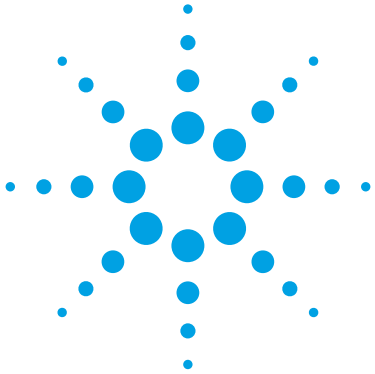


Agilent G2721AA/G2733AA Spectrum Mill MS Proteomics Workbench

Familiarization Guide

Agilent Technologies

Notices

Manual Part Number

G2721-90052

Edition

Eighth edition, October 2016

Agilent Technologies, Inc.
5301 Stevens Creek Blvd.
Santa Clara, CA USA 95052

Software Revision

This guide is valid for the B.06.00 revision or higher of the Agilent G2721AA/G2733AA Spectrum Mill MS Proteomics Workbench software, until superseded.

Warranty

The material contained in this document is provided "as is," and is subject to being changed, without notice, in future editions. Further, to the maximum extent permitted by applicable law, Agilent disclaims all warranties, either express or implied, with regard to this manual and any information contained herein, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Agilent shall not be liable for errors or for incidental or consequential damages in connection with the furnishing, use, or performance of this document or of any information contained herein. Should Agilent and the user have a separate written agreement with warranty terms covering the material in this document that conflict with these terms, the warranty terms in the separate agreement shall control.

Technology Licenses

The hardware and/or software described in this document are furnished under a license and may be used or copied only in accordance with the terms of such license.

Restricted Rights Legend

U.S. Government Restricted Rights. Software and technical data rights granted to the federal government include only those rights customarily provided to end user customers. Agilent provides this customary commercial license in Software and technical data pursuant to FAR 12.211 (Technical Data) and 12.212 (Computer Software) and, for the Department of Defense, DFARS 252.227-7015 (Technical Data - Commercial Items) and DFARS 227.7202-3 (Rights in Commercial Computer Software or Computer Software Documentation).

Safety Notices

CAUTION

A **CAUTION** notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in damage to the product or loss of important data. Do not proceed beyond a **CAUTION** notice until the indicated conditions are fully understood and met.

WARNING

A WARNING notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in personal injury or death. Do not proceed beyond a WARNING notice until the indicated conditions are fully understood and met.

In this Guide...

The *Familiarization Guide* presents step-by-step exercises to analyze example Agilent Q-TOF data files with the Spectrum Mill MS Proteomics Workbench. In this guide you process protein digest data to learn how to:

- Extract spectra from MS/MS data sets
- Run protein database searches
- Review and validate search results
- Use *de novo* sequencing
- Search and identify peptides with variable and unknown modifications
- Automate workflows

1 Normal Workflow – Interactive Processing of MS/MS Data

These exercises illustrate the basic processing steps you use for many types of samples. A basic workflow is useful for identification of most proteins in the sample with one round of search and autovalidation. You also do exercises to compare data from two yeast cell states.

2 Iterative Workflow – Alternative Processing of MS/MS Data

In these exercises, you process the yeast data with iterative searching and autovalidation, and further identify proteins with Spectrum Summary and *de novo* sequencing. This workflow is useful for more thorough searching and identification.

3 Unknown Modifications

These exercises show how to search MS/MS spectra to identify unsuspected modifications. The chapter illustrates this “mass gap” search with an Agilent Q-TOF/OGE data file. This OGE data comes from a single fraction of an OFFGEL fractionation.

4 Workflow Automation

These exercises show how to automate a typical data analysis for MS/MS data files from protein digests, from data extraction

through results summary, using the parameter files you create in Chapter 1.

5 Interactive Processing of MS-Only Data

These exercises show how to process Agilent TOF and Q-TOF .d files, using an Agilent Q-TOF .d file as an example.

Content

1 Normal Workflow – Interactive Processing of MS/MS Data

This chapter provides familiarization exercises to process an example set of MS/MS data files with a normal workflow using the Agilent Spectrum Mill MS Proteomics Workbench. This material elaborates on the first section of Chapter 1 of the companion *Spectrum Mill MS Proteomics Workbench Application Guide*. Use the normal workflow to identify most of the spectra in your sample with a single search and autovalidation.

This chapter is intended as an introduction to the basic operation of the software; you are encouraged to customize these steps for your samples and studies.

The exercises here assume that the Spectrum Mill workbench has already been installed on your server, the SwissProt database has been downloaded and indexed, and the server is ready to go. If this has not yet been done, see the

Installation Guide. The exercises also assume that you have set up your PC client, transferred data files to the server, and become familiar with the software. If this has not yet been done, see the *Quick Start Guide*.

Although the exercises in this chapter use Agilent Q-TOF data files, the general workflow applies broadly for any MS/MS data files. However, the specific settings, such as mass tolerance, may vary for MS/MS data from other types of instruments.

After you have worked through the exercises in this chapter and you have saved parameter files as described in the exercises, we recommend that you use the Spectrum Mill process automation tools to set up a workflow with the parameter files created here to automate the data processing steps. For more details see Chapter 4, Workflow Automation.

This chapter also describes the tasks you use for “interactive automation”, where you click commands, such as Extract, Start Search, Validate Files and Summarize, one right after the other, on their respective pages. The commands are then executed in that order after other commands placed in the queue before them have been executed. See Chapter 4, Workflow Automation, for an exercise on interactive automation.

Before you begin

The instructions in these exercises use the data directly under the **msdataSM\ExampleData** folder, which requires that you remove all prior results before doing a task.

To avoid having to remove prior results or for your convenience, transfer the **ExampleData** file folders to a new folder you create called **New_Examples**. Create this folder under the folder you created as the repository for your data files (**msdataSM\New_Examples**). Note that folders and file names cannot contain spaces.

1 From your PC client, open the **msdataSM** share to the server.

2 Create a folder and name it **New_Examples**.

3 Copy and paste the **\ExampleData\Agilent\Q-TOF\yeast\y1** and **y2** folders into **New_Examples.**

4 Copy and paste the **\ExampleData\Agilent\Q-TOF\OGE** folder into **New_Examples.**

Task 1. Search for proteins/peptides and autovalidate high-quality results

After you open Spectrum Mill MS Proteomics Workbench, you see the Spectrum Mill home page.

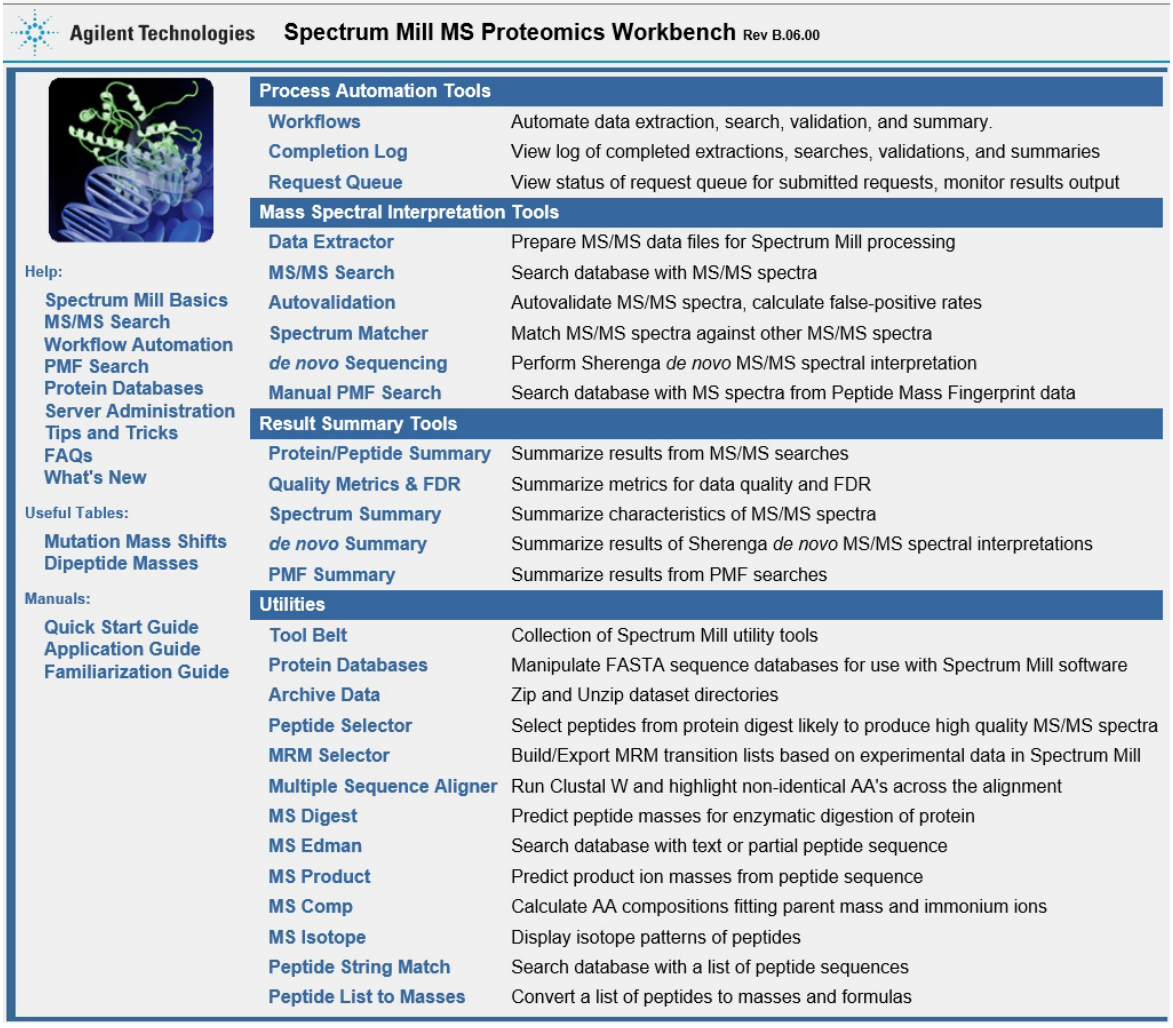


Figure 1 Spectrum Mill home page

Exercise 1. Extract the spectra

The Spectrum Mill Data Extractor preprocesses raw data files to extract high-quality spectra for database searches. The Data Extractor automatically detects which type of raw file you have and then invokes the appropriate extraction program (provided that it has been purchased and installed on your server).

The MS/MS raw file data extractors extract and merge nearby MS/MS spectra from the same precursor ion. They optionally apply MS/MS similarity criteria prior to merging scans, to avoid merging closely eluting or co-eluting isobaric peptides. For Agilent ion trap and Thermo Fisher Scientific *.raw ion trap data, the extractors optionally merge MS^2 and MS^3 scans from the same precursor.

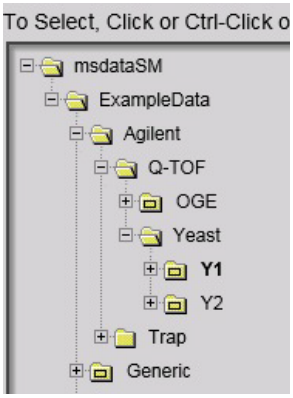
The extractors do these tasks:

- Assign precursor charges where possible
- Centroid the MS/MS spectra
- Calculate spectral features
- Filter MS/MS spectra by quality
- Calculate extracted ion chromatograms (EICs) for the intervening MS precursor scans. The latter are used for quantitation.

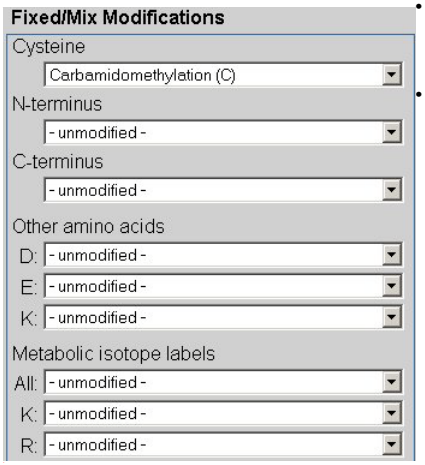
Exercise 1. Extract the spectra

Steps	Detailed Instructions	Comments
1 Navigate to the Data Extractor page.	• From the Spectrum Mill home page, click **Data Extractor**.	

Exercise 1. Extract the spectra (continued)

Steps	Detailed Instructions	Comments
2 Select the **msdataSM\ ExampleData\Agilent\Q-TOF\ yeast\y1** data directory or the y1 folder from the **New_Examples** folder (see above).	**a** Under Data Directories click **Select...**. **b** Expand the directory tree and click **y1** to select that data file directory.  **c** Make sure that the name of the **y1** directory changes to a **bold** font. This indicates that it has been selected. **d** Click **OK.** The name of the data directory appears as in Figure 2 on page 14.	• Directories are identified by different types of icons: ◦ Plain folders indicate directories that do not have data files directly beneath them. ◦ Folders with rectangles indicate data directories. ◦ Folders with line spectra (bar graphs) indicate data files (which for Agilent files are also directories in themselves). • Try clicking the names of each type to see which turn **bold**, indicating that they are selectable. • The software remembers your data file selection when you go to other Spectrum Mill pages. • If you mark the **Make Default** check box in this dialog, the software remembers your data directory even after you close your web browser. • To select multiple folders, press **Ctrl-Click**.

Exercise 1. Extract the spectra (continued)

Steps	Detailed Instructions	Comments
3 Choose the appropriate cysteine modification. • Carbamidomethylation is the default when you first install the Spectrum Mill workbench. • So if Carbamidomethylation is already the selection, skip this step.	**a** Click the **Choose...** button near the middle of the page.   **b** Click **OK.** The name of the modification appears as in Figure 2 on page 14.	• To view details about the modifications that are currently available on your server, click the **Details** button at the bottom right of the **Choose Modifications** dialog. • For more information about choosing modifications, see the online help. • Your system administrator can configure custom modifications.
4 Set parameters as shown in Figure 2 on page 14.	**a** Next to Retention time & m/z tolerance, type `30` sec. **b** If necessary, clear the check box for **Maximize CPUs**. Because this is a single data file, **Maximize CPUs** is not effective. **c** Examine the items in red text carefully, since these are the ones you may need to change when you process your own samples. **d** Click a blue section divider bar to display help for that section of the page.	• You use the **Sequence tag length** to filter out noisy spectra. This is the longest sequence of amino acids that is represented by the fragments in a spectrum. • If you set this to **>1**, you ensure that all possible good spectra are extracted. You have opportunities to set more stringent requirements later when you do the database search. Change the setting to **>0** if you expect post-translational modifications.

Exercise 1. Extract the spectra (continued)

Steps	Detailed Instructions	Comments

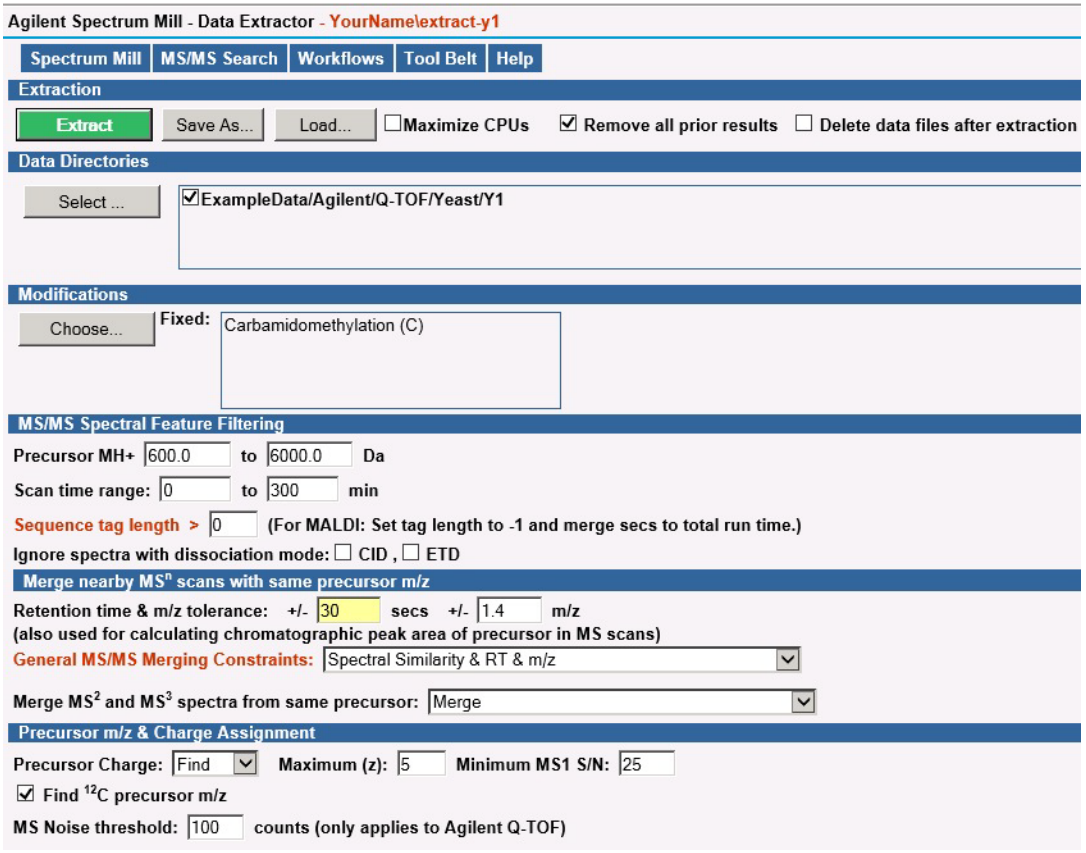
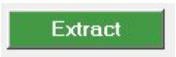
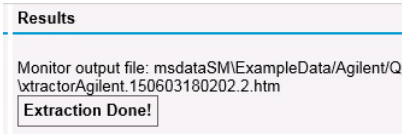


Figure 2 Settings for Data Extractor

Exercise 1. Extract the spectra (continued)

Steps	Detailed Instructions	Comments
5 Save the settings as a parameter file to a folder that you create. • Use a name to distinguish it from the folders of others doing these exercises. For example, use YourName. • Name the file, extract-y1.	**a** Click **Save As.** **b** Next to the **Folder** list, click the **New folder** icon. **c** For **New folder name**, type your own name. and click **OK**. **d** Next to **Name**, type **extract-y1**. **e** Click **Save**. Note that the name of the parameter file appears in red near the top of the browser window (Figure 2).	• A parameter file lets you reuse the settings later, either interactively or in an automated workflow. • A folder name must contain no spaces. • If you forget to save the parameter file, you can find this file under the Examples folder to use for the workflow automation exercises.
6 Start the extraction and monitor results.	**a** Click **Extract**. **b** In the **Results** area to the right of the Data Extractor page, click the **Monitor Results** link. **c** Scroll to the *top* of the **Results** area to see the message that indicates the extraction is finished.  	• The command is automatically placed in the queue for execution after other tasks in the queue have been executed. • The Data Extractor processes all files in the folder. • Spectra are extracted and written to an mzXML file in the folder. • Extraction time varies depending on the number and size of the files. • If you want to stop the extraction, click the red **Stop Extraction PID: xxx** link at the top of the **Results** section. Also go to the Tool Belt, click Stop process and make sure there are no other processes running. If there are, stop them. See the *Application Guide* for further instructions.

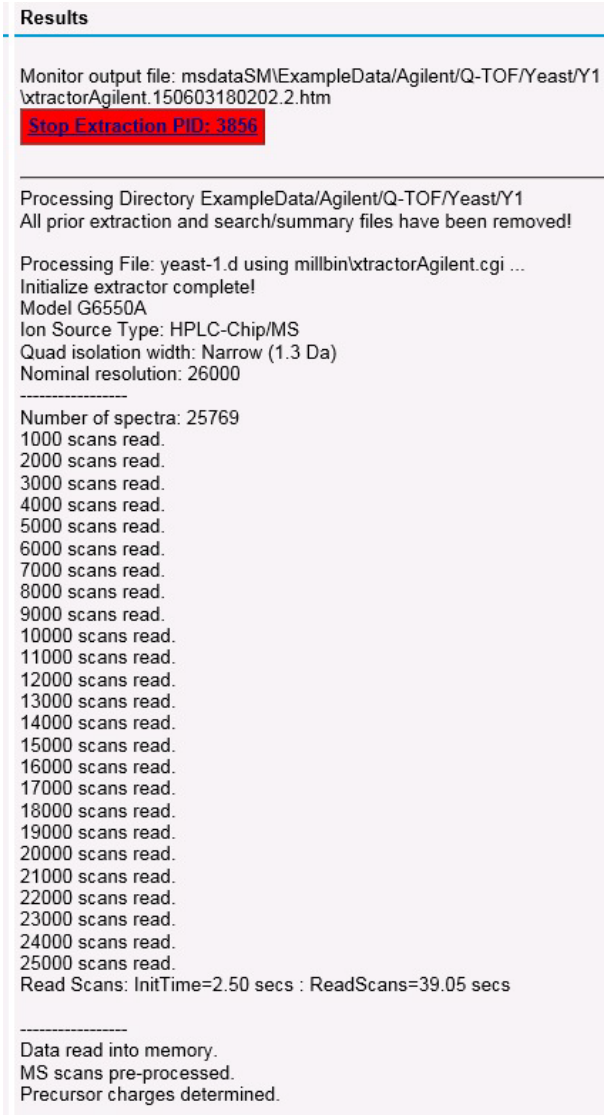


Figure 3 Data Extractor running

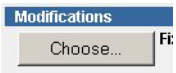
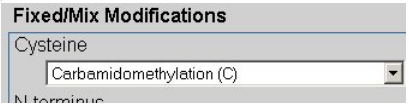
Exercise 2. Search the database

After you have extracted your spectra, you are ready to search each spectrum against a protein database. For this exercise the Search mode is Variable modifications.

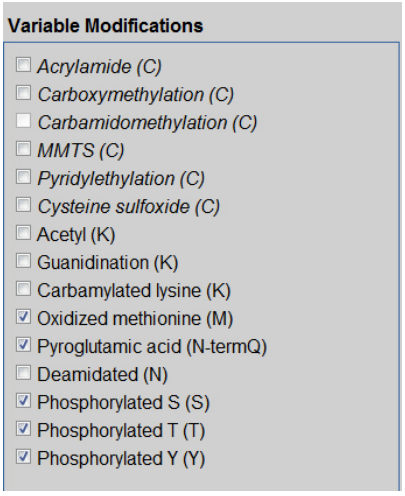
Exercise 2. Search the database

Steps	Detailed Instructions	Comments
1 Navigate to the MS/MS Search page.	• Do one of the following: ◦ From the Data Extractor page, click the **MS/MS Search** button.  ◦ From the Spectrum Mill home page, click the link to MS/MS Search. 	
2 Remove prior MS/MS search results and maximize use of available CPUs so that search (which is a two-step process) executes more quickly.	**a** Mark the **Remove all prior MS/MS Search results** check box. **b** Mark the **Maximize CPUs** check box.	• Someone else has probably done these exercises before. You remove prior results to start afresh, if you are using the data files directly from the **ExampleData** folder.
3 Check that your Data Directory is set to **y1**.	• If you have just completed the data extraction, your **y1** data directory should already be set correctly. If not, click **Select...** to select the **y1** folder.	

Exercise 2. Search the database (continued)

Steps	Detailed Instructions	Comments
4 Select the SwissProt.yeast database. • If you do not have this database, use the Protein Databases page to create it. See the System Administration chapter in the *Application Guide* for details.	• For **Database** (under **Search Parameters**), select the **SwissProt.yeast** database.	• If the database name does not appear, go to the Protein Databases page and click **Update Database List**. If the database still does not appear, then contact your system administrator. • While the official name of the database is "Swiss-Prot", the Spectrum Mill naming convention requires that the downloaded file be prefixed with "SwissProt" (no hyphen).
5 Set up the Fixed and Variable Modifications; • If the carbamidomethylation modification is set, set up the variable modifications. Variable Modifications: Oxidized methionine Pyroglutamic acid Phosphorylation S, T and Y	**a** Click **Choose...** near the middle of the page.  **b** Under the **Cysteine** heading, select **Carbamidomethylation**.  **c** For Variable Modifications, mark the check boxes for oxidized methionine, pyroglutamic acid, phosphorylation S, T and Y.	• Three types of modifications are available for MS/MS Search: ◦ The **fixed** modifications are assumed to apply universally and are searched in a single search cycle. ◦ The **mix** modifications trigger cyclic MS/MS searches, where a different part of the modification is searched in each cycle. ◦ The **variable** modifications allow for both modified and unmodified forms within a peptide. All variable modifications are searched in each cycle. • Note that when you select a variable modification, the **Search mode** changes automatically to **Variable modifications**.

