Agilent G2721AA/G2733AA Spectrum Mill MS Proteomics Workbench

Familiarization Guide

Agilent Technologies

Notices

Manual Part Number

G2721-90040

Edition

Revision A, January 2012

Printed in USA

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

Software Revision

This guide is valid for the B.04.00 revision or higher of the Agilent G2721AA/G2733AA Spectrum Mill MS Proteomics Workbench software, until superseded.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 and MS-only 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 Basic 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.

Contents

1 Basic Workflow – Interactive Processing of MS/MS Data

This chapter provides familiarization exercises to process an example set of MS/MS data files with a basic 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 basic 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 **NCBInr** species databases (yeast and *ecoli*) have 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 started, as well as 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 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\transfer file folder\New Examples**). See the *Quick Start Guide* for instructions on creating the transfer folder.

1. From your PC client, open the **msdataSM** share to the server.
2. In your **transfer file folder**, create a folder and name it **New_Examples**.
3. Copy and paste the **\ExampleData\Agilent\Q-TOF\yeast\y1** and **y2** folders into **\transfer file folder\New_Examples.**
4. Copy and paste the **\ExampleData\Agilent\Q-TOF\OGE** folder into **\transfer file folder\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.

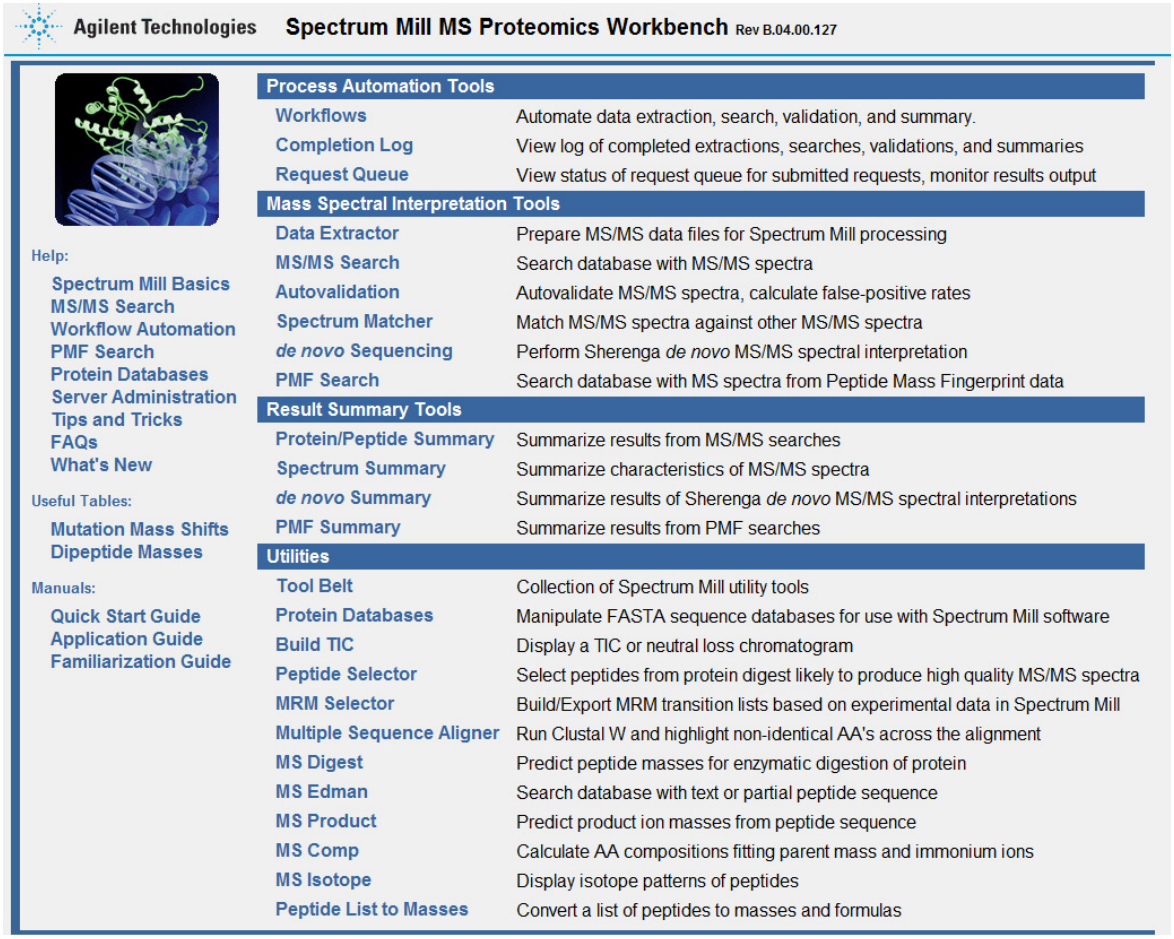


Figure 1 Spectrum Mill home page

Exercise 1. Extract the spectra

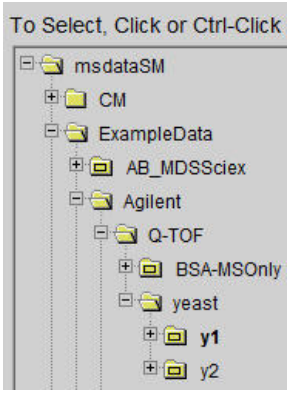
The Spectrum Mill Data Extractor preprocesses raw data files to extract high-quality spectra for database searches. The Data Extractor automatically detects which type of raw file you have and then invokes the appropriate extraction program (provided that it has been purchased and installed on your server).

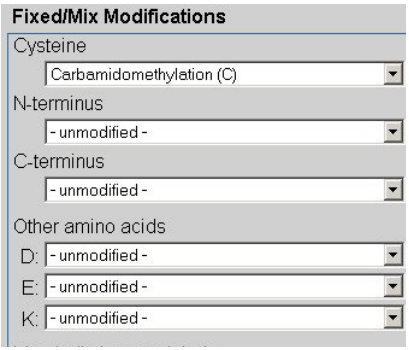
The MS/MS raw file data extractors extract and merge nearby MS/MS spectra from the same precursor ion. They optionally apply MS/MS similarity criteria prior to merging scans, to avoid merging closely eluting or co-eluting isobaric peptides. For Agilent ion trap and Thermo Fisher Scientific *.raw ion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

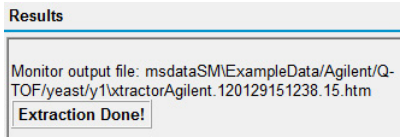
The extractors do these tasks:

- 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.

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

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.

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** Under the **Cysteine** heading, select **Carbamidomethylation** .  **c** 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** Examine the items in red text carefully, since these are the ones you may need to change when you process your own samples. **c** 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.
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, say, "your own name". • 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.

Steps	Detailed Instructions	Comments
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 directory. • 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.
7 (Optional) View the **cpick_in** subdirectory to see the files Data Extractor has created.	• Navigate to the folder **ExampleData\Agilent\Q-TOF\yeast\y1\cpick_in** on the **msdataSM** share to the server. Notice the new files created there. Each one represents an extracted spectrum from your raw data file. See the *Installation*Guide for instructions to set up the **msdataSM** share on the server and the *Quick Start Guide* for instructions to access the server msdataSM share from the PC client.	File names are in the format: Data_File_Name.aaaa.bbbb.c.pkl, where • aaaa = first merged scan • bbbb = last merged scan • c = assigned precursor charge (0 means charge was ambiguous) • Because the resolution for Agilent Q-TOF data is sufficient to resolve isotopes for most peptide precursors, the great majority of extracted spectra have assigned charges.

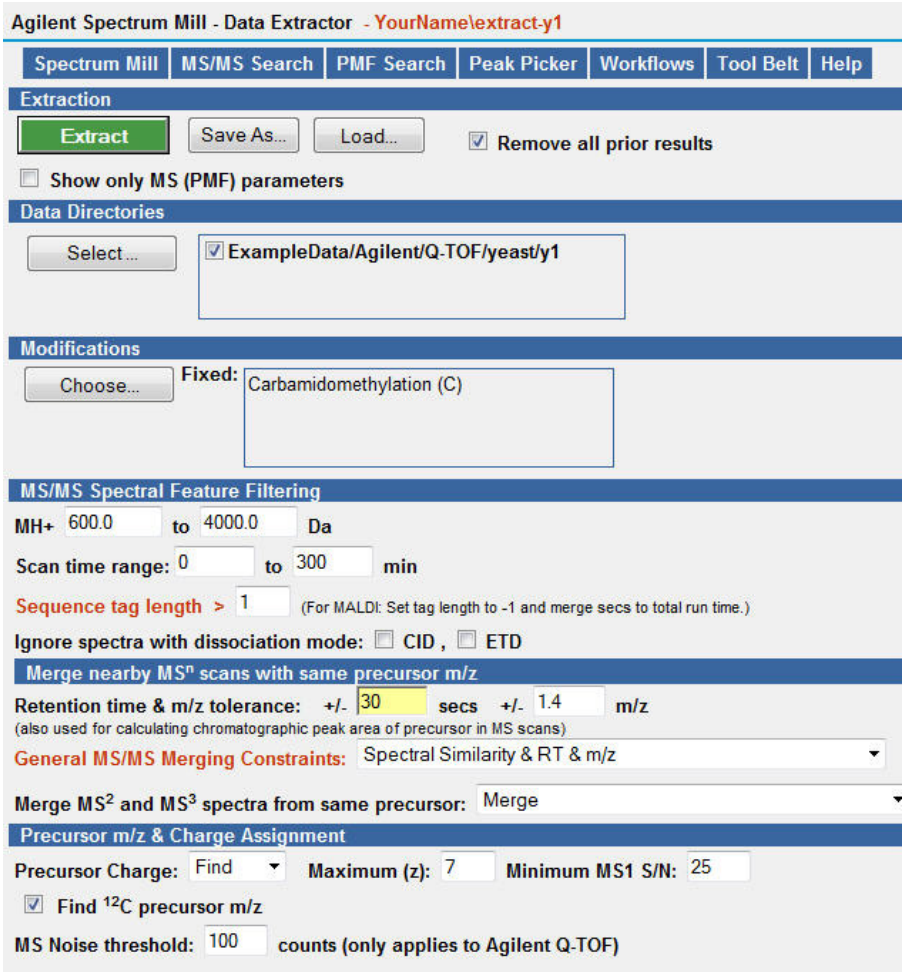


Figure 2 Settings for Data Extractor

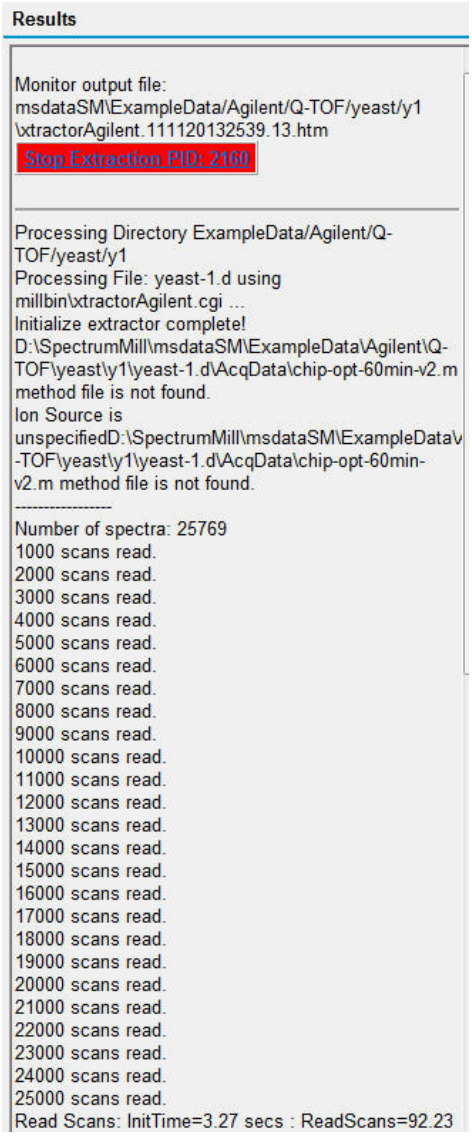


Figure 3 Data Extractor running

Exercise 2. Search the database

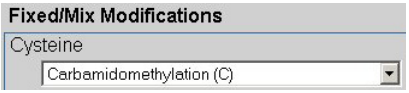
After you have extracted your spectra, you are ready to search each spectrum against a protein or DNA database. In the basic workflow, you do a single round of database search and results validation, with the goal of identifying most of the identifiable spectra. For this search, you use Variable Modifications mode, which means spectra may show modifications relative to the database sequence.

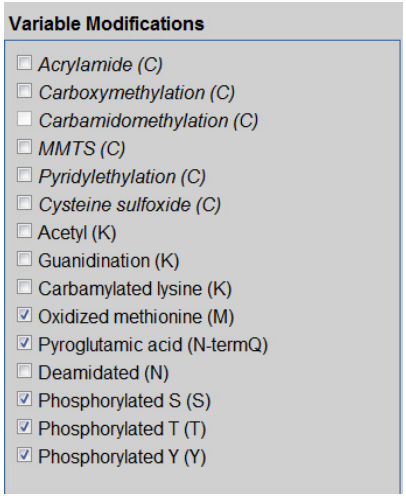
For single round searching, to speed up searching you can search fewer modifications or search a smaller database. Those are the only options. Iterative searching offers more options to speed up searching. Combining multiple searches lets you do a search with few modifications against a large database, and then do a search with more modifications against a smaller database, and combine the results.

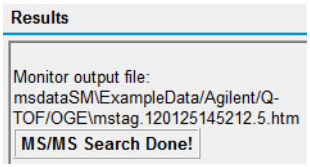
Steps	Detailed Instructions	Comments

1 Navigate to the MS/MS Search page.	• Do one of the following:
	  ◦ From the DataExtractor 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.	• Mark the **Remove all prior MS/MS Search results** 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.

Steps	Detailed Instructions	Comments
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.	
4 Select the NCBInr.yeast subset database.	• For **Database** (under **Search Parameters**), select the NCBInr.yeast subset 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.
5 Set up the Fixed and Variable Modifications; • If the carbamidomethlation 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 page 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.

Steps	Detailed Instructions	Comments
•	 **d** Click **OK.** The names of the modifications appear as in Figure 4 on page 20.	• 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**.
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** 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 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 for Agilent Q-TOF data.

Steps	Detailed Instructions	Comments
7 Set other parameters as shown in Figure 4 on page 20.	**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.	• Changes from defaults are highlighted in yellow. • 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.
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. **c** Scroll to the *top* of the **Results** area to see the message that indicates that the search is finished. 	• The command is automatically placed in the queue for execution. • 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.

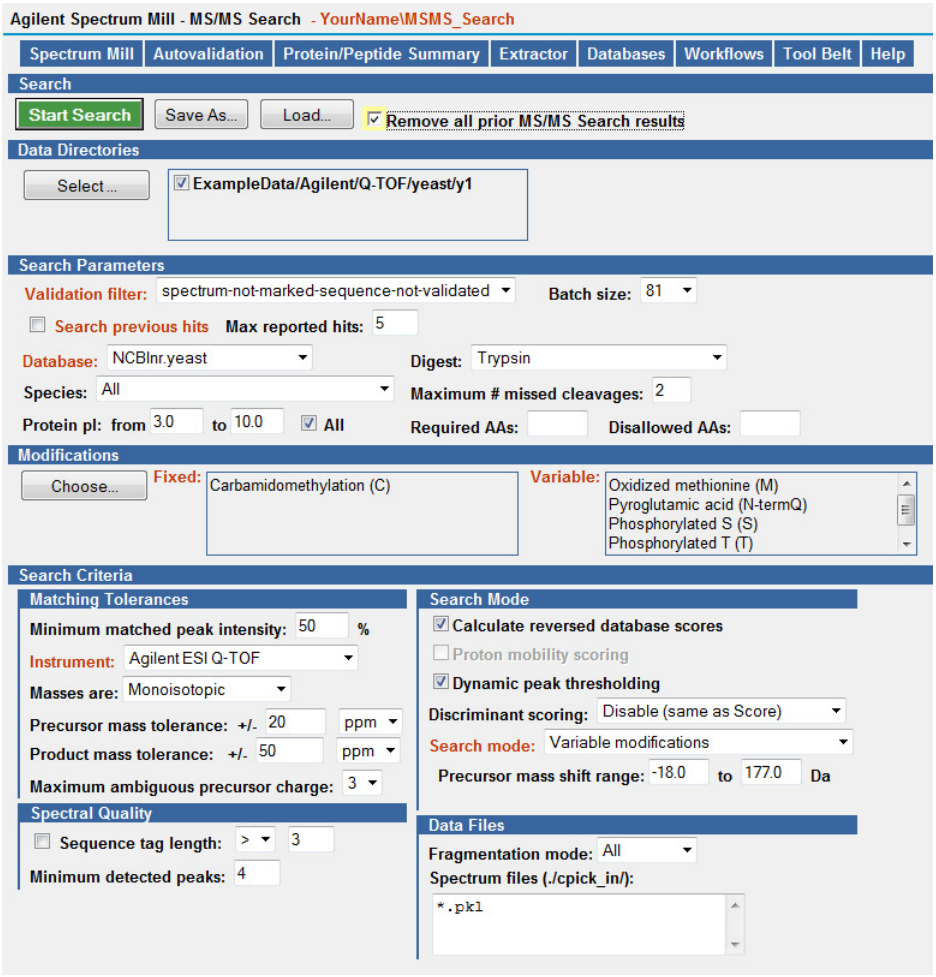
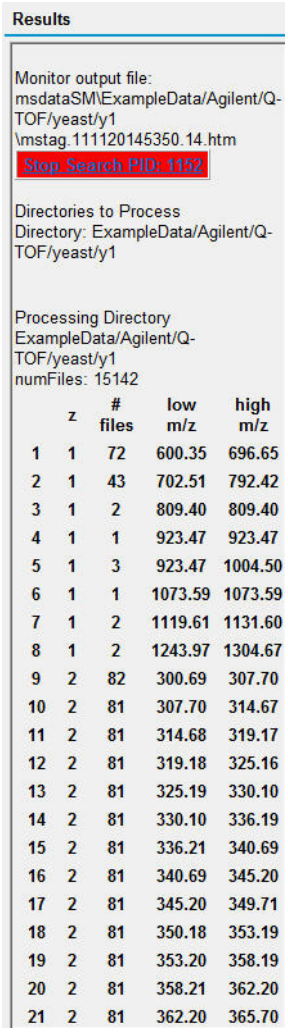


Figure 4 MS/MS Search settings



Results

Monitor output file:
msdataSM\ExampleData/Agilent/Q-TOF/yeast/y1
\mstag.111120145350.14.htm
Stop Search PID: 1152

Directories to Process
Directory: ExampleData/Agilent/Q-TOF/yeast/y1

Processing Directory
ExampleData/Agilent/Q-TOF/yeast/y1
numFiles: 15142

	z	# files	low m/z	high m/z
1	1	72	600.35	696.65
2	1	43	702.51	792.42
3	1	2	809.40	809.40
4	1	1	923.47	923.47
5	1	3	923.47	1004.50
6	1	1	1073.59	1073.59
7	1	2	1119.61	1131.60
8	1	2	1243.97	1304.67
9	2	82	300.69	307.70
10	2	81	307.70	314.67
11	2	81	314.68	319.17
12	2	81	319.18	325.16
13	2	81	325.19	330.10
14	2	81	330.10	336.19
15	2	81	336.21	340.69
16	2	81	340.69	345.20
17	2	81	345.20	349.71
18	2	81	350.18	353.19
19	2	81	353.20	358.19
20	2	81	358.21	362.20
21	2	81	362.20	365.70

Figure 5 MS/MS Search running

Exercise 3. Autovalidate hiqh-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.

