display range is enlarged.



Note: It is also possible to zoom in the following manner.

Check [Synchronize Axes] from the [Display] menu on the displayed heatmap.

Zooming on the heatmap will also zoom the 3D Display in tandem.


[1] Heatmap - re_5p_200fmol - yeast5_c0.msb
[1] 3D Display - re_5p_200fmol - yeast5_c0.msb
zoom with synchronizing

Chapter 2. Differential Analysis

This chapter describes how to perform differential analysis between data groups for biomarker discovery. Mass++ can search for differential peaks by using peak identification, statistical analysis, and multivariate analysis complementarily.

2.1. Create Peak Matrix

In biomarker discovery using MS1 data, "Peak Matrix" should be created using MS1 data measured from healthy person group and patient group as a first step.

"Peak Matrix" is a matrix consists of each MS data name, peak positions, peak values, identification results, statistical test results, and so on.

In the peak matrix, if peaks do not exist in healthy group commonly at a certain position but exist in patient group commonly (or vice versa), such peaks are "biomarker candidates" that discriminates the disease.

This section describes the steps for creating peak matrix using following sample data that supplied with Mass++.

(1) yeast5_c0～2.msb

Data consisting of mixed peptides derived form 5 kinds of yeasts.

(2) yeast5_bsa0～2.msb

Data adding peptides derived from BSA (bovine serum albumin) to (1).

In this case, the peptides' peaks derived from BSA are expected to exist only in (2).

In Mass++, wizard helps to create peak matrix.

Note: If you want to conduct database search to identify the peaks in peak matrix, you must establish database search settings before creating peak matrix.

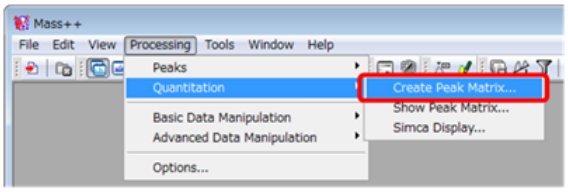
Refer to [Identification] section to check database search settings.

Note: Mass++ can crate peak matrix of MS data with LC (LC-MS (ESI), LC-MALDI), and without LC (MALDI).

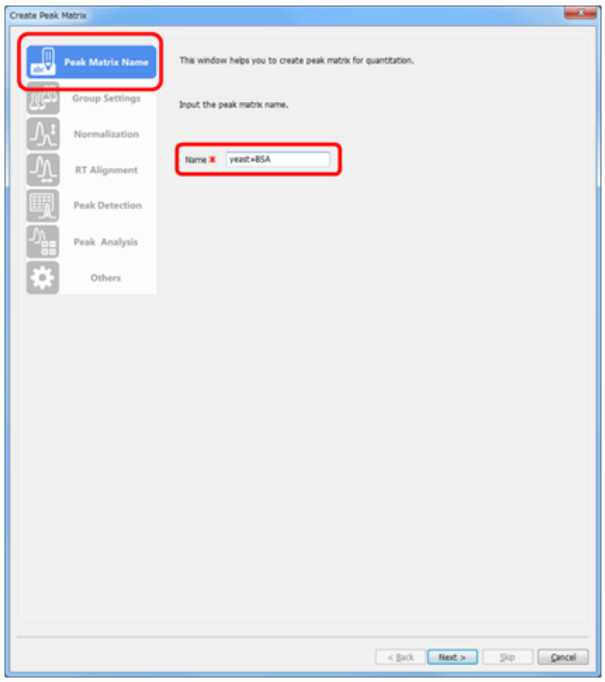
Note: It may take a long time to create peak matrix depending on data size.

2.1.1. Peak Matrix Name

Click [Quantitation] - [Create Peak Matrix] from the [Processing] menu.



[Create Peak Matrix] wizard is opened and [Peak Matrix Name] is displayed.



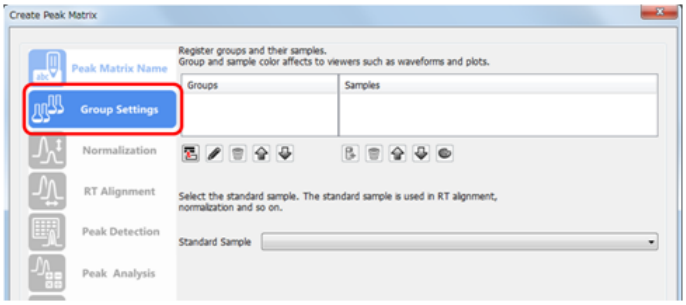
Enter peak matrix name in [Name].

Go to next step.

Click [OK].

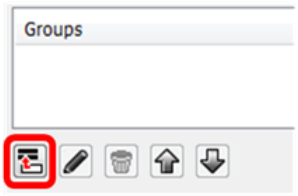
2.1.2. Group Settings

[Group Setting] is displayed.

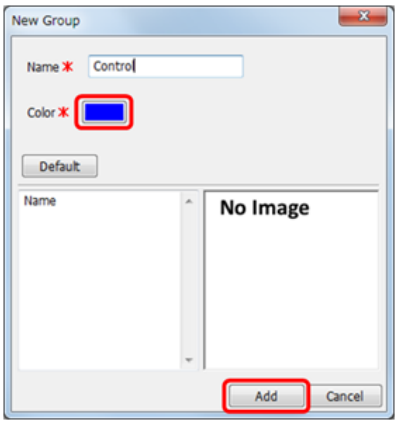


Register group name in [Groups].

Click the next icon.



[New Group] is opened.



Enter the group name in [Name].

Here, enter "Control".

Select [Color].

Click the place colored square.

[Color Setting] is opened.

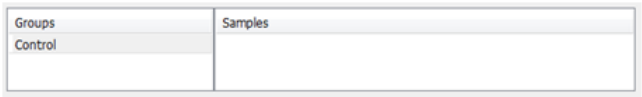
In this guide, select color blue.

After creating peak matrix, peak profiles around each peak position (RT or m/z) can be displayed.

In the peak profiles, each peak is colored with group color.

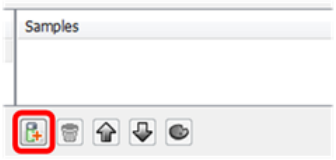
Click [Add] in [New Group].

Group name is registered.



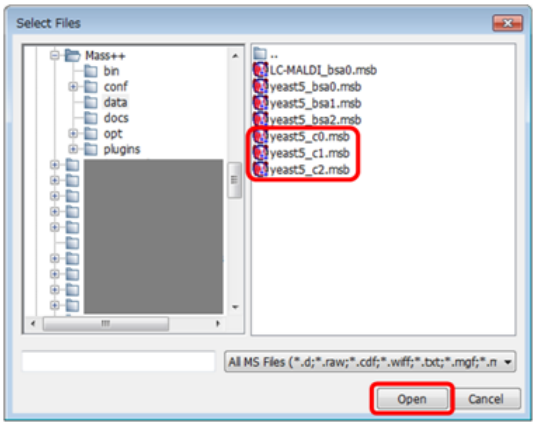
Register the samples of each group.

Click the next icon.



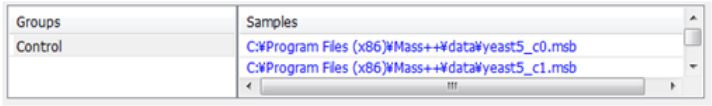
[Select Files] is opened.

Select "yeast5_c0.msb", "yeast5_c0.msb", and "yeast5_c0.msb" that come with Mass++.



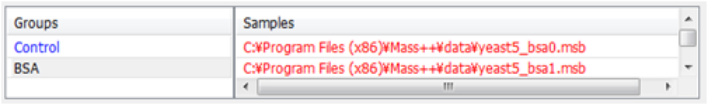
Click [Open].

Samples are registered.



Repeat these steps as many times as necessary.

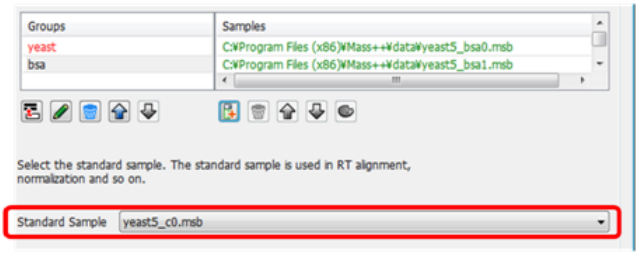
Register "yeast5_bsa0.msb", "yeast5_bsa1.msb" and "yeast5_bsa2.msb" to "bsa" group in a similar way.



Set color red.

After registering data, register standard sample which is used at RT alignment in [Standard Sample].

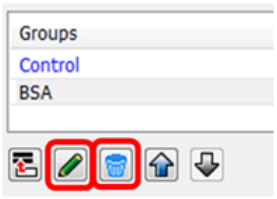
Here, select "yeast5_c0.msb".



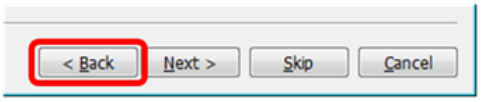
Go to next step.

Click [Next].

Note: If you want to change group name, click pencil mark icon. [Edit Group Name] is opened. If you want to delete registered groups or samples, click trash box mark icon.

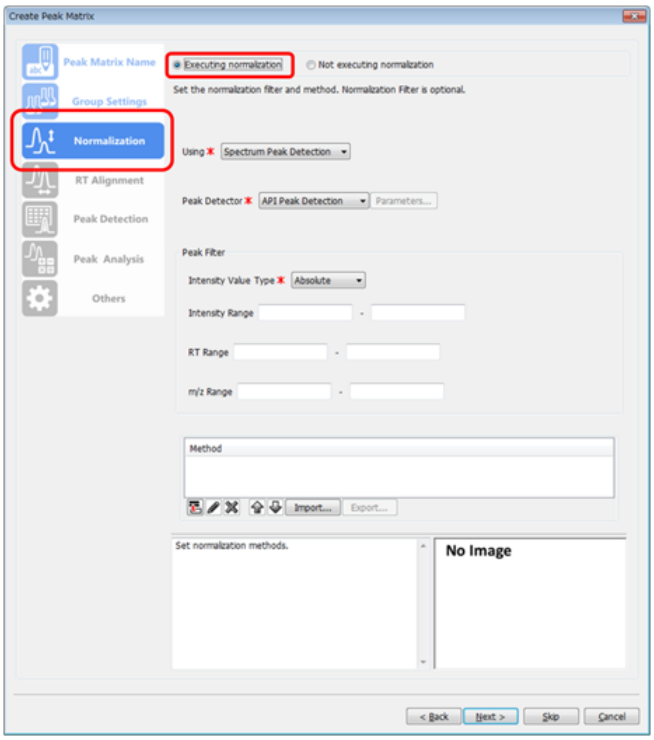


Note: If you want to go back to the previous setting step (ex. [Peak Matrix Name]), click [Back].



2.1.3. Normalization

[Normalization] is displayed.



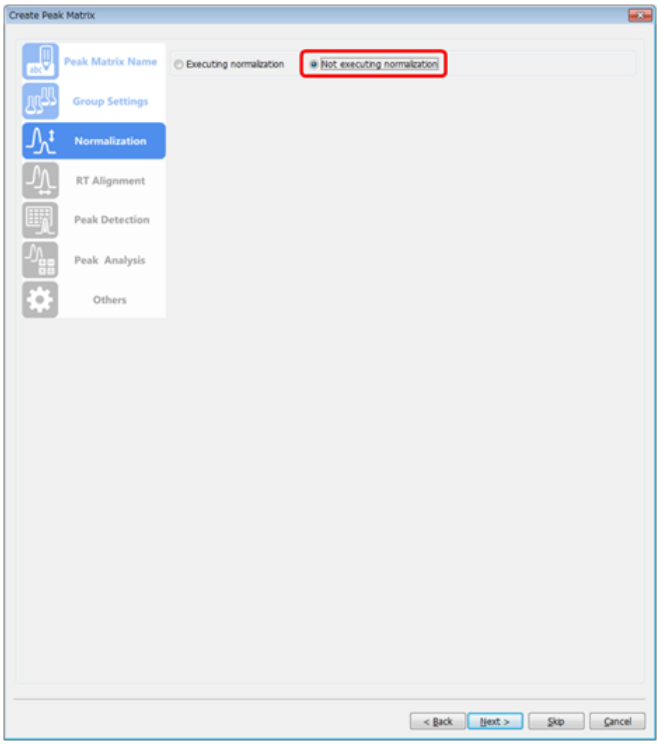
In this panel, set parameters for peak intensity normalization before peak detection.

If performing normalization is required, select [Executing normalization].

In this guide, normalization is not performed under the assumption that "reproducibility of the peak intensities is enough".

Select [Not executing normalization].

Controls on the panel disappear.



Go to next step.

Click [Next].

2.1.4. RT Alignment

[RT Alignment] is displayed.

reproducibility of LC is sometimes inferior to that of MS.

Here, Align RT (Retention Time) that is shifted a little for each data.

Mass++ aligns RT using similarities of TIC waveform among data.

For LC-MALDI data, scale TIC waveform in [Scaling method for TIC pattern] if necessary.

In this guide, RT alignment is not performed under the assumption that "reproducibility of RT direction is enough".

Click [Not Executing RT Alignment].

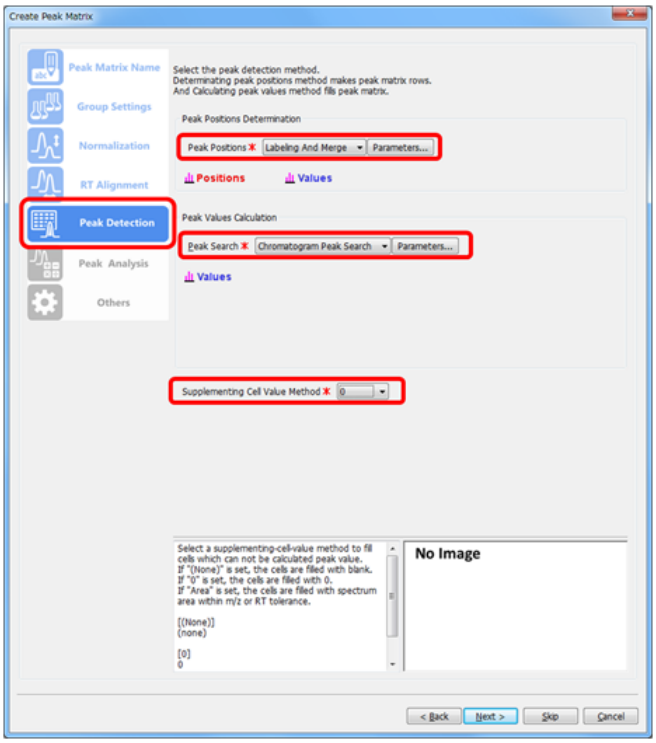
Note: In the case performing RT alignment is required, the parameters of RT alignment is adjusted beforehand in [RT Align] dialog. You can open [RT Align] dialog by clicking [Advanced Data Manipulation] - [RT Align] from the [Processing] menu.

Go to next step.

Click [Next].

2.1.5. Peak Detection

[Peak Detection] is displayed.



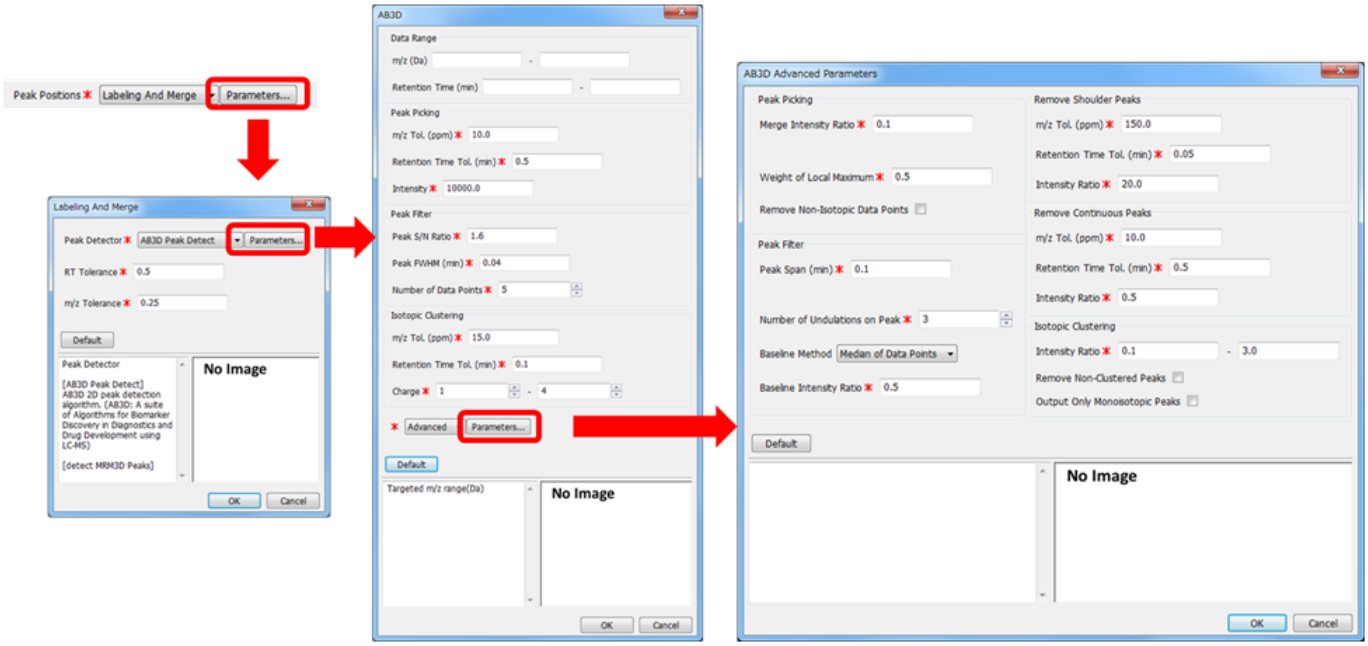
In this panel, set peak detection function.

In [Peak Positions Determination], detect peak positions.

In [Peak Values Determination], peak values (intensity or area) are calculated at detected peak positions.

In this guide, select [Labeling And Merge] as a peak detection function from [Peak Positions Determination].

Set [Labeling And Merge] parameters as follows.

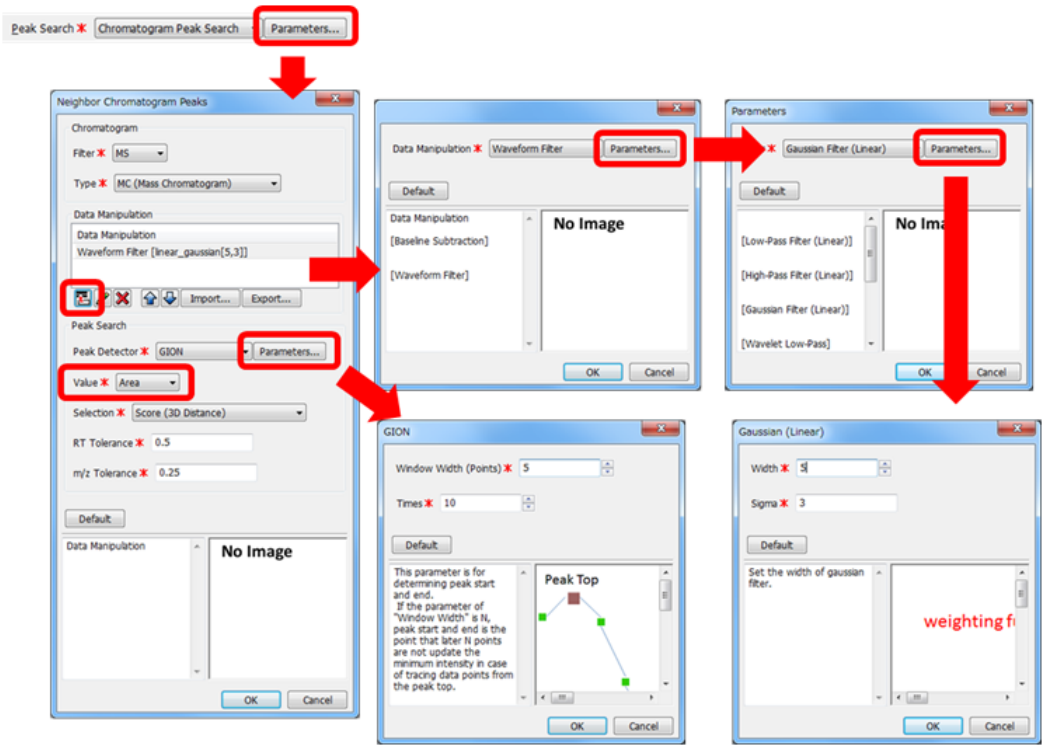


In Mass++, you can select peak type (spectrum peak or chromatogram peak) to create peak matrix,

In this guide, create peak matrix of chromatogram peaks.

Select [Chromatogram Peak Search] from [Peak Search] in [Peak Detection] panel.

Set parameters in [Chromatogram Peak Search] as follows.



Here, "Chromatogram Peak" means peaks of MC(Mass Chromatogram) derived from m/z range set at [RT Tolerance].

Mass++ can create peak matrix of peak intensities and that of peak areas.

Here, peak matrix of peak area is created as an example.

Select [Area] from [Peak Value].

If you select [0] from [Supplementing Cell Value Method] in the [Peak Detection] panel, the blank cell is supplemented with "0" to blank cells where no peaks are detected corresponding RT and m/z.

If you select [Area], blank cells are supplemented using peak values calculated from spectrum directly.

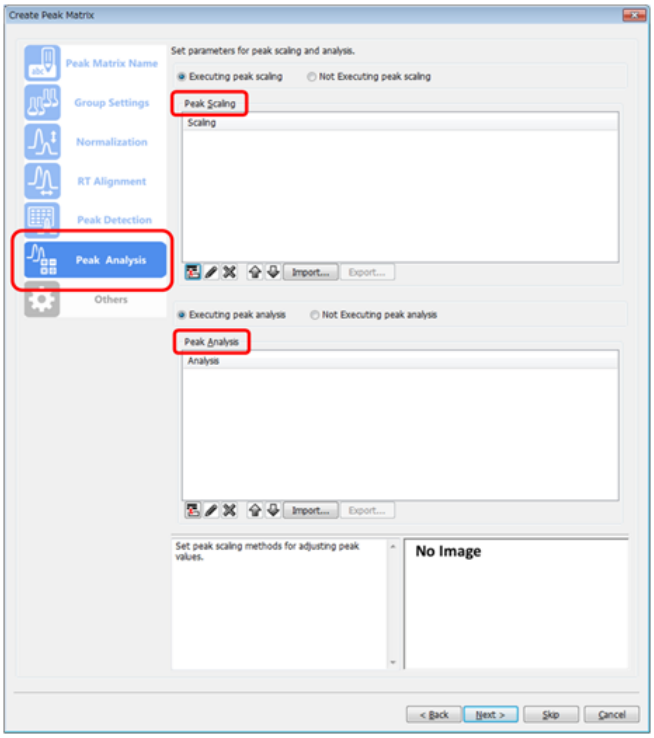
Here, select "0" to confirm that the blank cells are supplemented.

Go to next step.

Click [Next].

2.1.6. Peak Analysis (1) Peak Scaling

[Peak Analysis] is displayed.



In [Peak Scaling], set parameters of scaling for peak values in peak matrix.

In this guide, create peak matrix without scaling.

Select [Not Executing peak scaling].

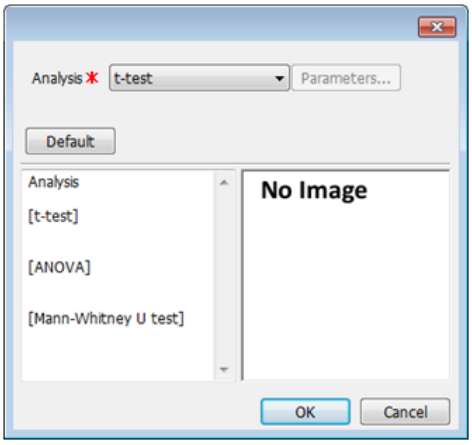
2.1.7. Peak Analysis (2) Statistical Analysis

In [Peak Analysis], set parameters for statistical test for between groups in peak matrix.

Here, perform "Welch's t test".

Click the following icon.

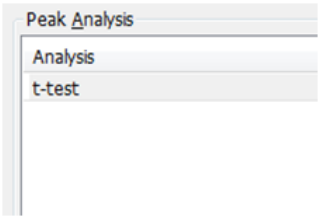
Statistical test selection dialog is opened.



Select [t-test].

Click [OK].

[t-test] is added to [Analysis] list.

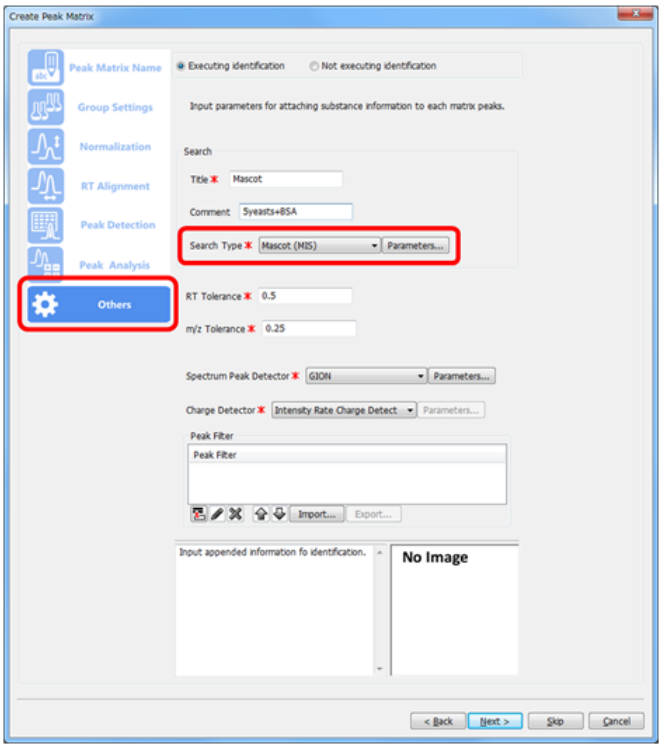


Go to next step.

Click [Next].

2.1.8. Others (1) Identification

[Others] is displayed.



In this panel, set parameters for database search.

Here, use Mascot MS/MS Ions Search.

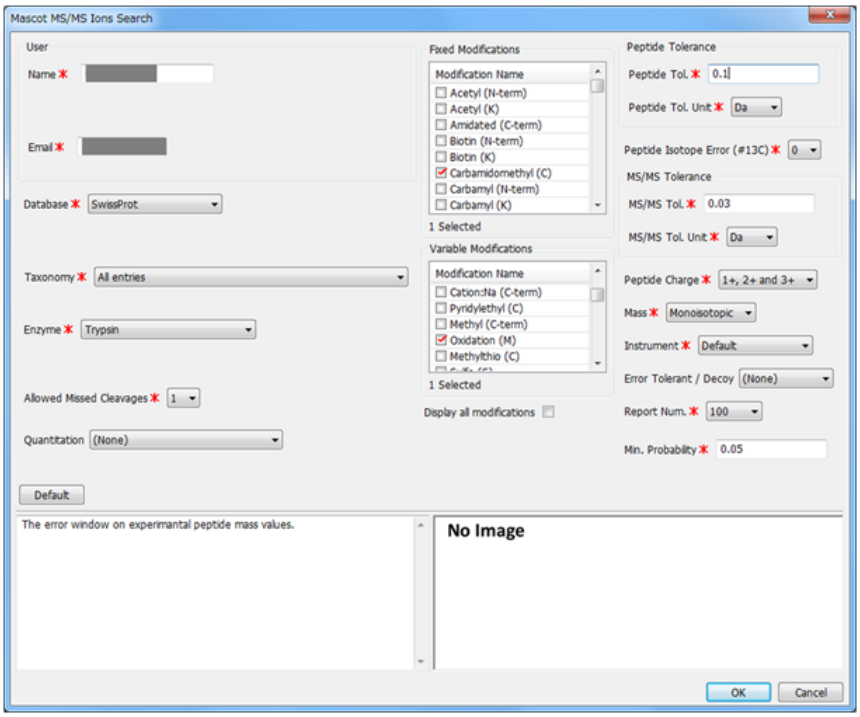
Select [Mascot (MIS)] from [Search Type].

Click [Parameters].

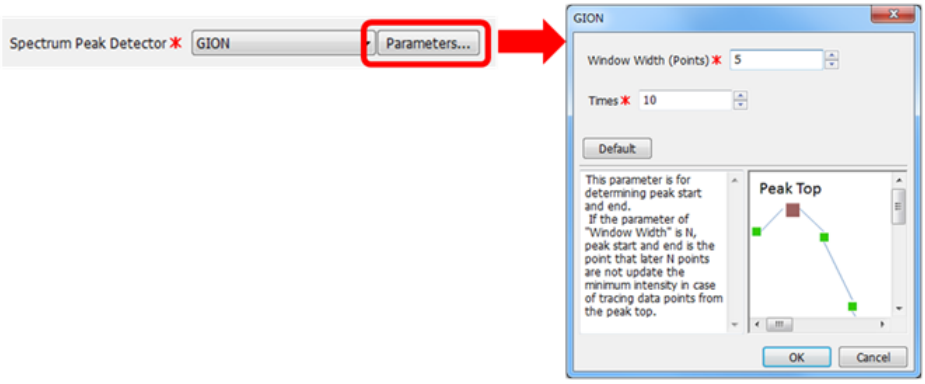
[Mascot MS/MS Ions Search] is opened.

Refer to website of Matrix Science (http://www.matrixscience.com/) for details of each parameter.

In this guide, search using following parameters.



In this guide, set parameters for peak detection function as follows.



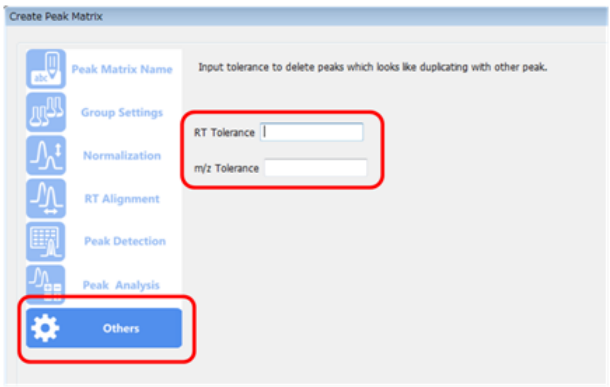
Set other parameters shown in [Others] panel figure.

Go to next step.

Click [Next].

2.1.9. Others (2) Peak Merge

Peak merge setting panel is displayed.



In this guide, crate peak matrix without peak merge.

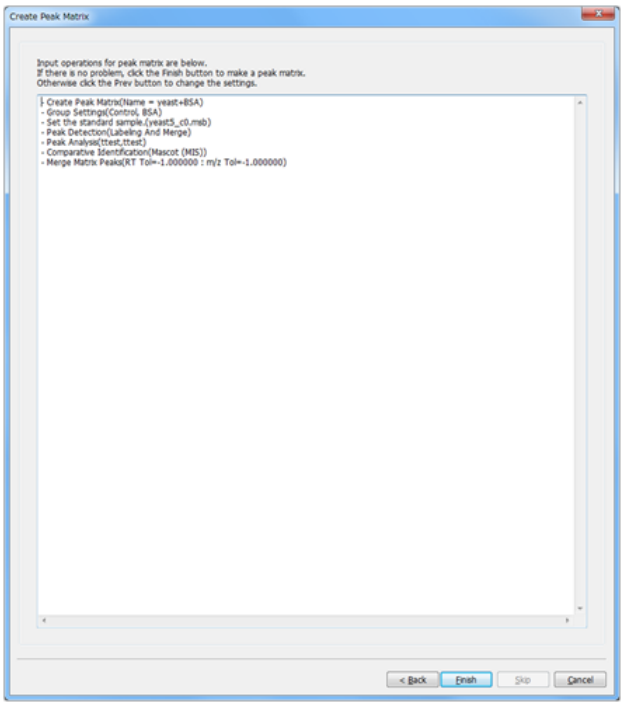
Leave [RT Tolerance] and [m/z Tolerance] blank.

That's the end of the parameter setting for creating the peak matrix.

Click [Next].

2.1.10. Setting Parameters List

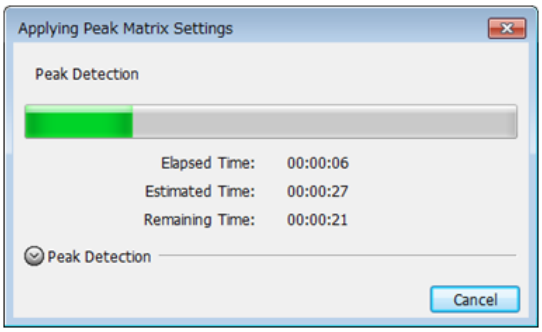
Setting parameter list is displayed.



If you want to change parameters, click [Back] and go back to the objective panel.

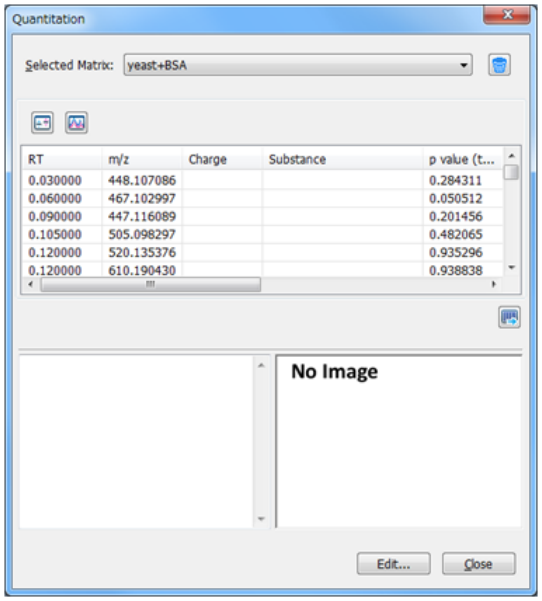
After parameter settings are completed, click [Finish].

Creating peak matrix is started.



2.1.11. Peak Matrix (1) Created Peak Matrix

Creating peak matrix is completed.