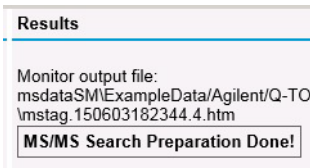
Exercise 2. Search the database (continued)

Steps	Detailed Instructions	Comments
	 **d** Click **OK.** The names of the modifications appear as in Figure 4 on page 22.	• For more information about choosing modifications, see the online help. • Your system administrator can configure custom modifications. • If you are not certain which modifications are present, do a mass gap or unknown modifications search first. See Chapter 3 of this guide. • Note that when you click OK, the **Search Mode** selection changes automatically to **Variable Modifications**.

Exercise 2. Search the database (continued)

Steps	Detailed Instructions	Comments
6 Set up the **Search Mode**.	**a** Make sure that the check box for **Calculate reversed database scores** is marked. (This is the default setting.) **b** Mark the check box for **Dynamic peak thresholding**. **c** From the **Discriminant scoring list**, select **Off**. (This is the default setting.) **d** Verify that the **Search mode** is set to **Variable Modifications**.	• When you **Calculate reversed database scores**, you search against peptide sequences in their forward and reverse directions. If you obtain similar scores for both searches, you may have a false positive. This setting is necessary to do autovalidation with a target false discovery rate (FDR). • **Dynamic peak thresholding** is a scoring enhancement that enables identification of more low-abundance and short-chain peptides for Agilent Q-TOF data. • By default, with Spectrum Mill B.06.00, peptides with lengths less than 5 amino acids are not considered unless you mark the check box for **Dynamic peak thresholding**.
7 Set other parameters as shown in Figure 4 on page 22.	**a** Examine the items in red text carefully, since these are the ones you may need to change when you process your own samples. **b** Click a blue section divider bar to display help for that section of the page.	• For Agilent Q-TOF and other instruments that acquire only CID data, always specify **All** for the **Fragmentation mode**. For instruments such as an ETD ion trap that provide additional fragmentation modes, specify the mode used during data acquisition.
8 Save the settings as a parameter file. • Use the YourName folder. • Name the file MSMS_Search.	**a** Click **Save As**. **b** For **Folder**, select **YourName**. **c** for **Name**, type **MSMS_Search**. **d** Click **Save**. Note that the name of the parameter file appears in red near the top of the browser window (Figure 4).	• A parameter file allows you to reuse the settings later, either interactively or in an automated workflow. • If you forget to save the parameter file, you can find this file under the Examples folder to use for the workflow automation exercises.

Exercise 2. Search the database (continued)

Steps	Detailed Instructions	Comments
9 Start the search.	**a** Click **Start Search**.  **b** In the **Results** area to the right of the Data Extractor page, click the **Monitor Results** link at the bottom, or use the **Request Queue**. **c** Scroll to the *top* of the **Results** area to see the message that indicates that search preparation is finished. 	• The command is automatically placed in the queue for execution. • The search is queued in two steps. The first step creates the set of batch files that are to be processed. The second step processes each batch search. • MS/MS Search processes all spectral files in the directory. • Search time varies depending on the size of the database searched. This search goes fairly quickly because you use a species subset database. • If you want to stop the search, click the red **Stop Search PID: xxx** link at the top of the **Results** section. Also go to the Tool Belt, click **Stop process** and make sure there are no other processes running. If there are, stop them. See the *Application Guide* for further instructions.

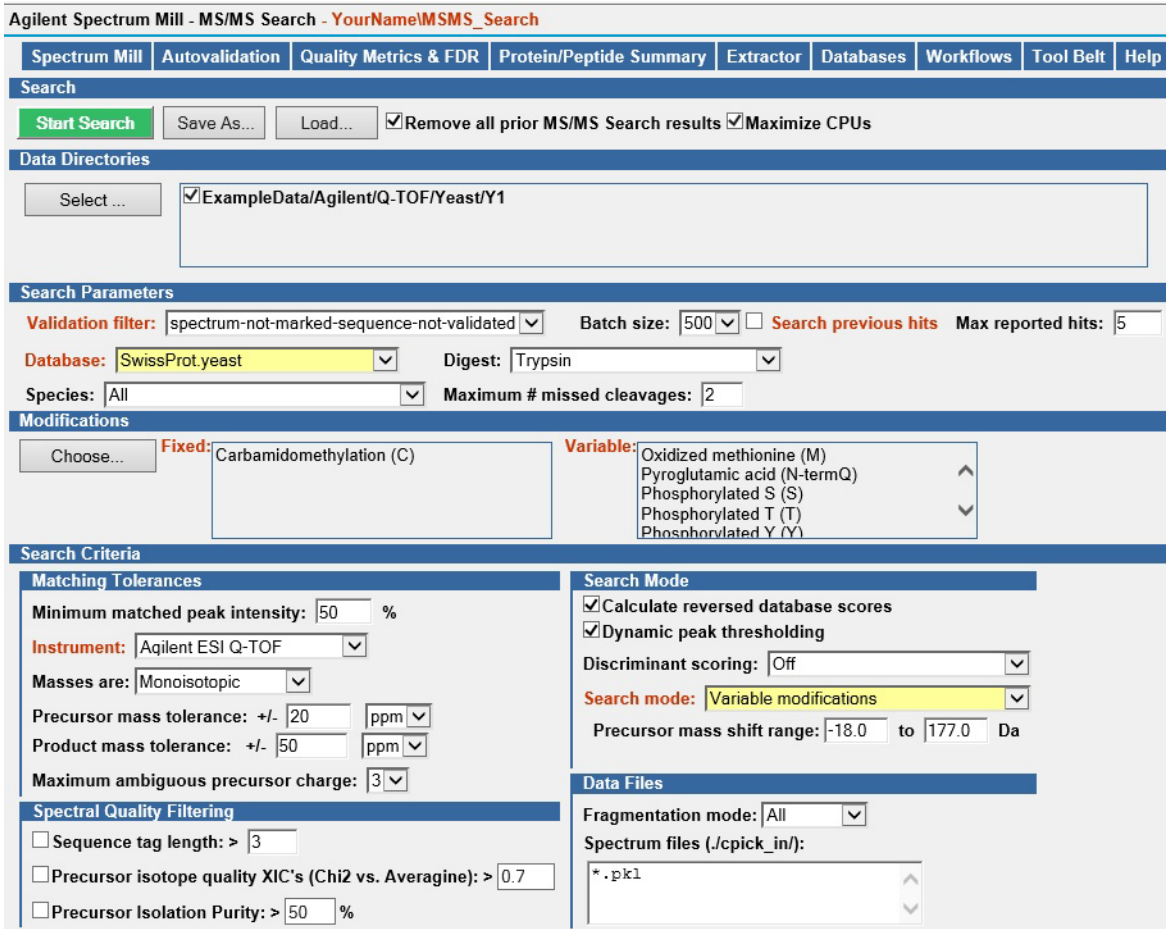
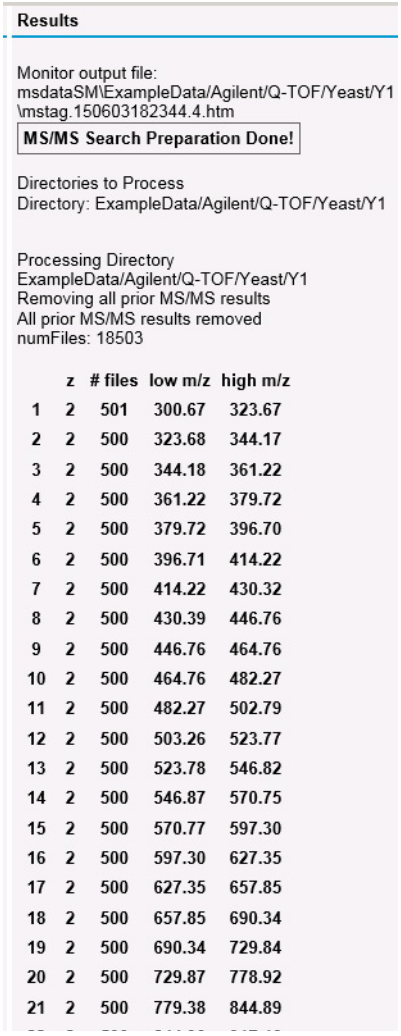


Figure 4 MS/MS Search settings



Results

Monitor output file:
msdataSM\ExampleData/Agilent/Q-TOF/Yeast/Y1
\mstag.150603182344.4.htm

MS/MS Search Preparation Done!

Directories to Process
Directory: ExampleData/Agilent/Q-TOF/Yeast/Y1

Processing Directory
ExampleData/Agilent/Q-TOF/Yeast/Y1
Removing all prior MS/MS results
All prior MS/MS results removed
numFiles: 18503

	z	# files	low m/z	high m/z
1	2	501	300.67	323.67
2	2	500	323.68	344.17
3	2	500	344.18	361.22
4	2	500	361.22	379.72
5	2	500	379.72	396.70
6	2	500	396.71	414.22
7	2	500	414.22	430.32
8	2	500	430.39	446.76
9	2	500	446.76	464.76
10	2	500	464.76	482.27
11	2	500	482.27	502.79
12	2	500	503.26	523.77
13	2	500	523.78	546.82
14	2	500	546.87	570.75
15	2	500	570.77	597.30
16	2	500	597.30	627.35
17	2	500	627.35	657.85
18	2	500	657.85	690.34
19	2	500	690.34	729.84
20	2	500	729.87	778.92
21	2	500	779.38	844.89

Figure 5 MS/MS Search running

Exercise 3. Autovalidate high-quality results

After you have completed a database search, you validate the good results. Validation means that you accept that the matches are correct.

The Spectrum Mill workbench provides two ways to autovalidate results. One way automatically validates matches at a certain FDR (False Discovery Rate). The other method of autovalidation uses fixed thresholds to validate the highest-scoring results, which then lets you use the Protein/Peptide Summary page for manual review and validation. This exercise shows you how to use the first method of autovalidation.

Exercise 3. Autovalidate high-quality results

Steps	Detailed Instructions	Comments

1 Navigate to the Autovalidation page.	• Navigate to this page from one of three pages:
	  ◦ Spectrum Mill home
	  ◦ MS/MS Search
	  ◦ Protein/Peptide Summary	

2 Check that your Data Directory is set to **y1**.	• If you have just completed the data extraction and MS/MS Search, your **y1** data directory should already be set correctly. If not, click the **Select...** button to select the **y1** folder.	
◦ Set up autovalidation first with the Auto thresholds strategy, Peptide mode. Keep defaults. See Figure 6 on page 26.	**a** For **Strategy**, click **Auto thresholds**. **b** For **Mode**, keep the default of **Peptide**.	• This strategy/mode optimizes the score and delta R1-R2 score thresholds up to a target False Discovery Rate (FDR) that you enter.
3 Save the settings as a parameter file. ◦ Save to the YourName folder. ◦ Name the file AV-auto-pep.	**a** Click **Save As**. **b** For **Folder**, select **YourName**. **c** For **Name**, type **AV-auto-pep**. **d** Click **Save**.	• If you forget to save the parameter file, you can find this file under the Examples folder to use for the workflow automation exercises.

Exercise 3. Autovalidate high-quality results (continued)

Steps	Detailed Instructions	Comments
4 Validate with Queue request turned off.	**a** Clear the **Queue request** check box, if necessary. **b** Click **Validate Files**. **c** Watch for a validation summary (at the bottom) that lists the hits and spectra that have been validated (Figure 7).	• If you do not mark Queue request, clicking this button validates your data immediately. • If you mark Queue request, clicking the button places the command in the queue, to be executed in sequence. To view where this command is in the queue, click **View Request Queue** when the results start appearing.
5 Set up autovalidation second with the Auto thresholds strategy-Protein polishing mode. • Use the settings shown in Figure 8 on page 28.	**a** For **Mode**, select **Protein polishing**. **b** Change **Minimum protein score** to 0. **c** Change **...maximum protein FDR** to 1.0%.	• After validating in Peptide mode, you refine the validation with the Protein polishing mode.
6 Save the settings as a parameter file. • Save to the YourName folder. • Name the file AV-auto-protpol.	**a** Click **Save As**. **b** For **Folder**, select **YourName**. **c** For **Name**, type **AV-auto-protpol**. **d** Click **Save**.	
7 Validate with Queue request turned off.	**a** Clear the **Queue request** check box, if necessary. **b** Click **Validate Files**. **c** Watch for a validation summary that lists the hits and spectra that have been validated, as well as the false positives (Figure 9).	• If you do not mark Queue request, clicking this button validates your data immediately. • If you mark Queue request, clicking the button places the command in the queue, to be executed after the other commands have been executed in sequence.
8 Close the Autovalidation window.	• Close the window only after you see the word “finished”.	

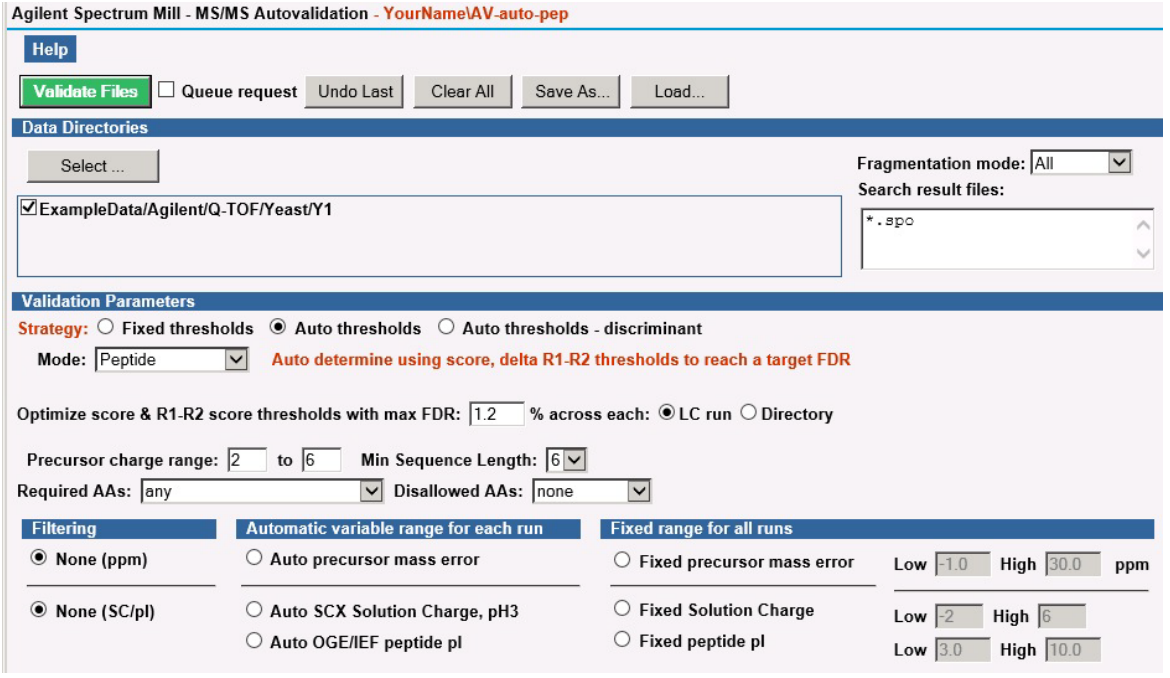


Figure 6 Autovalidation page, Auto thresholds strategy, Peptide mode

yeast-1 ruleNumber: 5 z: 6
initial FDR: 0.00 Score: 4.00 R1R2: 0.00 #hits: 5 #hits false: 0
optimal FDR: 0.00 Score: 4.00 R1R2: 0.00 #hits: 5 #hits false: 0

Run	Rule #	z	score	R1R2	FDR(%)	# hits	#hits False
yeast-1	1	2	5.4	0.4	1.19	3348	20
Subtotal	all	all			1.19	3348	20
yeast-1	2	3	5.0	0.8	1.15	1746	10
Subtotal	all	all			1.15	1746	10
yeast-1	3	4	5.6	0.1	1.08	186	1
Subtotal	all	all			1.08	186	1
yeast-1	4	5	4.8	0.7	0.00	22	0
Subtotal	all	all			0.00	22	0
yeast-1	5	6	4.0	0.0	0.00	5	0
Subtotal	all	all			0.00	5	0
Total	all	all			1.17	5307	31

Rule	z	Valid PSM's(#)	FP PSM's(#)	FDR(%)
1	2	3348	20	1.19
2	3	1746	10	1.15
3	4	186	1	1.08
4	5	22	0	0.00
5	6	5	0	0.00

Number of validated spectra : 5307 Number of False positives Fwd-Rev score < 0 : 31 FP rate (%): 1.17

ExampleData/Agilent/Q-TOF/Yeast/Y1 Number of validated spectra : 5307

Preparing new hit and spectra validation info.... finished
Resetting previously validated hits
Writing to hitTable.1.tsv
Validated Hits: 5307
Writing to spectrumTable.1.tsv
Validated Spectra: 5307

Figure 7 Autovalidation results for Auto thresholds strategy, Peptide mode (partial)

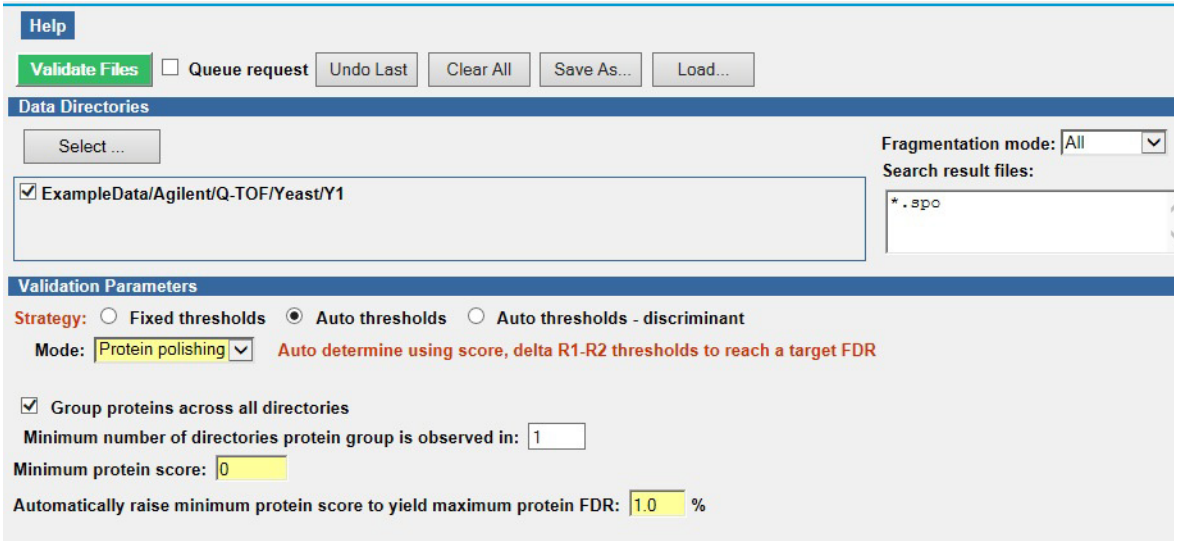


Figure 8 Autovalidation page, Auto thresholds strategy, Protein polishing mode

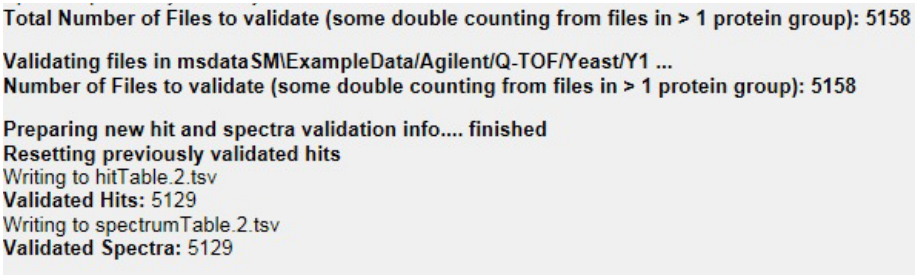


Figure 9 Autovalidation results for Auto thresholds, Protein polishing mode

Task 2. Summarize and compare valid results

Exercise 1. Examine the validated results

In this exercise you produce a Protein Summary Details report and a Peptide report to review the validated results.

Produce a Protein Summary Details Report

Exercise 1. Examine the validated results- Produce a Protein Summary Details Report

Steps	Detailed Instructions	Comments
1 Navigate to the Protein/Peptide Summary page, if you are not already there.	• From the MS/MS Search page or the Spectrum Mill home page, click the link to the Protein/Peptide Summary page.	There are links to this page from many other Spectrum Mill pages.
2 Check that your **Data Directory** is set to **y1**.	• If you have just completed autovalidation, your **y1** data directory should already be set correctly. If not, click **Select...** to select the **y1** folder.	
3 Set the **Mode** to **Protein Summary Details**.	• From the **Mode** list, select **Protein Summary Details**.	• The Protein/Peptide Summary page offers a variety of results display modes. • These allow you to organize results by proteins, peptides, or samples. • For more details, see Table 4 in the *Quick Start Guide* or the MS/MS Data Review and Validation chapter in the *Application Guide*.
4 Filter only valid results.	• From the **Filter results by** list, select **Valid**.	• The **Filter results by** setting selects from your data only the results that match your setting. • In this case, it selects only the results that match the **valid** setting (which was designated during autovalidation).

Exercise 1. Examine the validated results- Produce a Protein Summary Details Report (continued)

Steps	Detailed Instructions	Comments
5 Sort the proteins by Score (default setting).	**a** Keep the **Sort Proteins by** list selection as **Score**.	
6 Filter peptides by setting these parameters: • Score > 0 • % SPI > 0	**a** Under **Filter peptides by** for **Score, type 0**. **b** For **% SPI**, type **0**.	• By setting to zero, you ensure that all valid data is selected, regardless of score.
7 Set the **Review Fields**: • Var mod sites • Protein MW • Modification names	**a** Under **Review Fields**, mark the **Var mod sites, Protein MW**, and **Modification names** check boxes.	• **Review Fields** determine what information you see in the final results summary.
8 Keep the settings for other fields as in Figure 10.		• All fields are described and the default settings are shown in the online help. To access the online help, click the **Help** button for the page, or click the blue bar for a section.
9 Save the settings as a parameter file. • Save to the YourName folder. • Name the file prot-sum.	**a** Click **Save As**. **b** For **Folder**, select **YourName**. **c** For **Name**, type **prot-sum**. **d** Click **Save**.	• If you forget to save the parameter file, you can find this file under the Examples folder to use for the workflow automation exercises.