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.	
3 Set up autovalidation first with the Auto thresholds strategy, Peptide mode. ◦ Clear all previous validations. ◦ Keep defaults. See Figure 6 on page 24.	**a** For **Strategy**, click **Auto thresholds**. **b** For **Mode**, keep the default of **Peptide**. **c** Click **Clear All**.	• This strategy/mode optimizes the score and delta R1-R2 score thresholds up to a target False Discovery Rate (FDR) that you enter.
4 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.

Steps	Detailed Instructions	Comments
5 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 (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.
6 Set up autovalidation second with the Auto thresholds strategy-Protein Polishing mode. • Use the settings shown in Figure 8 on page 26	**a** For **Mode**, select **Protein Polishing**. **b** Change **Minimum protein score** to 0. **c** Change **...maximum protein FDR** to 1.0%	
7 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**.	
8 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.
9 Close the Autovalidation window.	• Close the window only after you see the word “finished”.	

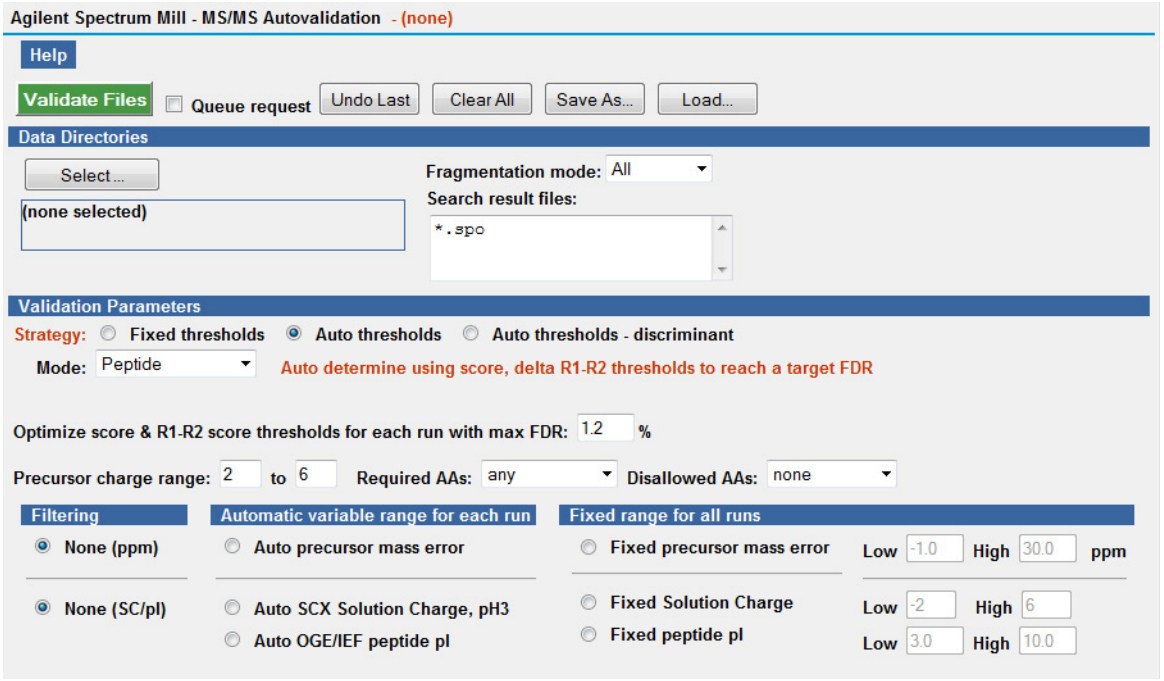


Figure 6 Autovalidation page, Auto thresholds strategy, Peptide mode

yeast-1 ruleNumber: 5 z: 6

Run	Rule #	z	score	R1R2	FDR(%)	# hits	#hits False
yeast-1	1	2	6.1	1.1	1.17	2730	16
Subtotal	all	all			1.17	2730	16
yeast-1	2	3	5.1	3.1	1.10	1457	8
Subtotal	all	all			1.10	1457	8
yeast-1	3	4	3.8	1.0	1.14	176	1
Subtotal	all	all			1.14	176	1
yeast-1	4	5	6.5	0.2	0.00	17	0
Subtotal	all	all			0.00	17	0
yeast-1	5	6	0.0	0.0	0.00	0	0
Subtotal	all	all			0.00	0	0
Total	all	all			1.14	4380	25

Rule: 1 Number of Files to validate: 2730 Num False positives Fwd-Rev score < 0: 16 FP rate (%): 1.17
Rule: 2 Number of Files to validate: 1457 Num False positives Fwd-Rev score < 0: 8 FP rate (%): 1.10
Rule: 3 Number of Files to validate: 176 Num False positives Fwd-Rev score < 0: 1 FP rate (%): 1.14
Rule: 4 Number of Files to validate: 17 Num False positives Fwd-Rev score < 0: 0 FP rate (%): 0.00
Rule: 5 Number of Files to validate: 1 Num False positives Fwd-Rev score < 0: 0 FP rate (%): 0.00

Number of validated spectra : 4381 Number of False positives Fwd-Rev score < 0 : 25 FP rate (%): 1.14

ExampleData/Agilent/Q-TOF/yeast/y1 Number of validated spectra : 4381

Preparing new hit and spectra validation info.... finished
Resetting previously validated hits
Validated Hits: 4382
Validated Spectra: 4382

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

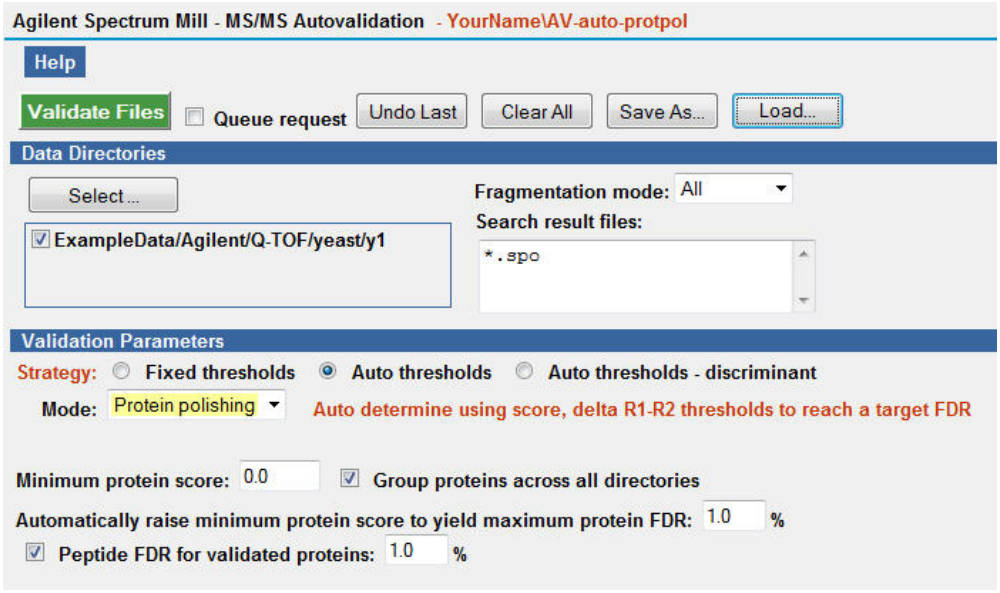


Figure 8 Autovalidation page, Auto thresholds strategy, Protein Polishing mode

```
Total Number of Files to validate (some double counting from files in > 1 protein group): 4241

Validating files in msdataSM\ExampleData/Agilent/Q-TOF/yeast/y1 ...
Number of Files to validate (some double counting from files in > 1 protein group): 4241

Preparing new hit and spectra validation info.... finished
Resetting previously validated hits
Validated Hits: 4242
Validated Spectra: 4242
```

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 report and a Peptide report to review the validated results.

Produce a Protein Summary 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.	• From the **Mode** list, select **Protein Summary**.	• 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 3 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).

Steps	Detailed Instructions	Comments
5 Sort the proteins by Score (default setting).	**a** Leave the **Sort Proteins by** list selection as **Score**.	
6 Filter peptides by changing 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.
7 Set the **Review Fields**: • Protein MW	**a** Under **Review Fields**, mark the **Protein** MW check box. Note that changes from defaults are highlighted in yellow.	• **Review Fields** determine what information you see in the final results summary.
8 Keep the settings for other fields as in Figure 10 on page 28.		• 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 pep-sum.	**a** Click **Save As**. **b** For **Folder**, select **YourName**. **c** For **Name**, type **pep-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.

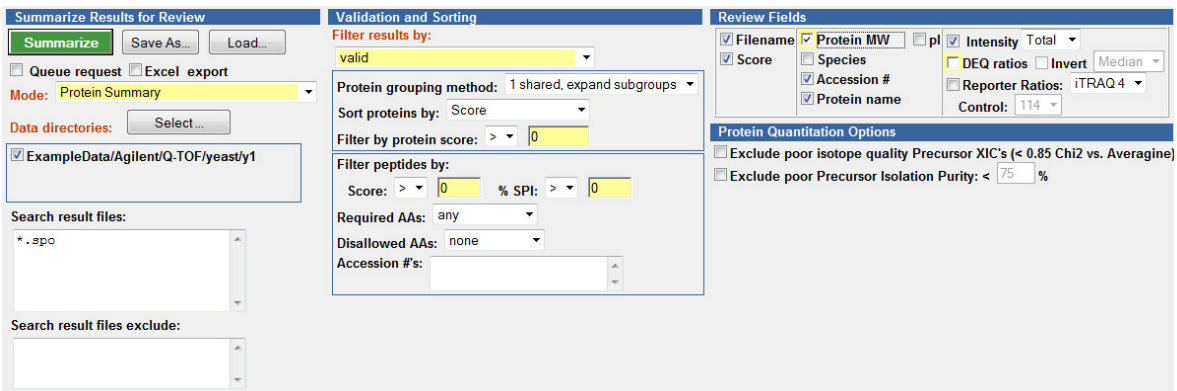


Figure 10 Protein Summary mode settings

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 have finished executing.
11 Examine the summary report. • Make sure results match those in Figure 11 on page 29. • Learn what headers mean.	**a** Check that your results are similar to those in Figure 11 on page 29. Since the public databases are constantly updated, your results may vary slightly. **b** Become familiar with the headers in Table 1 on page 30.	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	61	30	478.66	87.6	3.49e+008	35845.2	3720	reading frame
1	1.2	39	24	343.65	67.7	2.62e+008	35960.3	6322468	glyceraldehyde 3-phosphate dehydrogenase; Tdh2p
1	1.3	38	22	310.32	68.3	2.46e+008	35863.4	6322409	Glyceraldehyde-3-phosphate dehydrogenase 1; Tdh1p
2	2.1	58	33	473.79	80.5	3.75e+008	44794.6	10383781	3-phosphoglycerate kinase; Pgk1p
3	3.1	50	36	471.17	66.8	1.09e+008	69827.1	417149	HEAT SHOCK PROTEIN SSA1 (HEAT SHOCK PROTEIN YG100)
3	3.2	39	29	362.14	49.9	9.12e+007	69639.9	6323004	member of 70 kDa heat shock protein family; Ssa2p
3	3.3	16	11	146.12	17.8	3.03e+007	70659.6	6319396	heat-inducible cytosolic member of the 70 kDa heat shock protein family; Ssa3p
3	3.4	8	8	84.53	14.9	3.21e+006	74523.7	6322426	Involved in translocation of nascent polypeptides across the ER membrane; Kar2p
4	4.1	53	27	423.29	71.3	2.76e+008	46970.4	6321968	enolase; Eno2p
4	4.2	49	27	415.87	71.8	3.25e+008	46872.3	6321693	enolase I; Eno1p
5	5.1	38	30	399.62	68.8	8.84e+007	54943.0	6319279	Required for START A in the cell cycle and sporulation; Cdc19p
6	6.1	50	27	379.42	59.8	7.05e+007	61722.5	6323073	pyruvate decarboxylase; Pdc1p
6	6.2	14	10	120.52	20.4	2.43e+007	62139.1	6323163	pyruvate decarboxylase; Pdc5p
6	6.3	9	6	83.70	14	1.50e+007	61636.3	6321524	Third, minor isozyme of pyruvate decarboxylase; Pdc6p
7	7.1	31	27	360.47	44.1	2.25e+007	86029.5	6320936	cobalamin-independent methionine synthase; Met6p
8	8.1	39	26	356.58	63.3	3.71e+007	57007.9	6324950	Glucose repressed. Utilizes NADP+ or NAD+ as a coenzyme equally well. (sold by SIGMA under the ca
9	9.1	35	26	343.66	52.6	4.69e+007	66707.5	6324120	stress-seventy subfamily B; Ssb2p
9	9.2	35	26	338.02	51.3	4.73e+007	66771.5	6319972	Stress-Seventy Subfamily B; involved in translation, perhaps by guiding the nascent chain through the ri
10	10.1	33	27	343.27	43.2	3.43e+007	93743.9	6320593	translation elongation factor 2 (EF-2); Eft2p
11	11.1	28	25	321.90	17.7	9.44e+006	229542.8	6322666	pentafunctional enzyme consisting of the following domains : acetyl transferase, enoyl reductase, dehyc
12	12.1	29	24	317.90	32	2.84e+007	80898.2	6323840	constitutively expressed heat shock protein; Hsc82p
12	12.2	26	22	287.59	29	2.78e+007	81404.9	6325016	82 kDa heat shock protein; homolog of mammalian Hsp90; Hsp82p

Figure 11 Protein Summary results

Table 1 Protein Summary 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 whose sequences are different
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
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 Report

Steps	Detailed Instructions	Comments
1 Set the Mode to Peptide.	• From the **Mode** list, select **Peptide**.	•
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.

Steps	Detailed Instructions	Comments
4 Filter peptides by changing 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 mark these fields: • b/y map. • Delta mass • Var mod sites • VML score (s/t/y) • Var mod sequence • Modification names	**a** Under **Review Fields**, clear the **Sequence** check box. **b** Mark the **b/y map** check box. **c** Mark the **Delta mass, Var mod sites, Var mod score, Var mod sequence** and **Modification names** check boxes. Note that changes from defaults are highlighted in yellow.	• The modification sites, scores and sequences appear in the report.
6 Keep the settings for other fields in Figure 12 on page 31.		
7 Save the settings as a parameter file. • Save to the YourName folder. • Name the file pep-sum.	**a** Click **Save As**. **b** For **Folder**, select **YourName**. **c** For **Name**, type **pep-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.

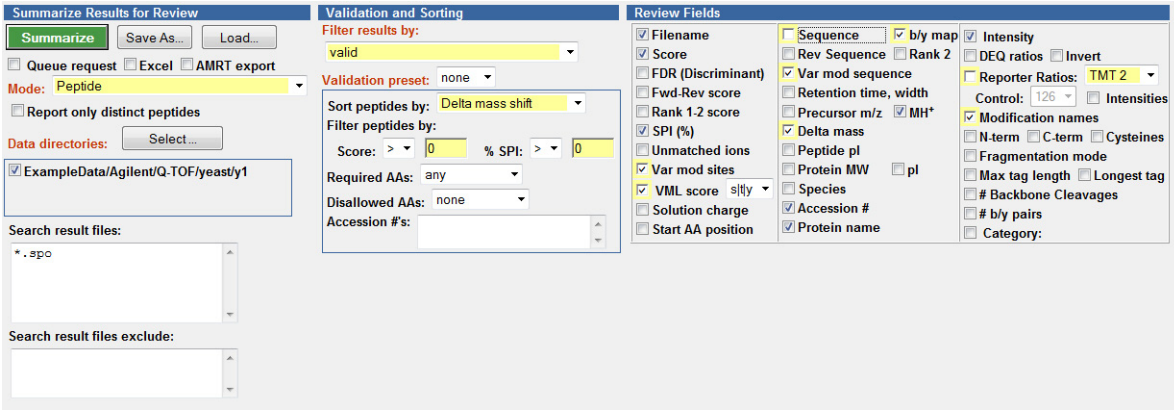


Figure 12 Peptide mode settings

Steps	Detailed Instructions	Comments
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 33. • Learn what headers mean.	**a** Check that your results are similar to those in Figure 13 on page 33. Since the public databases are constantly updated, your results may vary slightly. **b** Become familiar with the headers in Table 2 on page 33.	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)
10 To examine these results in more detail, go to "Exercise 2. Examine results with the Spectrum Viewer" on page 35. • 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.