Peak positions (m/z), identification results (Substance), statistical test results (p value) and peak values are listed.

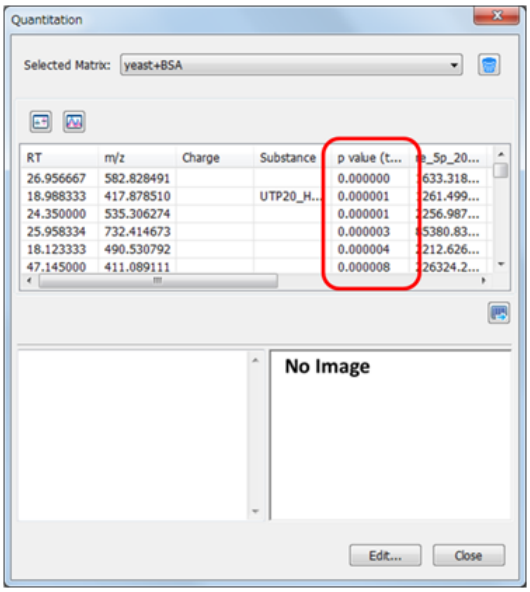
2.1.12. Peak Matrix (2) Sort

After creating peak matrix is completed, peak matrix is sorted in ascending sequence.

Here, re-sort in p value of U test.

Click [p value (u-Test)] header of peak matrix.

The peak matrix is sorted in p value of U test.



You can sort by keys in a similar way.

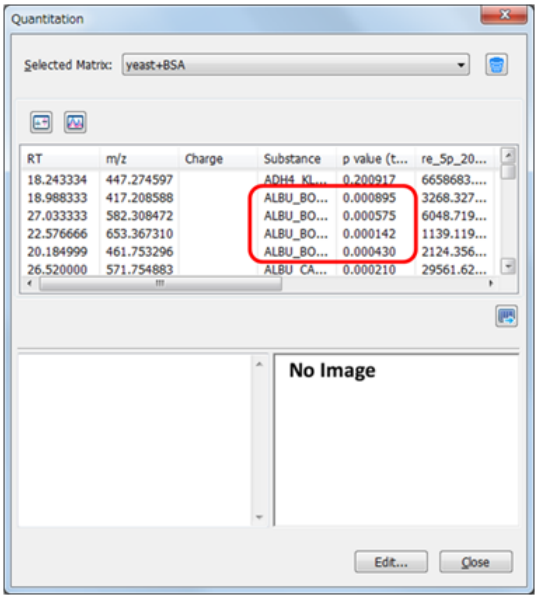
Note: If you click same header once again, the column is sorted in descending order.

2.1.13. Peak Matrix (3) Confirming Identificatoin Results (i)

Identification results of each peak is listed in [Substance].

Blank cell means that no compounds are identified at the corresponding peak.

Confirm that peptides that derived from BSA are identified with specified p value.



2.1.14. Peak Matrix (4) Confirming Identificatoin Results (ii)

You can confirm the details of database search results in [Search Engine Results] dialog.

Refer to "Identification" section for their details.

2.1.15. Peak Matrix (5) Displaying Peak Profiles and Box-Plot

It is not recommended to evaluate whether the peak is a biomarker or there are expression difference of substance that are derived from peak based only on p value.

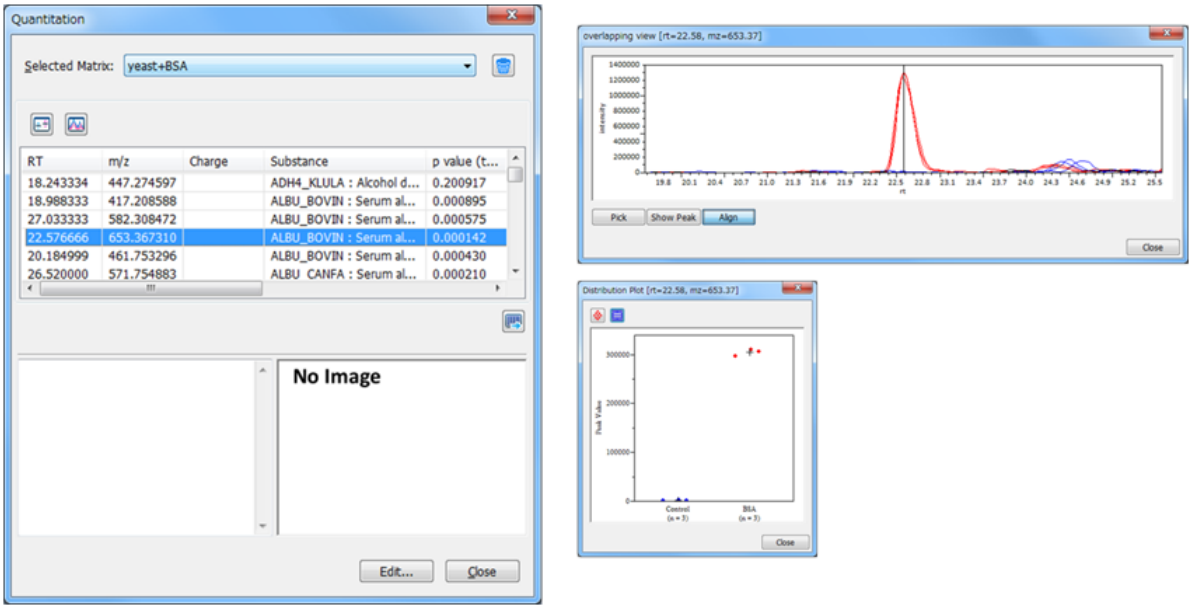
Even though p value is less than specific value, it is possibly that the dispersion among peaks is large, or noise is detected at peak detection.

It is important that checking peak profiles with the results of test.

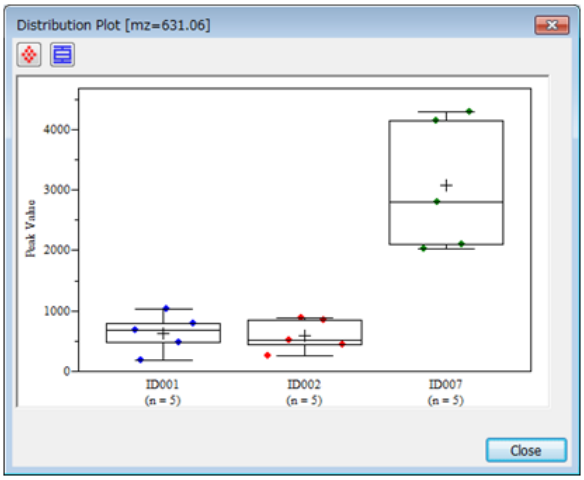
Mass++ can display peak profiles around its peak position in peak matrix and display peak value with box-plot.

Click the objective peak row.

Peak profile and box-plot are displayed.



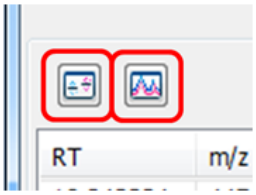
Note: In this guide, box-plot is not displayed, because this differential analysis is performed using 3 vs 3 data. In the case of 5 vs 5 vs 5 differential analysis as an example, box-plot is displayed as follows.



Black line in peak profile dialog is peak position.

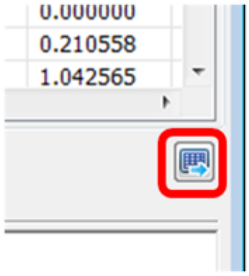
Confirm that peaks are shown with colors that are set at [Group Settings].

Note: If peak profile or box-plot are not displayed, click following icon.



2.1.16. Peak Matrix (6) Exporting Peak Matrix

If you want to export peak matrix as text format, click following icon in the peak matrix dialog.

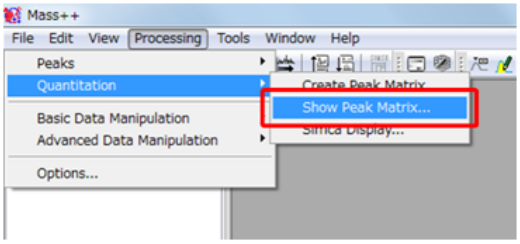


[Save] dialog opens, and input a name to export the peak matrix.

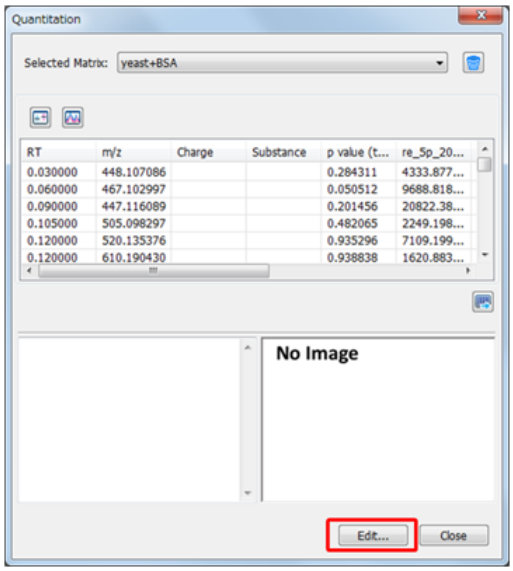
2.1.17. Show Peak Matrix

If you want to display and edit peak matrix that has been already created, do as follows.

Click [Quantitation] - [Show Peak Matrix] from the [Processing] menu.



[Quantiation] is opened.



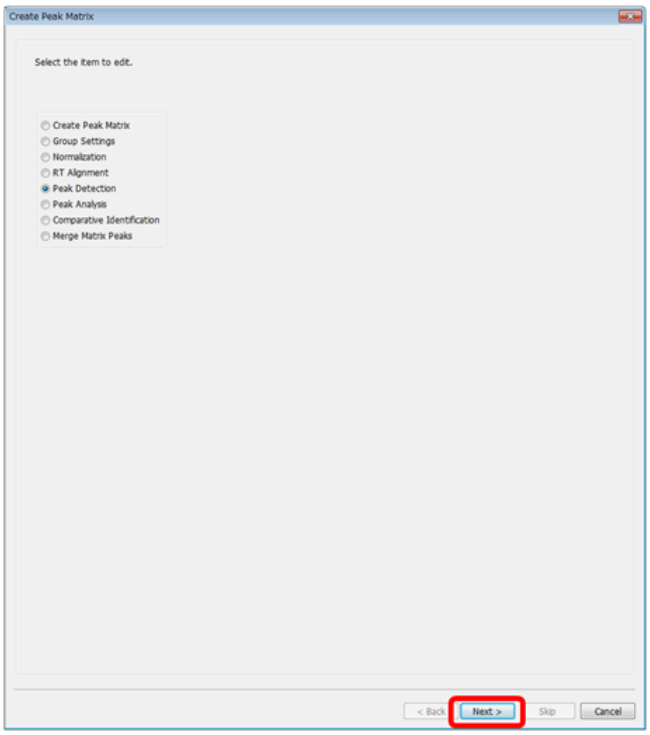
Select objective peak matrix form [Selected matrix].

Confirm that the objective peak matrix is displayed.

Edit displayed peak matrix as follows.

Click [Edit].

[Create Peak Matrix] is opened.



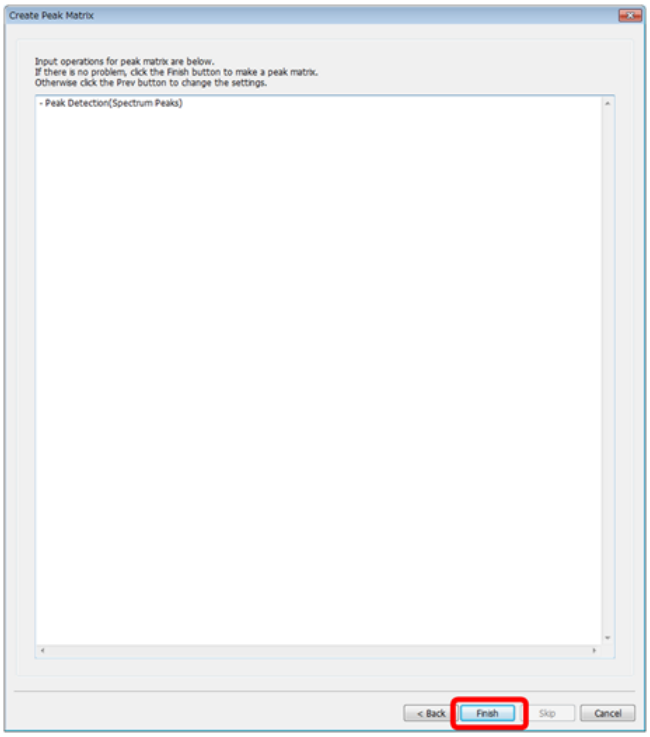
Select the objective control.

Click [Next].

The panel is changed.

After changing settings, click [Next].

Changing of the setting parameter list is displayed.



Click [Finish].

The peak matrix is recalculated.

2.2. SIMCA Display

SIMCA Display plug-in loads and displays three types of plots (Score plot, loading plot, and S plot), which are generated by analyzing “peak matrix” by statistic software SIMCA. Peak matrix is derived from MS data for biomarker discovery. SIMCA makes it possible to confirm the validity of biomarkers with reviewing the original spectrum of biomarker candidates, which are found through the analysis of SIMCA. This guide describes just an example of how to make use of SIMCA, see “SIMCA tutorial” for more information.

2.2.1. Pass Peak Matrix to SIMCA

Import “yeast+BSA.txt” file, having exported in “Create Peal Matrix” section, into Excel. Then, insert “Number” column at the initial column. The “Number” column should be numbered consecutively starting with “1” (as a result, the number of the last row equals to the number of samples).

Insert a new blank row and enter the values bellow to the cells of this line.

・Column A: “0” (This value is just a dummy and any string except for blank is acceptable)

・Column F: “y-val” (title stands for being objective variable)

・Column G - I: “0”

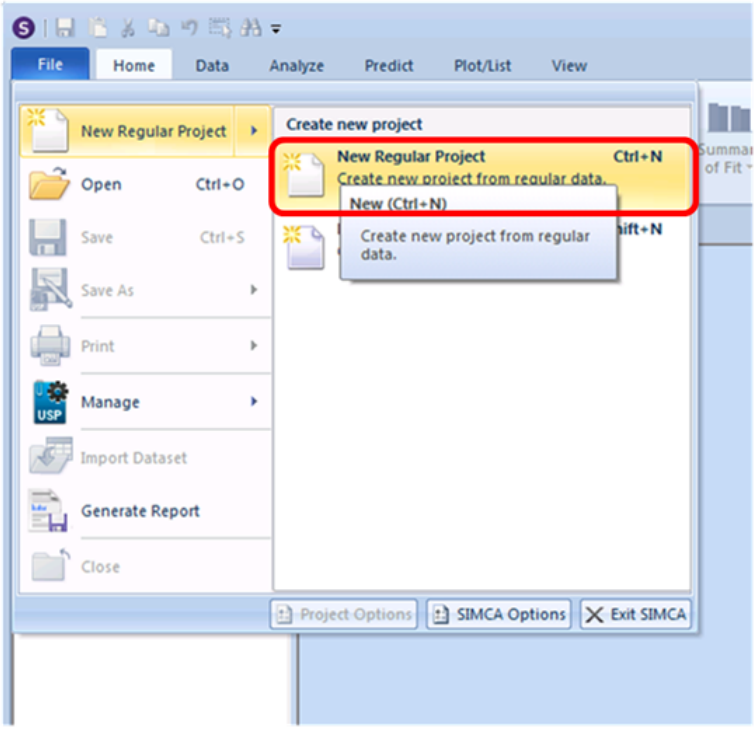
・Column J - L: “50” (contents of BSA)

yeast+BSA

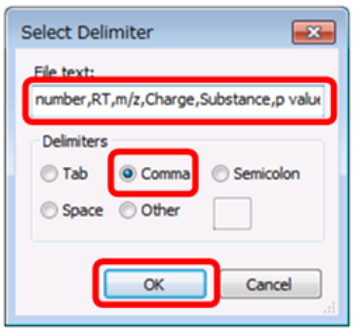
	A	B	C	D	E	F	G	H	I	J	K	L
1	number	RT	m/z	Charge	Substance	p value (t-	re_5p_200fi	re_5p_200fi	re_5p_200fi	5p_200fmo	5p_200fmo	5p_200fmol
2	0					y-val	0	0	0	50	50	50
3	1	0.03	448.1071			0.284311	4333.878	5443.824	7121.269	974.163	5930.426	3642.542
4	2	0.06	467.103			0.050512	9688.819	19310.07	22538.81	438.54	682.0787	2211.957
5	3	0.09	447.1161			0.201456	20822.39	60752.64	25254.99	17436.8	11162.99	8115.728
6	4	0.105	505.0983			0.482065	2249.198	1986.093	6804.04	7188.467	1824.193	7736.091
7	5	0.12	520.1354			0.935296	7109.199	11473.66	9799.122	3191.118	7113.829	19450.31
8	6	0.12	610.1904			0.938838	1620.883	3338.569	483.3511	1118.219	3167.845	886.4847
9	7	0.15	432.0914			0.887554	6138.941	1721.496	5573.773	1324.721	3911.425	9471.86
10	8	0.165	468.0956			0.276394	26892.21	2623.309	8187.103	1417.569	405.4763	3305.308
11	9	0.165	503.1002			0.260446	11642.61	36220.1	11890.12	3517.43	4934.131	13675.38
12	10	0.195	538.1711			0.854414	11609.47	7226.695	1826.942	13067.09	5432.136	4467.726
13	11	0.252366	519.7861	1		0.886468	292.7503	860.8229	0	497.8872	399.8185	379.0732
14	12	0.255	415.0301			0.388103	5338.359	980.6835	3476.376	11984.34	5052.619	2362.547
15	13	0.255	536.1736			0.465779	51969.5	19520.61	5409.316	23943.85	2484.313	12261.67
16	14	0.285	504.1093			0.291089	9835.941	10921.34	2875.269	1581.241	8535.655	555.1369
17	15	0.285	521.14			0.210638	19498.58	10329.72	3274.463	3574.897	2832.154	1246.723
18	16	0.33	519.1318			0.545923	107286.2	6224.151	3435.772	16752.59	2044.66	23877.03

After completing these modifications, save tables as a csv file.

Click [New Regular Project] from the [File] menu on SIMCA. Then select “yeast +BSA.csv” file, having been saved at the previous step, from the file open dialog.

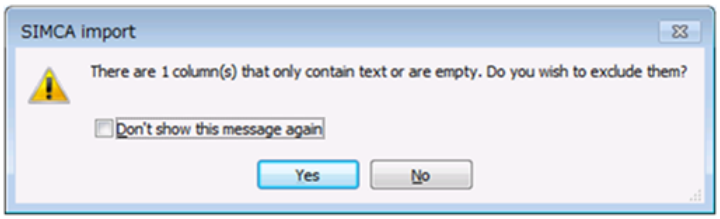


[Select Delimiter] is opened. After confirming that [Comma] is being selected at [Delimiter] and that the first line of the selected file is being displayed at [File Text:], Click [OK].

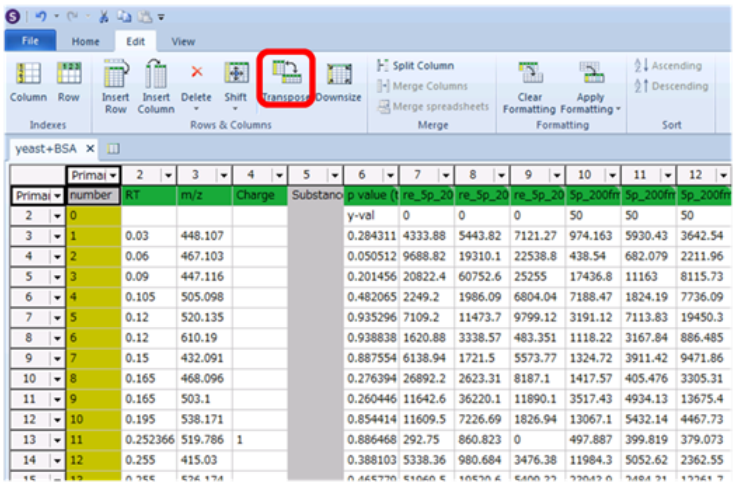


Note: You cannot load the file correctly if another software such as Excel is opening “yeast+BSA.csv”.

Click [Yes] with no regarding to the following warning.

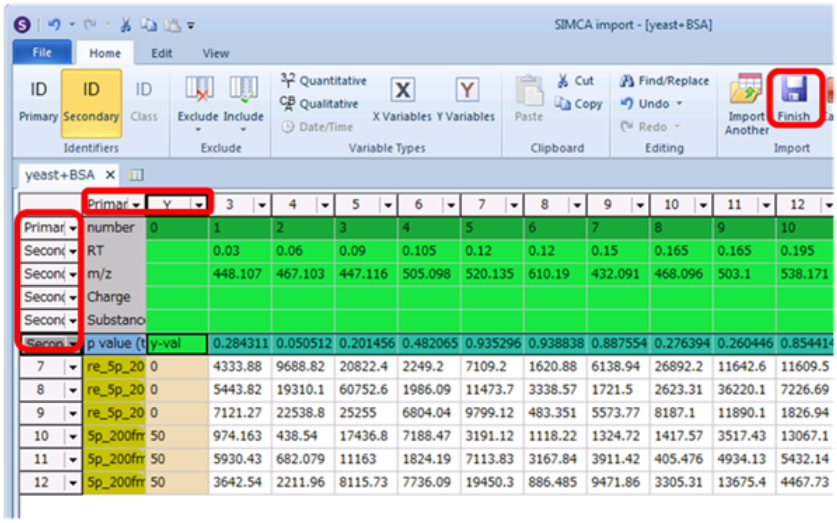


The content of “yeast+BSA.csv” files is shown in the grid of the SIMCA main window. Click [Transpose] from the [Edit] menu.



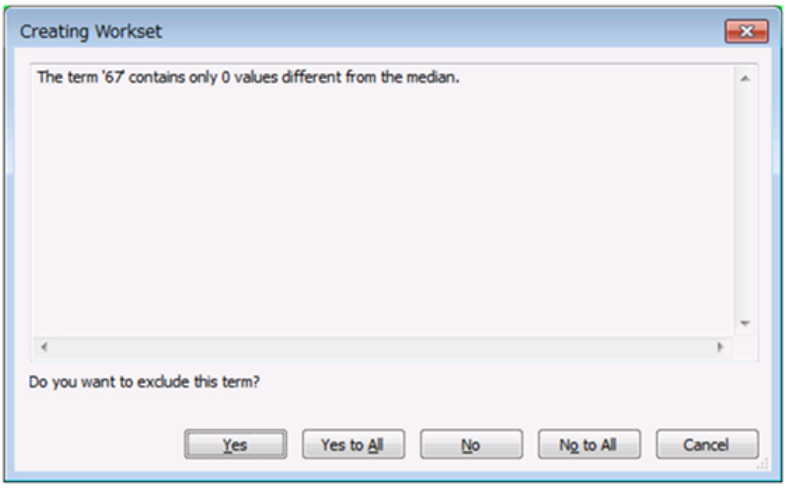
	Primai	2	3	4	5	6	7	8	9	10	11	12
Primai	number	RT	m/z	Charge	Substanc	p value (t	re_Sp_20	re_Sp_20	re_Sp_20	Sp_200fn	Sp_200fn	Sp_200fn
2	0					y-val	0	0	0	50	50	50
3	1	0.03	448.107			0.284311	4333.88	5443.82	7121.27	974.163	5930.43	3642.54
4	2	0.06	467.103			0.050512	9688.82	19310.1	22538.8	438.54	682.079	2211.96
5	3	0.09	447.116			0.201456	20822.4	60752.6	25255	17436.8	11163	8115.73
6	4	0.105	505.098			0.482065	2249.2	1986.09	6804.04	7188.47	1824.19	7736.09
7	5	0.12	520.135			0.935296	7109.2	11473.7	9799.12	3191.12	7113.83	19450.3
8	6	0.12	610.19			0.938838	1620.88	3338.57	483.351	1118.22	3167.84	886.485
9	7	0.15	432.091			0.887554	6138.94	1721.5	5573.77	1324.72	3911.42	9471.86
10	8	0.165	468.096			0.276394	26892.2	2623.31	8187.1	1417.57	405.476	3305.31
11	9	0.165	503.1			0.260446	11642.6	36220.1	11890.1	3517.43	4934.13	13675.4
12	10	0.195	538.171			0.854414	11609.5	7226.69	1826.94	13067.1	5432.14	4467.73
13	11	0.252366	519.786	1		0.886468	292.75	860.823	0	497.887	399.819	379.073
14	12	0.255	415.03			0.388103	5338.36	980.684	3476.38	11984.3	5052.62	2362.55

Then, select [Primary ID] and [Y] for the headers of the first and second columns respectively. Likewise select [Primary ID] for the header of the first row and [Secondary ID] for those of the second - sixth rows.

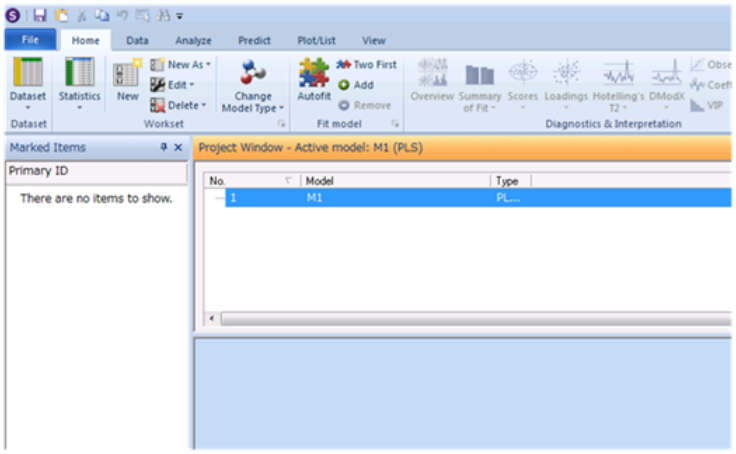


Click [Finish] from the [Home] menu. The file save dialog is opened. Save SIMCA project files (in this document “yeast+BSA.usp”).

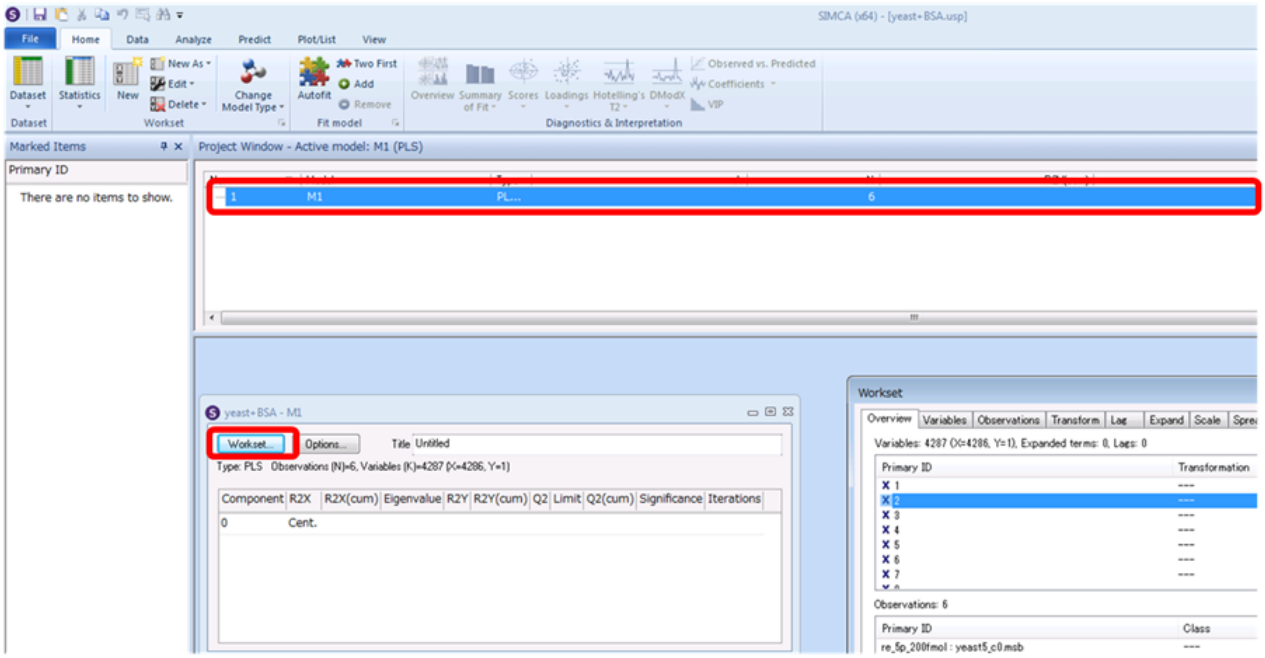
Click [Yes to All] with no regarding to the following warning.



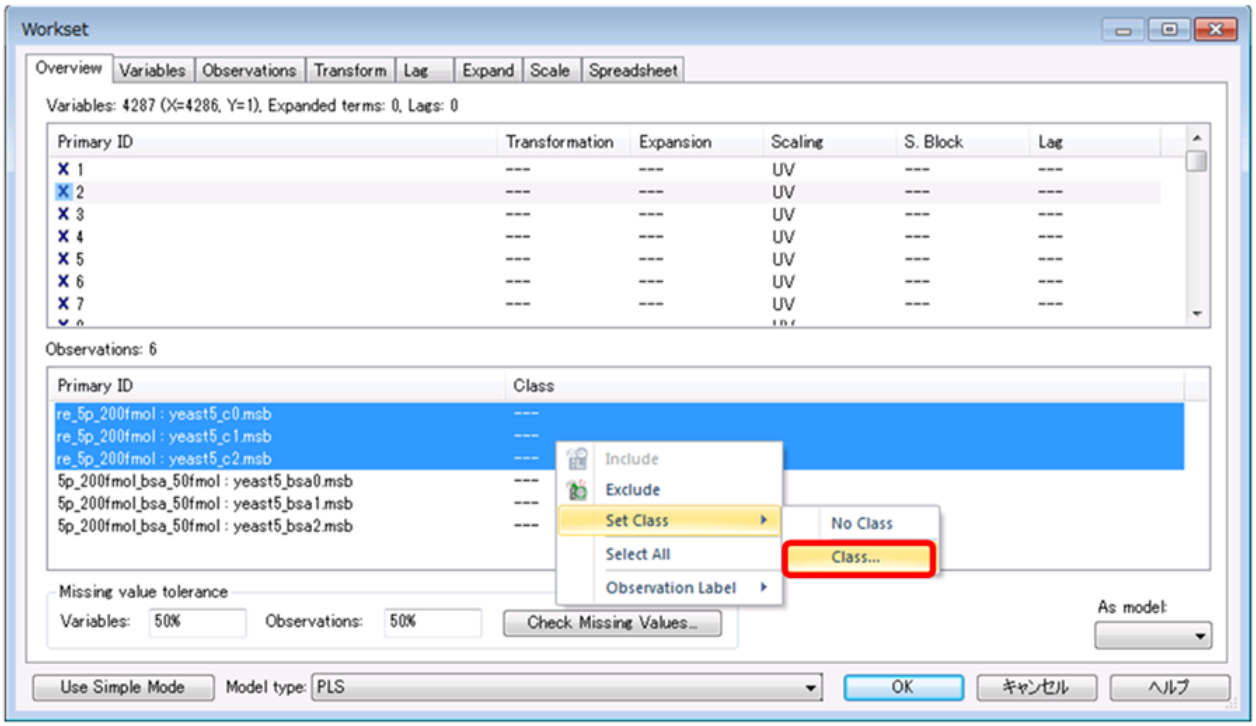
[Project Window] is opened.



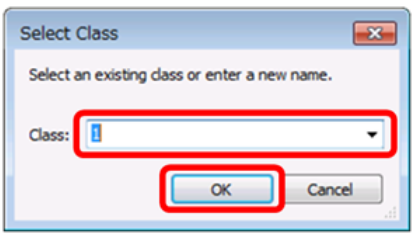
Double click the first row displayed in the [Project Window] list. [yeast+BSA] is opened. Click the [Workset] button. The [Workset] dialog is opened. Confirm that [Overview] tab is being selected.



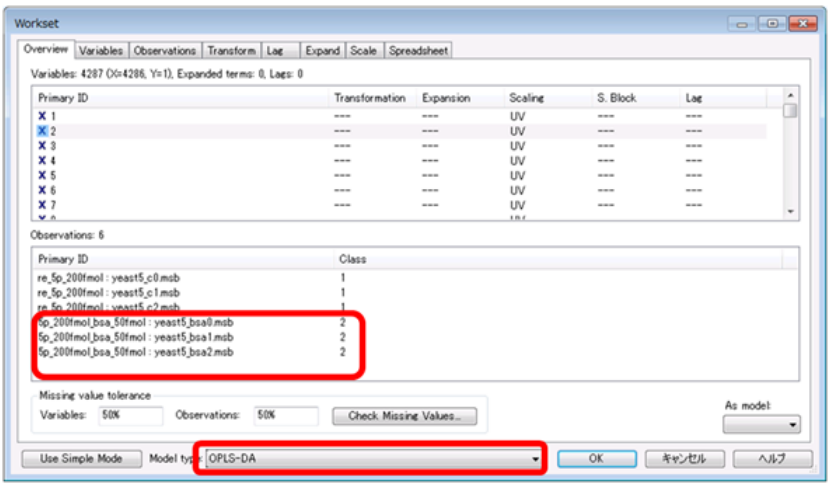
From the [Observations] lines in [Workset], select lines containing “yeast5_c0.msb” - “yeast5_c2.msb”, and click [Set Class] - [Class] from the right click menu.



[Select Class] is opened. Select “1” from the [Class] combobox and click [OK].