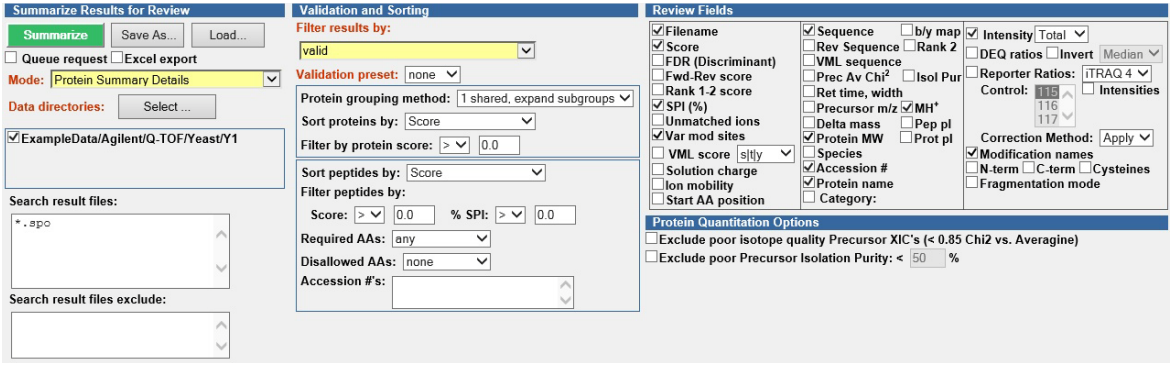


Figure 10 Protein Summary Details mode settings

Exercise 1. Examine the validated results- Produce a Protein Summary Details Report (continued)

Steps	Detailed Instructions	Comments
10 **Summarize** with the Queue request turned off.	**a** Clear the **Queue request** check box, if necessary. **b** Click **Summarize**.	• If you do not mark the Queue request check box, the valid data are immediately summarized upon clicking the button. • If you do mark the Queue request check box, the command is placed in the queue and will not execute until other commands ahead of it for the same data file have finished executing.
11 Examine the summary report. • Make sure results match those in Figure 11 on page 32. • Learn what headers mean.	**a** Check that your results are similar to those in Figure 11 on page 32. Since the public databases are constantly updated, your results may vary slightly. **b** Become familiar with the headers in Table 1 on page 32.	File names are in the format: Data_File_Name.aaaa.bbbb.c, where • aaaa = first merged scan • bbbb = last merged scan • c = assigned precursor charge (0 means charge was ambiguous)

Group (#)	Subgroup (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Total Protein Spectral Intensity	Protein MW (Da)	Database Accession #	Protein Name
1	1.1	55	40	503.35	67.6	1.22e+008	69827.1	P10591	Heat shock protein SSA1
1	1.2	45	33	397.45	50.8	1.05e+008	69639.9	P10592	Heat shock protein SSA2
1	1.3	21	16	182.37	24	4.25e+007	70659.6	P09435	Heat shock protein SSA3
1	1.4	13	12	121.78	19.9	1.08e+007	74523.7	P16474	78 kDa glucose-regulated protein homolog
2	2.1	69	33	489.08	80.5	4.91e+008	44794.6	P00560	Phosphoglycerate kinase
3	3.1	41	34	432.44	52.6	3.53e+007	86029.5	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
4	4.1	56	28	431.94	88.8	4.09e+008	35860.1	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3
4	4.2	46	25	356.97	69.8	3.55e+008	35960.3	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2
4	4.3	43	23	317.10	74	3.02e+008	35863.4	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1
5	5.1	59	27	417.27	73.4	3.20e+008	46970.4	P00925	Enolase 2
5	5.2	57	26	398.58	68.1	4.11e+008	46872.3	P00924	Enolase 1
6	6.1	42	32	395.23	46.6	4.01e+007	93743.9	P32324	Elongation factor 2
7	7.1	43	30	393.72	70.4	1.21e+008	54943.0	P00549	Pyruvate kinase 1
7	7.1	2	1	9.44	1.9	8.87e+006	55479.5	P52489	Pyruvate kinase 2
8	8.1	51	27	382.94	61.4	8.24e+007	61722.5	P06169	Pyruvate decarboxylase isozyme 1
8	8.1	2	1	11.88	2.9	5.90e+005	68992.7	Q07471	Thiamine metabolism regulatory protein THI3
8	8.2	15	9	108.72	20.2	1.99e+007	62139.1	P16467	Pyruvate decarboxylase isozyme 2
8	8.3	10	6	90.05	16.5	8.25e+006	61636.3	P26263	Pyruvate decarboxylase isozyme 3
9	9.1	36	33	374.09	32.6	1.51e+007	119319.2	P17255	V-type proton ATPase catalytic subunit A
10	10.1	39	28	366.25	50.8	6.78e+007	66771.5	P11484	Heat shock protein SSB1
10	10.2	38	28	365.02	52.2	6.57e+007	66707.5	P40150	Heat shock protein SSB2
11	11.1	35	30	363.27	38.8	3.25e+007	80898.2	P15108	ATP-dependent molecular chaperone HSC82
11	11.2	29	26	318.85	33.4	2.97e+007	81404.9	P02829	ATP-dependent molecular chaperone HSP82
12	12.1	40	27	363.06	63.3	4.11e+007	57007.9	P46367	Potassium-activated aldehyde dehydrogenase, mitochondrial

Figure 11 Protein Summary Details results

Table 1 Protein Summary Details Report column headers

Header	Definition
Group	Proteins related to at least one other protein in the group by at least 1 shared peptide
Subgroup	Distinct peptides that uniquely represent a lower-scoring member of the group (isoforms and family members)
Spectra	Number of spectra in the subgroup
Distinct Peptides	Number of peptides in each subgroup that are not shared by other subgroups
Distinct Summed MS/MS Search Score	Total score for all the distinct peptides in a subgroup
%AA Coverage	Percentage of amino acids in the protein hit covered by the spectral data

Table 1 Protein Summary Details Report column headers (continued)

Header	Definition
Total Protein Spectral Intensity	Spectral intensity of all the peptides in the subgroup protein
Protein MW	Molecular weight of the subgroup protein
Accession #	Database accession number for the top hit
Protein Name	Name of the top database hit

Produce a Peptide - Spectrum Match Report

Exercise 2. Examine results with the Spectrum Viewer - Produce a Peptide - Spectrum Match Report

Steps	Detailed Instructions	Comments
1 Set the Mode to Peptide - Spectrum Match.	• From the **Mode** list, select **Peptide - Spectrum Match**.	
2 Filter only valid results.	• From the **Filter results by** list, select **Valid**.	
3 Sort the peptides by delta mass shift.	**a** From the **Sort Peptides by** list select **Delta mass shift**.	• If trying to match a peptide to a specific chromatographic peak, sort peptides by Retention Time. Or sorting by Sequence charge can also be useful. Be sure to mark the check boxes in the Review Fields.
4 Filter peptides by confirming these parameters to: • Score > 0; % SPI > 0	**a** Under **Filter peptides by** for **Score, type 0**. **b** For **% SPI, type 0**.	• By setting to zero, you ensure that all valid data is selected, regardless of score.
5 In the **Review Fields,** clear Sequence and Protein MW and mark these fields: • b/y map • Delta mass • Var mod sites • VML score (s/t/y) • VML sequence • Modification names	**a** Under **Review Fields**, clear the **Sequence** and **Protein MW** check boxes. **b** Mark the **b/y map** check box. **c** Mark the **Delta mass, Var mod sites, VML score, VML sequence** and **Modification names** check boxes.	• The modification sites, scores and sequences appear in the report.

Exercise 2. Examine results with the Spectrum Viewer - Produce a Peptide - Spectrum Match Report (continued)

Steps	Detailed Instructions	Comments
6 Keep the settings for other fields in Figure 12.		
7 Save the settings as a parameter file. • Save to the YourName folder. • Name the file psm.	**a** Click **Save As**. **b** For **Folder**, select **YourName**. **c** For **Name**, type **psm**. **d** Click **Save**.	• If you forget to save the parameter file, you can find this file under the Examples folder to use for the workflow automation exercises.

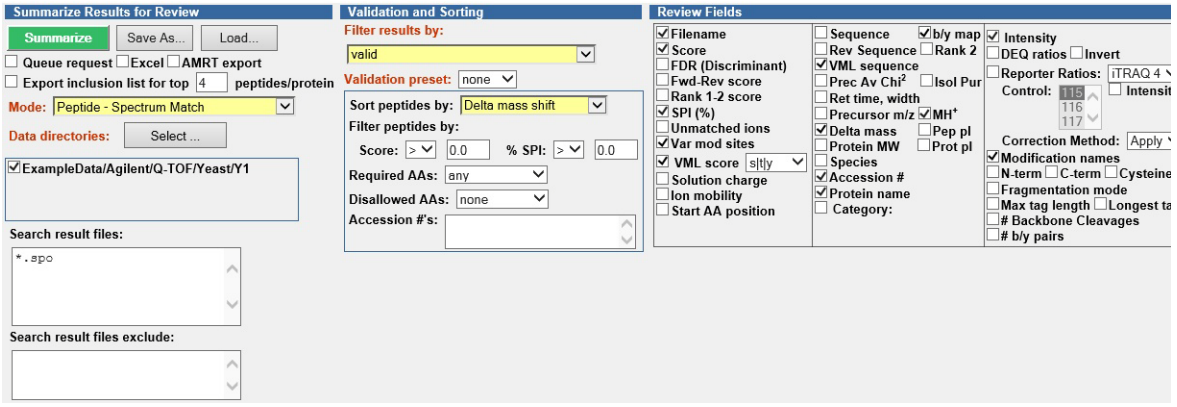


Figure 12 Peptide - Spectrum Match mode settings

8 **Summarize** with the Queue request turned off.	**a** Clear the **Queue request** check box, if necessary. **b** Click **Summarize**.	• If you do not mark the Queue request check box, the valid data are immediately summarized upon clicking the button. • If you do mark the Queue request check box, the command is placed in the queue and will not execute until other commands ahead of it have finished executing.
9 Examine the summary report. • Make sure results match those in Figure 13 on page 35. • Learn what headers mean.	**a** Check that your results are similar to those in Figure 13 on page 35. Since the public databases are constantly updated, your results may vary slightly. **b** Become familiar with the headers in Table 2 on page 35.	

Exercise 2. Examine results with the Spectrum Viewer - Produce a Peptide - Spectrum Match Report (continued)

Steps	Detailed Instructions	Comments
10 To examine these results in more detail, go to "Exercise 2. Examine results with the Spectrum Viewer" on page 37. • In particular, look at the sequence map for peptides #14 and 12.	• For peptide #14, the sequence map is NEAT(1.0)PEAEQVK. • For peptide #12, the sequence map is IVLT(0.50)GIS(0.50)K.	• The sequence for peptide #14 shows a (1.0) at the site because it is the only phosphorylation site possible. • Peptide #12 has two possible phosphorylation sites. Since the site cannot be determined from the MS/MS spectrum, (.5) is shown at each site.

#	Filename	z	Score	SPI (%)	Spectrum Intensity	Score VML	# STY	# sty Sites	# Loc sty Sites	# Ambig sty Sites	Variable Sites	Sequence Map	Modifications Map	MH+ Matched (Da)	MH+ Mass Shift (Da)	MH+ Error (ppm)	Accession #	Pr
1	yeast-1.19473.19473.4	4	6.72	63.1	8.68e+004	4.383	3	2	2	0	M489m T491t S498s	(K)E T W G K/m/N t F V L/H A L s K P L K F L E N L\N K(S) ET(0.0)WGKM(1.0)NT(0.99)FVLHALS(0.99)KPLKFLENLNK	m:Oxidized methionine tPhosphorylated T sPhosphorylated S	3058.655	175.9082	-6.0	P40458	Au
2	yeast-1.14853.14853.2	2	10.14	92.3	3.74e+005	0.000	4	2	1	1	S336s T341t	(R)L S L D/I s A/K H L\t T K(L) LS(0.0)LDIS(0.99)AKHLT(0.50)T(0.50)K	s:Phosphorylated S tPhosphorylated T	1426.826	159.9587	16.4	P47139	Ur
3	yeast-1.4934.4934.2	2	6.41	73.2	2.02e+005	0.000	3	2	1	1	Y197y S199s	(R)V/I y T s P I K(A) VIY(0.99)T(0.50)S(0.50)PIK	y:Phosphorylated Y sPhosphorylated S	920.545	159.9376	4.5	P47047	AT
4	yeast-1.19691.19691.3	3	5.28	66.5	1.20e+005	5.279	6	2	2	0	Y275y T293t	(K)y S K P A L L H/S R F F F P A L Q G S/t T K(M) Y(0.99)S(0.0)KPALLHS(0.0)RFFPALQGS(0.0)T(0.99)T(0.0)K	y:Phosphorylated Y tPhosphorylated T	2349.271	159.9347	0.8	Q12109	Tr
5	yeast-1.17766.17766.3	3	6.37	56.0	2.03e+005	0.000	3	2	0	2	S1381s S1392s	(K)K\F I L K s V Q N/V/E V E E A P s S D K K(N) KPILKS(0.0)VQNVEVEEAPS(0.50)S(0.50)DKK	s:Phosphorylated S	2325.266	159.9185	-5.7	P34244	Pr
6	yeast-1.12871.12871.4	4	7.60	58.6	2.18e+006	0.895	5	2	1	1	Y26y T45t	(R)S L T A/y A P/N L M L W G G A S M L G L F V F t E G W\P K(F) S(0.0)LT(0.0)AY(0.99)APNLMLWGGAS(0.50)MLGLFVFT(0.50)EGWPK	y:Phosphorylated Y tPhosphorylated T	3157.589	159.9166	-4.8	P37299	Cy
7	yeast-1.15330.15330.4	4	7.66	59.6	5.55e+006	0.462	8	2	0	2	T279t S289s	(K)F R L/A K M t F L/R P S S L I T s\E S A I V T T K(I) FRLAKMT(0.25)FLRPS(0.0)S(0.25)LIT(0.0)S(0.25)ES(0.25)AIVT(0.0)T(0.0)K	t:Phosphorylated T sPhosphorylated S	2797.564	159.8965	-12.2	P53916	Pr
8	yeast-1.15030.15030.3	3	8.88	76.9	1.30e+005	1.137	3	1	1	0	M282m Y285y	(K)L/F/P/m D A y L K L N H I I A N I N S F T/K(G) LPPM(1.0)DAY(0.99)LKLNHIIANINS(0.0)FT(0.0)K	m:Oxidized methionine yPhosphorylated Y	2513.358	95.9766	5.9	P36048	Pr
9	yeast-1.19045.19045.3	3	8.47	93.3	1.46e+005	0.000	4	1	0	1	Y246y M253m	(R)L S/D/L E R I T K y T G A H G V m A V R(G) LS(0.0)LERIT(0.33)KY(0.33)T(0.33)GAHGVM(1.0)AVR	y:Phosphorylated Y mOxidized methionine	2217.181	95.9596	-0.7	Q06063	tR
10	yeast-1.19393.19393.3	3	5.85	70.2	6.06e+005	0.000	4	1	0	1	M8m T11t	(R)I I m D/S t H F N E I R\S I I R S R(S) ILM(1.0)DS(0.25)T(0.25)HFNEIRS(0.25)IIRS(0.25)R	m:Oxidized methionine tPhosphorylated T	2188.166	95.9470	-6.2	P41807	V-A
11	yeast-1.19549.19549.3	3	5.11	57.4	7.80e+004	5.111	4	1	1	0	S291s	(K)N I P W Y G E T K N P V L F/S N A V E N L s R(T) NIPWY(0.0)GET(0.0)KNPVLFS(0.0)NAVENLS(0.99)R	s:Phosphorylated S	2648.347	80.0110	16.4	P32528	Ur
12	yeast-1.14544.14544.3	3	5.27	59.4	1.87e+006	2.515	3	1	1	0	Y233y	(K)S N L A G K F G F T A V V I y D N E P K(S) S(0.0)NLAGKFGFT(0.0)AVVIY(0.99)DNEPK	y:Phosphorylated Y	2170.118	79.9976	13.9	P37302	An
13	yeast-1.15719.15719.3	3	7.39	64.2	1.88e+005	99	1	1	1	0	T31t	(R)R L t G\F/L/P Q E I/K(S) RLT(1.0)GFLPQEIK	t:Phosphorylated T	1301.758	79.9936	19.7	P06738	Gl
14	yeast-1.14753.14753.2	2	6.61	78.2	3.98e+004	99	1	1	1	0	T629t	(R)I E/Q E Q V P G G A P G/E t V K(V) IEQEQVPGGAPGET(1.0)VK	t:Phosphorylated T	1638.833	79.9925	15.2	P36013	NA
15	yeast-1.16546.16546.2	2	8.41	76.3	1.09e+005	8.405	5	1	1	0	T204t	(R)I t/K A/D I E S Y L E K S S K(Q) IT(0.99)KADIES(0.0)Y(0.0)LEKS(0.0)S(0.0)K	t:Phosphorylated T	1711.911	79.9918	14.2	P12695	Dil

Figure 13 Peptide - Spectrum Match report

Table 2 Peptide - Spectrum Match Report column headers with variable modifications

Header	Definition
#	Table entry number
Filename	Name of the spectral file
z	Peptide charge
Score	Peptide database search score

Table 2 Peptide - Spectrum Match Report column headers with variable modifications

Header	Definition
SPI (%)	% scored peak intensity, which is the percentage of the ion current in the extracted MS/MS spectrum that is explained by the top database hit
Spectrum Intensity	Intensity calculated from extracted ion chromatogram of precursor ion
Score VML	Difference is Score of same identified sequences with different variable modification localizations
# 'Mod Type' (STY or Y or K or P)	Number of possible modification sites in the peptide for the modification chosen
# 'mod type' (sty or y or k or p) Sites	Number of modifications on the peptide based on precursor mass and fragment information
# Loc 'mod type' (sty or y or k or p) Sites	Number of sites where the modification site can be determined from the data; either only one site or fragmentation around the site makes it possible to localize the site
# Ambig 'mod type' (sty or y or k or p) Sites	Number of ambiguous modification sites
Variable Sites	The exact location and type of modification
Sequence Map	Sequence from top database hit with positions of b and y fragments; number in parentheses is the probability that the modification is localized at that particular location
Modifications Map	Name of the modification
MH^+ Matched (Da)	Measured precursor ion MH^+
MH^+ Mass Shift (Da)	Difference between the mass measured by the mass spec and the mass of the unmodified peptide; difference must be ascribed to either measurement error or some modification
MH^+Error (ppm)	Error (in ppm) between the observed and theoretical m/z
Accession #	Database accession number for the top hit
Protein Name	Name of the top database hit

Exercise 2. Examine results with the Spectrum Viewer

The Spectrum Viewer is a visual tool to evaluate MS/MS search results (Protein/Peptide Summary), to assess MS/MS spectral quality (Spectrum Summary), and to evaluate *de novo* sequencing results (Sherenga *de novo* Summary). This exercise focuses on use of the Spectrum Viewer for evaluation of MS/MS search results. In this capacity, the Spectrum Viewer shows your extracted spectrum annotated with theoretical fragments from the top database search result. The Spectrum Viewer has flexible scaling and labeling options to let you better visualize your results.

Exercise 2. Examine results with the Spectrum Viewer

Steps	Detailed Instructions	Comments
1 If you have not already done so, display a Peptide - Spectrum Match report.	• Follow steps in "Produce a Peptide - Spectrum Match Report" on page 33.	
2 Display the peptide #24 spectrum to view an example of a single phosphorylation site.	**a** In the summary report, click the number **24** under the **#** header. Tip: The file name is yeast-1.5485.5485.2. **b** Check the bottom of the browser window, where you see the spectrum displayed in the Spectrum Viewer. **c** Check that your results are similar to those in Figure 14 on page 38. Since the public databases are constantly updated, your results may vary. For example, the numbers of the peptides may be slightly different.	• See "Exercise 3. Examine details of search results" on page 42 to examine the phosphorylation sites in the sequence map.
3 Display the peptide #22 spectrum to view a spectrum with two possible phosphorylation sites (S and T).	**a** In the summary report, click the number **22** under the **#** header. Tip: The file name is yeast-1.11447.11447.2 **b** Check the bottom of the browser window, where you see the spectrum displayed in the Spectrum Viewer. **c** Check that your results are similar to those in Figure 15 on page 38.	• See "Exercise 3. Examine details of search results" on page 42 to examine the phosphorylation sites in the sequence map.

Exercise 2. Examine results with the Spectrum Viewer (continued)

Steps	Detailed Instructions	Comments

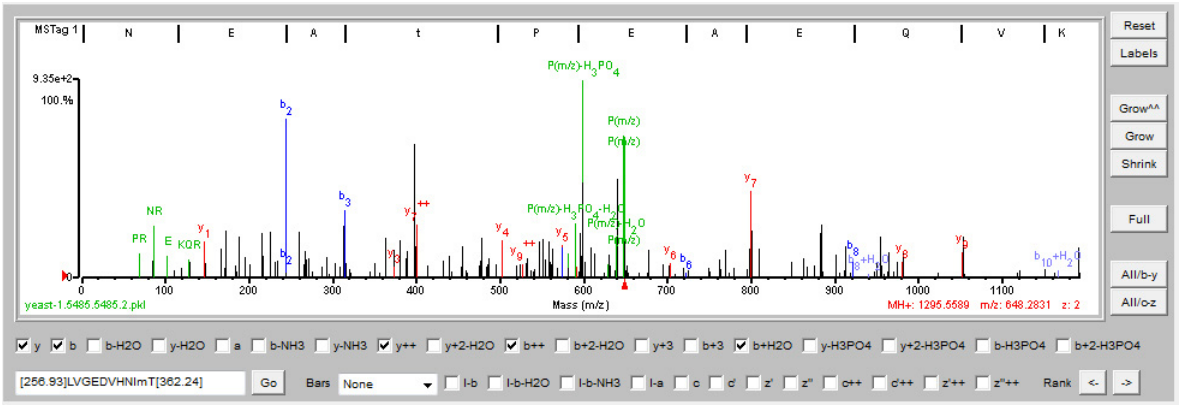


Figure 14 Spectrum #24 in Spectrum Viewer

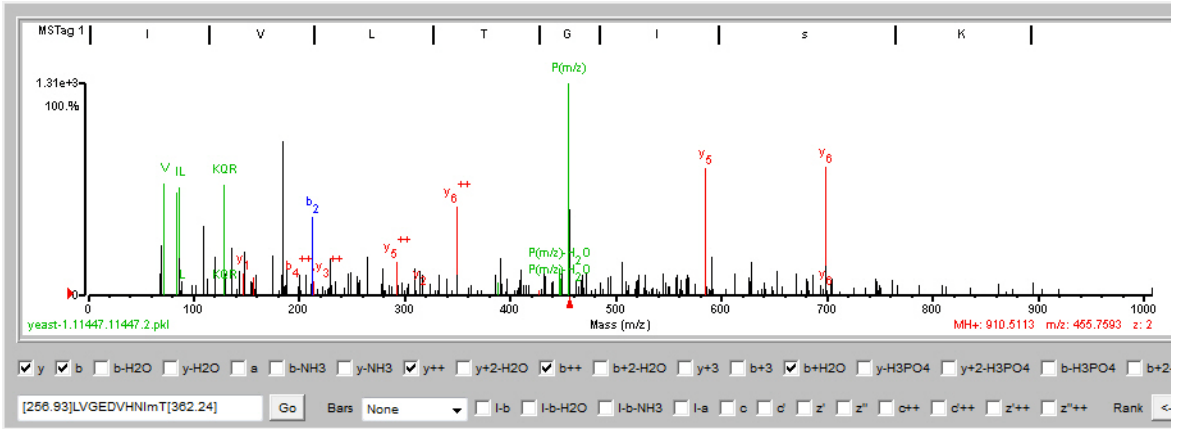


Figure 15 Spectrum #22 in Spectrum Viewer

Exercise 2. Examine results with the Spectrum Viewer (continued)

Steps	Detailed Instructions	Comments

4 Use the **Labels** and **All/b-y** buttons to change the spectrum annotations.
• Use this step and the next ones for either or both peptide #24 or #22.
	a Click four times on the **Labels** button on the right side of the Spectrum Viewer.
	b Check that the peak labeling changes as you click. It cycles through four options: 1) ion types 2) all masses and ion types 3) only assigned masses and ion types 4) none.
	c Be sure the labels are visible before proceeding.
	d Click twice on the **All/b-y** button on the right side of the Spectrum Viewer.
	e Check that as you click, the marks in the check boxes immediately below the spectrum change, as do the spectrum labels.
		• In the spectrum, unmatched ions are shown in black, while matched ions are displayed in red and blue.
		• The **Labels** button toggles among different peak labeling options. These are:
		  • The default of interpreted peaks (b, y, etc.)
		  • Interpreted peaks plus mass labels of *all* peaks
		  • Interpreted peaks plus mass labels of interpreted peaks
		  • No labels
		• The **All/b-y** button toggles marking of check boxes in the first row under the spectrum and labels peaks accordingly. The toggle either marks all the check boxes or resets to defaults.

