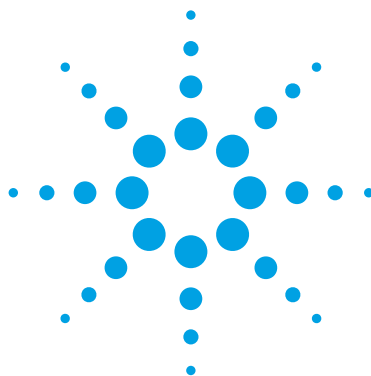




**Agilent
G2721AA/G2733AA
Spectrum Mill MS
Proteomics Workbench**

Application Guide



Agilent Technologies

Notices

© Agilent Technologies, Inc. 2016

No part of this manual may be reproduced in any form or by any means (including electronic storage and retrieval or translation into a foreign language) without prior agreement and written consent from Agilent Technologies, Inc. as governed by United States and international copyright laws.

Manual Part Number

G2721-90049

Edition

Revision B, October 2016

Agilent Technologies, Inc.
3501 Stevens Creek Blvd.
Santa Clara, CA USA 95052

Software Revision

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for the B.06.00 revision or higher of the Agilent G2721AA/G2733AA Spectrum Mill MS Proteomics Workbench software, until superseded.

Windows® is a U.S. registered trademark of Microsoft Corporation.

UNIX® is a registered trademark of the Open Group.

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 *Application Guide* presents instructions to analyze protein and peptide data with the Spectrum Mill MS Proteomics Workbench.

This guide assumes that the Spectrum Mill workbench has already been installed on your server, databases have been downloaded and indexed, and the server is ready to go (*Installation Guide*).

It also assumes you have set up your PC client, transferred files to the server and started the software (*Quick Start Guide*).

1 Processing MS/MS Data Interactively

Learn basic step-by-step procedures to process MS/MS data using a basic workflow and an iterative workflow.

2 Reviewing and Validating MS/MS Data

Learn details to help you customize MS/MS data review and validation and to do manual validation and review.

3 Automating Workflows

Learn how to set up and execute workflows containing sequential Spectrum Mill tasks (parameter files), e.g., extraction, search, autovalidation, quality metrics, summary, and data archive.

4 Sherenga *de novo* Sequencing

Learn details to help you understand and use the software module for Sherenga *de novo* sequencing.

5 Processing for Differential Expression Quantitation

Learn details to process quantitative data for differential profiling studies.

6 Processing MS-Only Data Interactively

Learn basic step-by-step procedures to process MS-only data.

7 Using the Tool Belt

Learn how to use the tools on the Tool Belt page.

8 Using Spectrum Mill Utilities

Learn how to use the many useful protein/peptide utilities included with the Spectrum Mill workbench.

9 System Administration

Learn common system administration tasks, including how to install and update databases.

10 Files Created during Spectrum Mill Data Processing

Learn about the files the Spectrum Mill workbench creates. This information is useful to troubleshoot data processing, to selectively remove some of the data processing, and to make decisions about data archives.

Content

1	Processing MS/MS Data Interactively	11
	Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively	13
	Step 1. Run the Data Extractor	13
	Step 2. Search in variable modifications mode	20
	Step 3. Validate high quality results automatically	25
	Step 4. Examine the validated results for variable modifications mode	30
	Iterative Workflow – Alternative Interactive Processing	34
	Overview	34
	Step 1. Do additional searches and autovalidations	35
	Step 2. Check for remaining high-quality spectra	37
	Step 3. Make a decision regarding further processing	41
	Step 4. Summarize valid results	41
	To search in no enzyme mode	47
	To search a larger database	47
	To search in homology modes, or do a mass gap search	48
	To use Sherenga <i>de novo</i> sequencing	50
	To review Sherenga <i>de novo</i> sequencing results	52
	To use the Spectrum Matcher	55
2	Reviewing and Validating MS/MS Data	59
	Setting Up Autovalidation	61
	Step 1. Select Data Directory and/or parameter file	61
	Step 2. Select the Strategy and Mode to autovalidate results	62
	Step 3. Save parameters and validate data	68
	Setting Up the Protein/Peptide Summary Page	69
	Step 1. Set data directory and mode	69
	Step 2. Set filtering, sorting, and validation parameters	70
	Step 3. Choose review fields	72

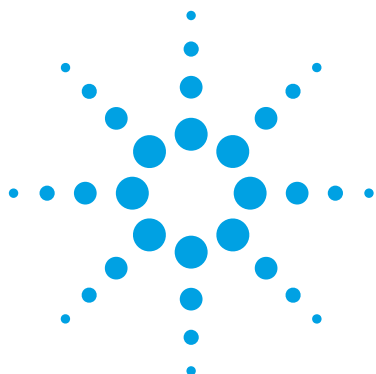
Step 4. Select Protein Quantitation Options	74
Step 5. Save parameters and summarize data	74
Reviewing and Validating Results Interactively	76
To display results in preparation for manual validation	76
To manually review and validate results	80
To use the Spectrum Viewer	83
Protein/Peptide Summary Displays	90
Peptide - Spectrum Match	91
Peptide - Distinct	92
Protein Summary Details	93
Protein-Protein Comparison	95
Protein-Peptide Comparison	98
Protein-Var Mod Site Comparison	99
Protein-Single Peptide ID	101
Protein-Peptide Distribution Columns	102
Identification Statistics and False Discovery Rate	103
To create a summary table of MS/MS identification statistics	103
To report FDR and search statistics	106
3 Automating Workflows	111
Setting Up Automated Workflows	112
To create parameter files or workflow tasks	112
To view already existing workflows and their tasks	112
To create/edit an automated workflow	114
Executing and Monitoring Workflows	116
To execute an automated workflow	116
To monitor results	117
To view workflow progress	119
To check the Completion Log	120

4	Sherenga <i>de novo</i> Sequencing	121
	To set parameters and run the <i>de novo</i> sequencing algorithm	122
	To generate a Sherenga report	125
	To view detailed Sherenga results	127
	To compare Sherenga results with MS/MS Search results	129
5	Processing Data for Differential Expression Quantitation	131
	To use the Data Extractor for a differential profiling study	133
	To use MS/MS Search for a differential profiling study	135
	To calculate DEQ ratios for isotopic labels using the Protein/Peptide Summary page	136
	To interpret DEQ results for isotopic labels in Peptide - Spectrum Match mode	137
	To interpret DEQ results for isotopic labels in protein modes	139
	To calculate iTRAQ and TMT ratios using the Protein/Peptide Summary page	140
	To interpret results for iTRAQ labels in peptide mode	141
	To interpret results for iTRAQ labels in protein mode	142
	To view light/heavy results on the Spectrum Summary page	143
6	Processing MS-Only Data Interactively	145
	Step 1. Find the compounds of interest in MassHunter Qualitative Analysis	146
	Step 2. Copy and paste compound masses into PMF Search	147
	Step 3. Search the MS-only data and display results	148
	Step 4. Examine the results	151
7	Using the Tool Belt	153
	To terminate a process	154
	To create a saved results file so you can search previous hits	156
	To create an MS/MS Search summary file if search terminated abnormally	158
	To create a summary table of previously-used parameters	159
	To list details about amino acid modifications	161

	To create a file of reporter ion correction factors	163
	To apply a file of reporter ion correction factors	165
	To calculate discriminant scoring coefficients	166
	To copy files to your collections directory	168
	To export PepXML files	169
	To convert spectra	171
8	Using Spectrum Mill Utilities	173
	To archive data	174
	To identify digest peptides likely to meet specific experimental goals (Peptide Selector)	176
	To build MRM lists for triple quadrupole instruments	179
	To align sequences (Multiple Sequence Aligner)	181
	To list peptides that correspond to a theoretical protein digest (MS Digest)	182
	To retrieve database entries using text searches (MS Edman)	184
	To list theoretical fragment ion masses for peptides (MS Product)	186
	To list amino acid compositions that fit precursor mass and partial composition (MS Comp)	188
	To show isotope patterns of peptides (MS Isotope)	190
	To search a database with a list of peptide sequences (Peptide String Match)	191
	To calculate the masses and formulas for peptides	193
9	System Administration	195
	Manipulating sequence databases	196
	To install or update databases	196
	To create database indices	198
	To re-index an existing database	200
	To create a species and protein molecular weight subset database	201
	To generate a database summary report	202
	To concatenate databases (FASTA files)	203
	To create a subset database from saved hits	204

To make a subset FASTA file from Accession Numbers	205
To create a user (proprietary) database or add sequences to one	206
To compare two databases	207
To calculate statistics	208
To remove redundant database entries	209
To create a category file	210
Other system administration tasks	211
To add custom amino acid modifications	211
To change the URLs of HTML links in the search results	212
To enable the HTML link to BLAST search	212
To add/change options related to biology/chemistry	213
To add instrument types	213
To configure IEF Selector defaults	214
To maintain server performance	215
To use server administration scripts	215
To avoid problems with connection time-outs	216
To configure server with drives other than the default configuration	216
To remove and reinstall the Spectrum Mill workbench	217
To limit the number of parallel workflow processes to less than the CPU count	218
To start and stop the Spectrum Mill Workflow Manager Service	219
10 Files Created during Spectrum Mill Data Processing	221
Architecture Overview	222
Data Extractor (MS/MS raw data)	223
Data Extractor (generic data)	225
Data Extractor (MS-only raw data)	227
MS/MS Search	228
Protein/Peptide Summary, Spectrum Summary, and Autovalidation	230
Tool Belt	232
Sherenga <i>de novo</i> Sequencing	234
PMF Search	235
MS Edman	236

Protein Databases	237
-------------------	-----



1

Processing MS/MS Data Interactively

Normal Workflow – Setting Up Parameter Files and Processing Tasks
Interactively 13

Iterative Workflow – Alternative Interactive Processing 34

This chapter gives more details for the steps in normal and iterative MS/MS workflows (Figure 1) and shows you how to set up parameter files for automating workflows, how to automate tasks interactively and how to execute some single tasks immediately.

The normal workflow is a one-pass search and validation workflow that lets you identify most spectra in your sample. If you want to gather more information to characterize identified proteins, use an iterative workflow.

First read the *Quick Start Guide* to learn this basic information:

- Features new to the current version of software
- How to find other information for using Spectrum Mill successfully
- How to transfer spectral files to the server
- How to configure the PC client and start the software

Both this chapter and the *Quick Start Guide* assume that the Spectrum Mill workbench has already been installed on your server and databases have been downloaded and indexed. If this has not yet been done, see the *Installation Guide*.

- Descriptions of the Spectrum Mill workflows that Agilent recommends for analyzing protein/peptide data
- How to process data automatically and interactively



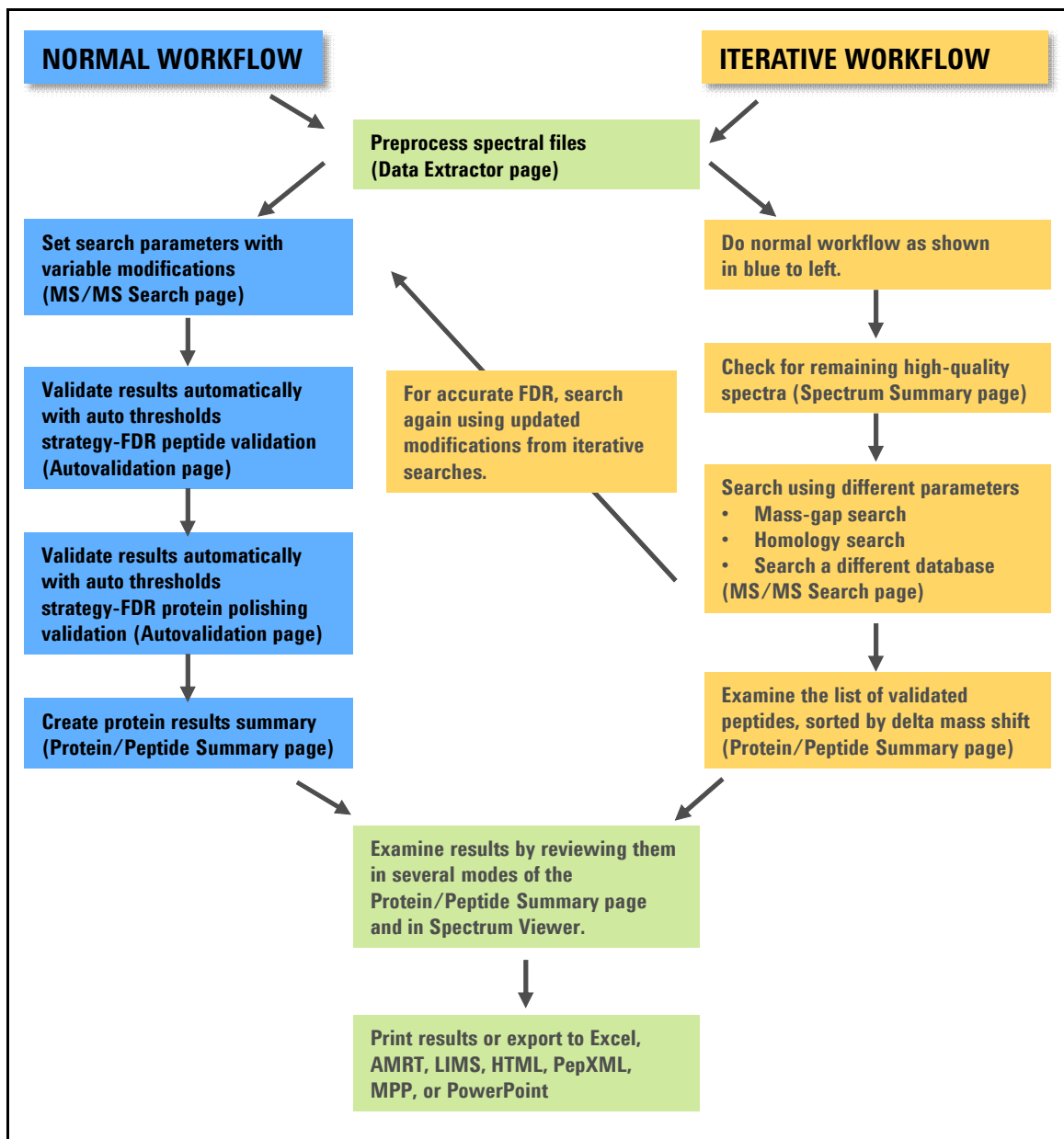


Figure 1 Summary of processing for MS/MS data

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

This section gives you the instructions for setting up the parameter files needed to create automated normal workflows and to run the workflow tasks interactively. See Chapter 3 for instructions on creating automated workflows from the parameter files you set up here.

When started interactively, extraction and search tasks are automatically submitted to the workflow manager, or queue, to be executed. You have the option to select a Queue request for Autovalidation, Quality Metrics, Protein/Peptide Summary, and Archive Data, but Autovalidation cannot start (Queue request unmarked) before searching is complete. While extraction and searching are being executed, start Autovalidation and Protein/Peptide Summary after marking the Request queue check boxes. This is called *interactive automation*.

Step 1. Run the Data Extractor

The Spectrum Mill Data Extractor software has two fundamental types of MS/MS data extractors: raw and generic.

The Spectrum Mill workbench recognizes the type of MS/MS data file and chooses the appropriate data extractor. The Data Extractor page changes accordingly.

To preprocess raw data files (Agilent Q-TOF or ion trap or Thermo Fisher Scientific data)

The raw data extractors operate on *.d and *.raw files. They 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 data, and Thermo Fisher Scientific *.raw ion trap data, the extractors optionally merge MS² and MS³ scans from the same precursor. The extractors assign precursor charges where possible, centroid the MS/MS

1 Processing MS/MS Data Interactively

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

spectra, calculate spectral features, filter MS/MS spectra by quality, and calculate extracted ion chromatograms (EICs) for the intervening MS precursor scans. The latter are used for quantitation.

- 1 Navigate to the Data Extractor page.
- 2 In the **Data Directories** section, click the **Select...** button to select the folder or folders that contain your files.
- 3 For Agilent Q-TOF data and most extractions, fill in the boxes as shown in [Figure 2](#).
- 4 If you want the extraction to take advantage of all available processors, mark the check box for **Maximize CPUs**.
 - If you clear the check box, extraction uses only a single CPU per directory so that the other CPUs are available for other processes/users).
 - If your folder contains only one data file, clear the check box.
 - If you process multiple directories at the same time, (where “multiple” means more than the number of CPUs on the server), do *not* mark the check box for **Maximize CPUs**.
 - If you mark this check box for a workflow, the request queue shows two requests: the initial one to create the batch of files and a second one to show the progress and extractor results.

Agilent Spectrum Mill - Data Extractor - YourNameextract-y1

Spectrum Mill | **MS/MS Search** | **Workflows** | **Tool Belt** | **Help**

Extraction

Extract | Save As... | Load... | ☐ Maximize CPUs | ☒ Remove all prior results | ☐ Delete data files after extraction

Data Directories

Select ... | ☒ ExampleData/Agilent/Q-TOF/Yeast/Y1

Modifications

Choose... | Fixed: Carbamidomethylation (C)

MS/MS Spectral Feature Filtering

Precursor MH+ to Da

Scan time range: to min

Sequence tag length > (For MALDI: Set tag length to -1 and merge secs to total run time.)

Ignore spectra with dissociation mode: ☐ CID, ☐ ETD

Merge nearby MSⁿ scans with same precursor m/z

Retention time & m/z tolerance: +/- secs +/- m/z
(also used for calculating chromatographic peak area of precursor in MS scans)

General MS/MS Merging Constraints:

Merge MS² and MS³ spectra from same precursor:

Precursor m/z & Charge Assignment

Precursor Charge: Maximum (z): Minimum MS1 S/N:

☒ Find ¹²C precursor m/z

MS Noise threshold: counts (only applies to Agilent Q-TOF)

Figure 2 Data Extractor for .d files, shown with typical settings for Agilent Q-TOF

- 5** If you have a different data type, or want to change other parameters, choose from the optional steps below.

To find explanations of all the fields in a particular category, click the blue divider bar to bring up that category in the online help.

- Click the **Choose...** button to select the modifications that match the chemistry for your samples.
- 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.

1 Processing MS/MS Data Interactively

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

- If you are extracting SILAC or other isotope labeled data for differential analysis, select the corresponding modification with the “-mix” suffix. For more information about choosing modifications, click the blue bar labeled **Modifications** to access the online help.
- Note that your system administrator can add custom modifications.
- If you are extracting MALDI MS/MS data, change the **Sequence tag length** to -1.
Otherwise, leave the default value. The sequence tag length represents the longest sequence of amino acids that can be located in the spectrum. The default of >0 for non-MALDI data removes most of the noisy spectra without removing good data.
- If necessary, change the **Retention Time & m/z tolerance** time range to be compatible with your chromatographic data.
For MALDI MS/MS data, change the time range to 2000 sec, or to the spot analysis time.
- For complex samples, keep the default selection for **General MS/MS Merging Constraints**. See the online help for descriptions of the other options.
If you suspect a large amount of chemical background noise, you may want to increase the **MS Noise threshold**.
- Set any other parameters. In general, you should use default values, and change only the parameters highlighted in red text.

6 Click Save As to save the parameter file, enter a name and click Save.

You can now use this parameter file as a “task” in the list of tasks for an automated workflow. See Chapter 3, “Automating Workflows”.

7 Click Extract.

The extraction is submitted to the workflow manager to be executed.

8 Click the Monitor Results link to monitor the extractor results, or the View Request Queue link to view and monitor tasks.

```

Results

Monitor output file: msdataSM\ExampleData\Agilent\Q-TOF\Ye
\xttractorAgilent.150603180202.2.htm
Stop Extraction PID: 3856

Processing Directory ExampleData\Agilent\Q-TOF\Yeast\Y1
All prior extraction and search/summary files have been remov

Processing File: yeast-1.d using millbin\xttractorAgilent.cgi ...
Initialize extractor complete!
Model G6550A
Ion Source Type: HPLC-Chip/MS
Quad isolation width: Narrow (1.3 Da)
Nominal resolution: 26000
-----
Number of spectra: 25769
1000 scans read.
2000 scans read.
3000 scans read.
4000 scans read.
5000 scans read.
6000 scans read.
7000 scans read.
8000 scans read.
9000 scans read.
10000 scans read.
11000 scans read.

```

Figure 3 Display while the raw file Data Extractor runs

The exact display depends upon the type of file you extract. The Data Extractor extracts all files in the subdirectory. Extraction time varies depending on the number and size of the files. A message informs you when extraction is complete.

If, after extracting, you put other single tasks in the queue, one right after the other, such as MS/MS Search, Autovalidation and Protein/Peptide Summary, this process is called *interactive automation*.

To preprocess generic peak list files (e.g., Waters Micromass Q-ToF or MALDI-TOF-TOF or Agilent IM-Q-TOF data)

The generic data extractor processes peak list files, such as centroided peaks from Waters Micromass (Q-ToF) *.pkl and *.dta files, and most *.mgf files.

If the *.pkl file came from IM-MS Browser, the generic data extractor gets the retention time (RT), drift time (DT), and optional collision cross section (CCS) values and writes them to the mzXML file.

1 Processing MS/MS Data Interactively

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

Because this data extractor operates on exported peak list files rather than raw files, it does not have access to chromatographic time information and mass information contained in MS precursor scans. This limits quantitation capabilities. Like the raw data extractor, the generic data extractor filters MS/MS spectra by quality and calculates spectral features.

- 1 Navigate to the Data Extractor page, shown in [Figure 4](#).

The screenshot shows the 'Data Extractor for generic (peak list) files' window. At the top is a navigation bar with tabs: 'Spectrum Mill', 'MS/MS Search', 'Workflows', 'Tool Belt', and 'Help'. The 'Extraction' section contains buttons for 'Extract' (highlighted in green), 'Save As...', and 'Load...', along with checkboxes for 'Maximize CPUs', 'Remove all prior results' (checked), and 'Delete data files after extract'. Below these is a checkbox for 'Generate spectral features file only' and a dropdown menu for 'Instrument' set to 'ESI QTOF Generic'. The 'Data Directories' section has a 'Select...' button and a list box containing 'ExampleData/Generic' which is checked. The 'Modifications' section includes a 'Choose...' button and a 'Fixed:' dropdown set to 'Carbamidomethylation (C)'. The 'MS/MS Spectral Feature Filtering' section has input fields for 'Precursor MH+' (600.0) and 'Da' (6000.0), and a 'Sequence tag length' dropdown set to '0' with a note '(For MALDI: Set tag length to -1 and merge secs to total run time.)'.

Figure 4 Data Extractor for generic (peak list) files

- 2 In the **Data Directories** section, click the **Select...** button to select the folder or folders that contain your files. The Data Extractor page changes to match [Figure 4](#).
- 3 If you want the extraction to take advantage of all available processors, mark the check box for **Maximize CPUs**.
- 4 For concatenated *.pkl or *.mgf files, do *not* mark **Generate spectral features file only**. For *.pkl files of individual spectra that have been placed in a **cpick_in** subfolder, mark the check box.

This option appears when you select a directory that contains peak list files but no raw data file. When you have *.dta files, or *.pkl files that represent individual spectra, you put your files in the **cpick_in** folder, and then you must mark this check box. (When you have *.mgf files or appended *.pkl files, where each file contains multiple spectra, then you put your file in the root sample directory and you do not mark the check box.)