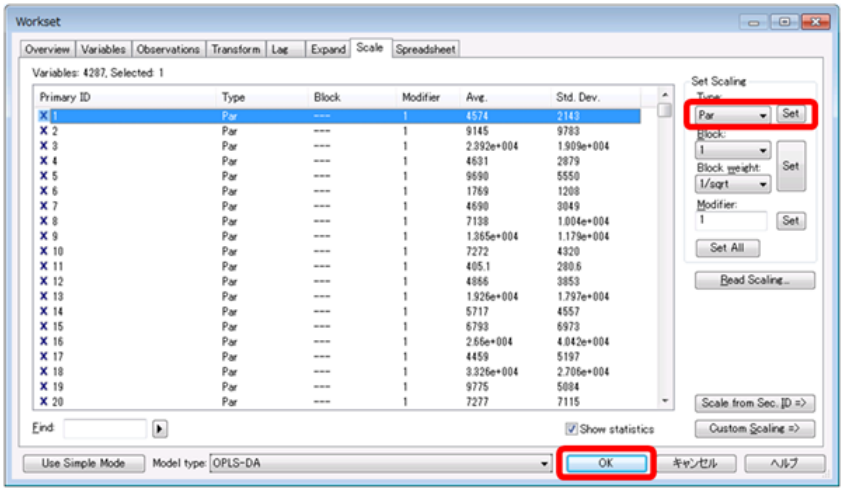
Consequently, [Class] of “yeast5_c0.msb” - “yeast5_c2.msb” are set to “1”. Likewise, set [class] of lines corresponding to “yeast5_bsa0.msb” - “yeast5_bsa2.msb” to “2”.



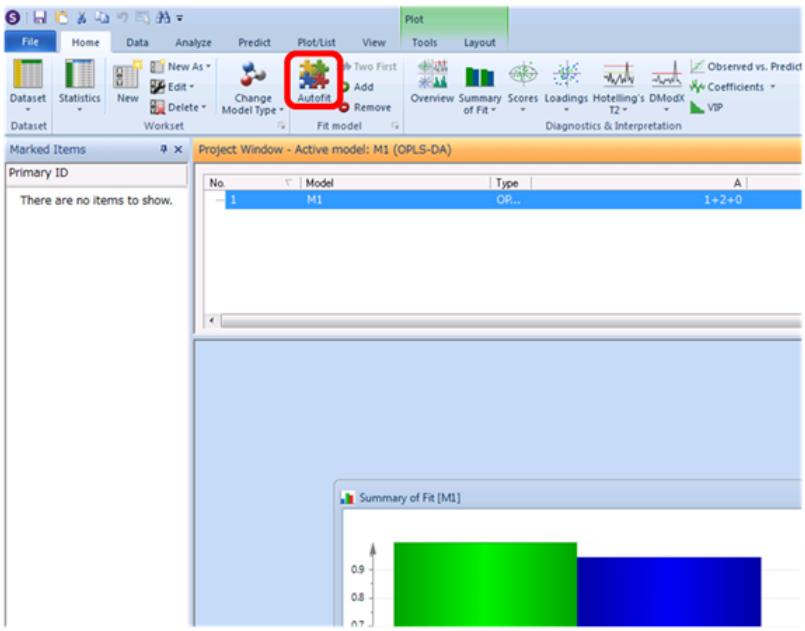
You can select the algorithm of statistical analysis via [Model Type]. See “SIMCA 13 Tutorial” for detail of each algorithm. Here, select OPLS-DA.

Open [Scale] tab. After setting [Type] in the [Set Scaling] region to [Par], select all rows in the [Variables] list. Click [set] at the right side of [Par].

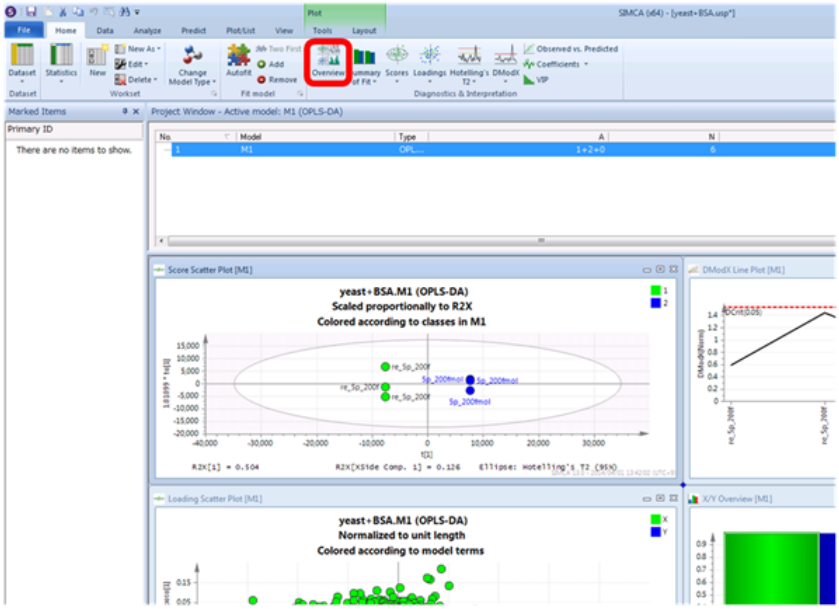
Consequently, [Type] of all lines of the Variable] list are set to [Par]. Click [OK].



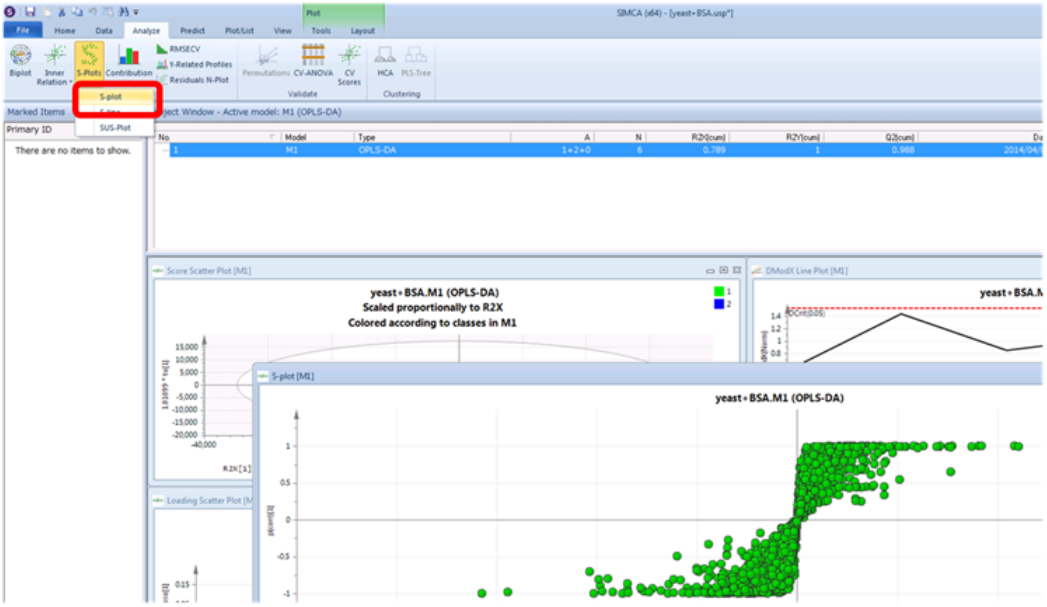
Click [Autofit] from the [Home] menu. [Autofit] selects automatically the required size of dimensions for modeling. In this example, three dimensional modeling is regard as optimal.



Click [Overview] from the [Home] menu. Score plot and loading plot are displayed at the left-top and left-bottom respectively.

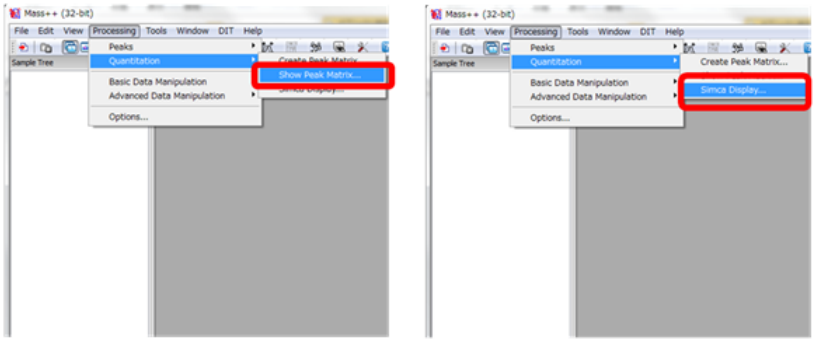


You can display S plot by clicking [S-Plots] - [S-Plot] from the [Analysis] menu.

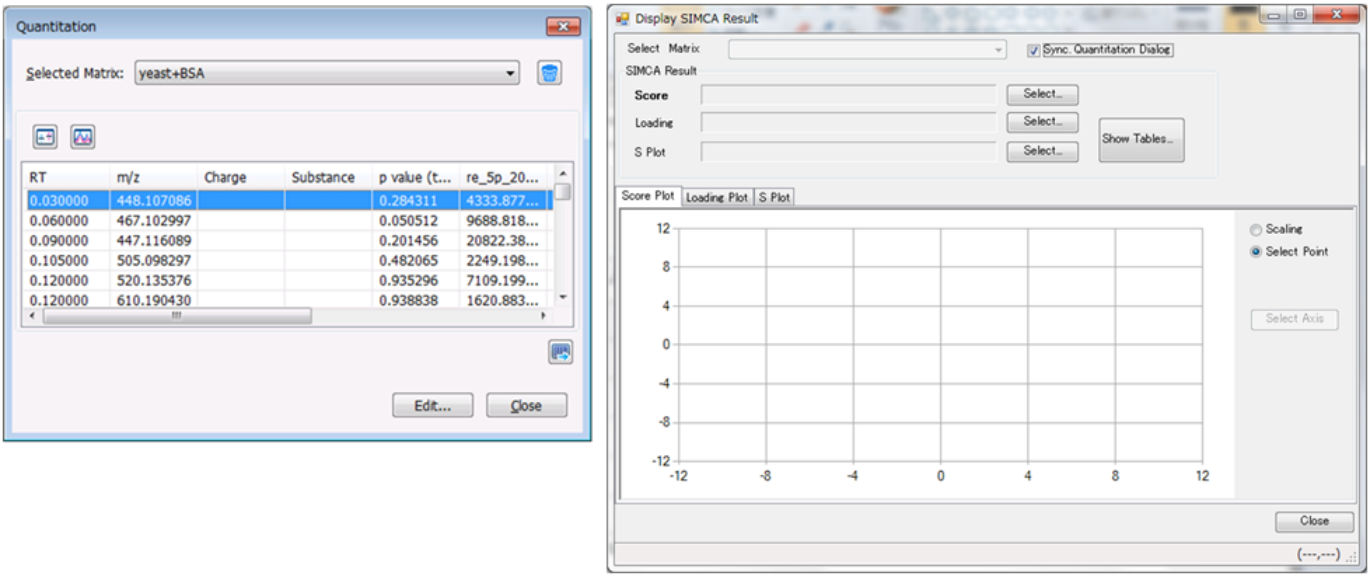


2.2.2. Import the Results of Anaylysis by SIMCA

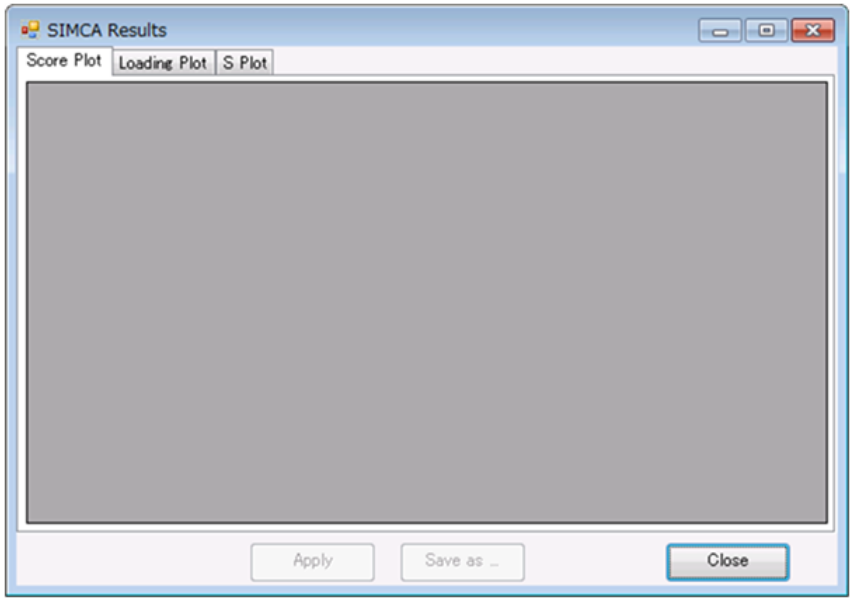
Click [Quantitation] - [Show Peak Matrix] and [SIMCA Display] from the [Processing] menu of Mass++.



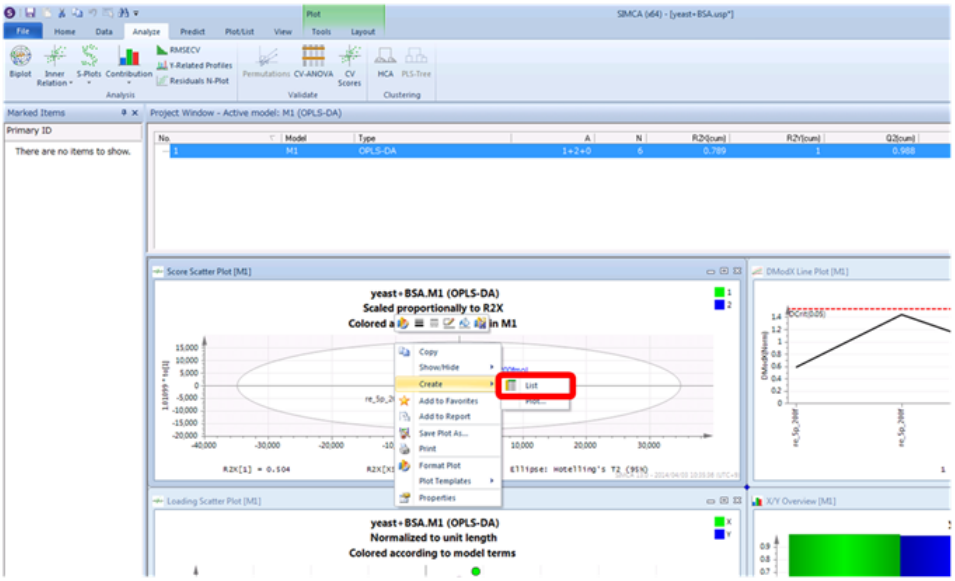
[Quantitation] and [Display SIMCA Result] are opened.



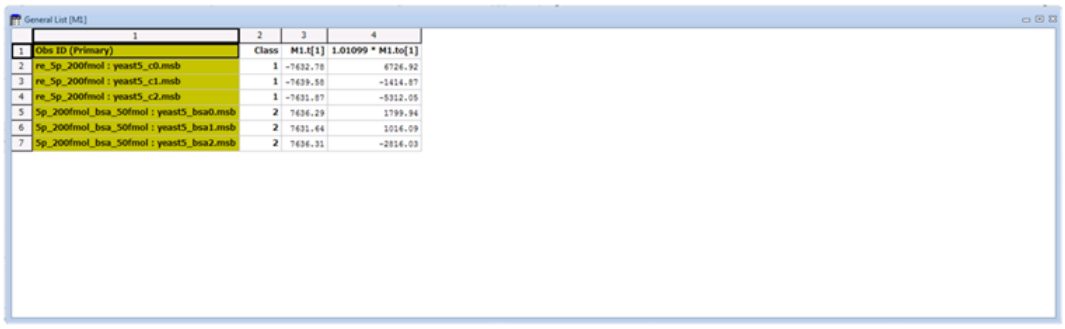
Click [Show Tables]. [SIMCA Results] is opened.



While on SIMCA, keep score plot (or loading plot, S plot) displayed, on which click [Create - List] from the right click menu.

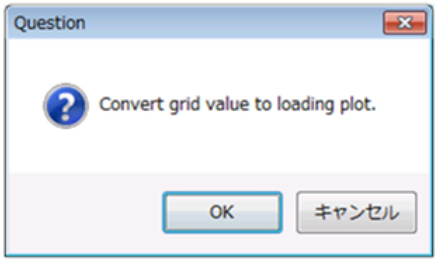


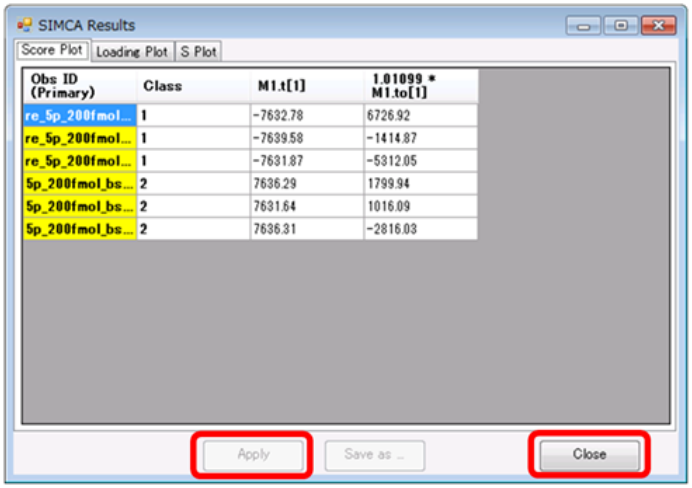
The data of plotted points is displayed as a table. Select whole cells in the table and copy (Ctrl-C) cells to clipboard.



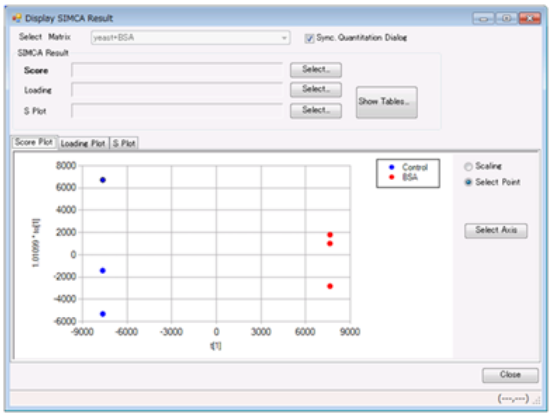
	1	2	3	4
1	Obs ID (Primary)	Class	M1.t[1]	1.01099 * M1.to[1]
2	re_5p_200fmol : yeast5_c0.msb	1	-7632.78	6724.92
3	re_5p_200fmol : yeast5_c1.msb	1	-7639.58	-1414.87
4	re_5p_200fmol : yeast5_c2.msb	1	-7631.87	-9312.05
5	5p_200fmol_bsa_50fmol : yeast5_bsa0.msb	2	7636.29	1799.94
6	5p_200fmol_bsa_50fmol : yeast5_bsa1.msb	2	7631.64	1014.09
7	5p_200fmol_bsa_50fmol : yeast5_bsa2.msb	2	7636.31	-2816.03

Paste (Ctrl-V) the contents of clipboard on the [Score Plot] tab of [SIMCA Results]. The confirming dialog bellow is opened. Click [OK]. Then, the data of score plot appears to the [Score Plot] tab.





Click [Apply] to reflect score plot to [Display SIMCA Result] and click [Close] to close [SIMCA Results].

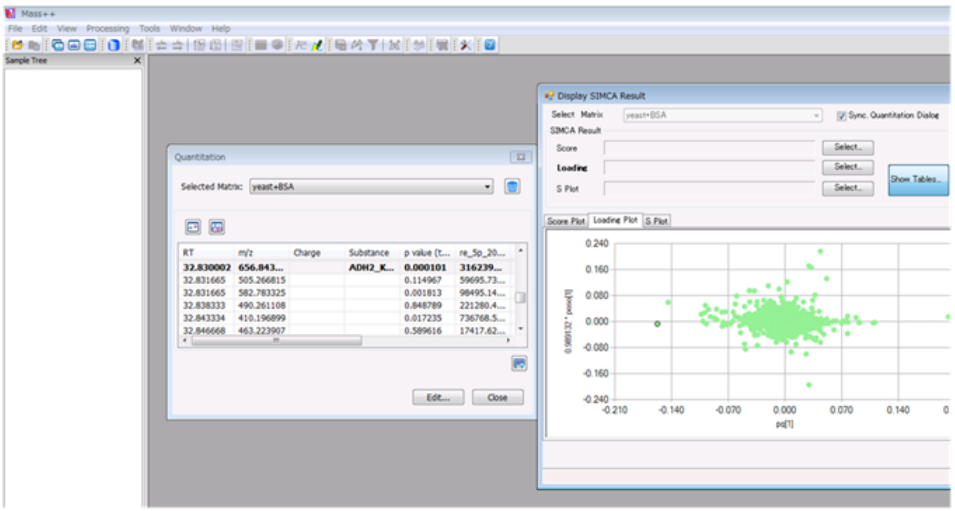


You can reflect loading plot and S plot, which are being displaying to [SIMCA Results] in a similar way.

Note: You can save the data of points plotted on SIMCA as a csv file and reflect the data to SIMCA plug-in by loading the saved file. For this procedure, click [Select] buttons and select csv files from file open dialog.

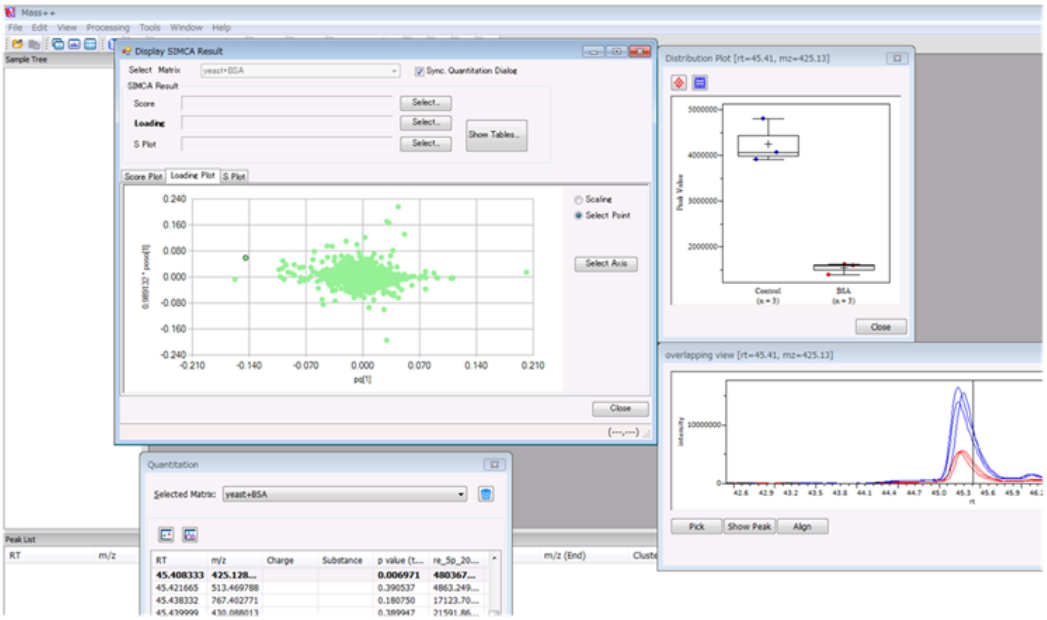
2.2.3. Validation of Biomarker Candidates

When loading plot or S plot is being displayed on [Display SIMCA Result], double click one of plotted points. Then the corresponding row of peak lists shown in [Quantitation] is highlighted.



Click the following icons at the top of [Quantitation]. The box plot of peak value of highlighted peak, MC (Mass Chromatogram) around the corresponding RT (retention time), or MS spectrum around the corresponding m/z. See “Create Peak Matrix” section for more details.

You can easily confirm whether the spectrum around the peaks of biomarker candidates, discovered by SIMCA, have really significant difference or not.



Chapter 3. Data Analysis Tools

Mass++ offers many purpose-specific data processing functions, including identification (data base searching) and, de novo sequencing. This chapter describes how to operate these data processing features.

3.1. Identification

This section describes the steps to conduct database search and identify the peak list of spectra.

Mass++ can perform database search directly using following database search:

・Mascot

・X!Tandem

・MassBank

This section describes the steps for Mascot search (MS/MS Ions Search:MIS).

"Mascot" is a product of MatrixScience Ltd.

In this section, LC-MALDI TOF/TOF MS/MS data of trypsin digested BSA (bovine serum albumin) is used.

You can use other database search in a similar way.

3.1.1. Peparation (1) Setting for Search Engine

Before conducting database search, configure database search setting.

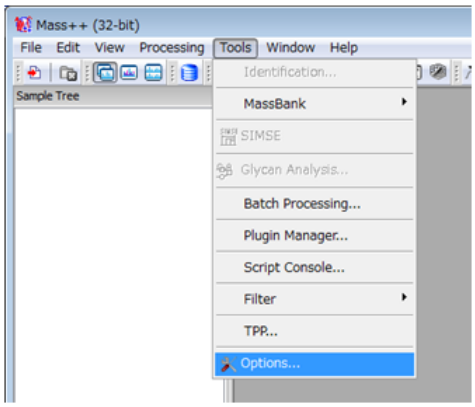
First, configure proxy server setting.

In the following case, proxy server setting is not required.

・Not using proxy server.

・The proxy server setting on IE (Internet Explorer) is already established.

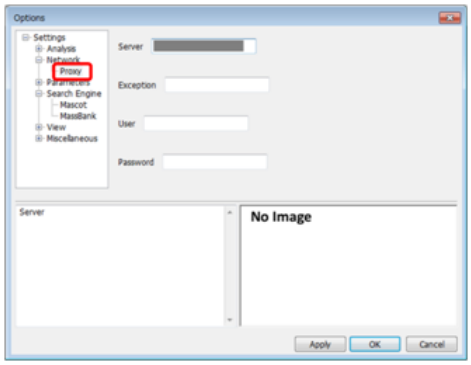
Click [Options] from the [Tools] menu.



[Options] opens.

Click [Network] - [Proxy] from the [Setting] menu.

[Proxy] setting window is displayed.



Enter the server address in [Server].

Enter port number ("1111", for example) after the server address by connecting with ":" as follows:

http://xxx.xxx/:1111

Enter the server address and website address those are accessed without using proxy server.

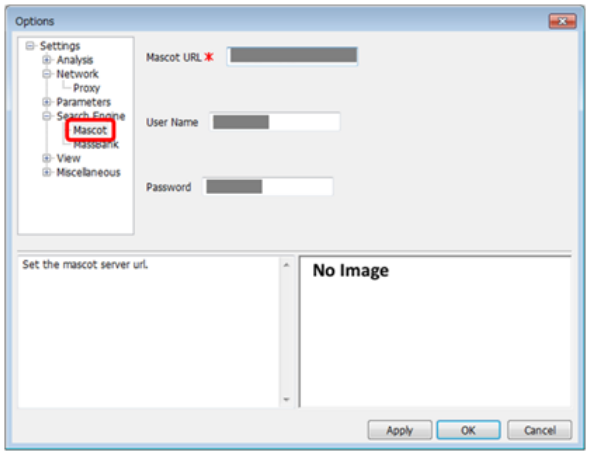
In the case of local Mascot server is installed and you want to access Mascot server without using a proxy server, just enter Mascot server address.

If necessary, enter the user name and the password.

Second, register website address of search engine.

Click [Search Engine] - [Mascot].

[Mascot] setting panel is displayed.



If you want to conduct Mascot search using website of Matrix Science, enter the following address in [Mascot URL].

http://www.matrixscience.com/

In the case of Mascot server is installed, enter Mascot server address.

If necessary, enter the user name and password.

Preparation is completed.

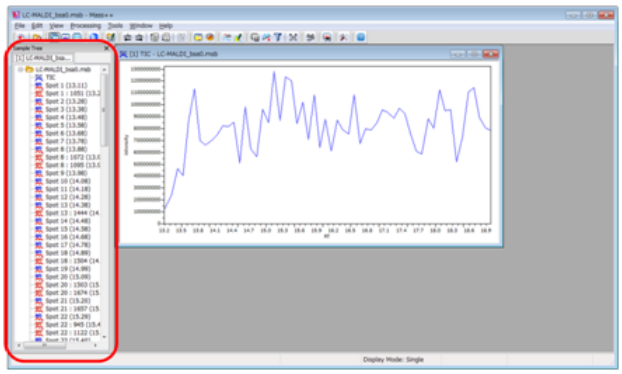
Click [OK] to close [Options].

3.1.2. Peparation (2) Opening a File

Open objective data for database search.

Here, open "LC-MALDI_bsa0.msb"that supplied with Mass++.

A data opened in [Sample Tree] is a target of database search.



Note: If multi samples are opened in [Sample Tree], only active data is the object of database search.

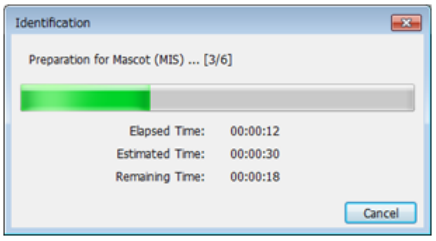
"Active data" is the data whose spectra are listed in [Sample Tree].

3.1.3. Mascot Search

Conduct Mascot search.

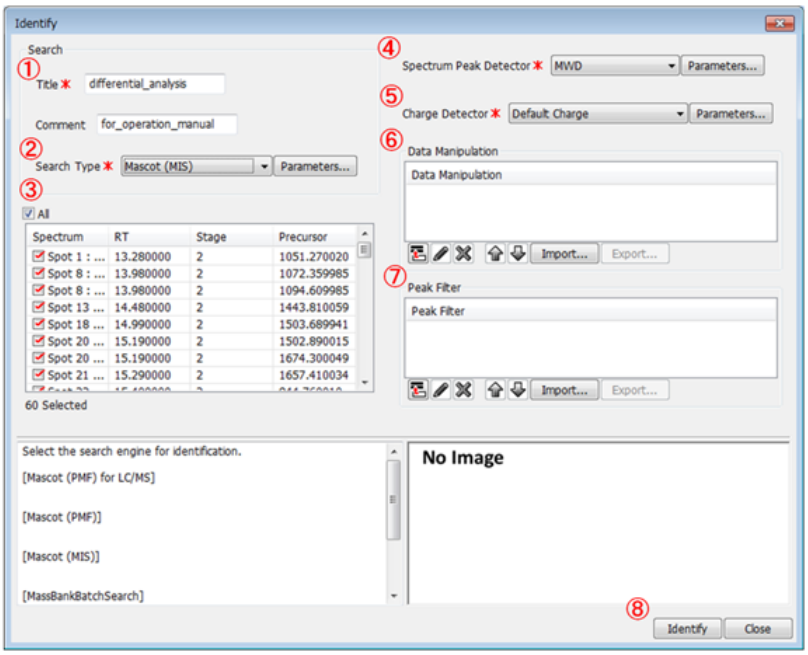
Click [Identification] from the [Tools] menu.

Access to the server or website starts.



Note: If the following error message is displayed, check the settings of proxy server and search engine.

[Identify] is opened.

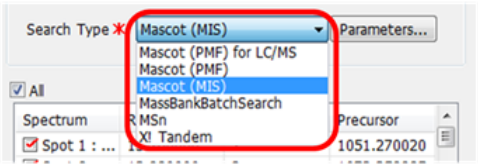


Enter [Title] and [Comment](①)

In this guide, conducting identification for peptides using MS/MS peaks.

Select [Mascot (MIS)] from [Search Type](②)

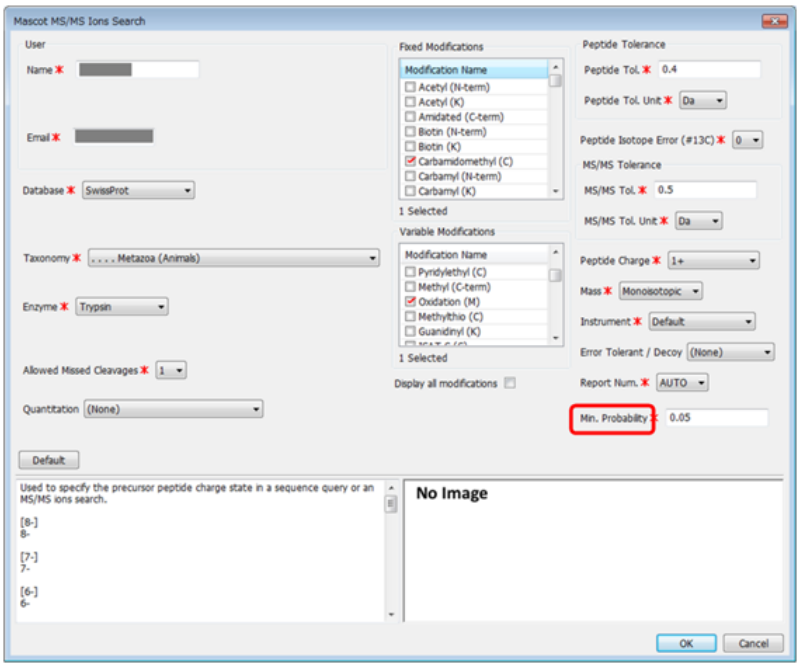
Note: In [Search Type], you can select various database searches (such as MassBankBatchSearch). Refer manuals of each database search engine to check the details of parameters.



3.1.4. Settings for Dababase Search.

Click [Parameters].

[Mascot MS/MS Ions Search] is opened.



In this guide, conduct database search using parameters shown in the following figure.

Set the threshold for identification in [Min. Probability].

After database search is completed, proteins (or peptides) which have probabilities below setting threshold are listed.

Click [OK] to close [Mascot MS/MS Ions Search].

Select target spectra for database search.

The spectra which are subjected to a database search are listed in [Identify] dialog(③).

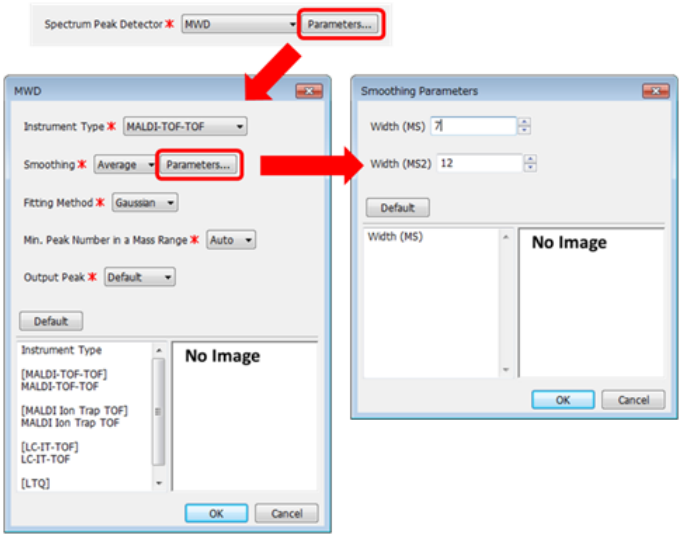
Since [Mascot (MIS)] is selected from [Search Type], [Stage] is set as 2, namely only MS2 spectra are listed.