5 Use the **Grow**, **Grow^^, Shrink**, and **Reset** buttons to expand and shrink the spectrum in the vertical (y) direction.	**a** Click twice on the **Grow** button on the right side of the Spectrum Viewer. **b** Check that the spectrum expands vertically. **c** Click twice on the **Shrink** button on the right side of the Spectrum Viewer. **d** Check that the spectrum returns to original size. **e** Click twice again on the **Grow** button. **f** Click once on the **Reset** button on the right side of the Spectrum Viewer. **g** Click once on the **Grow^^** button. **h** Click once on the **Reset** button. **i** Check that the spectrum returns to original size.	

Exercise 2. Examine results with the Spectrum Viewer (continued)

Steps	Detailed Instructions	Comments
6 Use the mouse and the **Full** and **Reset** buttons to expand and shrink the spectrum in the horizontal (x) direction.	**a** Use the cursor to expand a portion of the x-axis of the spectrum. Move your mouse over the spectrum. When a crosshair is displayed, select the portion of the spectrum you wish to expand. **b** Double-click the spectrum or click the **Reset** button to return to the original display.	• The **Full** button displays the spectrum with the full x-axis values. • If you see a black arrow at the lower right-hand corner of the spectrum, this may mean that some of the spectrum is not displayed because there were no significant peaks in that region. Click the **Full** button to display the full x-axis range of the spectrum.
7 Select **Bars** options to locate all potential b- and y-ions.	**a** Find the **Bars** options at the bottom of the Spectrum Viewer.  **b** Select **y** and observe the gray lines that indicate locations of all possible y-ions from the database match. **c** Select **b** and observe the gray lines that indicate locations of all possible b-ions from the database match. Notice that the amino acid sequence reverses order when you switch from **y** to **b**. **d** Select **None** to return to the default setting.	• By choosing **y** or **b** for **Bars**, you quickly see how well the sequence from the summary table matches the spectrum.

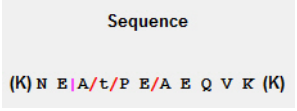
Exercise 2. Examine results with the Spectrum Viewer (continued)

Steps	Detailed Instructions	Comments
8 Type in a sequence and check for alignment with the spectrum.	**a** Find the sequence box in the lower left-hand corner of the Spectrum Viewer.  **b** Type in your own sequence. **c** Click the **Go** button. **d** Select the **b** and **y Bars** options and observe whether your trial sequence is consistent with the mass information in the spectrum. **e** Click the left **Rank** arrow to go back to prior sequences, if present. 	• This feature is useful when you know part of a sequence, or you want to test a possible sequence. • The sequence in the box is initially set to a default. • Mass gaps shown in brackets indicate segments of the sequence that are unknown. Perhaps there are portions of your MS/MS spectrum where fragmentation is insufficient to provide an amino acid sequence. • You can enter mass gaps in the middle of the sequence as well as at the ends.

Exercise 3. Examine details of search results

While the Spectrum Viewer is your primary tool to determine how well your database hit matches your spectrum, you may occasionally want to examine search results in more detail. In this exercise, you become familiar with the additional database search details that are available.

Exercise 3. Examine details of search results

Steps	Detailed Instructions	Comments
1 If you have not already done so, display a Peptide- Spectrum Match report.	• Follow steps in "Produce a Peptide - Spectrum Match Report" on page 33.	
2 Display the MS/MS Search results for peptide #24.	**a** In the summary report, click the 24th entry under the **Filename** header. **b** Check the bottom of the browser window, where you see the MS/MS Search results for the 24th spectrum. See Figure 16 on page 43 as an example.	• At this point, you may see only the spectrum.
3 Expand the section of the page that shows the MS/MS Search results.	**a** To adjust the size of this section, rest your mouse over the split line. **b** When the pointer becomes a double-sided arrow, drag the arrow to move the split line. 	
4 Examine the MS/MS Search hit list for modification sites.	**a** Scroll up to see the hit list at the top of the MS/MS Search results section (top of Figure 16 on page 43.) **b** Note the color-coded lines in the sequence and read the comment to the right in this table.  Note that since the public databases are constantly updated, your results may vary.	• The color-coding of the amino acid sequence of the matched peptide from the database search shows: ◦ Red forward-slashes for locations of y-ions ◦ Blue backslashes for locations of b-ions ◦ Magenta pipes (vertical lines) for locations of both b- and y-ions ◦ The sequence map shows that there is a cleavage on either side of the phosphothreonine: (K)N E|A/t/P E/A E Q V K(K)

Exercise 3. Examine details of search results (continued)

Steps	Detailed Instructions	Comments

Sample ID (comment): **yeast-1.5485.5485.2.pkl**
Search Time: **0 secs / 1 searches, 0.000 secs/search** Parameters used in Search
Database searched: **SwissProt.yeast**
Number of entries from database used for search as limited by results in original SPO file: **1**
Number of sequences passing through parent mass filter: **1**
MS/MS Search search selects **1** entry.

Parent mass: **648.2827**$^{+2}$ **(+/- 0.0259 Da)**
Parent mass shift range (multiple mutations): **-18.0 to 177.0 Da**

scoreVML: 99 sequenceVML: NEAT(1.0)PEAEQVK

Result Summary

Rank	Score	SPI (%)	BCS	# Unmatched Ions	Variable sites	Sequence	MH$^+$ Calculated (Da)	MH$^+$ Error (Da)	MH$^+$ Error (ppm)	Protein MW/pI (Da)	Species	Accession #	Protein Name
1	8.20	59.9	4	10/24	T49t	(K)NEAtPEAEQVK(K)	1215.5852	79.9730	5.2	12588.8/5.43	YEAST	Q08245	Protein ZEO1

Detailed Results

Rank	Score	SPI (%)	BCS	# Unmatched Ions	Variable sites	Sequence	MH$^+$ Calculated (Da)	MH$^+$ Error (Da)	MH$^+$ Error (ppm)	Protein MW/pI (Da)	Species	Accession #	MS-Digest Index #	Protein Name
1	8.20	59.9	4	10/24	T49t	(K) N E A/t/P E A E Q V K (K)	1215.5852	79.9730	5.2	12588.8/5.43	YEAST	Q08245	6706	Protein ZEO1

Fragment-ion (m/z)	87.057	173.092	216.094	226.079	244.091	260.088	314.182	363.135	398.161	400.712	475.285	547.264^{+2}	559.261	574.327	581.273^{+2}	590.773^{+2}	596.759^{+2}	599.293^{+2}	612.801^{+2}	800.409	883.439	954.476	981.416	1052.455
Frac. Inten.(% of TIC)	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
Rel. Inten.(% of BP)	31.79	28.95	27.05	27.77	95.97	28.09	81.65	38.77	100.00	32.49	49.64	37.23	31.56	36.93	0.96	39.37	39.91	5.32	29.94	81.21	65.95	31.61	24.84	43.68
Signal/Noise	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Score	0.500	-0.289	0.500	0.250	0.500	-0.281	-0.817	-0.388	0.500	1.500	-0.496	-0.372	-0.316	1.500		-0.394	-0.399	1.500	-0.299	1.500	0.500	0.500	1.500	1.500
Ion-type	a_1		a_2	b_2-H_2O	b_2				b_4-H_3PO_4	y^{++}_{11}				y_3	sty^{+2}			y_{11}-$H_3PO_4$$^{+2}$		y-	y_8-H_3PO_4	y_5-H_3PO_4	y_8-H_3PO_4	y_8	y_9
Delta ppm	19.2		-17.5	-13.6	-7.6				-15.9	2.6				13.3				2.8		-6.9	-14.2	-13.7	-13.4	-10.1
	0.50																	1.50						
	NR																	sty^{+2}						
	19.2																	2.8						

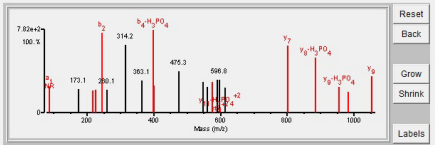


Figure 16 MS/MS Search results for peptide #24

5 View details of the search results.

a Scroll below the hit list to see details of the search results (Figure 16).

b Inspect the table for the rank-one hit. Table 3 on page 44 shows in detail how the ions were assigned.

c Inspect the spectrum just below the table to see visually how well the interpretation matches the spectrum.

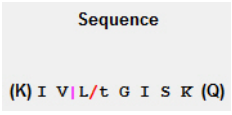
Table 3 Ion assignments

Header	Definition
Fragment-ion (m/z):	Ions from the experimental spectrum
Frac. Inten. (% of TIC):	Fractional intensities for each of the fragment ions. The final % SPI for the database match is the sum of the fractional intensities for the assigned ions.
Rel. Inten. (% of BP):	Relative intensities for each of the fragment ions, measured as a percentage of the base peak
Score:	Scores for each of the fragment ions. The final score for the database match is the sum of the scores for each of the ions. (The unassigned ions subtract from the score.)
Ion-type:	Ion type for theoretical fragment from database match
Delta Da (or ppm):	Difference between theoretical and experimental MH^+

Exercise 3. Examine details of search results

Steps	Detailed Instructions	Comments
6 (Optional) Scroll back up to the top of the MS/MS Search Result Summary table to discover how all the links work.	• To jump to a particular search result, click a number under the **Rank** heading in the **Result Summary**. • To view an MS Product listing that shows all theoretical fragment ions, click an amino acid sequence in the **Result Summary**. • To view information in the database you searched, click an accession number in the **Result Summary**. • To view MS Digest results that show peptide masses from a theoretical digestion, click an MS Digest index number in the **Detailed Results**.	• When you check links, click your browser's **back** arrow to return to the previous display. • When you process your data, you only click the links you feel are necessary to understand the quality of the results. These steps are presented here only so you know what information is available to help you with your decision.
7 Navigate back to the main Peptide report.		You may need to readjust the page partition so you can see the top section again.

Exercise 3. Examine details of search results (continued)

Steps	Detailed Instructions	Comments
8 (Optional) Click the remaining Accession # link here to see what it does.	• To link to the database you searched, click a link under the **Accession #** header.	If you wish to enable a link from the amino acid sequence to a BLAST database, see the system administration chapter in the *Application Guide*.
9 Repeat steps 2-4 for peptide #22. • (Optional) Repeat steps 5-8 for peptide #22.	**a** In the summary report, click the 22nd entry under the **Filename** header. **b** Check the bottom of the browser window, where you see the MS/MS Search results for the 22nd spectrum. **c** Note the color-coded lines in the sequence and read the comment to the right in this table. 	• Color-coding of the sequence allows you to assess b- and y-ion coverage without inspecting the spectrum, and for variable modifications results, color-coding aids in identifying the site of a post-translational modification (PTM) by highlighting existence of surrounding b/y ions. • The sequence map shows that there is no cleavage between two of the sites; so it is not possible to localize the modification to the S or the T. Therefore, Spectrum Mill just assigns it to one of them.

Exercise 4. Compare results from two samples

The example data includes two *yeast* samples. The goal of this exercise is to compare the proteins in the two samples, to see which are up-regulated and which are down-regulated in the cells. To compare the two, you first need to process the y2 sample in the same way you processed the y1 sample.

Exercise 4. Compare results from two samples

Steps	Detailed Instructions	Comments
1 Extract the **y2** data.	• Repeat "Exercise 1. Extract the spectra" on page 11, except when you **Select** the **Data Directory**, select **y2** rather than **y1**.	
2 Search the **y2** data.	• Repeat "Exercise 2. Search the database" on page 17, except when you **Select** the **Data Directory**, select **y2** rather than **y1**.	
3 Autovalidate the **y2** data.	• Repeat "Exercise 3. Autovalidate high-quality results" on page 24, except when you **Select** the **Data Directory**, select **y2** rather than **y1**.	
4 Summarize and compare the **y1** and the **y2** data. ◦ Mode: Protein-Protein Comparison ◦ Group columns by: Directory	**a** Repeat "Produce a Protein Summary Details Report" on page 29, with the following changes: ◦ When you **Select** the **Data Directory**, select *both* **y2** and **y1**. (Control-click the second directory. Do not use shift-click.) ◦ For **Mode**, select **Protein-Protein Comparison**. ◦ For **Group columns by**, choose **Directory**. **b** Check to see that the completed page looks like that in Figure 17. **c** Click **Summarize**.	The **Group results by** parameter controls how the sample results are displayed. • If you select **File**, each sample fraction is displayed individually. • If you select **Directory**, all fractions present in a data directory are consolidated into a single sample.

Exercise 4. Compare results from two samples (continued)

Steps	Detailed Instructions	Comments

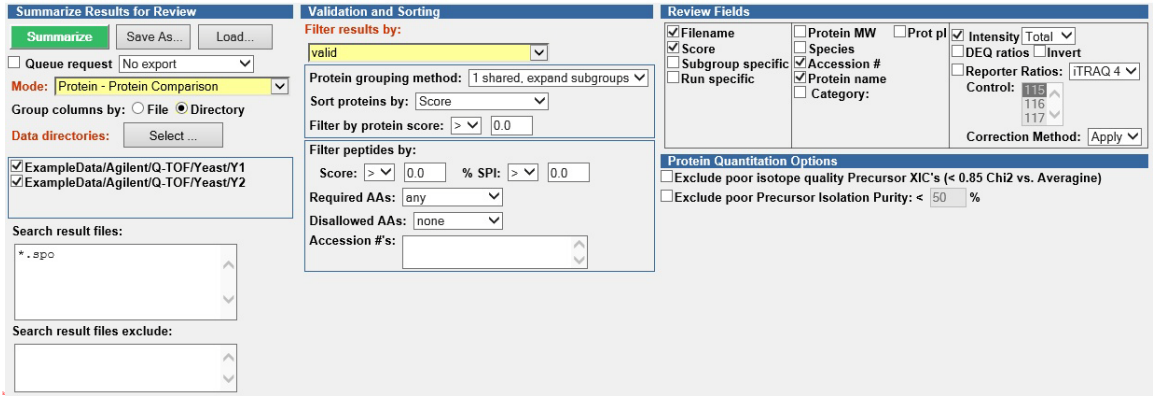


Figure 17 Protein/Peptide Summary settings to compare two samples

5 Examine the overall summary report.

- Check that your results are similar to those in Figure 18 on page 48.
  - Since the public databases are constantly updated, your results may vary,
  - Your results also vary depending on which peptides you manually validated.

Note how the colored cells in the summary report make it easy to visualize sample differences.

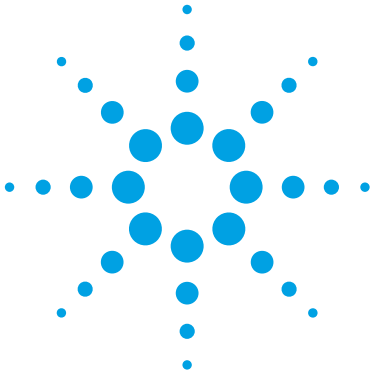
- The color-code in the left columns represents the relative protein concentrations. Darker colors represent higher concentrations.
- The top number in each cell is the number of peptides detected for each protein, while the bottom number is the Total Protein intensity, the total of the intensities for all peptides detected for that protein. These intensities are calculated from extracted ion chromatograms from the precursor ions.
- These intensity results are sufficient for studies where you are interested in differences of two-fold or more.

Exercise 4. Compare results from two samples (continued)

Steps	Detailed Instructions	Comments
6 (Optional) Vary the summary settings (particularly the **Mode** setting) and investigate the results.		• See the MS/MS Data Review and Validation chapter of the *Spectrum Mill MS Proteomics Workbench Application Guide*.

ExampleData/Agilent/Q-TOF/Yeast/Y1 # spectra total intensity	ExampleData/Agilent/Q-TOF/Yeast/Y2 # spectra total intensity	Columns Protein Observed (#)	Spectra Protein Observed (#)	Database Accession #	%AA Coverage	Distinct Peptides (#CI)	Distinct Summed MS/MS Search Score	Group #	Subgroup #	Protein Name
57 4.11e+008	147 1.47e+009	2	204	P00924	88.7	48	753.64	1	1.1	2 X Enolase 1
59 3.20e+008	97 9.39e+008	2	156	P00925	86.4	39	615.17		1.2	Enolase 2
55 1.22e+008	58 2.00e+008	2	113	P10591	73.8	46	626.84	2	2.1	5 X Heat shock protein SSA1
45 1.05e+008	45 1.58e+008	2	90	P10592	60.5	38	499.56		2.2	Heat shock protein SSA2
13 1.08e+007	18 2.69e+007	2	31	P16474	31.3	20	226.40		2.3	78 kDa glucose-regulated protein homolog
21 4.25e+007	19 5.58e+007	2	40	P09435	27.1	18	235.57		2.4	Heat shock protein SSA3
42 4.01e+007	51 7.89e+007	2	93	P32324	58.3	41	540.18	3	3.1	Elongation factor 2
69 4.91e+008	76 6.49e+008	2	145	P00560	80.5	33	528.43	4	4.1	Phosphoglycerate kinase
43 1.21e+008	45 1.81e+008	2	88	P00549	82.8	36	499.84	5	5.1	Pyruvate kinase 1
41 3.53e+007	41 5.52e+007	2	82	P05694	54.3	36	495.05	6	6.1	5-methyltetrahydropteroyltriglutamate--homocysteine meth
31 1.06e+007	33 1.64e+007	2	64	P07149	24.7	39	489.17	7	7.1	Fatty acid synthase subunit beta
35 3.25e+007	36 5.09e+007	2	71	P15108	47.8	36	482.13	8	8.1	2 X ATP-dependent molecular chaperone HSC82
29 2.97e+007	32 4.69e+007	2	61	P02829	41.7	32	419.57		8.2	ATP-dependent molecular chaperone HSP82
56 4.09e+008	83 7.21e+008	2	139	P00359	88.8	29	470.02	9	9.1	3 X Glyceraldehyde-3-phosphate dehydrogenase 3
43 3.02e+008	38 5.04e+008	2	81	P00360	87.3	27	371.47		9.2	Glyceraldehyde-3-phosphate dehydrogenase 1
46 3.55e+008	48 5.76e+008	2	94	P00358	80.4	26	401.41		9.3	Glyceraldehyde-3-phosphate dehydrogenase 2
36 1.51e+007	37 3.41e+007	2	73	P17255	40.5	42	467.50	10	10.1	V-type proton ATPase catalytic subunit A
39 6.78e+007	42 1.00e+008	2	81	P11484	64.2	33	465.62	11	11.1	2 X Heat shock protein SSB1
38 6.57e+007	41 9.57e+007	2	79	P40150	65.5	33	461.53		11.2	Heat shock protein SSB2
30 2.44e+007	35 8.75e+007	2	65	P19414	48.4	35	445.76	12	12.1	Aconitate hydratase, mitochondrial
51 8.24e+007	53 2.05e+008	2	104	P06169	62.8	28	439.71	13	13.1	3 X Pyruvate decarboxylase isozyme 1

Figure 18 Comparison of proteins from two *yeast* cell states (partial listing)



2 Iterative Workflow – Alternative Processing for MS/MS Data

This chapter provides familiarization exercises to process an example set of Q-TOF MS/MS data files using the Agilent Spectrum Mill MS Proteomics Workbench with an alternative workflow: an iterative search/autovalidation strategy. These exercises correspond with the iterative processing described in the second half of Chapter 1 of the companion *Spectrum Mill MS Proteomics Workbench Application Guide*. Use this workflow to obtain more detailed information to help characterize the identified proteins.

In these exercises you first do a variable modifications search, as described in Chapter 1 of this guide. You then do additional searches. When you are satisfied with the amount of information you have gleaned from your samples, you generate a final results summary.

Because the exact steps you take are sample- and study-dependent, you are encouraged to vary the processing to fit your samples and information needs.

After you summarize final validated results, you can check for remaining high-quality spectra with Spectrum Summary. Then you identify the good spectra with Sherenga *de novo* sequencing.

In these exercises you do not save parameter files unless you choose to do so.

While the exercises in this chapter use Agilent Q-TOF data files, the general workflow applies broadly for MS/MS data files.

Task 1. Do iterative searches and autovalidations

The MS/MS Search software provides identity, variable modifications, and homology search modes.

- In identity mode, you search for exact matches in the selected database.
- In variable modifications mode, you search for post-translational and chemical modifications that occur at a fraction of the available sites.
- In homology modes, you search for single amino acid substitutions relative to the database (as well as any variable modifications you have selected). In homology modes, you can also search for an unknown or unexpected modification. For details, see Chapter 3, “Unknown Modifications: Mass Gap Search”.

For this series of exercises, you extract the data, do a variable modifications search, autovalidate and summarize all the valid results. If you have already completed the exercises in Chapter 1, you are done with this step. You then do additional processing, as described in this chapter.

Exercise 1. Do initial processing of the yeast data

In this exercise, you extract the yeast data, do a variable modifications search, autovalidate and summarize all the valid results. If you have already completed Chapter 1, you are done with this exercise. If not, see the instructions there.

Exercise 2. Do a second round of processing

If you want additional information, follow one of these paths:

- Search saved results for other possible modifications.
- Search saved results and do a mass gap search against these saved results so you can find unexpected modifications. Chapter 3 describes a mass gap search.
- Search a different database. In this case, you do *not* search saved results. You must do a *new* search.

- Search an additional species. Once again, you do *not* search saved results. Instead, you do a *new* search.

Because you will choose a custom path, detailed steps are not provided here.

As you process this data, Spectrum Mill retains a cumulative summary of all the results you have validated. At the end, your final results summary includes everything you have validated through multiple interpretation/review/validation steps. However, if you have done iterative searches on saved results, for accurate FDR calculations, do a new search using all modifications.

Exercise 3. Check for remaining high-quality spectra

In this exercise, you learn to use the Spectrum Summary page to view spectra that do not yet have a valid interpretation, and to decide whether they are good spectra worthy of further interpretation.

One of the most useful measures of spectral quality is the maximum sequence tag length. The maximum sequence tag length represents the longest sequence of amino acids that can be located in the spectrum. The **Maximum tag length** filter retains spectra that are likely to be peptide spectra and that have a fragmentation pattern that lends itself to interpretation.

The Spectrum Summary results show spectra annotated with the amino acid sequences that correspond with the maximum sequence tag lengths. These are not intended as *de novo* interpretations, but rather show how much sequence information *could* be contained in the spectra. This exercise illustrates use of this filter to locate remaining good spectra.

Exercise 3. Check for remaining high-quality spectra

Steps	Detailed Instructions	Comments
1 Navigate to the Spectrum Summary page.	• Navigate to this page from the Spectrum Mill home page, the Protein/Peptide Summary page or the Tool Belt page.	
2 Check that your **Data Directory** is set to **yeast\y1**.	• If you have just been working with this data set, your data directory should already be set correctly. If not, click **Select...** to select the **y1** folder.	
3 Filter those spectra that do not yet have a search result that you have designated as valid.	• From the **Spectrum validation filter** list, select **spectrum-not-marked-sequence-not-validated**.	• The **Spectrum validation filter** selects from your data only the results that match your setting.

Exercise 3. Check for remaining high-quality spectra (continued)

Steps	Detailed Instructions	Comments
4 Set the **Validation preset** to **good-spectrum.**	• From the **Validation preset** list, select **good-spectrum**.	When you process your own data, you have three choices: • If in the next step you set the **Filter by** parameter such that poor spectra are likely to be excluded, select **good-spectrum**. • Otherwise, select **bad-spectrum.** • Select **reset** if you want to undo a previous designation.
5 Filter by sequence tag length • Set a minimum tag length of 7.	• For this exercise, mark the **Sequence tag length** option and set a filter of **>7**.	• With the **Sequence tag length** filter, higher numbers correlate with spectra of better quality. The setting of **7** filters and keeps only the best spectra. • The software has a number of additional filters. These include intensity and number of b/y pairs.
6 Check to see that your settings look like those in Figure 19.	• Change any settings that do not match Figure 19.	