- 5 Select the instrument type you used for the analyses.

This option only appears when you select a data directory with generic peak list files.

6 For most extractions, fill in the boxes as shown in [Figure 4](#).

7 If you want to change other parameters, choose from the optional steps below.

To find explanations of all the fields in a particular category, click the blue divider bar to bring up that category in the online help.

- Click the **Choose...** button to select the modifications that match the chemistry for your samples.
 - 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, click the blue bar labeled **Modifications** to access the online help.
 - Note that your system administrator can add custom modifications.
- If you are extracting MALDI MS/MS data, change the **Sequence tag length** to -1.

Otherwise, leave the default value. The sequence tag length represents the longest sequence of amino acids that can be located in the spectrum. The default of >1 for non-MALDI data removes most of the noisy spectra without removing good data.

- Set the other parameters. In general, you should use default values, and change only the parameters highlighted in red text.

8 Click **Save As** to save the parameter file, enter a name and click **Save**.

You can now use this parameter file as a “task” in the list of tasks for an automated workflow. See Chapter 3, “Automating Workflows”.

9 Click **Extract**.

The extraction is submitted to the workflow manager to be executed.

10 Click on the **Monitor Results** link to monitor the extractor results, or the **View Request Queue** link to view and monitor tasks.

1 Processing MS/MS Data Interactively

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

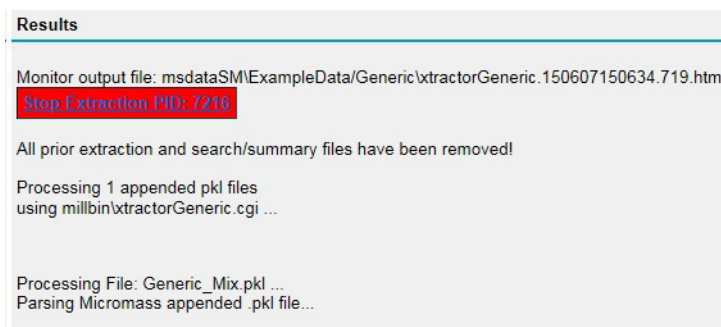


Figure 5 Display while monitoring the results for a generic (peak list) extraction

The Data Extractor extracts all files in the subdirectory. Extraction time varies depending on the number and size of the files. When monitoring results, a message informs you when extraction is complete.

If, after extracting, you put other single tasks in the queue, one right after the other, such as MS/MS Search, Autovalidation and Protein/Peptide Summary, this process is called *interactive automation*.

Step 2. Search in variable modifications mode

NOTE

If you have not yet created a database, see “[Manipulating sequence databases](#)” on page 196.

- 1 Navigate to the MS/MS Search page, as shown in [Figure 6](#).
- 2 Click the **Select...** button to choose the **Data Directories** where your files reside.
- 3 To load settings from an existing parameter file, click **Load**.
- 4 If you want the search to take advantage of all available processors, mark the check box for **Maximize CPUs**.

If you search multiple directories (where the number is greater than the number of CPUs), it is best to *not* mark **Maximize CPUs**. Then the processes will appear more orderly in the Workflow Request Queue. In this case, the directories will be processed in parallel, up to the number of configured CPUs.

If you have only a single data file, **Maximize CPUs** is an effective tool for search (even though it is not effective for data extraction in this situation).

5 Set the **Validation filter** to **spectrum-not-marked-sequence-not-validated**.

6 Select a **Database**.

If you want to search a species-specific subset database, select the database name for a subset database you previously created.

If you have not yet created a species subset database, see [“To create a species and protein molecular weight subset database”](#) on page 201.

7 Click the **Choose...** button to select **Fixed** or **Mix Modifications**.

8 Add any **Variable Modifications** that you wish to search, and click **OK**.

- Note that when you select any variable modifications, the Search mode changes to **Variable modifications**.
- 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, click the blue bar labeled **Modifications** to access the online help.
- Note that your system administrator can add custom modifications.

1 Processing MS/MS Data Interactively

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

NOTE

If you are not sure what variable modifications to choose, do a mass gap search to find the unknown modifications. See “[To search in homology modes, or do a mass gap search](#)” on page 48.

In a normal workflow you combine multiple variable modifications into a single search, but as the number of permutations increases, the search time increases and the number of hits decreases (as a result of FDR-based validation to limit the number of false positives). So the best strategy for a normal workflow is to use the variable modifications you think are present. Remember that the goal of a normal workflow is to obtain as much information as possible in a single round of processing.

You can add the following common modifications into the same search: oxidized methionine (methionine sulfoxide) and pyroglutamic acid.

If you suspect phosphorylation, add the following modifications into the same search: phosphorylated S, phosphorylated T, and phosphorylated Y.

Note that you can search only a single type of variable modification for a given amino acid. For example, you cannot select both guanidination and carbamylation of lysine as variable modifications within the same search. Metabolic modifications such as SILAC and ¹⁵N-mix are the exception, where you can select an additional fixed or variable modification that applies to the same amino acid. For example, if the SILAC modifies K, you can still select carbamylated lysine.

After you have determined all the necessary modifications, you then remove prior search results and search again with all needed modifications. This new search ensures that false discovery rate (FDR) calculations are accurate.

9 From the **Instrument** list, select the correct instrument.

10 To help limit false positives and allow FDR calculations when you review the data, verify that the check box for **Calculate reversed database scores** is marked.

11 If you have ion trap data, decide whether to use **Proton mobility scoring**.

- In most situations, you mark this check box because this scoring refinement for ion trap data provides better results.
- But do not mark the check box if you have peptides that are modified by iTRAQ, lysine mass tagging, guanidination or phosphorylation, because these modifications change the expected fragmentation pattern.

12 To enable identification of short peptides (< 5 amino acids), mark the check box for **Dynamic peak thresholding**, which is a scoring enhancement that

enables identification of more low-abundance and short-chain peptides. (For most experiments, you clear this check box.)

13 To search the variable modification(s) that you chose in [step 8](#) set the **Precursor mass shift range** to a value that is appropriate for the types and numbers of modification(s) that you are searching. Larger values may produce longer searches.

14 Under **Data Files**, select the **Fragmentation** mode to match the one within the instrument settings.

For CID-only instruments, be sure to select **All**.

15 Set the other parameters. See [Figure 6](#).

In general, you should keep the defaults except for the settings that are highlighted in red. Click the blue divider bars for more information in the online help.

1 Processing MS/MS Data Interactively

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

Agilent Spectrum Mill - MS/MS Search - (none)

Spectrum Mill	Autovalidation	Quality Metrics & FDR	Protein/Peptide Summary	Extractor	Databases	Workflows	Tool Belt	Help
---------------	----------------	-----------------------	-------------------------	-----------	-----------	-----------	-----------	------

Search

Start Search Save As... Load... ☒ Remove all prior MS/MS Search results ☒ Maximize CPUs

Data Directories

Select ... ☒ ExampleData/Agilent/Q-TOF/Yeast/Y1

Search Parameters

Validation filter: spectrum-not-marked-sequence-not-validated Batch size: 500 ☐ Search previous hits Max reported hits: 5

Database: SwissProt.yeast Digest: Trypsin

Species: All Maximum # missed cleavages: 2

Modifications

Choose... Fixed: Carbamidomethylation (C) Variable: Oxidized methionine (M)
Pyroglutamic acid (N-termQ)
Phosphorylated S (S)
Phosphorylated T (T)
Phosphorylated Y (Y)

Search Criteria

Matching Tolerances Minimum matched peak intensity: 30 % Instrument: Agilent ESI Q-TOF Masses are: Monoisotopic Precursor mass tolerance: +/- 20 ppm Product mass tolerance: +/- 50 ppm Maximum ambiguous precursor charge: 3 Spectral Quality Filtering ☐ Sequence tag length: > 3 ☐ Precursor isotope quality XIC's (Chi2 vs. Average): > 0.7 ☐ Precursor Isolation Purity: > 50 %	Search Mode ☒ Calculate reversed database scores ☒ Dynamic peak thresholding Discriminant scoring: Off Search mode: Variable modifications Precursor mass shift range: -18.0 to 177.0 Da Data Files Fragmentation mode: All Spectrum files (/cpick_in/): *.pkl
---	---

Figure 6 MS/MS Search page, set up for variable modifications mode search

16 Click **Save As** to save the settings to the existing parameter file or a new one, enter a name and click **Save**.

You can now use this parameter file as a “task” in the list of tasks for an automated workflow. See Chapter 3, “Automating Workflows”.

17 Click **Start Search**.

The search is submitted to the workflow manager to be executed.

- 18** Click on the **Monitor Results** link to monitor the search results, or the **View Request Queue** link to view and monitor tasks.

If the **Maximize CPUs** check box is marked, the first “monitor” displays the creation of the batch files for searches. At the bottom of those results, you can click links to monitor the search results.

- 19** When the search is complete, autovalidate high-quality results in the next step.

Step 3. Validate high quality results automatically

Validation means you accept that the database match is good. The Spectrum Mill workbench provides two ways to validate MS/MS search results. One way uses the Autovalidation page, and is totally automated. You use this method to validate high quality results. The other method uses the Protein/Peptide Summary page for manual review and validation. See “[Reviewing and Validating Results Interactively](#)” on page 76.

- 1** Navigate to the **MS/MS Autovalidation** page, as shown in [Figure 7](#).

You navigate to this page from the Spectrum Mill Home page, the MS/MS Search page or the Protein/Peptide Summary page.

When the page first appears, you see the Auto threshold strategy and the Peptide mode.

1 Processing MS/MS Data Interactively

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

Help

Validate Files ☐ Queue request **Undo Last** **Clear All** **Save As...** **Load...**

Data Directories

Select ...

☒ ExampleData/Agilent/Q-TOF/Yeast/Y1

Fragmentation mode: All

Search result files:

*.spo

Validation Parameters

Strategy: ☐ Fixed thresholds ☒ Auto thresholds ☐ Auto thresholds - discriminant

Mode: Peptide **Auto determine using score, delta R1-R2 thresholds to reach a target FDR**

Optimize score & R1-R2 score thresholds with max FDR: 1.2 % across each: ☒ LC run ☐ Directory

Precursor charge range: 2 to 6 Min Sequence Length: 6

Required AAs: any Disallowed AAs: none

Filtering	Automatic variable range for each run	Fixed range for all runs
☒ None (ppm)	☐ Auto precursor mass error	☐ Fixed precursor mass error Low -1.0 High 30.0 ppm
☒ None (SC/pl)	☐ Auto SCX Solution Charge, pH3	☐ Fixed Solution Charge Low -2 High 6
	☐ Auto OGE/IEF peptide pl	☐ Fixed peptide pl Low 3.0 High 10.0

Figure 7 Autovalidation of MS/MS Search results – Auto thresholds strategy/Peptide mode

- 2 Select the **Data Directories** for which you want to validate results.
- 3 If you want to use the settings from an existing parameter file, click **Load**.
If you are not sure of a field's meaning, click the category divider bar for the section containing the field to see the online help explanation for the field.
- 4 Choose the **Strategy** for autovalidation and the **Mode** that suits your study best. (See [Table 1.](#))

Table 1 Choose the Autovalidation Strategy and Mode

Select this Strategy:	Select this Mode:	What happens:	Pay attention to this:
Auto thresholds	Peptide	- Automatically optimizes the score and R1-R2 score thresholds until they reach a specified target FDR (false discovery rate) - See the online help to learn more about the false discovery rate (FDR)	- Use this mode before the Protein Polishing mode. - You may want to review the results before autovalidation again in the Protein polishing mode. 1.2% is set as the Target FDR instead of 1% because, in practice, a target FDR of 1.2% results in an actual FDR of 1%.
	Protein polishing (optional)	- Part 1: Minimum protein score and "... raise minimum protein score to yield maximum protein FDR %" - Controls the maximum allowed protein FDR. - Part 2: Peptide FDR for validated proteins % - Recalculates the peptide FDR based on validated proteins only. This recalculation usually increases the number of validated peptides, thereby increasing the sequence coverage of the validated proteins but leaving their number unchanged.	Part 1 is accomplished by "invalidating" peptides from the lowest-scoring proteins. Part 1 enables you to be aggressive with peptide level FDR thresholds in peptide mode and then remove suspicious proteins with Protein polishing.
Auto thresholds - Discriminant	Peptide	Automatically optimizes the discriminant score thresholds until they reach a specified global or 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. - Use this mode before the Protein Polishing mode. - You must have searched your data using a Discriminant mode other than "Off".

Table 1 Choose the Autovalidation Strategy and Mode (continued)

Select this Strategy:	Select this Mode:	What happens:	Pay attention to this:
	Protein polishing (optional)	• See the “What happens” description for protein polishing above.	• See the “Pay attention to this” description for protein polishing above.
Fixed thresholds	Protein details	• Summarizes results by protein and considers all the peptides that belong to a given protein by entering fixed score thresholds beyond which the matches are not validated.	• Individual peptides typically have scores > 6 to 12; the cumulative protein score is typically > 20. • You must use this mode before the peptide mode; Protein details is step 1 and Peptide mode is step 2. • You can use the fixed thresholds strategy in an iterative workflow.
	Peptide	• Summarizes results by peptide by entering fixed score thresholds beyond which the matches are not validated.	• Using the default scoring, individual peptides typically have scores greater than 11 to 15 (depending on charge state). If the score is high enough, this mode validates the search results even if it finds only a single peptide corresponding to a protein. • Use the Quality Metrics tool to determine the resulting FDR. • The Fixed thresholds strategy was the only strategy that was initially available. Agilent now recommends you use the Auto thresholds strategy.

5 After selecting a strategy and mode, change any parameter values you choose.

Agilent recommends that for Agilent Q-TOF .d data you select the Auto thresholds strategy and validate in the peptide mode first, then the protein polishing mode (Figure 8). See Chapter 1 of the *Familiarization Guide* for the recommended parameter values. You may want to try other strategies with other instrument data.

NOTE

Click the blue divider bar for Validation Parameters to learn more about the parameters for each strategy and mode.

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

- 6 Click **Save As** to save settings to the original or a different parameter file, then click **Save**.

You can now use this parameter file as a “task” in the list of tasks for an automated workflow. See Chapter 3, “Automating Workflows”.

- 7 If you want to see results immediately, make sure the **Queue Request** check box is clear. If you want to queue the task as part of an *interactive automation*, mark the check box.

- 8 Click **Validate Files**.

If the Queue Request check box is clear, you quickly see a validation summary that lists the hits and spectra that have been validated. If it is marked, the task is submitted to the workflow manager to be executed.

Help

Validate Files ☐ Queue request

Data Directories

Select ...

☒ ExampleData/Agilent/Q-TOF/Yeast/Y1

Fragmentation mode: All

Search result files: *.spo

Validation Parameters

Strategy: ☐ Fixed thresholds ☒ Auto thresholds ☐ Auto thresholds - discriminant

Mode: Protein polishing

☒ Group proteins across all directories

Minimum number of directories protein group is observed in: 1

Minimum protein score: 0

Automatically raise minimum protein score to yield maximum protein FDR: 1.0 %

Figure 8 Autovalidation of MS/MS Search results – Auto thresholds/Protein polishing mode

Step 4. Examine the validated results for variable modifications mode

After autovalidating your results, you now have a choice of viewing different types of protein and peptide summary reports. See [Table 2](#) on page 42. For a normal workflow review, however, we want to see a peptide report showing the site locations of the variable modifications.

- 1 Navigate to the Protein/Peptide Summary page.
- 2 Set **Mode** to **Peptide - Spectrum Match**.
- 3 Click the **Select...** button to select the folder or folders that contain your data files.
- 4 If you want to use the settings from an existing parameter file, click **Load**.
If you are not sure of a field's meaning, click the category divider bar for the section containing the field to see the online help explanation for the field.
- 5 Set **Validation and Sorting** parameters as shown in [Figure 9](#), or click **Load** to load a parameter file with these settings. These settings filter and present all the validated data. Note that only the settings highlighted in yellow have been changed from defaults.
- 6 Make sure the **Filter Results by** is set to **valid** and the **Validation preset** is set to **none**.
- 7 To see the variable modifications grouped at the top and/or bottom of the results table, for **Sort peptides by**, select **Delta Mass Shift**.
- 8 If you want to see the variable modifications and/or amino acid substitutions, additionally mark the check boxes for **Var mod sites**, **VML score**, **Delta mass**, and **Modification Names**, as shown in [Figure 9](#).
 - If the variable modifications occur on the N-terminus, C-terminus, or cysteines, also mark the appropriate check box(es) in the last column of the **Review Fields**.
 - **VML score** assigns probabilities when there are multiple possible variable sites.
 - To use the pI filter for modified peptides, ask your server administrator to first verify that the pK of the modified amino acid is specified in **smconfig.std.xml** or **smconfig.custom.xml**.
 - To see a b/y map instead of a regular sequence, clear **Sequence** and mark the **b/y** check box.

Figure 9 Settings to filter and keep valid results

- 9 Click **Save As** to save settings to the original or a different parameter file, then click **Save**.

You can now use this parameter file as a “task” in the list of tasks for an automated workflow. See Chapter 3, “Automating Workflows”.

- 10 If you want to see results immediately, make sure the **Queue Request** check box is clear. If you want to queue the task as part of an *interactive automation*, mark the check box.

- 11 Click **Summarize**.

If the Queue Request check box is clear, you quickly see a protein/peptide summary table that lists the hits and spectra that have been validated. Queuing a summary is useful when you want the results saved to an html file that can be accessed via the Completion Log.

NOTE

Once you click the **Summarize** button, the button is disabled until the results appear.

- 12 After this task is complete, make sure that you see a display similar to that shown in [Figure 10](#).

Note that the filenames are in the format Data_File_Name.aaaa.bbbb.c, where

- aaaa = first merged scan
- bbbb = last merged scan

1 Processing MS/MS Data Interactively

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

- c = assigned precursor charge (0 means charge was ambiguous)

#	Filename	z	Score	SPI (%)	Spectrum Intensity	Score VML	# STY	# sty Sites	# Loc sty Sites	# Ambig sty Sites	Variable Sites
1	yeast-1.19473.19473.4	4	6.72	63.1	8.68e+004	4.383	3	2	2	0	M489m T491t S498s
2	yeast-1.14853.14853.2	2	10.14	92.3	3.74e+005	0.000	4	2	1	1	S336s T341t
3	yeast-1.4934.4934.2	2	6.41	73.2	2.02e+005	0.000	3	2	1	1	Y197y S199s
4	yeast-1.19691.19691.3	3	5.28	66.5	1.20e+005	5.279	6	2	2	0	Y275y T293t
5	yeast-1.17766.17766.3	3	6.37	56.0	2.03e+005	0.000	3	2	0	2	S1381s S1392s
6	yeast-1.12871.12871.4	4	7.60	58.6	2.18e+006	0.895	5	2	1	1	Y26y T45t
7	yeast-1.15330.15330.4	4	7.66	59.6	5.55e+006	0.462	8	2	0	2	T279t S289s
8	yeast-1.15030.15030.3	3	8.88	76.9	1.30e+005	1.137	3	1	1	0	M282m Y285y
9	yeast-1.19045.19045.3	3	8.47	93.3	1.46e+005	0.000	4	1	0	1	Y246y M253m
10	yeast-1.19393.19393.3	3	5.85	70.2	6.06e+005	0.000	4	1	0	1	M8m T11t
11	yeast-1.19549.19549.3	3	5.11	57.4	7.80e+004	5.111	4	1	1	0	S291s
12	yeast-1.14544.14544.3	3	5.27	59.4	1.87e+006	2.515	3	1	1	0	Y233y
13	yeast-1.15719.15719.3	3	7.39	64.2	1.88e+005	99	1	1	1	0	T31t
14	yeast-1.14753.14753.2	2	6.61	78.2	3.98e+004	99	1	1	1	0	T629t
15	yeast-1.16546.16546.2	2	8.41	76.3	1.09e+005	8.405	5	1	1	0	T204t

Figure 10 Protein/Peptide Summary report - Peptide Spectrum Match mode - first part

Note that the resulting data display includes the additional headings **Variable Sites**, **Modifications Map**, **MH⁺ mass shift (Da)**, and **MH⁺ error (ppm)**, as shown in [Figure 10](#) and [Figure 11](#). These columns summarize the variable modifications (and for homology modes the amino acid substitutions). The variable modifications are indicated as lower-case letters under **Variable Sites**.

Normal Workflow – Setting Up Parameter Files and Processing Tasks Interactively

Sequence Map	Modifications Map	MH ⁺ Matched (Da)	MH ⁺ Mass Shift (Da)	MH ⁺ Error (ppm)	Accession #	Protein Name
(K)ETW G K/m/N t F V L/H A L s K P L K F L E N L \N K(S) ET(0.0)WGKM(1.0)NT(0.99)FVLHALS(0.99)KPLKFLENLNK	m:Oxidized methionine tPhosphorylated T sPhosphorylated S	3058.655	175.9082	-6.0	P40458	Autophagy-rel
(R)L S L D I s A/K H L \t T K(L) LS(0.0)LDIS(0.99)AKHLT(0.50)T(0.50)K	s:Phosphorylated S tPhosphorylated T	1426.826	159.9587	16.4	P47139	Uncharacterize
(R)V I y T s P I K(A) VIY(0.99)T(0.50)S(0.50)PIK	y:Phosphorylated Y sPhosphorylated S	920.545	159.9376	4.5	P47047	ATP-depender
(R)Y S K P A L L H/S R F F P A L Q G S/t T K(M) Y(0.99)S(0.0)KPALLHS(0.0)RFFPALQGS(0.0)T(0.99)T(0.0)K	y:Phosphorylated Y tPhosphorylated T	2349.271	159.9347	0.8	Q12109	Tryptophan-tf
(K)K P I L K s V Q N/V/E V E E A P s S D K K(N) KPIKLS(0.0)VQNVEVEEAPS(0.50)S(0.50)DKK	s:Phosphorylated S	2325.266	159.9185	-5.7	P34244	Probable serin
(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)APNMLWGGAS(0.50)MLGLFVFT(0.50)EGWPK	y:Phosphorylated Y tPhosphorylated T	3157.589	159.9166	-4.8	P37299	Cytochrome b-
(K)F R L/A K M t F L/R P S S L I T s E S A I V T T K(I) FRLAKMT(0.25)FLRPS(0.0)S(0.25)LIT(0.0)S(0.25)ES(0.25)AIVT(0.0)T(0.0)K	t:Phosphorylated T sPhosphorylated S	2797.564	159.8965	-12.2	P53916	Probable phos
(K)L/P/P/m D A y L K L N H I I A N I N S F T/K(G) LPPM(1.0)DAY(0.99)LKLNHIINANINS(0.0)FT(0.0)K	m:Oxidized methionine yPhosphorylated Y	2513.358	95.9766	5.9	P36048	Pre-mRNA-spl
(K)L S D I 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 mOxidized methionine	2217.181	95.9596	-0.7	Q06063	tRNA-dihydrou
(K)I L m D/s t H F N E I R S I I R S R(S) ILM(1.0)DS(0.25)T(0.25)HFNEIRS(0.25)IIRS(0.25)R	m:Oxidized methionine tPhosphorylated T	2188.166	95.9470	-6.2	P41807	V-type proton
(K)N I P W Y G E T K N P V L F/S N A V E N L s R(T) NIPWY(0.0)GET(0.0)KNPVLFS(0.0)NAVENLS(0.99)R	s:Phosphorylated S	2648.347	80.0110	16.4	P32528	Urea amidolya
(K)S N L A G K F G F T A V V I y D N E P K(S) S(0.0)NLAGKFGFT(0.0)AVVIY(0.99)DNEPK	y:Phosphorylated Y	2170.118	79.9976	13.9	P37302	Aminopeptidas
(R)R L t G F L P Q E I/K(S) RLT(1.0)GFLPQEI	t:Phosphorylated T	1301.758	79.9936	19.7	P06738	Glycogen phos
(R)I E Q E Q V P G G A P G/E t V K(V) IEEQVPGGAPGET(1.0)VK	t:Phosphorylated T	1638.833	79.9925	15.2	P36013	NAD-depende
(R)I t/K A/D I E S Y L E K S S K(Q) IT(0.99)KADIES(0.0)Y(0.0)LEKS(0.0)S(0.0)K	t:Phosphorylated T	1711.911	79.9918	14.2	P12695	Dihydrolipoyly

Figure 11 Additional columns identifying and locating the sites of selected variable modifications**NOTE**

If you click the SPI (%) value for any hit, the standalone Spectrum Viewer appears. You must have marked the SPI (%) check box under Review Fields for this to be possible.