Select target spectra for database search.

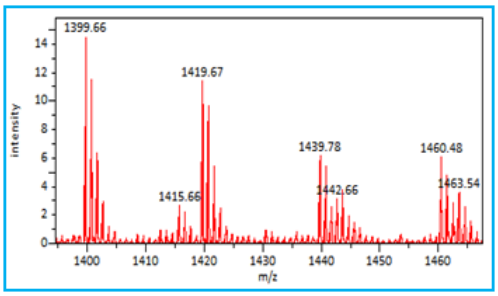
Here, conduct database search using all MS/MS spectra.

Select peak detection method(④).

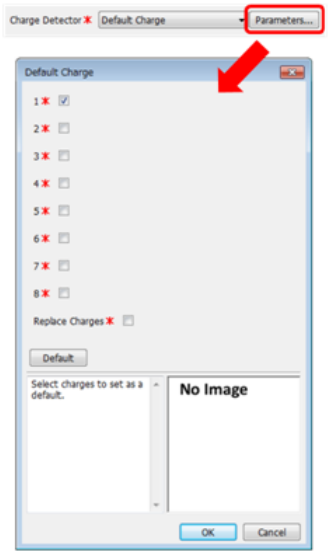
Here, select [MWD] from [Spectrum Peak Detector]



Note: [MWD] is the unique peak detection function in Mass++ to detect monoisotopic peaks.



Set charge detection(⑤) as follows.



In [Data Manipulation], set processing methods (such as smoothing, baseline subtraction, etc) before peak detection (⑥).

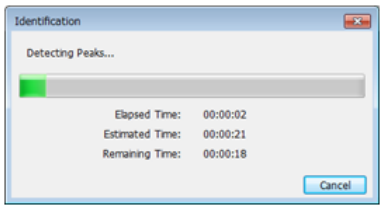
In this guide, no processing is set.

In [Peak Filter], set m/z range (and so on) of target peaks for database search (⑦).

In this guide, no filer is set.

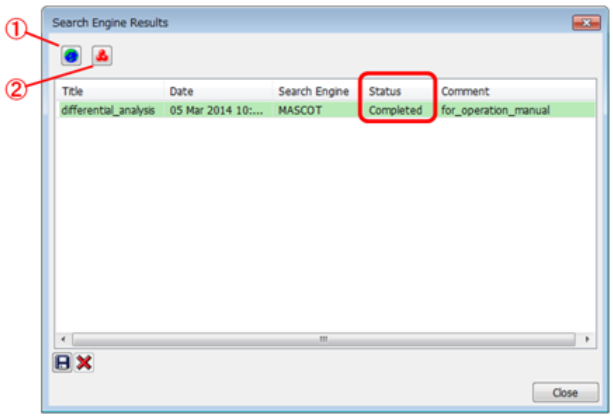
Click [Identify](⑧).

Peak detection and database search start.



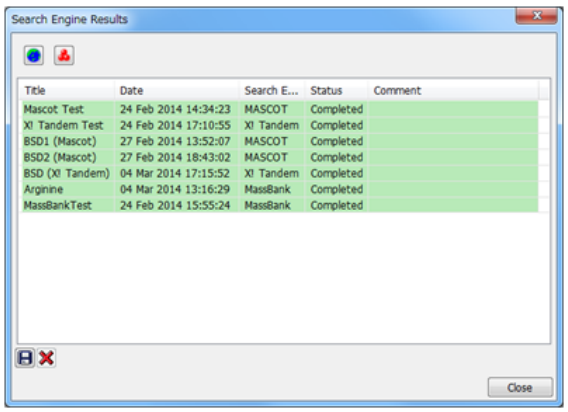
3.1.5. Confirming the Search Result

After database search is completed, [Search Engine Results] is opened automatically.



Confirm that [Status] of the database search is [Completed].

Note: In this dialog, database search that was conducted in the past are listed.



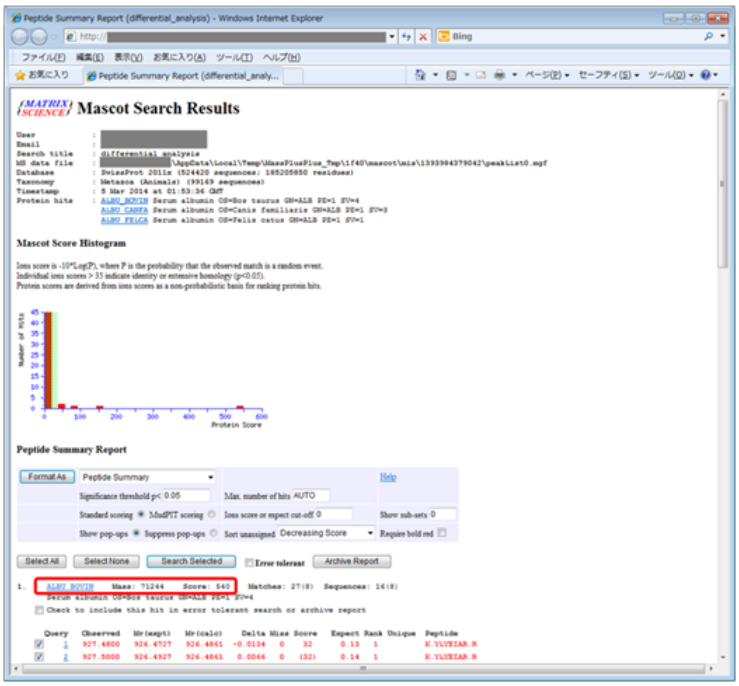
If you want to open this dialog to confirm the past database search results, click [Results] - [Search Engine] from the [View] menu.

Confirm the details of database search result.

Click target row of database search result.

Click ① icon.

The details of database search result are shown with web browser.



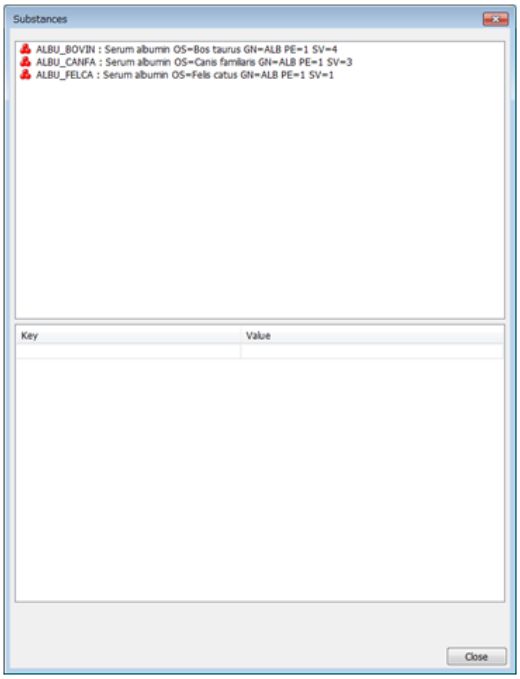
"ALBU_BOVIN" is identified with score 540.

Note: Identification results including score may change by database type or update of the database.

Usually, multi candidates of substances identified as a database search are listed.

Click ② to check candidates-list of substances that is identified with threshold set at [Min. Probability] in the [Mascot MS/MS Ions Search] dialog.

[Substance] is opened.



3.2. SIMSE(de novo sequencing)

This section describes how to analyze an MS/MS spectrum by de novo sequencing using SIMSE.

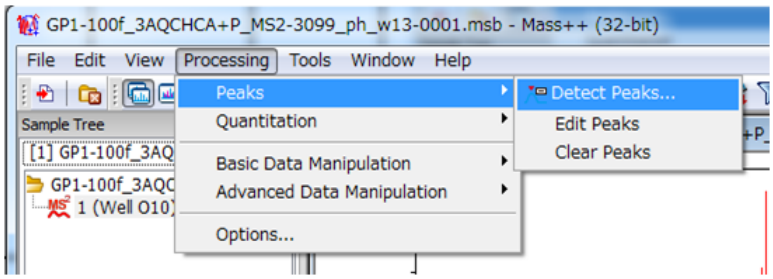
SIMSE enables to analyze peptides, nucleic acids and glycans.

3.2.1. Opening a file

First, open the target MS/MS data. As an example here, that supplied with Mass++, "GP1-100f_3AQCHCA+P_MS2-3099_ph_w13-0001.msb" is used.

3.2.2. Detecting peaks

Click [Peaks]-[Detect Peaks] from the [Processing] menu



[Detect Peaks] is opened.

Select "MWD" peak detection algorithms in this example.

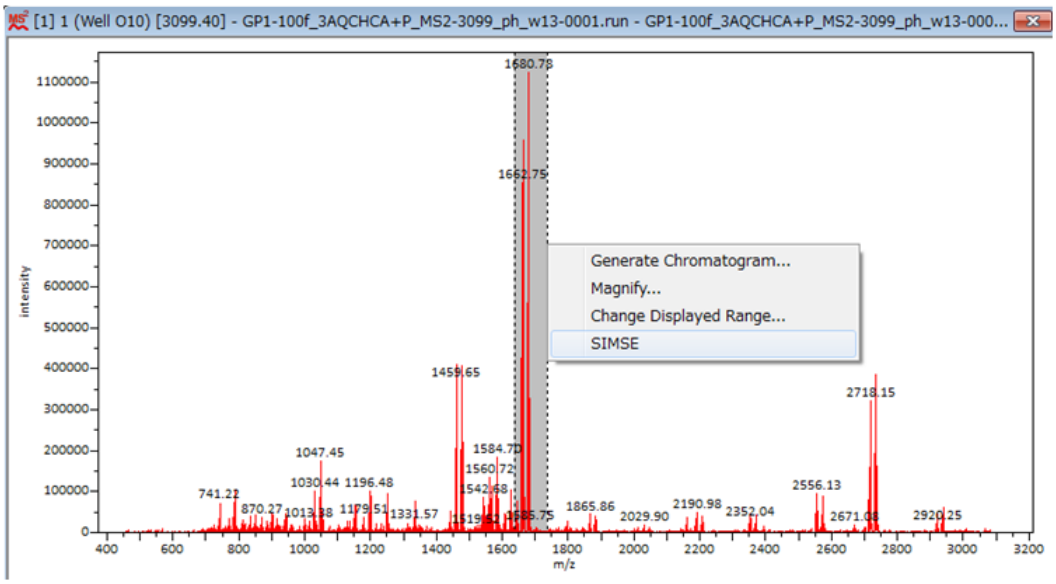
"MWD" is a unique algorithm in Mass++ to detect monoisotopic peaks from mass spectra. In general, de novo sequencing is applicable only for monoisotopic peak list.

Click [Detect]. Peak detection is processed for an active MS/MS spectrum.

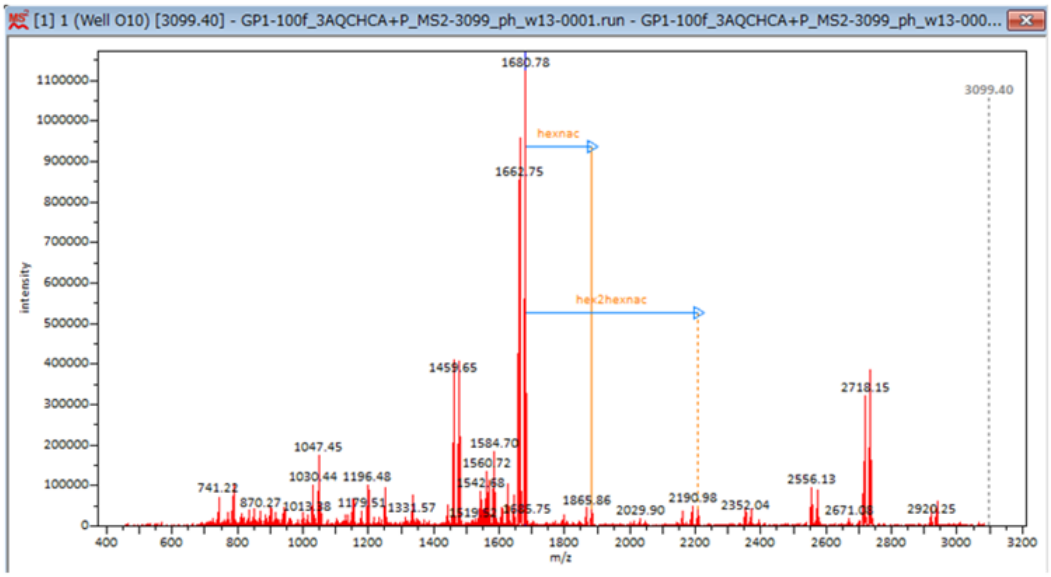
Preparation is complete.

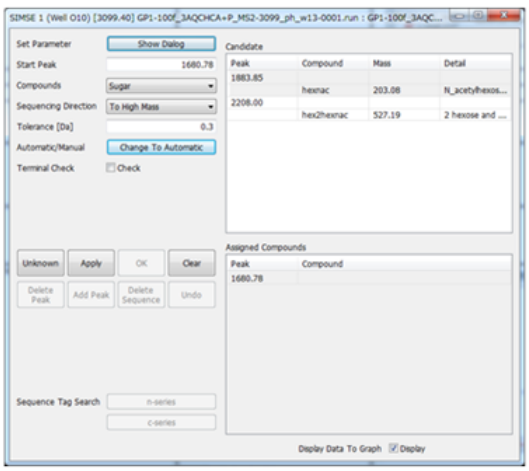
3.2.3. Starting de novo sequencing

Right-drug over start peak for de novo sequencing and click [SIMSE]



[SIMSE] is displayed and de novo sequencing is started on the spectrum.



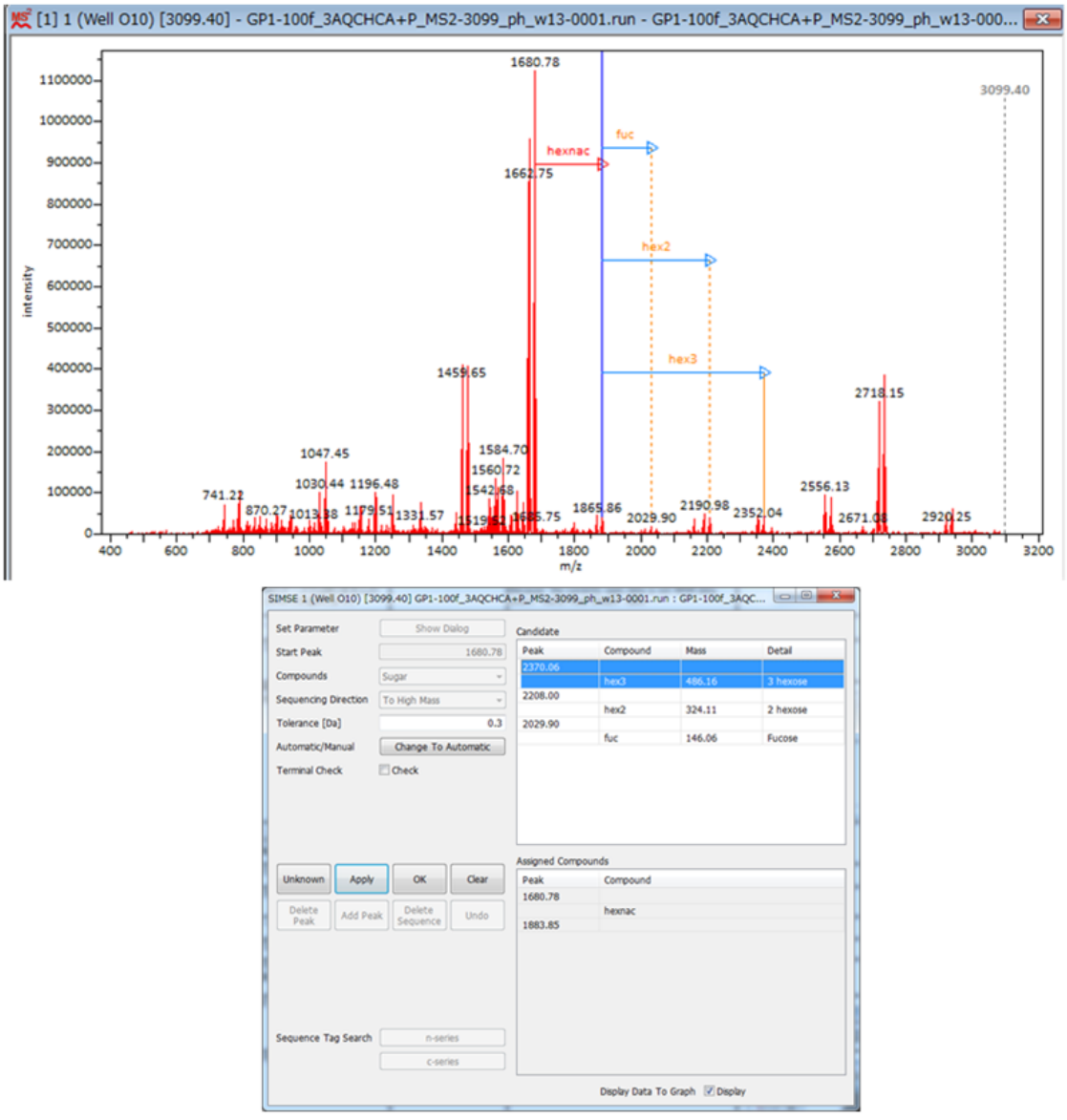


3.2.4. Processing de novo sequencing manually

De novo sequencing is processed to high mass direction from the start peak at m/z = 1680.79. Select "Sugar" on [Compound] drop-down list and 2 compounds are shown as candidates to assign space between any peak and start peak.

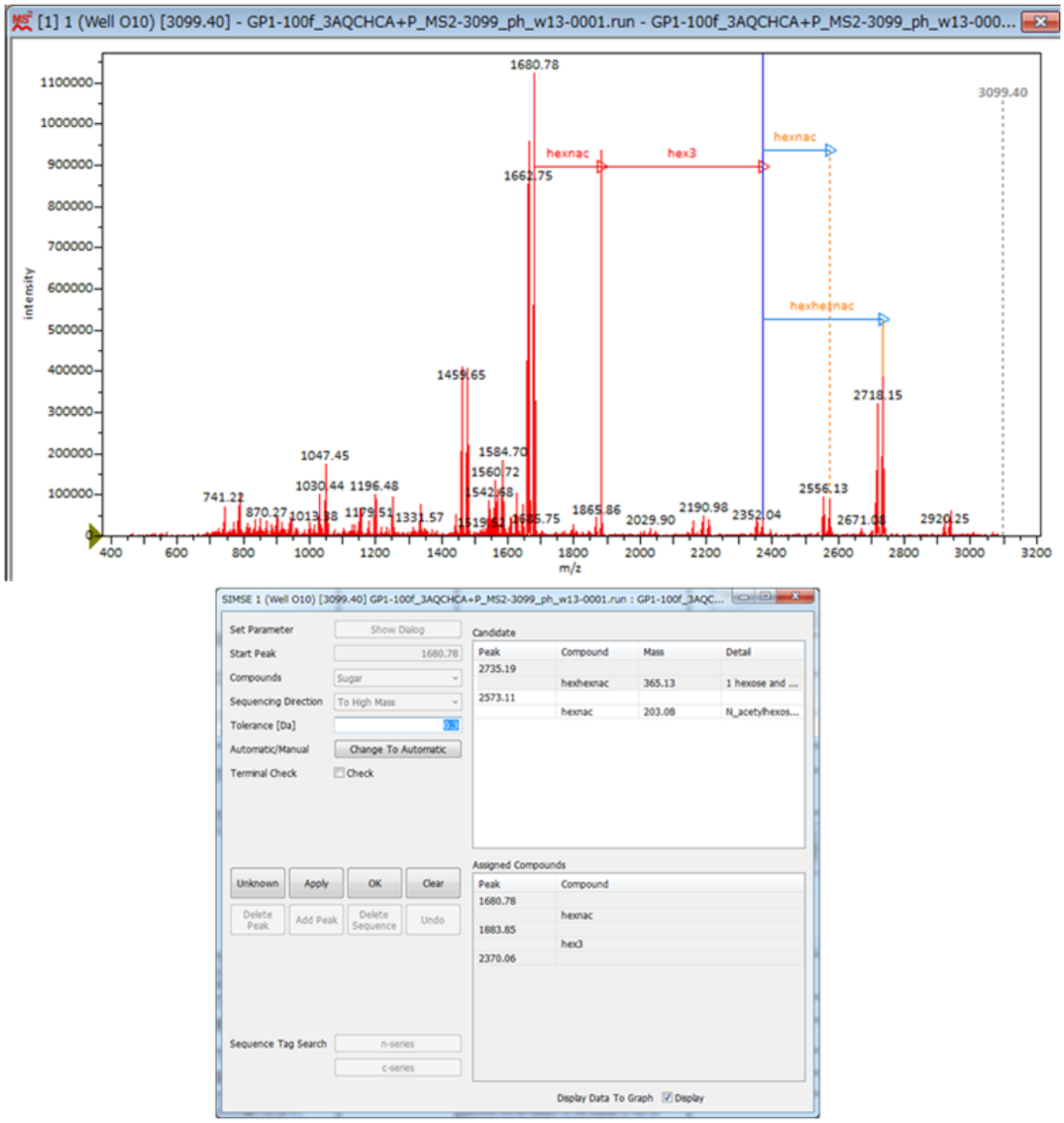
Compound "hexnac" is selected in [Candidate] list. Click [Apply] to search next candidates.

First candidates "hex3", "hex2" and "fuc" are displayed in [Candidate] list.



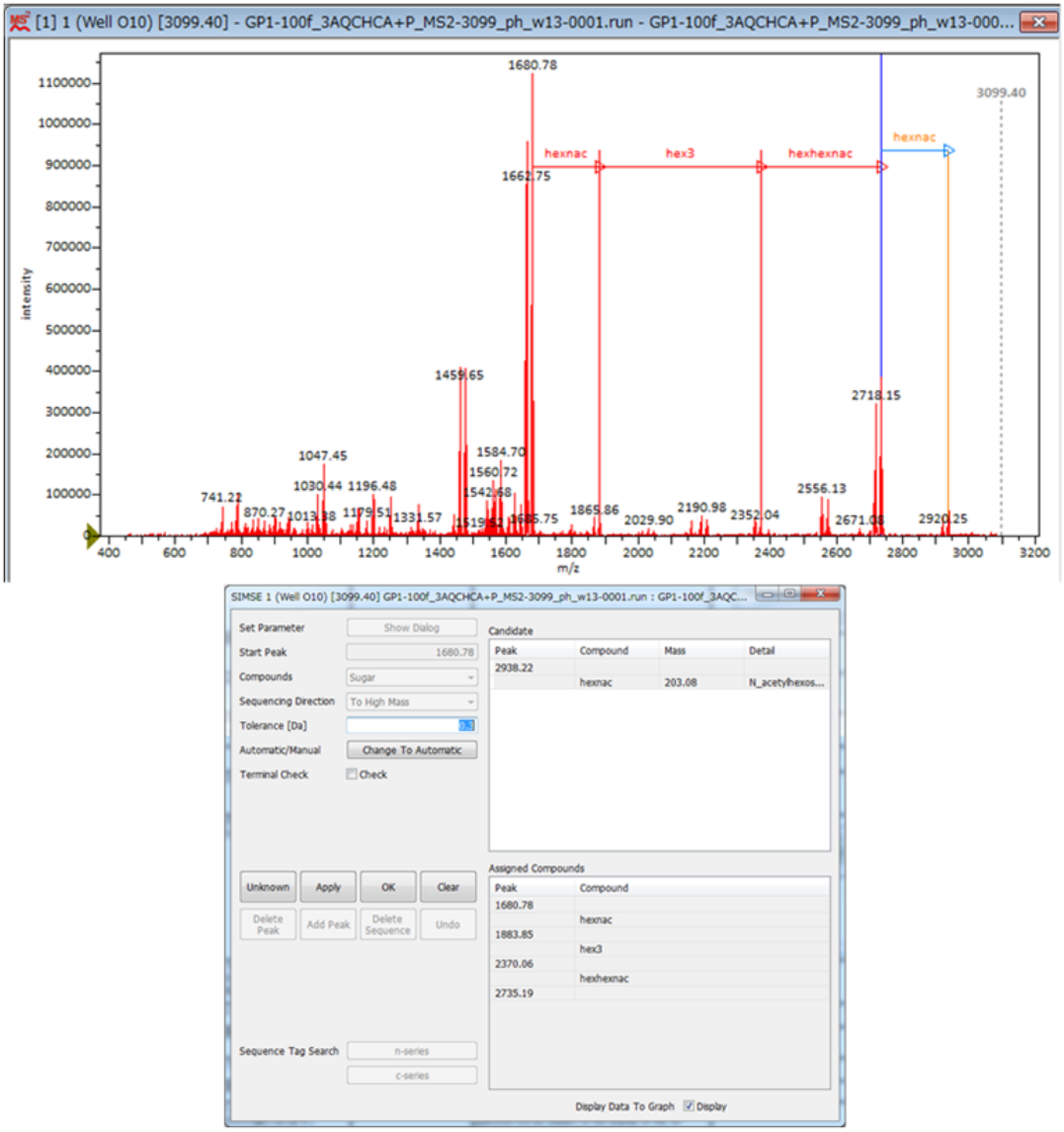
Compound "hex3" is selected in [Candidate] list. Click [Apply] to search next candidates.

Second candidates "hexhexnac" and "hexnac" are displayed in [Candidate] list.



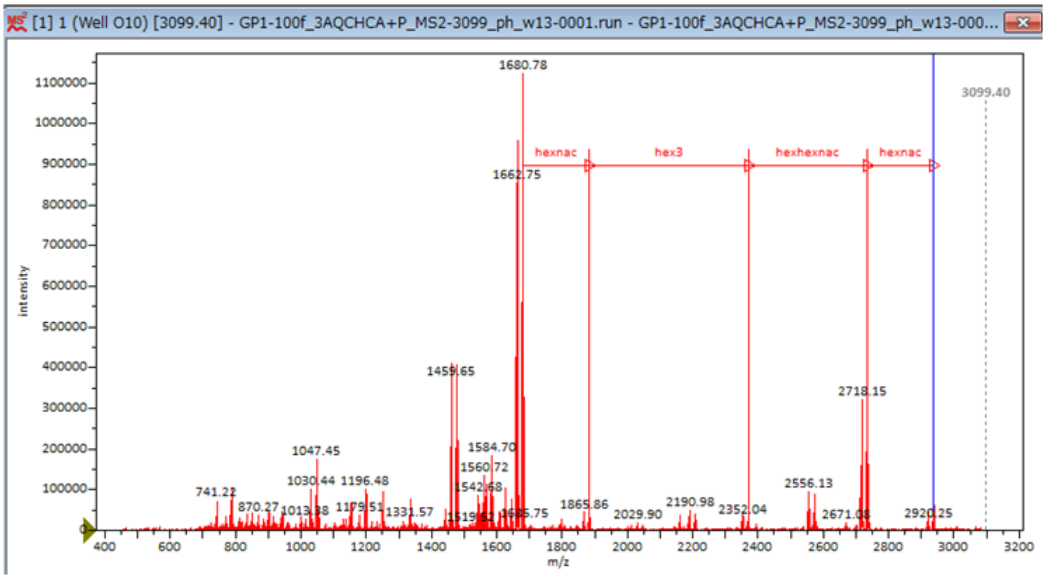
Compound "hexhexnac" is selected in [Candidate] list. Click [Apply] to search next candidates.

Third candidate "hexnac" are displayed in [Candidate] list.



Click [Apply] to search next candidates.

De novo sequencing is complete when no more candidates are found.

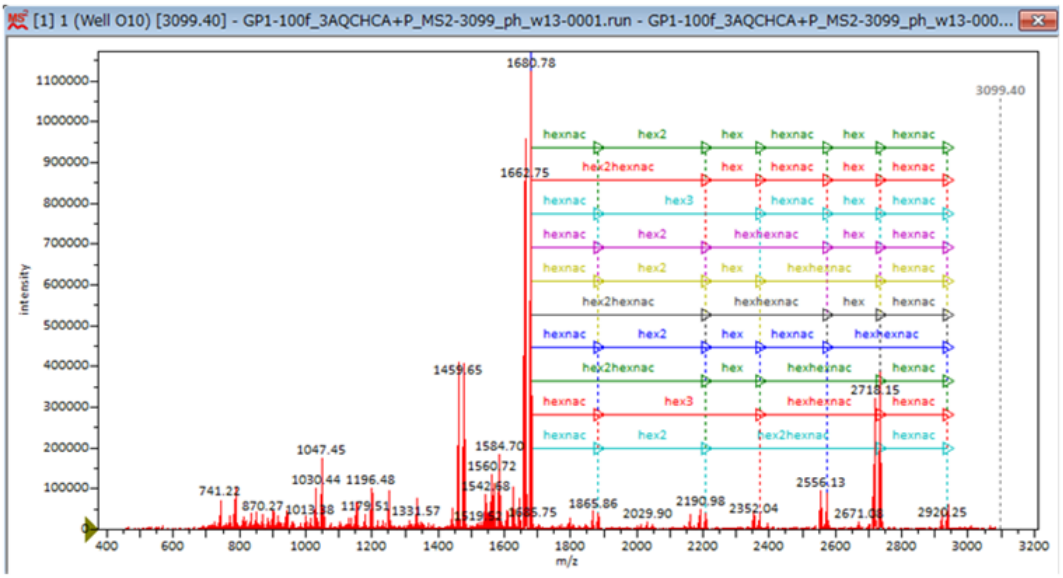


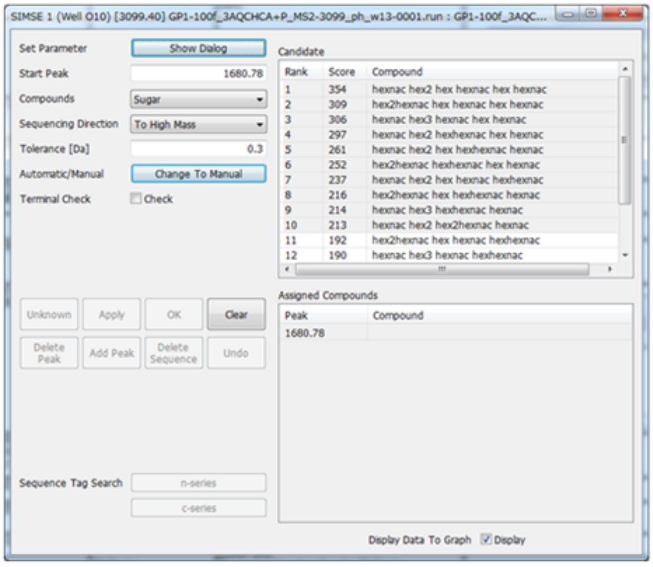
3.2.5. Processing de novo sequencing automatically

Automatic de novo sequencing is also applicable. De novo sequencing is started from the same start peak. First, select start peak by right-drug in the same way

Click [Change to automatic]

Result of automatic de novo sequencing is displayed.





3.3. Script Console

This section describes how to use script console.

Script console is a tool to extend Mass++'s functionality by executing programs which are called scripts. Scripts are executed at anytime you need without restarting Mass++ or compilation to executable binaries. Script is written in a script language and its grammar is easy to learn. By script console, you can extend Mass++'s functionality using Mass++'s API for .NET Framework.

3.3.1. System Requirements of the Script Console

Script console (of Mass++ ver.2.7.2) just supports IronPython 2.7.x. If you don't have IronPython in your computer, download latest IronPython 2.7 installer from http://ironpython.net/ and install it.

3.3.2. About Iron Python

IronPython (http://ironpython.net/) is an open-source implementation of python scripting language to .NET Framework on Windows. Python is a widely used scripting language to develop various softwares including web server. IronPython provides higher source code compatibility to python language. IronPython allows to use many Windows API provided by .NET Framework. See following web site to know details of grammar and API.

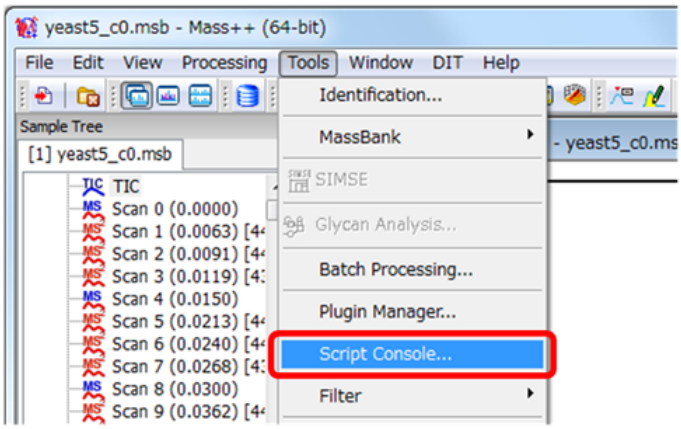
http://ironpython.net/documentation/

If you are not familiar with python language, see python tutorial:

http://docs.python.org/3/tutorial/index.html

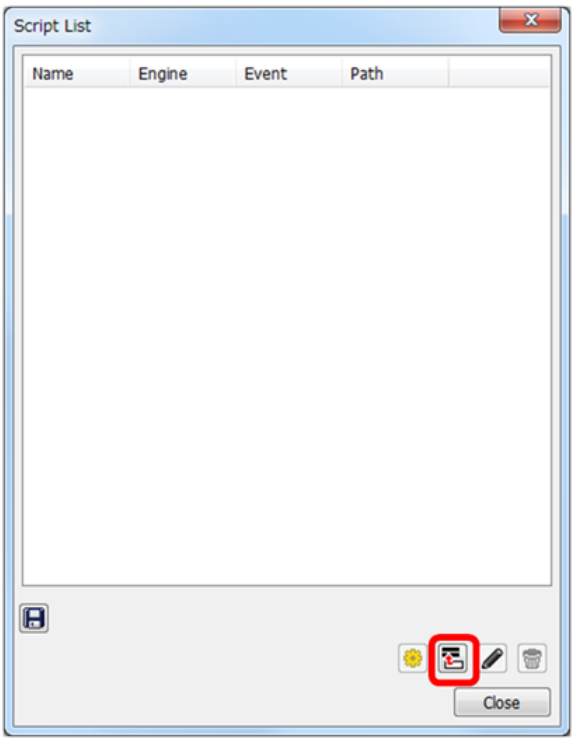
3.3.3. Opening the Script Console

Click [Script Console] from the [Tools] menu to open script console.