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)
0.64	3	2	1	1	T201t Y207y	(K)I V|H D Q V t/G M S K G y V L V K(F) IVHDQVT(0.99)GMS(0.50)KGY(0.50)VLVK	t:Phosphorylated T y:Phosphorylated Y	1874.020	159.9678	17.3
0.89	5	2	1	1	Y29y T48t	(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 t:Phosphorylated T	3157.589	159.9212	-3.4
0.00	4	1	0	1	M103m S112s	(K)C L/V/T m L L G V A Y E F s S K(E) CLVT(0.0)M(1.0)LLGVAY(0.33)EFS(0.33)S(0.33)K	C:Carbamidomethylation m:Oxidized methionine s:Phosphorylated S	1817.918	95.9836	11.7
0.00	4	1	0	1	T88t M89m	(R)V/D|S V A/S/R V Q T A V t m R(Q) VDS(0.0)VAS(0.0)RVQT(0.50)AVT(0.50)M(1.0)R	t:Phosphorylated T m:Oxidized methionine	1619.853	95.9765	8.9
0.00	4	1	0	1	Y246y M253m	(K)L S|D|L E R I T K y T G A H G V m A V R(G) LS(0.0)DLERIT(0.33)KY(0.33)T(0.33)GAHGVM(1.0)AVR	y:Phosphorylated Y m:Oxidized methionine	2217.181	95.9652	1.7
0.38	9	1	0	1	S423s	(R)F E Y S K Y R P L T P S S E/V/Y S P I E K E s P W/K(I) FEY(0.0)S(0.0)KY(0.0)RPLT(0.0)PS(0.0)S(0.0)EVY(0.33)S(0.33)PIEKES(0.33)PWK	s:Phosphorylated S	3147.567	80.0201	16.6
3.83	2	1	1	0	S355s	(K)L I K N I A A G G V S M N E L D s R(I) LIKNIAAGGVS(0.0)MNELDS(0.99)R	s:Phosphorylated S	1887.996	80.0046	19.4
3.87	3	1	1	0	S235s	(K)V I\D V/D T L D I V F S L A V F G L V M N A L s A K R N V D/K(Y) VIDVDT(0.0)LDIVFS(0.0)LAVFGLVMNALS(0.99)AKRNVDK	s:Phosphorylated S	3362.839	79.9920	7.5
99	1	1	1	0	S100s	(R)V/I|L P E R s D/E/K(K) VILPERS(1.0)DEK	s:Phosphorylated S	1185.647	79.9899	18.7
3.96	2	1	1	0	Y230y	(R)A Q Q V P\P Y G|V/y H P P F I K(Y) AQQVPPY(0.0)GVY(0.99)HPPFIK	y:Phosphorylated Y	1840.975	79.9888	11.7
1.27	7	1	1	0	T1705t	(K)L S/S A S T A P Y V E/E Y t K L L A R C F L K(Q) LS(0.0)S(0.0)AS(0.0)T(0.0)APY(0.0)VEEY(0.0)T(0.99)KLLARCFLK	C:Carbamidomethylation t:Phosphorylated T	2647.380	79.9835	6.3
0.00	2	1	0	1	S35s	(K)I V|L/T G I s K(Q) IVLT(0.50)GIS(0.50)K	s:Phosphorylated S	830.535	79.9767	11.4
1.31	2	1	1	0	T410t	(K)D F H\A E\L/S t/P L L K/P V N K(G) DFHAELS(0.0)T(0.99)PLLKPVNK	t:Phosphorylated T	1808.991	79.9749	4.5
99	1	1	1	0	T49t	(K)N E|A/t/P E/A E Q V K(K) NEAT(1.0)PEAEQVK	t:Phosphorylated T	1215.585	79.9738	5.7
-0.10	4	1	0	1	T104t	(K)t E A P Q L R D E|Y/K T Y K(I) T(0.0)EAPQLRDEY(0.0)KT(0.0)Y(0.0)K	t:Phosphorylated T	1741.876	79.9734	3.9
0.00	5	1	0	1	T250t	(R)E T T/A L V/E S Y E L P D G R t I/K(V) ET(0.0)T(0.0)ALVES(0.33)Y(0.33)ELPDGRT(0.33)IK	t:Phosphorylated T	2022.039	79.9671	0.4

Figure 13 Peptide summary report

Table 2 Peptide Summary 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
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

Table 2 Peptide Summary Report Column Headers with Variable Modifications

Header	Definition
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 between the observed and theoretical m/z in ppm
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.

Steps	Detailed Instructions	Comments
1 If you have not already done so, display a Peptide report.	• Follow steps in "Produce a Peptide Report" on page 30.	
2 Display the peptide #14 spectrum to view an example of a single phosphorylation site.	**a** In the summary report, click the number **14** under the **#** header. **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 36. 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 39 to examine the phosphorylation sites in the sequence map.
3 Display the peptide #12 spectrum to view a spectrum with two possible phosphorylation sites.	**a** In the summary report, click the number **12** under the **#** header. **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 36.	• See "Exercise 3. Examine details of search results" on page 39 to examine the phosphorylation sites in the sequence map.

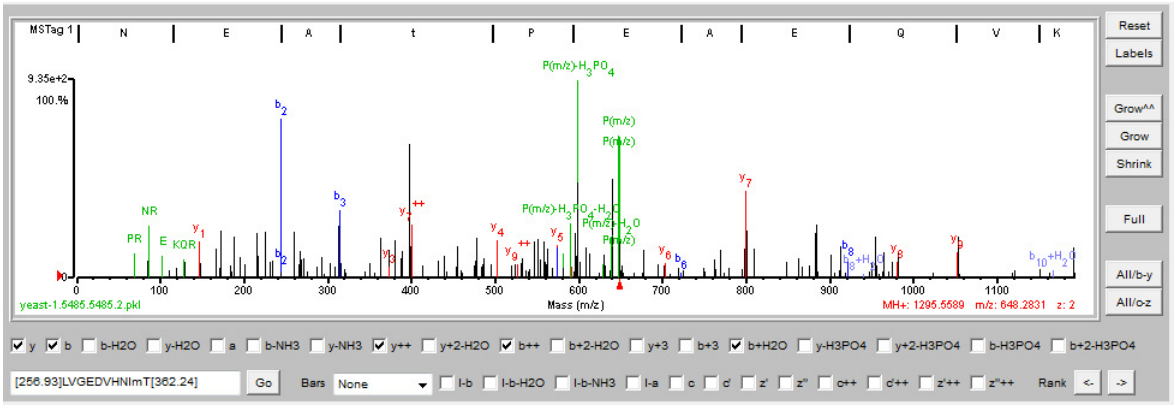


Figure 14 Spectrum #14 in Spectrum Viewer

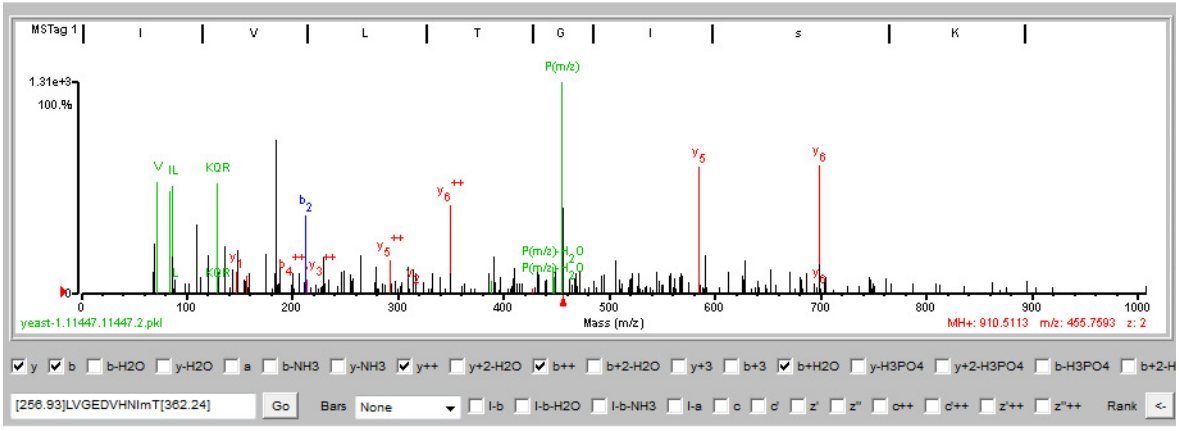
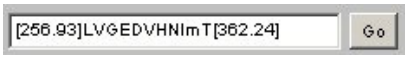
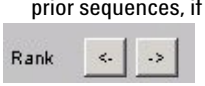


Figure 15 Spectrum #12 in Spectrum Viewer

Steps	Detailed Instructions	Comments

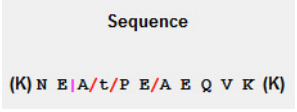
4 Use the **Labels** and **All/b-y** buttons to change the spectrum annotations.
• Use this step and the next ones for either or both peptide #14 or #12.
	a Click several times on the **Labels** button on the right side of the Spectrum Viewer.
	b Check that the peak labeling changes as you click.
	c 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.

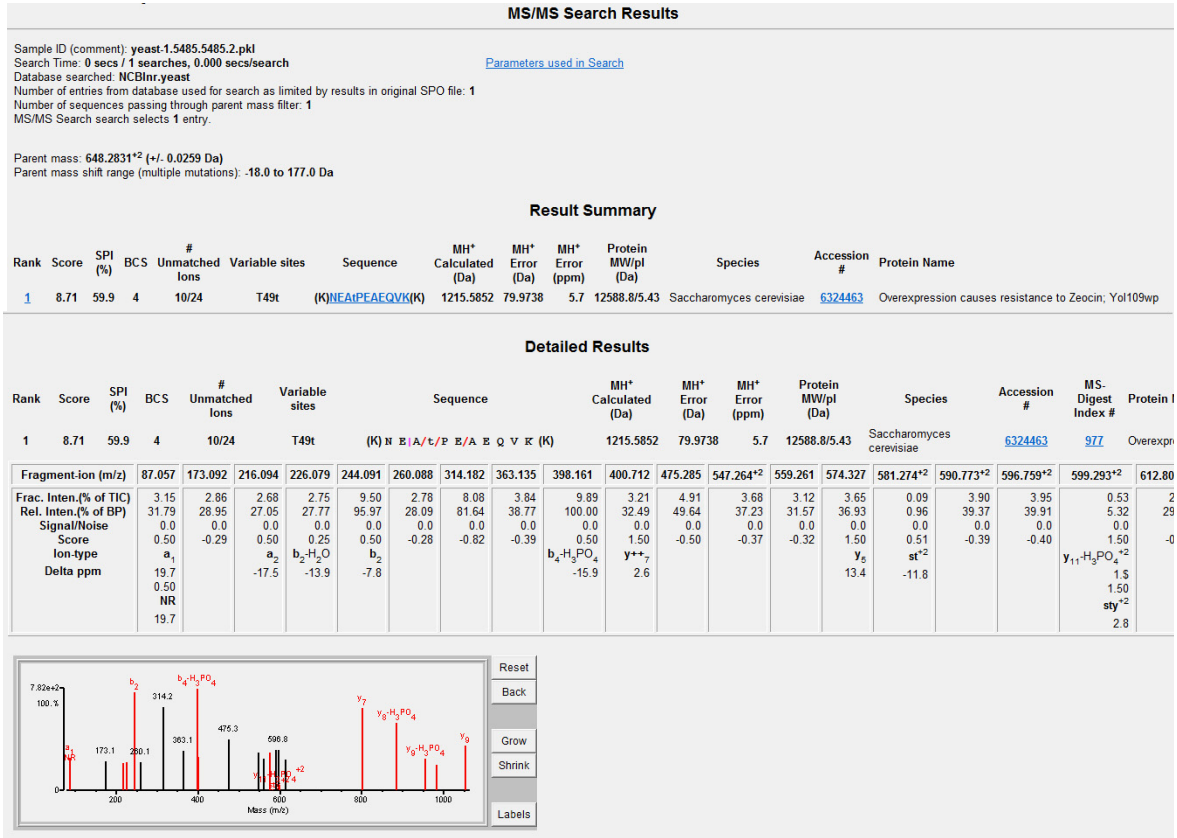
5 Use the **Grow**, **Grow^^**, **Shrink**, and **Reset** buttons to expand and shrink the spectrum in the vertical (y) direction.	**a** Click twice on the **Grow** button on the right side of the Spectrum Viewer. **b** Check that the spectrum expands vertically. **c** Click twice on the **Shrink** button 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.	

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.
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.

Steps	Detailed Instructions	Comments
1 If you have not already done so, display a Peptide report.	• Follow steps in "Produce a Peptide Report" on page 30.	
2 Display the MS/MS Search results for peptide #14	**a** In the summary report, click the 14th entry under the **Filename** header. **b** Check the bottom of the browser window, where you see the MS/MS Search results for the 14th spectrum. See Figure 16 on page 40 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 40.) **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)



MS/MS Search Results

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: **NCBInr.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: **648.2831**^{*2} **(+/- 0.0259 Da)**
Parent mass shift range (multiple mutations): **-18.0 to 177.0 Da**

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	8.71	59.9	4	10/24	T49t	(K)NEAtPEAEQVK(K)	1215.5852	79.9738	5.7	12588.8/5.43	Saccharomyces cerevisiae	6324463	Overexpression causes resistance to Zeocin; Yol109wp

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 I
1	8.71	59.9	4	10/24	T49t	(K) N E|A/t/P E/A E Q V K (K)	1215.5852	79.9738	5.7	12588.8/5.43	Saccharomyces cerevisiae	6324463	977	Overexpr

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	581.274^{*2}	590.773^{*2}	596.759^{*2}	599.293^{*2}	612.80
Frac. Inten.(% of TIC)	3.15	2.86	2.68	2.75	9.50	2.78	8.08	3.84	9.89	3.21	4.91	3.68	3.12	3.65	0.09	3.90	3.95	0.53	2
Rel. Inten.(% of BP)	31.79	28.95	27.05	27.77	95.97	28.09	81.64	38.77	100.00	32.49	49.64	37.23	31.57	36.93	0.96	39.37	39.91	5.32	29
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	
Score	0.50	-0.29	0.50	0.25	0.50	-0.28	-0.82	-0.39	0.50	1.50	-0.50	-0.37	-0.32	1.50	0.51	-0.39	-0.40	1.50	-0
Ion-type	a_1		a_2	b_2-H_2O	b_2				b_4-H_3PO_4	y^{++}_7				y_5	st^{+2}			y_{11}-$H_3PO_4^{+2}$	
Delta ppm	19.7		-17.5	-13.9	-7.8				-15.9	2.6				13.4	-11.8			1.5	
	0.50																	1.50	
	NR																	sty^{+2}	
	19.7																	2.8	

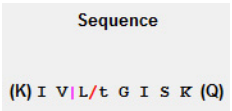
Figure 16 MS/MS Search results for peptide #14

Steps	Detailed Instructions	Comments
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 41 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.	Fragment-ion (m/z) Frac. Inten.(% of TIC) Rel. Inten.(% of BP) Score Ion-type Delta Da

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^+

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.

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 #12. • (Optional) Repeats steps 5-8 for peptide #12.	**a** In the summary report, click the 12th entry under the **Filename** header. **b** Check the bottom of the browser window, where you see the MS/MS Search results for the 12th 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 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: (K)I V | L/T G I s K(Q).

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.

Steps	Detailed Instructions	Comments
1 Extract the **y2** data.	• Repeat "Exercise 1. Extract the spectra" on page 10, 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 16, except when you **Select** the **Data Directory**, select **y2** rather than **y1**.	
3 Autovalidate the **y2** data.	• Repeat "Exercise 3. Autovalidate hiqh-quality results" on page 22, 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 Columns • Group results by: Directory	**a** Repeat "Produce a Protein Summary Report" on page 27, 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 Columns**. • For **Group results 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.

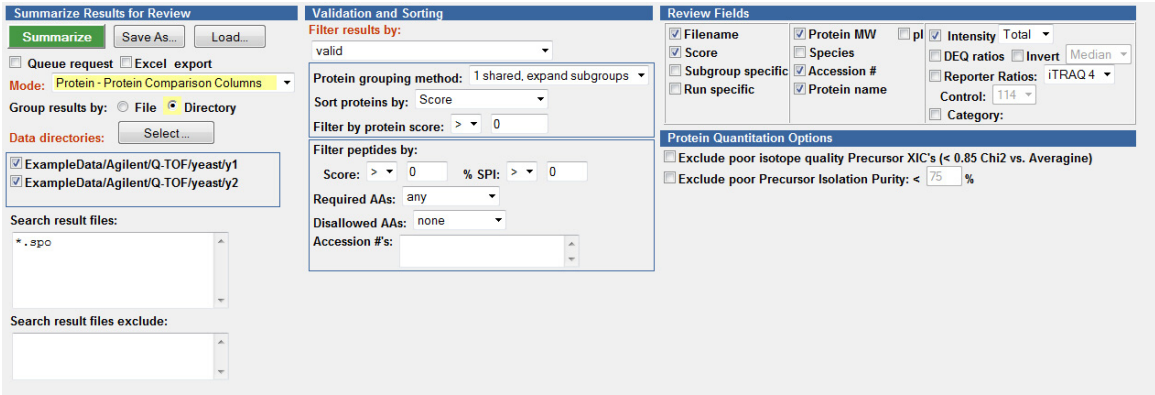


Figure 17 Protein/Peptide Summary settings to compare two samples

Steps	Detailed Instructions	Comments

5 Examine the overall summary report.	• Check that your results are similar to those in Figure 18 on page 45.
	  ◦ 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 TotalProtein 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.