Iterative Workflow – Alternative Interactive Processing

With an iterative workflow, you extract the data, then do a series of searches and autovalidations and then summarize all the valid results. You use the normal workflow to identify most of the MS/MS spectra in one round. When you occasionally must obtain more information to help characterize the identified proteins, use an iterative workflow.

An iterative workflow is helpful when:

- You want to look for an atypical modification or you suspect a sample handling artifact that you do not usually consider.
- A variable modification has a particularly large delta mass, which would slow a search.
- You search a second database.

Overview

The Spectrum Mill workbench retains a cumulative summary of valid results, so if you elect to further process and validate the data, the software keeps track of all validation states for you. At the end, you can generate a report that contains everything you have validated through multiple interpretation/review/validation steps.

When you do iterative searches, the false discovery rate (FDR) is no longer accurate. So first determine what modifications are necessary. Use a mass gap or homology or additional variable modifications search to do that. Then do a new search that includes those modifications. The FDR from the new search will be correct.

To segregate the valid search results, the software keeps track of both the validation states of the spectra and their interpretations in a coordinated way. Furthermore, it permits spectra to be segregated according to quality (via Spectrum Summary) without regard to the validation states of their database interpretations. Because the software keeps track of which spectra you have segregated as “good” and which database interpretations you have designated as “valid,” you can intelligently apply processing steps to segregated groups of spectra. For example, you can:

- Restrict further database searches to spectra that do not yet have a valid interpretation
- Restrict further database searches to only “good” spectra that do not yet have a valid interpretation
- Summarize only the valid results

Steps for iterative processing

The iterative data processing steps you choose depend upon your sample, the available information in the databases you search, and how important it is to extract all available information from a sample set. The steps shown below are just an example.

Step 1. Do additional searches and autovalidations

NOTE

For accurate FDR, search all expected modifications in a single search. Mark **Maximize CPUs** in this situation.

Possible iterative workflow steps are:

- Search saved results for other possible modifications.
- Do a mass gap search against saved results, to find unexpected modifications.
- Search an additional species. In this case, you do *not* search saved results. You must do a new search so the FDR calculations are accurate.
- Search previous hits with a no-enzyme or half-enzyme (trypsin nonspecific C term or trypsin nonspecific N term) digest and review/validate results. See [“To search in no enzyme mode”](#) on page 47.
- Search a larger database in identity mode and review/validate results. In this case, you do *not* search saved results. You must do a new search so the FDR calculations are accurate. See [“To search a larger database”](#) on page 47.
- Search previous hits in homology mode and review/validate results. See [“To search in homology modes, or do a mass gap search”](#) on page 48.
- Check for remaining high-quality spectra. See [“Step 2. Check for remaining high-quality spectra”](#) on page 37.

- Create a final results summary. See “[Step 4. Summarize valid results](#)” on page 41.
- Use Sherenga *de novo* sequencing on remaining high quality spectra and review results. See “[To use Sherenga de novo sequencing](#)” on page 50.

You can also manually interpret, review and validate the data if you wish to mine the data more deeply or to confirm autovalidation results. See “[Reviewing and Validating Results Interactively](#)” on page 76.

Search modes

The following modes are available for MS/MS Search:

- Identity: Matches MS/MS spectra to sequences directly created from the databases (no modifications)
- Variable modifications: Searches for variable modifications you have selected. This mode tests all permutations of the combined set of modifications, within the limits of the precursor mass shift that you specify.
- Homology - All mutations: Searches for matches that are consistent with a single amino acid substitution. The substitution does not need to be one that would result from a point mutation. If you have selected variable modifications for an amino acid, they are also tested. The search also matches to sequences of unmodified peptides.
- Homology - Single base pair mutations: Searches for matches that are consistent with a single amino acid substitution that would result from a point mutation (single base substitution within a codon). If you have selected variable modifications for an amino acid, they are also tested. The search also matches to sequences of unmodified peptides.
- For the two homology modes, you can enable a search that allows an **Unassigned single mass gap**. You can use this type of search to identify unknown or unexpected modifications. This search also matches to sequences of unmodified peptides. If you enable this mode, you cannot search for variable modifications in the same search.

Types of databases

You can run a variable modifications or homology mode search against a full database, a species-specific subset database, or a small database of proteins that have already been identified in your sample. To create a species-specific subset database, see “[To create a species and protein molecular weight subset database](#)” on page 201. The broader the database search, the longer the search

time. A search of previous hits is the most efficient.

If you search species-subset databases, you may want to run more than one variable modifications or homology mode search. For example, if you study proteins from a rare species, you may want to search first a species-specific subset database and second, a more general subset database.

Step 2. Check for remaining high-quality spectra

- 1 Navigate to the Spectrum Summary page.

This page provides a convenient means to find good spectra that have not yet been interpreted.

One of the most useful measures of spectral quality is the maximum sequence tag length, which is the longest sequence of amino acids that can be located in the spectrum. 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.

- 2 Set parameters similar to those shown in Figure 12, or click **Load** to load an existing parameter file.

The steps below discuss important parameters.

Agilent Spectrum Mill - Spectrum Summary - (none)

Summarize Results for Review

Summarize Save As... Load...

☐ Excel Export

Spectrum Files:

*.pk1

Sorting

Sort by: Maximum tag length

Data Directory

Select ... ExampleData\Agilent\Q-TOF\Yeast\Y1

Spectral Quality Filtering

☒ Sequence tag length: > 7

☐ Precursor isotope quality XIC's (Chi2 vs. Average): > 0.7

☐ Precursor Isolation Purity: > 50 %

Filter by: None > 3

Fragmentation mode: All

Spectrum validation filter:

spectrum-not-marked-sequence-not-validated

Validation preset: good-spectrum

☒ DB search Result

☒ Longest sequence tag

☒ Precursor charge

☐ Fragmentation mode

☒ Precursor MH+ m/z

☐ Acquired precursor m/z

☐ Collision energy ☐ RT

☐ Ion mobility

☐ # b/y pairs

☒ MS precursor EIC intensity

☐ MS L/H EIC intensity

☐ Reporter ion intensity

☐ MS/MS TIC intensity

☒ # peaks Detected

☐ # peaks Centroid ☐ Profile

☐ Noise Mean ☐ Std dev

☐ Base peak intensity

☐ Base peak / TIC ratio

Figure 12 Spectrum Summary page

- 3** In the **Data Directory** section, click the **Select...** button to select the folder or folders that contain your data files.
- 4** Set the **Spectrum validation filter** as follows:
 - If this is the first round of processing using the Spectrum Summary page, set to **spectrum-not-marked-sequence-not-validated**.
 - For later rounds of processing (after you have already designated spectra as “good”), select one of the options that start with **good-spectrum**.
- 5** Choose one of the four choices for **Validation preset**:
 - If you set **Spectral Quality Filtering** parameters such that poor spectra are likely to be excluded (for example, **Sequence tag length > 7**), select **good-spectrum**.

The **Sequence tag length** represents the longest sequence of amino acids that were located in the spectrum during data extraction. Larger numbers filter out more spectra, keeping those of best quality.
 - Otherwise, select **bad-spectrum** or **reset** or **none**.

As you review the spectra later, you can select **reset** to undo a **good-spectrum** or **bad-spectrum** designation.
- 6** Select the desired **Filter by:** parameter and **Fragmentation mode**.
- 7** Mark the **MS L/H EIC intensity** check box for additional information on ICAT and other isotopically labeled data. For more information, see “[To view light/heavy results on the Spectrum Summary page](#)” on page 143 of [Chapter 5](#).
- 8** Click **Save Settings** if you want to save the parameters.
- 9** Click **Summarize**.

NOTE

Once you click the **Summarize** button, the button is disabled until the results appear. If you need to re-enable the button, click the **Summary Settings** button to reload the Spectrum Summary page.

- 10** Check that you see a display similar to the one shown in [Figure 13](#).

Note that the filenames 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)

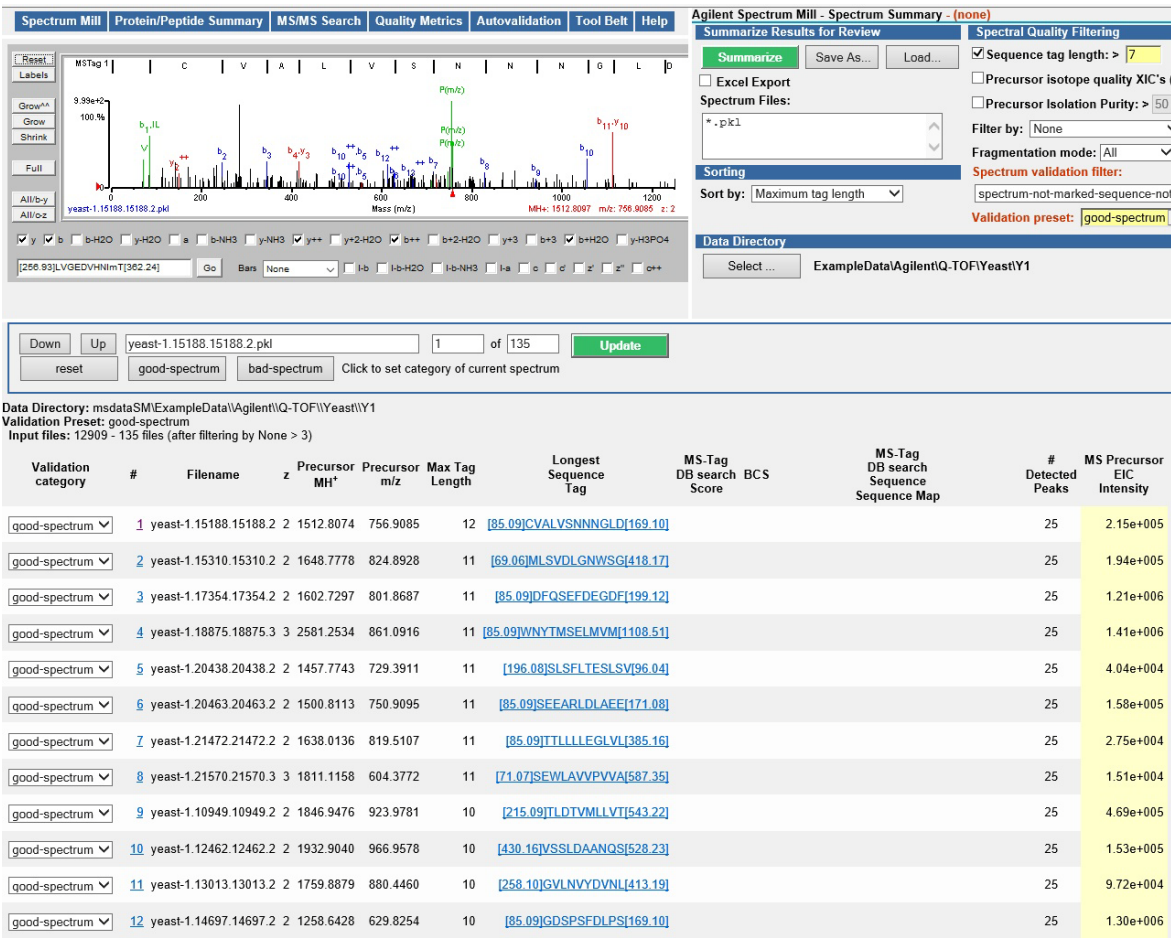


Figure 13 Spectrum Summary results

11 Now examine the spectra. The Spectrum Viewer, shown at the top left of Figure 13, automatically displays the first spectrum. To display others, click a link under the # or **Longest Sequence Tag** headers, or click the **Up** or **Down** button at the top of the page.

- If you click a link under the # header or the **Up** or **Down** button, the spectrum is displayed at the top left of the page.

1 Processing MS/MS Data Interactively

Iterative Workflow – Alternative Interactive Processing

- If you click a link under the **Longest Sequence Tag** header, the spectrum is displayed in a new window.
- For more details about manipulating the spectral display, see “To use the Spectrum Viewer” on page 83 of [Chapter 2](#).

12 After you review each spectrum, designate each one as “good” or “bad,” or “reset.” Use either the drop-down list under the **Validation category** heading, or the **good-spectrum**, **bad-spectrum**, and **reset** buttons near the top of the page.

13 Click **Update** to save your designations for future reference.

NOTE

If you find you have designated spectra as “good” or “bad” in error, and you have already clicked the **Update** button, you can undo the action. Change the designation(s) to **Reset** and click the **Update** button again.

If at any point you wish to change parameters and create a new summary table, click the **Summary Settings** button to start over.

Step 3. Make a decision regarding further processing

- 1 After you use the Spectrum Summary page, assess whether you still have a lot of good spectra that lack valid interpretations and decide whether you want to further process the data.
If you do not wish to further process the data, summarize the valid results.
- 2 If you do wish to further process the data, decide how much you wish to do. The possible next steps are:
 - a Search for additional variable modifications
 - b Search in no enzyme mode against previous protein hits. (See [“To search in no enzyme mode”](#) on page 47.)
 - c Search in one of the homology modes against previous protein hits (See [“To search in homology modes, or do a mass gap search”](#) on page 48.)
 - d Search a larger database. (See [“To search a larger database”](#) on page 47.)
- 3 Process the data with one or more of the choices above.
- 4 If you do not wish to further process the data with the above choices, create a final results summary. See [“Step 4. Summarize valid results”](#) on page 41.
- 5 Continue with manual review and validation, if you choose. (See [“Reviewing and Validating Results Interactively”](#) on page 76.)

Step 4. Summarize valid results

At this point you summarize all the valid results.

- 1 If you have been using links to examine the data, click the **Summary Settings** button to display the settings at the bottom of the page.
- 2 Make sure that your data folder is still selected. (Note that if the **Validation preset** is set to **none**, some data display modes permit selection of multiple directories.)
- 3 Select the data display mode that organizes data in the best way for your study, or click **Load** to load an already existing parameter file that includes the display mode and other parameters you want.

Use the following table as a guide. Note that display modes determine which other options are available.

1 Processing MS/MS Data Interactively

Iterative Workflow – Alternative Interactive Processing

Table 2 Protein/Peptide Summary Modes

If you want to:	And you want results organized by:	Then use this mode:	Example application
Validate results	Peptides	Peptide - Spectrum Match	Manual review and validation of MS/MS search results, organized by peptide
	Proteins, then peptides	Protein Summary Details	Manual review and validation of MS/MS search results, organized by protein
Summarize results by proteins	Proteins, then samples	Protein-Protein Comparison	Compare two or more samples, each of which may contain multiple fractions.
	Proteins, then peptides	Protein Summary Details	View proteins, with supporting peptide details
Summarize results by peptides	Peptides only	Peptide - Spectrum Match	List of all peptides identified in the data
	Peptides, then samples	Protein-Peptide Distribution Columns	Method development (evaluation of 2D LC/MS/MS or other fractionation scheme)
	Rows: peptides clustered by protein Columns: samples	Protein-Peptide Comparison	Table of distinct peptides, case-insensitive (CI). Collapse all peptide spectrum matches (PSMs) that contain same peptide into one row. Combine quantitation for each peptide: median of all PSMs.
	Distinct peptides only	Peptide-Distinct	Provides filtering options to view peptides with variable modifications. Collapse peptide list to best-scoring representative PSM per distinct peptide.
Summarize results by variable modification sites	Rows: variable modification sites clustered by protein Columns: samples	Protein - Var Mod Site Comparison	Compare variable modifications across samples. Provides table of phosphorylation sites. Collapse all PSMs that contain same phosphorylation site into one row. Combine quantitation for each site: median of all PSMs.
View a list of proteins identified via a single peptide	Peptides only	Protein-Single Peptide ID	Examination of results where a single peptide was used to generate a protein identification

- 4 In some modes, you have the option to display results by file or directory. Decide how you want to display data and select the **Group results by** option accordingly.

- If you want to summarize all data files in the directory as a single sample, select **Directory**. This is useful when the samples in your directory are salt fractions from a 2D LC/MS/MS analysis or spots from a 2D gel, and you want to summarize fractions or spots as a single sample.
 - If you want to display results individually for each sample in the directory, select **File**.
- 5 Mark **DEQ ratios** to display ratios for differential expression quantitation for ICAT and SILAC reagents, or mark **Report ratios** for iTRAQ and TMT data. See [Chapter 5](#) for details.
 - 6 Set the remaining parameters as shown in [Figure 14](#), or use your customized settings. Note that in [Figure 14](#), only the settings highlighted in yellow have been changed from defaults.

See the online help for descriptions of all the fields.

Figure 14 Example of settings to summarize all valid results

- 7 Click **Save As** to save settings to the same or different parameter file, then click **Save**.
You can now use this parameter file as a “task” in the list of tasks for an automated workflow. See [Chapter 3](#), “Automating Workflows”.
- 8 If you want to see results immediately, make sure the **Queue Request** check box is clear. If you want to queue the task as part of an *interactive automation*, mark the check box.
- 9 Click **Summarize**.

1 Processing MS/MS Data Interactively

Iterative Workflow – Alternative Interactive Processing

If the Queue Request check box is clear, you quickly see a protein/peptide summary table that lists the hits and spectra that have been validated. Queuing a summary is useful when you want the results saved to an html file that can be accessed via the Completion Log.

10 Check that you see a summary display.

The display shown in Figure 15 was generated with the settings shown in Figure 14. The score and % SPI filters were set to zero to capture all data marked **valid**.

ExampleData/Agilent/Q-TOF/Yeast/Y1 # spectra total intensity	ExampleData/Agilent/Q-TOF/Yeast/Y2 # spectra total intensity	Columns Protein Observed (#)	Spectra Protein Observed (#)	Database Accession #	%AA Coverage	Distinct Peptides (#C1)	Distinct Summed MS/MS Search Score	Group #	Subgroup #	Protein Name
57 4.11e+008	147 1.47e+009	2	204	P00924	88.7	48	753.64	1	1.1	2X Enolase 1
59 3.20e+008	97 9.39e+008	2	156	P00925	86.4	39	615.17		1.2	Enolase 2
55 1.22e+008	58 2.00e+008	2	113	P10591	73.8	46	626.84	2	2.1	5X Heat shock protein SSA
45 1.05e+008	45 1.58e+008	2	90	P10592	60.5	38	499.56		2.2	Heat shock protein SSA2
13 1.08e+007	18 2.69e+007	2	31	P16474	31.3	20	226.40		2.3	78 kDa glucose-regulated protein homolog
21 4.25e+007	19 5.58e+007	2	40	P09435	27.1	18	235.57		2.4	Heat shock protein SSA3
42 4.01e+007	51 7.89e+007	2	93	P32324	58.3	41	540.18	3	3.1	Elongation factor 2
69 4.91e+008	76 6.49e+008	2	145	P00560	80.5	33	528.43	4	4.1	Phosphoglycerate kinase
43 1.21e+008	45 1.81e+008	2	88	P00549	82.8	36	499.84	5	5.1	Pyruvate kinase 1
41 3.53e+007	41 5.52e+007	2	82	P05694	54.3	36	495.05	6	6.1	5-methyltetrahydropteroyltriglutamate-homocysteine
31 1.06e+007	33 1.64e+007	2	64	P07149	24.7	39	489.17	7	7.1	Fatty acid synthase subunit beta
35 3.25e+007	36 5.09e+007	2	71	P15108	47.8	36	482.13	8	8.1	2X ATP-dependent molecular chaperone HSC82
29 2.97e+007	32 4.69e+007	2	61	P02829	41.7	32	419.57		8.2	ATP-dependent molecular chaperone HSP82
56 4.09e+008	83 7.21e+008	2	139	P00359	88.8	29	470.02	9	9.1	3X Glyceraldehyde-3-phosphate dehydrogenase 3
43 3.02e+008	38 5.04e+008	2	81	P00360	87.3	27	371.47		9.2	Glyceraldehyde-3-phosphate dehydrogenase 1
46 3.55e+008	48 5.76e+008	2	94	P00358	80.4	26	401.41		9.3	Glyceraldehyde-3-phosphate dehydrogenase 2
36 1.51e+007	37 3.41e+007	2	73	P17255	40.5	42	467.50	10	10.1	V-type proton ATPase catalytic subunit A
39 6.78e+007	42 1.00e+008	2	81	P11484	64.2	33	465.62	11	11.1	2X Heat shock protein SSB1
38 6.57e+007	41 9.57e+007	2	79	P40150	65.5	33	461.53		11.2	Heat shock protein SSB2
30 2.44e+007	35 8.75e+007	2	65	P19414	48.4	35	445.76	12	12.1	Aconitate hydratase, mitochondrial
51 8.24e+007	53 2.05e+008	2	104	P06169	62.8	28	439.71	13	13.1	3X Pyruvate decarboxylase isozyme 1
45 1.11e+007	41 5.52e+007	2	82	P05694	54.3	36	495.05	6	6.1	5-methyltetrahydropteroyltriglutamate-homocysteine

Figure 15 Summary of valid results- comparison of two samples

The example in Figure 15 includes two *yeast* samples. The summary results allow you to compare the proteins in the two samples, to see which are up-regulated and which are down-regulated in the cells.

The color-code in the left columns represents the relative protein concentrations. Dark red is highest, orange is intermediate, and yellow is lowest. The top number in each cell is the number of peptides detected for each protein, while the bottom number is the mean peptide intensity. The mean peptide intensity is an average of the intensities for all the peptides assigned to that protein. The peptide intensities are calculated from extracted ion chromatograms from the precursor ions.

These intensity results are sufficient for studies where you are interested in sample differences of two-fold or more.

11 To import data into Excel or MPP or LIMS:

- a** Mark the **Excel** or **MPP export** check box.

You can add this change to the parameter file, or not.

- b** Click **Summarize**.