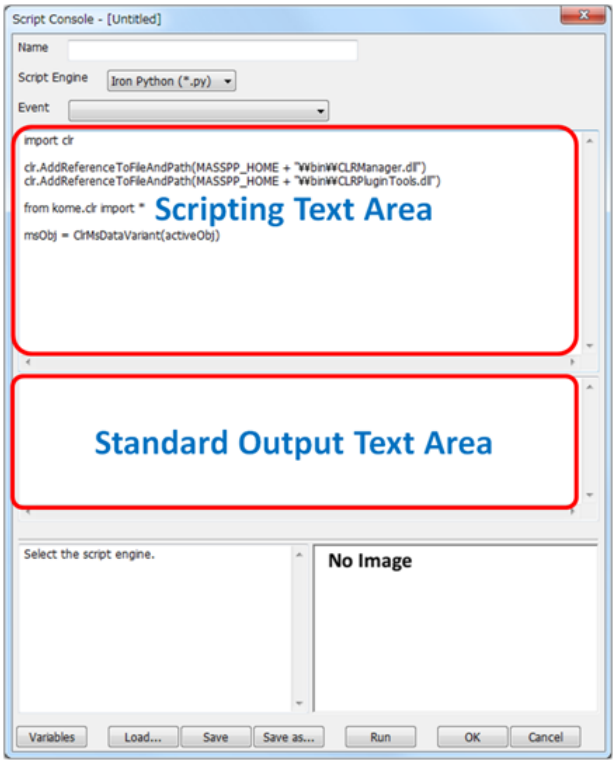
[Script List] opens. [Script List] allows script files to register, execute, and edit.

Then click [Add Script].



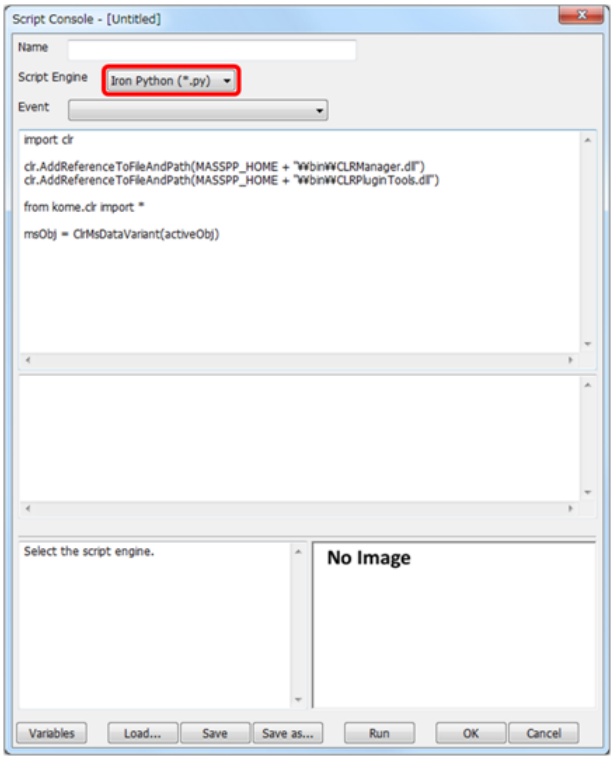
[Script Console] opens.

Using script console, you can edit and execute scripts which are displayed in [Scripting Text Area]. [Standard Output Text Area] is used for standard output of the script.

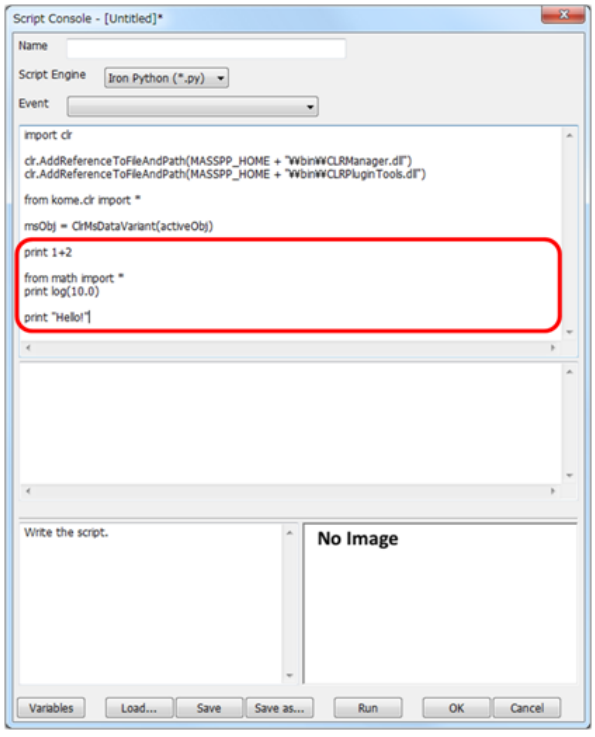


3.3.4. Writing a Simple Script

Simple example of IronPython script is described. Select "Iron Python (*.py)" at [Script Engine] list.



Add following script at the end of existing script in [Scripting Text Area].

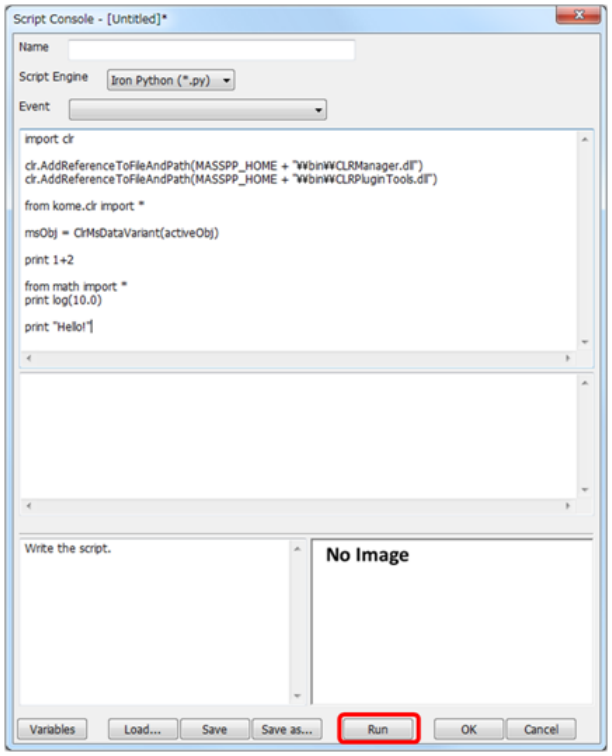


print 1+2 # print out calculation result of '1+2' to standard output

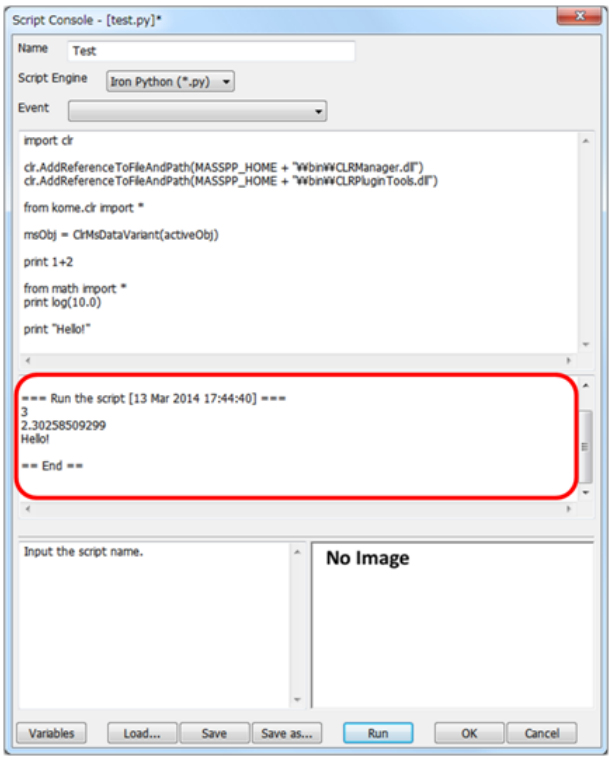
```
from math import * # load all classes and functions from math module to use log function.

print log(10.0) # print out log(10.0).

print "Hello!" # print out string "Hello!" to standard output.
```

3.3.5. Executing Scripts

After writing script, click [Run] to execute script.

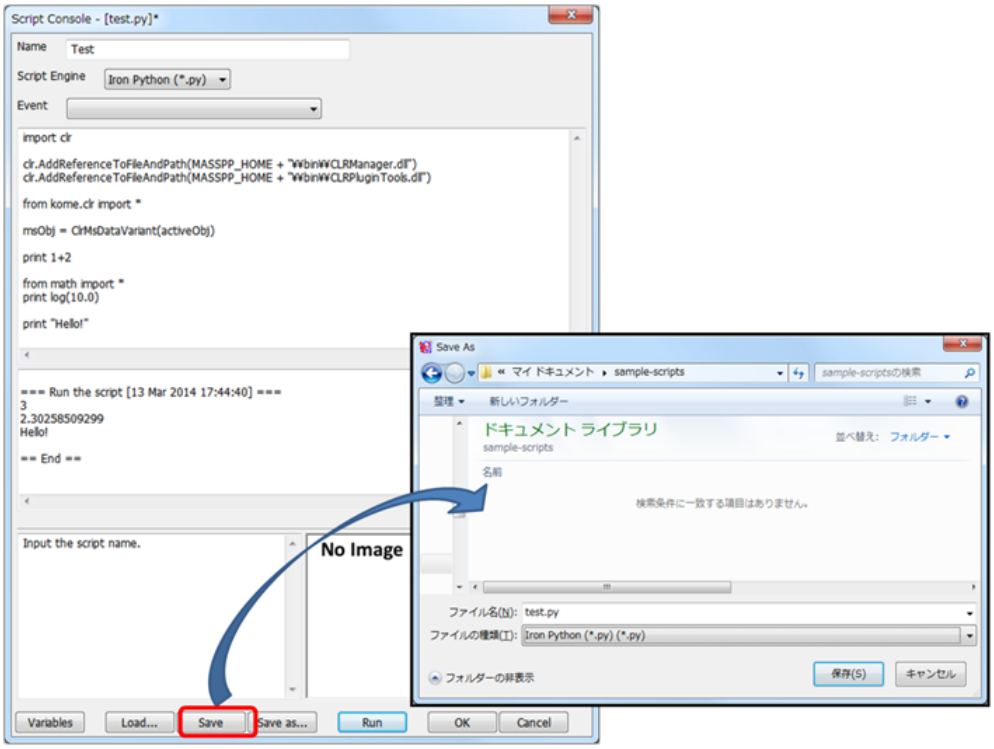


The script is executed and the results of print statements are displayed at [Standard Out Text Area]. The date and time of execution are also outputted. You can use Mass++ as a mathematical calculator using script console.



3.3.6. Registering Scripts

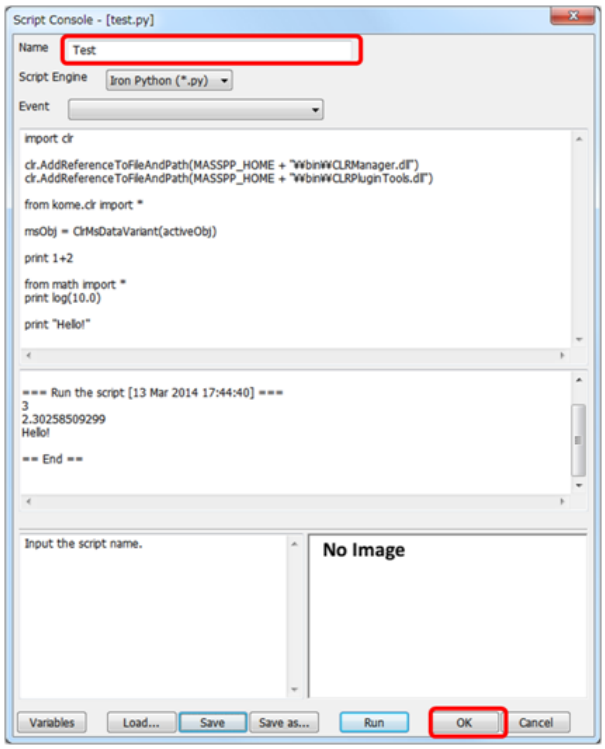
Click [Save] to store the script as a file.



Enter the name of the script file in [File Name] in [Save File].

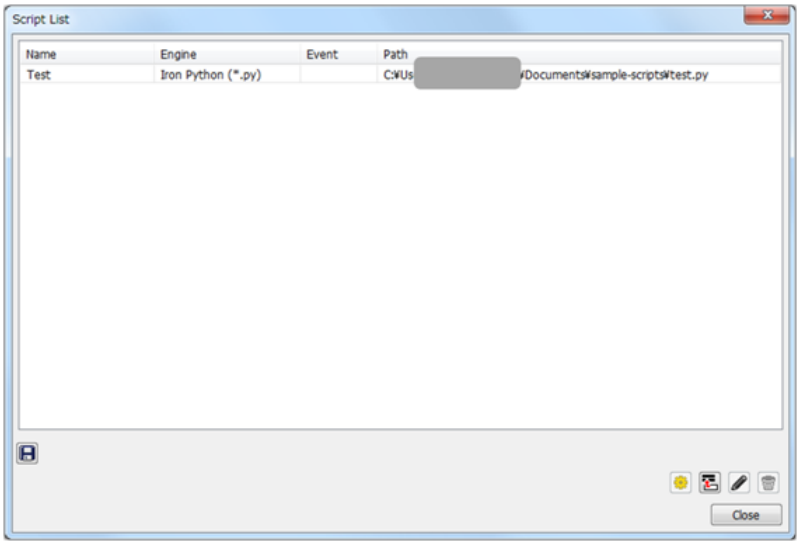
Click [Save], then the script was saved into a file.

Enter the name of script in [Name] for script registration. "Test" is specified in this example.



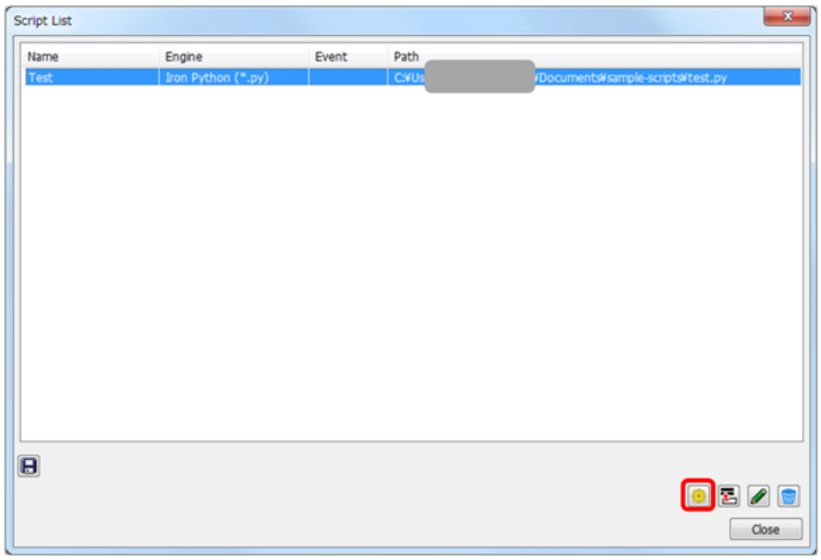
Click [OK].

Script "Test" is registered on [Script List]. [Name] column shows the name of the registered script and [Path] column shows its absolute path.

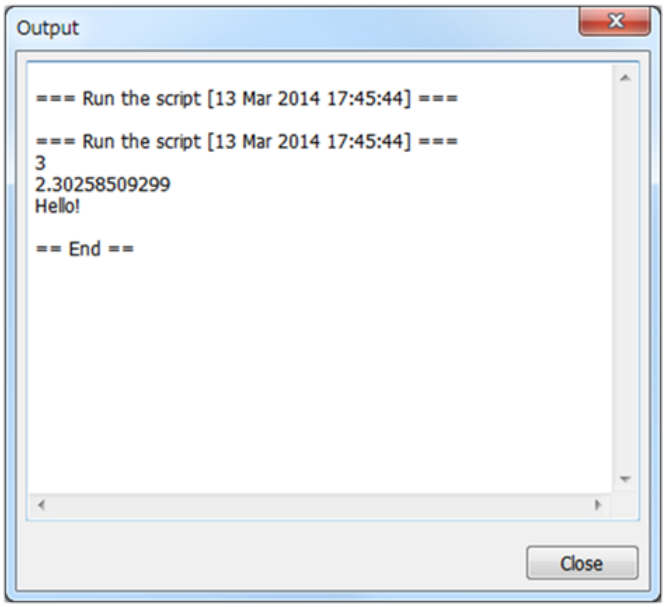


3.3.7. Executing a Registered Script

Click registered script "Test", then click [Run Script].



Output of the selected script is shown in [Output].



After confirmation of the result, click [Close].

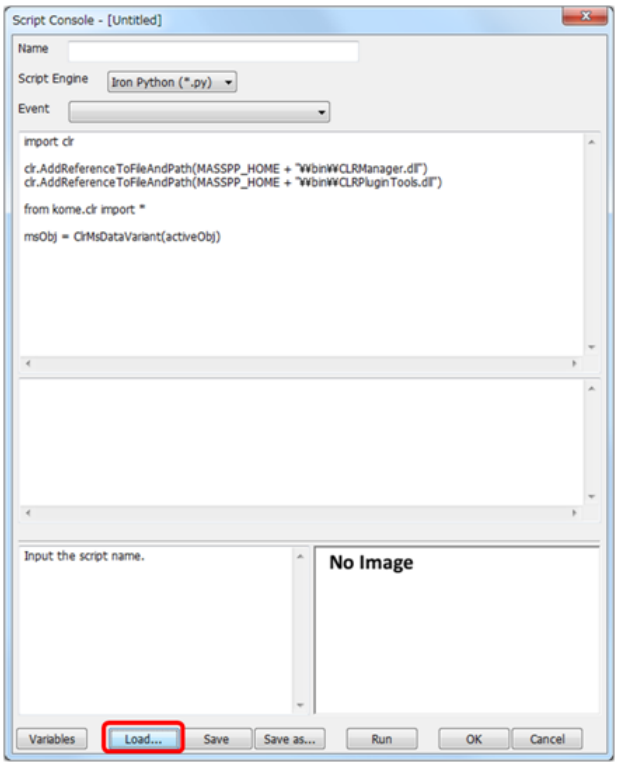
3.3.8. Loading Scripts from a File and Executing It

Loading and executing of script files are described. In this example, a script is loaded to display brief summary of MS/MS data.

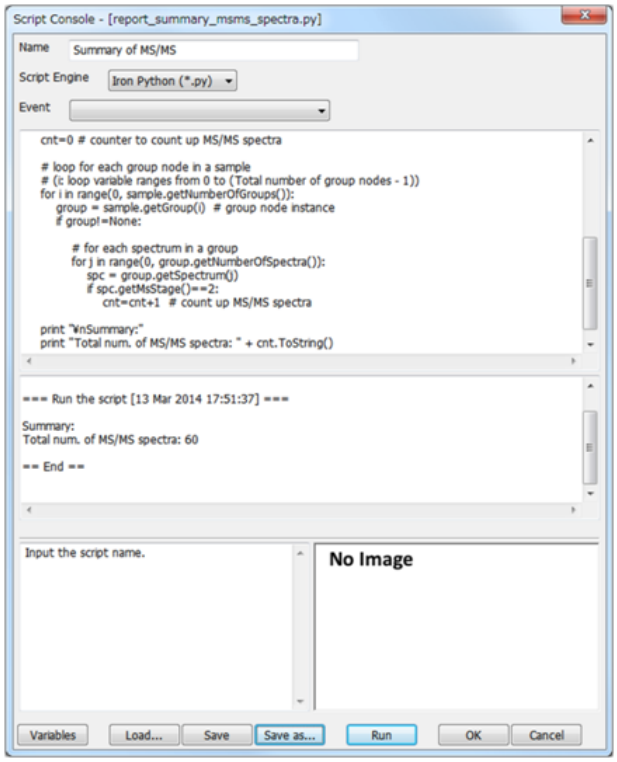
As an MS/MS data, open LC-MALDI data "LC-MALDI_base1.msb" that supplied with Mass++.

Click [Add Script] in [Script List].

[Script Console] opens. Then click [Load].

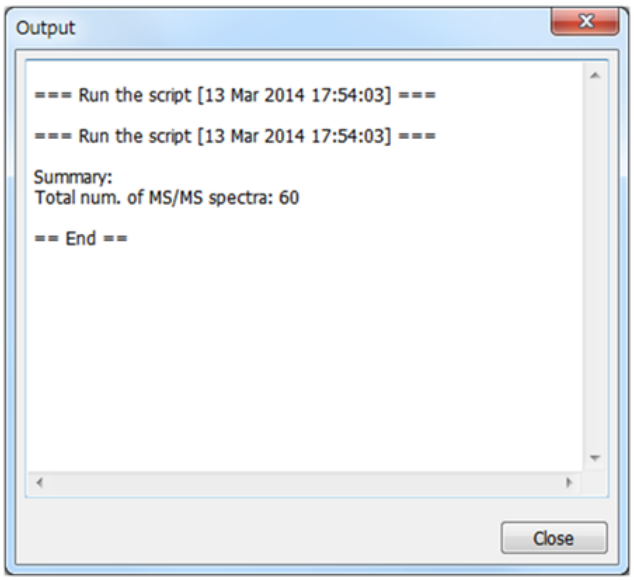


Select "report_summary_msms_spectra1.py" that supplied with Mass++. Then, click [Open].



The selected script is displayed in [Scripting Text Area] of script console. Then register the script.

Select registered script and click [Run Script]. Then, the number of MS/MS spectra is shown in [Output].



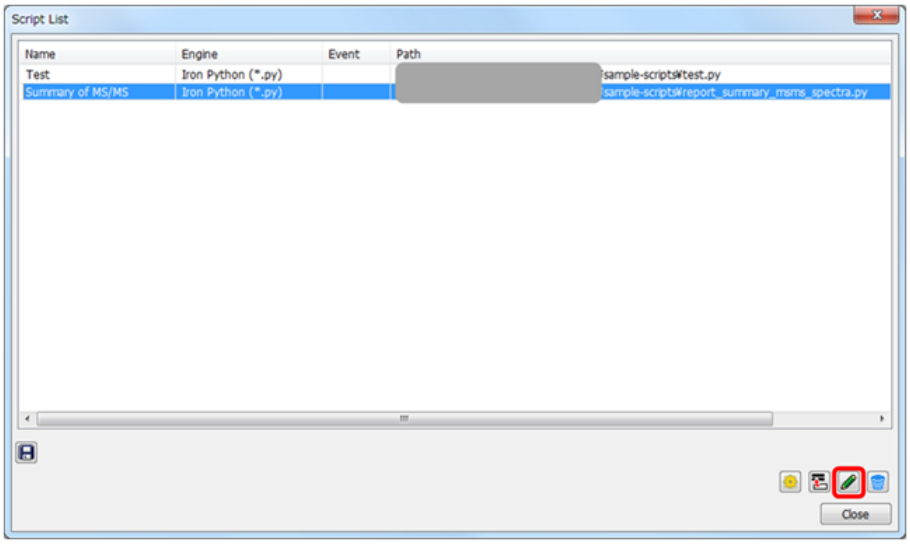
Click [Close] to close [Output].

3.3.9. Editing Scripts

Edit the script to show following information stored in data.

- spectrum name
- precursor ion mass
- lower/upper mass range of precursor ions
- retention time

Click [Edit script] in [Script List].



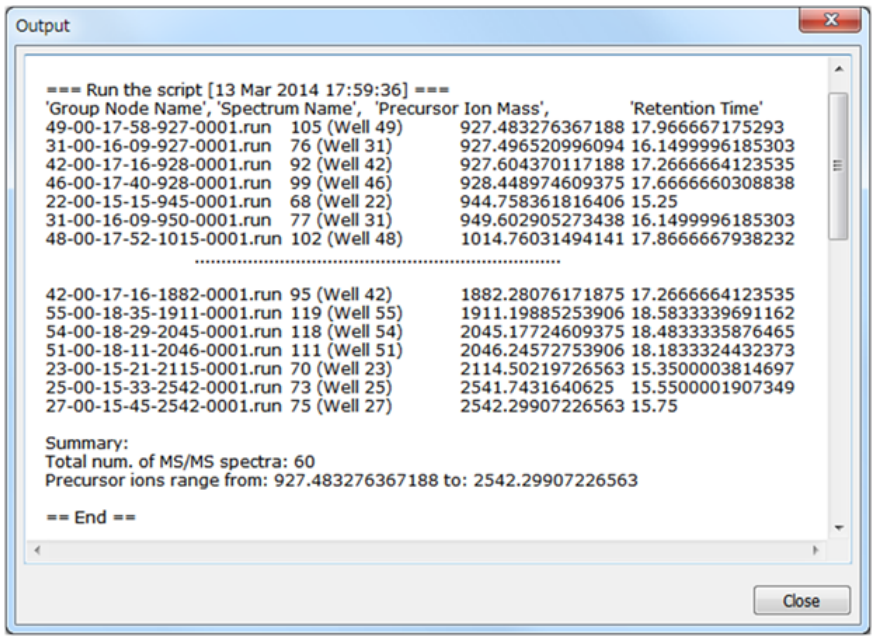
[Script Console] opens with displaying the registered script. Then, insert lines just after the line "from kome.clr import *" (see following figure). The inserted lines define container class to store attributes of MS/MS spectra: group node name, spectrum name, precursor ion mass and retention time. Modified script also supplied with Mass++, by the name of "report_summary_msms_spectra_2.py".

```
import clr  # load .NET library
# load Mass++'s Application Programming Interface for .NET
clr.AddReferenceToFileAndPath(MASSPP_HOME + "\\bin\\CLRManager.dll")
clr.AddReferenceToFileAndPath(MASSPP_HOME + "\\bin\\CLRPluginTools.dll")
from kome.clr import *
# definition of container class to store properties of a spectrum
class spec_feature:
    def __init__(self, group_name, name, pc_mass, rt):
        self.group_name =  group_name
        self.name       =  name
        self.pc_mass    =  pc_mass
        self.rt         =  rt
    def __repr__(self):  # default string representation of the instance
        return str(self.group_name + '\t' +  self.name + '\t'
                   + self.pc_mass.ToString() +'\t' + self.rt.ToString())
list_spec_feature = []  # container to store properties of spectra
feature_sort_key = lambda spec_feature: spec_feature.pc_mass  # sort key for spec_feature
# access MS/MS spectra in active MS data instance of Mass++, and then
# store properties of spectra into list_spec_feature
msObj = ClrMsDataVariant(activeObj)  # active MS data instance of Mass++
sample = msObj.getSample()           # sample node of MS data
if sample == None:  # active sample doesn't exist
    print "cannot obtain active sample."
```

Next, edit lines after 'cnt = cnt + 1' as shown in following figure. These lines make attributes of each MS/MS spectrum be stored into the container previously defined and all information be printed to standard output.

```
            # for each spectrum in a group
            for j in range(0, group.getNumberOfSpectra()):
                spc = group.getSpectrum(j)
                if spc.getMsStage()==2:
                    cnt=cnt+1  # count up MS/MS spectra

                    # store properties of MS/MS spectra into list
                    sf = spec_feature( group.getName(), spc.getName(),
                                       spc.getPrecursor(), spc.getRt() )
                    list_spec_feature.append(sf)

    # print out properties of MS/MS spectra
    # sorted by precursor ion mass in ascending order
    print "'Group Node Name',\t'Spectrum Name',\t'Precursor Ion Mass',\t'Retention Time'"
    for spc in sorted(list_spec_feature, key=feature_sort_key):
        print spc
    print "\nSummary:"
    print "Total num. of MS/MS spectra: " + cnt.ToString()
    print ("Precursor ions range from: "
           + min(list_spec_feature, key=feature_sort_key).pc_mass.ToString() + " to: "
           + max(list_spec_feature, key=feature_sort_key).pc_mass.ToString())
```

Run scripts after edition. In addition to total number of MS/MS spectra, spectrum name, precursor ion mass and retention time of each spectrum are listed and sorted by precursor ion mass in ascending order. Lower/upper mass range of precursor ion mass in the data are also shown.



3.3.10. Executing Scripts by Event

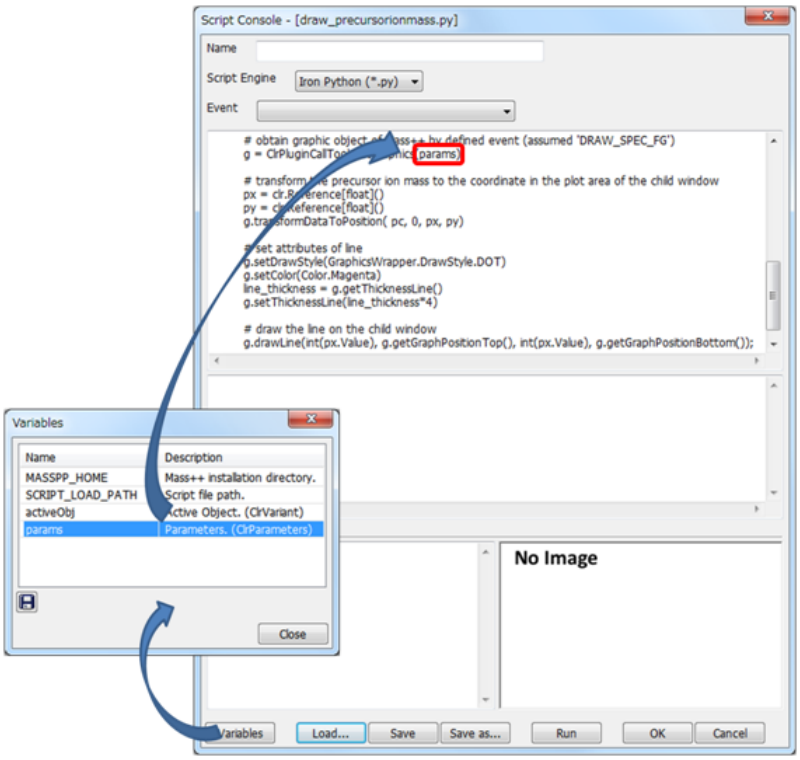
Mass++ can execute scripts using defined plug-in events such as re-drawing spectra or peak detection. Here, how to execute scripts by event is described using an example of drawing lines in MS/MS spectra.

Open script console and load "draw_precursorionmass.py" that supplied with Mass++.

Click [Variables] on [Script Console]. [Variables] lists pre-defined parameters such as 'params' which is used in sample script. 'params' is a set of parameters of plug-in. Types of the parameters are dependent on an event type of Mass++ plug-in.

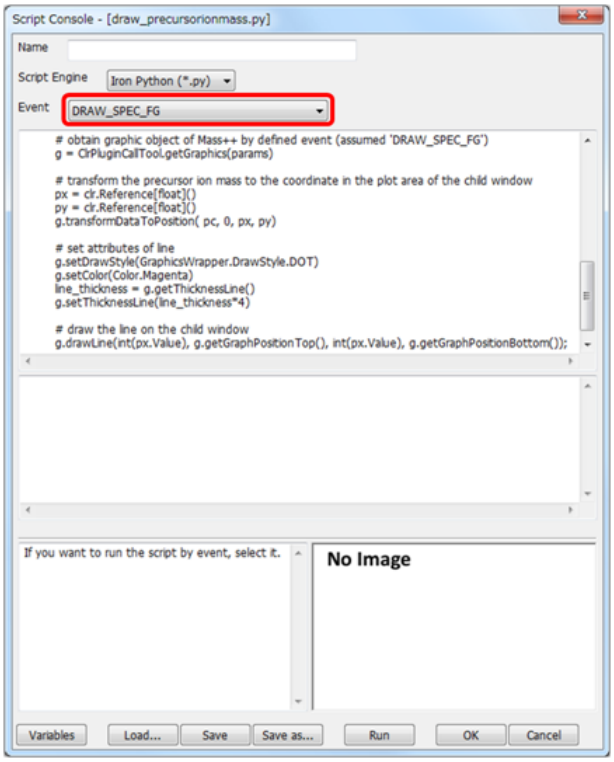
[Variables] lists other variables. "MSPP_HOME" is a start-up folder of Mass++ as string. "activeObj" is an instance of active data class. "SCRIPT_LOAD_PATH" is an absolute path of a loaded script and used for loading other modules.

When a listed variable is double-clicked, then the clicked variable is inserted to [Script Text Area] of script console.

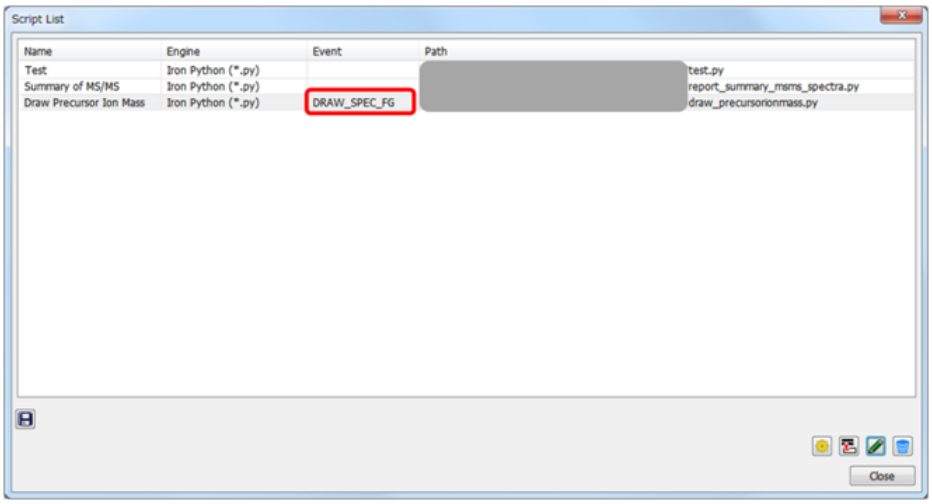


Click [Close] to close [Variables].