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/y2 # spectra total intensity	ExampleData/Agilent/Q-TOF/yeast/y1 # spectra total intensity	Protein MW (Da)	Database Accession #	%AA Coverage	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	Group #	Subgroup #	Protein Name
109 1.04e+009	49 3.25e+008	46872.3	6321693	76.4	41	665.66	1	1.1	2 X enolase I; Eno1p
84 7.79e+008	53 2.76e+008	46970.4	6321968	77.5	37	586.40		1.2	enolase; Eno2p
54 1.49e+008	50 1.09e+008	69827.1	417149	71.8	42	587.55	2	2.1	5 X HEAT SHOCK PROTEIN SSA1 (HEAT SHOC
44 1.20e+008	39 9.12e+007	69639.9	6323004	60.4	36	474.51		2.2	member of 70 kDa heat shock protein family; Ssa2
18 3.37e+007	16 3.03e+007	70659.6	6319396	23.8	16	212.83		2.3	heat-inducible cytosolic member of the 70 kDa he
10 3.92e+006	8 3.21e+006	74523.7	6322426	19.9	12	141.96		2.4	Involved in translocation of nascent polypeptides
82 7.13e+008	58 3.49e+008	35845.2	3720	89.7	32	538.50	3	3.1	6 X (2128) reading frame
40 4.50e+008	39 2.62e+008	35960.3	6322468	80.4	26	401.83		3.2	glyceraldehyde 3-phosphate dehydrogenase; Tc
32 3.88e+008	38 2.46e+008	35863.4	6322409	84.6	26	368.25		3.3	Glyceraldehyde-3-phosphate dehydrogenase 1;
11 2.03e+008	13 1.24e+008	35609.1	1362244	30.5	9	133.49		3.4	glyceraldehyde-3-phosphate dehydrogenase (ph
14 2.60e+008	13 1.53e+008	35637.1	1362246	28.7	9	129.25		3.5	glyceraldehyde-3-phosphate dehydrogenase (ph
58 4.66e+008	58 3.75e+008	44794.6	10383781	80.5	33	514.41	4	4.1	3-phosphoglycerate kinase; Pgk1p
45 7.51e+007	33 3.43e+007	93743.9	6320593	54.2	37	491.72	5	5.1	translation elongation factor 2 (EF-2); Eft2p
29 1.19e+007	28 9.44e+006	229542.8	6322666	23.6	37	474.03	6	6.1	pentafunctional enzyme consisting of the following
40 1.55e+008	38 8.84e+007	54943.0	6319279	70.8	33	467.03	7	7.1	Required for START A in the cell cycle and sporu
35 5.27e+007	31 2.25e+007	86029.5	6320936	50	33	450.14	8	8.1	cobalamin-independent methionine synthase; Me
51 1.41e+008	51 7.19e+007	61722.5	6323073	64.6	30	449.55	9	9.1	4 X pyruvate decarboxylase; Pdc1p

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 an identity search, autovalidate, then create a saved results file against which you do a search with variable modifications and autovalidate once more. You may also continue searching against other databases.

When you are satisfied with the amount of information you have gleaned from your samples, you generate a final results summary

As you continue to process your 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.

Since 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 complete summarizing final validated results, you can check for remaining high-quality spectra with Spectrum Summary. Then you identify the validated peptides with Sherenga *de novo* sequencing.

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

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”.

For this series of exercises, you extract the data, do an identity search, autovalidate, do a variable modifications search, autovalidate and summarize all the valid results.

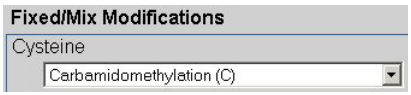
You can do a variable modifications search against a full database, a species-specific subset database, or a list of proteins that you have already identified in your sample. In this series of exercises, you search against the latter. This saves time because the modified peptides you find in variable modifications mode are often derived from the proteins you already identified in identity mode. The variable modifications search is much faster when you search a smaller list.

Another option for an iterative workflow is to search a second database. For example, if you first searched against a human species-specific database, try the larger mammals database. While not described in detail here, you may also conduct a no enzyme (or half-enzyme) search against the list of proteins that you have already identified in your sample. Such a search identifies any partial tryptic or non-tryptic peptides.

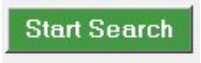
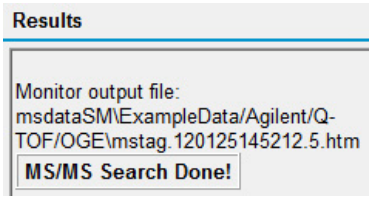
Exercise 1: Extract the data and do an identity search and validation

In this exercise you extract the data just as you did with the basic workflow, but instead of doing a variable modifications search after the extraction, you do an identity search.

Steps	Detailed Instructions	Comments
1 Extract the yeast **y1** data.	• Use the instructions in "Exercise 1. Extract the spectra" on page 10.	
2 Navigate to the MS/MS Search page.	• Do one of the following: ◦ From the DataExtractor page, click the **MS/MS Search** button.  ◦ From the Spectrum Mill home page, click the link to MS/MS Search.  	
3 Remove prior MS/MS search results.	• Mark the **Remove all prior MS/MS Search results** 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.
4 Check that your Data Directory is set to **y1** and marked.	**a** 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. **b** Mark the check box, if necessary.	

Steps	Detailed Instructions	Comments
5 Select the NCBInr.yeast subset database.	• For **Database** (under **Search Parameters**), select the NCBInr.yeast subset 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.
6 Set the appropriate cysteine modification. ◦ If the carbamidomethylation modification is set, move on to the next step.	**a** Click **Choose...** near the middle of the page.  **b** Under the **Cysteine** heading, select **Carbamidomethylation**.  **c** Click **OK.** The name of the modification appear as in Figure 19 on page 54.	• 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 form 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. • For more information about choosing modifications, see the online help. • Your system administrator can configure custom modifications.

Steps	Detailed Instructions	Comments
7 Set up the **Search Mode**. • Calculate reversed database scores. • Turn on dynamic peak thresholding. • Turn on discriminant scoring - Agilent Q-TOF standard. • Search Mode: Identity	**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 **Agilent Q-TOF Standard**. • Agilent recommends that you turn discriminant scoring on for an iterative search in order to calculate the most accurate local FDR %, which is used in the iterative autovalidation. • When discriminant scoring is disabled, the score is used for all FDR calculations. **d** Verify that the **Search mode** is set to **Identity mode**.	• 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 for Agilent Q-TOF data. • 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.
8 Set other parameters as shown in Figure 19 on page 54.	**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.	• Changes from defaults are highlighted in yellow. • 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.

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. **c** Scroll to the *top* of the **Results** area to see the message that indicates that the search is finished.  	• The command is automatically placed in the queue for execution after other tasks in the queue have been executed. • 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.

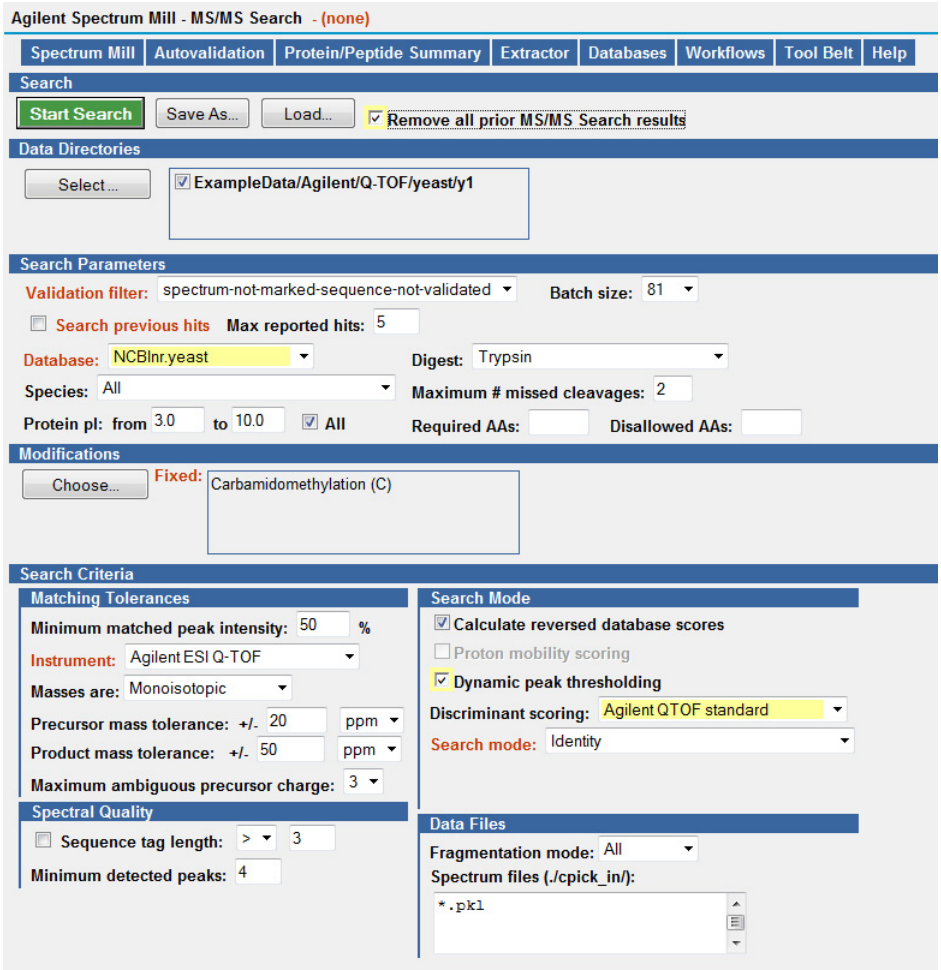


Figure 19 MS/MS Search settings

Steps	Detailed Instructions	Comments

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

11 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.	
12 Set up autovalidation with the Auto thresholds - Discriminant strategy, Peptide mode. • Clear all previous validations. • Select Local FDR. • Keep the default of 1%. See Figure 20 on page 56.	**a** For **Strategy**, click **Auto thresholds - Discriminant**. **b** For **Mode**, keep the default of **Peptide**. **c** Click **Clear All**. **d** Click **Local FDR**.	• In combining results from multiple searches (from successive iterations of the iterative workflow), it is preferable to use the local FDR rather than the global FDR. • The local FDR measures the error rate of an individual validated spectrum, while the global FDR measures the error rate of an entire set of validated spectra.
13 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 (Figure 21 on page 56).	• 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.
14 Close the Autovalidation window.	• Close the window only after you see the word “finished”.	

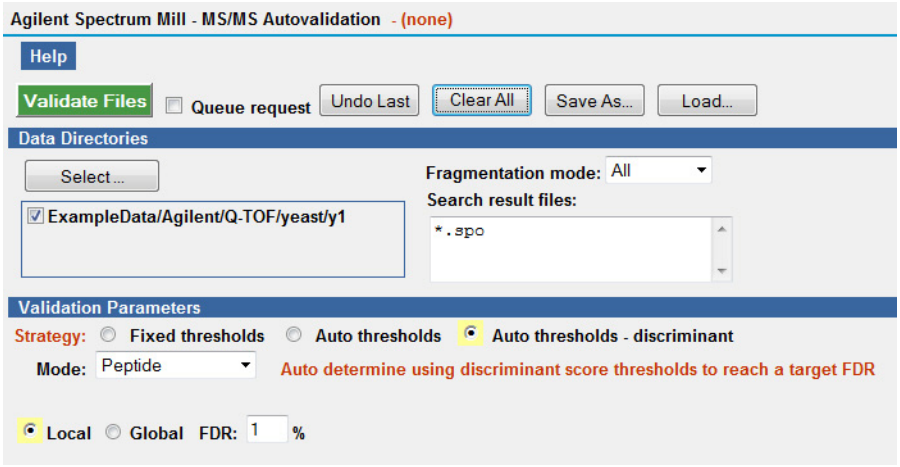


Figure 20 Autovalidation settings

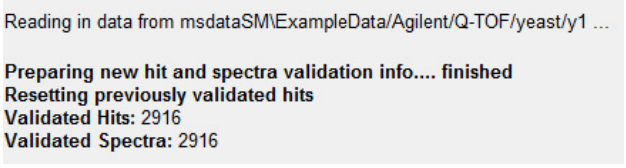


Figure 21 Autovalidation results from Identity search

Exercise 2. Create a list of identified proteins for variable modifications search

This exercise shows how to create a file of previously validated hits, which is a small database of proteins that you have already identified in your sample. In this exercise, you create the file using the Tool Belt page.

Alternatively, you can create the file automatically from the MS/MS Search page. 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. (On the page, be sure to select the database that you searched previously.)

Steps	Detailed Instructions	Comments
1 Navigate to the Tool Belt page.	• From the Spectrum Mill home page, click the link to the **Tool Belt** page.	• There are other ways to navigate to the Tool Belt page. Check the top of your current page for a **Tool Belt** button.
2 Create a saved results file. • Fill in the page as shown in Figure 22.	**a** Click the **Create saved results file** option near the top of the page. **a** Select the data directory that you previously searched (**yeast\y1**). **b** Select the database you previously searched (**NCBInr.yeast**). **c** Click **Create File**.	• You will create a file of previously validated hits from the search results from "Exercise 1: Extract the data and do an identity search and validation" on page 50. • You will later search this protein list in variable modifications mode.

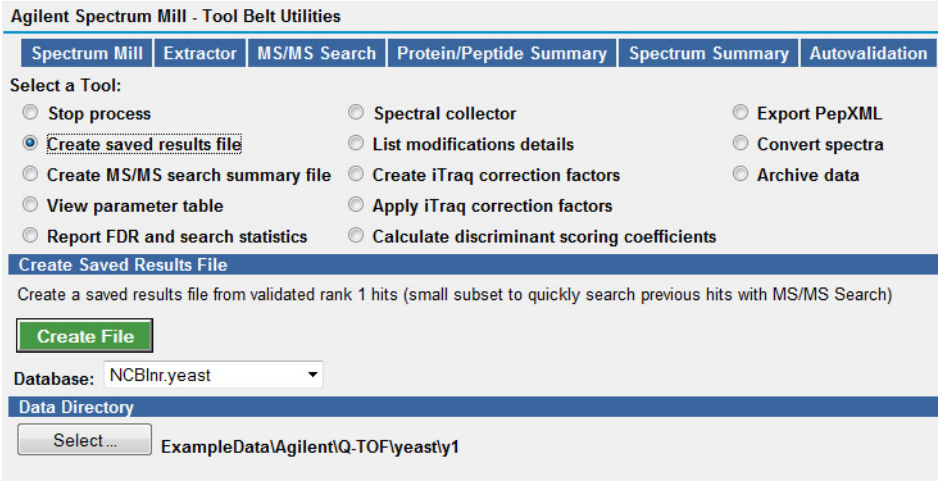


Figure 22 Create file of previously validated proteins

Exercise 3. Search in variable modifications mode and validate

This exercise describes a search in variable modifications mode. Choosing variable modifications against which to search is a trade-off. If you do not search with a modification, you cannot find it. If you search with too many modifications, the search runs too slowly.

To avoid having the search run too slowly you can take one or both of two actions:

- Search a smaller database.
- Choose fewer variable modifications for each iterative search.

In this exercise, you search in variable modifications mode against database hits that you validated previously. When you process your own samples, you could instead search in variable modifications mode against a full or species subset database. If you search a species subset database, you may want to start with a well-defined species (e.g., human), and then do a second variable modifications search on a broader range of species (e.g., mammal). The broader the database search, the longer the search time. A search of previous hits as in this exercise is very efficient.

When you process your own samples, you may conduct multiple iterative searches, with autovalidation between each set of search conditions:

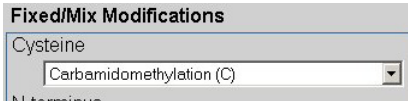
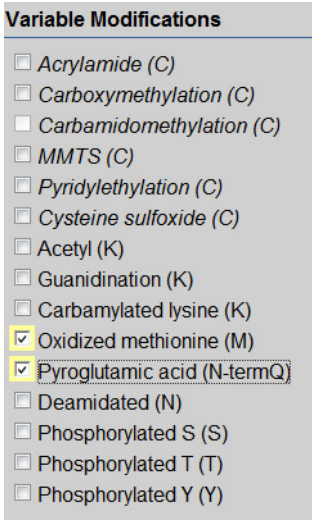
- Combine the following common modifications into a single search: oxidized methionine (methionine sulfoxide) and pyroglutamic acid.
- If you suspect phosphorylation, combine the following modifications into a single search: phosphorylated S, phosphorylated T, and phosphorylated Y.
- Search for any other modifications you suspect for your sample.

NOTE

To identify the modifications to use in your variable modifications search when you are unsure of their identity, do a separate mass gap search for unknown modifications.

See Chapter 3 of the *Familiarization Guide* for exercises to help you do mass gap searches.

Steps	Detailed Instructions	Comments
1 Navigate to the MS/MS Search page.	• From the Tool Belt page or 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 **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 Set **Search Parameters**. • Database: NCBInr.yeast • Search previous hits	**a** For **Database**, select **NCBInr.yeast**, as before. **b** Mark the check box next to **Search previous hits.**	• This searches the file of previous hits you created for this data directory. (See "Exercise 2. Create a list of identified proteins for variable modifications search" on page 57.)

Steps	Detailed Instructions	Comments
4 Choose the appropriate modifications. • Fixed - Carbamidomethylation, if necessary • Variable: Oxidized methionine Pyroglutamic acid	**a** Click **Choose...** near the middle of the page.  **b** Choose **Fixed/Mix Modifications**. Under the **Cysteine** heading, select **Carbamidomethylation** .  **c** Choose **Variable Modifications**. Select **Oxidized methionine** and **Pyroglutamic acid**.  **d** Click **OK.** The names of the modifications appear as in Figure 23 on page 63.	• 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 form 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**. • For more information about choosing modifications, see the online help. • Your system administrator can configure custom modifications.

Steps	Detailed Instructions	Comments
5 Set up the **Search Mode**. • Calculate reversed database scores • Turn on dynamic peak thresholding • Turn on discriminant scoring - Agilent Q-TOF standard.	**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 **Agilent Q-TOF Standard**.	
6 Set the other parameters as shown in Figure 23 on page 63.	**a** Clear the **Remove all prior MS/MS results** check box, if necessary. **b** Mark the **Search previous hits** check box. **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.	• Changes are highlighted in yellow. • 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.
7 Start the search.	**a** Click **Start Search**. **b** When **Monitor Results** appears, click the link. **c** Scroll to the *top* of the **Results** area to see the message that indicates that the search is finished.	• MS/MS Search processes all spectral files in the directory. • Search time varies depending on the size of the database searched. • For the same data set and database, variable modifications and 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.