- c** Check that you see a display with two buttons, as in [Figure 16](#). At this point, a new file with extension ***.ssv** has been created. You can import this file into Excel or upload it to a LIMS system (if configured by your system administrator), or display it on your screen.

- d** Do one of the following:

- To import the data into Excel, import as semicolon-delimited data.
- To import the data into MPP, use the MPP Generic Data Import or MPP APR export (Agilent data only). With MPP APR export, if you have multiple sample folders, each with multiple fraction data files (.d folders), for **Group columns by**, select **Directory**.
- To upload to LIMS, see the Server Administration online help.

1 Processing MS/MS Data Interactively

Iterative Workflow – Alternative Interactive Processing

Agilent Spectrum Mill - Protein/Peptide Summary - Defaults\pep_sum_defaults_valid

Spectrum Mill	Summary Settings	Autovalidation	MRM Selector	MS/MS Search	Spectrum Summary	Wo
---------------	------------------	----------------	--------------	--------------	------------------	----

Preparing results for Excel Export.....
Data Directory: msdataSM\ExampleData\Agilent\Q-TOF\Yeast\Y2
hit table read 5396 files - SpecFeatures read parent_charge excluded
valid hits read from tagSummary file - **Files: 5396 Hits: 6461**
Data Directory: msdataSM\ExampleData\Agilent\Q-TOF\Yeast\Y1
hit table read 4782 files - SpecFeatures read parent_charge excluded
valid hits read from tagSummary file - **Files: 4782 Hits: 5766**
beginning to assemble proteins proteins assembled 0.81122 sec
proteins filtered by distinct peptides 0.226866 sec
proteins filtered by score
calculated protein coverage maps 1.763405 sec
beginning to roll up 679 proteins into groups ...
Sorted peptides by length to accelerate peptide roll up 0.052876 sec
removing P0CW40 from group; 39 with index: 2 to group 64
removing P0CW41 from group; 39 with index: 3 to group 64
removing Q08295 from group; 39 with index: 4 to group 64
completed primary peptide roll up 3276760 sequence comparisons 5.305346 sec
The protein (P22202) in group 2 (sort by score) subgroup can be subsumed by earlier subgroups.
completed post roll up subgrouping and sorting 0.078616 sec
Finished protein roll up into 557 groups 5.448971 sec
protein groups ready for display
proteinGroupingMethod: oneSharedPeptideSubgrouped Subgroups expanded 0.462346 sec
Output File created msdataSM\ExampleData\Agilent\Q-TOF\Yeast\Y2\proteinProteinCentricColumnsExport.1.ssv
writing results
Generated Protein Comparison Column Table: 0.872402 sec

Upload Path [msdataSM\ExampleData\Agilent\Q-TOF\Yeast\Y2\proteinProteinCentricColumnsExport.1.ssv](#)

[Upload File to LIMS](#) [Display Created File Below](#)

Figure 16 Display when you mark the Excel export check box

12 To print results, use your browser's print function:

- Click **File > Page Setup**, and mark the check boxes for **Print background colors and images** and for **Landscape**.
- Click **OK**.
- Click in the results frame.
- Select **File > Print Preview**.
- At the top of the Print Preview window, select **Only the selected frame**.
- Click the **Print...** button.

NOTE

To export results into HTML or PowerPoint files, see the Frequently Asked Questions section of the online help.

To search in no enzyme mode

- 1 Navigate to the MS/MS Search page.
- 2 Fill in the search page as described in “[Step 2. Search in variable modifications mode](#)” on page 20, except:
 - Mark the check box to **Search previous hits**.
 - Set **Digest** to **No enzyme**.
 - To reduce the memory impact of each batch search in **No enzyme** mode, set the **Batch size** to **20** or less.
- 3 Run the search.
- 4 Use the Autovalidation page and the Protein/Peptide Summary page to review and validate results. Follow instructions on [page 25](#) through [page 41](#).

To search a larger database

- 1 Navigate to the MS/MS Search page.
- 2 Fill in the search page as described in “[Step 2. Search in variable modifications mode](#)” on page 20, except select a different database this time.

If you first searched a species subset database (e.g., human), search a larger subset (e.g., mammals). This is most important when the original subspecies is not well-represented in the database.
- 3 Run the search.
- 4 Use the Autovalidation page and the Protein/Peptide Summary page to review and validate results. Follow instructions on [page 25](#) through [page 41](#).

To search in homology modes, or do a mass gap search

- 1 Follow the instructions in “[Step 2. Search in variable modifications mode](#)” on page 20, but make these substitutions depending on what you want to do:
 - To search for matches that are consistent with a single amino acid substitution that would result from a point mutation (single base substitution within a codon), select **Homology - Single base pair mutations**. This mode also searches for any variable modifications you have selected.
 - To search for matches that are consistent with a single amino acid substitution, where the substitution does not need to be one that would result from a point mutation, select **Homology - All mutations**. This mode also searches for any variable modifications you have selected.
 - To search for an unknown or unexpected modification, select one of the homology modes described above, then click **Unassigned single mass gap**. This search looks for an unexpected modification (a mass gap) or an amino acid substitution (but not both within the same peptide). Note that you cannot simultaneously search variable modifications in this mode.
- 2 If you do a homology mode search, autovalidate and review results as described in “[Step 3. Validate high quality results automatically](#)” on page 25 and in “[Step 4. Summarize valid results](#)” on page 41

Or you can create a peptide report with modifications fields selected; see “[Step 4. Examine the validated results for variable modifications mode](#)” on page 30.

Autovalidation is not recommended for the homology modes with the Fixed threshold strategy because of the increased possibility of false positives when so many permutations are considered. But using the Auto threshold strategies (including discriminant) is fine because the FDR automatically adjusts thresholds as the search space gets larger and larger.

NOTE

If you attempt a homology search against a large database, the search will take a very long time and the chance of random matches will increase. If at all possible, conduct a homology search against a file of previously validated hits. Or select **Search previous hits** in the MS/MS Search page. If a file of previously validated results has not yet been created, the program will automatically create the file when the search starts.

Depending on the size of the data and the database, it may also be necessary to reduce the **Batch size** for the search. Do this to avoid out-of-memory problems due to the increased number of hits.

To use Sherenga *de novo* sequencing

Sherenga *de novo* sequencing is useful as a first step in manual spectral interpretation, or as a confirmation of MS/MS Search results. This section describes briefly how to use the Sherenga *de novo* sequencing software. For additional details and background information, see Chapter 5.

- 1 Navigate from the Spectrum Mill home page to the Sherenga *de novo* Sequencing page, shown in Figure 17.

Figure 17 Sherenga *de novo* Sequencing page

- 2 To use an already existing parameter file, click **Load**.
- 3 Set the **Validation filter**:
 - If you took the time to designate your spectra as “good” or “bad,” as described in “[Step 2. Check for remaining high-quality spectra](#)” on page 37, set the **Validation filter** to **good-spectrum-sequence-not-validated**. This ensures that only the uninterpreted spectra you designated as “good” are sequenced.

- Otherwise, set the **Validation filter** to **spectrum-not-marked-sequence-not-validated**.
- 4 Click the **Select...** button to select the data directory where your files reside.
 - 5 Click the **Choose...** button to select the modifications that match the chemistry for your samples.
 - 6 Under **Sequencing Parameters**, for **Scoring**, select your instrument.
 - 7 Set the other parameters. Click the blue divider bars for more information in the online help.
 - 8 To save the settings to a parameter file click **Save As**.
You can use these parameter files for workflow automation.
 - 9 Click **Sequence**.
 - 10 After clicking **Monitor Results**, check that you see a display similar to that shown in [Figure 18](#). The sequencing program processes the entire directory. A message alerts you when processing is complete.

```
Monitor output file:
msdataSM\ExampleData\Agilent\Q-TOF\Yeast\Y1
\sherenga.150604165215.308.htm
Stop Search PID: 4112

Removing all prior Sherenga results
All prior Sherenga results removed
-comb -opt
E:\SpectrumMill\msparams_mill\Agilent.qtof.trypsin.sherc
-sma 10 -osp -ntop 0 -ctop 0 -swi 100 -vsf 2 -gth -gbr -spo
-pap2 -grr2 -fm carbamidomethylation -vm Oxidized-
Methionine -ilm 1 -kqm Q

Processing first of 301 files ...

1. EC: 0 yeast-1.3572.3572.2.pkl.txt
2. EC: 0 yeast-1.3794.3794.2.pkl.txt
3. EC: 0 yeast-1.3817.3817.2.pkl.txt
4. EC: 0 yeast-1.4185.4185.2.pkl.txt
5. EC: 0 yeast-1.4789.4789.2.pkl.txt
6. EC: 0 yeast-1.5023.5023.2.pkl.txt
7. EC: 0 yeast-1.5323.5323.2.pkl.txt
8. EC: 0 yeast-1.5613.5613.4.pkl.txt
9. EC: 0 yeast-1.5645.5645.2.pkl.txt
10. EC: 0 yeast-1.5860.5860.2.pkl.txt
11. EC: 0 yeast-1.6072.6072.2.pkl.txt
12. EC: 0 yeast-1.6112.6112.2.pkl.txt
13. EC: 0 yeast-1.6276.6293.2.pkl.txt
14. EC: 0 yeast-1.6444.6444.2.pkl.txt
15. EC: 0 yeast-1.6470.6470.2.pkl.txt
16. EC: 0 yeast-1.6510.6510.2.pkl.txt
17. EC: 0 yeast-1.6622.6622.2.pkl.txt
```

Figure 18 Display while Sherenga runs

To review Sherenga *de novo* sequencing results

- 1 Click the ***de novo* Summary** button to navigate from the Sherenga *de novo* Sequencing page to the Sherenga *de novo* Summary page.
- 2 Check that you see the page shown in [Figure 19](#).

The screenshot shows the 'Summarize Results for Review' window. It has a title bar and three buttons: 'Summarize' (green), 'Save As...', and 'Load...'. Below is the 'Data Directory' section with a 'Select ...' button and the text 'ExampleData\Agilent\Q-TOF\Yeast\Y1'. The 'Parameters' section includes 'Min. vertex score: 2' (with a yellow highlight), 'Show aligned sequence tags' (checkbox), 'Force tryptic digest' (checkbox), 'Sort by: Sherenga Score' (dropdown), and 'Result files (./results_sherenga/)' (text box). At the bottom is a file list showing '*.txt'.

Figure 19 Sherenga *de novo* Summary page

- 3 You should generally keep the default settings, but if you want to load a set of different parameters, click **Load** to load a parameter file.
- 4 If you change any settings, click **Save As** to save the new settings to an existing parameter file or a new one.
- 5 Click **Summarize**.
- 6 Check that you see a display like that shown in [Figure 20](#).

Note that the filenames are in the format
Data_File_Name.aaaa.bbbb.c.pkl.txt, where

- aaaa = first merged scan
- bbbb = last merged scan
- c = assigned precursor charge (0 means charge was ambiguous)

Results						
SS	TL	TS	MS/MS Search Delta MH ⁺	Filename	Sherenga Sequence Tag	MS/MS Search Result
59	9	21.3	0.0	yeast-1.11439.11888.2.pkl.txt	[171.06]PTV[251.14]VmVV[234.09]	GNPTVEVELTTEK
59	8			yeast-1.17013.17262.2.pkl.txt	[213.11]IDTDCEYAYIT[326.20]	
59	9			yeast-1.17135.17361.2.pkl.txt	[257.10]VYmDVEA[300.11][210.13][213.13]	
58	10	19.2	0.0	yeast-1.16122.16320.2.pkl.txt	IGIDCA[174.06]Em[275.16]	IGLDCASSEFFK
57	8	18.5	-0.0	yeast-1.14137.14362.2.pkl.txt	ISVQDID[241.18]	LSVQDLDLK
57	11	23.4	0.0	yeast-1.20456.20456.2.pkl.txt	[214.13]AETIAEm[208.11][185.09][199.13]	TIAETLAELINAAK
54	11	20.3	0.0	yeast-1.19422.19422.2.pkl.txt	[273.14]IADAIN[184.12][556.26]	SSVLADALNAINNAEK
54	11	22.0	-0.0	yeast-1.20458.20458.2.pkl.txt	[200.12]mIQEIG[284.08][199.13]	TVFIQELINNAK
52	9	20.1	-0.0	yeast-1.20174.20174.2.pkl.txt	[699.33]AIAAIQIED[768.37]	IEAALSDAALQIEDPS
51	9			yeast-1.19114.19321.3.pkl.txt	[212.15][244.07][260.12]IEEG[188.09][357.19]H	
49	9	16.6	-0.0	yeast-1.17897.18082.2.pkl.txt	[188.09]mmQAIHQ[198.11]	TSFFQALGVPTK
49	8			yeast-1.18517.18517.2.pkl.txt	[602.29]MIT[228.13]ATAD[566.27]	
48	9	19.2	0.0	yeast-1.10534.10534.2.pkl.txt	[416.13]SVT[226.10][158.09][246.11][168.06]	ADDDSVTIISAGNDK
48	13	20.3	0.0	yeast-1.19511.19511.2.pkl.txt	[200.12]IWVIE[199.09][410.14][154.11]	TVIGEVLLEQAYGGMR
47	8	5.7	0.0	yeast-1.17237.17267.2.pkl.txt	mEEI[231.06]FC[224.12]	LEFEQLYDDK
47	8			yeast-1.17464.17464.2.pkl.txt	[229.12][170.09][241.14]V[361.12][154.10]EA[170.11][268.12]	
47	8	18.0	-0.0	yeast-1.18014.18014.2.pkl.txt	[230.09]EISIIQ[172.12][184.09]	ETELSLQLSLR
47	9			yeast-1.19496.19496.2.pkl.txt	[200.12]IWVII[328.14]Y[401.19]	
47	11			yeast-1.20438.20438.2.pkl.txt	PVSISET[242.15][236.18][247.10]	
47	11			yeast-1.20463.20463.2.pkl.txt	[172.10]EERA[228.11][154.03][329.17]	
47	9	17.8	0.0	yeast-1.9018.9018.2.pkl.txt	[202.10]EEINN[202.10]PDR	TTEEQVENAVDR
46	11	16.3	0.0	yeast-1.14895.14895.2.pkl.txt	[953.52]YWSVII[224.09]	NQILVSGEIPSTLNEESK
46	10			yeast-1.21302.21302.2.pkl.txt	[1021.48][227.13]AIAAY[200.11][499.19]	

Figure 20 Sherenga top-level report

7 Review Sherenga results:

- Click a link under the **Filename** header. Check that you see the Spectrum Viewer as shown in [Figure 21](#) and detailed Sherenga results as shown in [Figure 22](#). The detailed 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.
- Click the **Rank** arrow buttons shown in the lower right-hand corner of [Figure 21](#) to navigate through the *de novo* possibilities. See “To use the Spectrum Viewer” on page 83 of [Chapter 2](#) for more details.

1

- C**

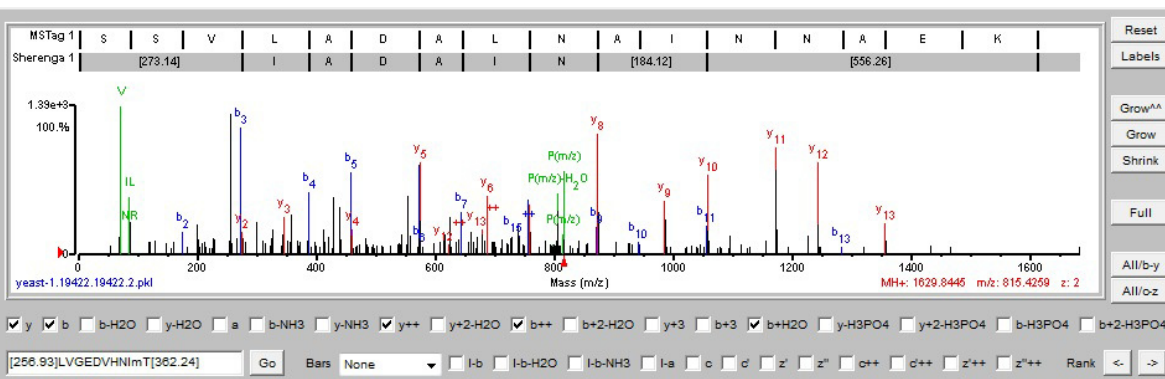


Figure 21

[illegible]

Figure 22

- 8 Compare Sherenga results with MS/MS Search results. When database search results are available, the software displays them along with the Sherenga interpretations, so you can compare the Sherenga and MS/MS Search sequences. When MS/MS Search results exist, the database sequence (labeled **MSTag 1**) is displayed just above the Sherenga sequence (labeled **Sherenga 1**). The latter is shown at the top of [Figure 21](#).

To use the Spectrum Matcher

You use the Spectrum Matcher to search unknown spectra against another set of library spectra on the Spectrum Mill server. The library spectra may reside in a **Collections** directory, or any other data directory on the server.

You can also use Spectrum Matcher to compare spectra acquired with different acquisition methods to evaluate any improvements, and to evaluate the quality of spectra using the spectral quality filters.

To create a library in your collections directory, see [“To copy spectra to your collections directory”](#) on page 190.

To search the spectra

- 1 On the Spectrum Mill home page, click the **Spectrum Matcher** link.
- 2 Check that you see the page shown in [Figure 23](#).
- 3 Set the **Search mode** to either **Identity** or **Precursor mass shift**.
 - An **Identity** search generates a **specMatcherSummary.ID.#.txt** file.
 - A **Precursor mass shift** search generates a **specMatcherSummary.HM.#.txt** file. After the search completes, you can return to the Spectrum Matcher page and click **Mass shift histogram from last search** to generate a **Parent Shift Histogram** that lists the number of times each parent shift was observed.
- 4 Click **Select** to select the **Query Set** data directory; then click **Select** to select the **Library Set** data directory.
- 5 Fill in the options. They are self-explanatory or are explained in the online Help.
- 6 Click the **Match** button.

1 Processing MS/MS Data Interactively

Iterative Workflow – Alternative Interactive Processing

- 7 If you have a large data set and you see an error that the CGI input string is too long, use the **Search result files** box under **Query Set** to search a subset of the file names.

For example, you can search just a single data file, typed into the Search result files box.

- 8 After time for processing, check that you see results like those shown in Figure 24 (which shows results of an **Identity** search).

Agilent Spectrum Mill - Spectrum Matcher - (none)

Spectrum Mill | Match Settings | Spectrum Summary | Protein/Peptide Summary | *de novo* | Help

Match Spectra

Match Save As... Load...

Instrument: Agilent ESI Q-TOF

Search Criteria

Search Mode
Search mode: Identity

Spectral Quality Filtering (instrument-specific peak detection used in Extractor)
☐ Sequence tag length: > 3
☐ Precursor isotope quality XIC's (Chi2 vs. Averagine): > 0.7
☐ Precursor Isolation Purity: > 70 %

Matching Tolerances
Minimum matched peak intensity: 50 %
Precursor m/z: +/- 50 ppm
Product m/z: +/- 20

MS/MS Peak Detection (over-ride instrument-specific peak detection for matching)
Minimum S/N: 2 < 10 Peaks (most intense ¹²C)
Minimum # of peaks: 4

Data Sets

Query Set	Library Set
Select ... ExampleData/Agilent/Q-TOF/Yeast/Y1	Select ... ExampleData/Agilent/Q-TOF/Yeast/Y2
Validation state: spectrum-not-marked-sequence-not-validated	Validation state: spectrum-not-marked-sequence-not-validated
Search result files: *.pk1	Search result files: *.pk1

Figure 23 Spectrum Matcher

Agilent Spectrum Mill - Spectrum Matcher - (none)

Spectrum Mill	Match Settings	Spectrum Summary	Protein/Peptide Summary	de novo	Help
		Query	Library		
Directory:	ExampleData/Agilent/Q-TOF/Yeast/Y1		ExampleData/Agilent/Q-TOF/Yeast/Y2		
Validation State:	spectrum-not-marked-sequence-not-validated		spectrum-not-marked-sequence-not-validated		
# spectra (VS filtered):	12909		14766		
# spectra (SQ filtered):	12909		14766		

Results will be written to specMatcherSummary.ID.3.txt
Fragment mass tolerance: 20.0000
Parent mass tolerance: 50.0000
Search Type: identity
X tolerance units: 1
Opened output file: E:\SpectrumMill\msdataSM\ExampleData\Agilent\Q-TOF\Yeast\Y1\specMatcherSummary.ID.3.txt

Matched Spectra

Score	SPI (%)	# Query Peaks	Qry num	query file	Query Parent m/z	Lib Parent m/z	Delta Parent m/z	Query Parent MH ⁺	Lib Parent MH ⁺	Delta Parent MH ⁺	Lib num	Library file
1	8.0	82.4	10	1 yeast-1.3679.3679.3.pkl	275.1493	275.1526	-0.0033	823.4324	823.4421	-0.0098	0	yeast-2.3678.3678.3.pkl
2	5.5	66.3	10	1 yeast-1.3679.3679.3.pkl	275.1493	275.1634	-0.0140	823.4324	823.4745	-0.0421	3	yeast-2.2932.2932.3.pkl
3	8.3	82.8	10	3 yeast-1.2365.2365.3.pkl	275.1568	275.1612	-0.0044	823.4547	823.4678	-0.0131	1	yeast-2.2451.2451.3.pkl
4	5.7	51.4	10	9 yeast-1.4961.4961.3.pkl	275.5047	275.5044	0.0003	824.4986	824.4976	0.0010	5	yeast-2.5036.5036.3.pkl
5	5.3	52.1	10	13 yeast-1.2423.2423.3.pkl	276.1620	276.1603	0.0017	826.4704	826.4653	0.0051	7	yeast-2.2690.2690.3.pkl
6	5.3	60.9	10	13 yeast-1.2423.2423.3.pkl	276.1620	276.1645	-0.0025	826.4704	826.4777	-0.0074	8	yeast-2.2524.2524.3.pkl
7	8.4	82.7	10	16 yeast-1.10460.10460.3.pkl	276.4826	276.4823	0.0003	827.4321	827.4311	0.0009	9	yeast-2.11715.11715.3.pkl

Figure 24 Spectrum Matcher results

To examine the results

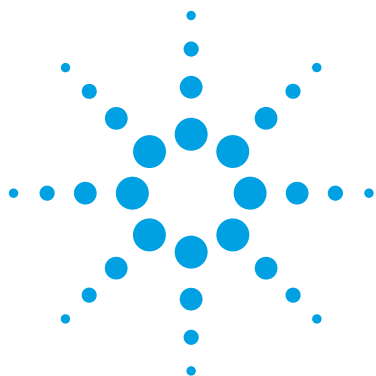
Click the links in the report to view and compare the spectra:

- Click a number link on the left side of the report to view both the query spectrum and the library spectrum in side-by-side Spectrum Viewer windows at the bottom of the page.
- Click a **query file** link to view the query spectrum in the Spectrum Viewer at the bottom left side of the page.
- Click a **library file** link to view the library spectrum in the Spectrum Viewer at the bottom right side of the page.

You cannot view these results with the Protein/Peptide Summary page.