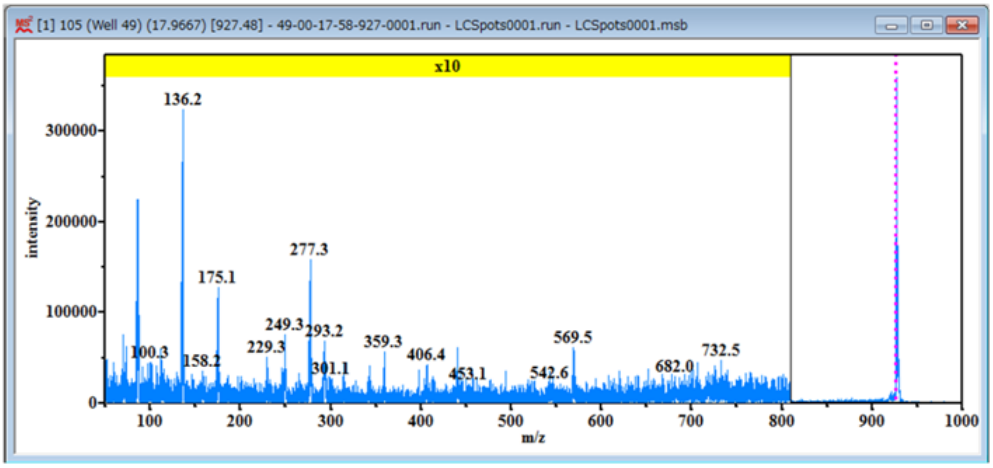
To execute a script by an event, select a target event type at [Event] list. In this example, select "DRAW_SPEC_FG" which is the event type for re-drawing the foreground in the spectrum drawing area. Then, register the example script.



"DRAW_SPEC_FG" is shown in [Event] column of [Script List]. Then, open a MS/MS spectrum of "LC-MALDI_base1.msb" which has been loaded.



Vertical dotted line in magenta is drawn at precursor ion mass.



For more details of Mass++'s API, read "Developers' manual" that will be published with Mass++.

Chapter 4. MassBank

MassBank is a public metabolomics data base developed by a group comprising the Nara Institute of Science and Technology and Professor Takaaki Nishioka. Mass ++ makes it possible to register data efficiently and perform database search directly to MassBank.

4.1. Create Spectrum Records

Create spectrum records for submission to MassBank. Sample data in this section were provided by Prof. Shigehiko Kanaya at Nara Institute of Science and Technology (http://kanaya.naist.jp/joomla/).

4.1.1. Preface

In order to submit data to MassBank, you have to translate your data into the "MassBank" format using "Record Editor" attached to the MassBank server package, and then register them on the MassBank server with the web application "Administration Tool," a part of the MassBank server system. However, one MassBank record contains more than 120 items, which makes it difficult to input all items manually.

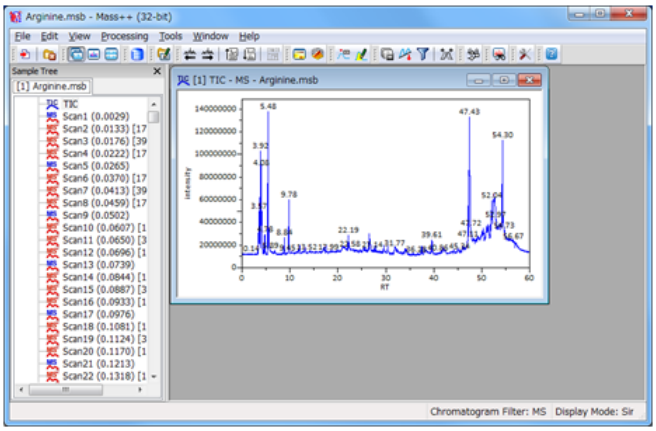
Hence, Mass++ is equipped with the "Creating MassBank record file" function for the automatic generation of MassBank-formatted data, mainly the file format transformation of raw data. See the MassBank manuals for details.

You can either publish the data on the internet (public server) or share the data only in the intranet (in-house server). For the construction of a public server, you (now a contributor) have to obtain three-letter code for the accession (described in the following section) specific to each contributor from the MassBank project.

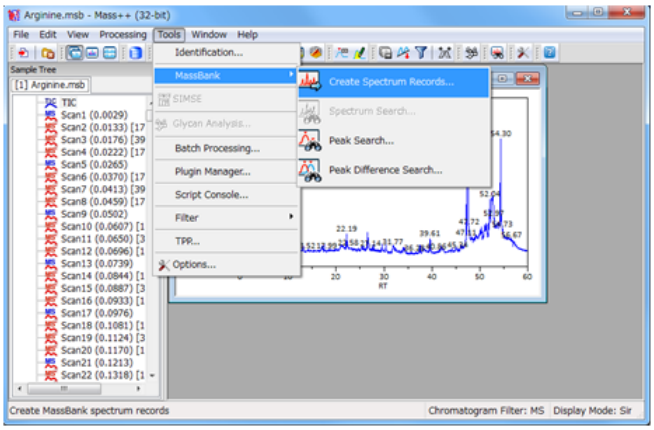
This users' guide is described for the construction of the in-house server. See the MassBank manuals for more information about the MassBank spectrum records.

4.1.2. Open the [Create] wizard ("Create Spectrum Records").

Open "Arginine.msb."

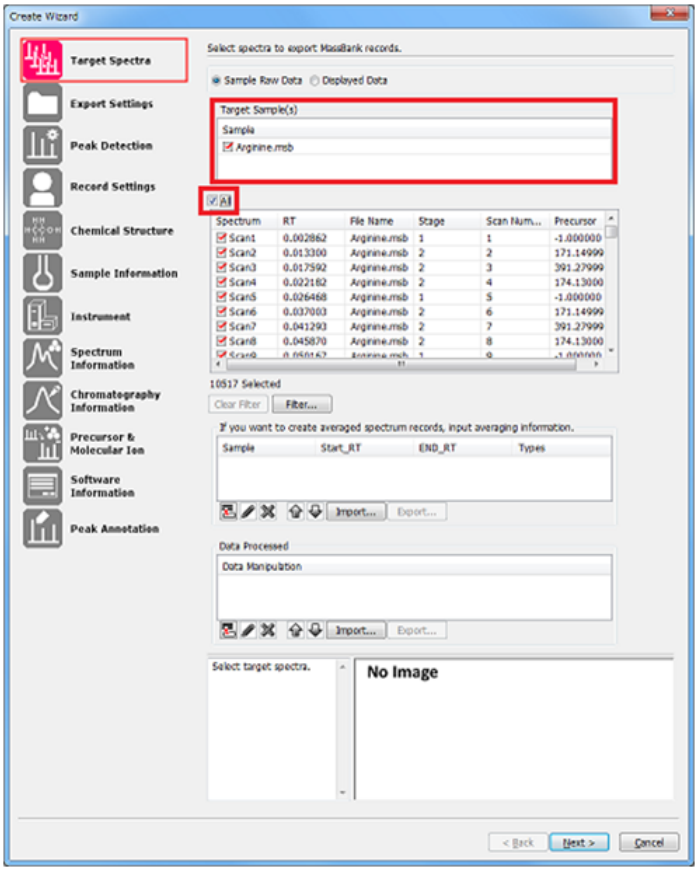


Click [MassBank] - [Create Spectrum Records] from the [Tools] menu.

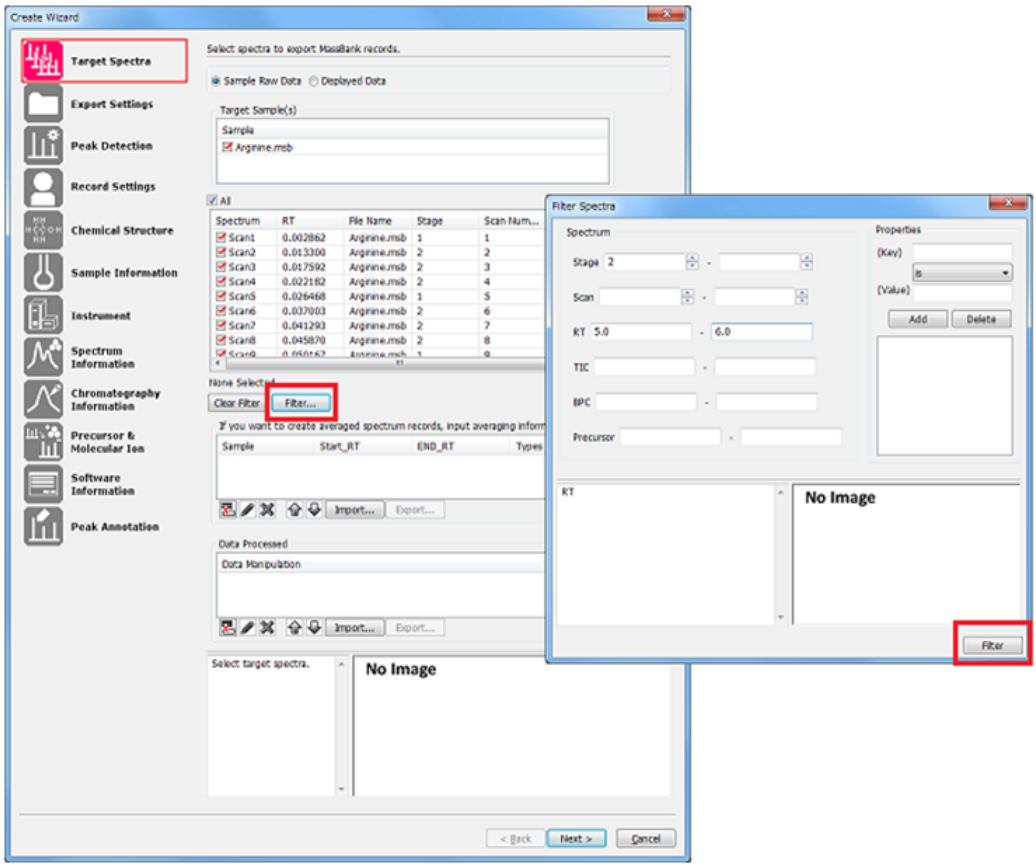


[Create] wizard opens.

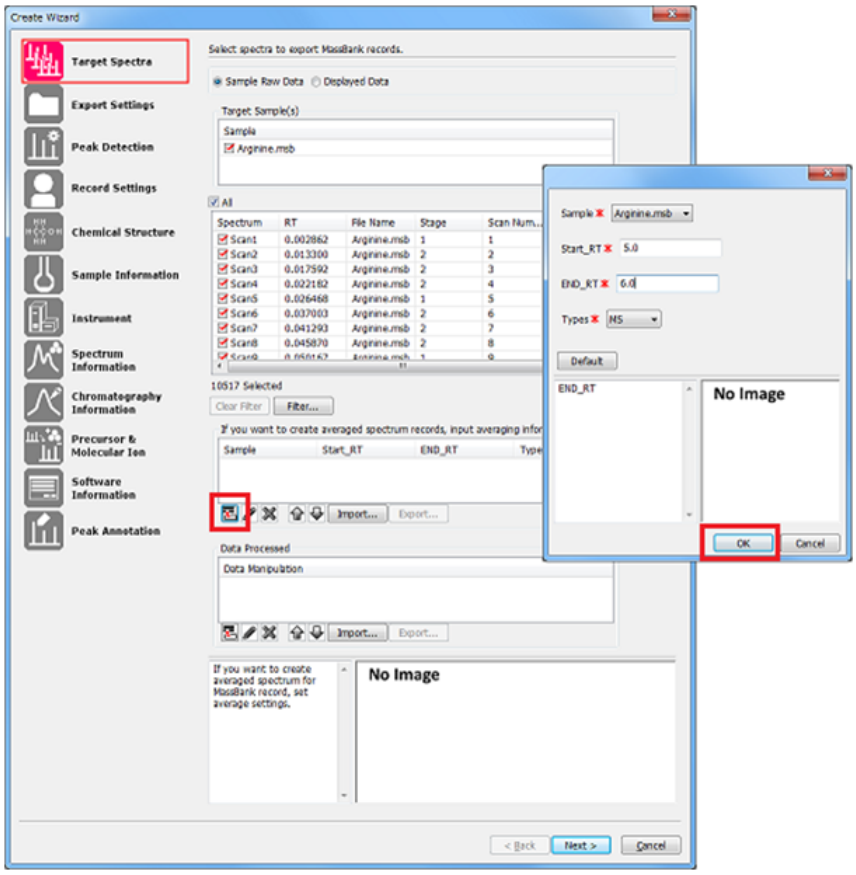
Select all spectra. First, check the "Arginine.msb" in [Target Sample(s)] area. All spectra contained in "Arginine.msb" are listed below [Target Sample(s)] section. Then check [All].



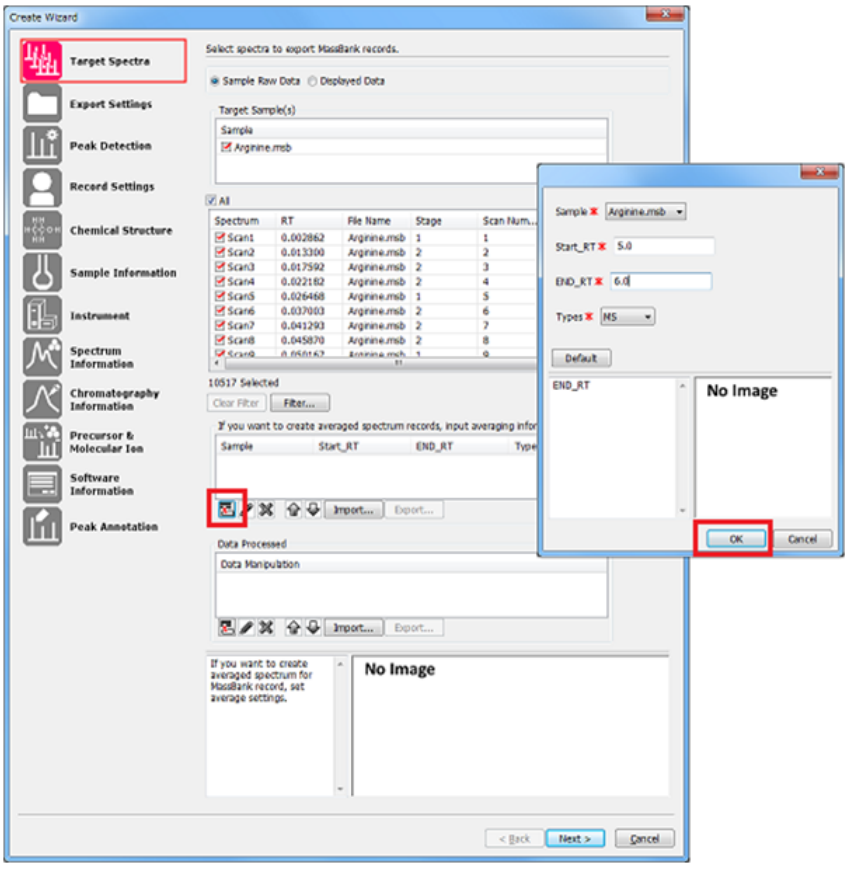
Note: When you want to focus on spectra meeting your conditions, click [Filter]. [Filter Spectra] opens. Then input filtering condition and click [Filter].



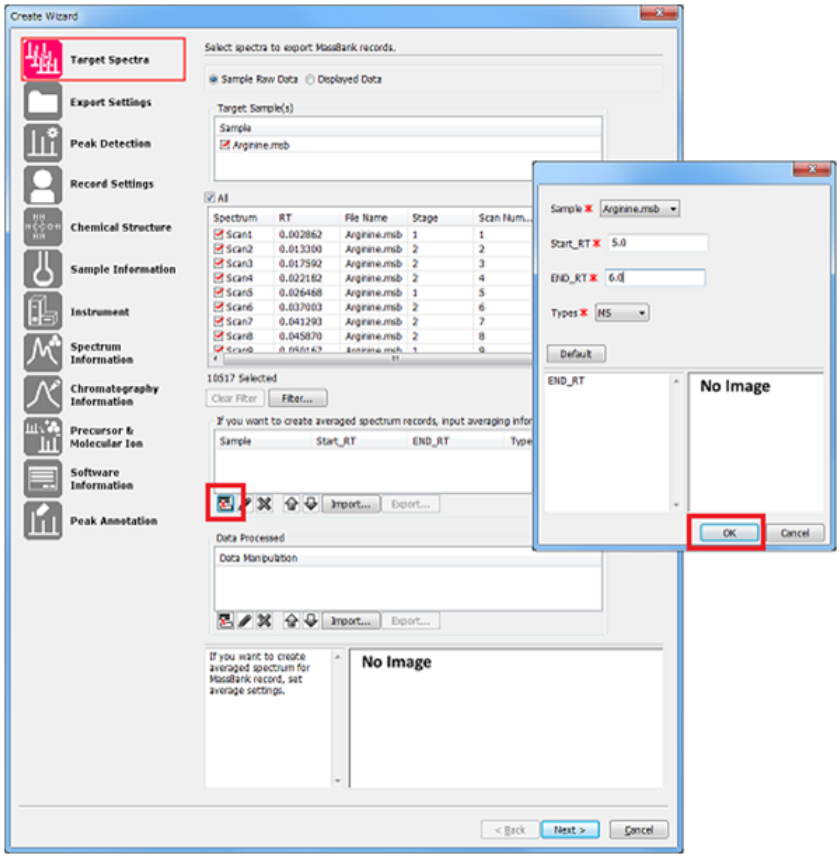
Note: When you want to create records from averaged spectrum, click [Add] of the "averaged spectra list." [Filter Spectra] opens. Input conditions of spectra for averaging process:



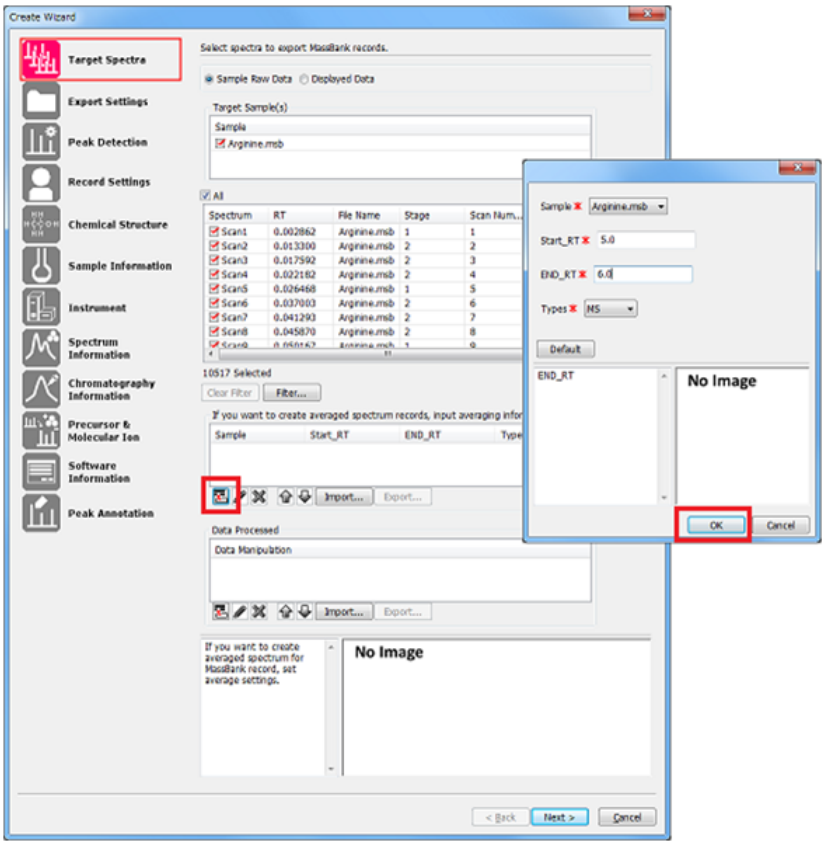
・target sample



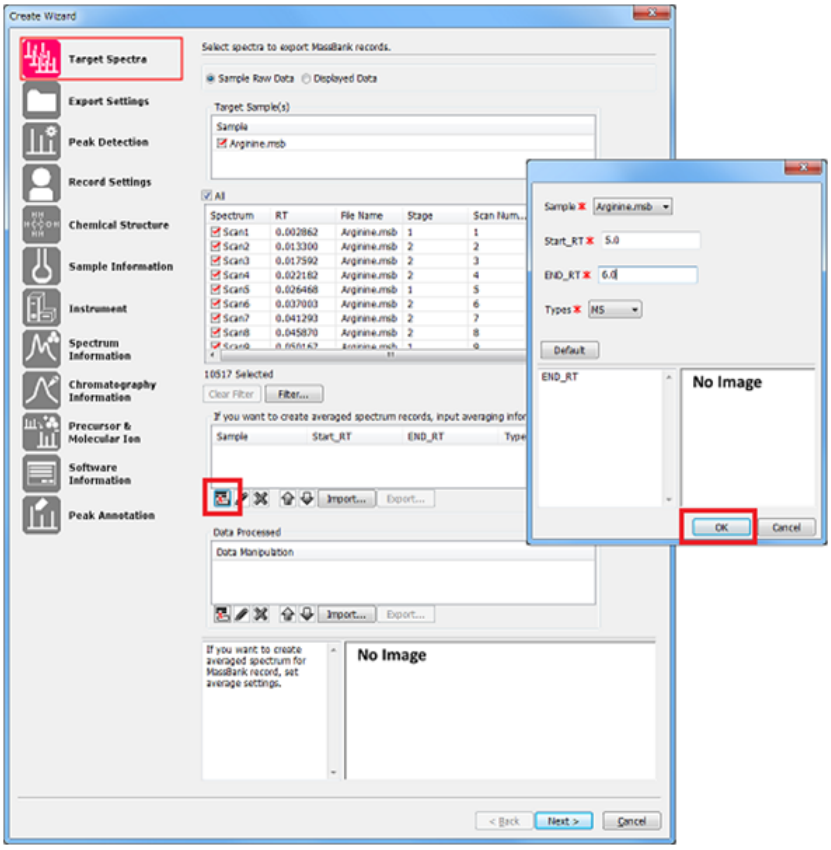
・range of RT



・type of spectra (All: any spectra, MS: only MS1 spectra, MS/MS: MSn spectra. n is 2 or more)

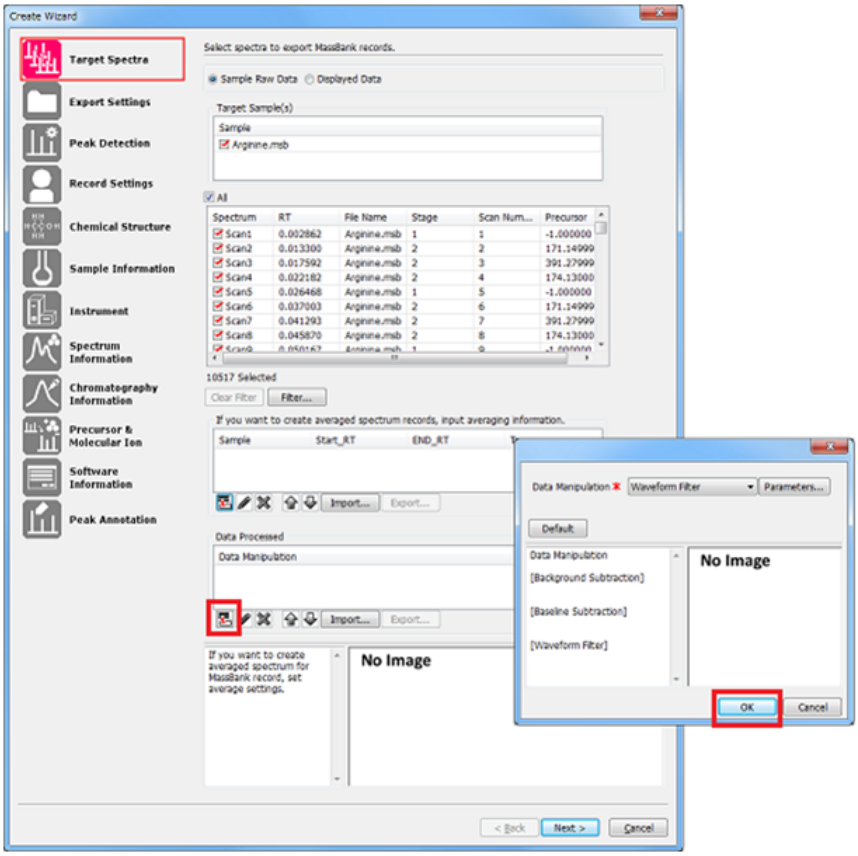


and click [OK].

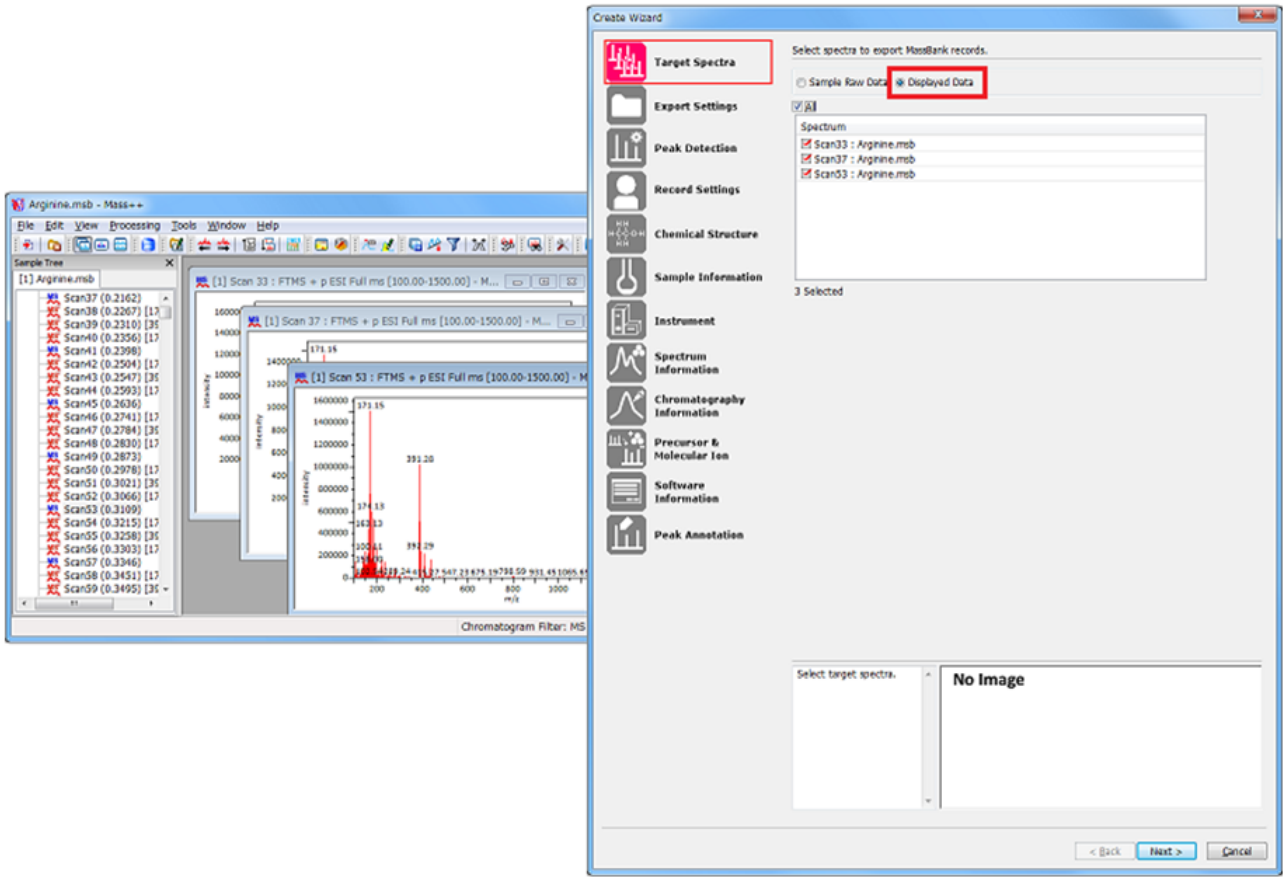


Note: When you want to create records from manipulated spectra, click [Add] below [Data manipulation] list. A dialog opens. Select manipulation process and click

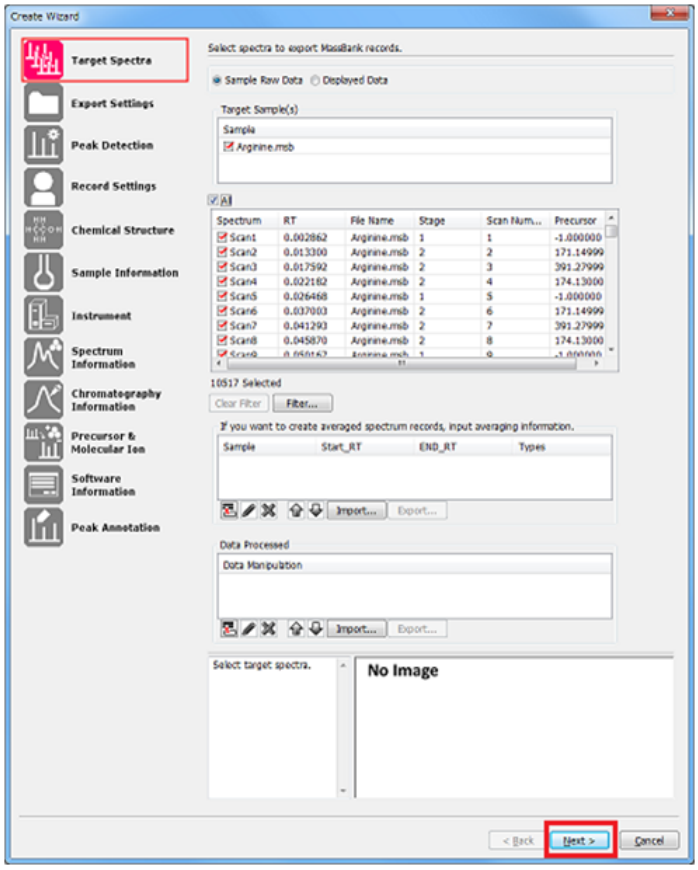
[Parameters]. Another dialog opens. Set parameter information in a similar way. Click [OK] to close each dialog.



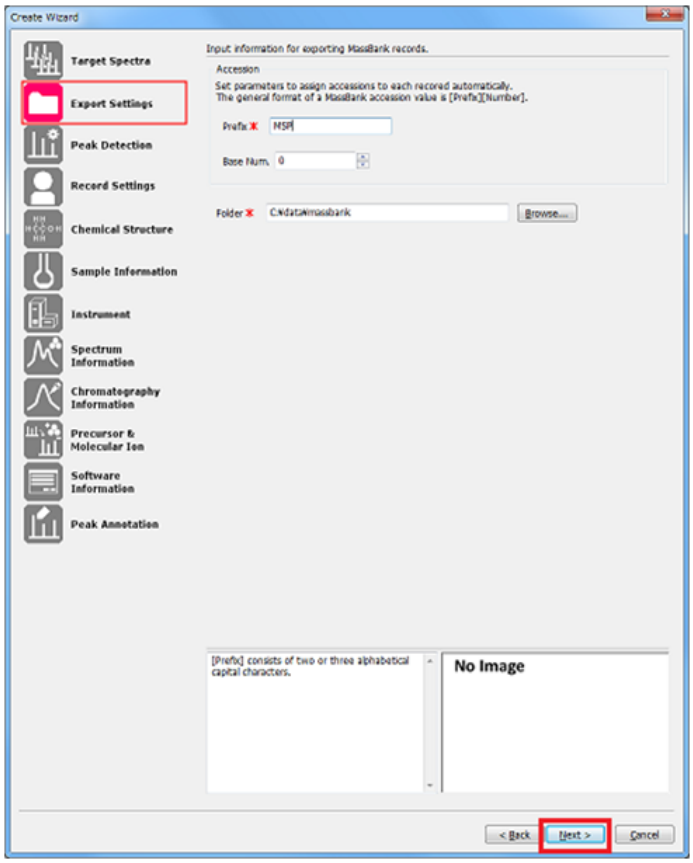
Note: When you want to create records from spectra currently displayed, select [Displayed Data] radio button and click the target spectra to select.



Click [Next] to go to the next page.

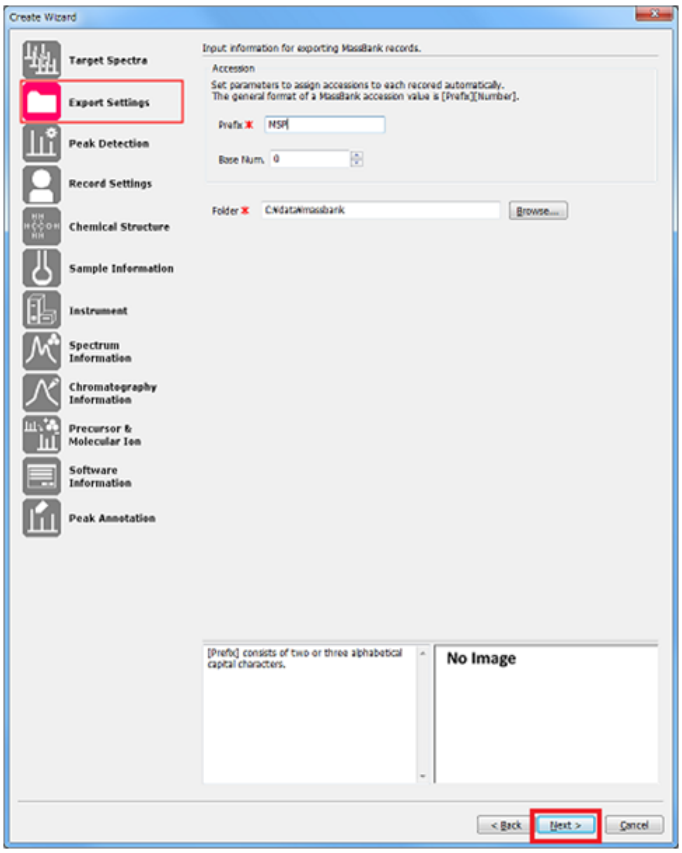


Set parameters for data exporting in [Export Settings] page. Each MassBank record is assigned to a unique "Accession," namely, an identifier. An accession consists of two or three uppercase alphabetic characters follower by five digits (See the MassBank manuals for more information).