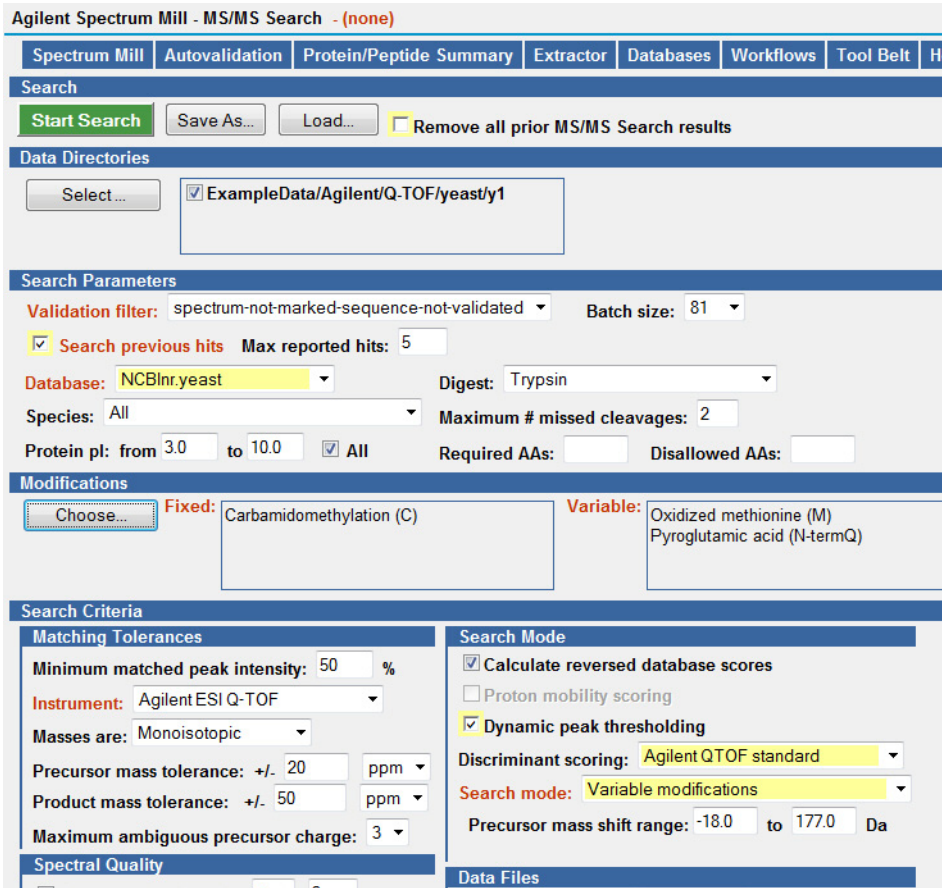


Figure 23 MS/MS Search settings for search of previous hits for variable modifications

Steps	Detailed Instructions	Comments

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

Steps	Detailed Instructions	Comments
9 Check that your Data Directory is set to **y1**.	• If you have just completed the MS/MS Search, your **y1** data directory should already be set correctly. If not, click the **Select...** button to select the **y1** folder.	
10 Set up autovalidation with the Auto thresholds - Discriminant strategy, Peptide mode. • Select Local FDR. • Keep the default of 1%. See Figure 20 on page 56.	**a** For **Strategy**, click **Auto thresholds - Discriminant**. **b** For **Mode**, keep the default of **Peptide**. **c** Click **Local FDR**.	• You do not clear previous validation results here because in an iterative workflow all results are summarized after you are satisfied your searches have elicited enough information on your sample.
11 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 (Figure 24 on page 64).	• 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.
12 Close the Autovalidation window.	• Close the window only after you see the word "finished".	

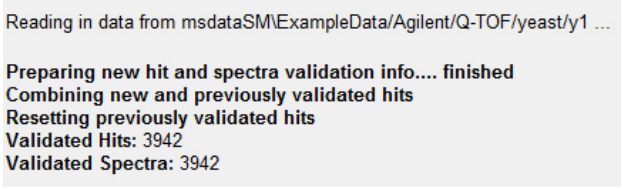


Figure 24 Autovalidation results from Variable Modifications search

Task 2. Review valid results and remaining spectra

Exercise 1. Review valid results

The Spectrum Mill workbench provides a means to segregate search results that contain a valid interpretation of an MS/MS spectrum from those that do not. Results that are *not* validated can then be subjected to subsequent rounds of searches (against other databases, for example). Results that *are* validated can be summarized in a results table.

Steps	Detailed Instructions	Comments
1 Navigate to the Protein/Peptide Summary page.	• Click **Protein/Peptide Summary** from either the MS/MS Search page or the Spectrum Mill home page.	
2 Set up the parameters for a Summary Details report. • Sort peptides by Score. • Filter all scores by 0.0. • Add these to the Review Fields: FDR (Discriminant) Var mod sites Modification names	**a** From the mode list, select **Protein Summary Details**. **b** From the **Sort peptides by** list, select **Score**, if necessary. **c** For **Filter by protein score**, type `0.0`. **d** For **Filter peptides by**, type `0.0` for **Score** and for **%SPI**. **e** Under **Review Fields**, mark the **FDR (Discriminant)**, **Var mod sites** and **Modification names** check boxes, as shown in Figure 25 on page 66.	• This summary includes results validated from both identity mode and variable modifications searches. • If the variable modifications occur on the N-terminus, C-terminus, or cysteines, additionally mark the appropriate check box(es) in the last column of the **Review Fields**.

Steps	Detailed Instructions	Comments
3 Summarize the results. • Review report for Group #6. • Check that the display contains the names of the variable modifications found and the appropriate headers.	**a** Click **Summarize**. **b** Review the report for Group #6. Note the peptides containing the oxidized methionine sites and the pyroglutamic acid site. **c** Check that the data display for the groups includes the additional headings **Discriminant**, **Local FDR (%)**, **Global FDR (%)**, **FDR Search #**, **Variable Sites** and **Modifications Map**, as shown in Figure 26 on page 67.	• This report shows a Protein Summary section and a section for reporting the peptides for each group. • Discriminant - score 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. • Local FDR (%) - expected error rate of an individual spectrum. • Global FDR (%) - expected error rate for the set of spectra at or above the given discriminant score. It is calculated and displayed here but not used in autovalidation unless specified. • Variable Sites - abbreviations for the variable modifications locations and types. • Modifications Map - name of variable modification at that site

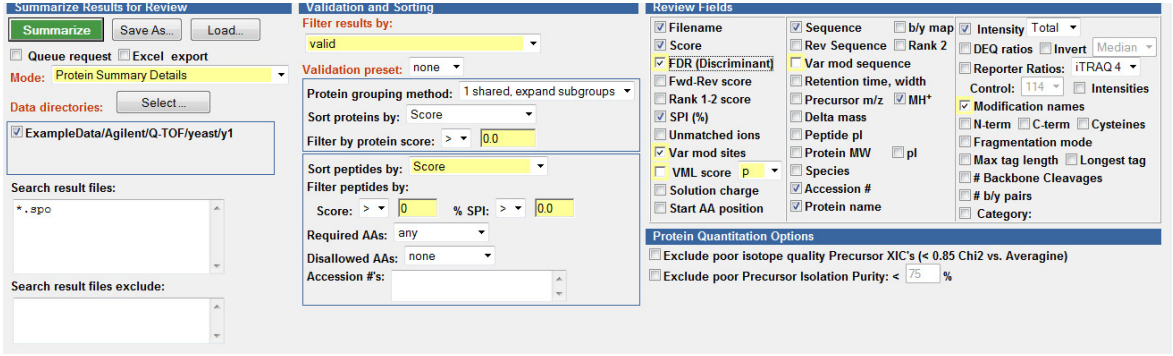


Figure 25 Protein Summary Details settings

Group (#)	Subgroup (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Total Protein Spectral Intensity	Database Accession #	Protein Name
6	6.1	47	26	358.27	57.9	5.89e+007	6323073	pyruvate decarboxylase; Pdc1p

#	Filename	z	Score	Discriminant	Local FDR (%)	Global FDR (%)	FDR Search #	SPI (%)	Spectrum Intensity	Variable Sites	Sequence	Modifications Map
1	yeast-1.13773.13773.2	2	21.07	23.4401	<0.1%	<0.1%	1	85.5	8.43e+005		(R)WAGNANELNAAYAADGYAR(I)	
2	yeast-1.20640.20640.3	3	20.80	21.3927	<0.1%	<0.1%	1	97.7	3.00e+005		(R)MIEIMLPVFDAPQNLVEQAK(L)	
3	yeast-1.19425.19521.3	3	20.62	20.3754	<0.1%	<0.1%	1	93.6	2.32e+006		(K)AQYNEIQGWDHLSLLPTFGAK(D)	
4	yeast-1.20437.20437.3	3	19.75	14.9302	<0.1%	<0.1%	1	96.2	2.16e+004		(K)EVIDTILALVK(D)	
5	yeast-1.20148.20148.3	3	19.23	19.4799	<0.1%	<0.1%	1	92.6	1.21e+006		(K)LIDLTQFPAFVTPMGK(G)	
6	yeast-1.20426.20426.2	2	19.10	15.2844	<0.1%	<0.1%	1	91.1	1.95e+005		(K)EVIDTILALVK(D)	
7	yeast-1.19203.19342.3	3	18.07	19.5553	<0.1%	<0.1%	1	94.1	4.55e+006		(R)TTYVTQRPVYLGLPANLVDLNVPAK(L)	
8	yeast-1.19655.19655.3	3	18.04	19.0360	<0.1%	<0.1%	2	94.5	2.76e+005	M247m	(K)LIDLTQFPAFVTPmGK(G)	m:Oxidized methionine
9	yeast-1.20635.20635.2	2	16.94	14.7662	<0.1%	<0.1%	1	81.0	5.16e+005		(R)MIEIMLPVFDAPQNLVEQAK(L)	
10	yeast-1.11568.11568.3	3	16.09	16.5193	<0.1%	<0.1%	1	80.5	6.96e+006		(R)YGGVYVGTLSKPEVK(E)	
11	yeast-1.19886.19886.2	2	15.72	19.8115	<0.1%	<0.1%	1	79.3	7.20e+005		(R)MSANISETTAMITDIATAPAEIDR(C)	
12	yeast-1.19513.19513.3	3	15.17	17.1574	<0.1%	<0.1%	1	100.0	7.22e+005		(K)KLIDLTQFPAFVTPMGK(G)	
13	yeast-1.10516.10516.3	3	15.05	15.2542	<0.1%	<0.1%	1	77.8	1.47e+006		(K)NPVILADACCSR(H)	C:Carbamidomethylation
14	yeast-1.2691.2806.2	2	14.92	12.8055	<0.1%	<0.1%	1	88.9	1.62e+006		(K)SFNDNSK(I)	
15	yeast-1.19906.19906.3	3	14.64	16.5382	<0.1%	<0.1%	1	71.2	6.86e+005		(R)MSANISETTAMITDIATAPAEIDR(C)	
16	yeast-1.19209.19396.4	4	14.51	16.1476	<0.1%	<0.1%	1	96.1	1.05e+006		(R)TTYVTQRPVYLGLPANLVDLNVPAK(L)	
17	yeast-1.9669.9734.2	2	14.42	13.9073	<0.1%	<0.1%	1	92.2	5.02e+006		(K)IYEVEGMR(W)	
18	yeast-1.21057.21057.2	2	14.26	16.2047	<0.1%	<0.1%	2	92.6	4.14e+004	Q16q	(K)qVNVNTVFGLPGDFNLSLLDK(I)	q:Pyroglutamic acid
19	yeast-1.9482.9482.2	2	14.04	16.4979	<0.1%	<0.1%	1	85.8	4.65e+004		(R)TPANAAVPASTPLK(Q)	
20	yeast-1.19631.19631.2	2	13.52	15.4284	<0.1%	<0.1%	2	77.1	1.26e+005	M247m	(K)LIDLTQFPAFVTPmGK(G)	m:Oxidized methionine
21	yeast-1.20550.20550.2	2	13.44	15.8405	<0.1%	<0.1%	1	80.4	4.08e+004		(K)QVNVNTVFGLPGDFNLSLLDK(I)	
22	yeast-1.20534.20554.3	3	13.06	11.0267	0.6%	<0.1%	1	86.9	2.62e+004		(K)QVNVNTVFGLPGDFNLSLLDK(I)	
23	yeast-1.8167.8167.2	2	12.88	13.8301	<0.1%	<0.1%	1	84.0	1.36e+005		(K)NIVEFHSDHMK(I)	

Figure 26 ProteinSummary Details report for identity and variable modifications searches for Group #6.

Exercise 2. 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.

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.

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 Sort by and filter by maximum tag length • Set a maximum tag length of 7.	• For this exercise, keep the **Maximum tag length** option and set a filter of **>7**.	• With the **Maximum 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.
a Check to see that your settings look like those in Figure 27 on page 69.	• Change any settings that do not match Figure 27 on page 69.	

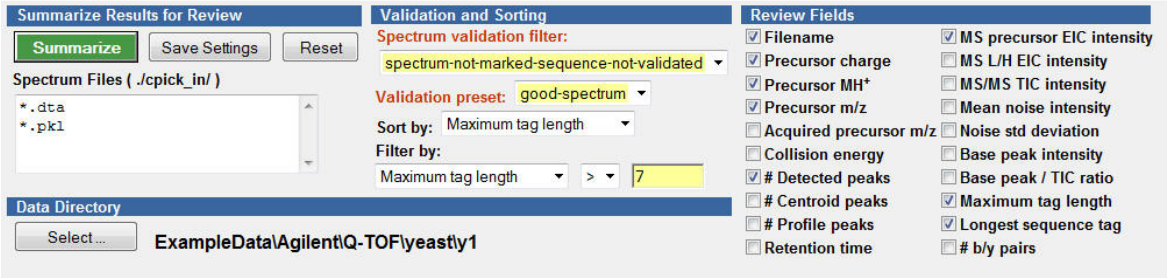


Figure 27 Spectrum Summary settings

Steps	Detailed Instructions	Comments
6 Summarize the spectra. • Results should look like those in Figure 28 on page 71.	**a** Click **Summarize**. **b** Check that your results are similar to those in Figure 28 on page 71. Check that your results show: • A table at the top of the page • The Spectrum Viewer with the first spectrum at the bottom 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.
7 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 bottom 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 35.
8 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.

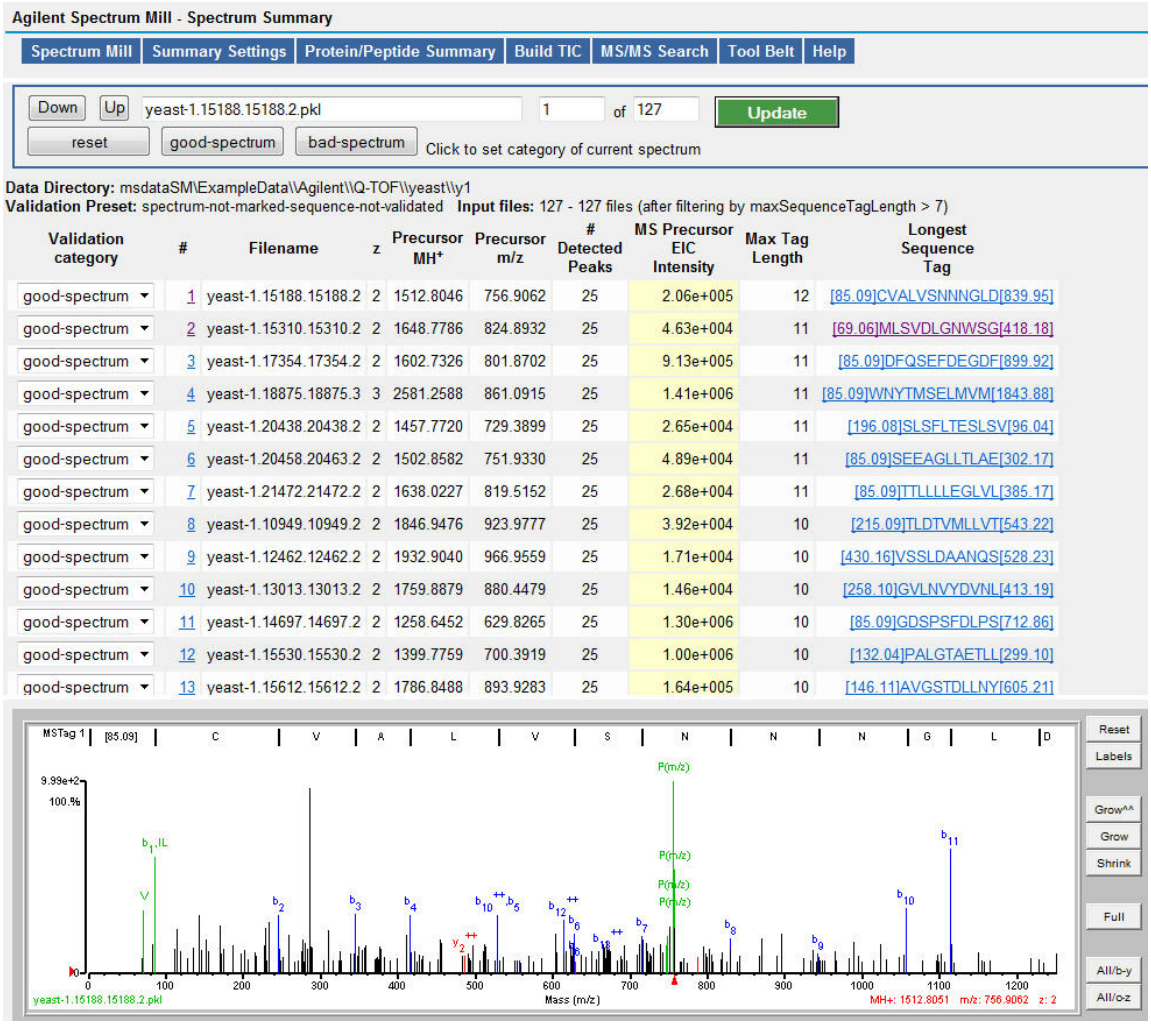


Figure 28 Spectrum Summary results

Steps	Detailed Instructions	Comments

9 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 bottom of the page.
	  ◦ In the summary report, click the number **2** under the **#** header. Again, this displays the spectrum at the bottom 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.