1 Processing MS/MS Data Interactively

Iterative Workflow – Alternative Interactive Processing



2 Reviewing and Validating MS/MS Data

Setting Up Autovalidation	61
Setting Up the Protein/Peptide Summary Page	69
Reviewing and Validating Results Interactively	76
Protein/Peptide Summary Displays	90
Identification Statistics and False Discovery Rate	103

In this chapter you learn in detail how to review and validate database search results from MS/MS analyses of peptides and protein digests. Validation means that you accept the database search results for a particular MS/MS spectrum or group of spectra.

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. You can then subject results that are *not* validated to subsequent rounds of searches (against other databases or in variable modifications mode, for example). You can summarize results that *are* validated in a results table.

To segregate the valid search results, the software keeps track of both the spectra and their interpretations in a coordinated way. Furthermore, the software permits spectra to be segregated according to quality (via Spectrum Summary) without regard to the validation states of their database interpretations. Because the software keeps track of which spectra you have segregated as “good” and which database interpretations you have designated as “valid,” you can intelligently apply processing steps to individual groups of spectra.

In this chapter, you learn how to use the MS/MS Autovalidation page to quickly validate the search results with high scores. You learn how to use the Protein/Peptide Summary page to manually review and validate search results that have not been autovalidated. You learn how the various Protein/Peptide Summary data displays are organized to provide insight into different types of



2 Reviewing and Validating MS/MS Data

scientific problems. You also learn how to manipulate results in the Spectrum Viewer, an indispensable tool to visualize the sequence information contained in MS/MS spectra.

If you want a quick introduction to this subject matter, without a lot of details, see the relevant sections of [Chapter 1](#), “Processing MS/MS Data Interactively”.

Setting Up Autovalidation

Automatic results validation saves time and lets you use different threshold strategies to meet the data quality needs of your experiment.

Step 1. Select Data Directory and/or parameter file

- 1 Navigate to the **MS/MS Autovalidation** page, as shown in [Figure 25](#).

You navigate to this page from the Spectrum Mill Home page, the MS/MS Search page or the Protein/Peptide Summary page.

Figure 25 Default Autovalidation page: Auto thresholds strategy - Peptide mode

- 2 Select the **Data Directory** for which you want to validate results.
- 3 To use an already existing set of parameters, click **Load**.

Step 2. Select the Strategy and Mode to autovalidate results

For more detailed explanations of the strategies/modes and false discovery rate (FDR), click the Validation Parameters blue bar.

To use the Auto thresholds strategy

You can use this strategy in a normal workflow, but not an iterative one. After validating in peptide mode first (always first), you have the option of refining the validation with the Protein polishing mode.

Peptide mode This strategy optimizes the score threshold and the delta R1-R2 threshold to reach a target FDR.

- 1 Select the **Peptide** mode, if necessary. (Figure 25)
- 2 Type in a new **max FDR** if necessary.
1.2% was chosen because it typically produces an actual FDR of 1%.
- 3 Select a filter of interest.
- 4 Move to “[Step 3. Save parameters and validate data](#)” on page 68.

Protein polishing mode (optional) This strategy optimizes the score threshold and the delta R1-R2 threshold to reach a target FDR. Use this mode to remove suspicious proteins only after you have autovalidated in Peptide mode.

- 1 Select the **Protein polishing** mode. (Figure 26)
- 2 Change the parameters as you intend.
A **Minimum protein score** of 0 and a **maximum protein FDR** of 1.0% is typically used for Agilent Q-TOF data.
See the Online Help for explanations of each field.
- 3 Move to “[Step 3. Save parameters and validate data](#)” on page 68.

The screenshot shows the 'Autovalidation' window in the Spectrum Mill Workbench. At the top, there is a 'Help' button and a row of buttons: 'Validate Files' (highlighted in green), 'Queue request', 'Undo Last', 'Clear All', 'Save As...', and 'Load...'. Below this is the 'Data Directories' section with a 'Select ...' button and a list box containing 'ExampleData/Agilent/Q-TOF/Yeast/Y1'. To the right of the list box, there is a 'Fragmentation mode:' dropdown set to 'All' and a 'Search result files:' text box containing '*.spo'. The 'Validation Parameters' section is below, featuring three radio buttons for 'Strategy': 'Fixed thresholds', 'Auto thresholds' (selected), and 'Auto thresholds - discriminant'. The 'Mode:' dropdown is set to 'Protein polishing', with a note 'Auto determine using score, delta R1-R2 thresholds to reach a target FDR'. Other parameters include a checked box for 'Group proteins across all directories', a 'Minimum number of directories protein group is observed in:' field set to '1', a 'Minimum protein score:' field set to '0', and an 'Automatically raise minimum protein score to yield maximum protein FDR:' field set to '1.0 %'.

Figure 26 Auto thresholds strategy - Protein polishing mode

To use the Auto thresholds - discriminant strategy

You can use this strategy in either a normal workflow or an iterative one. It is typically used in an iterative workflow. After validating in peptide mode first (always first), you have the option of refining the validation with the Protein Polishing mode.

To use this strategy, the data must have been searched with **Discriminant scoring** set to a valid discriminant coefficients file.

Peptide mode This strategy optimizes the discriminant score thresholds until a target FDR is reached.

- 1 Select the **Peptide** mode. (Figure 27)
- 2 Click **Local**, and type an **FDR %**.

The Auto thresholds - discriminant strategy with Local FDR % marked in the Peptide mode is the only Auto thresholds strategy that can be used in iterative workflows. You cannot use this strategy with Global % marked in an iterative workflow.

- 3 Move on to “Step 3. Save parameters and validate data” on page 68.

2 Reviewing and Validating MS/MS Data

Setting Up Autovalidation

The screenshot shows the 'Validation Parameters' dialog box. Under the 'Strategy' section, 'Auto thresholds - discriminant' is selected with a radio button. The 'Mode' dropdown is set to 'Peptide'. A red text label reads 'Auto determine using discriminant score thresholds to reach a target FDR'. Under the 'Local' section, 'Local' is selected with a radio button, and the 'FDR' is set to '1 %'.

Figure 27 Auto thresholds - discriminant strategy - Peptide mode

Protein polishing mode (optional) This strategy optimizes the discriminant score threshold and the delta R1-R2 threshold to reach a target FDR. Use this mode to remove suspicious proteins only after you have autovalidated in Peptide mode.

- 1 Select **Protein polishing** mode. (Figure 28)
- 2 Change the parameters as you intend.
A **Minimum protein score** of 0 and a **maximum protein FDR** of 1.0% is typically used for Agilent Q-TOF data.
- 3 Move on to “Step 3. Save parameters and validate data” on page 68.

The screenshot shows the 'Validation Parameters' dialog box with 'Protein polishing' selected in the 'Mode' dropdown. The 'Strategy' remains 'Auto thresholds - discriminant'. A red text label reads 'Auto determine using discriminant score thresholds to reach a target FDR'. The 'Group proteins across all directories' checkbox is checked. The 'Minimum number of directories protein group is observed in:' is set to '1'. The 'Minimum protein score:' is set to '20.0'. The 'Automatically raise minimum protein score to yield maximum protein FDR:' is set to '0.0 %'. The 'Peptide FDR for validated proteins:' checkbox is checked, and the 'FDR' is set to '1.0 %'.

Figure 28 Auto thresholds - discriminant strategy - Protein polishing mode

To use the Fixed thresholds strategy

This strategy summarizes results by protein and peptide, uses fixed thresholds for validation and can calculate a false discovery rate (FDR) using reversed hits. If you want the program to calculate an FDR using reversed hits, you must have marked the **Calculate reversed database scores** check box on the MS/MS Search page.

The two modes for this strategy are Protein details and Peptide. For high quality results you autovalidate first with Protein details mode and then with Peptide mode. These are not optional steps.

You can use both of these modes in an iterative workflow.

While you can use the Fixed thresholds strategy, the Auto thresholds strategy is preferred, especially if you want to validate using FDR.

Protein details mode Using this mode is the first step in a complete validation.

- In the Protein details mode, the software summarizes results by protein, and considers all the peptides that belong to a given protein.
 - Using the default scoring, individual peptides typically have scores greater than 6 to 12 (depending on charge state), and the cumulative protein score typically is greater than 20.
- 1 Select **Protein details** mode (Figure 29 on page 66).
 - 2 Change the **Minimum protein score** if necessary.
 - 3 If the data files from a single sample occupy multiple directories, mark the check box for **Group proteins across all directories**.
 - 4 If you intend to filter the data, click **Fixed precursor mass error** and type in the limits.
 - 5 Change the **Protein Rules** thresholds, if necessary.
 - 6 Move on to “Step 3. Save parameters and validate data” on page 68.

Peptide mode For this strategy, using Peptide mode is the second step for a complete autovalidation. You can also use this mode in an iterative workflow.

- In this mode, the software summarizes results by peptide. Even if it finds only a single peptide that corresponds to a protein, it validates the search results if the peptide score is high enough.
 - Using the default scoring, individual peptides typically have scores greater than 11 to 15 (depending on charge state). This score threshold is higher than in the **Protein details** mode, where you have the additional assurance of knowing you have identified more than one peptide per protein.
- 1 Select the **Peptide** mode (Figure 30 on page 67).
 - 2 Enter abbreviations for the amino acids you intend to allow or disallow.
 - 3 Select filters you intend to use.

If you wish to autovalidate only the peptides in a given pI range, select **Auto OGE/IEF Peptide pI**, or select **Filter Peptide pI** and type in a pI range.

2 Reviewing and Validating MS/MS Data

Setting Up Autovalidation

If you ran an OFFGEL fractionation, and the data files from all pI fractions are in the same directory, then use the **Search result files** filter to select the specific file(s) for which that pI range applies.

- 4 Change the **Peptide Rules**, if necessary.
- 5 Move on to “Step 3. Save parameters and validate data” on page 68.

NOTE

If you wish to use the pI filter for modified peptides, ask your server administrator to first verify that the pK of the modified amino acid is specified in **smconfig.std.xml** or **smconfig.custom.xml**.

Agilent Spectrum Mill - MS/MS Autovalidation - YourName\AV-Q-TOF-OGE

Help

Validate Files ☐ Queue request

Data Directories

☒ ExampleData\Agilent\Q-TOF\OGE

Fragmentation mode:

Search result files:

Validation Parameters

Strategy: ☒ Fixed thresholds ☐ Auto thresholds ☐ Auto thresholds - discriminant

Mode: Fixed score thresholds, calculate an FDR

☐ Group proteins across all directories

Minimum number of directories protein group is observed in:

Minimum protein score:

☒ Calculate FDR using reversed hits Min Sequence Length:

Filtering ☒ Fixed range for all runs

☒ None (ppm) ☐ Fixed precursor mass error Low High ppm

Protein Rules

Rule	Precursor Charge	Score Threshold	% SPI Threshold	Fwd - Rev Score Threshold	Rank 1-2 score Threshold
1.	2	6.0	60.0	☐	☐
2.	1	6.0	70.0	☐	☐
3.	3	8.0	70.0	☐	☐
4.	4	8.0	70.0	☐	☐
5.	5	12.0	70.0	☐	☐
6.	2	6.0	90.0	☐	☐

Figure 29 Fixed thresholds strategy - Protein details mode

Agilent Spectrum Mill - MS/MS Autovalidation - YourName\AV-Q-TOF-OGE

Help

Validate Files ☐ Queue request

Data Directories

☒ ExampleData\Agilent\Q-TOF\OGE

Fragmentation mode:

Search result files:

Validation Parameters

Strategy: ☒ Fixed thresholds ☐ Auto thresholds ☐ Auto thresholds - discriminant

Mode: **Fixed score thresholds, calculate an FDR**

☒ Calculate FDR using reversed hits **Min Sequence Length:**

Required AAs: **Disallowed AAs:**

Filtering **Automatic variable range for each run** **Fixed range for all runs**

☒ None (ppm) ☐ Auto precursor mass error ☐ Fixed precursor mass error Low High ppm

☒ None (SC/pl) ☐ Auto SCX Solution Charge, pH3 ☐ Fixed Solution Charge Low High

☐ Auto OGE/IEF peptide pl ☐ Fixed peptide pl Low High

Peptide Rules

Rule	Precursor Charge	Score Threshold	% SPI Threshold	Fwd - Rev Score Threshold	Rank 1-2 score Threshold
1.	2	8.0	70.0	☐	☐
2.	1	7.0	70.0	☐	☐
3.	3	9.0	70.0	☐	☐
4.	4	9.0	70.0	☐	☐
5.	5	15.0	70.0	☐	☐

Figure 30 Fixed thresholds strategy - Peptide mode

Step 3. Save parameters and validate data

Before you validate data, you can choose to save any parameters you changed to the same or different parameter file and you can choose to place this command in the queue.

- 1** To save the settings you changed to a parameter file, click **Save As**, name the file, then click **Save**.

This file can now be used as a task in workflow automation.

- 2** To delay execution of the command or have it be part of the interactive automation process, mark **Queue request** to place the command in the queue.

To see results right away, do not mark Queue request.

- 3** Click **Validate Files**.

You see a validation summary that lists the hits and spectra that have been validated.

- 4** If you have completed autovalidation, move on to “[Setting Up the Protein/Peptide Summary Page](#)” on page 69.

If you intend to autovalidate with the second mode in a strategy, repeat “[Step 2. Select the Strategy and Mode to autovalidate results](#)” on page 62 with the second mode.

- Auto thresholds strategy - “[Protein polishing mode \(optional\)](#)” on page 62
- Auto thresholds - discriminant strategy - “[Protein polishing mode \(optional\)](#)” on page 64
- Fixed thresholds strategy - “[Peptide mode](#)” on page 65

Setting Up the Protein/Peptide Summary Page

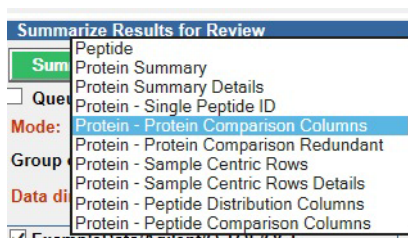
To set up the Protein/Peptide Summary page, you select options in three categories:

- **Summarize Results for Review** - Select data directory and results display mode.
- **Validation and Sorting** - Select validation, sorting and filtering options.
- **Review Fields** - Mark check boxes to choose columns to include in the report.

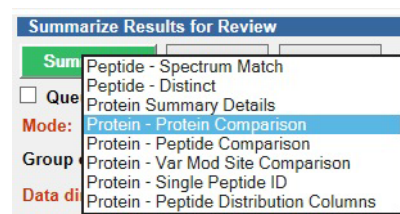
Then you click the **Summarize** button to summarize results.

Step 1. Set data directory and mode

- 1 Navigate to the Protein/Peptide Summary page.
- 2 Click the **Select...** button to select the data directory where your files reside.
- 3 To upload a set of existing parameter values, click **Load**.
- 4 Click the **Mode** drop-down list and select one of the summary modes shown in [Figure 31](#). Use [Table 2](#) on page 42 to guide you.



"Classic" modes from some previous releases



Modes from this release

Figure 31 Selecting the data display mode

NOTE

If you updated from Spectrum Mill B.04.01 or prior, and you want to configure the "classic" data display modes shown on the left of [Figure 31](#), see the help for Spectrum Mill Server Administration. The current modes are shown on the right.

For more details on the displays, see “Protein/Peptide Summary Displays” on page 90.

Step 2. Set filtering, sorting, and validation parameters

In this step, you determine which data are displayed, how the data are sorted, and the initial validation state of the data. As you set parameters in this section, it is useful to keep in mind the following:

- Filters limit the data that is viewed. For example, they can be used to show only results where search scores exceed a chosen value.
- Sorting parameters determines the order of the data that is viewed.
- Validation presets determine the initial settings of the validation states for the data that is viewed. You can always change these presets after you review the data. They are not made a permanent part of the data record at this point.
- A validation state represents your judgment as to whether a particular data interpretation (e.g., a search result) is satisfactory.

Validation and Sorting

Filter results by: valid

Validation preset: none

Protein grouping method: 1 shared, expand subgroups

Sort proteins by: Score

Filter by protein score: > 0

Sort peptides by: Score

Filter peptides by:

Score: > 0 % SPI: > 0

Required AAs: any

Disallowed AAs: none

Accession #'s:

Figure 32 Validation, sorting and filtering parameters - Protein Summary Details mode

- Set the validation, sorting, and filtering parameters (Figure 32).

To set these parameters, refer to the instructions below and the *Online Help* by clicking the Validation and Sorting blue bar.

Note that the specific options that are available depend upon the mode you selected earlier.

- Select the **Filter results by** validation filter to determine which data are displayed: sequence-not-validated, valid, good-spectrum-sequence-not-validated or all.
- Select the **Validation preset** to determine how you categorize your data upon initial display: none, valid, reset or status.
- Select the **Protein-grouping method** to determine how proteins are grouped in certain protein summary modes: 1 shared peptide or 1 shared peptide, expand subgroups.
- From the **Sort proteins by** list, select a data sorting method for your report.

Choices include score, number of peptides, protein molecular weight, category and function.

- Set a **Filter by protein score**.

Scores >15-18 almost always represent valid results for Q-TOF data.

- From the **Sort peptides by** list, select a sorting method.

Note that when you sort by accession number, the sort is alphabetical rather than numerical. This is because some databases do not have strictly numeric accession numbers.

- Set a **Filter by peptide score**.

For ion trap data, peptide scores greater than 13 almost always represent valid results. Scores below 6 generally represent poor results. For Agilent Q-TOF data, you search with a narrower mass tolerance; so there is a better chance that lower-scoring results are valid. It is not unusual for a score of 5 to represent a valid result, but only if the peptide is short or in low abundance.

- Set a **%SPI**, or filter for scored peak intensity.

2 Reviewing and Validating MS/MS Data

Setting Up the Protein/Peptide Summary Page

For ion trap data, values greater than 70 (or 60% for doubly-charged precursor ions) generally represent high-quality results. For Agilent Q-TOF spectra, % **SPI** below 60 may represent a poor interpretation. The lower the value, the greater the likelihood that the MS/MS spectrum contains significant peaks not explained by the database search results.

- If you have already validated the data and wish to see all valid results, instead of using the settings in the three bullets above, set these filters all to **0**: **Filter by protein score**, **Filter by peptide score**, and **% SPI**.
- From the **Required AAs** list select the amino acid(s) that must be present in the peptides that are shown. To disable, select **any**.
- From the **Disallowed AAs** list select the amino acid(s) that cannot be present in the peptides that are shown. To disable, select **none**.
- Type a range of peptide pI values. Mark **All** to ignore the range.
This field appears only when you have marked **Peptide pI** in the Review Fields.
- Type or paste a list of accession numbers in various formats (space-separated, separated by '|', comma-separated, etc.). To disable, leave the box blank.

NOTE

Filters that are more stringent produce higher confidence results. Set the score filters in a way that makes sense for your study. For some studies, it may be better to provide fewer results, albeit with greater probability of being correct. For others, it may be better to provide more results and take the risk that some false positives could be reported.

Your system administrator can configure the options for **Required AAs** and **Disallowed AAs**.

If you wish to use the pI filter for modified peptides, ask your server administrator to first verify that the pK of the modified amino acid is specified in **smconfig.std.xml** or **smconfig.custom.xml**.

Step 3. Choose review fields

In this step you mark the check boxes for data you wish to display.

Click the Review Fields blue bar to see the descriptions of all the review fields for every data display mode.

Note that some fields are unavailable in certain data display modes. For example, fields that pertain only to peptides are available only in the modes that display peptide data.

Review Fields		
☒ Filename	☒ Sequence ☐ b/y map	☒ Intensity
☒ Score	☐ Rev Sequence ☐ Rank 2	☐ DEQ ratios ☐ Invert
☐ FDR (Discriminant)	☐ VML sequence	☐ Reporter Ratios: iTRAQ 4 ▾
☐ Fwd-Rev score	☐ Prec Av Chi ² ☐ Isol Pur	Control: 115 ▾ 116 ▾ 117 ▾
☐ Rank 1-2 score	☐ Ret time, width	☐ Intensities
☒ SPI (%)	☐ Precursor m/z ☒ MH ⁺	Correction Method: Apply ▾
☐ Unmatched ions	☐ Delta mass ☐ Pep pI	☐ Modification names
☐ Var mod sites	☐ Protein MW ☐ Prot pI	☐ N-term ☐ C-term ☐ Cysteines
☐ VML score s t y ▾	☐ Species	☐ Fragmentation mode
☐ Solution charge	☒ Accession #	☐ Max tag length ☐ Longest tag
☐ Ion mobility	☒ Protein name	☐ # Backbone Cleavages
☐ Start AA position	☐ Category:	☐ # b/y pairs

Figure 33 Fields for display in Peptide - Spectrum Match page

1 Choose from items shown in [Figure 33](#).

The marked check boxes are ones you would typically use.

2 Mark any others that suit your experiment.

Here are instructions for changing a few.

- Mark **Fwd-Rev score** to see the difference between scores for top hits from forward and reversed database searches.
If you obtain similar scores for both searches, there is an increased likelihood of an incorrect assignment.
- Mark **Rank 1-2** score to see the difference between rank 1 and rank 2 database search scores.
- Mark **Var mod sites**, **VML score** and/or **VML sequence** to display sites of variable modifications and amino acid substitutions, the sequences around the sites, and the scores.
- Mark **Peptide pI** to display peptide pI and enable filtering by peptide pI.
- Mark **DEQ ratios** to calculate and display ratios for differential expression quantitation, such as light/heavy ratios for ICAT reagents or other reagents that use isotopic labels.

See “To calculate DEQ ratios for isotopic labels using the Protein/Peptide Summary page” on page 136 of Chapter 5.

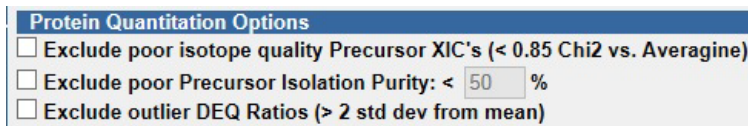
- Mark **Reporter Ratios** and from the iTRAQ/TMT list select the tag type used for quantitation. Mark **Intensities** to display intensities for the marker ion masses.
- Mark **Ratios control**, then from the Ratios control list, select the control ions to use for ratio calculations. To select multiple control ions, select using the **Ctrl** key.
- Mark **Fragmentation mode** to show which fragmentation mode, for example, collision-induced dissociation (CID) or electron transfer dissociation (ETD), was used for searching the results.

Step 4. Select Protein Quantitation Options

- Select one or more of the options in [Figure 34](#) in order to filter out data for protein quantitation.

Note that the third option appears only when you mark the DEQ ratios check box in the Review Fields panel.

For complete descriptions of each of these options, click the Protein Quantitation Options blue bar.



Protein Quantitation Options

- ☐ Exclude poor isotope quality Precursor XIC's (< 0.85 Chi2 vs. Averagine)
- ☐ Exclude poor Precursor Isolation Purity: < 50 %
- ☐ Exclude outlier DEQ Ratios (> 2 std dev from mean)

Figure 34 Protein Quantitation Options when DEQ ratios are marked in Review Fields

Step 5. Save parameters and summarize data

Before you summarize data, you can save the changed parameters, place the command in the queue and/or choose to export the summarized data to an Excel spreadsheet or indirectly to a MassHunter database.

- 1 To save the settings you changed to a parameter file, click **Save As**, name the file, then click **Save**.

This file can now be used as a task in workflow automation.