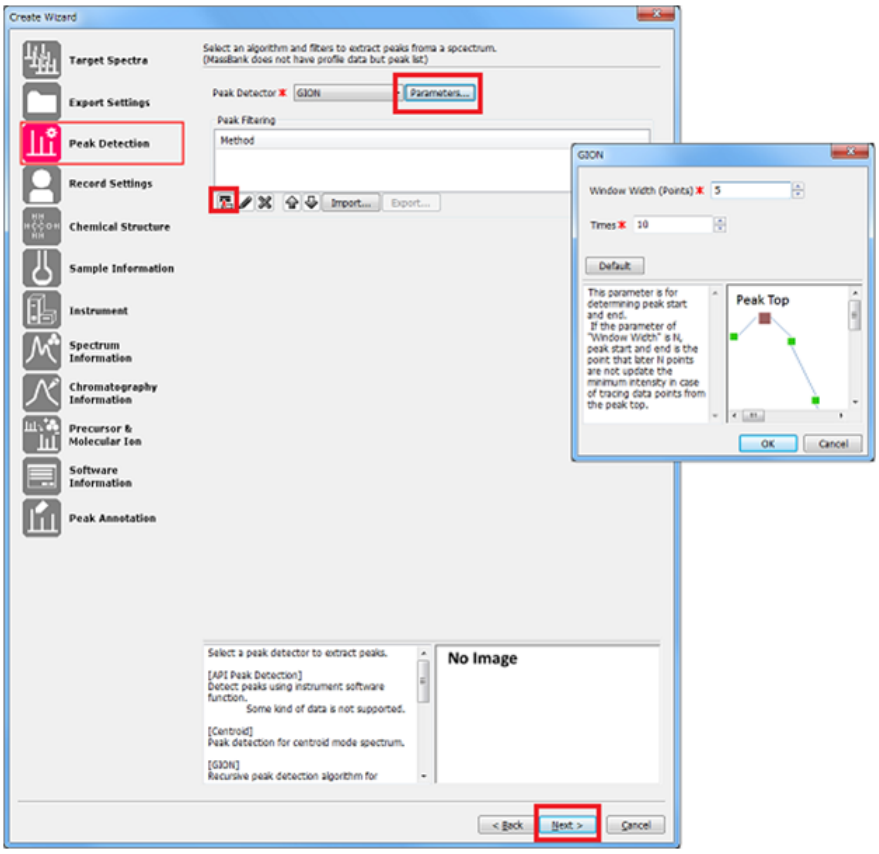


In this page, input uppercase characters in [Prefix], and digits in [Base Num] by specifying the first of the sequential number.

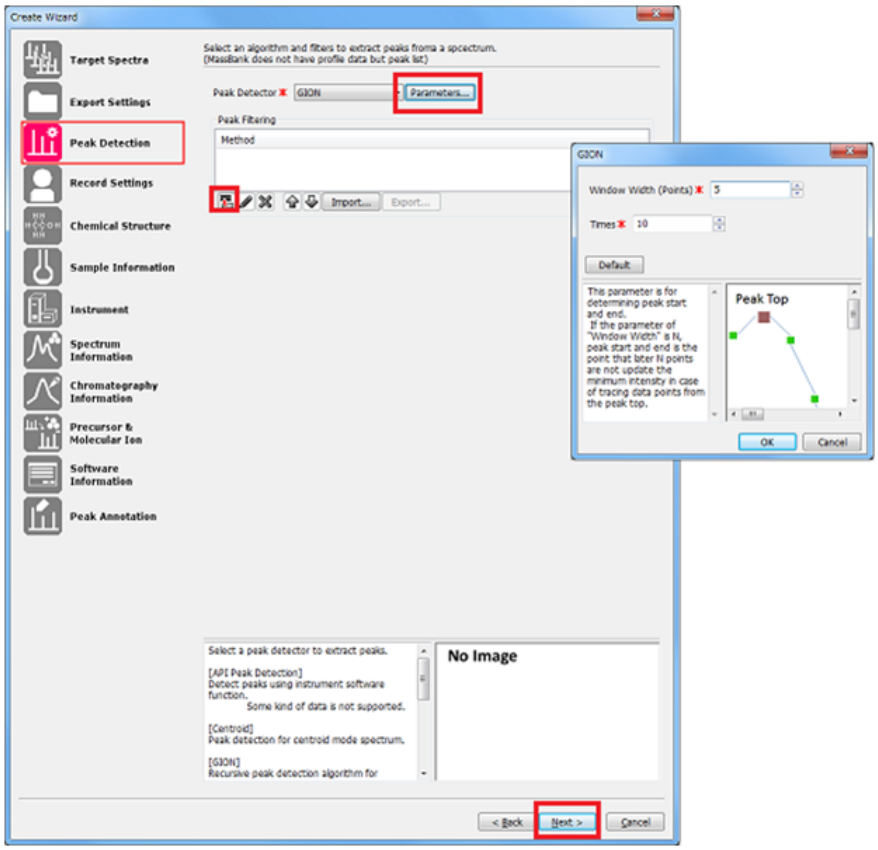
In this example, we input "MSP" in the [Prefix] and specify 0 in the [Base Num]. MSP00000 is assigned as the accession for the first spectrum, MSP00001 is for the second spectrum, and MSP00002, MSP00003, and continuing accessions are assigned to the following spectra in a similar way. In [Folder], output folder is specified. To complete inputting, click [Next].



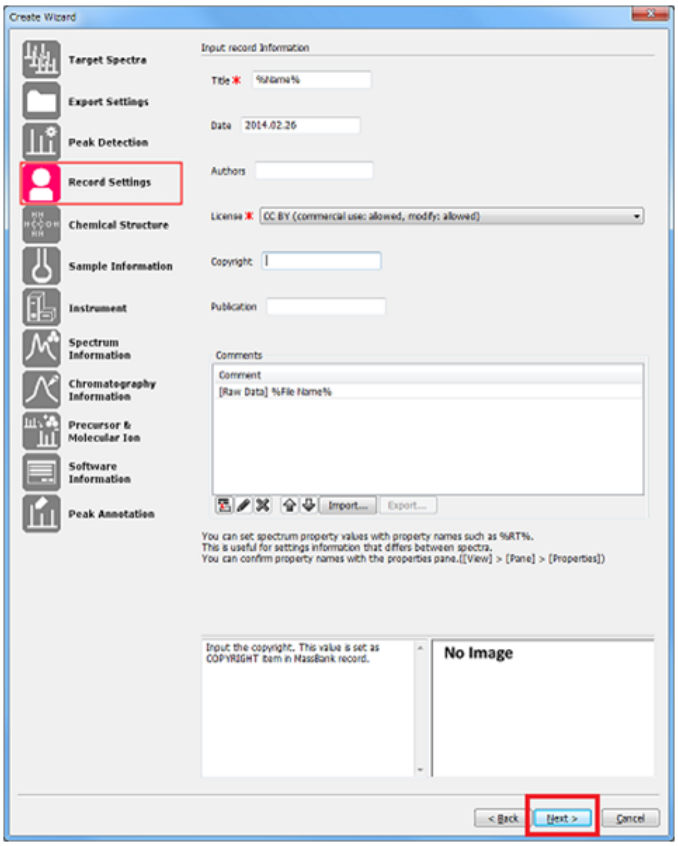
Set peak detection parameters in [Peak Detection] page. In this case, select "GION" in [Peak Detector] menu and click [Parameters].



[GION] opens. Set "5" as [Window Width (Points)] and "10" as [Times]. In case that you want to set peak filters, click [Add] in [Peak Filtering]. After completing to set all parameters, and then click [Next].

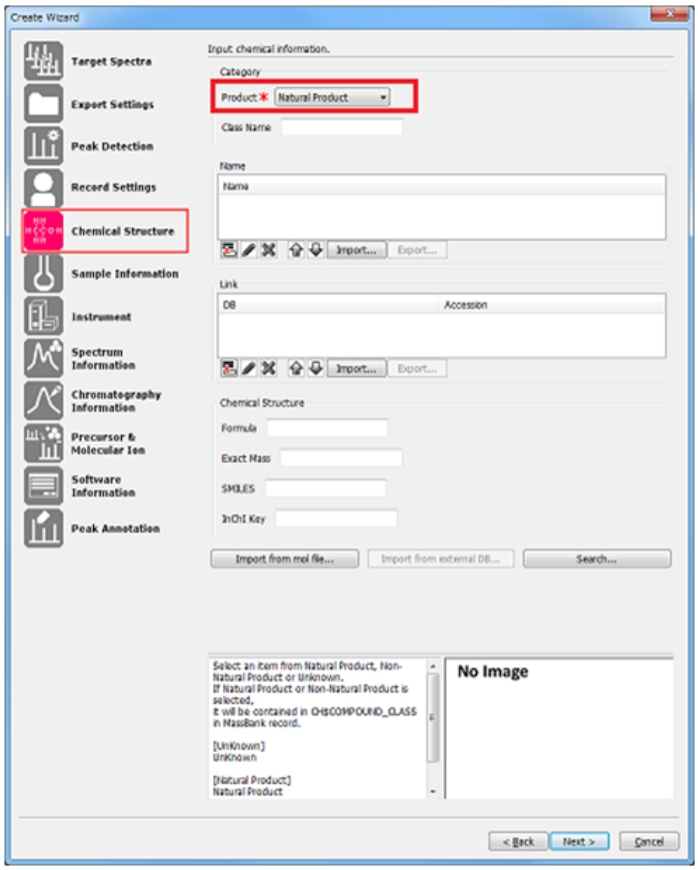


Input basic information about the record: data title, date of the data creation, authors of the data, copyright license for the data, copyright holder, data publication information, and comments. The data title and the copyright license are the "required field" (cannot be omitted). In this example, click [Next] leaving parameters default/blank.

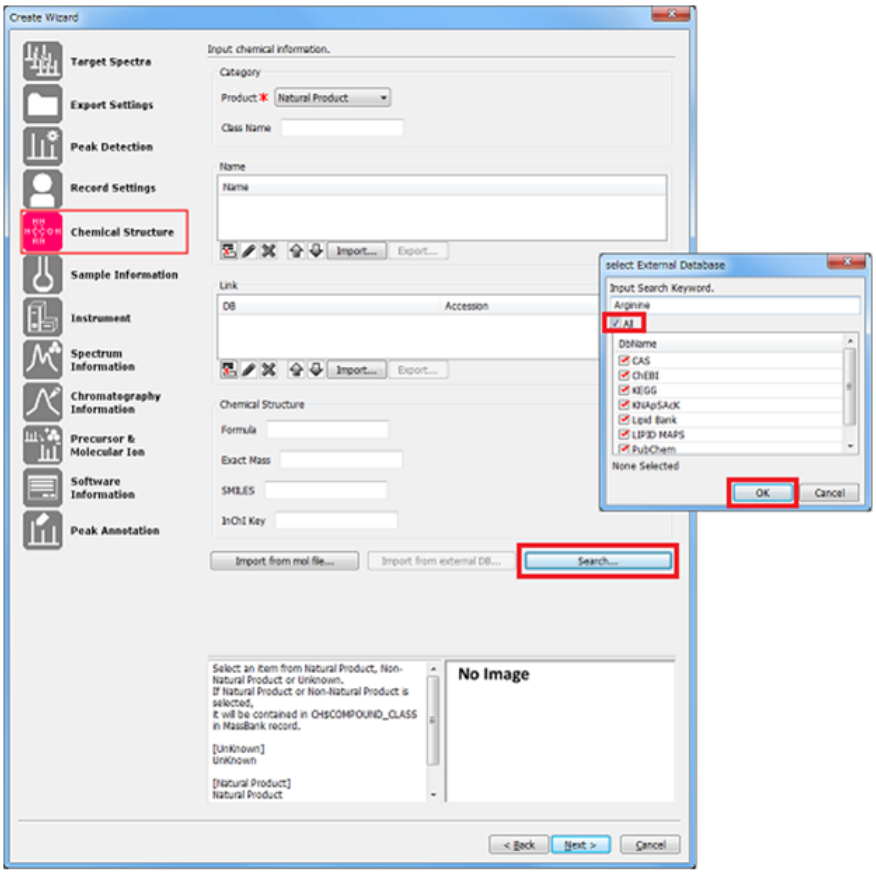


Note: All data stored in public version of MassBank are required to publish under "Creative Commons" license, thus it is necessary to specify the subcategory of Creative Commons for each data entry at the data submission (See [Parameter Hint] for the subcategories of Creative Commons. Also see MassBank manuals for more detailed information).

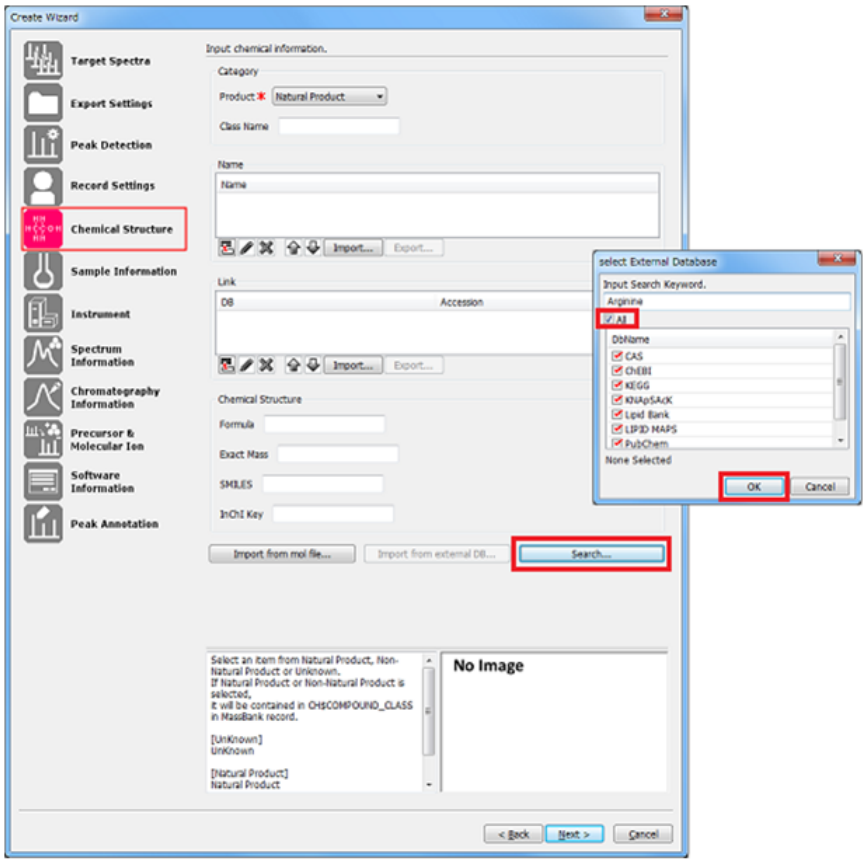
Input the chemical structure information of the substance assigned to the data in the [Chemical Structure] page. First, select "Natural Product" or "Non Natural Product" in the [Product]. This item is a "required field." In this example, select "Natural Product."



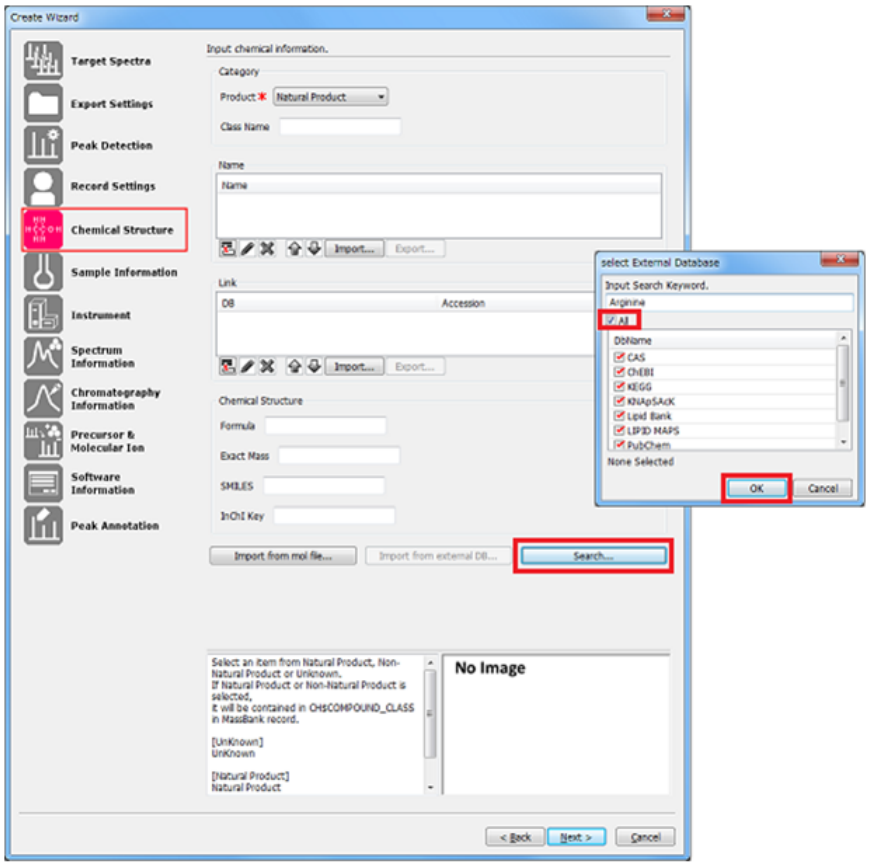
Then, input the chemical compound name, chemical structure information, and the link information to the public databases containing further information about the target compound (hereinafter referred to as "external databases").



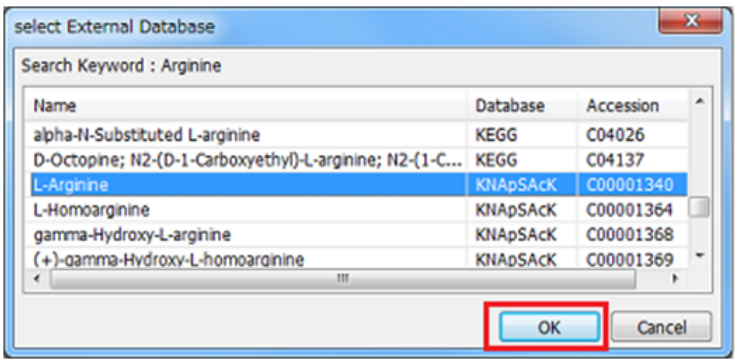
It is necessary to input so many numbers of items that makes the operation cumbersome; hence in this example, we exhibit a recommended approach to input data by extracting information obtained from the external database.



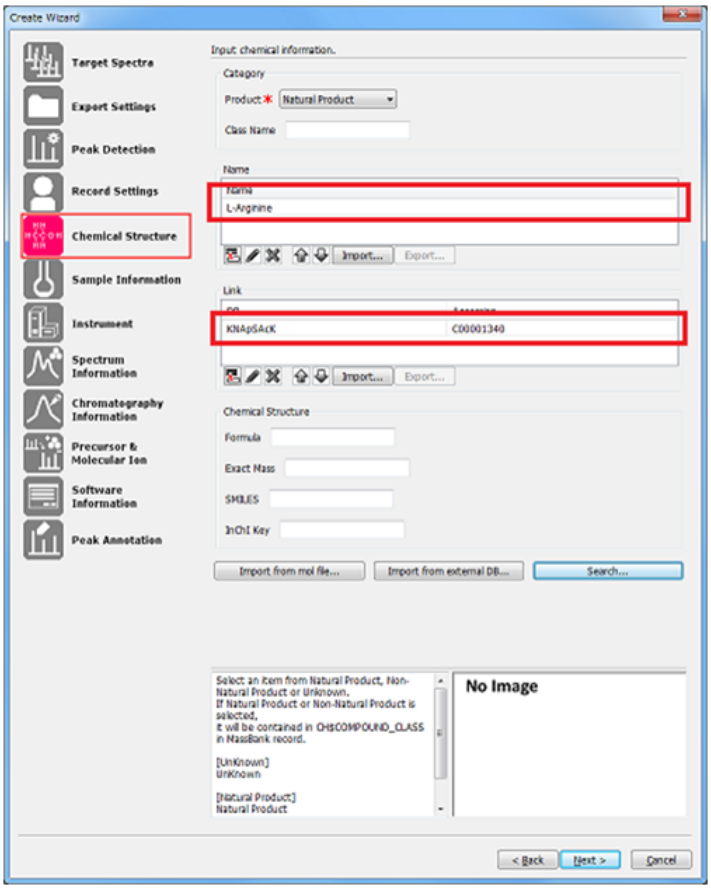
Click [Search] to search the external databases. Then [select External Database] opens. Input "Arginine" as the keyword and check [All] to search against all preset external databases. Set all information, then click [OK].



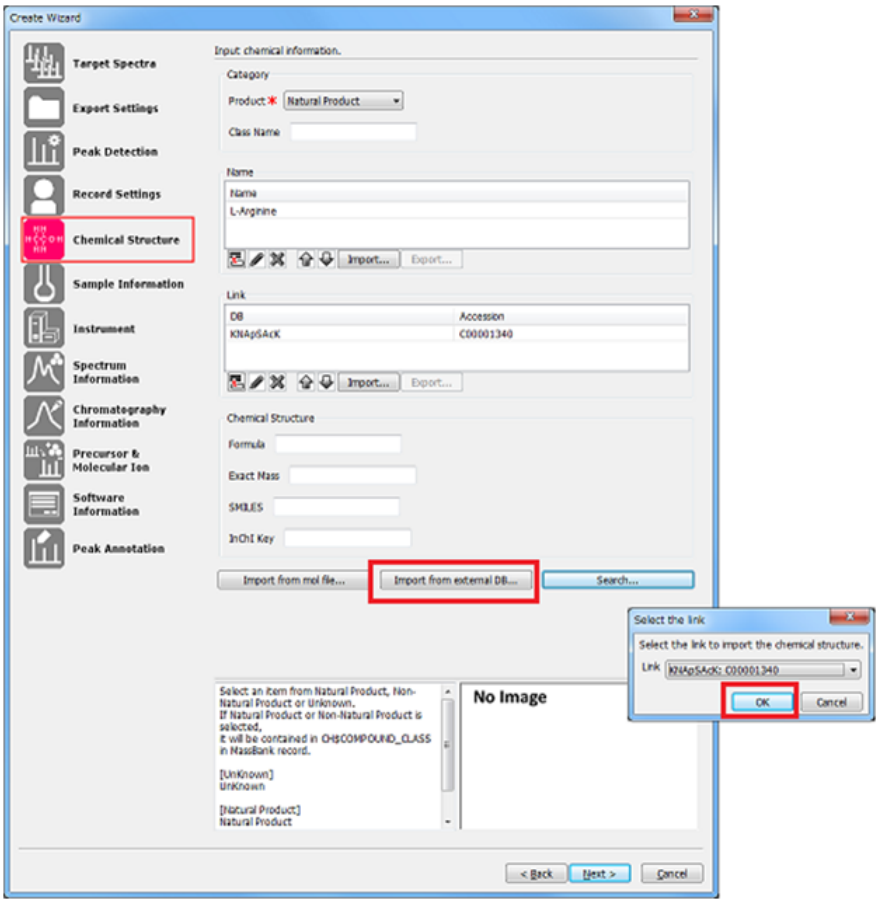
Search results are displayed. In this example, select item "L-Arginine" in "KNApSAcK" whose accession is "C00001340" and click [OK].



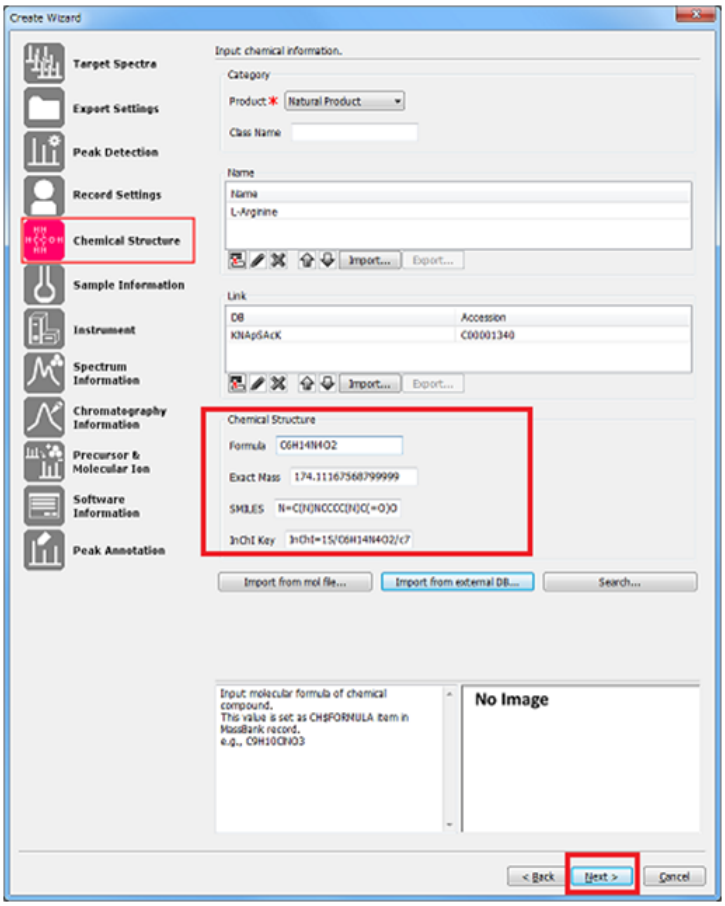
Confirm that the compound name in [Name], and the external database name and the accession information in [Link] are automatically set.



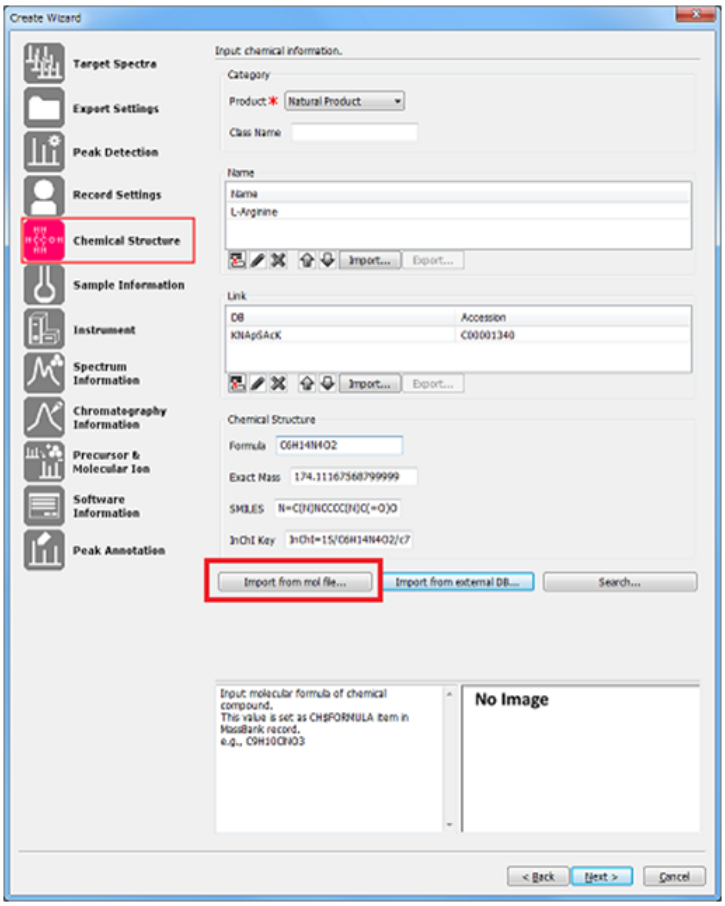
Next, click [Import from external DB] to obtain the chemical structure information. [Select the link] opens. Select "KNApSAcK: C00001340" and click [OK].



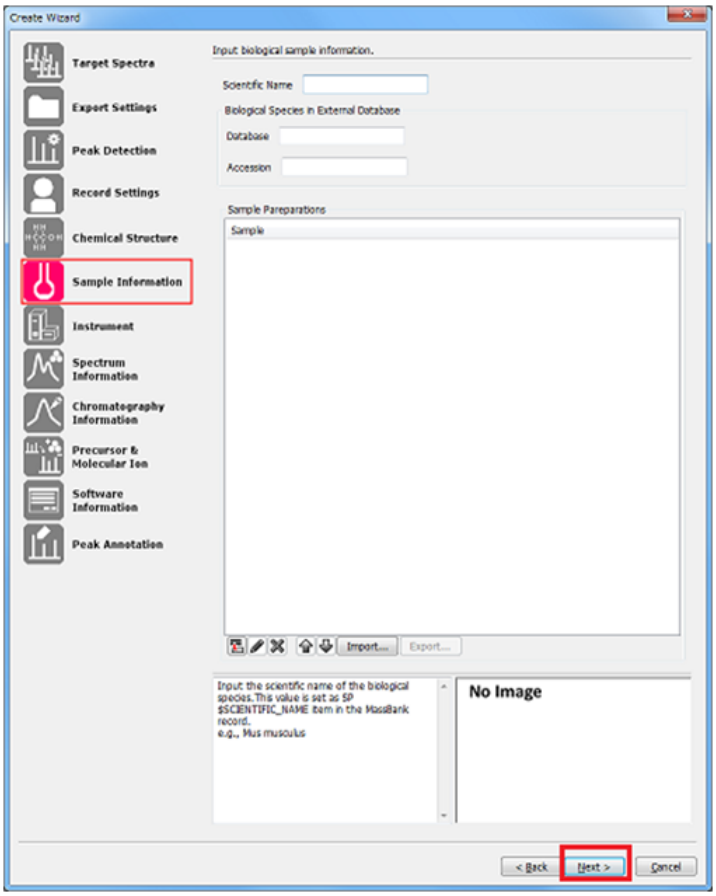
Confirm the structure information is automatically set in [Chemical Structure] area and click [Next].



Note: When have the corresponding .mol file to the target compound, click [Import from mol file] and specify it.



In [Sample Information] page, input the information about the organism from which the sample derived: Refer to [parameter hint].

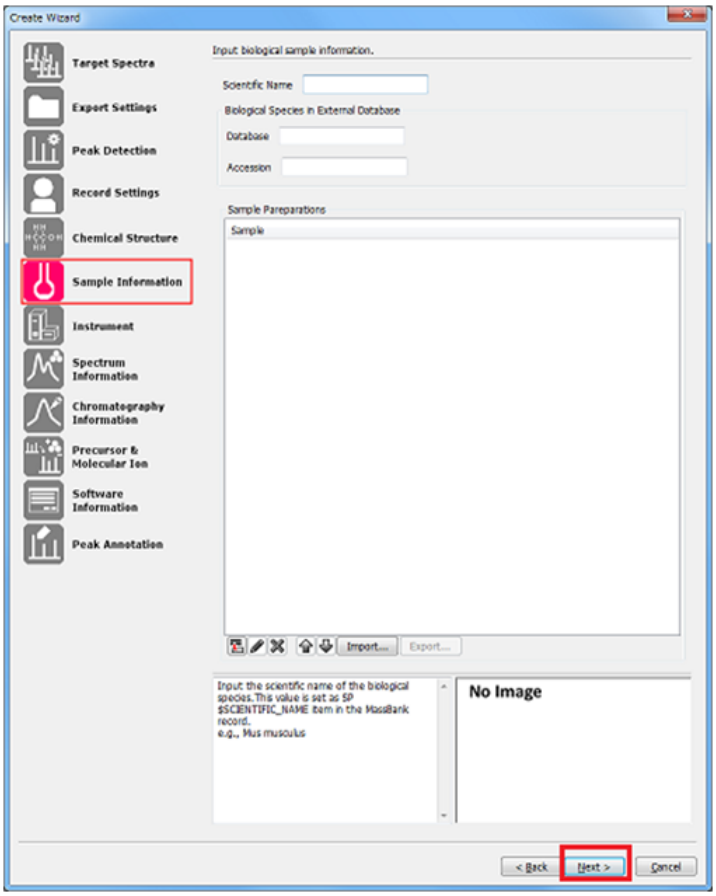


Input the scientific name of the organism in [Scientific Name]. Also input the database name and the accession of the organism in [Biological Species in External Database] area in order to obtain the taxonomy information.

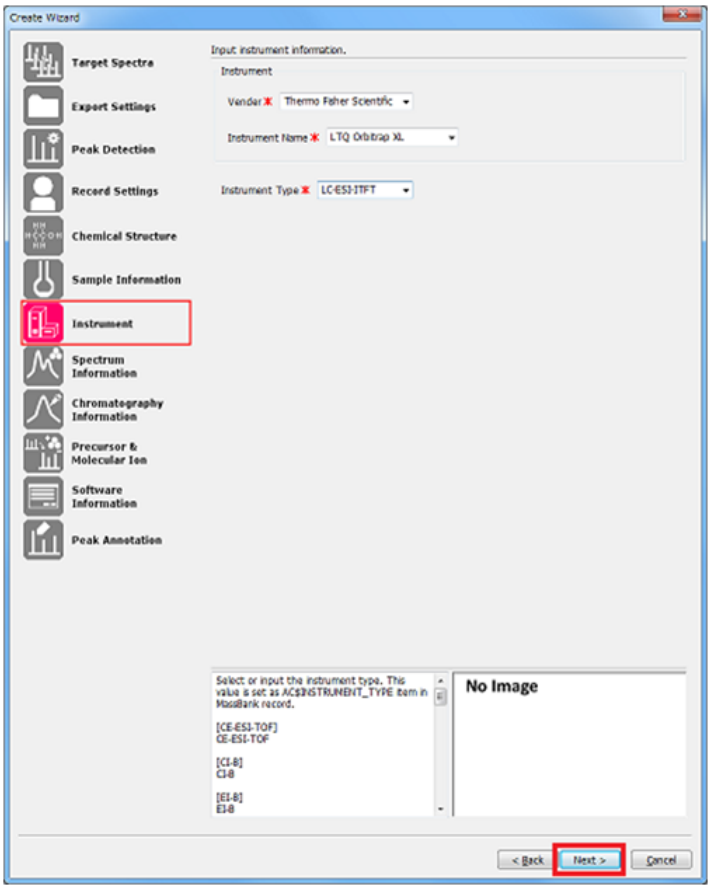
In [Sample Preparations], any information string can be input. e.g. tissue information, sample preparation procedures.