10 Examine and categorize the second spectrum as a good spectrum.	• In the **Validation category**, keep the **good-spectrum** designation.	
11 Repeat step 9 and step 10 for as many spectra as you care to review.		
12 Update the spectral designations.	• Click **Update**.	• This makes your designations a permanent part of the data record. • You have now designated approximately 127-128 spectra as good (if you did not change any to "bad-spectrum". • Your results may vary slightly.
13 (Optional) Vary input parameters and investigate the results.	**a** To change parameters and create a new summary table, click the **Summary Settings** button to 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 3. 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*.

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 29 on page 74.	**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.

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 30 on page 75. **c** Scroll to the *top* of the **Results** area to see the message that indicates that sequencing is finished.	

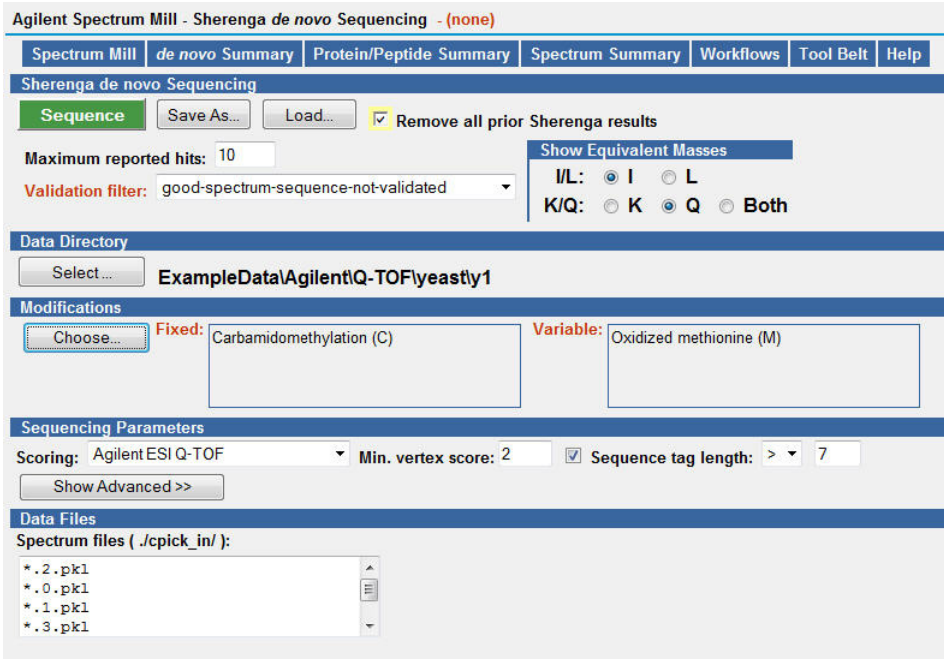


Figure 29 Sherenga *de novo* Sequencing settings

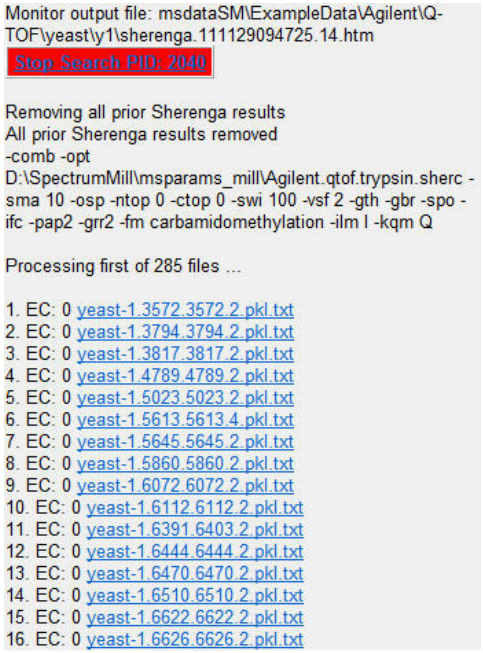


Figure 30 Display while Sherenga *de novo* Sequencing is in process

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.

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 31 on page 76.	• Set Min. vertex score to `2`.	

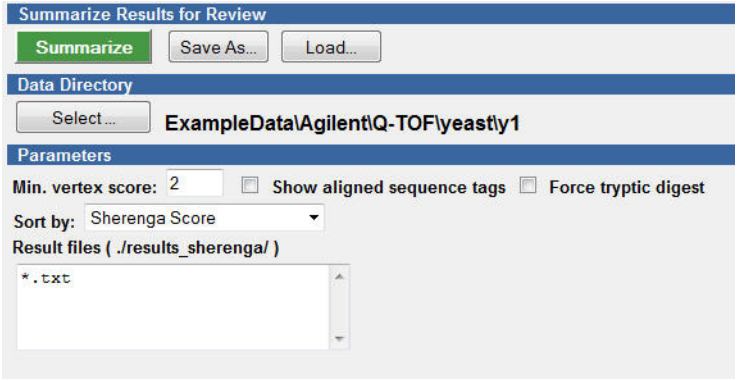


Figure 31 Settings for Sherenga Summary page

Steps	Detailed Instructions	Comments
4 Summarize the results.	• Click **Summarize**.	
5 View the overall results table.	**a** Check to see that your results look like the Sherenga Summary results shown in Figure 32 on page 77. **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)

SS	TL	TS	MS/MS Search Delta MH^+	Filename	Sherenga Sequence Tag	MS/MS Search Result
87	8	22.1	0.0	yeast-1.17489.17882.2.pkl.txt	VPTVDVSVVDITVQ	VPTVDVSVVDLTVK
76	13	22.5	0.0	yeast-1.16410.16787.2.pkl.txt	[213.12]IDTDCQYAYITIGR	VNLDTDCQYAYLTGIR
73	11	20.0	0.0	yeast-1.19422.19515.2.pkl.txt	[255.12]MADAINAI[301.13][255.13]	SSVLADALNAINNAEK
67	13	23.7	0.0	yeast-1.19496.19511.2.pkl.txt	[200.12]IWVIIEQAY[247.07][154.11]	TVIGEVLLEQAYGGMR
61	10	8.9	0.0	yeast-1.16412.16777.3.pkl.txt	[142.08]AM[214.12]YAY[288.07]DP[184.12]A[172.10]	VNLDTDCQYAYLTGIR
61	9			yeast-1.17135.17361.2.pkl.txt	[257.11]VYmDVEA[172.07]Q[210.12][213.13]	
59	9			yeast-1.16078.16100.2.pkl.txt	[228.12]SEASVD[270.12][238.03]I[225.15]	
58	8	19.3	0.0	yeast-1.11439.11888.2.pkl.txt	[171.07]PTVEVP[172.10]VV[234.08]	GNPTVEVELTTEK
58	11	20.5	-0.0	yeast-1.14343.14406.2.pkl.txt	[214.13]ADYDAIDIA[270.15]	LTADYDALDIANR
58	11	23.4	0.0	yeast-1.20456.20456.2.pkl.txt	[214.13]AETIAEm[208.11][185.09][199.13]	TIAETLAEELINAAK
57	8	13.1	-0.0	yeast-1.10000.10416.2.pkl.txt	[241.15]EGVGS[217.15]PVQ	QLENVSSNIVK
57	11			yeast-1.20458.20463.2.pkl.txt	[200.10]mIQEIIG[226.07]EQ	
56	9	17.9	0.0	yeast-1.17897.18106.2.pkl.txt	[234.11]TmQAIRR[170.10]	TSFFQALGVPTK
55	8			yeast-1.17013.17262.2.pkl.txt	[213.12]IDTDCEYAYIT[326.20]	
53	9	20.1	-0.0	yeast-1.20174.20174.2.pkl.txt	[699.33]AIAAIQIED[768.36]	IEAALSDALAALQIEDPSADELR
52	8			yeast-1.21507.21573.3.pkl.txt	[404.20][265.14]V[273.09]W[345.13]VEIISHII[225.07]	
51	8	18.6	0.0	yeast-1.18520.18763.2.pkl.txt	[170.11]D[262.10]G[200.13]IVI[430.15]	AVDDFLLSLDGTANK
50	9			yeast-1.13123.13130.2.pkl.txt	[213.11]EIGT[248.13]AI[225.08]	
49	8			yeast-1.14137.14362.2.pkl.txt	ISVQD[228.11][241.18]	
49	8			yeast-1.18517.18517.2.pkl.txt	[602.29]MIT[228.13]ATAD[566.27]	
49	10	17.6	0.0	yeast-1.5860.5860.2.pkl.txt	[276.10]N[316.11]AEWA[210.16]Am[226.06]	KAEASGEAADEADEADEE
48	9	16.7	0.0	yeast-1.17912.18171.2.pkl.txt	[200.12]GWV[260.15]DmTQ	SIGGEVFIDFTK
48	11			yeast-1.20438.20438.2.pkl.txt	PVSISETIE[236.19][247.10]	
48	9	18.5	0.0	yeast-1.9018.9018.2.pkl.txt	[202.10]EEINN[202.10]PDR	TTEEQVENAVDR
47	9	19.2	0.0	yeast-1.10534.10534.2.pkl.txt	[416.13]SVT[226.10][158.08][246.12][168.05]	ADDDSVTIISAGNDK
47	10			yeast-1.12462.12462.2.pkl.txt	W[343.15]SQNAWIQ[241.14][316.13]	

Figure 32 Sherenga Summary results

Steps	Detailed Instructions	Comments
6 Review detailed results for a single spectrum. • Review the MS for the 3rd 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 33 on page 79. (You may need to move the page section split line again.) • Detailed Sherenga results, as shown in Figure 34 on page 79. **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** button. 	• 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 35.
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.

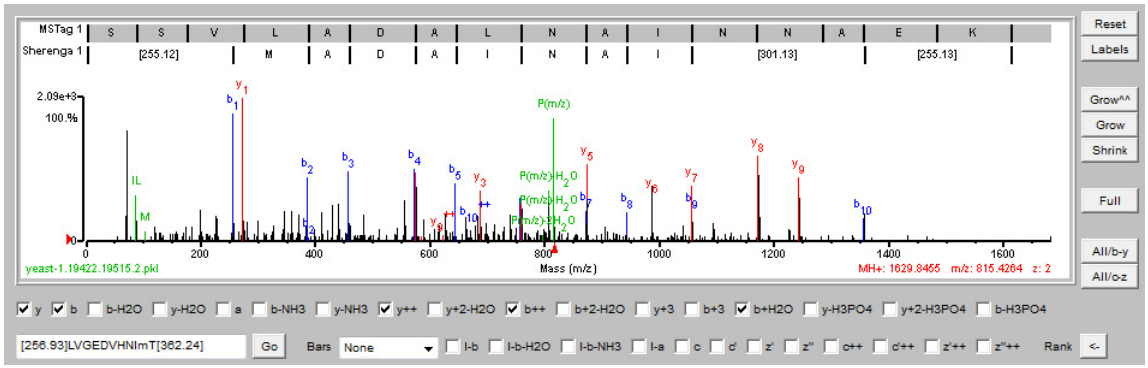


Figure 33 Spectrum Viewer with Sherenga results

ExampleData\Agilent\Q-TOF\yeast\y1 yeast-1.19422.19515.2.pkl.txt yeast-1.19422.19515.2.pkl.spo

Rank	Orig. Score	Score	MS-Tag Sequence	Sherenga Sequence Tag	Matched Precursor MH+
		20.02	SSVLADALNAINNAEK	[255.12]MADAINAI[301.13][255.13]	1629.8442
			S S V L A D A L N A I N N A E K		
1.	74.6	74.6	[255.12] M A D A I N A I [301.13] [255.13]		1629.85
2.	73.7	73.7	[255.12] M A D A I N A I [188.05] I [255.13]		
3.	72.4	72.4	[272.14] N A D A I N A I [301.13] [255.13]		
4.	71.9	71.9	[255.12] [202.13] D A I N A I [188.05] I [255.13]		
5.	71.5	71.5	[272.14] N A D A I N A I [188.05] I [255.13]		
6.	71.3	71.3	[273.14] I A D A I N A I [556.25]		
7.	70.7	70.7	[255.12] M A D A I N A I [556.25]		
8.	70.7	70.7	[386.22] A D A I N A I [301.13] [255.13]		
9.	70.4	70.4	[273.14] I A D A I N A I [188.05] [368.21]		
10.	70.3	70.3	[273.14] I A D A I N A I [216.04] [340.21]		

Figure 34 Detailed results from Sherenga *de novo* Sequencing

3 Unknown Modifications

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 3. Search in variable modifications mode and validate” on page 59 of Chapter 2.) 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.

Unlike Chapters 1 and 2 this chapter uses the Agilent Q-TOF/OGE example data to illustrate the mass gap search. This search takes much less time for the single fraction OGE data than for the more complex yeast data. 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.

As with the variable modifications described in Chapter 2, it is best to conduct a mass gap search against a file of previously validated hits. That way, you decrease search time and minimize the chance of random matches.

You conduct a mass gap search as part of an iterative workflow with an identity search and validation run first. You can next run a variable modifications search and validation for those modifications you know are present and then a mass gap search. After you conduct a mass gap search, you autovalidate the results. The Spectrum Mill workbench adds the newly validated results to those you validated previously.

The exercises in this chapter include only the identity search and validation followed by the mass gap search and validation.

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 identity search.

Steps	Detailed Instructions	Comments
1 Extract the QTOF/OGE folder. • Remove prior results	• Follow the steps for extraction in Chapter 2, "Exercise 1: Extract the data and do an identity search and validation" on page 50. • 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 Identity mode. • Remove prior MS/MS search results • Substitute NCBInr.ecoli database for the yeast database.	• Follow the steps for the MS/MS identity search in Chapter 2, "Exercise 1: Extract the data and do an identity search and validation" on page 50. • Mark the **Remove all prior MS/MS search results** check box. • From the Database list, select **NCBInr.ecoli**.	
3 Autovalidate the results. • Clear all prior validated results.	• Follow the steps for autovalidation in Chapter 2, "Exercise 1: Extract the data and do an identity search and validation" on page 50. • 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.

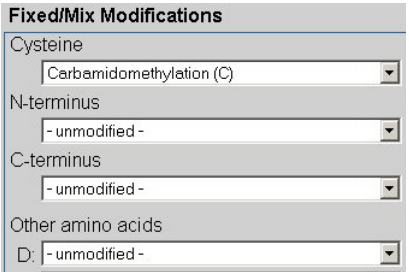
In identity mode, you search for exact matches in the selected database.

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.

Steps	Detailed Instructions	Comments
1 Navigate to the MS/MS Search page.	• From the Tool Belt page or 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: NCBInr.ecoli • Search previous hits.	**a** For **Database**, select **NCBInr.ecoli**, as before. **b** 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.

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 35 on page 87.	• This is the default modification. • For more information about choosing modifications, see the online help. • Your system administrator can configure custom modifications.
5 Set up the **Search Mode**. • Calculate reversed database scores. • Turn on dynamic peak thresholding. • Enable Discriminant Scoring with Agilent Q-TOF Standard. • 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 **Agilent Q-TOF Standard**. **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.

Steps	Detailed Instructions	Comments
6 Set the other parameters as shown in Figure 35 on page 87. • Note that +/- 130 Da for the precursor mass shift is a reasonable default in most cases.	**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.	• Changes are highlighted in yellow. • 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. • This search takes about 20 minutes for the OGE data. • 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.

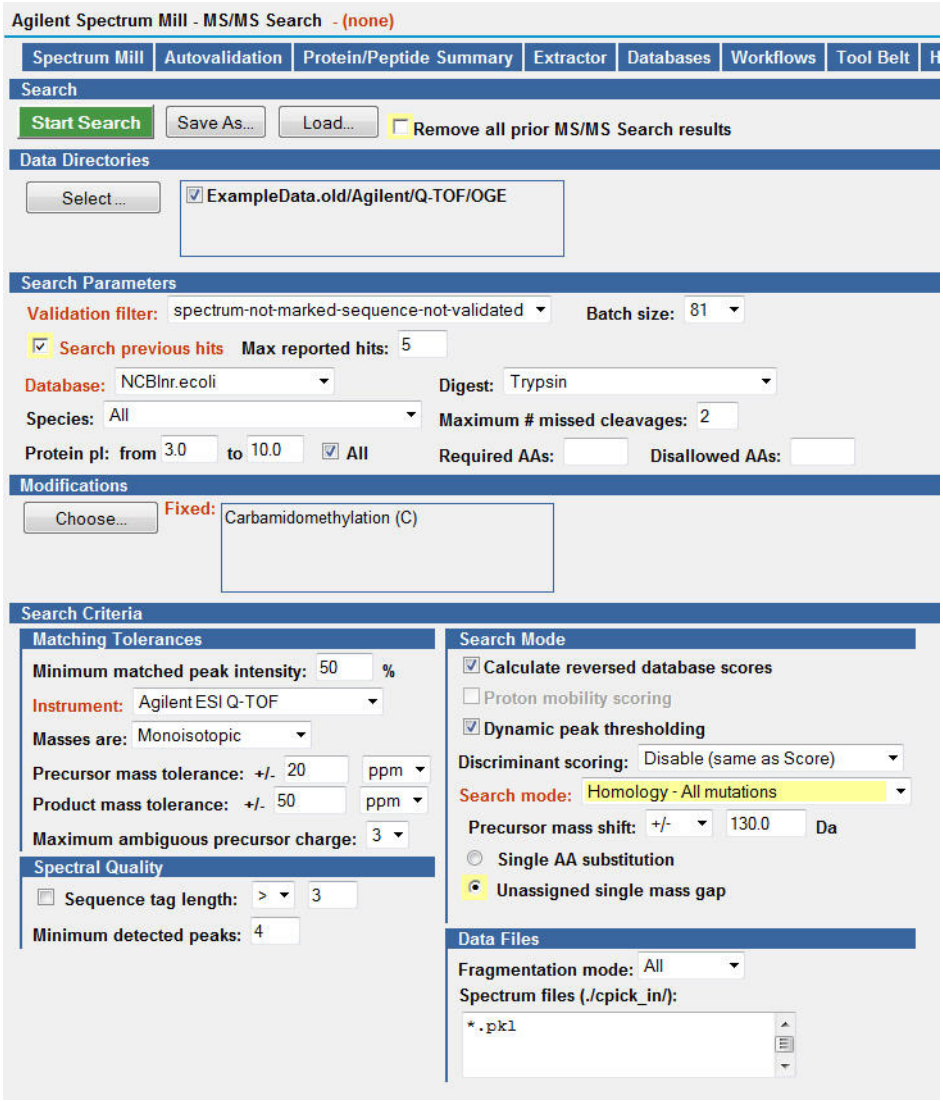


Figure 35 MS/MS Search settings for search of previous hits for unknown modifications

Exercise 3. Examine the list of validated peptides

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

Steps	Detailed Instructions	Comments
1 Autovalidate using the Auto thresholds - Discriminant strategy with the Peptide step. • Do not clear the previous autovalidation results.	• Follow the instructions for autovalidation in Chapter 2, "Exercise 1: Extract the data and do an identity search and validation" on page 50.	• 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.	• From the **Mode** list, select **Peptide**.	• 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).