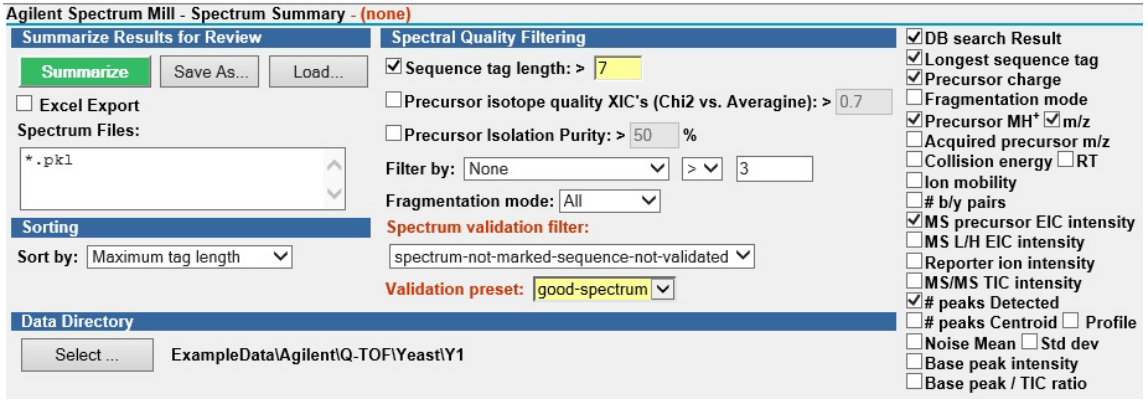


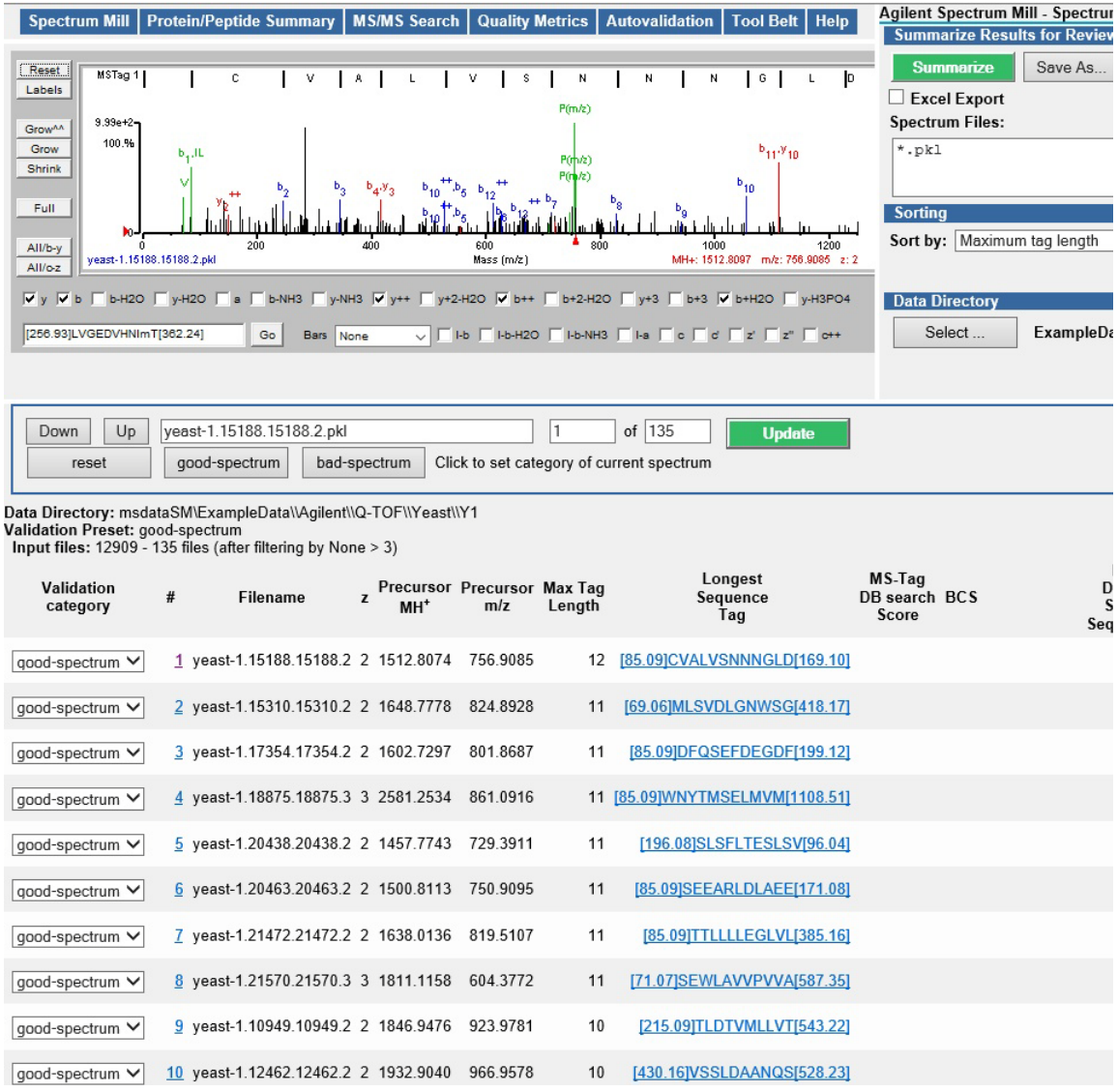
Figure 19 Spectrum Summary settings

Exercise 3. Check for remaining high-quality spectra (continued)

Steps	Detailed Instructions	Comments
7 Summarize the spectra. • Results should look like those in Figure 20 on page 56.	**a** Click **Summarize**. **b** Check that your results are similar to those in Figure 20 on page 56. Check that your results show: • A table at the bottom of the page • The Spectrum Viewer with the first spectrum at the top of the page	• Your results may vary slightly. • Note that text near the top of the summary table (just under the blue buttons) indicates that you have displayed about 127-128 potentially good spectra.
8 Use the Spectrum Viewer to determine how much potential fragmentation information is contained in the spectrum. • Tip: the information is under the # for each spectrum.	**a** In the summary report, click the number **1** under the # header. **b** Check the top left of the browser window, where you see the first spectrum displayed in the Spectrum Viewer.	• The Spectrum Viewer shows your extracted MS/MS spectrum annotated with an amino acid sequence that could explain the mass gaps between the ions. • In blue and red, you see potential b and y ions. The other ions are in black. • This is not intended to be a *de novo* interpretation, but merely shows how much plausible sequence information may be contained in the spectrum. • To learn more about the Spectrum Viewer, see "Exercise 2. Examine results with the Spectrum Viewer" on page 37.
9 Categorize the spectrum as a good-spectrum.	• In the **Validation category**, keep the **good-spectrum** designation.	• If you wanted to change the designation to bad-spectrum, you would use one of two methods: • Select **bad-spectrum** from the drop-down list under the **Validation category** heading. • Click the **bad-spectrum** button near the top of the page.

Exercise 3. Check for remaining high-quality spectra (continued)

Steps	Detailed Instructions	Comments



Validation category	#	Filename	z	Precursor MH^+	Precursor m/z	Max Tag Length	Longest Sequence Tag	MS-Tag DB search Score	BCS
good-spectrum	1	yeast-1.15188.15188.2	2	1512.8074	756.9085	12	[85.09]CVALVSNNNGLD[169.10]		
good-spectrum	2	yeast-1.15310.15310.2	2	1648.7778	824.8928	11	[69.06]MLSVDLGNWSG[418.17]		
good-spectrum	3	yeast-1.17354.17354.2	2	1602.7297	801.8687	11	[85.09]DFQSEFDEGDF[199.12]		
good-spectrum	4	yeast-1.18875.18875.3	3	2581.2534	861.0916	11	[85.09]WNYTMSELMVM[1108.51]		
good-spectrum	5	yeast-1.20438.20438.2	2	1457.7743	729.3911	11	[196.08]SLSFLTESLSV[96.04]		
good-spectrum	6	yeast-1.20463.20463.2	2	1500.8113	750.9095	11	[85.09]SEEARLDLAEE[171.08]		
good-spectrum	7	yeast-1.21472.21472.2	2	1638.0136	819.5107	11	[85.09]TTLLLLEGLVL[385.16]		
good-spectrum	8	yeast-1.21570.21570.3	3	1811.1158	604.3772	11	[71.07]SEWLAVVPVVA[587.35]		
good-spectrum	9	yeast-1.10949.10949.2	2	1846.9476	923.9781	10	[215.09]TLDTVMLLVT[543.22]		
good-spectrum	10	yeast-1.12462.12462.2	2	1932.9040	966.9578	10	[430.16]VSSLDAANQS[528.23]		

Figure 20 Spectrum Summary results

Exercise 3. Check for remaining high-quality spectra (continued)

Steps	Detailed Instructions	Comments

10 Display the next spectrum.	• Do one of the following:
	  ◦ Click the **Down** button at the top of the page to display the next spectrum at the top of the page.
	  ◦ In the summary report, click the number **2** under the **#** header. Again, this displays the spectrum at the top of the page.
		• In the summary report, click the second entry under the **Longest Sequence Tag** header. This displays the Spectrum Viewer in a new window.

11 Examine and categorize the second spectrum as a good spectrum.	• In the **Validation category**, keep the **good-spectrum** designation.	
12 Repeat step 10 and step 11 for as many spectra as you care to review.		
13 Update the spectral designations.	• Click **Update**.	• This makes your designations a permanent part of the data record. • You have now designated approximately 135 spectra as good (if you did not change any to "bad-spectrum"). Your results may vary slightly.
14 (Optional) Vary input parameters and investigate the results.	**a** To change parameters and create a new summary table, change settings and start over. **b** For this step, do not click the **Update** button again. If you do, your results will not match in the next exercise.	

Task 2. Use *de novo* sequencing and review results

Exercise 1. Use Sherenga *de novo* sequencing

Now that you have found some additional good spectra, you decide to try to identify them. If these were your samples, you might search another database in identity mode, conduct additional variable modifications or homology searches, or try *de novo* sequencing. In this exercise, you use Sherenga *de novo* sequencing to gain insight into possible amino acid sequences for these peptides. For more information about *de novo* sequencing, see the Sherenga *de novo* Sequencing chapter in the *Spectrum Mill MS Proteomics Workbench Application Guide*.

Exercise 1. Use Sherenga de novo sequencing

Steps	Detailed Instructions	Comments
1 Navigate to the Sherenga *de novo* Sequencing page.	• From the Spectrum Mill home page, click the link to the ***de novo* Sequencing** page.	
2 Check that your **Data Directory** is set to **yeast\y1**.	• If you have just been working with this data, the data directory should already be set correctly. If not, click **Select...** to select the **y1** folder.	• Note that the variable modification is still there. Don't remove it.
3 Filter **good spectra whose sequences have not been validated.**	• From the **Validation Filter** list, select **good-spectrum-sequence not validated**.	• The **Validation filter** selects from your data only the results that match your setting. • In this case, it selects only the spectra that you just designated as "good" with the Spectrum Summary page.
4 Set the other parameters as shown in Figure 21 on page 59.	**a** Mark the **Remove all prior Sherenga results** checkbox. **b** For Min. vertex score, type `2`. **c** For **Sequence tag length**, type `7`.	• You may want to remove prior results if others have set up Sherenga *de novo* sequencing before you began the exercise.

Exercise 1. Use Sherenga de novo sequencing

Steps	Detailed Instructions	Comments
5 Start the sequencing, and monitor results.	**a** Click **Sequence**. **b** In the **Results** panel, click the **Monitor Results** link. See Figure 22 on page 60. **c** Scroll to the *top* of the **Results** area to see the message that indicates that sequencing is finished.	

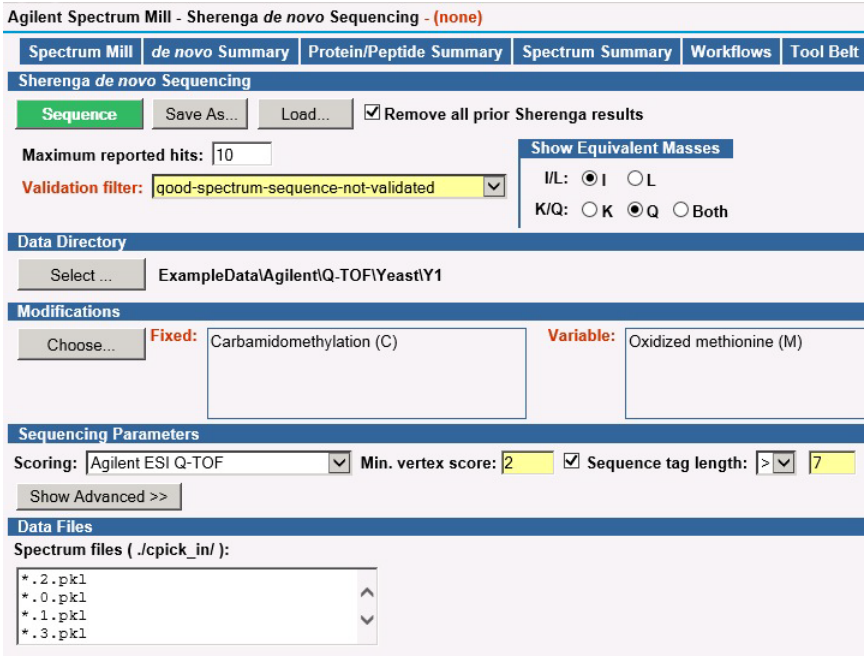


Figure 21 Sherenga *de novo* Sequencing settings

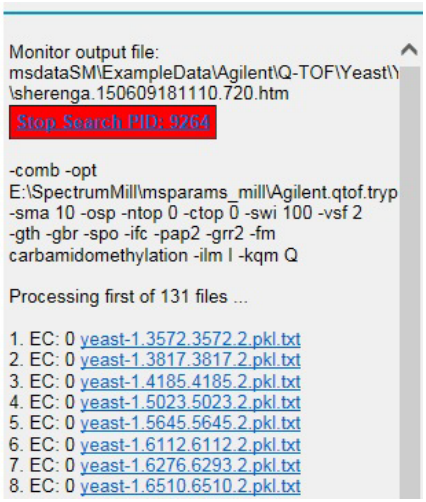


Figure 22 Display while Sherenga *de novo* Sequencing is in process

Exercise 2. Review results from Sherenga *de novo* sequencing

In this exercise, you review the *de novo* interpretations that you generated in the previous exercise.

Exercise 2. Review results from Sherenga de novo sequencing

Steps	Detailed Instructions	Comments
1 Navigate to the Sherenga *de novo* Summary page.	• From the Sherenga *de novo* Sequencing page or the Spectrum Mill home page, click the link to the ***de novo* Summary** page.	
2 Check that your **Data Directory** is set to **yeast\y1**.	• If you have just been working with this data, the data directory should already be set correctly. If not, click **Select...** to select the **y1** folder.	
3 Set the parameters as shown in Figure 23.	• Set Min. vertex score to 2.	

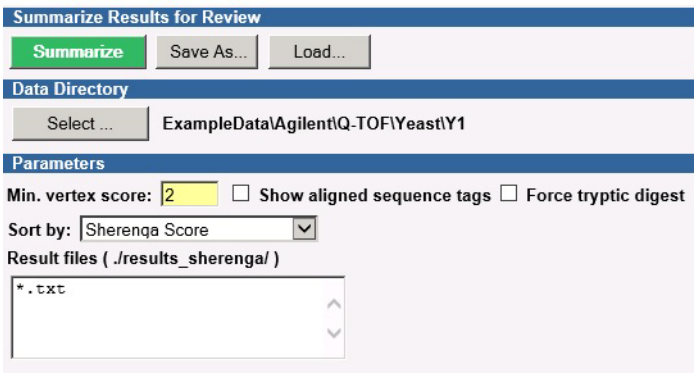


Figure 23 Settings for Sherenga Summary page

4 Summarize the results.	• Click **Summarize**.	

Exercise 2. Review results from Sherenga de novo sequencing (continued)

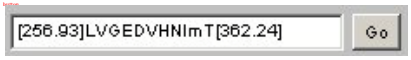
Steps	Detailed Instructions	Comments
5 View the overall results table.	**a** Check to see that your results look like the Sherenga Summary results shown in Figure 24 on page 62. **b** Adjust the size of this section to see the results better. • Rest your mouse over the page section split line. • When the pointer becomes a double-sided arrow, drag the arrow to move the split line.	File names are in the format: Data_File_Name.aaaa.bbbb.c.pkl.txt, where: • aaaa = first merged scan • bbbb = last merged scan • c = assigned precursor charge (0 means charge was ambiguous)

Results

SS	TL	TS	MS/MS Search Delta MH^+	Filename	Sherenga Sequence Tag	MS/MS Search Result
61	9			yeast-1.17135.17361.2.pkl.txt	[257.10]VYFDVEA[300.11][210.13][213.13]	
59	8			yeast-1.17013.17262.2.pkl.txt	[213.11]IDTDCEYAYIT[326.20]	
52	9			yeast-1.19114.19321.3.pkl.txt	[212.15][300.11][204.08]IEEGIM[227.16][211.06]	
49	8			yeast-1.18517.18517.2.pkl.txt	[602.29]MIT[228.13]ATAD[566.27]	
48	8	5.7	0.0	yeast-1.17237.17267.2.pkl.txt	FEEI[231.06]FC[224.12]	LEFEQLYDDK
47	8			yeast-1.17464.17464.2.pkl.txt	[229.12][170.09][241.14]V[361.12][154.10]EA[170.11][268.12]	
47	9			yeast-1.19496.19496.2.pkl.txt	[200.12]IWVII[328.14]Y[401.19]	
47	11			yeast-1.20438.20438.2.pkl.txt	PVSISET[242.15][236.18][247.10]	
47	11			yeast-1.20463.20463.2.pkl.txt	[172.10]EERA[228.11]I[154.03][329.17]	
47	10			yeast-1.21302.21302.2.pkl.txt	[1021.48][227.13]IAIAAY[200.11][499.19]	
47	8			yeast-1.7888.7888.2.pkl.txt	[242.13]ET[225.14][244.06]V[225.15]	
45	10			yeast-1.12462.12462.2.pkl.txt	[334.14][282.12]SII[144.12]N[215.10]V[412.16]	
42	11			yeast-1.21472.21472.2.pkl.txt	[386.18]IVIG[468.34]I[269.19]	
41	10			yeast-1.15530.15530.2.pkl.txt	[300.11][226.17]T[301.15]G[225.09][170.09]	
41	8			yeast-1.6276.6293.2.pkl.txt	[142.11]SQ[184.10]PMR	
39	8			yeast-1.22021.22021.2.pkl.txt	[673.30][142.07]YIVA[184.12]I[355.29][154.08]V[455.19]	
39	8			yeast-1.4185.4185.2.pkl.txt	[200.08]TAA[251.11][322.16][154.07]	
38	10			yeast-1.10949.10949.2.pkl.txt	[302.14][242.08]TVIIG[800.40]	
38	10			yeast-1.15612.15612.2.pkl.txt	[248.12]D[243.10]YNIID[170.11][174.06][199.13]	
37	9			yeast-1.13123.13123.2.pkl.txt	[213.11]EIGT[265.16][196.11][196.10]	
37	8			yeast-1.14340.14340.2.pkl.txt	[201.07]IEE[226.14]A[248.11][194.10]	

Figure 24 Sherenga Summary results

Exercise 2. Review results from Sherenga de novo sequencing (continued)

Steps	Detailed Instructions	Comments
6 Review detailed results for a single spectrum. • Review the MS for the seventh filename in the Spectrum Viewer. • Make sure the Sherenga 1 sequence is in white. • Display your own sequence if you choose. • The Results, Spectrum Viewer and detailed Sherenga results are each in separate panes. You may have to move the Results pane to see the Spectrum Viewer, for example.	**a** Click a link under the 3rd **Filename** header. Check that you see: • The Spectrum Viewer, as shown in Figure 25 on page 64. (You may need to move the page section split line again.) • Detailed Sherenga results, as shown in Figure 26 on page 64. **b** If you see an **MSTag 1** sequence at the top of the Spectrum Viewer, click the **Sherenga 1** sequence and make sure the sequence turns from gray to white background. If you do not see an **MSTag1** sequence, skip this step. **c** If there are multiple *de novo* interpretations, use the Rank arrow buttons to cycle through them.  **d** To display your own sequence above the spectrum, type the sequence in the box and then click the **Go.** 	• Note that the detailed Sherenga results list multiple Sherenga interpretations for the same spectrum. The most probable interpretation (Sherenga 1) is initially displayed at the top of the Spectrum Viewer. • If there are MS/MS Search results for these spectra, these are displayed along with the Sherenga interpretation. This makes it easy to compare the two. • To learn more about the Spectrum Viewer, see "Exercise 2. Examine results with the Spectrum Viewer" on page 37.
7 Repeat step 6 for as many spectra as you care to view.		• While *de novo* sequencing gives you valuable insight into possible spectral interpretations, you do not validate these results, and they are not included in final results summaries. • However, the Sherenga results are saved and can be reviewed at any time from the Sherenga *de novo* Summary page.

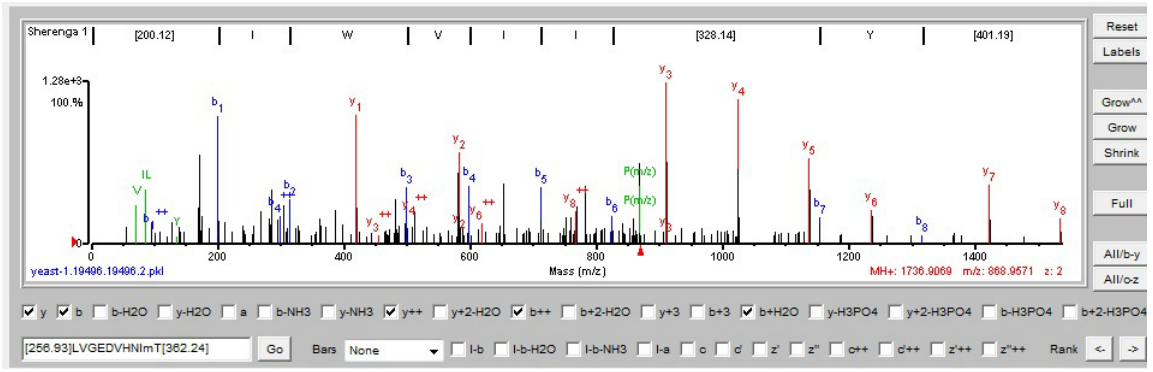


Figure 25 Spectrum Viewer with Sherenga results

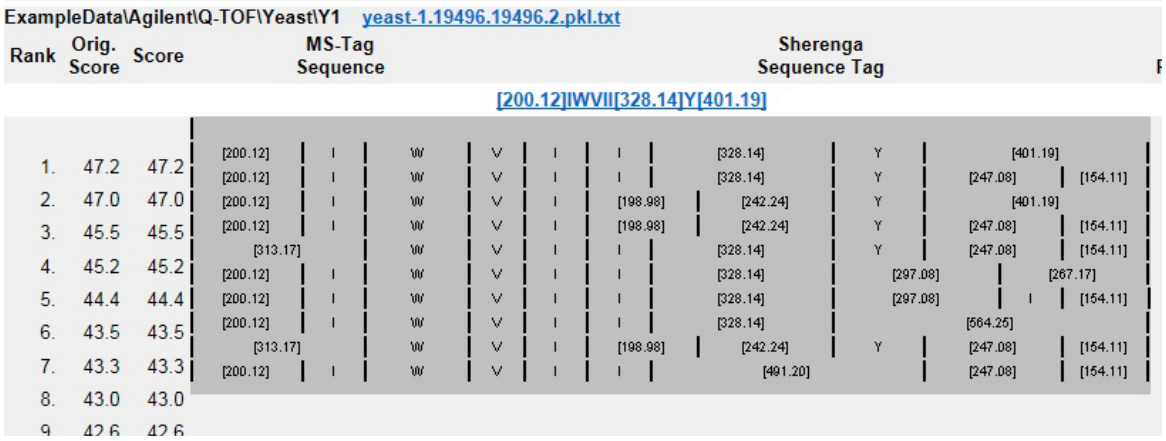


Figure 26 Detailed results from Sherenga *de novo* Sequencing

3 Unknown Modifications: Mass Gap Search

This chapter shows how to search MS/MS spectra to identify unsuspected modifications. To search for modifications that you do not suspect, you conduct a “mass gap” search. This search is named “mass gap” because the modification produces a gap (a mass shift) between the masses in the experimental MS/MS spectrum and the spectrum that would result from the unmodified sequence in the database.