Note: Currently any database name is acceptable for these textboxes in [Biological Species in External Database] area; however, the input database should be one of databases that MassBank system is supporting. See MassBank manuals (two databases are supported at April 1, 2014: NCBI Taxonomy, which is a de facto standard, and KNApSAcK).

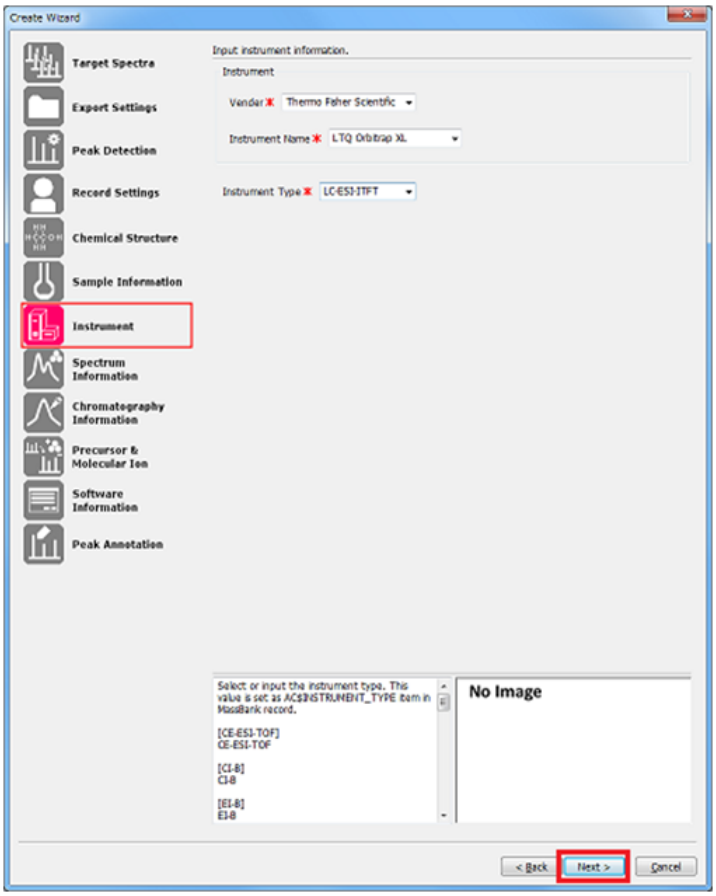
Note that no items should be input in case of non-organism-derived samples. Therefore, no item is a "required field." In this example, click [Next] leaving items blank.



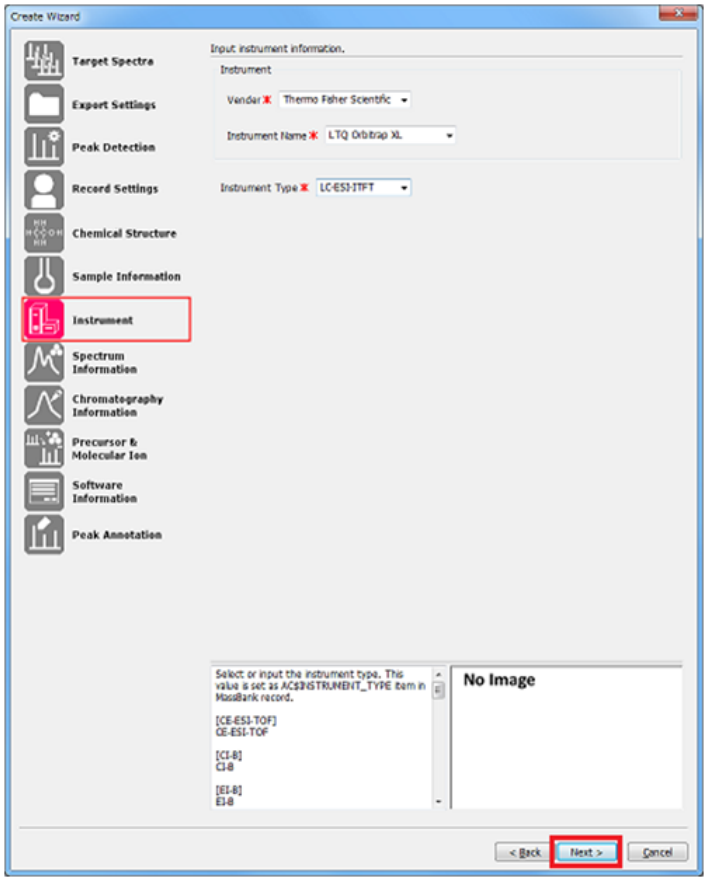
Input mass spectrometer information. No items can be omitted. In this example, select as follows:



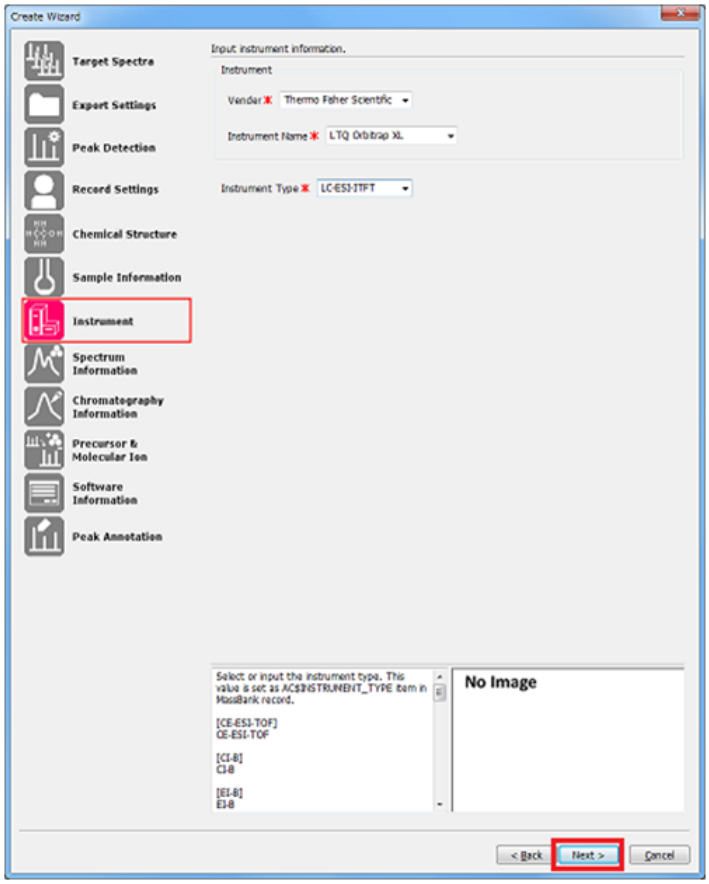
・"Thermo Fisher Scientific" as [Vender] menu,



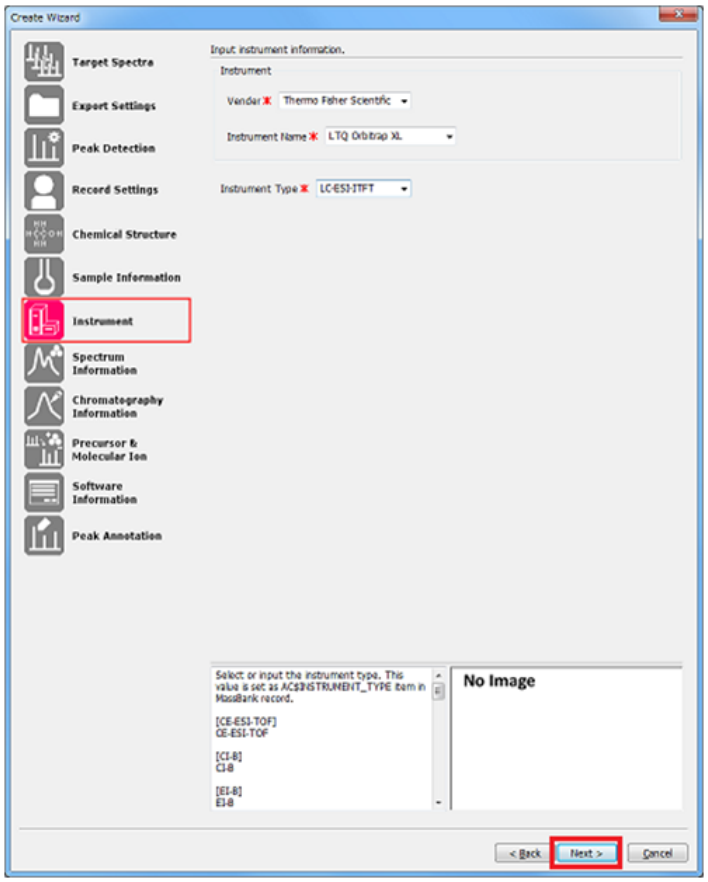
・ "LTQ Orbitrap XML" as [Instrument Name] menu, and



・"LC-ESI_IT FT" as [Instrument Type] menu.

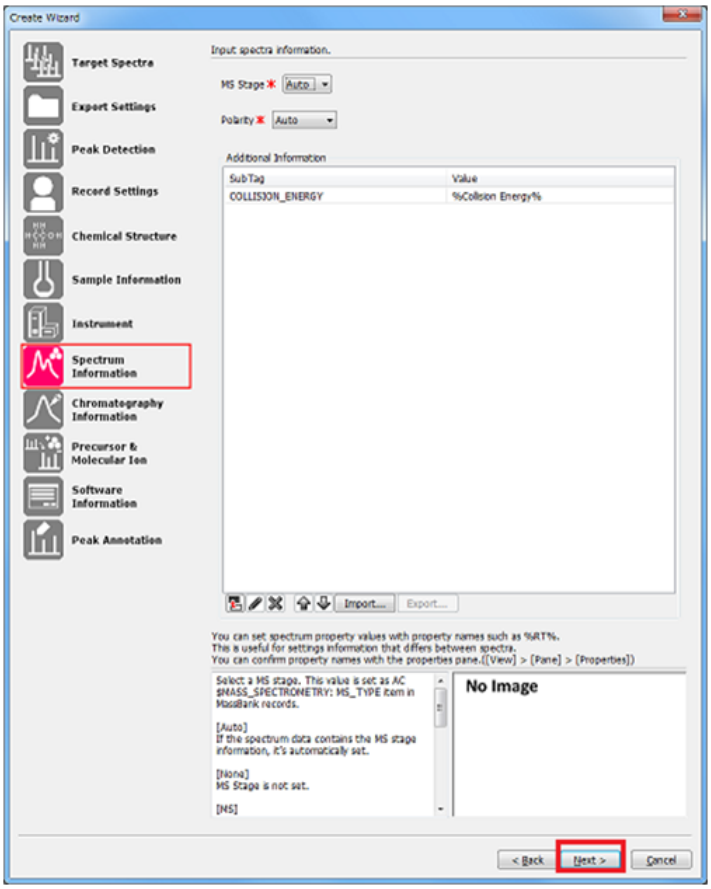


After completing to set all parameters, then click [Next].

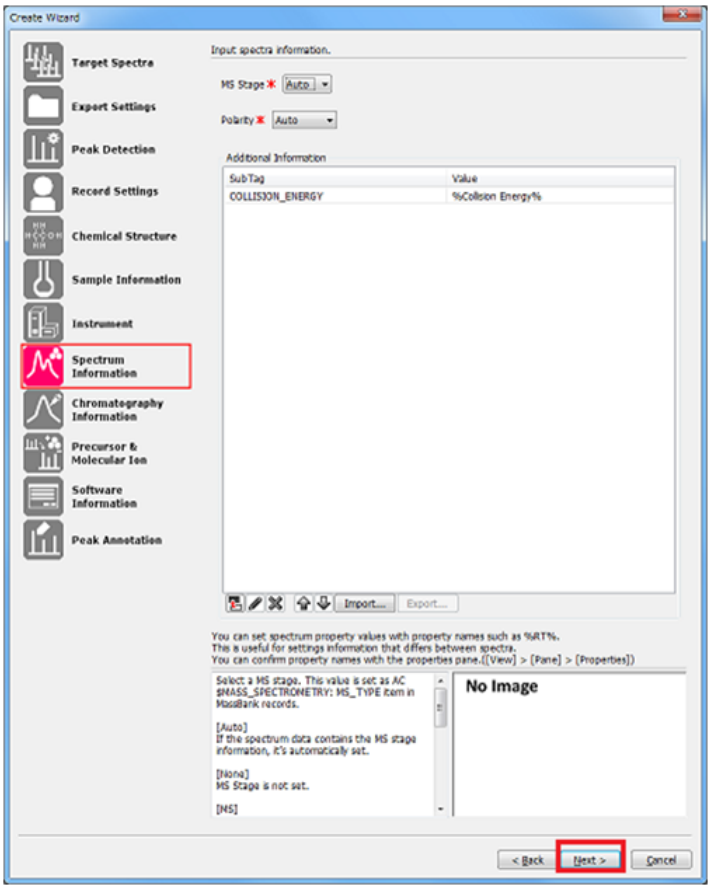


Note: The information of [Vendor] and [Instrument Name] are extracted automatically from the metadata in the spectrum data. Note that there is no guarantee that the preset instrument name is always correct because hundreds of names are currently used in the instruments. Verify the name.

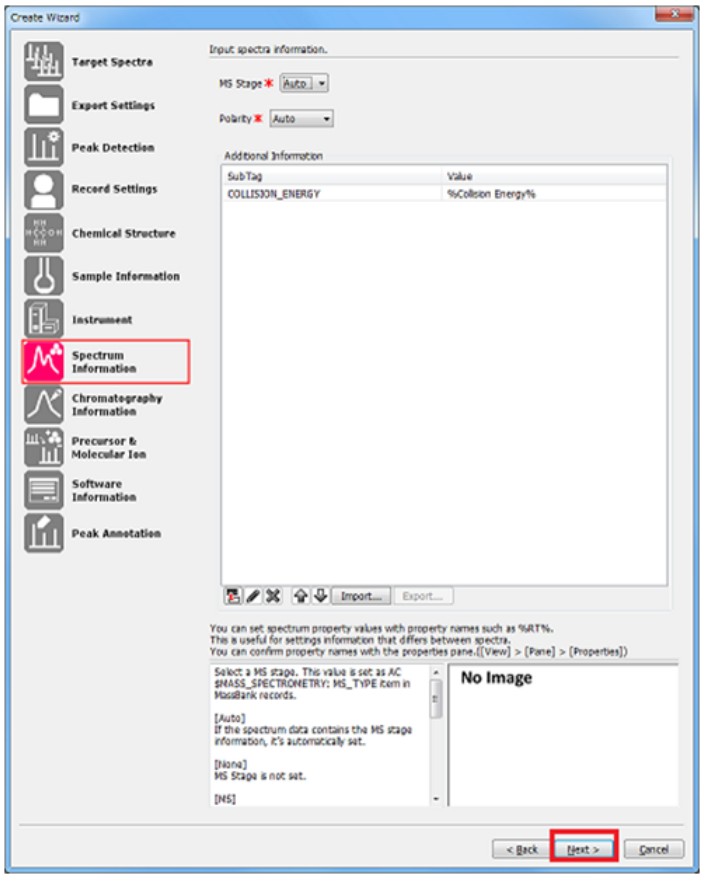
Set spectrum information in [Spectrum Information] page.



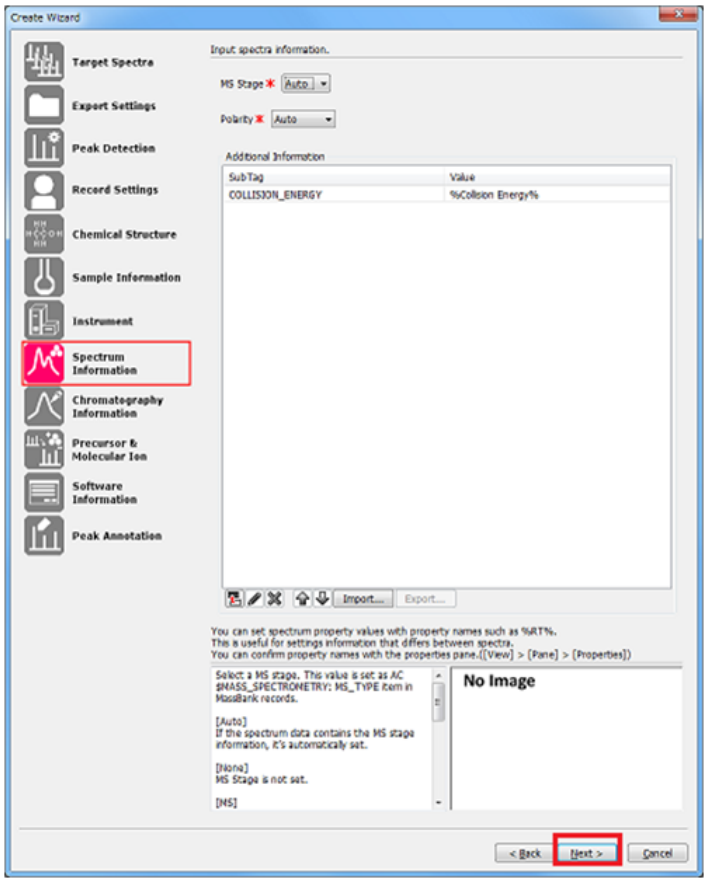
Select the ion generation in [MS Stage] menu and the ion mode in [Polarity] menu. These two items are "required fields"; however, by setting [Auto], these data can be automatically extracted from the metadata in the spectrum data.



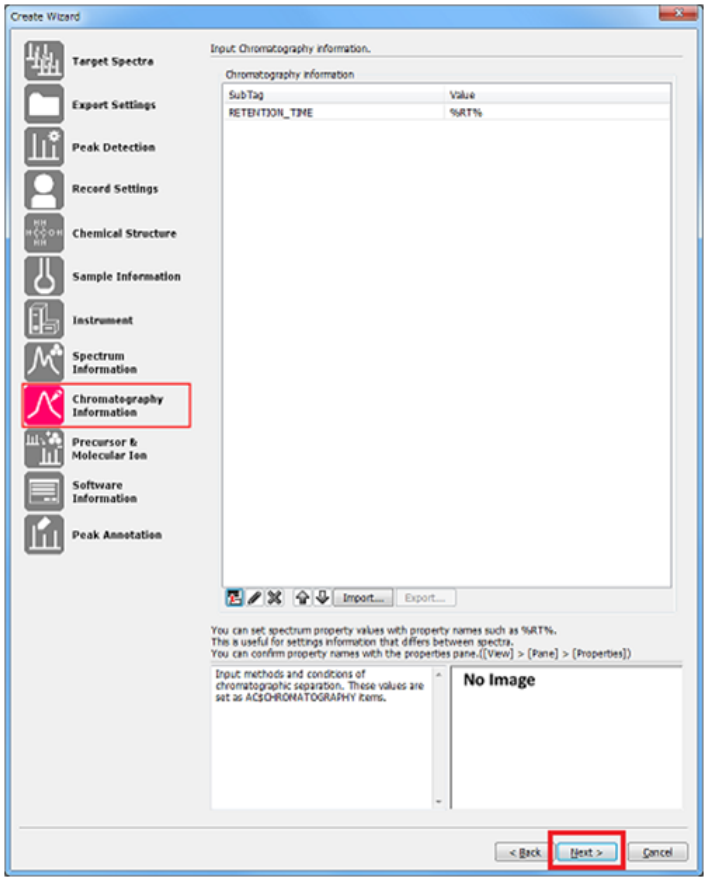
Other additional information can be input in the [Additional Information] area. "COLLISION_ENERGY" information is automatically extracted from the metadata.



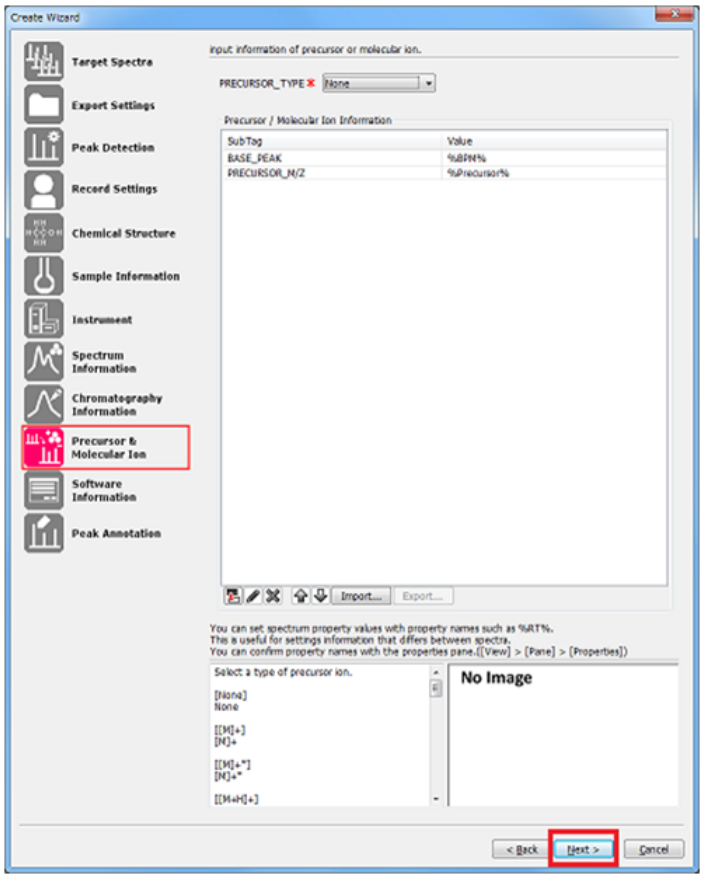
In this example, click [Next] leaving parameters default.



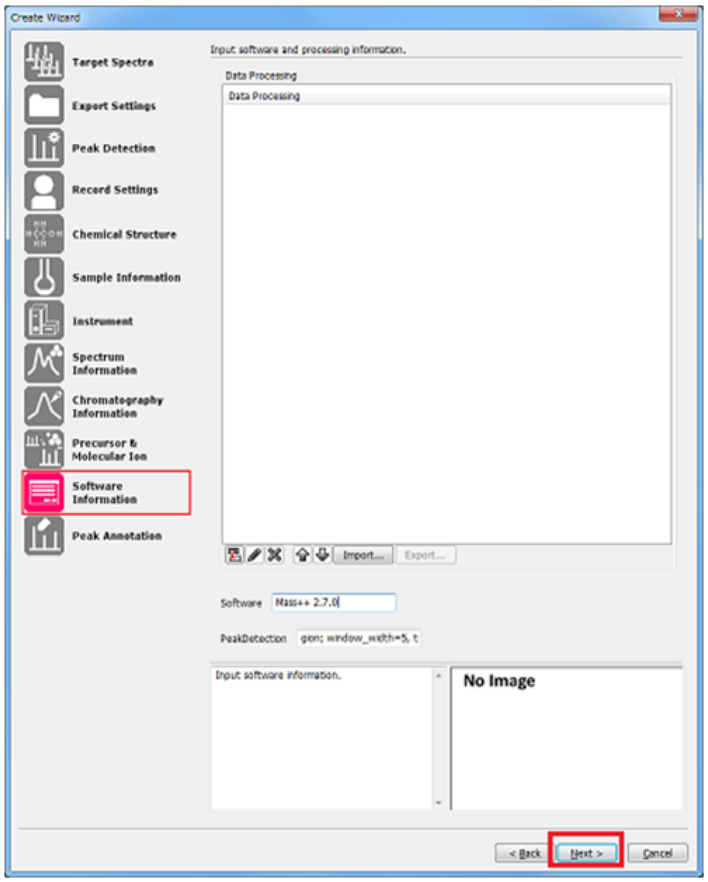
Input the chromatogram information in [Chromatography Information] page. In this example, click [Next] leaving parameters default.



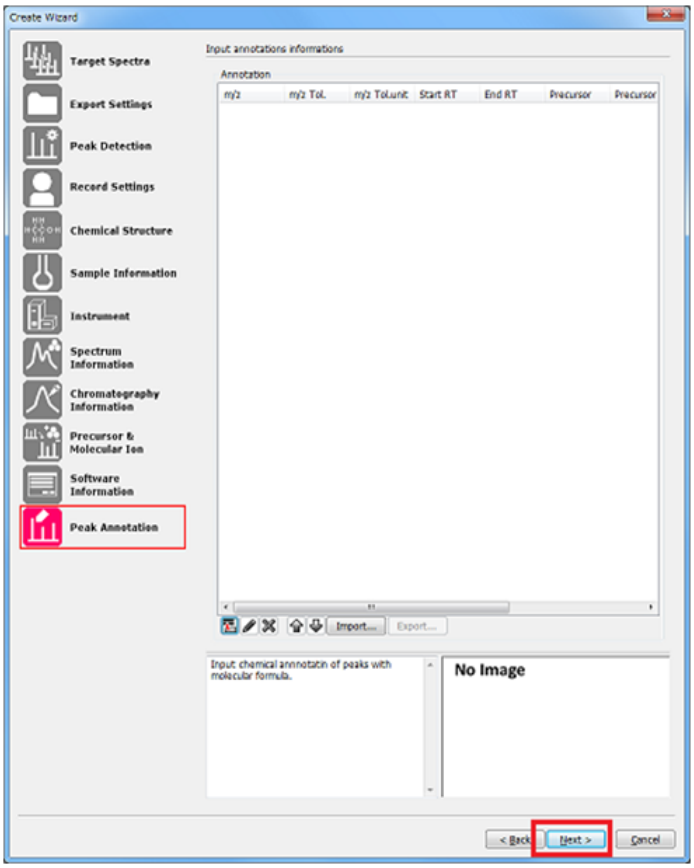
Input the information of precursor and molecular ions in [Precursor > Molecular Ions] page. Select precursor ion species. This item is a "required field." In this example, click [Next] leaving parameters default.



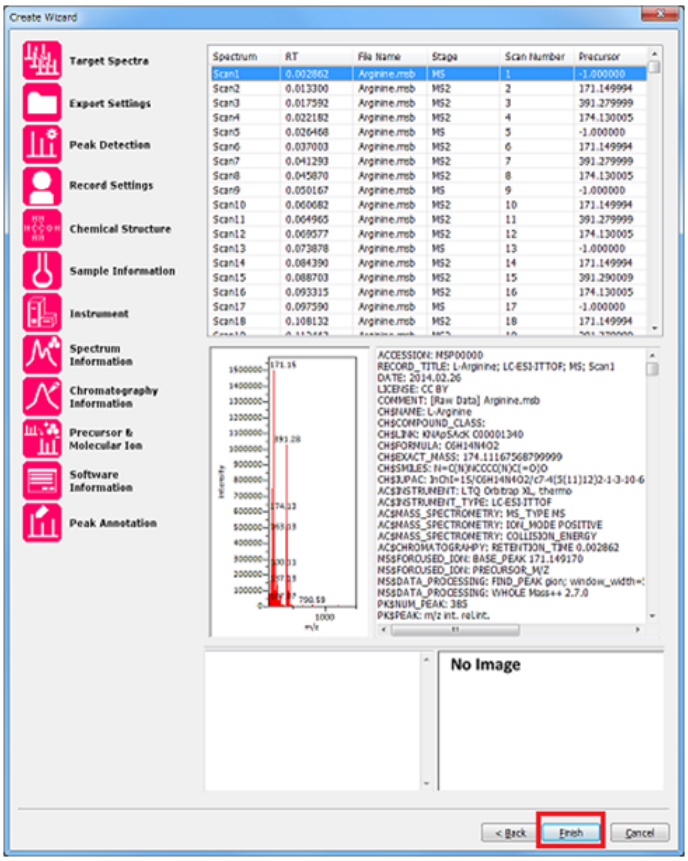
Input the utilized software information in [Software Information] page. In case that spectrum data were preprocessed before the peak detection, input related information in [Data Processing]. The initial value of [Software] is "Mass++" and the initial values of [PeakDetection] are the parameters input in [Peak Detection] page. If necessary, re-write them directly. In this example, click [Next] without editing parameters.



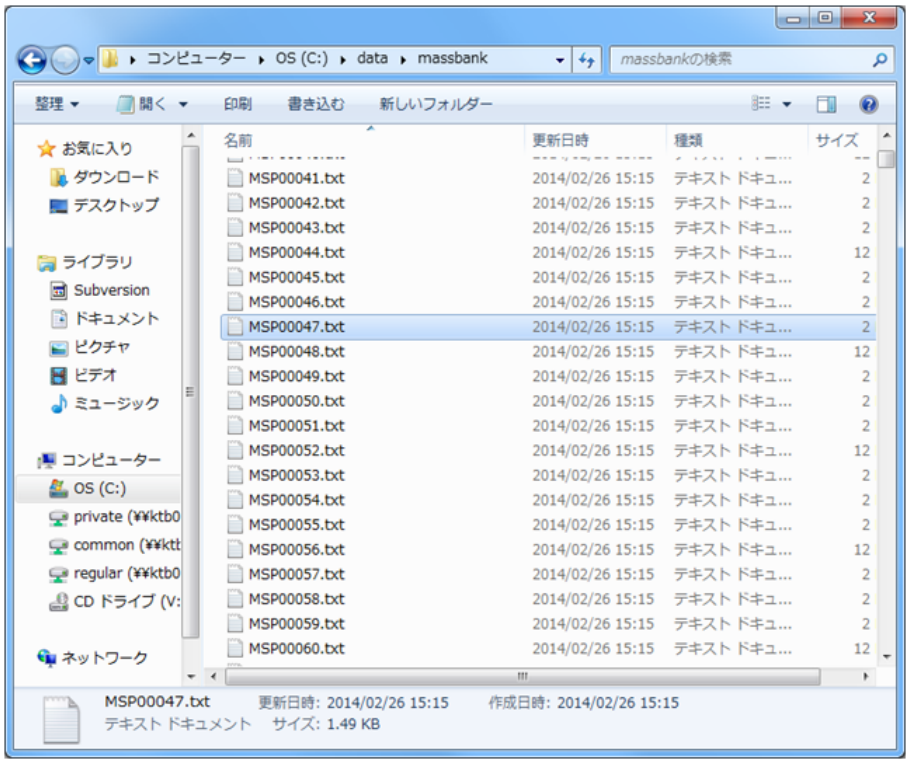
Input the chemical annotation information of peaks in [Peak Annotation] page. In this example, click [Next] leaving parameters blank.



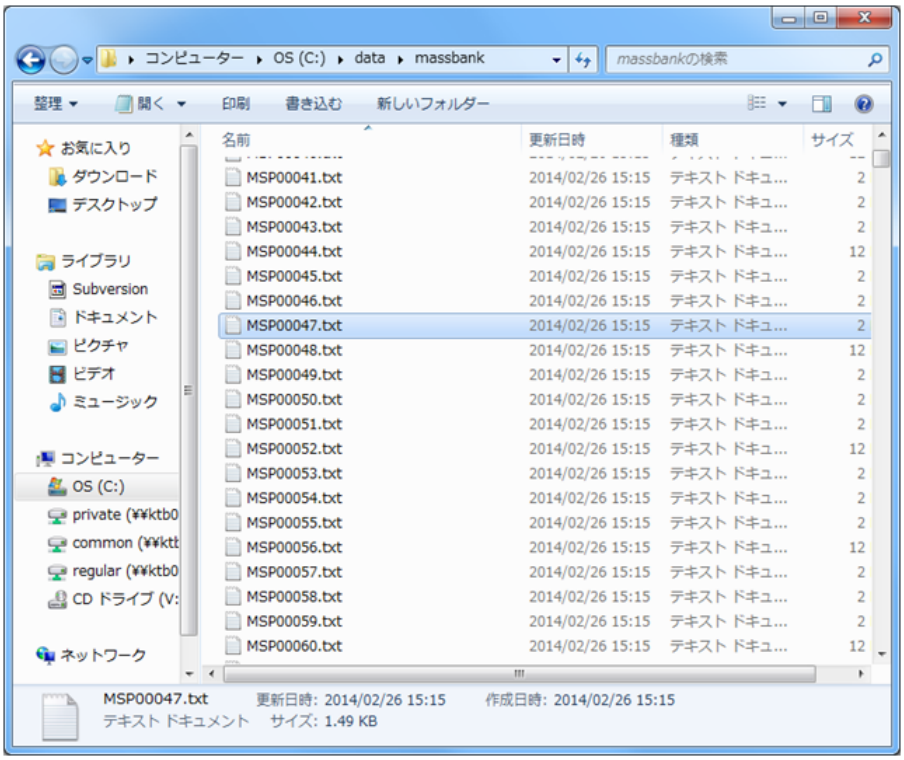
In last page, the exporting record information is listed. Click to select a row in the list on the upper side of the page, and you can preview the exporting record. Click [Finish] when properly done.



The mouse cursor is turned to the "wait cursor" during creating spectrum records. A MassBank record file is a text file, of which name is formatted in "(MassBank Accession).txt."



Open the folder specified in [Export Settings] page, and confirm the record files. Use the MassBank attached "Record Editor" and "Administration Tool" of MassBank via a web browser for the following operation. See the following note and the MassBank manuals.



Note: This "Create MassBank record file" function has been developed under the cooperation with the MassBank development team, though it cannot replace "Record Editor" completely. For data submission, multiple files are zipped, and the Record Editor operates this compression.

Note: Therefore, in order to create a MassBank record for submission, after the translation with Mass++, it is necessary to open the generated data file to examine by "Record Editor," continue the operation if necessary, and complete the file creation for data submission.