- 2 To delay execution of the Summarize command or to place the command in a queue to execute later as part of interactive automation, mark **Queue request**.

- 3 In some modes, you may export data. To export protein or peptide data to an Excel spreadsheet, mark **Excel export**, or

To export peptide data indirectly (through a CSV file) to an AMRT (Accurate Mass and Retention Time) database, mark **AMRT export**.

To export protein results to Mass Profiler Professional, choose the Protein-Protein Comparison mode, and select MPP Generic export or MPP APR export. MPP APR export is supported only for data acquired with Agilent instruments.

Which option is available depends on the data display mode.

Instead of check boxes, Protein-Protein Comparison mode has a drop-down menu that includes:

- No export
- Excel export
- MPP Generic export
- MPP APR export

- 4 Click **Summarize** to display results in tabular format.

Reviewing and Validating Results Interactively

After you have completed a workflow of searches and autovalidations you can manually check to see if lower quality results are, in fact, valid.

To display results in preparation for manual validation

- 1 Navigate to the Protein/Peptide Summary page. If this page is already open, click the **Summary Settings** button to display the parameter settings at the bottom of the page.
- 2 Make sure **Mode** is set to **Peptide - Spectrum Match**.
- 3 If your **Data Directories** are not selected, click the **Select...** button to select the folder or folders that contain your data files.
- 4 Set parameters for **Validation and Sorting** as shown in [Figure 35](#). These settings locate results that have not yet been validated. See the following steps for more details.

Figure 35 Settings to prepare for a manual validation

- 5 From the **Filter results by** list, select **sequence-not-validated**.
- 6 Use the following guidelines for setting the **Filter peptides by** parameters:

High-resolution mass spectrometers

Table 3 Q-TOF and other high-resolution mass spectrometers: settings for **Filter peptides by**

Peptide score	Quality	Peptide fragmentation	Likelihood of valid interpretation
> 15	Outstanding, well-assigned	Thorough	When combined with % SPI of 60 or greater, very likely to be valid
> 9	Good	Substantial	When combined with % SPI of 60 or greater, likely to be valid
> 5	Mixed Quality	Moderate	Review results to determine whether interpretation is valid; more likely to be valid than same scores for ion trap data
< 5	Generally poor	Less	Not likely to be valid

Because the mass accuracy is better than with ion trap data, you search with a narrower mass tolerance, so there is a better chance that lower-scoring results are valid.

Low-resolution mass spectrometers

Table 4 Ion trap and other low-resolution mass spectrometers: settings for **Filter peptides by**

Peptide score	Quality	Peptide fragmentation	Likelihood of valid interpretation
> 15	Outstanding, well-assigned	Thorough	When combined with % SPI of 70 or greater, very likely to be valid
> 10	Good	Substantial	When combined with % SPI of 70 or greater, likely to be valid
> 5	Mediocre	Few ions, look for mostly b's or y's	Review results to determine whether interpretation is valid; less likely to be valid than same scores for Q-TOF data
< 5	Poor	Little to none	Not likely to be valid

- If you have Agilent Q-TOF data, set the **Score** to **5**.
Because the mass accuracy is better than with ion trap data, you search with a narrower mass tolerance, so there is a better chance that lower-scoring results are valid.
- If you have ion trap data, for the **Score** settings, see the guidelines above.
- **% SPI** (Scored Peak Intensity) is the percentage of the MS/MS peak-detected spectral ion current explained by the search interpretation. For Agilent Q-TOF spectra, **% SPI** below 60 may represent poor interpretations; so do not set lower values. For ion trap spectra, **% SPI** below 70 (or 60 for doubly-charged precursor ions) are more likely to represent poor interpretations.
- If you want to limit your data review, set the **Filter peptides by** parameters to higher values to filter out the poorer-quality results.

7 Set the **Validation preset** in a way that minimizes manual changes later:

- If you have Agilent Q-TOF data and you set **Filter peptides by** to a score of 7 or greater, select **valid**.
- If you have Agilent Q-TOF data and you set **Filter peptides by** to a value less than 7, select **reset**.
- If you have ion trap data and you set **Filter peptides by** to a score of 9 or greater, select **valid**.
- If you have ion trap data and you set **Filter peptides by** to a value less than 9, select **reset**.

8 In the **Review Fields**:

- a** Mark the check box for **Rank 1-2 score**. Verify that the check box for **Fwd-Rev score** is also marked. These parameters help you rule out false positives. Small differences between forward and reversed scores, or between rank 1 and rank 2 scores, may indicate an incorrect identification.
- b** Mark the check box for **# Backbone Cleavages**, which is the number of cleavages of the amino acid backbone that are represented in the spectrum for the assigned sequence. For CID spectra, these cleavages include the b- and y-fragments, and for ETD spectra, they include the c- and z-fragments. Other ions, such as neutral losses from b and y ions, are not included.
- c** If you used **Dynamic peak thresholding** in the search, mark the check box for **Unmatched ions**. These results tell you both the number of unmatched ions and the number of ions that were included in the search

score that you see in the report. If the latter number is less than 15, check that the peptide is either short or in low abundance. If not, be suspicious of the results.

NOTE

In the **Review Fields**, if you mark the check box for **Peptide pI**, then under **Validation and Sorting** you have the option to display only peptides that meet your pI criteria. This feature is useful if you need to validate peptides from OFFGEL fractionation.

If you wish to use the pI filter for modified peptides, ask your server administrator to first verify that the pK of the modified amino acid is specified in **smconfig.std.xml** or **smconfig.custom.xml**.

- 9 Click **Save As** to save settings to the original or a different parameter file, then click **Save**.

You can now use this parameter file as a “task” in the list of tasks for an automated workflow. See Chapter 3, “Automating Workflows”.

- 10 If you want to see results immediately, make sure the **Queue Request** check box is clear. If you want to queue the task as part of an *interactive automation*, mark the check box.

- 11 Click **Summarize**.

If the Queue Request check box is clear, you quickly see a protein/peptide summary table that lists the hits and spectra that have been validated. Queuing a summary is useful when you want the results saved to an html file that can be accessed via the Completion Log.

NOTE

Once you click the **Summarize** button, the button is disabled until the results appear. If you need to re-enable the button, click the **Summary Settings** button to reload the Protein/Peptide Summary page.

- 12 After this task is complete, make sure that you see a display similar to that shown in [Figure 36](#) on page 80. Note that in this example, where the score threshold was set to only 5, results have been categorized as **R** for **reset**. These results require manual review to avoid the risk of false positives.
- 13 Manually review the results, as described in the next section.

To manually review and validate results

- 1 Depending on how you set validation and filtering parameters, you will want to manually review some or all of the results. Start by examining the overall results, as shown in [Figure 36](#).

Agilent Spectrum Mill - Protein/Peptide Summary - Defaults\pep_sum_defaults

Spectrum Mill	Summary Settings	AutovValidation	MRM Selector	MS/MS Search	Spectrum Summary	Workflows	Tool Belt	Help
---------------	------------------	-----------------	--------------	--------------	------------------	-----------	-----------	------

Perform Validation Results Shown Filtered by Validation Category: sequence-not-validated
Data Directory: ms\data\SMExampleData\AgilentQ-TOF\Yeas\Y1
hit table read 4782 files - SpecFeatures read
all hits read from tagSummary file num hits: 6157
original hits subset determined from files: 1375 Files: 1375 Hits: 622

Validation category	#	Filename	z Score	Fwd-Rev Score	Rank1-Rank2 Score	SPI (%)	# Backbone Cleavages	Missed Cleavages	#KR	Un-matched Ions	Spectrum Intensity	Sequence	Protein MW (Da)	Accession #	Protein Name
☐ V ☒ R	1	yeast-1.15446.15446.2	2	19.07	19.07	19.07	91.9	10	0	1	2 / 22	2.80e+005 (R) NESMLTFETSQIQSAK (D)	14452.9	P33331	Nuclear transport factor 2
☐ V ☒ R	2	yeast-1.19917.19917.2	2	18.16	10.42	18.16	90.1	7	0	1	3 / 22	9.25e+004 (K) DAVDDLESLEK (E)	110456.7	P40825	Alanine-tRNA ligase, mitochondrial
☐ V ☒ R	3	yeast-1.10420.10420.3	3	17.89	17.89	17.89	80.9	10	1	2	6 / 25	1.20e+005 (K) VINLDTNKDDDDGGIADGR (L)	17303.1	P53834	Hsp90 co-chaperone HCH1
☐ V ☒ R	4	yeast-1.10960.10960.2	2	16.26	16.26	16.26	83.9	7	0	1	5 / 21	1.40e+005 (R) SENFDVTFGQVGR (I)	11475.8	P11633	Non-histone chromosomal protein 6B
☐ V ☒ R	5	yeast-1.20340.20340.2	2	16.25	9.78	10.47	83.2	7	0	1	5 / 22	6.73e+004 (K) VEFLGGLDAIFPK (Q)	11028.3	P40554	Ubiquitin-related modifier 1
☐ V ☒ R	6	yeast-1.19903.19903.3	3	16.03	16.03	16.03	87.7	10	0	1	4 / 23	1.47e+005 (R) VDSFTNAFFAILEIPSK (D)	13460.9	P39111	V-type proton ATPase subunit F
☐ V ☒ R	7	yeast-1.4847.4847.4	4	15.37	15.37	9.03	82.1	7	1	2	7 / 22	3.18e+005 (K) TTDADSEKHNYGEQYNK (L)	7707.3	P01094	Protease A inhibitor 3
☐ V ☒ R	8	yeast-1.19544.19544.2	2	15.09	15.09	15.09	83.9	7	0	1	4 / 21	1.21e+005 (K) ELFPVDFPIGK (K)	62426.5	P40413	T-complex protein 1 subunit epsilon
☐ V ☒ R	9	yeast-1.19239.19239.2	2	14.98	6.72	14.98	87.7	5	0	1	9 / 25	1.47e+005 (R) IITLPTFILEK (K)	51758.2	Q02201	Oxysterol-binding protein homolog 6
☐ V ☒ R	10	yeast-1.20661.20661.2	2	14.46	14.46	14.46	89.7	8	0	1	2 / 17	3.91e+004 (K) IGVDAINELLEPIR (Q)	44247.4	P36421	Tyrosine-tRNA ligase, cytoplasmic
☐ V ☒ R	11	yeast-1.20403.20403.2	2	14.40	14.40	14.40	78.0	6	0	1	6 / 19	3.65e+004 (R) FGFFDILR (D)	32765.8	Q03835	Monothiol glutaredoxin-3
☐ V ☒ R	12	yeast-1.18231.18231.2	2	14.39	14.39	7.28	84.2	7	0	1	8 / 24	2.11e+005 (K) VISLEDLK (I)	39605.9	Q12449	Hsp90 co-chaperone AHA1

Figure 36 Peptide - Spectrum Match display

- 2 To review spectra, click links as described on the next few pages. As you encounter valid interpretations, select the **V** under the **Validation category** heading to change the validation state to **valid**. For guidelines on how to decide whether a database match is reasonable, see [“To use the Spectrum Viewer”](#) on page 83 and [Table 5](#), “Guidelines for manual results validation,” on page 87.
- 3 When you are satisfied that the validation states are set correctly, click the **Perform Validation** button. This saves your validation states for further reference.
- 4 Note that you quickly see lists of validated hits and spectra. These are cumulative and include both the new hits and spectra you just validated, as well as those you validated previously via autovValidation or manual validation.

NOTE

If you find you have designated valid results in error, and you have already clicked the **Perform Validation** button, you can reverse the action. Change the validation state(s) and click the **Perform Validation** button again.

To compare theoretical and experimental spectra

Your primary review tool is the Spectrum Viewer. In many cases, this gives you all the information you need to evaluate database search results. To display the Spectrum Viewer (shown in [Figure 37](#)), click a link under the # or **SPI** headers.

- If you click a link under the # header, the spectrum is displayed at the bottom of the page.
- If you click a link under the **SPI** header, the spectrum is displayed in a new window.
- For details on use of the Spectrum Viewer, see “To use the Spectrum Viewer” on page 83.

The Spectrum Viewer shows your extracted spectrum annotated with theoretical fragments from the top database search result.

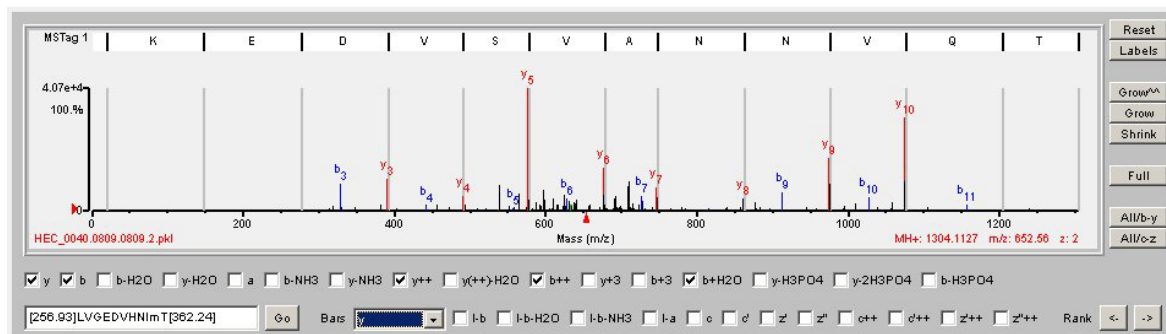


Figure 37 Spectrum Viewer, with y-ions highlighted using the **Bars** feature

To view MS/MS Search results

If you want more detailed information than is presented in the Spectrum Viewer, you view the MS/MS search details, as shown in [Figure 38](#).

- 1 To view these details, click a link under the **Filename** header shown in [Figure 36](#). Note that you may need to use your mouse to move the white page divider to see more than the spectrum.
- 2 Examine the MS/MS Search results, where you see up to five database hits with their scores, as well as the species for each protein. (While five is the default, you can increase the number of displayed hits with the **Max reported hits** setting on the MS/MS Search page.)

2 Reviewing and Validating MS/MS Data

Reviewing and Validating Results Interactively

- 3 Scroll through the **Detailed Results** to see how the ions were assigned for each of the database hits.
- The final **Score** from the database search is the sum of the scores for each of the ions. (The unassigned ions subtract from the score.)
 - The % **SPI** is the sum of the fractional intensities – **Frac.Inten. (% of TIC)** – for each of the assigned fragment ions.
 - For phosphorylated peptides, the **Rank 2**, **Rank 3**, etc. hits may show other forms of the phosphopeptide where the modification is on other sites. Examine the ions that constitute the evidence for those locations to judge which is the best interpretation. If there is ambiguity in the location of the modification, the various hits have similar scores.

NOTE

If you enable discriminant scoring, you also see the Discriminant Score in the Detailed Results.

MS/MS Search Results

Sample ID (comment): yeast-1.5485.5485.2.pkl
Search Time: 0 secs / 1 searches, 0.000 secs/search
Database searched: NCBI nr.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

Parameters used in Search

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)NEAIPEAEQVK(K)	1215.5852	79.9738	5.7	12588.8/5.43	Saccharomyces cerevisiae	6324463	Overexpression causes resistance to Zeocin; Yoi109v

Detailed Results

Rank	Score	SPI (%)	BCS	# Unmatched Ions	Variable sites	Sequence	MH ⁺ Calculated (Da)	MH ⁺ Error (Da)	MH ⁺ Error (ppm)	Protein MW/pI (Da)	Species	Accession #	MS-Digest Index #	Protein Name
1	8.71	59.9	4	10/24	T49t	(K)N E I A / t / P E / A E Q V K (K)	1215.5852	79.9738	5.7	12588.8/5.43	Saccharomyces cerevisiae	6324463	977	Overexpression causes resistance to Zeocin; Yoi109v

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	648.2831 ⁺²
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	0.53
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	5.32
Signal/Noise	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
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	1.50
Ion-type	a ₁		a ₂	b ₂ -H ₂ O	b ₂				b ₄ -H ₃ PO ₄	y ⁺⁺				y ₅	st ⁺²			y ₁₁ -H ₃ PO ₄ ⁺²	
Delta ppm	19.7		-17.5	-13.9	-7.8				-15.9	2.6				13.4	-11.8			1.5	1.5
	0.50																	1.50	1.50
	NR																	sty ⁺²	2.8
	19.7																		

Figure 38 MS/MS Search results

- 4 Click additional links shown in the MS/MS Search results in Figure 38:

- To view an MS Product listing that shows all theoretical fragment ions, click an amino acid sequence.
- To view information in the database you searched, click an accession number in the MS/MS Search report.

To view MS Digest results that show peptide masses from a theoretical digestion, click an MS Digest index number in the MS/MS Search report.

To use the Spectrum Viewer

Figure 39 shows the Spectrum Viewer, a tool to visualize the sequence information contained in MS/MS spectra and to evaluate spectral interpretations from MS/MS database searches or Sherenga *de novo* sequencing. To manipulate the Spectrum Viewer, see the descriptions below.

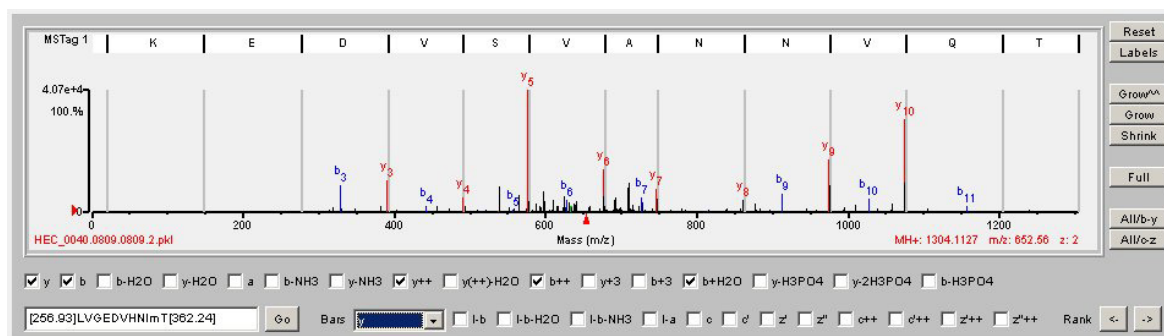


Figure 39 Spectrum Viewer, with y-ions highlighted using the **Bars** feature

Color-coding

In the Spectrum Viewer, the software color-codes the ions as follows:

- Red: C-terminal fragments (y-ions and z-ions)
- Blue: N-terminal fragments (b-ions and c-ions)
- Green: Precursor ions, neutral losses from precursor ions, immonium ions

To use the buttons

- Use the buttons on the right side of the Spectrum Viewer to manipulate the display. The buttons do the following:

Reset	Resets the spectrum to the original x- and y-axis values.
Labels	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, and no labels. To see the spectrum without the peak interpretations, click the red file name under the sequence at the top of the Spectrum Viewer. Now click the Labels button to turn the mass labels on and off. Note that you need to re-display the spectrum to see the interpretations again.
Grow^^	Expands the spectrum by ten times in the vertical axis
Grow	Expands the spectrum in the vertical axis
Shrink	Reduces the spectrum in the vertical axis
Full	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 means 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.
All/b-y	Toggles marking of check boxes for b- and y-ions in the first row under the spectrum and labels peaks accordingly. The toggle either marks all the check boxes or resets to defaults.
All/c-z	Toggles marking of check boxes for c- and z-ions in the second row under the spectrum and labels peaks accordingly. The toggle either marks all the check boxes or resets to a default subset. In addition, the All/c-z button labels precursor ions that have reduced charges, as are typically observed in ETD spectra. When you analyze ETD data, you must click the All/c-z button to see the appropriate labels for the fragment ions.

To use the features above the spectrum

- Spectra are annotated with interpretations based on the sequence shown in white. When both **MSTag** and **Sherenga** sequence bars are displayed, click the sequence for which you wish to see annotations.

- To see the spectrum without the peak interpretations, click the red file name under the sequence at the top of the Spectrum Viewer. Now click the **Labels** button to turn the mass labels on and off. Note that you need to re-display the spectrum to display the interpretations again.

To use the check boxes in the first row below the spectrum

- Mark check boxes for the ion types you want labeled in the spectrum. If you mark check boxes but do not see any labels, click the **Labels** button to turn the labels back on.

To use the features in the second row below the spectrum

- Use these features to manipulate the display. These features do the following:

Go To display a sequence above the spectrum, type the sequence in the box to the left and then click the **Go** button.

The sequence in the box is initially set to a default sequence. Mass gaps shown in brackets indicate portions of the spectrum where there was insufficient fragmentation to provide an amino acid sequence. You can enter mass gaps in the middle of the sequence as well as at the ends.

Note that in addition to single-letter capitalized abbreviations for the 20 amino acids, you can type the following lower-case abbreviations for modified amino acids:

Symbol	Modified amino acid
k	Carbamylated lysine
m	Oxidized methionine
q	Pyroglutamic acid (only at N-terminus of peptide)
s	Phosphorylated serine
t	Phosphorylated threonine
y	Phosphorylated tyrosine

2 Reviewing and Validating MS/MS Data

Reviewing and Validating Results Interactively

The variable modifications **kmqsty** are defined by default for the Spectrum Viewer. But if in MS/MS Search, you specified a different variable modification for K, M, Q, S, T, or Y (for example, guanidination of K), then that modification is used instead.

Use the **Rank** arrow buttons (<- and ->) to go from the sequences that were identified by MS/MS Search or Sherenga *de novo* Sequencing to the sequence that you typed. For MS/MS Search, the arrow buttons cycle between the peptide from the highest-scoring MS/MS search result and the sequence you typed. For Sherenga, the arrow buttons cycle through all the Sherenga result sequences. If you add a custom sequence, the software appends it to the list of Sherenga sequences that can be cycled.

- | | |
|--------------------------------------|--|
| Bars | Select an ion type to highlight. As shown in Figure 39 , this allows you to more easily visualize how the ions align with the amino acid sequence displayed at the top of the Spectrum Viewer. |
| Check boxes labeled l-x | Mark these if you want to label peaks with amino acid sequences combined with common losses. |
| Check boxes for c- and z-ions | Mark check boxes for the ion types you want to see labeled in the spectrum. If you mark check boxes but do not see any labels, click the Labels button to turn the labels back on. |
| Rank | Click arrows to display additional Sherenga results. Note that when both MS-Tag and Sherenga sequence bars are displayed in the Spectrum Viewer, you must first click the Sherenga sequence bar and make sure it turns from gray to white before you click the Rank arrows. |

To expand the x-axis

- Use the cursor to expand a portion of the spectrum in the x-axis. Move your mouse over the spectrum. When a crosshair is displayed, select the portion of the spectrum you wish to expand.
- Double-click the spectrum or click the **Reset** button to return to the original display.

To change the threshold for mass labeling

- The red triangle to the left of the y-axis indicates the threshold for peak labeling. Click anywhere on the y-axis to change this threshold.

Table 5 Guidelines for manual results validation

Guideline	Example
The best spectra are generally found by autovalidation, so do not expect to see spectra like the one to the right.	
- Loss of NH_3 occurs from R, K, Q and N residues. - Loss of H_2O occurs from S, T, E and D. - R, H, K and N are charge-bearing residues and increase the maximum charge state allowed for a peptide fragment. - Enhanced fragmentation may be observed at the following bonds: - His – Xaa - Xaa – Gly - Xaa – Ser	
The peptide bond on the N-terminal side of Pro is particularly labile (high local proton affinity) and usually results in an MS/MS spectrum dominated by the y-ion ending in Pro.*	
It is typical to lack fragmentation between the first two N-terminal amino acids (i.e., no b_1 or y_{x-1} ions).	