If you expect that certain modifications are present in your sample, then you search specifically for those. (See “Exercise 2. Search the database” on page 17 of Chapter 1.) It is best to do variable modifications searches prior to a mass gap search because variable modifications searches are more specific and reduce the chance of false positive identifications.

This chapter uses the Agilent Q-TOF/OGE example data to illustrate the mass gap search. You can conduct a mass gap search with any MS/MS data, but the chance of a false positive result diminishes greatly when MS/MS spectra exhibit greater mass accuracy.

You conduct a mass gap search as part of an iterative workflow. You first run a variable modifications search and validation for those modifications you know are present.You then run a mass gap search. The exercises in this chapter demonstrate these steps.

Exercise 1. Extract, search and autovalidate Q-TOF/OGE data

To search for unknown modifications with the Q-TOF/OGE data, you must first extract, search and autovalidate the data as you did in Exercise 1 of Chapter 2 of this guide for the yeast data. This is the first step of an iterative workflow that uses the variable modifications search.

As you do this exercise, you do not need to save settings in parameter files.

Exercise 1. Extract, search and autovalidate Q-TOF/OGE data

Steps	Detailed Instructions	Comments
1 Extract the QTOF/OGE folder. • Remove prior results	• Follow the steps for extraction in Chapter 1, "Exercise 1. Extract the spectra" on page 11, except substitute the QTOF/OGE data. • Mark the **Remove all prior results** check box. • Under **Data Directories**, click **Select** and open the **msdataSM/ExampleData/QTOF/OGE folder.**	
2 Search the MS/MS spectra in Variable modifications mode. • Remove prior MS/MS search results • Search the SwissProt.ecoli database.	• Follow the steps for the MS/MS variable modifications search in Chapter 1 "Exercise 2. Search the database" on page 17, except substitute the QTOF/OGE data and the SwissProt.ecoli database. If the SwissProt.ecoli database is not listed, use the Protein Databases page to create the ecoli subset. • Mark the **Remove all prior MS/MS search results** check box. • From the Database list, select **SwissProt.ecoli**.	
3 Autovalidate the results. • Clear all prior validated results.	• Follow the steps for autovalidation in Chapter 1, "Exercise 3. Autovalidate high-quality results" on page 24, except substitute the QTOF/OGE data. • Click **Clear All**.	

Exercise 2. Search for unknown modifications

In this exercise, you conduct a mass gap search against database hits that you validated previously.

The MS/MS Search software provides identity, variable modifications, and homology search modes.

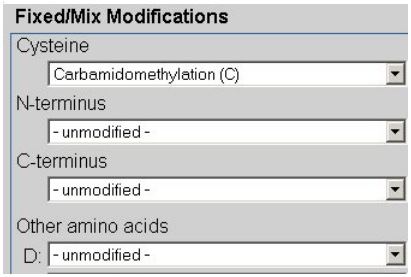
- In identity mode, you search for exact matches in the selected database.
- In variable modifications mode, you search for post-translational and chemical modifications that occur at a fraction of the available sites.
- In homology modes, you search for single amino acid substitutions relative to the database (as well as any variable modifications you have selected).

When you select one of the homology modes and click the option for **Unassigned single mass gap**, the search looks for an unexpected modification (a mass gap). Note that you cannot simultaneously search variable modifications in combination with the unassigned single mass gap.

Exercise 2. Search for unknown modifications

Steps	Detailed Instructions	Comments
1 Navigate to the MS/MS Search page.	• From the Spectrum Mill home page, click the link to MS/MS Search.	• There are a number of ways to navigate to the MS/MS Search page.
2 Check that your **Data Directory** is set to **ExampleData\Agilent\ Q-TOF\OGE**	• If you have just been working with the Agilent Q-TOF data set, your data directory should already be set correctly. If not, click **Select...** to select the **OGE** folder.	
3 Set **Search Parameters**. • Database: SwissProt.ecoli • Search previous hits.	**a** For **Database**, select **SwissProt.ecoli**, as before. **b** Clear the check box for **Remove all prior MS/MS Search results**. **c** Mark the check box next to **Search previous hits.**	• When you mark the check box to Search previous hits in MS/MS Search, if you have not created a saved results file, then the search page automatically creates one for you.

Exercise 2. Search for unknown modifications (continued)

Steps	Detailed Instructions	Comments
4 Choose the appropriate cysteine modification, if necessary.	**a** Click **Choose...** near the middle of the page.  **b** Under the **Cysteine** heading, select **Carbamidomethylation**  **c** Click **OK.** The name of the modification appears as in Figure 27 on page 71.	• This is the default modification. • For more information about choosing modifications, see the online help. • Your system administrator can configure custom modifications.

Exercise 2. Search for unknown modifications (continued)

Steps	Detailed Instructions	Comments
5 Set up the **Search Mode**. • Calculate reversed database scores. • Turn on dynamic peak thresholding. • Turnoff Discriminant Scoring. • Set up the Homology mode for all mutations and an unassigned single mass gap.	**a** Make sure that the check box for **Calculate reversed database scores** is marked. (This is the default setting.) **b** Mark the check box for **Dynamic peak thresholding**. **c** From the **Discriminant Scoring** list, select **Off**. (This is the default setting.) **d** Change the **Search mode** to **Homology - All mutations.** **e** Click the option for **Unassigned single mass gap**.	• When you **Calculate reversed database scores**, you search against peptide sequences in their forward and inverted directions. If you obtain similar scores for both searches, you may have a false positive. • **Dynamic peak thresholding** is a scoring enhancement that enables identification of more low-abundance and short-chain peptides. • The **discriminant score** is calculated as a weighted linear combination of a number of metrics (such as score, SPI, etc.), where the weights determine the relative importance of each of the component metrics. Select **Off** unless you want to use the Autothresholds - Discriminant mode to autovalidate.

Exercise 2. Search for unknown modifications (continued)

Steps	Detailed Instructions	Comments
6 Set the other parameters as shown in Figure 27 on page 71. • Note that +/- 130 Da for the precursor mass shift is a reasonable default in most cases. • Maximize CPUs.	**a** Examine the items in red text carefully, since these are the ones you may need to change when you process your own samples. **b** Click a blue section divider bar to display help for that section of the page.	• You do not mark the check box to **Remove all prior MS/MS results** because you want to combine the results of this search with the results of the previous one. • If the precursor mass shift range is outside +/- 130 Da, the unknown modification will not be detected during the search, but increasing the range value leads to a significant increase in search time.
7 Start the search.	**a** Click **Start Search**. **b** View progress in the area to the right of the MS/MS Search page. **c** When **Monitor Results** appears, click the link. **d** Scroll to the *top* of the **Results** area to see the message that indicates that the search is finished.	• For the same data set and database, homology searches take longer than identity searches because there are more permutations to consider. • If you want to stop the search, click the red **Stop Search PID: xxx** link at the top of the **Results** section. Also go to the Tool Belt, click **Stop process** and make sure there are no other processes running. If there are, stop them. See the *Application Guide* for further instructions. Or you can use the Request Queue to delete the search.

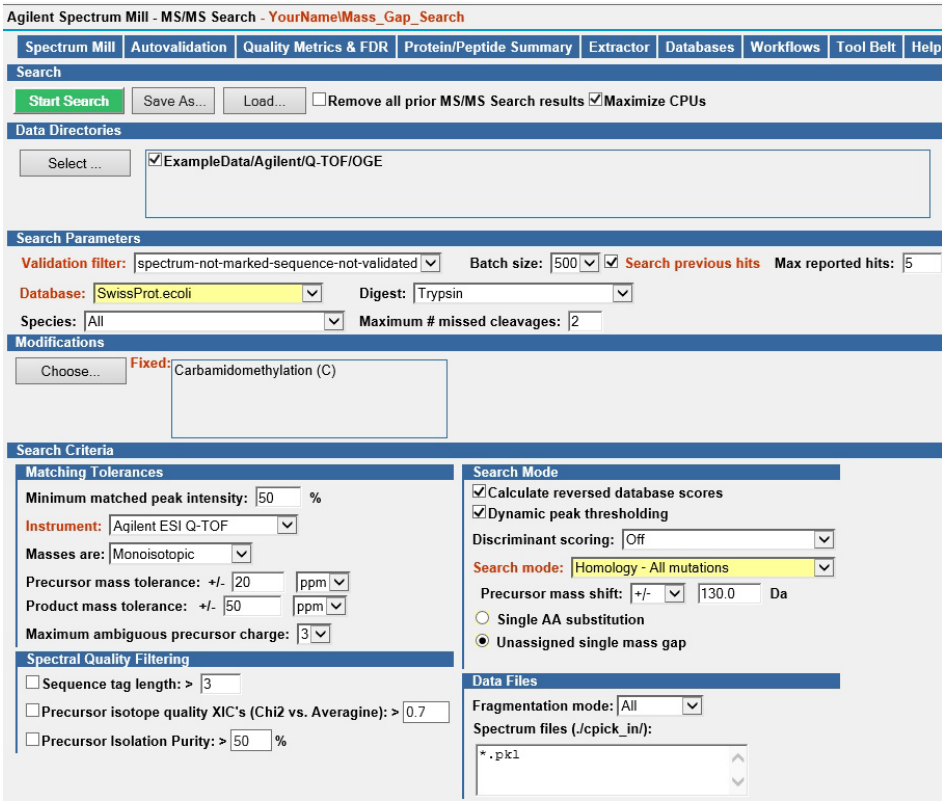


Figure 27 MS/MS Search settings for search of previous hits for unknown modifications

Exercise 3. Examine the list of validated peptides

In this exercise, you examine the list of peptides you have validated from both the variable modifications mode search and the mass gap search.

Exercise 3. Examine the list of validated peptides

Steps	Detailed Instructions	Comments
1 Autovalidate using the Auto thresholds strategy with the Peptide step. • Do not clear the previous autovalidation results.	• Follow the instructions for autovalidation in Chapter 1, "Exercise 3. Autovalidate high-quality results" on page 24.	• Autovalidation is not usually recommended for the homology modes because of the increased possibility of false positives when so many permutations are considered. But you can monitor the false positives with the FDR calculations.
2 Navigate to the Protein/Peptide Summary page.	• From the Spectrum Mill home page, click the link to the Protein/Peptide Summary page.	• There are links to this page from many other Spectrum Mill pages.
3 Check that your **Data Directory** is set to **QTOF\OGE**.	• If you have just completed autovalidation, your **OGE** data directory should already be set correctly. If not, click **Select...** to select the **OGE** folder.	
4 Set the Mode to Peptide - Spectrum Match.	• From the **Mode** list, select **Peptide - Spectrum Match**.	• The Protein/Peptide Summary page offers a variety of results display modes. • These allow you to organize results by proteins, peptides, or samples. • For more details, see the MS/MS Data Review and Validation chapter in the *Application Guide*.
5 Filter only valid results.	• From the **Filter results by** list, select **Valid**.	• The **Filter results by** setting selects from your data only the results that match your setting. • In this case, it selects only the results that match the **valid** setting (which was designated during autovalidation).

Exercise 3. Examine the list of validated peptides (continued)

Steps	Detailed Instructions	Comments
6 Set the **Sort** parameters such that the unknown modifications are grouped at the top and/or bottom of the results table.	• For **Sort peptides by**, select **Delta mass shift**.	• The **Sort** parameters determine how the data are sorted in the results display. • In other modes, you may have an option for **Sort proteins by**. If you temporarily select the **Protein Summary Details** mode, you see this option. (Be sure to reset **Mode** to **Peptide - Spectrum Match** before you continue.)
7 Filter peptides. Ensure that these parameters are set as follows: • Score > 0 • % SPI > 0	**a** Under **Filter peptides by** for **Score, type 0**. **b** For **% SPI,** type **0**.	• By setting the filters to zero, you ensure that you display all valid hits, regardless of score.
8 Set the **Validation preset** to **None**.		
9 Set the **Review Fields**. Add these to the Review Fields: • Fwd-Rev score • Rank 1-2 score • Var mod sites • Delta mass	**a** Mark the check box for **For-Rev score** and **Rank 1-2 score**. **b** To see the modifications, mark the check boxes for **Var mod sites** and **Delta mass**. **c** Mark and clear check boxes so that your page looks like that in Figure 28 on page 74.	• **Review Fields** determine what information you see in the final results summary. • The default settings are shown in the online help. To access the online help, click the **Help** button for the page.

Exercise 3. Examine the list of validated peptides (continued)

Steps	Detailed Instructions	Comments
10 Summarize all results.	• Click **Summarize.**	
11 Examine the summary report. • Make sure headings are consistent with marked Review Fields. • Look for unknowns in the Variable Sites column.	**a** Check that your results are similar to those in Figure 29 on page 75 and Figure 30 on page 75. Since the public databases are constantly updated, your results may vary slightly. **b** Check that the data display includes the additional headings **Variable Sites**, **MH$^+$ Mass Shift (Da)**, and **MH$^+$ Error (ppm)**. **c** Note that the columns for **Variable Sites** and **MH$^+$ Mass Shift (Da)** summarize the modifications. To see results of the mass gap search, look for **Unknown** in the column for **Variable Sites**.	• This summary includes results validated from both identity mode and homology mode/mass gap searches. • Because you sorted by **Delta mass shift**, the software groups the results of the mass gap search at the top and bottom of the results table. • For a mass gap search, the column for **MH$^+$ Error (ppm)** shows the mass error relative to the unmodified peptide. For a variable modifications search (where the identity and mass of the modification are known), this column displays the mass error relative to the modified peptide.

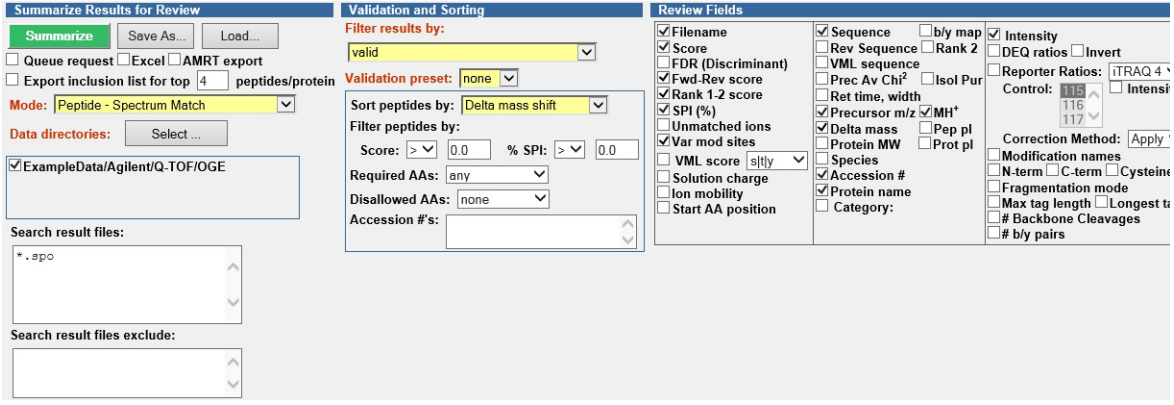


Figure 28 Settings to display results of mass gap search

#	Filename	z	Score	Fwd-Rev Score	Rank1-Rank2 Score	SPI (%)	Spectrum Intensity	Variable Sites	Sequence	m/z Measured (Da)	MH+ Matched (Da)	MH+ Mass Shift (Da)	MH+ Error (ppm)	Accession #	Protein Name
1	ec-oge-2.2049.2058.2	2	13.07	4.03	1.40	71.3	4.64e+005	Unknown	(R)LWDKETLEK(M)	645.8306	1161.615	129.0388	99979.4	P0A7D7	Phosphoribosylaminoimidazole-succinocarb
2	ec-oge-2.1737.1746.2	2	10.66	3.02	5.80	73.2	3.77e+005	Unknown	(R)GVEILK(G)	387.2221	658.413	115.0234	148717.3	P0A7E5	CTP synthase
3	ec-oge-2.2954.2954.2	2	9.63	1.30	2.43	57.6	2.46e+005	Unknown	(R)WLELGLTDEEK(Q)	717.8737	1332.668	102.0719	71143.1	P0A8V2	DNA-directed RNA polymerase subunit beta
4	ec-oge-2.1044.1053.2	2	7.54	3.60	1.80	64.9	3.29e+005	Unknown	(K)NQASNDLPN(-)	536.7651	972.438	100.0847	93317.1	P0AET2	Acid stress chaperone HdeB
5	ec-oge-2.1645.1645.2	2	7.90	1.61	1.25	75.4	4.00e+005	Unknown	(R)EADLSLK(R)	438.0069	775.420	99.5869	113812.7	P07395	Phenylalanine--tRNA ligase beta subunit
6	ec-oge-2.2868.2876.2	2	8.97	1.07	8.97	57.8	2.92e+005	Unknown	(K)DAGRIAGLEVK(R)	614.3455	1128.637	99.0466	80677.6	P0A6Y8	Chaperone protein DnaK
7	ec-oge-2.0887.0887.2	2	7.54	1.84	1.98	64.5	1.91e+004	Unknown	(K)LAKDDAEK(Y)	494.7490	889.462	99.0282	100181.2	P0A6Z3	Chaperone protein HtpG
8	ec-oge-2.1173.1179.2	2	7.51	7.51	3.11	65.3	7.43e+005	Unknown	(K)LTDDLAK(A)	437.7272	775.420	99.0274	113245.8	P46837	Protein YhgF
9	ec-oge-2.3477.3484.2	2	9.66	3.64	3.74	69.1	9.79e+004	Unknown	(R)LYAAGRAGERFGVR(N)	809.9388	1522.824	96.0466	59329.4	P09831	Glutamate synthase [NADPH] large chain
10	ec-oge-2.2866.2874.2	2	7.67	2.13	3.03	71.0	4.94e+005	Unknown	(R)MEFPEPVISIAVEPK(T)	890.9622	1685.882	95.0352	53363.0	P0A6M8	Elongation factor G
11	ec-oge-2.1137.1137.2	2	7.46	7.46	7.46	62.5	1.47e+005	Unknown	(K)GFADARLAK(L)	522.2236	948.526	94.9137	90962.3	P00968	Carbamoyl-phosphate synthase large chain
12	ec-oge-2.2960.2968.2	2	8.86	1.17	2.10	60.7	1.01e+006	Unknown	(R)RNGEKQLIK(T)	587.3120	1085.643	87.9741	74959.8	P16659	Proline--tRNA ligase
13	ec-oge-2.1629.1638.2	2	8.36	2.04	8.36	63.7	2.21e+005	Unknown	(K)VIDGQINLR(D)	557.3088	1027.590	86.0208	77245.0	P08997	Malate synthase A
14	ec-oge-2.1229.1232.2	2	13.01	4.74	7.32	74.8	4.99e+005	Unknown	(R)DYEFASK(Y)	473.1947	859.383	85.9988	90967.3	P07813	Leucine--tRNA ligase
15	ec-oge-2.2425.2432.2	2	10.23	2.47	5.10	72.9	4.49e+005	Unknown	(R)IDESGFDIR(L)	568.7884	1051.506	85.0641	74842.8	P0AG07	Ribulose-phosphate 3-epimerase
16	ec-oge-2.1568.1600.2	2	12.74	3.15	3.10	69.0	5.79e+006	Unknown	(R)ITEYNVELVK(Y)	643.8080	1207.657	78.9518	61364.3	P09151	2-isopropylmalate synthase
17	ec-oge-2.1906.1906.2	2	10.17	4.12	1.16	90.4	2.71e+005	Unknown	(K)DIALGEEFVNK(-)	654.8374	1234.631	74.0362	56573.7	P61889	Malate dehydrogenase
18	ec-oge-2.4383.4383.3	3	9.99	3.34	9.99	75.7	8.34e+004	Unknown	(K)SGFINKEGTRPAVVIPTNEELVIAQDASR(L)	1058.2356	3111.643	61.0490	19242.0	P0A6A3	Acetate kinase
19	ec-oge-2.1045.1052.2	2	10.64	6.80	10.64	75.2	7.34e+005	Unknown	(R)TADEQQAPRR(E)	616.2976	1171.581	60.0066	48722.9	P21513	Ribonuclease E
20	ec-oge-2.2089.2096.2	2	8.60	3.04	8.60	55.2	1.08e+005	Unknown	(K)DLSDVTLGQFAGK(R)	704.3620	1350.690	57.0268	40510.2	P0A862	Thiol peroxidase
21	ec-oge-2.3070.3070.2	2	10.57	5.82	10.57	90.1	3.16e+005	Unknown	(K)FADVACAGPLLAAELDALGK(A)	1030.0318	2002.031	57.0248	27694.7	P0A799	Phosphoglycerate kinase

Figure 29 Mass gap search results, top

423	ec-oge-2.2821.2828.2	2	7.73	7.73	3.13	65.2	3.74e+006	Unknown	(K)YAKVKPIR(D)	473.8202	974.615	-27.9815	-29558.9	P09373	Formate acetyltransferase 1
424	ec-oge-2.2604.2612.2	2	8.31	1.54	3.63	54.5	4.14e+005	Unknown	(R)IEPSLEAAFVDYGAER(H)	868.4287	1766.860	-31.0094	-17864.1	P21513	Ribonuclease E
425	ec-oge-2.1866.1872.2	2	7.83	1.98	7.83	58.7	9.81e+005	Unknown	(R)ELDAQPTGFLDSR(I)	708.8681	1448.702	-31.9726	-22567.9	P30850	Exoribonuclease 2
426	ec-oge-2.2724.2728.2	2	8.51	1.43	1.56	55.9	3.00e+005	Unknown	(-)MSVIKMTDLDLAGK(R)	743.3882	1521.802	-36.0323	-24251.6	P0A799	Phosphoglycerate kinase
427	ec-oge-2.0957.0963.2	2	9.19	2.69	3.61	64.2	1.38e+005	Unknown	(K)VNDEGIIEDAR(F)	597.2560	1230.596	-37.0912	-31077.6	P0ACD4	Iron-sulfur cluster assembly scaffold protein
428	ec-oge-2.3966.3966.3	3	7.75	1.31	7.75	65.6	1.89e+005	Unknown	(R)VIDGADGARFITIINNTLSDIRR(L)	830.1346	2530.374	-41.9845	-16872.1	P06959	Dihydrolipoyllysine-residue acetyltransferase
429	ec-oge-2.1514.1521.2	2	12.62	3.96	1.70	76.3	4.13e+005	Unknown	(R)LNNAPADSWR(K)	548.7825	1143.554	-46.9964	-42858.1	P46837	Protein YhgF
430	ec-oge-2.2020.2028.2	2	11.97	8.47	11.97	72.1	2.42e+006	Unknown	(K)CSEFGDAIIENM(-)	669.2868	1385.571	-48.0049	-35889.7	P08200	Isocitrate dehydrogenase [NADP]
431	ec-oge-2.2405.2412.2	2	8.21	8.21	8.21	68.4	4.25e+006	Unknown	(R)ELNEQLEENLGYK(L)	763.8594	1578.765	-52.0531	-34094.9	P00864	Phosphoenolpyruvate carboxylase
432	ec-oge-2.4403.4403.3	3	7.43	7.43	7.43	81.0	1.48e+005	Unknown	(-)MISGILASPGIAFGKALLLKEDEIVIDR(K)	972.5262	2969.674	-54.1105	-18559.2	P08839	Phosphoenolpyruvate-protein phosphotrans
433	ec-oge-2.1120.1129.2	2	9.48	0.62	9.48	50.4	3.59e+005	Unknown	(R)FGVEDGSVEGLR(A)	605.2715	1264.617	-55.0810	-45539.0	P0A850	Trigger factor
434	ec-oge-2.1969.1974.2	2	7.32	1.38	1.19	65.3	2.32e+005	Unknown	(R)ELSFLNSGVSIR(L)	633.3203	1321.711	-56.0778	-44308.1	P0AES6	DNA gyrase subunit B
435	ec-oge-2.2493.2500.2	2	7.34	0.09	2.08	60.3	3.64e+005	Unknown	(K)VSTKLTVLDSLSK(I)	667.3850	1390.815	-57.0525	-42775.6	P0A799	Phosphoglycerate kinase
436	ec-oge-2.3225.3225.2	2	7.70	2.28	2.95	71.2	1.77e+004	Unknown	(K)ANRSRMNVPFQLADSALDK(L)	1036.5326	2133.087	-61.0290	-29453.3	P23721	Phosphoserine aminotransferase
437	ec-oge-2.1901.1910.2	2	10.11	2.02	1.39	68.6	1.73e+005	Unknown	(R)LIAEAMDKVGK(E)	553.7936	1174.650	-68.0701	-61513.9	P0A6F5	60 kDa chaperonin
438	ec-oge-2.2966.2972.2	2	8.20	1.10	2.79	62.8	3.31e+005	Unknown	(R)LEKAFQELNAIDVL(-)	763.4240	1602.874	-77.0330	-50485.6	P0AB71	Fructose-bisphosphate aldolase class 2
439	ec-oge-2.1752.1762.2	2	7.83	1.09	1.23	63.8	4.71e+005	Unknown	(R)DAEANAEADRK(F)	555.2470	1189.544	-80.0576	-72157.3	P0A6Y8	Chaperone protein DnaK
440	ec-oge-2.1061.1067.2	2	7.92	1.71	2.90	58.5	5.43e+005	Unknown	(K)NYPDRVLPDR(Q)	581.2607	1244.638	-83.1242	-71565.4	P08997	Malate synthase A
441	ec-oge-2.1902.1902.2	2	9.17	4.28	9.17	78.0	1.87e+005	Unknown	(-)MSEQHAQGADAVVDLNNELK(T)	1040.4998	2169.024	-89.0317	-42803.8	P0A8N3	Lysine--tRNA ligase
442	ec-oge-2.1905.1905.2	2	9.78	9.78	9.78	68.3	1.87e+005	Unknown	(-)MSEQHAQGADAVVDLNNELK(T)	1040.4998	2169.024	-89.0317	-42803.8	P0A8N3	Lysine--tRNA ligase
443	ec-oge-2.1582.1586.2	2	8.84	1.32	8.84	68.8	3.71e+005	Unknown	(K)IIDAARAREAAR(R)	608.3131	1312.744	-97.1256	-79898.0	P0AES6	DNA gyrase subunit B
444	ec-oge-2.2973.2976.2	2	10.45	2.03	2.21	81.6	3.76e+005	Unknown	(K)LAIVPDHPLLK(D)	555.3323	1215.746	-106.0886	-95604.8	P30850	Exoribonuclease 2

Figure 30 Mass gap search results, bottom

Exercise 4. Examine details of search results to determine modification sites

In this exercise, you examine details that help you localize the site of the peptide modification.