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** before you continue.)
7 Filter peptides by changing these parameters to: ◦ Score > 0 ◦ % SPI > 0	**a** Under **Filter peptides by** for **Score, type 0**. **b** For **% SPI, type 0**.	
8 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 36 on page 90. Note that changes are highlighted in yellow.	• **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.
9 Summarize all results.	• Click **Summarize.**	

Steps	Detailed Instructions	Comments
10 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 37 on page 91 and Figure 38 on page 91. 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.

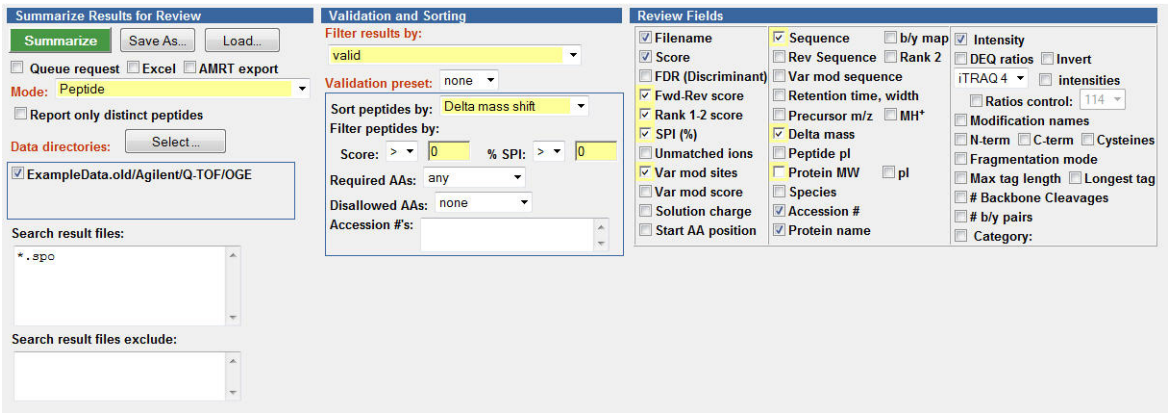


Figure 36 Settings to display results of mass gap search

#	Filename	z	Score	Discriminant	Fwd-Rev Score	Rank1-Rank2 Score	SPI (%)	Spectrum Intensity	Variable Sites	Sequence	MH+ Matched (Da)	MH+ Mass Shift (Da)	MH+ Error (ppm)	Accession #	Protein Nam
1	ec-oge-2.3070.3082.2	2	16.87	16.2562	9.53	11.92	94.6	3.06e+005	Unknown	(K) FADVACAGPLLAAELDALGK (A)	2002.031	57.0233	27693.9	16130827	phosphoglyce
2	ec-oge-2.1982.1989.2	2	15.65	12.3748	7.44	2.41	90.2	6.12e+005	Unknown	(K) DIALGEEFVNK (-)	1234.631	57.0205	44145.4	26249817	Malate dehyd
3	ec-oge-2.1276.1284.2	2	16.71	15.5746	8.12	12.82	83.3	1.70e+006	Unknown	(R) AFDQIDNAPEEK (A)	1376.633	57.0185	39771.5	15804570	protein chain
4	ec-oge-2.1862.1870.2	2	14.14	13.5272	11.14	9.40	69.7	5.27e+005	Unknown	(R) DWYVVDATGK (T)	1153.552	57.0171	47099.4	15803765	50S ribosoma
5	ec-oge-2.0968.0973.2	2	10.46	11.3107	10.46	10.46	50.6	1.07e+005	Unknown	(R) FQDEEVQR (D)	1050.485	57.0159	51481.5	15799694	chaperone Hs
6	ec-oge-2.2485.2492.2	2	15.82	12.2911	11.76	15.82	80.2	4.48e+005	Unknown	(R) ECTLETLEEMLEK (L)	1624.745	43.0085	25788.3	15801465	DNA-binding
7	ec-oge-2.3097.3104.2	2	22.28	20.8990	15.22	16.57	97.5	3.14e+005	Unknown	(R) EMLIADGIDPNELLNSLAAVK (S)	2226.169	43.0057	18952.1	15801465	DNA-binding
8	ec-oge-2.2670.2677.2	2	17.75	11.9755	13.40	12.41	89.6	3.92e+005	Unknown	(K) AFQELNAIDVL (-)	1232.652	21.9824	17521.0	15803459	fructose-bisph
9	ec-oge-2.3057.3065.2	2	18.17	17.3282	11.40	13.83	88.7	3.16e+005	Unknown	(R) EMLIADGIDPNELLNSLAAVK (S)	2226.169	15.9989	7135.5	15801465	DNA-binding
10	ec-oge-2.1317.1323.2	2	18.00	12.8835	13.99	18.00	91.8	5.38e+005	Unknown	(R) AFDQIDNAPEEK (A)	1376.633	14.0115	10075.5	15804570	protein chain
11	ec-oge-2.1360.1403.2	2	19.91	17.9973	13.39	14.62	90.1	1.58e+007	Unknown	(K) ALEGDAEWEAK (I)	1218.564	1.0014	821.1	15804570	protein chain
12	ec-oge-2.2530.2538.2	2	14.02	15.2955	10.10	14.02	76.6	8.60e+004	Unknown	(R) ESAEPEELVEPFLTR (F)	1745.859	0.9999	572.4	15831113	RNase II, mR
13	ec-oge-2.2470.2474.2	2	11.94	13.7685	11.94	11.94	55.8	3.38e+004	Unknown	(K) ETIDEGLWDTLSAR (L)	1605.776	0.9988	621.6	15802265	glucose-6-pho
14	ec-oge-2.1340.1340.2	2	12.20	11.9228	5.37	6.18	78.5	3.48e+005	Unknown	(K) EDEIVIDR (K)	988.495	0.9896	1000.1	16130342	PEP-protein p
15	ec-oge-2.1424.1432.2	2	15.19	11.9396	15.19	11.34	82.6	3.85e+005	Unknown	(R) VYDALEVQNGNER (L)	1506.718	0.9848	653.2	15804332	membrane-bo
16	ec-oge-2.2820.2829.2	2	19.13	14.3261	12.97	12.32	87.5	1.51e+006	Unknown	(R) LPSDWLLSAGLINGR (N)	1611.885	0.9832	609.6	15804421	tetrahydropte
17	ec-oge-2.1909.1918.2	2	18.47	16.2609	9.75	14.97	90.2	5.30e+005	Unknown	(K) IIAADNGDAWVEVK (G)	1500.769	0.9818	653.8	15799694	chaperone Hs
18	ec-oge-2.0921.0928.2	2	14.86	11.7735	5.51	10.90	75.9	2.30e+005	Unknown	(K) FTDVNGETK (T)	1010.479	0.9800	968.9	15803187	DNA-binding
19	ec-oge-2.1349.1356.2	2	13.57	13.2426	5.81	13.57	80.3	9.93e+005	Unknown	(R) NGAEGVGLYR (T)	1035.522	0.9797	945.2	16130342	PEP-protein p
20	ec-oge-2.2578.2578.2	2	9.51	11.8985	9.51	9.51	53.8	2.24e+005		(K) SLSDTLEEVLSSSGEK (S)	1680.817	0.0257	15.3	15803639	orf, hypothetic
21	ec-oge-2.2249.2258.2	2	11.17	13.4635	11.17	11.17	88.1	1.82e+005		(R) DSGVVSELFDER (N)	1352.633	0.0171	12.6	16129658	phosphoenolp

Figure 37 Mass gap search results, top

369	ec-oge-2.2718.2718.2	2	6.12	9.5251	6.12	6.12	58.8	1.95e+004		(R) SLDVLEGYLVDGTLK (T)	1621.868	-0.0268	-16.5	16131463	putative S-tra
370	ec-oge-2.2394.2402.2	2	18.83	17.4394	9.94	10.19	84.6	5.46e+005	Unknown	(R) IEPSLEAAFVDYGAER (H)	1766.860	-0.9604	-543.9	3183563	Ribonuclease
371	ec-oge-2.2574.2574.2	2	11.93	11.5818	8.00	1.70	89.9	6.42e+005	Unknown	(K) ELLEIEGLDEPTVEALR (E)	1926.007	-0.9817	-510.0	42144	NusA protein
372	ec-oge-2.1874.1881.2	2	15.68	15.0821	9.10	12.14	85.2	3.18e+006	Unknown	(R) ELDAQPTGFLDSR (I)	1448.702	-0.9891	-683.2	15831113	RNase II, mR
373	ec-oge-2.2853.2853.2	2	11.16	13.1690	3.68	11.16	71.2	1.51e+006	Unknown	(K) ESVEAILGETLDLPK (E)	1613.863	-0.9947	-616.7	15799684	threonine syr
374	ec-oge-2.2490.2497.2	2	13.84	14.5697	8.21	10.78	82.7	1.76e+005	Unknown	(R) DVNVFCAPYDLVK (T)	1539.751	-3.9997	-2604.4	4972324	UDP-N-acety
375	ec-oge-2.0881.0881.2	2	13.18	14.0984	8.36	13.18	84.3	2.51e+004	Unknown	(R) IPDGQAEDAR (K)	1071.507	-5.0378	-4723.8	15801728	aldehyde deh
376	ec-oge-2.2357.2357.2	2	13.92	14.3127	3.72	13.92	100.0	4.76e+005	Unknown	(R) CTQELLFGK (G)	1095.550	-17.0268	-15787.1	26246955	Aspartate am
377	ec-oge-2.2182.2194.2	2	19.71	19.0818	19.71	19.71	82.6	6.31e+005	Unknown	(R) QIVLNCGEEPSVVANTVK (G)	1957.006	-17.0270	-8776.9	15834378	chaperonin G
378	ec-oge-2.3133.3162.2	2	18.88	14.7520	8.54	12.14	94.6	3.41e+005	Unknown	(R) QSLGGLIEAYEAVAR (R)	1576.833	-17.0278	-10916.6	15802999	phosphoribos
379	ec-oge-2.2341.2350.2	2	16.41	14.7270	9.67	11.49	79.5	2.78e+005	Unknown	(R) CVEKDIDVAAESLK (V)	1576.789	-17.0300	-10918.3	15799758	2-isopropylm
380	ec-oge-2.1339.1347.2	2	19.74	18.8113	19.74	19.74	86.5	6.75e+005	Unknown	(R) QQIEEATSDYDREK (L)	1711.777	-17.0302	-10048.8	15834378	chaperonin G
381	ec-oge-2.0977.0980.2	2	18.09	12.7870	7.37	6.54	93.4	1.29e+005	Unknown	(K) QAQQELR (D)	872.458	-17.0305	-19908.8	15800768	3-phosphose
382	ec-oge-2.2174.2174.2	2	18.87	15.6533	13.59	13.88	84.6	3.82e+005	Unknown	(R) QLDPLVVGQEHYDTAR (G)	1840.919	-17.0305	-9337.5	15804332	membrane-bo
383	ec-oge-2.1256.1263.2	2	12.00	13.0615	12.00	8.50	100.0	1.15e+006	Unknown	(K) QQYGEGLAR (A)	1021.506	-17.0308	-16954.9	16131332	high-affinity a
384	ec-oge-2.0847.0854.2	2	15.47	11.8553	6.81	8.24	87.0	7.59e+004	Unknown	(R) QAEKEGQEPDPR (V)	1383.650	-17.0316	-12462.6	15804552	phosphoenol
385	ec-oge-2.2310.2310.2	2	14.70	11.9041	14.70	7.49	86.0	2.87e+005	Unknown	(R) QQFAQVTNPPIDPLR (E)	1723.913	-17.0597	-9994.8	15833345	glutamate sy
386	ec-oge-2.3050.3056.2	2	9.56	11.4518	2.28	9.56	73.9	2.31e+005	Unknown	(R) IFISTKGNEVTTMRVLK (F)	1937.089	-41.9877	-22155.9	26245920	Aspartokinas
387	ec-oge-2.1902.1905.2	2	12.97	11.8231	8.63	12.97	79.9	1.87e+005	Unknown	(-) MSEQHAQGADAVVDLNNELK (T)	2169.024	-89.0289	-42802.5	15803426	lysine tRNA
388	ec-oge-2.2729.2746.2	2	12.15	11.3214	5.28	4.53	60.7	2.37e+005	Unknown	(K) LSEDAFDDQCTGANPR (Y)	1795.755	-109.9082	-65194.7	15830995	CoA-linked a

Figure 38 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.

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 88.	
2 Display MS/MS Search results.	**a** Under the **Filename** header, click the third entry. **b** Check the bottom of the browser window, where you see the annotated spectrum, as shown in Figure 39 on page 93. **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 40 on page 94.	
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.

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 40 on page 94 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 39. **b** See the information about manual review and validation in Chapter 1 of the *Application Guide*.	

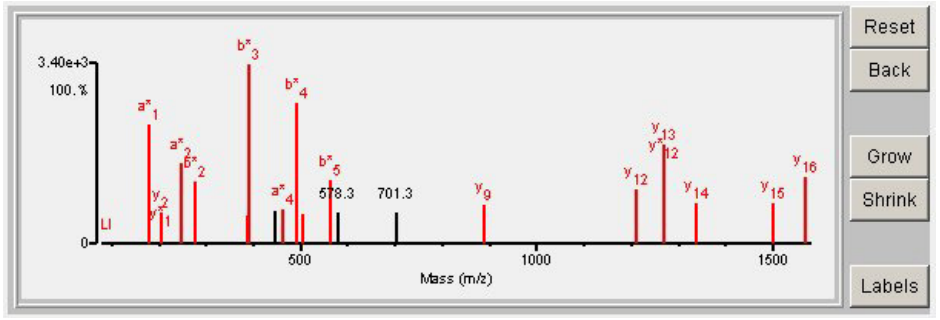


Figure 39 Spectrum from MS/MS Search results, showing star-ions

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	16.87	94.6	9	2/19	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	57.0233	27693.9	41288.6/5.08	Escherichia coli K12	16130827	10014	phosphoglycerate kinase
1	16.87	94.6	9	2/19	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	57.0233	27693.9	41300.7/5.08	Escherichia coli O157:H7 EDL933	15803460	11601	phosphoglycerate kinase

Fragment-ion (m/z)	86.100	177.099	204.127	248.134	276.133	388.254	391.154	462.239	490.225	503.292	561.255	578.301	701.336	887.486	1210.679	1267.722	1338.751
Frac. Inten.(% of TIC)	0.11	10.12	2.65	6.91	5.24	2.38	15.31	2.94	12.03	2.53	5.45	2.72	2.65	3.33	4.59	8.45	3.44
Rel. Inten.(% of BP)	0.74	66.10	17.29	45.13	34.22	15.57	100.00	19.19	78.58	16.51	35.62	17.77	17.31	21.76	29.95	55.17	22.46
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
Score	0.22	0.50	1.50	0.50	0.50	1.50	0.50	0.50	0.50	1.50	0.50	-0.18	-0.17	1.50	1.50	1.50	1.50
Ion type	LI	a^*_1	y_2	a^*_2	b^*_2	y_4	b^*_3	a^*_4	b^*_4	y_5	b^*_5			y_9	y_{12}	y_{13}	y_{14}
Delta ppm	35.7	-1999955.8	-37.6	-1999960.6	-1999983.9	-4.0	-1999971.6	-2000005.0	-1999985.5	19.9	-1999972.6			2.7	-21.0	-2.7	-9.1
			y^*_1 -1999935.2													y^*_{12} -1999995.6	

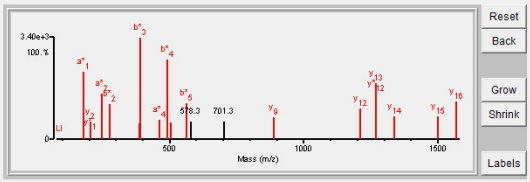


Figure 40 MS/MS Search results, showing the spectrum, table of fragment-ion assignments, and sequence

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

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, as does Sherenga *de novo* sequencing.

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.