2 Reviewing and Validating MS/MS Data

Reviewing and Validating Results Interactively

Table 5 Guidelines for manual results validation (continued)

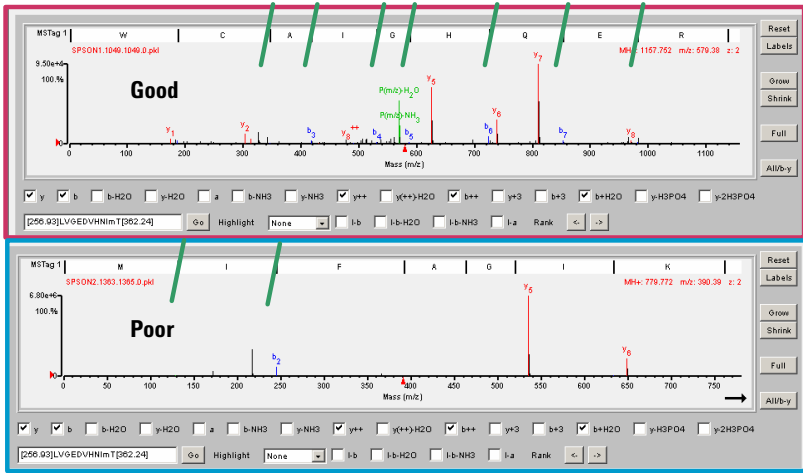
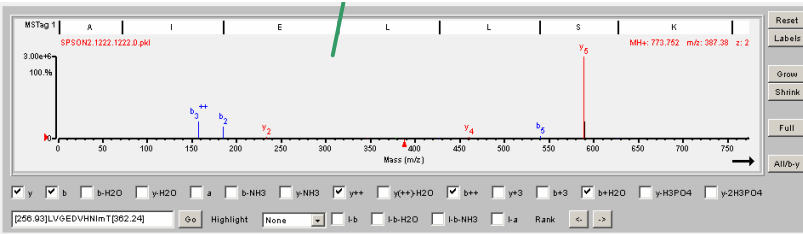
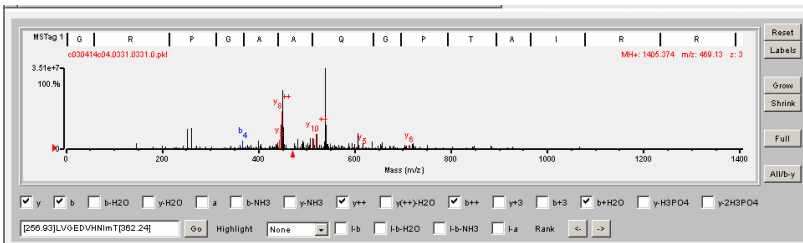
Guideline	Example
Look for coverage over a significant portion of the backbone. A series of consecutive b or y fragments is better than a random distribution.	
The peptide bond on the C-terminal side of acidic residues (D, E) may show enhanced cleavage. [†]	
Be suspicious of a clumped pattern that occurs frequently throughout the LC run (usually a background ion).	

Table 5 Guidelines for manual results validation (continued)

Guideline	Example
Be suspicious if no ions appear above the precursor ion (when precursor charge state is >1).	

Remember - When in doubt, *don't* validate.

- The goal is *no* false positives.
- Validating a questionable hit removes it from subsequent searches.

* Proton mobility scoring accounts for enhanced fragmentation around P.

† Proton mobility scoring accounts for enhanced fragmentation around D and E.

Protein/Peptide Summary Displays

The data display modes on the Protein/Peptide Summary page allow you to view data in a number of ways, so you can select the mode that best fits your experiment. The list of display modes is shown in [Figure 40](#). This section shows examples of data displayed in each mode. The modes are discussed in the order that they appear in the drop-down list.

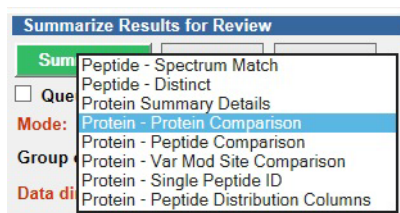


Figure 40 Data display modes

Peptide - Spectrum Match

Select this mode to display results by peptide. This mode is particularly useful to validate the database identifications from each MS/MS spectrum. The output looks like that in [Figure 41](#).

Note that 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)

#	Filename	z	Score	SPI (%)	Spectrum Intensity	Sequence	Accession #	Protein Name
1	yeast-1.17537.17537.2	2	25.62	96.4	5.71e+005	(K) LLYVDYQELPAEAINR (L)	Q08911	Formate dehydrogenase 1
2	yeast-1.21279.21279.3	3	25.49	94.5	7.50e+006	(R) YGASAGNVGDEGGVAPNIQTAAEALDLIVDAIK (A)	P00924	Enolase 1
3	yeast-1.20324.20500.3	3	25.12	98.5	3.74e+006	(K) VINDAFGIEEGLMTTVHSLTATQK (T)	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3
4	yeast-1.18605.18605.3	3	24.65	96.9	2.36e+006	(K) FADGFLIHSGQPVNDFIDTATR (H)	P05750	40S ribosomal protein S3
5	yeast-1.17684.17882.2	2	24.45	100.0	4.59e+007	(R) VFTVDVSVVDLTVK (L)	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1
6	yeast-1.18150.18150.4	4	24.43	93.8	1.93e+005	(K) LADDIPSNSPMAINCEYGSFDNEHLVLPK (T)	P04806	Hexokinase-1
7	yeast-1.19622.19622.2	2	24.31	96.9	6.23e+005	(R) SINPDEAVAYGAAVQAAILTGDESSK (T)	P10591	Heat shock protein SSA1
8	yeast-1.17757.17757.2	2	24.17	91.1	1.47e+005	(K) EYFQDVYDLTGSEPIK (E)	P07149	Fatty acid synthase subunit beta
9	yeast-1.14605.14605.2	2	23.66	96.2	3.93e+006	(K) LVSWYDNEYGYSTR (V)	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2
10	yeast-1.20693.20693.2	2	23.44	98.4	2.72e+005	(K) ETNEGTDVIVSSVESVLAHL (-)	P38013	Peroxiredoxin type-2
11	yeast-1.20456.20456.2	2	23.41	95.3	9.17e+005	(K) TIAETLAEELINAAK (G)	P26783	40S ribosomal protein S5
12	yeast-1.8159.8159.2	2	23.32	97.7	7.71e+005	(K) SASEYDSALASGDK (L)	P22803	Thioredoxin-2
13	yeast-1.17415.17415.3	3	23.27	95.7	3.25e+005	(K) ILPEPTEGFAADDEFTIPAEVLGK (F)	P07149	Fatty acid synthase subunit beta
14	yeast-1.20290.20290.3	3	23.27	97.7	1.18e+005	(K) AVGENDPIDVLEIGETIAYTGQVK (Q)	P00817	Inorganic pyrophosphatase
15	yeast-1.17560.17560.3	3	23.02	97.8	1.41e+005	(K) AEQGEHDENISPAQAAELVGEDLSR (R)	P27616	Phosphoribosylaminoimidazole-succinocarboxamide synthetase
16	yeast-1.16983.17203.4	4	23.01	93.2	1.38e+007	(K) TLLEAIDAIEQPSRPTDKPLR (L)	P02994	Elongation factor 1-alpha

Figure 41 Peptide - Spectrum Match display (partial)

Peptide - Distinct

Select this mode when you want to summarize results by distinct peptides only. It provides filtering options to view peptides with variable modifications. It collapses the peptide list to the best-scoring representative peptide spectrum match per distinct peptide.

Distinct can mean any one of these user-selectable options:

- Case insensitive - Both modified and unmodified peptides are considered the same peptide.
- Case sensitive (CS) - Modified peptides are considered distinct from unmodified peptides.
- Charge file CS - This mode keeps the best-scoring peptide spectrum match for each precursor charge for each LC/MS/MS run.

Validation category	#	Filename	z	Score	SPI (%)	Spectrum Intensity	Sequence	Accession #	Protein Name
☐ V ☒ R	1	yeast-1.17537.17537.2	2	25.62	96.4	5.71e+005	(K) LLYYDYQELPAEAINR (L)	P008911	Formate dehydrogenase 1
☐ V ☒ R	2	yeast-1.21279.21279.3	3	25.49	94.5	7.50e+006	(R) YGASAGNVGDEGGVAPNIQTAEALDLIVDAIK (A)	P00924	Enolase 1
☐ V ☒ R	3	yeast-1.20324.20500.3	3	25.12	98.5	3.74e+006	(K) VINDAFGIEEGLMTTVHSLTATQR (T)	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3
☐ V ☒ R	4	yeast-1.18605.18605.3	3	24.65	96.9	2.36e+006	(K) FADGFLIHSGQFVNDFTIATR (H)	P05750	40S ribosomal protein S3
☐ V ☒ R	5	yeast-1.17684.17882.2	2	24.45	100.0	4.59e+007	(R) VPTVDVSVVDLTVK (L)	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1
☐ V ☒ R	6	yeast-1.18150.18150.4	4	24.43	93.8	1.93e+005	(K) LADDIPSNSPMAINCEYGSFDNEHLVLR (T)	P04806	Hexokinase-1
☐ V ☒ R	7	yeast-1.19622.19622.2	2	24.31	96.9	6.23e+005	(R) SINPDEAVAYGAQVQAAILTGDESSK (T)	P10591	Heat shock protein SSA1
☐ V ☒ R	8	yeast-1.17757.17757.2	2	24.17	91.1	1.47e+005	(K) EYFQDYYDLTGSEPIK (E)	P07149	Fatty acid synthase subunit beta
☐ V ☒ R	9	yeast-1.14605.14605.2	2	23.66	96.2	3.93e+006	(K) LVSWYDNEYGYSTR (V)	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2
☐ V ☒ R	10	yeast-1.20693.20693.2	2	23.44	98.4	2.72e+005	(K) ETNPGIDVTVSSVESVLAHL (-)	P38013	Peroxisomal type-2
☐ V ☒ R	11	yeast-1.20456.20456.2	2	23.41	95.3	9.17e+005	(K) TIAETLAELINAAK (G)	P26783	40S ribosomal protein S5
☐ V ☒ R	12	yeast-1.8159.8159.2	2	23.32	97.7	7.71e+005	(K) SASEYDSALASGDK (L)	P22803	Thioredoxin-2
☐ V ☒ R	13	yeast-1.17415.17415.3	3	23.27	95.7	3.25e+005	(K) ILPEPTEGFAADDEFTTFAELVGK (F)	P07149	Fatty acid synthase subunit beta
☐ V ☒ R	14	yeast-1.20290.20290.3	3	23.27	97.7	1.18e+005	(K) AVGDNDPIDVLEIGETIATGQVK (Q)	P00817	Inorganic pyrophosphatase
☐ V ☒ R	15	yeast-1.17560.17560.3	3	23.02	97.8	1.41e+005	(K) AEQGHENISPAQAAELVGEDLSR (R)	P27616	Phosphoribosylaminoimidazole-succinocarboxamide
☐ V ☒ R	16	yeast-1.16983.17203.4	4	23.01	93.2	1.38e+007	(K) TLLEAIDAIEQPSRPTDKPLR (L)	P02994	Elongation factor 1-alpha
☐ V ☒ R	17	yeast-1.18564.18564.3	3	22.70	94.4	1.74e+006	(K) NDTFALIGYGSQYQGGLMLR (D)	P06168	Ketol-acid reductoisomerase, mitochondrial
☐ V ☒ R	18	yeast-1.19110.19110.2	2	22.68	89.7	2.09e+006	(R) IEDDPFNEEDTDIFQK (D)	P04806	Hexokinase-1

Figure 42 Peptide - Distinct display

Protein Summary Details

Select this mode to see results organized by proteins, with all the supporting peptide information. This mode is useful when your sample fractions are organized in a single data directory, and you want the peptides from a single protein organized together even if they originate from different data files.

The top of the summary shows the proteins, as in [Figure 43](#), while the bottom of the summary shows the individual peptides for each protein, as in [Figure 44](#). This mode is very useful for manual review and validation of MS/MS search results. If your sample has multiple fractions, you sometimes see the same peptide sequences in multiple data files. You also see this if you chose not to merge spectra during extraction.

Group (#)	Subgroup (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Total Protein Spectral Intensity	Database Accession #	Protein Name
1	1.1	41	25	412.64	70.1	3.09e+008	P00560	Phosphoglycerate kinase
2	2.1	37	22	369.72	62.7	2.33e+008	P00925	Enolase 2
2	2.2	37	21	353.79	59.7	3.00e+008	P00924	Enolase 1
3	3.1	30	26	366.09	43.8	2.90e+007	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine
4	4.1	35	21	365.97	77.7	3.03e+008	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3
4	4.2	28	18	291.84	58.7	2.57e+008	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2
4	4.3	28	16	252.47	57.5	2.55e+008	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1
5	5.1	30	24	357.32	46.5	7.27e+007	P10591	Heat shock protein SSA1
5	5.1	7	6	89.82	10.2	2.04e+007	P22202	Heat shock protein SSA4
5	5.2	23	18	264.16	32.7	6.11e+007	P10592	Heat shock protein SSA2
5	5.3	10	8	114.22	13.2	2.34e+007	P09435	Heat shock protein SSA3
6	6.1	34	21	326.59	55.4	5.44e+007	P06169	Pyruvate decarboxylase isozyme 1
6	6.2	9	6	88.42	16.5	8.03e+006	P26263	Pyruvate decarboxylase isozyme 3
6	6.3	6	5	69.93	12.7	1.10e+007	P16467	Pyruvate decarboxylase isozyme 2
7	7.1	30	21	317.00	54.4	8.62e+007	P00549	Pyruvate kinase 1
8	8.1	26	21	308.04	55.1	3.00e+007	P46367	Potassium-activated aldehyde dehydrogenase, mito
9	9.1	25	21	302.76	40.2	4.63e+007	P11484	Heat shock protein SSB1
9	9.2	25	21	301.53	41.5	4.47e+007	P40150	Heat shock protein SSB2
10	10.1	25	20	292.89	35.3	2.30e+007	P32324	Elongation factor 2
11	11.1	22	18	274.60	51.5	2.83e+007	P04806	Hexokinase-1
12	12.1	23	19	273.03	26.9	2.18e+007	P15108	ATP-dependent molecular chaperone HSC82
12	12.2	18	16	234.32	22.8	2.05e+007	P02829	ATP-dependent molecular chaperone HSP82
13	13.1	19	17	257.97	12.8	4.90e+006	P07149	Fatty acid synthase subunit beta

Figure 43 Protein Summary Details display - top

2 Reviewing and Validating MS/MS Data

Protein/Peptide Summary Displays

Group (#)	Subgroup (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Total Protein Spectral Intensity	Database Accession #	Protein Name
1	1.1	41	25	412.64	70.1	3.09e+008	P00560	Phosphoglycerate kinase ▾
#	Filename	z	Score	SPI (%)	Spectrum Intensity	Sequence		
1	yeast-1.17794.17970.4	4	22.26	90.6	6.22e+006	(R) HELSSLADVYINDAFGTAHR (A)		
2	yeast-1.14068.14068.2	2	20.53	94.5	4.05e+006	(K) EGIPAGWQGLDNGPESR (K)		
3	yeast-1.15601.15814.2	2	20.05	97.7	2.61e+007	(K) VLENTEIGDSIFDK (A)		
4	yeast-1.17053.17266.2	2	19.59	96.4	3.47e+007	(K) ELPGVAFLEK (K)		
5	yeast-1.18984.18984.2	2	19.43	87.6	4.45e+006	(K) TIWNNGPPGVFEFEK (F)		
6	yeast-1.13768.13768.2	2	19.25	96.7	2.47e+007	(R) VDFNVFLDGK (K)		
7	yeast-1.17816.17816.3	3	19.17	89.3	4.22e+006	(R) HELSSLADVYINDAFGTAHR (A)		
8	yeast-1.17772.17948.2	2	19.15	90.8	1.77e+007	(K) ASAPGSVILENLR (Y)		
9	yeast-1.13316.13316.2	2	18.68	88.8	3.33e+006	(K) ISHVSTGGGASLELEK (E)		
10	yeast-1.17906.18102.2	2	18.62	87.7	8.78e+006	(K) DVTFLNDCVGPVEAAVK (A)		
11	yeast-1.13284.13284.3	3	18.61	85.5	1.74e+006	(K) IVTDKEGIPAGWQGLDNGPESR (K)		
12	yeast-1.14137.14362.2	2	18.50	87.9	3.89e+007	(K) LSVQDLDLK (D)		
13	yeast-1.18052.18052.4	4	18.04	93.3	1.10e+005	(K) FRHELSSLADVYINDAFGTAHR (A)		
14	yeast-1.13289.13746.3	3	17.87	91.5	2.92e+007	(K) ISHVSTGGGASLELEK (E)		
15	yeast-1.21214.21214.3	3	17.74	83.7	1.95e+005	(K) GVEVVLFPVDFIADAFSADANTK (T)		
16	yeast-1.18030.18030.5	5	16.84	89.8	2.47e+005	(K) FRHELSSLADVYINDAFGTAHR (A)		
17	yeast-1.9445.9445.4	4	16.78	79.9	4.76e+005	(R) YVVLASHLGRPNR (N)		
18	yeast-1.11678.11678.2	2	16.41	86.0	1.55e+007	(K) ELQSLLGK (D)		
19	yeast-1.17949.17949.3	3	16.26	80.4	4.64e+006	(K) DVTFLNDCVGPVEAAVK (A)		

Figure 44 Protein Summary Details display - bottom

To view sequence coverage, click a link under the % AA Coverage heading.

Protein-Protein Comparison

Select this mode to organize your results by proteins, especially if you want to compare proteins across multiple samples, as in a differential expression study.

Note that MPP export (MPP Generic format) is supported in this mode.

This mode is useful when your sample fractions are organized in a single data directory, and you want the peptides from a single protein organized together even if they originate from different data files.

If you have multiple fractions for each sample, and each sample has its own subdirectory on the Spectrum Mill server, then you can use the **Group results by** parameter to control whether the individual fractions or the overall sample results (with all fractions consolidated) are displayed.

Figure 45 and Figure 46 show examples of each type of display. Figure 45 shows the distribution of proteins across a 2D LC/MS/MS run (individual fractions), while Figure 46 combines the results of all salt fractions into a single sample.

2 Reviewing and Validating MS/MS Data Protein/Peptide Summary Displays

Agilent Spectrum Mill - Protein/Peptide Summary

Spectrum MillSummary SettingsAutovalidationEasy MS/MSMS/MS SearchSpectrum SummaryBuild TICTool BeltHelp

Results Shown Filtered by Validation Category: valid
Data Directory: msdataSMExampleDataAgilentTrap/ecoli/heatshock
hit table read - SpecFeatures read
valid hits read from tagSummary file - Files: 376 Hits: 2557
beginning to assemble proteins proteins assembled 0.380508 sec
proteins filtered by unique peptides 0.149504 sec
proteins filtered by score
calculated protein coverage maps 0.448646 sec
beginning to roll up proteins into groups ... proteins rolled up into groups 0.187058 sec
protein groups ready for display
proteinGroupingMethod: oneSharedPeptide
Mean intensity: sum of intensity for all spectra of peptides belonging to protein / # spectra

HEC_0000 # spectra mean intensity	HEC_0010 # spectra mean intensity	HEC_0020 # spectra mean intensity	HEC_0040 # spectra mean intensity	HEC_0060 # spectra mean intensity	HEC_0080 # spectra mean intensity	HEC_0100 # spectra mean intensity	HEC_0150 # spectra mean intensity	HEC_0250 # spectra mean intensity	HEC_0500 # spectra mean intensity	HEC_1000 # spectra mean intensity	Database Accession #	%AA Coverage	Distinct Peptides (#)	Distin Summ MS/M Searc Score
0 0.00e+000	2 1.61e+006	31 2.05e+007	2 1.20e+007	43 1.27e+007	8 8.86e+006	5 3.14e+006	0 0.00e+000	1 3.10e+006	3 2.33e+006	3 2.06e+006	45686198	18	7	125
0 0.00e+000	0 0.00e+000	4 3.14e+007	1 4.87e+007	4 6.67e+006	3 3.75e+006	4 6.89e+006	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	15799694	19	7	104
0 0.00e+000	2 1.33e+007	9 4.23e+007	3 2.71e+007	1 1.96e+007	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	15803853	9	4	76
0 0.00e+000	3 4.72e+006	2 1.95e+007	0 0.00e+000	0 0.00e+000	2 1.02e+007	2 7.18e+006	1 4.59e+006	1 1.94e+006	0 0.00e+000	0 0.00e+000	15800756	9	3	49
0 0.00e+000	0 0.00e+000	4 5.42e+007	2 1.53e+007	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	15803363	15	3	49
0 0.00e+000	4 6.79e+006	6 5.75e+007	1 2.44e+007	1 6.36e+006	1 2.67e+006	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	15800320	21	3	46
0 0.00e+000	0 0.00e+000	6 1.34e+007	2 3.63e+007	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	15804515	8	3	46
0 0.00e+000	0 0.00e+000	2 5.63e+007	21 2.62e+007	2 1.73e+007	0 0.00e+000	0 0.00e+000	1 2.51e+006	0 0.00e+000	0 0.00e+000	0 0.00e+000	26249817	13	2	45
0 0.00e+000	0 0.00e+000	4 1.66e+007	1 2.22e+007	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	15803459	10	2	34

Figure 45 Protein-Protein Comparison display mode, **Group results by File**

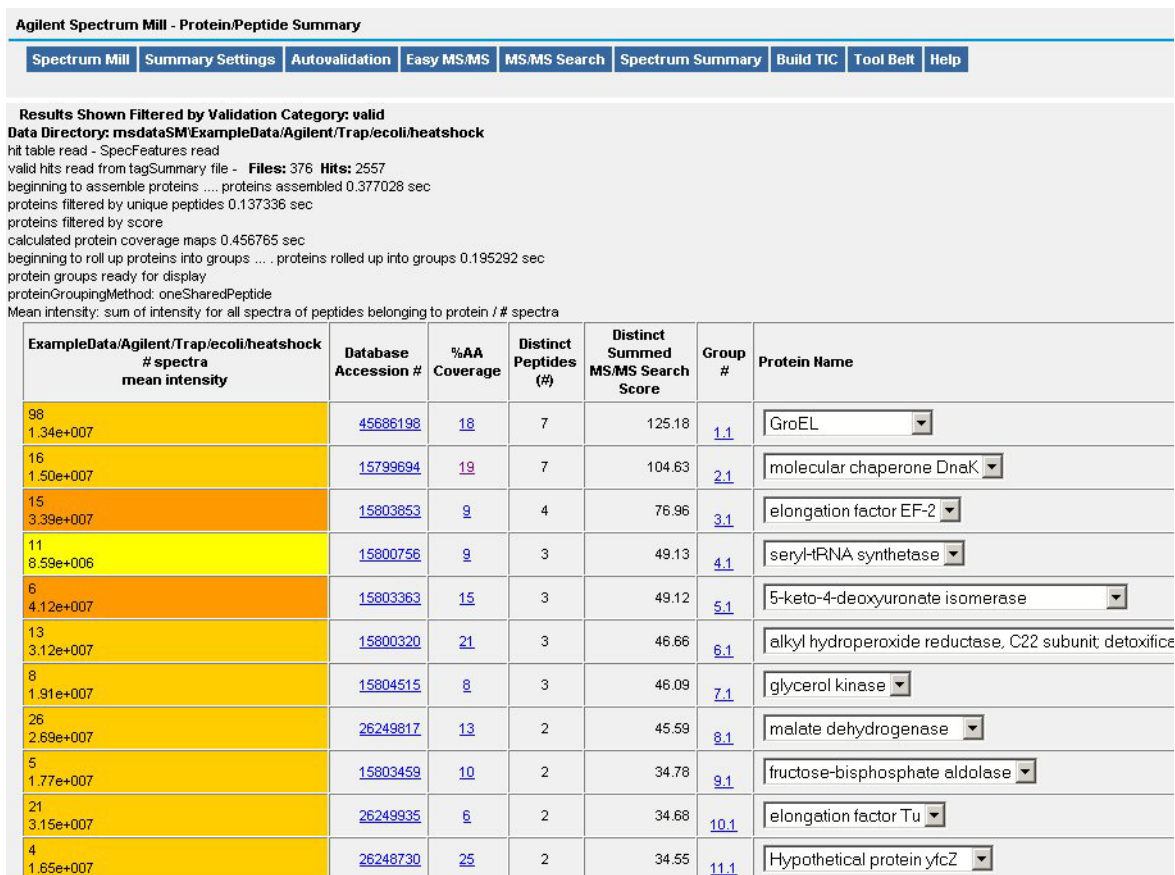


Figure 46 Protein-Protein Comparison display mode, **Group results by Directory**

If you want the software to align and compare amino acid sequences of proteins within a protein group, for **Protein grouping method**, select **1 shared, expand subgroups**.)