Exercise 4. Examine details of search results to determine modification sites

Steps	Detailed Instructions	Comments
1 If you have not already done so, display a Protein/Peptide Summary report.	• Follow steps in "Exercise 3. Examine the list of validated peptides" on page 72.	
2 Display MS/MS Search results.	**a** Under the **Filename** header, click the 21st entry. **b** Check the bottom of the browser window, where you see the annotated spectrum, as shown in Figure 31 on page 77. **c** Note the presence of star-ions (**b^*_2**, **b^*_3**, **b^*_4**, **b^*_5**). These are the complements of the detected y-ions, and they contain the modification. In other words, they are the b-ions plus the mass shift.	• Note that **MH^+ Mass Shift** shows an addition of 57 Da.
3 Expand the section of the page that shows the MS/MS Search results.	**a** To adjust the size of this section, rest your mouse over the split line. **b** When the pointer becomes a double-sided arrow, drag the arrow to move the split line. 	
4 Scroll up to see more results.	**a** Scroll up until you see the heading for **Detailed Results**. **b** Check that your display resembles that in Figure 32 on page 78.	
5 Examine the table where the ions are assigned.	**a** Note that the table shows y-ions and b^*-ions, but no b-ions. This means that the modification is on the N-terminus.	• The addition of 57 Da is likely carbamidomethylation of the N-terminus, an artifact of cysteine alkylation.

Exercise 4. Examine details of search results to determine modification sites (continued)

Steps	Detailed Instructions	Comments
6 Examine the sequence of the database hit.	**a** Look above the table of ion assignments to see a sequence or list of sequences. **b** Note the color-coded lines in the sequence below (from an identity-mode search) and read the comment to the right in this table. Sequence (K) F A D V/A/C/A/G/P L L/A A E L/D/A L/G K (A) **c** Examine the sequence in Figure 32 on page 78 to see the locations of identified y-ions in our example. (Again, there are no b-ions because the N-terminus is modified.)	• The color-coding of the amino acid sequence of the matched peptide from the database search shows: ◦ Red forward-slashes for locations of y-ions ◦ Blue backslashes for locations of b-ions (not b*-ions) ◦ Magenta pipes (vertical lines) for locations of both b- and y-ions • For mass gap results, the color-coded sequence helps you to identify the site of a modification by highlighting the existence of surrounding b/y ions.
7 (Optional) Learn more details about this view of the MS/MS Search results.	**a** See "Exercise 3. Examine details of search results" on page 42. **b** See the information about manual review and validation in Chapter 1 of the *Application Guide*.	

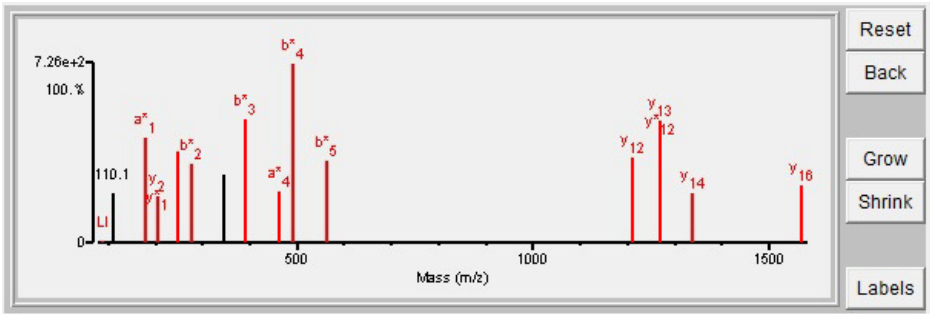
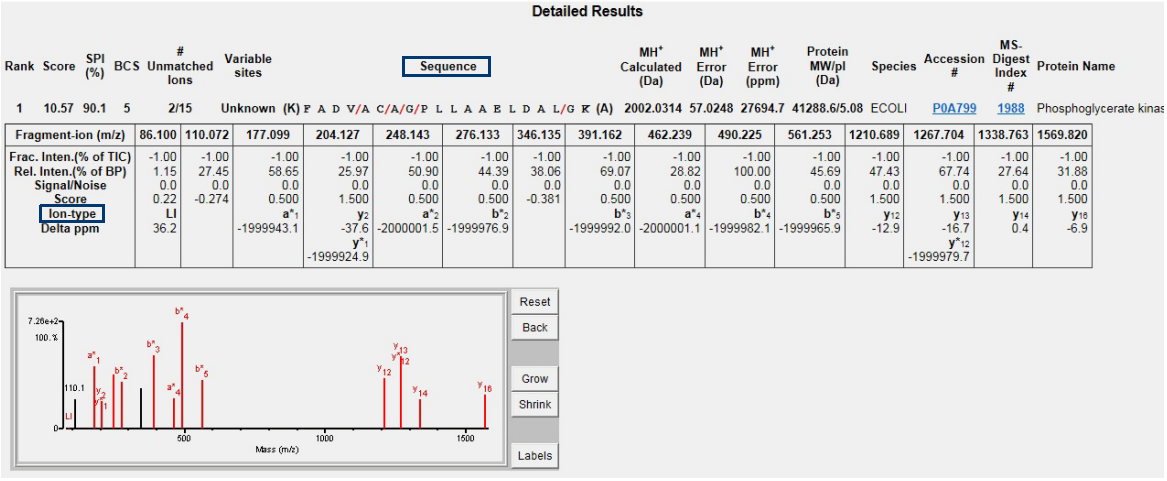


Figure 31 Spectrum from MS/MS Search results, showing star-ions



Detailed Results

Rank	Score	SPI (%)	BCS	# Unmatched Ions	Variable sites	Sequence	MH+ Calculated (Da)	MH+ Error (Da)	MH+ Error (ppm)	Protein MW/pI (Da)	Species	Accession #	MS-Digest Index #	Protein Name
1	10.57	90.1	5	2/15	Unknown (K)	F A D V/A C/A/G/P L L A A E L D A L/G K (A)	2002.0314	57.0248	27694.7	41288.6/5.08	ECOLI	P0A799	1988	Phosphoglycerate kinas

Fragment-ion (m/z)	86.100	110.072	177.099	204.127	248.143	276.133	346.135	391.162	462.239	490.225	561.253	1210.689	1267.704	1338.763	1569.820
Frac. Inten.(% of TIC)	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
Rel. Inten.(% of BP)	1.15	27.45	58.65	25.97	50.90	44.39	38.06	69.07	28.82	100.00	45.69	47.43	67.74	27.64	31.88
Signal/Noise	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Score	0.22	-0.274	0.500	1.500	0.500	0.500	-0.381	0.500	0.500	0.500	0.500	1.500	1.500	1.500	1.500
Ion-type	LI		a^*_1	y_2	a^*_2	b^*_2		b^*_3	a^*_4	b^*_4	b^*_5	y_{12}	y_{13}	y_{14}	y_{16}
Delta ppm	36.2		-1999943.1	-37.6	-2000001.5	-1999976.9		-1999992.0	-2000001.1	-1999982.1	-1999965.9	-12.9	-16.7	0.4	-6.9
				y^*_1									y^*_{12}		
				-1999924.9									-1999979.7		

Figure 32 MS/MS Search results, showing the spectrum, table of fragment-ion assignments, and sequence

Once you have used a mass gap search to find the unexpected modifications, the modifications, to get an accurate FDR, you must repeat the search and autovalidation steps from "Exercise 1. Extract, search and autovalidate Q-TOF/OGE data" on page 66. Be sure to remove prior search results and include your newly discovered modification in this new search.

4
Workflow Automation

This chapter provides familiarization exercises for automation of a basic workflow with the Agilent Spectrum Mill MS Proteomics Workbench. Workflow automation allows you to place a series of tasks into a queue; then the Spectrum Mill workbench executes them in the proper order. For example, you can automate a typical basic workflow for MS/MS data files from protein digests:

- Spectral extraction
- MS/MS search
- Autovalidation
- Protein/Peptide Summary
- False discovery rate (FDR) and quality metrics
- Data archive

You can also automate iterative workflows where you add additional searches and autovalidations, then summarize all the results.

Exercise 1. Create parameter files

The first step to set up a workflow is to save the settings you want to use in parameter files. Most Spectrum Mill pages allow you to save and load parameter files associated with the page. For example, Data Extractor, MS/MS Search, Autovalidation, and Protein/Peptide Summary all include buttons to save settings in parameter files, and to load settings from parameter files.

In all the cases mentioned above, you use the parameter files to build and execute workflows. Peptide Selector, MRM Selector, and Sherenga *de novo* Summary also let you save parameter files, but you do not use them with workflow automation.

Parameter files are stored on the Spectrum Mill server under **\SpectrumMill\millauto\your_own_folder**. Only one level of subfolder is allowed. We recommend that you assign subfolders to users or types of experiments or studies. Each task type stores (in a cookie) the last parameter file that you selected for that task. The program shows the selected parameter file in red in the title portion of the page.

Exercise 1. Create parameter files

Steps	Detailed Instructions	Comments
1 Create these parameter files, if you have not already done so: • extract-y1 • MSMS_Search • AV-auto-pep • AV-auto-protpol • psm • prot-sum	• Create parameter files as described in Chapter 1, "Normal Workflow – Interactive Processing of MS/MS Data". • The specific exercises where you create parameter files are: • "Exercise 1. Extract the spectra" on page 11 • "Exercise 2. Search the database" on page 17 • "Exercise 3. Autovalidate high-quality results" on page 24 • "Produce a Protein Summary Details Report" on page 29	• To create your own folder for the parameter files for these exercises, see Chapter 1, "Exercise 1. Extract the spectra", step 5 on page 15.
2 Verify that the parameter files are present on the server under **SpectrumMill\millauto\YourName**.	**a** On your server, open the **SpectrumMill\millauto\YourName** folder. **b** Verify that you see the following files: • **xtractor,extract-y1.params** • **mstag.MSMS_Search.params** • **autovalidation.AV-auto-pep.params** • **autovalidation.AV-auto-protpol.params** • **ppsummary.psm.params** • **ppsummary.prot-sum.params**	

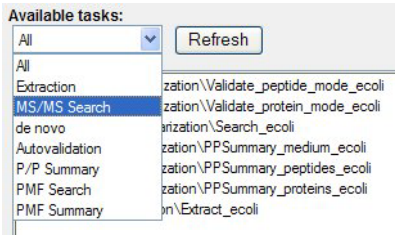
Exercise 2. Set up a workflow for automation

Workflow automation uses a Workflows page, an Edit Workflow page, and a Request Queue / Completion Log viewer. You access these pages through the Process Automation Tools section of the Spectrum Mill home page. This exercise primarily uses the Edit Workflow page.

Exercise 2. Set up a workflow for automation

Steps	Detailed Instructions	Comments
1 Navigate to the Workflows page.	• From the Spectrum Mill home page, click the **Workflows** link. 	
2 Display the Edit Workflow page.	**a** Select any workflow from which to create another workflow. **b** Click **Edit Workflow**.  **c** Verify that the Edit Workflow window appears.	• If no workflows are present, just click Edit Workflows and create the first workflow.
3 Add a Data Extraction task to the workflow. • Remove all tasks. • Add Extraction - YourName\extract-y1	**a** Select a **Workflow task**, then click **Remove**. **b** Repeat the first step until you have removed all tasks. **c** From the Available Tasks list, click **Extraction - YourName\extract-y1** **d** (Optional) Click the **View** button to view (but not change) the extraction parameters. **e** Verify that **Extraction - YourName\extract-y1** is still selected and click **Add**.	• This exercise assumes that you have already created the necessary parameter files (See Chapter 1 or "Exercise 1. Create parameter files" on page 80).

Exercise 2. Set up a workflow for automation (continued)

Steps	Detailed Instructions	Comments
4 Add an MS/MS Search task to the workflow. • YourName\MSMS_Search	**a** Click **MS/MS Search - YourName\ MSMS_Search**. **b** (Optional) Click the **View** button to view the search parameters. **c** Verify that **MS/MS Search - YourName\MSMS_Search** is still selected and click the **Add** button.	• If the list of available parameter files is long, you may select to view only a subset. Under **Available tasks**, select the subset you need. Be sure to change the setting back to **All** for the next step. 
5 Add the two Autovalidation tasks to the workflow. • YourName\AV-auto-pep • YourName\AV-auto-protpol	**a** Add the following tasks: • Autovalidation - YourName\ AV-auto-pep • Autovalidation - YourName\ AV-auto-protpol **b** Make sure these two tasks are in the correct order, with peptide mode first. If not, click a task and then click the **Up** or **Down** button to achieve the correct order.	• If you set up new parameter files after the Edit Workflow window is already open, click the **Refresh** button to cause them to appear in the list of available tasks.
6 Add a Protein/Peptide Summary task to the workflow. • Your Name\prot-sum	• Add the following task: • P/P Summary - YourName\ prot-sum	
7 Save the workflow. • Call it Fam_Workflow.	**a** Click **Save As**. **b** For **Folder**, select **YourName**. **c** For **Name**, type **Fam_Workflow.** **d** Click **Save**. Note that the name of the workflow appears in red near the top of the browser window. (Figure 33 on page 84)	
8 Close the Edit Workflow window.	**a** Click **Close** in the upper right corner. 	

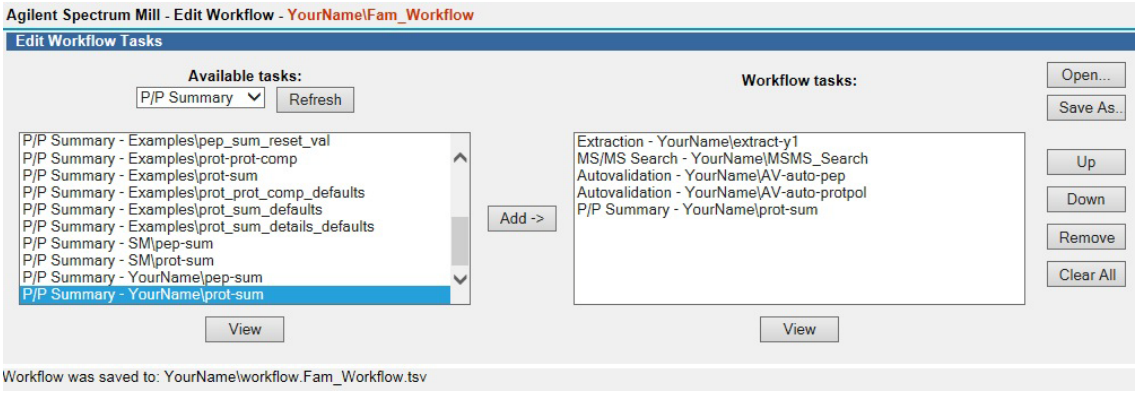
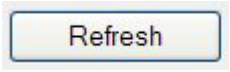


Figure 33 Edit Workflow window after new workflow is saved

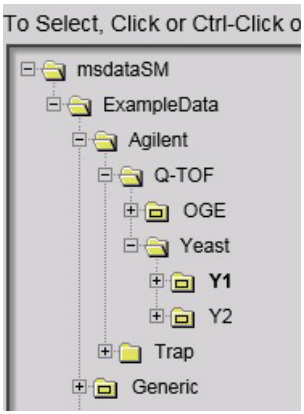
Exercise 3. Run the workflow

In this exercise, you use the Workflows page to view and run the workflow you created in the previous task.

Exercise 3. Run the workflow

Steps	Detailed Instructions	Comments
1 Go back to the Workflows page.	• This page should still be displayed. If not, display it again.	• See Figure 34 on page 86.
2 Make sure all the available workflows appear.	• Click the **Refresh** button. 	• If you create a new workflow while the Workflows page is already displayed, you will not see the new workflow in the list unless you refresh the screen.
3 Verify that the workflow you just created contains the correct tasks in the correct order.	**a** Click the name of the workflow (**YourName\Fam_Workflow**.) **b** Verify that it contains the correct tasks, in the correct order. **c** If necessary, click any of the tasks to view (but not change) the parameters.	

Exercise 3. Run the workflow (continued)

Steps	Detailed Instructions	Comments
4 Select the **msdataSM\ExampleData\Agilent\Q-TOF\yeast\y1** data directory. • Make sure that the name of the **y1** directory changes to a **bold** font. This indicates that it has been selected.	**a** Under **Data Directories**, click **Select...**. **b** Expand the directory tree and click **y1** to select that data file directory.  **c** Click **OK.** The name of the data directory appears as in Figure 34 on page 86.	• If you specify multiple data folders, the program processes them in parallel for extraction and search. After search is complete, the program autovalidates the data independently for each folder on the same CPU unless "Group proteins across..." is marked. After autovalidation, the program summarizes results for all folders for P/P Summary together. • All data folders whose results are summarized together must contain the same type of data (for example, all **.raw**, all **.d**, etc.). • If you have selected a group of folders that are to be processed independently, submit each folder one-at-a-time by marking only one folder so PP Summary does not process them together.
5 Execute the workflow.	**a** Click **Execute**.	

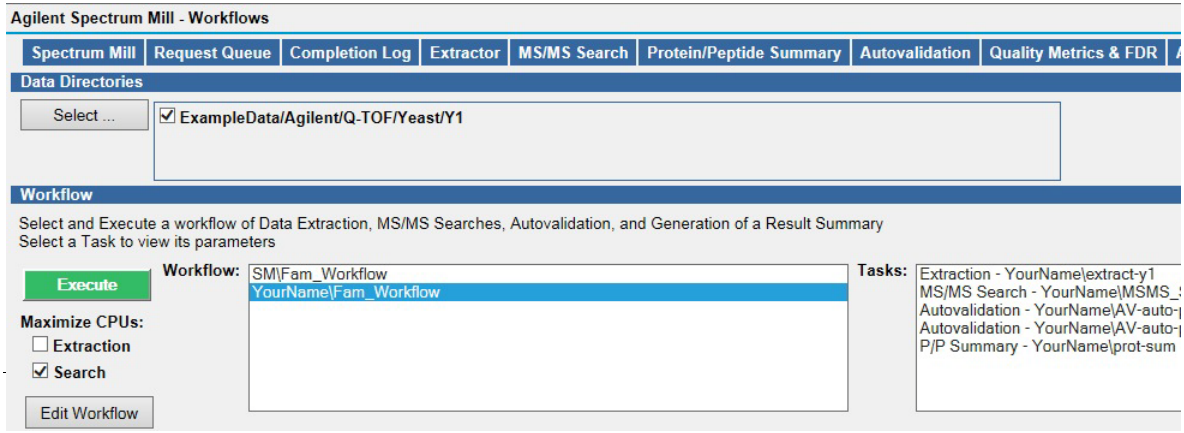
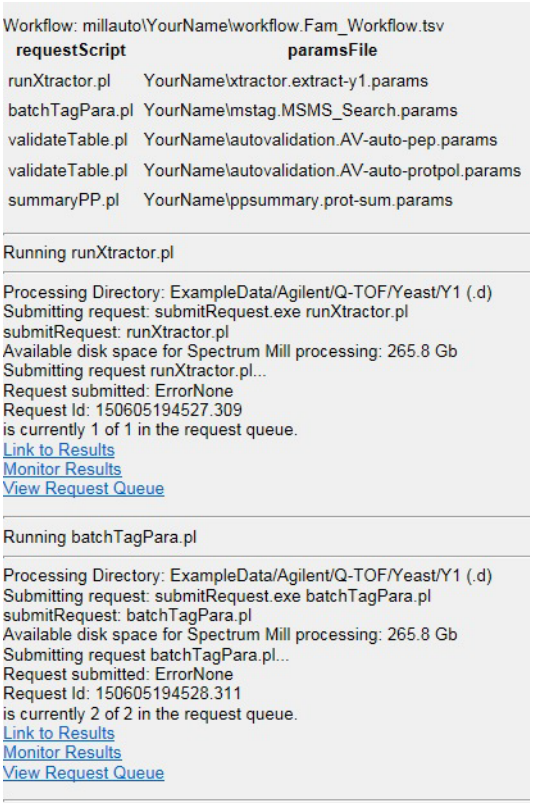


Figure 34 Workflows window



Workflow: millauto\YourName\workflow.Fam_Workflow.tsv

requestScript	paramsFile
runXtractor.pl	YourName\xtractor.extract-y1.params
batchTagPara.pl	YourName\mstag.MSMS_Search.params
validateTable.pl	YourName\autovalidation.AV-auto-pep.params
validateTable.pl	YourName\autovalidation.AV-auto-protpol.params
summaryPP.pl	YourName\ppsummary.prot-sum.params

Running runXtractor.pl

Processing Directory: ExampleData/Agilent/Q-TOF/Yeast/Y1 (.d)
Submitting request: submitRequest.exe runXtractor.pl
submitRequest: runXtractor.pl
Available disk space for Spectrum Mill processing: 265.8 Gb
Submitting request runXtractor.pl...
Request submitted: ErrorNone
Request Id: 150605194527.309
is currently 1 of 1 in the request queue.
Link to Results
Monitor Results
View Request Queue

Running batchTagPara.pl

Processing Directory: ExampleData/Agilent/Q-TOF/Yeast/Y1 (.d)
Submitting request: submitRequest.exe batchTagPara.pl
submitRequest: batchTagPara.pl
Available disk space for Spectrum Mill processing: 265.8 Gb
Submitting request batchTagPara.pl...
Request submitted: ErrorNone
Request Id: 150605194528.311
is currently 2 of 2 in the request queue.
Link to Results
Monitor Results
View Request Queue

Figure 35 Running Workflow

Exercise 4. View progress via the Request Queue and Completion Log viewers

The Request Queue Viewer shows a list of all tasks that are currently executing and those that are queued for execution. It lists the tasks in the order they were queued. Because some tasks depend upon earlier tasks, the tasks that are currently executing do not always appear at the top of the list.