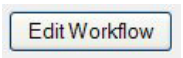
Parameter files are stored on the Spectrum Mill server under **\SpectrumMill\millauto\your own folder**. Only one level of subfolder is allowed. We recommend that you assign subfolders to users or types of experiments or studies. Each task type stores (in a cookie) the last parameter file that you selected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

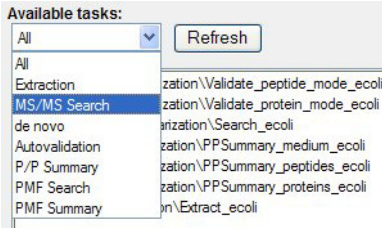
Steps	Detailed Instructions	Comments
1 Create these parameter files, if you have not already done so: • extract-y1 • MSMS_Search • AV-auto-pep • AV-auto-protpol • prot-sum	• Create parameter files as described in Chapter 1, "Basic Workflow – Interactive Processing of MS/MS Data". • The specific exercises where you create parameter files are: • "Exercise 1. Extract the spectra" on page 10 • "Exercise 2. Search the database" on page 16 • "Exercise 3. Autovalidate high-quality results" on page 22 • "Produce a Protein Summary Report" on page 27	• To create your own folder for the parameter files for these exercises, see Chapter 1, "Exercise 1. Extract the spectra", step 5 on page 12.

Steps	Detailed Instructions	Comments
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: • **extract-y1** • **MSMS_Search** • **AV-auto-pep** • **AV-auto-protpol** • **prot-sum**	

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.

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 96).

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 41 on page 100)	
8 Close the Edit Workflow window.	**a** Click **Close** in the upper right corner. 	

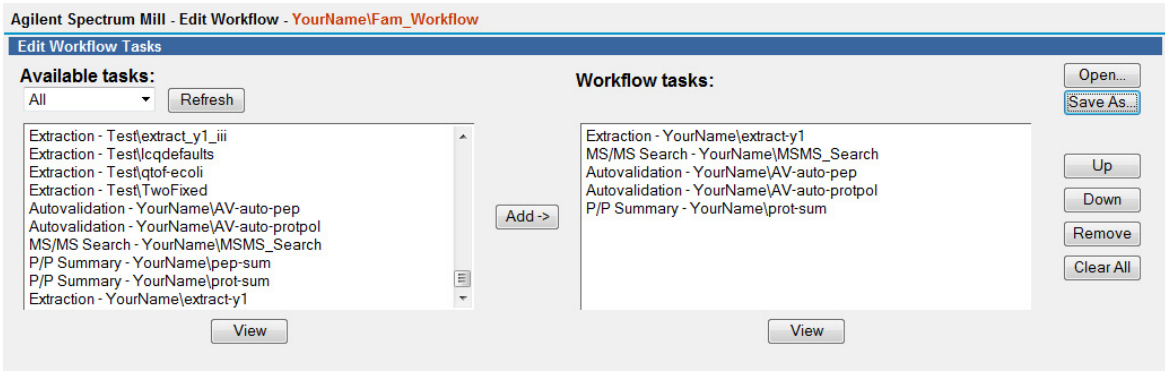
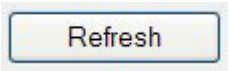
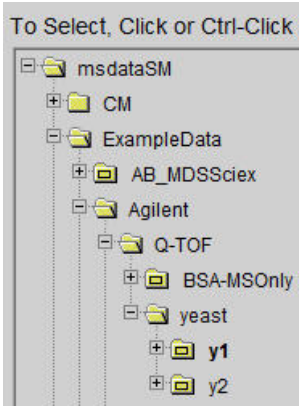
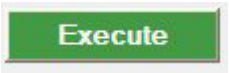


Figure 41 Edit Workflow window after new workflow saved

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.

Steps	Detailed Instructions	Comments
1 Go back to the Workflows page.	• This page should still be displayed. If not, display it again.	• See Figure 42 on page 102.
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.	
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 42 on page 102.	• 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.

Steps	Detailed Instructions	Comments
5 Execute the workflow.	**a** Click **Execute**..  **b** Briefly note the information that appears at the bottom of the window. A table lists the requests and the parameter files. Then each request appears in greater detail. (Figure 43 on page 103) **c** While the workflow executes, go to the next exercise.	

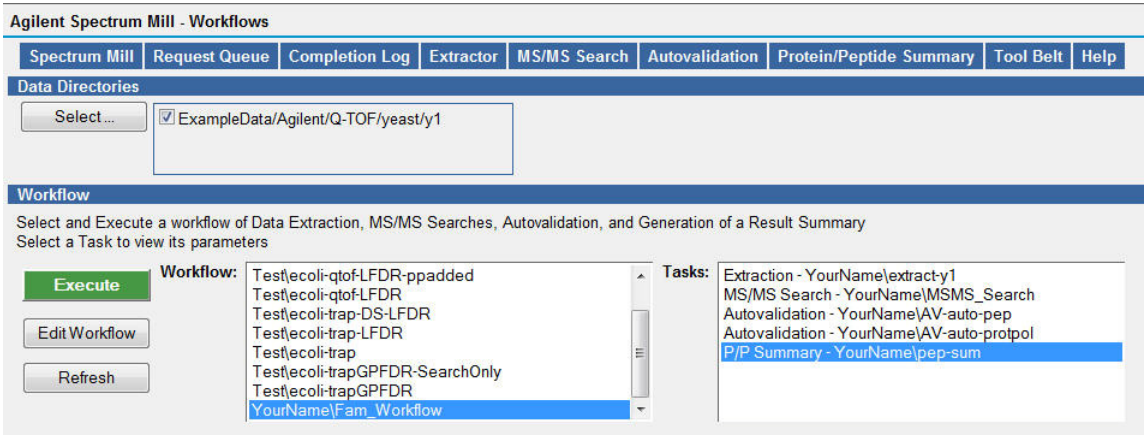
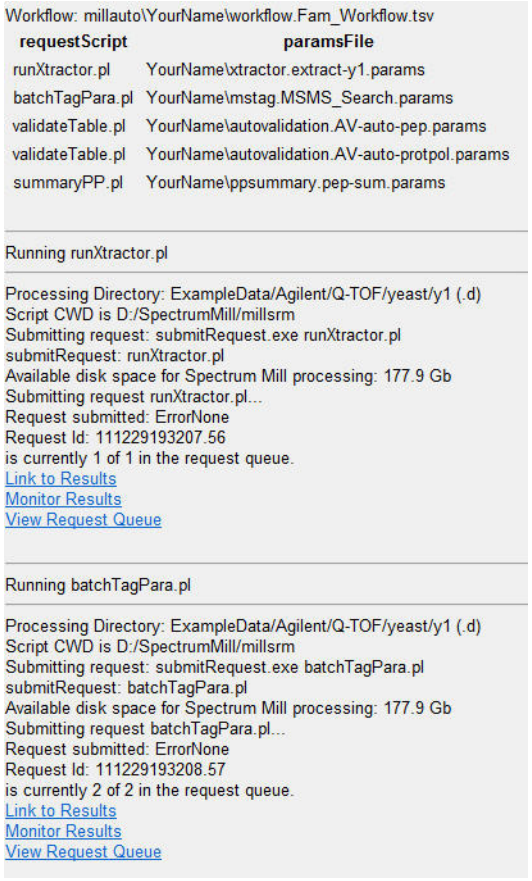


Figure 42 Workflows window


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.pep-sum.params

Running runXtractor.pl

Processing Directory: ExampleData/Agilent/Q-TOF/yeast/y1 (.d)
Script CWD is D:/SpectrumMill/millsrm
Submitting request: submitRequest.exe runXtractor.pl
submitRequest: runXtractor.pl
Available disk space for Spectrum Mill processing: 177.9 Gb
Submitting request runXtractor.pl...
Request submitted: ErrorNone
Request Id: 111229193207.56
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)
Script CWD is D:/SpectrumMill/millsrm
Submitting request: submitRequest.exe batchTagPara.pl
submitRequest: batchTagPara.pl
Available disk space for Spectrum Mill processing: 177.9 Gb
Submitting request batchTagPara.pl...
Request submitted: ErrorNone
Request Id: 111229193208.57
is currently 2 of 2 in the request queue.
Link to Results
Monitor Results
View Request Queue

Figure 43 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.

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 44 on page 105.	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**.

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 45 on page 106. **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 Report" on page 27. (Figure 11 on page 29)

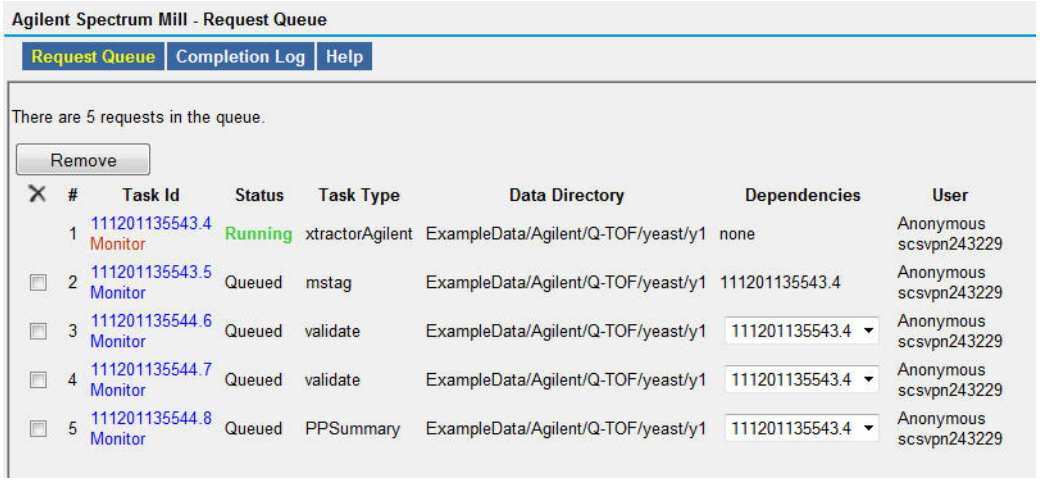


Figure 44 Request Queue Viewer

Task Id	Task Type	Data Set Parameters	Status	Completion Time	User	Client
111201135544.8	PPSummary	ExampleData/Agilent/Q-TOF/yeast/y1	NoError	Thu Dec 01 14:08:21 2011	Anonymous	scsvpn243229
111201135544.7	validate	ExampleData/Agilent/Q-TOF/yeast/y1	NoError	Thu Dec 01 14:07:49 2011	Anonymous	scsvpn243229
111201135544.6	validate	ExampleData/Agilent/Q-TOF/yeast/y1	NoError	Thu Dec 01 14:07:16 2011	Anonymous	scsvpn243229
111201135543.5	mstag	ExampleData/Agilent/Q-TOF/yeast/y1	NoError	Thu Dec 01 14:07:09 2011	Anonymous	scsvpn243229
111201135543.4	xtractorAgilent	ExampleData/Agilent/Q-TOF/yeast/y1	NoError	Thu Dec 01 14:02:08 2011	Anonymous	scsvpn243229
111130131012.3	mstag	ExampleData/Agilent/Q-TOF/yeast/y1	NoError	Wed Nov 30 16:54:32 2011	Anonymous	scsvpn243229

Figure 45 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 PP 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 PP Summary as Extraction or Search is running means those tasks automatically execute when appropriate.

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**.	

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 45 on page 106 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 Report" on page 27. (Figure 11 on page 29)
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.

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 (chr) 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 (chr)**. **c** Click **Use peaks with height >=,** type `100`, if necessary. **d** Click **Ion 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 46 on page 111. • If you do not see the complete table, click **Automatically show columns**.

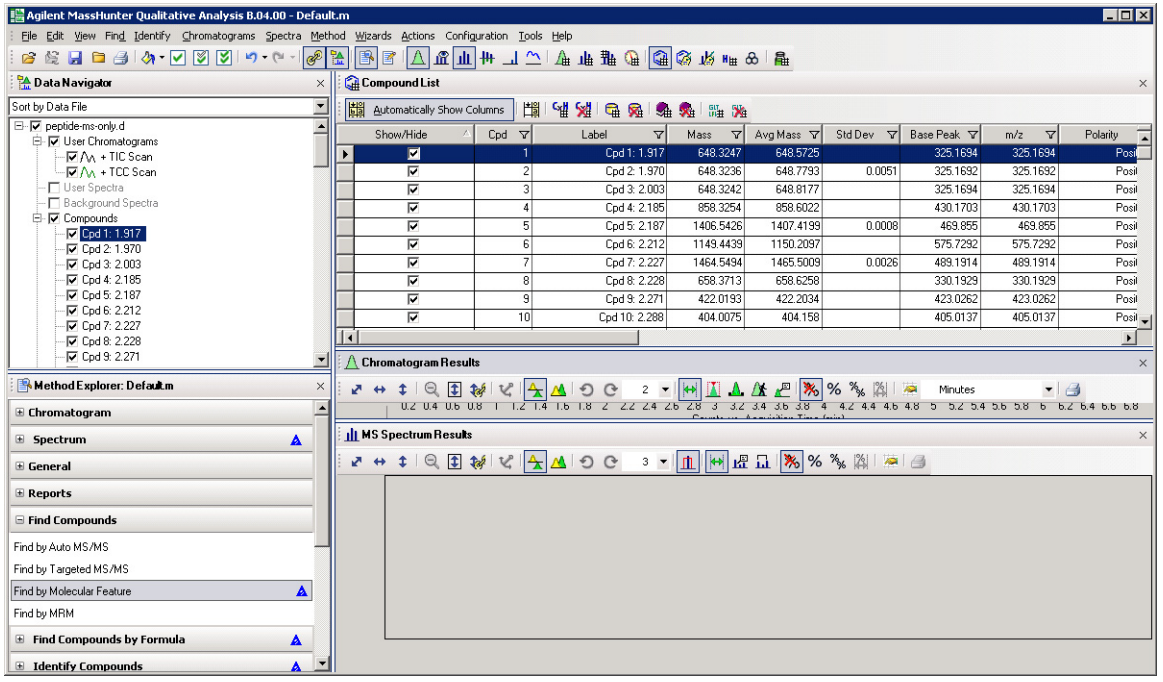


Figure 46 MassHunter compound list

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 PMF Search page.	**a** Open **Spectrum Mill Workbench**. **b** From the Spectrum Mill home page, click the link to **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 47 on page 114.	

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.

Steps	Detailed Instructions	Comments
4 Make sure the page looks like Figure 47 on page 114.	**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 SpectrumMill 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 48 on page 115.	• Manual PMF Search processes all the peaks in the list. • Search time varies depending on the size of the database searched.

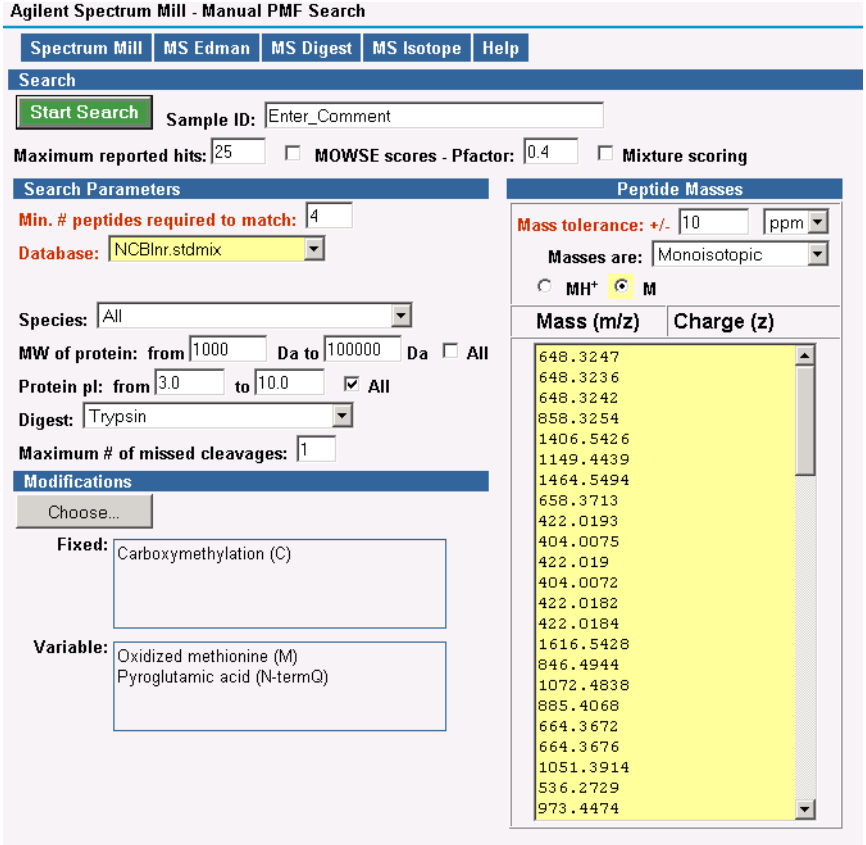


Figure 47 PMF Search settings for Q-TOF MS-only data

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 48 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.

Steps	Detailed Instructions	Comments
1 View the potential match(es).	• See potential match(es), as shown in Figure 48 on page 115. Again, your results may vary.	• The detailed results shown in Figure 48 list only one match for the data.
2 View a coverage map. • Coverage should be 59%. • See Figure 49 on page 117.	**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** In Internet Explorer (IE) 8.0 or 9.0 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.	

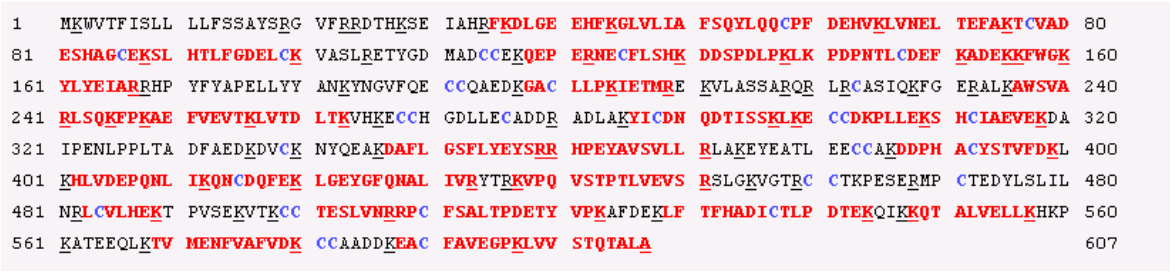


Figure 49 Coverage map for protein-MS-only.d

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:

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

Printed in USA
Revision A, January 2012

G2721-90040

Agilent Technologies