The Completion Log Viewer shows a list of all tasks that have completed, with the most recent shown at the top.

Exercise 4. View progress via the Request Queue and Completion Log viewers

Steps	Detailed Instructions	Comments
1 Display the Request Queue Viewer.	**a** At the bottom of the Workflows window, click any **View Request Queue** link. **b** Verify that you see something like Figure 36 on page 89.	The following links are available: • Link to Results - displays the output html (up to whatever content is available). • Monitor Results - shows a window that continually updates as more results are available. • View Request Queue - shows the Request Queue
2 Become familiar with the Request Queue Viewer.	**a** Note that under **Task ID,** you see the following: ◦ The **Task ID** number, which you can click to see the results that have been created so far ◦ A link to **Monitor** the results **b** Note that under **Dependencies**, you see a list of tasks that must be completed before a given task can execute. For example, an MS/MS Search cannot execute until Data Extraction is complete.	• To see the status of requests after you first open the Request Queue Viewer, you must refresh the screen. Right-click the Viewer and select **Refresh**. • When you mark **Maximize CPUs**, tasks labelled with "P" process the batches, while those without "P" create the batches.

Exercise 4. View progress via the Request Queue and Completion Log viewers (continued)

Steps	Detailed Instructions	Comments
3 Monitor the task that is currently running.	**a** Find the task for which **Status** is **Running**. (Look for the word **Running** with green letters.) **b** Click the **Monitor** link. **c** Observe the progress for a few seconds.	
4 Display the Completion Log Viewer.	**a** In the Request Queue window, click **Completion Log**. See Figure 37 on page 90. **b** Click a **Task ID** to view the results of that task. **c** After the workflow is complete, click the final **Task ID** to see the final results.	• The protein summary report should be identical to the one produced in "Produce a Protein Summary Details Report" on page 29. (Figure 11 on page 32)

There are 5 requests in the queue. Available memory: 6.7 Gb of 12.9 Gb

Remove

✕	#	Task Id	Status	Task Type	Data Directory	Dependencies
	1	150821093216.141 Monitor	Running	Extraction	ExampleData/Agilent/Q-TOF/Yeast/Y1	none
☐	2	150821093218.142P	Queued	MS/MS Search	ExampleData/Agilent/Q-TOF/Yeast/Y1	150606093439.470
☐	3	150821093219.144	Queued	Autovalidation	ExampleData/Agilent/Q-TOF/Yeast/Y1	150606093439.470

Figure 36 Request Queue Viewer

Request Queue | Completion Log | Help

Task Id	Task Type	Data Set Parameters	Status	Completion Time	User
150606093444.475	P/P Summary	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Sat Jun 06 09:38:14 2015	Anonymous
150606093442.474	Autovalidation	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Sat Jun 06 09:38:04 2015	Anonymous
150606093441.473	Autovalidation	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Sat Jun 06 09:37:57 2015	Anonymous
150606093440.471P	MS/MS Search	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Sat Jun 06 09:37:55 2015	Anonymous
150606093440.472	MS/MS Search	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Sat Jun 06 09:36:47 2015	Anonymous
150606093439.470	Extraction	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Sat Jun 06 09:36:43 2015	Anonymous
150605194531.314	P/P Summary	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Fri Jun 05 19:49:00 2015	Anonymous
150605194530.313	Autovalidation	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Fri Jun 05 19:48:50 2015	Anonymous
150605194529.312	Autovalidation	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Fri Jun 05 19:48:43 2015	Anonymous
150605194528.310P	MS/MS Search	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Fri Jun 05 19:48:40 2015	Anonymous
150605194528.311	MS/MS Search	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Fri Jun 05 19:47:32 2015	Anonymous
150605194527.309	Extraction	ExampleData/Agilent/Q-TOF/Yeast/Y1	NoError	Fri Jun 05 19:47:27 2015	Anonymous

Figure 37 Completion Log

Exercise 5. Run interactive automation

You can queue portions of a workflow one step at a time from each task's page. When you queue tasks in workflow order, this is called interactive automation. Extraction and Search are queued when started, but Autovalidation and Protein/Peptide Summary must be queued by marking their respective Queue request check boxes. In fact, since you cannot start Autovalidation until Search is complete, queuing Autovalidation and Protein/Peptide Summary as Extraction or Search is running means those tasks automatically execute when appropriate.

Exercise 5. Run interactive automation

Steps	Detailed Instructions	Comments
1 Start the Data Extractor with the parameters in YourName\ extract-y1.	**a** Navigate to the Data Extractor page. **b** If it is not already loaded, load the parameter file **YourName\extract-y1**. **c** Check that your **Data Directory** is set to **y1**. **d** Click **Extract**.	• When you click **Extract**, the task is automatically queued.
2 Display the Request Queue Viewer so you can see that the Data Extractor is in the queue.	**a** In the Results frame, click the link to **View Request Queue**. **b** Verify that in the **Task Type** column, you see **xtractorAgilent**.	
3 Start the MS/MS Search on the same data with the parameters in YourName\MSMS_Search.	**a** Navigate to the MS/MS Search page. **b** If it is not already loaded, load the parameter file **YourName\ MSMS_Search**. **c** Click **Start Search**.	• You do not need to wait for data extraction to finish. • When you click **Start Search**, the task is automatically queued.
4 (Optional) Display the Request Queue Viewer so you can see that MS/MS Search is in the queue.	• If the Request Queue Viewer window is already displayed, right-click in the window and click **Refresh**. • Otherwise, follow the instructions as in step 2, but look for a **Task Type** of **mstag**.	

Exercise 5. Run interactive automation (continued)

Steps	Detailed Instructions	Comments
5 Queue the request to validate in the Auto strategy, Peptide mode. • Use parameters in AV-auto-pep file.	**a** Navigate to the Autovalidation page. **b** Load the parameter file **YourName\AV-auto-pep**. **c** Mark the check box for **Queue request**. **d** Click **Validate Files**.	• To have autovalidation run after search is complete, queue the autovalidation and PP summary tasks. • You should not start autovalidation while the search is still running.
6 Queue the request to validate in the Auto strategy, Protein polishing mode. • Use parameters in AV-auto-protpol file.	**a** Load the parameter file **YourName\AV-auto-protpol.** **b** Make sure the check box for **Queue request** is still marked. **c** Click **Validate Files**.	
7 Queue the request to generate a results summary. • Use parameters in prot-sum file.	**a** Navigate to the Protein/Peptide Summary page. **b** Load the parameter file **YourName\prot-sum.** **c** Mark the check box for **Queue request**. **d** Click **Summarize**.	
8 View the summary result	**a** Navigate to the Completion Log Viewer. (If you already have the Request Queue Viewer displayed, click **Completion Log**.) **b** Note when the Protein/Peptide Summary is complete. (You must refresh the window to see the updated progress.) See Figure 37 on page 90 as an example. **c** Click the **Task ID** number for that task. **d** View the results.	• If you click the **Task ID** number before the task is complete (for example, from the Request Queue Viewer), you will see a "page not found" error. • The protein summary report should be identical to the one produced in "Produce a Protein Summary Details Report" on page 29. (Figure 11 on page 32)
9 Find the summary results in the data file folder.	**a** On the server navigate to the folder **msdataSM\ExampleData\Agilent\Q-TOF\yeast\y1**. **b** Find the summary file. Look for a file that starts with **PPSummary** and ends with **htm**. **c** Double-click the file to display it.	• If you mark the **Queue request** check box, the program stores the html file that contains the summary results within the data file folder.

5
Interactive Processing MS-Only Data

This chapter provides familiarization exercises to process example MS-only Q-TOF data files with the Agilent Spectrum Mill MS Proteomics Workbench.

You use an example data file, peptide-ms-only.d, found in the example data folder with the MassHunter Qualitative Analysis software. Instructions in this chapter show you how to obtain and copy the appropriate masses to paste into the Spectrum Mill Manual PMF Search page in order to search Q-TOF MS-only data with Spectrum Mill workbench.

You can use the procedure described in these exercises for both Agilent TOF and Q-TOF .d files.

This familiarization chapter is intended as an introduction to the basic operation of the software; you are encouraged to customize these steps for your samples and studies.

Exercise 1. Find the compounds of interest

You do not have to extract the data you are to use in these exercises. Rather, you obtain the appropriate masses for a manual PMF search by selecting them from an MS-only Q-TOF data file through the MassHunter Qualitative Analysis software.

Exercise 1. Find the compounds of interest

Steps	Detailed Instructions	Comments
1 Open the peptide-ms-only.d data file in MassHunter Qualitative Analysis.	**a** Open MassHunter Qualitative Analysis. **b** Under the **Data** folder, select **peptide-ms-only.d**, and click **Open**.	• This file contains data on the peptides detected in a tryptic digest of bovine serum albumin (BSA).
2 Open Method Explorer, if necessary.	**a** From the top menu bar, click **Method**. **b** Select **Method Explorer**.	
3 Find compounds with these molecular features: • Extraction - Select small molecules (chromatographic) and use peaks with height >100 counts • Ion Species - +H and -H • Charge State - 7 for peptides • Compound Filters - only use Absolute Height - 1500 • Results - Clear check boxes for Chromatograms and spectra, and for Display limits, if necessary.	**a** In Method Explorer, click **Find by Molecular Feature**. **b** Under the **Extraction** tab and from the **Target data type** list, select **small molecules (chromatographic)**. **c** Click **Use peaks with height >=,** type `100`, if necessary. **d** Click I**on Species** and clear all check boxes except for **+H** and **-H**. **e** Click **Charge State** and from the **Isotope Model** list, select **Peptides**. **f** Mark **Limit assigned charge states to a maximum of** and type `7`. **g** Click **Compound Filters**, and clear **Relative height**; for **Absolute height**, type 1500. **h** Click **Results** and clear all check boxes in the tab except for the first one. **i** Click **Find Compounds by Molecular Feature**. **j** Close the dialog box.	• After changing these values and starting the subroutine, you should see the MassHunter main window look like Figure 38 on page 95. • If you do not see the complete table, click **Automatically show columns**.

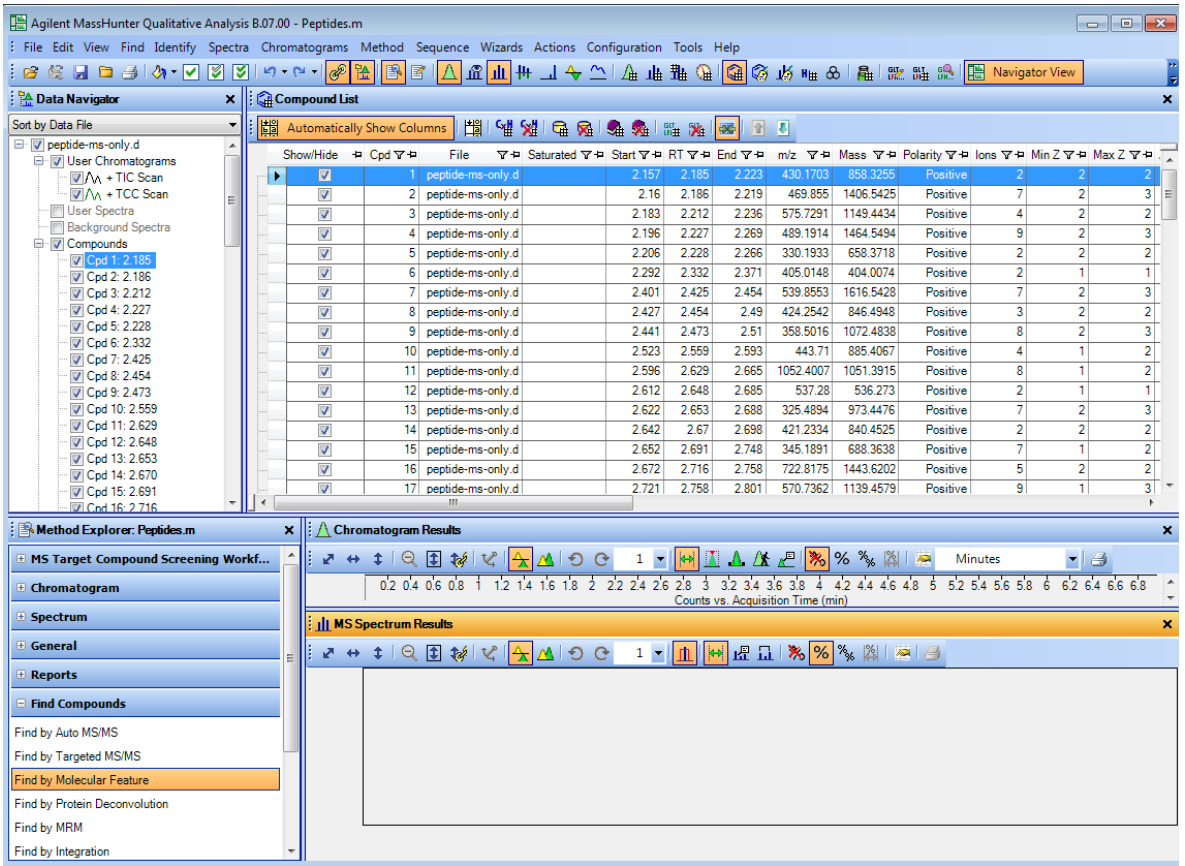


Figure 38 MassHunter compound list

Exercise 2. Copy and paste compound masses into PMF Search

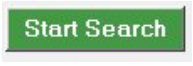
In this exercise you copy only the masses from all the compound features and then paste them into the Manual PMF Search page.

Steps	Detailed Instructions	Comments
1 Copy the neutral masses from the list, and close the MassHunter Qualitative Analysis main window.	**a** Right-click the **Mass** heading, and select **Copy Column to Clipboard > Using Newline separator**. **b** Close the window.	• Note that the list should contain 91 compounds.
2 Navigate to the Manual PMF Search page.	**a** Open **Spectrum Mill Workbench**. **b** From the Spectrum Mill home page, click the link to **Manual PMF Search**. . 	
3 Paste the masses into the Manual PMF Search page.	**a** Click the **Manual PMF** button at the top of the PMF Search page. **b** Select all the masses in the **Peptide Masses List,** and press **Delete**. **c** Press **Ctrl V**. You should now see the new masses in the list. See Figure 39 on page 98.	

Exercise 3. Search the Q-TOF MS-only data and display results

After you paste the masses into the Manual PMF Search page, you set up the parameters to prepare for and then search the data.

Exercise 3. Search the Q-TOF MS-only data and display results

Steps	Detailed Instructions	Comments
4 Make sure the page looks like Figure 39 on page 98.	**a** To identify the masses as neutral masses, click the **M** radio button. **b** From the **Database** list, select **NCBInr.stdmix**. **c** Click **Choose**, and select **Carboxymethylation** as the Fixed Modification. **d** If oxidized methionine and pyroglutamic acid are not listed as variable modifications on the Manual PMF Search page, choose them. **e** Click **OK**.	• The NCBInr.stdmix database is installed with Spectrum Mill but must be indexed before using it here. • See the *Installation Guide* for indexing instructions.
5 Start the search.	• Click **Start Search**  The PMF Search Results display appears in a separate window. See Figure 40 on page 99.	• Manual PMF Search processes all the peaks in the list. • Search time varies depending on the size of the database searched.

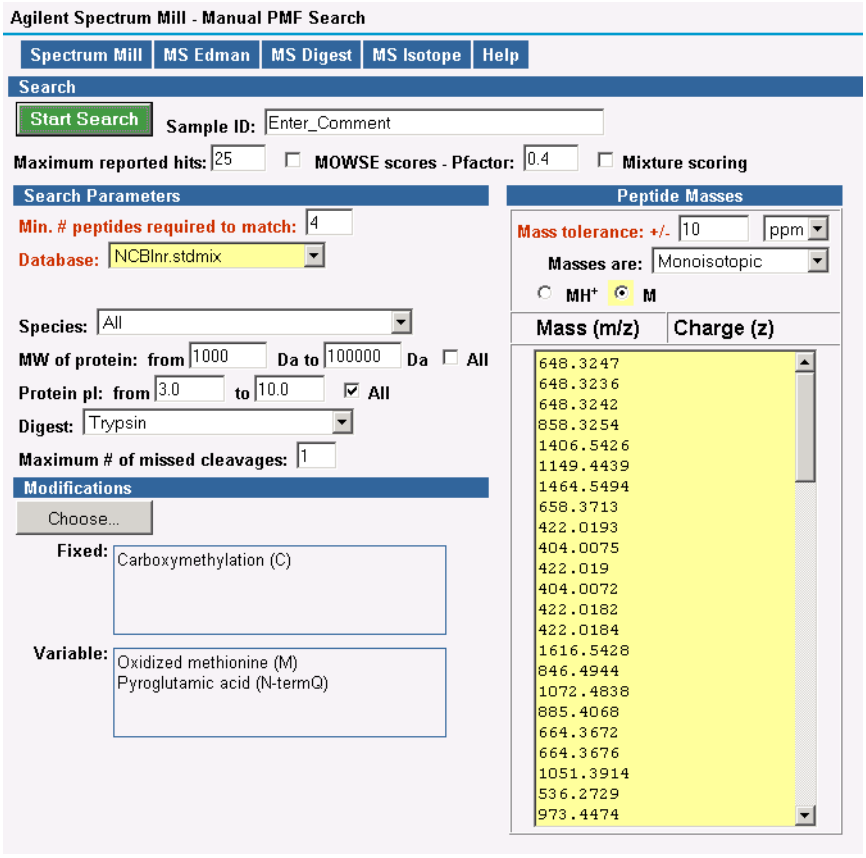


Figure 39 PMF Search settings for Q-TOF MS-only data

PMF Search Results

Sample ID (comment): **Enter_Comment**
Database searched: **NCBInr.stdmix** Parameters used in Search
Molecular weight search (**1000** - **100000 Da**) selects **17** entries.
Full pI range: **18** entries.
Combined molecular weight and pI searches select **17** entries.
PMF search selects **1** entry.

Rank	Dynamic Probability Score	Static Probability Score	# (%) Masses Matched	Mass Error Mean (Std Dev) (ppm)	Protein Coverage	Protein MW (Da)/pI	Species	Accession #	Protein Name
1	0	0	50/90 (55%)	-1.0 (2.8)	59%	71323.5/5.82	bovine	1351907	bovine serum albumin

Detailed Results

1. 50/90 matches (55%). bovine. bovine serum albumin (71323.5 Da) pI = 5.82

m/z submitted	MH^+ matched	Delta ppm	Score Counted	Tolerance Bin Index (0.0)	start	end	Peptide Sequence	Modifications
537.2802	537.2820	-3.4	1	**0**	157	160	**(K)FWGK(Y)**	
649.3315	649.3338	-3.5	1	**0**	205	209	**(K)IETMR(E)**	
649.3309	649.3338	-4.5	1	**0**	205	209	**(K)IETMR(E)**	
649.3320	649.3338	-2.8	1	**0**	205	209	**(K)IETMR(E)**	
665.3745	665.3770	-3.7	1	**0**	156	160	**(K)KFWGK(Y)**	
665.3749	665.3770	-3.1	1	**0**	156	160	**(K)KFWGK(Y)**	
689.3712	689.3729	-2.6	1	**0**	236	241	**(K)AWSVAR(L)**	
759.4058	759.4069	-1.5	1	**0**	198	204	**(K)GACLLPK(I)**	
759.4052	759.4069	-2.3	1	**0**	198	204	**(K)GACLLPK(I)**	
759.4048	759.4069	-2.8	1	**0**	198	204	**(K)GACLLPK(I)**	
789.4697	789.4716	-2.5	1	**0**	257	263	**(K)LVTDLTK(V)**	
789.4719	789.4716	0.3	1	1	257	263	(K)LVTDLTK(V)	
847.5017	847.5036	-2.3	1	**0**	242	248	**(R)LSQKFPK(A)**	
886.4141	886.4153	-1.3	1	**0**	131	138	**(K)DDSPDLPK(L)**	
899.4629	899.4655	-2.9	1	**0**	483	489	**(R)LCVLHEK(T)**	
922.4870	922.4880	-1.1	1	**0**	249	256	**(K)AEFVEVTK(L)**	
927.4934	927.4934	-0.1	1	**0**	161	167	**(K)YLYEIAR(R)**	
974.4547	974.4578	-3.2	1	**0**	37	44	**(K)DLGEEHFK(G)**	
997.5963	997.5928	3.5	1	1	549	557	(K)QTALVELLK(H)	**pyroGlu**
1002.5826	1002.5830	-0.4	1	**0**	598	607	**(K)LVVSTQTALA(-)**	

Figure 40 Manual PMF Search results with MS-only Q-TOF data

Exercise 4. Review Q-TOF MS-only PMF Search results

After you display PMF Search results, you manually review them. In this exercise, you examine the PMF Search results to determine how well they fit the peak list that was submitted.

Exercise 4. Review Q-TOF MS-only PMF Search results

Steps	Detailed Instructions	Comments
1 View the potential match(es).	• See potential match(es), as shown in Figure 40 on page 99. Again, your results may vary.	• The detailed results shown in Figure 40 list only one match for the data.
2 View a coverage map. • Coverage should be 59%. • See Figure 41 on page 101.	**a** Click the percentage link under the **Protein Coverage** header to view amino acid coverage. **b** Close the coverage window. Or just look at the bottom of the Detailed Results section to see the amino acid coverage.	In the coverage map: • Red = matched • Black = not matched • Blue = modified amino acid, e.g., cysteine • Green = consensus N-linked glycosylation site
3 View information in the database.	**a** Click a link under the **Accession #** header to see the database information. **b** Close the accession number window.	
4 (Optional) Generate a printout of the results.	**a** Select **File > Page Setup...**, set the page to **Landscape** mode and mark the check box for **Print background colors and images**. Click **OK**. **b** Click in the frame you wish to print. **c** Select **File > Print Preview...**. **d** Click the **Print icon**.	

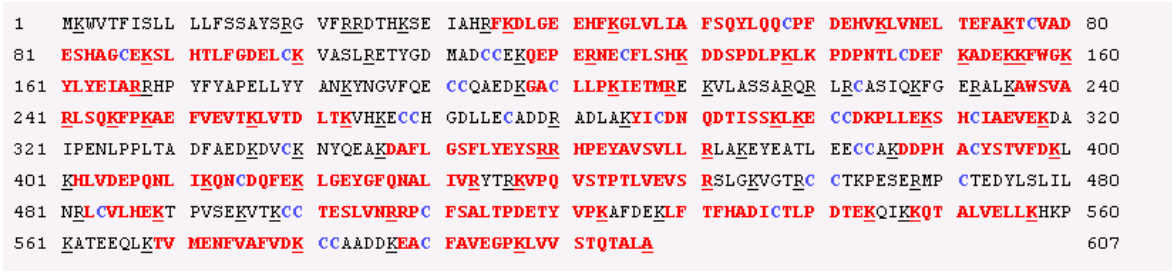


Figure 41 Coverage map for protein-MS-only.d

www.agilent.com

In This Book

The *Familiarization Guide* presents exercises to process example data with the Spectrum Mill MS Proteomics Workbench. In this guide you learn how to:

- Run protein database searches
- Review and validate search results
- Use *de novo* sequencing
- Work with variable and unknown modifications
- Automate workflows

Printed in USA
Eighth edition, October 2016

G2721-90052

Agilent Technologies