To view the aligned sequences, click a link for **Group #** or **Subgroup #** in the report. The software highlights the detected peptides within the sequences. For more information, see the online help about the multiple sequence alignment tool within Protein/Peptide Summary (in **Spectrum Mill Basics**).

Protein-Peptide Comparison

This display mode, shown in Figure 47, is similar to Protein-Peptide Distribution Columns, but allows filename, score, SPI, and intensity to be reported together, and enables easier Excel export. This mode ensures reporting of ambiguous locations of modifications (for example, phosphorylated serine presence not distinguishable between two or more possible serine residues in a peptide).

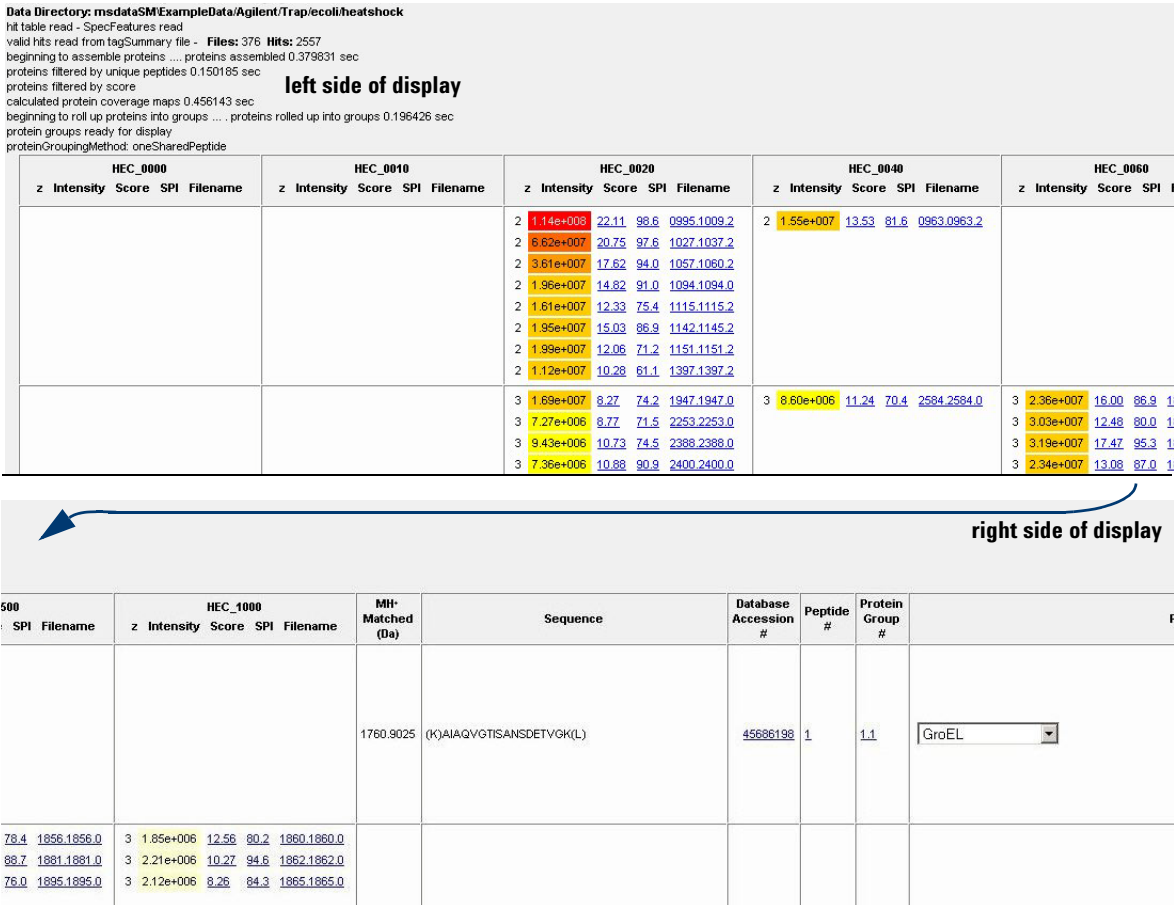


Figure 47 Protein-Peptide Comparison display mode, Group Results by File

Protein-Var Mod Site Comparison

Select this mode when you want to summarize results by variable modification sites. The rows show the variable modification sites clustered by protein, while the columns are organized by sample.

You can use this mode to compare variable modifications across samples. For example, you can generate a table of phosphorylation sites. Spectrum Mill collapses all the peptide spectrum matches that contain the same phosphorylation site into one row. It combines the quantitation for each site as the median of all the relevant peptide spectrum matches.

2 Reviewing and Validating MS/MS Data Protein/Peptide Summary Displays

yeast-1										Columns VM site Feature Observed (#)	Spectra VM site Feature Observed (#)	accessionNumber numVMsitesPresent numVMsitesLocalizedBest earliestVMsiteAA latestVMsiteAA	Best Filename
# spectra	z	Intensity	Score	SPI	Filename	accessionNumber	numVMsitesPresent	numVMsitesLocalized	Sequence				
					StartAA	EndAA							
					earliestVMsiteAA	latestVMsiteAA							
2	2	8.53e+006	15.35	84.7	4318.4318.2	P00560_9_49_56_48_48	YVLEHHPR			1	2	P00560_9_48_48	yeast-1.4318.4318.2
1	2	1.15e+007	10.84	78.0	9214.9434.2	P00560_9_75_82_74_74	YSLAPVAK			1	1	P00560_9_74_74	yeast-1.9214.9434.2
1	2	1.55e+007	16.41	86.0	11678.11678.2	P00560_9_83_90_82_82	ELQSLLGK			1	1	P00560_9_82_82	yeast-1.11678.11678.2
1	2	1.42e+007	11.86	71.9	11987.12206.2	P00560_9_183_190_182_182	AAGFLLEK			1	1	P00560_9_182_182	yeast-1.11987.12206.2
1	2	4.68e+006	10.27	78.6	9346.9346.2	P00560_9_321_328_320_320	LFAATVAK			1	1	P00560_9_320_320	yeast-1.9346.9346.2
2	2	3.95e+007	18.50	87.9	14137.14362.2	P00560_10_7_15_6_6	LSVQDLDLK			1	2	P00560_10_6_6	yeast-1.14137.14362.2
1	3	7.87e+005	11.06	80.4	13599.13599.3	P00560_10_40_48_39_39	IYAALPTIK			1	1	P00560_10_39_39	yeast-1.13599.13599.3
2	2	2.59e+007	19.25	96.7	13768.13768.2	P00560_11_23_32_22_22	VDFNVPLDGK			1	2	P00560_11_22_22	yeast-1.13768.13768.2
1	3	5.08e+006	14.19	77.0	11855.11855.3	P00560_12_23_33_22_22	VDFNVPLDGKK			1	1	P00560_12_22_22	yeast-1.11855.11855.3
1	2	3.47e+007	19.59	96.4	17053.17266.2	P00560_12_405_415_404_404	ELPGVAFLEK			1	1	P00560_12_404_404	yeast-1.17053.17266.2
3	3	9.50e+006	15.79	87.8	12382.12382.3	P00560_14_170_182_169_169	AHSSMVGFDLPQR			1	3	P00560_14_169_169	yeast-1.12382.12382.3
2	2	2.25e+007	19.15	90.8	17772.17948.2	P00560_15_109_122_108_108	ASAPGSVILENLR			1	2	P00560_15_108_108	yeast-1.17772.17948.2
1	2	2.61e+007	20.05	97.7	15601.15814.2	P00560_15_245_258_244_244	VLENTEIGDSIFDK			1	1	P00560_15_244_244	yeast-1.15601.15814.2
1	4	4.76e+005	16.78	79.9	9445.9445.4	P00560_16_57_71_56_56	YVVLASHLGRPNR			1	1	P00560_16_56_56	yeast-1.9445.9445.4
1	3	2.74e+006	10.18	79.8	13945.13945.3	P00560_16_244_258_243_243	KVLENTEIGDSIFDK			1	1	P00560_16_243_243	yeast-1.13945.13945.3

left side of display

right side of display

Best Filename	Best Score	Start AA	Sequence	Database Accession #	Peptide #	Protein Group #	Protein Score	Protein Name
yeast-1.4318.4318.2	15.35	49	(K)YVLEHHPR(Y)	P00560	1	1.1	412.64	Phosphoglycerate kinase
yeast-1.9214.9434.2	10.84	75	(K)YSLAPVAK(E)	P00560	2	1.1	412.64	Phosphoglycerate kinase
yeast-1.11678.11678.2	16.41	83	(K)ELQSLLGK(D)	P00560	3	1.1	412.64	Phosphoglycerate kinase
yeast-1.11987.12206.2	11.86	183	(R)AAGFLLEK(E)	P00560	4	1.1	412.64	Phosphoglycerate kinase
yeast-1.9346.9346.2	10.27	321	(K)LFAATVAK(A)	P00560	5	1.1	412.64	Phosphoglycerate kinase
yeast-1.14137.14362.2	18.50	7	(K)LSVQDLDLK(D)	P00560	6	1.1	412.64	Phosphoglycerate kinase
yeast-1.13599.13599.3	11.06	40	(R)IYAALPTIK(Y)	P00560	7	1.1	412.64	Phosphoglycerate kinase
yeast-1.13768.13768.2	19.25	23	(R)VDFNVPLDGK(K)	P00560	8	1.1	412.64	Phosphoglycerate kinase
yeast-1.11855.11855.3	14.19	23	(R)VDFNVPLDGKK(I)	P00560	9	1.1	412.64	Phosphoglycerate kinase
yeast-1.17053.17266.2	19.59	405	(K)ELPGVAFLEK(K)	P00560	10	1.1	412.64	Phosphoglycerate kinase
yeast-1.12382.12382.3	15.79	170	(R)AHSSMVGFDLPQR(A)	P00560	11	1.1	412.64	Phosphoglycerate kinase
yeast-1.17772.17948.2	19.15	109	(K)ASAPGSVILENLR(Y)	P00560	12	1.1	412.64	Phosphoglycerate kinase
yeast-1.15601.15814.2	20.05	245	(K)VLENTEIGDSIFDK(A)	P00560	13	1.1	412.64	Phosphoglycerate kinase
yeast-1.9445.9445.4	16.78	57	(R)YVVLASHLGRPNR(N)	P00560	14	1.1	412.64	Phosphoglycerate kinase
yeast-1.13945.13945.3	10.18	244	(K)KVLENTEIGDSIFDK(A)	P00560	15	1.1	412.64	Phosphoglycerate kinase

Figure 48 Protein-Var Mod Site Comparison display

Protein-Single Peptide ID

Select this mode to display proteins that have been identified via a single peptide. This mode enables review of the annotated MS/MS spectra that led to these identifications. At least one journal (*Molecular & Cellular Proteomics*) requires that these annotated spectra be submitted with the manuscript.

An example Protein-Single Peptide ID report is shown in Figure 49. Each row corresponds to the highest scoring instance of the single peptide used to identify the protein, with a link to the corresponding spectrum. The fourth column shows how many spectra correspond to that peptide. (The other spectra arise from additional charge states, same peptide from one of the other raw files in the directory, etc.)

Group #	Distinct Summed MS/MS Search Score	Num Unique Peps	Num Spectra	#	Filename	z	Score	SPI (%)	Spectrum Intensity	Sequence	MH ⁺ Matched (Da)	Accession #	Protein Name
1	23.44	1	1	1	yeast-1.20693.20693.2	2	23.44	98.4	1.54e+004	(K) ETNPGTDVTVSSVESVLAHL (-)	2055.024	6323138	thioredoxin peroxidase; Ahp1p
2	22.34	1	1	1	yeast-1.6725.6725.3	3	22.34	82.2	2.33e+004	(K) SSYGSNNDDSYGSNNDDSYGSSNK (K)	2533.946	6323826	DNA damage inducible; implicated in the p
3	22.03	1	1	1	yeast-1.20082.20082.2	2	22.03	96.4	3.08e+005	(K) MAIFDELLQMPETK (L)	1665.823	6322565	adenosine kinase; Ado1p
4	20.52	1	1	1	yeast-1.14343.14406.2	2	20.52	90.5	2.28e+006	(R) LTADYDALDIANR (I)	1450.717	320176	ribosomal protein L23a.e
5	20.38	1	1	1	yeast-1.18109.18109.2	2	20.38	93.4	5.04e+005	(K) EFDEAYDLFEVQR (V)	1660.749	6321842	subunit VI of cytochrome c oxidase; Cox6p
6	20.21	1	1	1	yeast-1.20704.20704.3	3	20.21	90.0	6.63e+004	(K) NAPAIIFIDEIDSIAFK (R)	1826.990	6320077	Microsomal protein of CDC48/PAS1/SEC1
7	20.19	1	1	1	yeast-1.17335.17335.2	2	20.19	85.9	1.70e+006	(R) SGVVADESITAFNDLK (L)	1736.870	6322978	yeast cortical cytoskeleton component; m
8	20.13	1	1	1	yeast-1.20174.20174.2	2	20.13	91.4	3.82e+005	(K) IEAALSDAALQIEDPSADELR (K)	2411.230	6319972	Stress-Seventy Subfamily B; involved in tr
9	20.02	1	1	1	yeast-1.19422.19515.2	2	20.02	97.8	2.30e+006	(R) SSVLADALNAINNAEK (T)	1629.844	6322271	Homology to rat S15a; Rps22ap
10	19.60	1	1	1	yeast-1.9833.9833.3	3	19.60	91.3	5.57e+005	(K) HGENSELLSGTISQDSLRL (E)	1787.852	6321679	phosphofructokinase alpha subunit Pfk1p

Figure 49 Protein-Single Peptide ID display

Protein-Peptide Distribution Columns

Select this mode to troubleshoot your fractionation method. This display mode, shown in [Figure 50](#), allows you to look at individual analyses that comprise your sample. The individual analyses could be 2D LC fractions, gel slices, or other types of fractions. This display gives insight into whether you need to adjust analysis parameters for better separation.

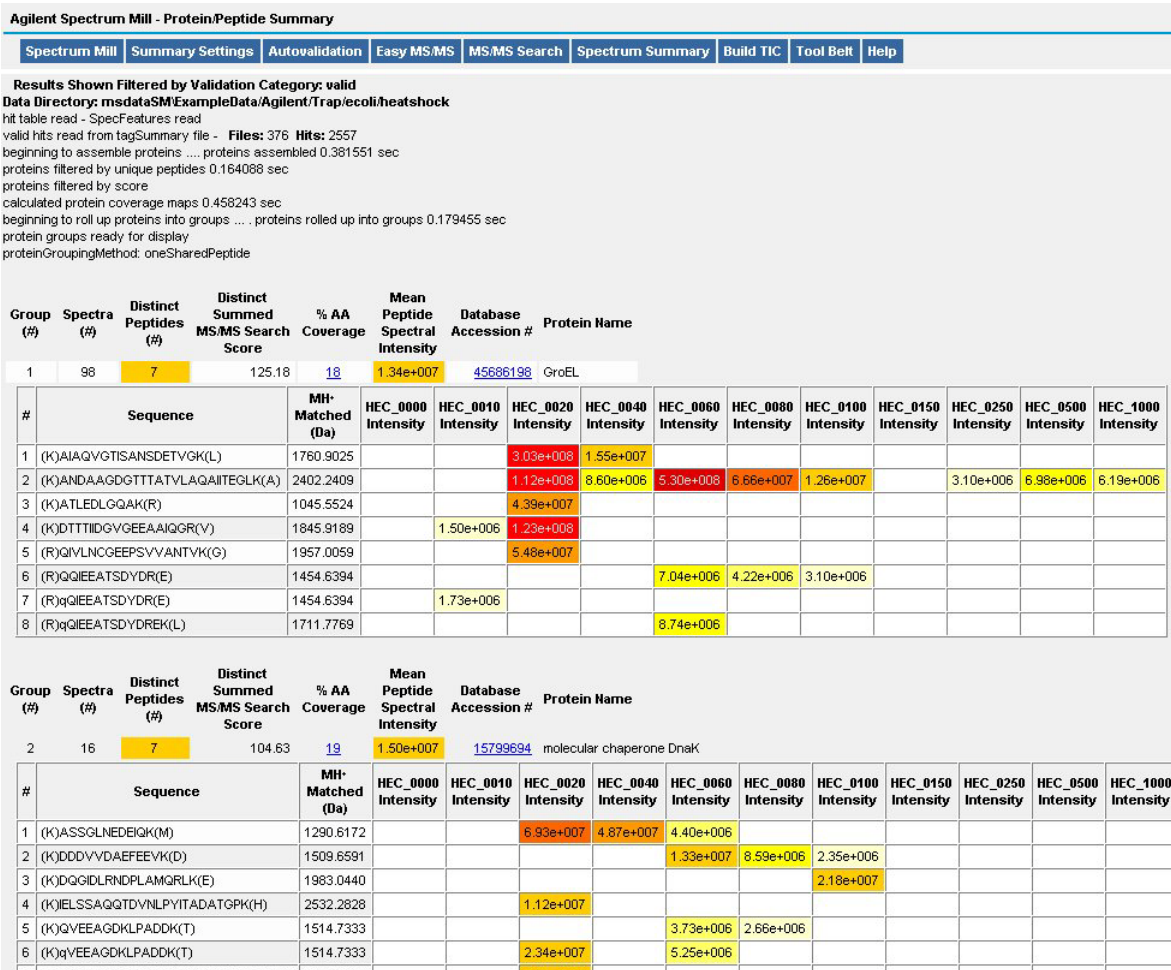


Figure 50 Protein-Peptide Distribution Columns display mode, Group results by File

Identification Statistics and False Discovery Rate

To create a summary table of MS/MS identification statistics

The Spectrum Mill workbench compiles identification statistics, such as the percentage of MS/MS spectra that have been interpreted for a set of data files. This is the same set of statistics and report that you accessed previously using the View MS/MS search statistics option but now produce using the Quality Metrics and FDR option.

To create the summary table for basic search statistics:

- 1 From the Spectrum Mill home page, click **Quality Metrics & FDR**.

Agilent Spectrum Mill - Quality Metrics and False Discovery Rates (FDR) - (none)

Spectrum Mill | Extractor | MS/MS Search | Protein/Peptide Summary | Spectrum Summary | Autovalidation | Workflows | Tool Belt | Help

Report ☐ Queue request ☐ Excel Export Save As... Load...

FDR Metrics (spectra, peptide, protein)
☐ FDR at the peptide & spectra level (from valid hits)
☐ FDR at the protein level
☐ Group proteins across all directories
 Grouping method: 1 shared, expand subgroups

MS/MS Interpretation Metrics
☐ Identification Scores
☐ Fragmentation Mode
☐ Variable modification site localization: slightly

Peptide Separation Metrics
☐ Chromatography metrics for each run
 Middle RT portion of validated spectra in LC run: 90 (%)
☐ Peptide pI median for each run

Precursor Ion Metrics
☐ Precursor mass error mean (ppm)
☐ Precursor charge count (from valid spectra)
☐ Precursor Isolation Purity & Average Chi²
☐ Precursor Acquisition Uncertainty: m/z and z
☐ Precursor Ion Fragmentation Table

MS/MS Spectral Identifiability Metrics (define thresholds below)
☐ Sequence tag length: > 3
☐ Precursor isotope quality XIC's (Chi² vs. Average): > 0.7
☐ Precursor Isolation Purity: > 70 %

Sample Handling Metrics
☐ Isobaric label incorporation for each run: ITRAQ 4 Control: 115 116 116
☐ Digestion completeness
☐ Observed modifications by: Distinct Peptides

Peptide Fraction Overlap
☐ Distinct Peptide Fraction Overlap Table

Data Directories
 Select ...
☒ ExampleData/Agilent/Q-TOF/Yeast/Y1
☒ ExampleData/Agilent/Q-TOF/Yeast/Y2

Distinct peptide comparison method: Case Insensitive(CI)
 For fraction overlap, sample handling, or pI choose CS or CI.
 FDR calculations always use CI.
 Filtering to distinct peptides retains each highest scoring representative after CS or CI string comparison of sequences. Variable mods are lowercase.

Figure 51 Report Quality Metrics and FDR - with no check boxes marked

- 2 Do NOT mark any of the check boxes. Clear any check marks that you see. You see MS/MS identification statistics by just clicking the Report button without marking any other check boxes. See [“To report FDR and search statistics”](#) on page 106 to learn how to report FDR calculations and other statistics in addition to the basic set described here.
- 3 Select the **Data Directories** for which you want to summarize statistics.
- 4 Click **Report**.

2 Reviewing and Validating MS/MS Data

Identification Statistics and False Discovery Rate

5 Check that you see a report like that shown in Figure 52.

Data Directory: ExampleData/Agilent/Q-TOF/Yeast/Y1
retrieved list of 4451 validated spectra.
Calculated ID-independent metrics.
Calculated ID-dependent metrics.

Data Directory: ExampleData/Agilent/Q-TOF/Yeast/Y2
retrieved list of 4912 validated spectra.
Calculated ID-independent metrics.
Calculated ID-dependent metrics.

File	Raw Files	MS/MS spectra collected C	MS/MS spectra merged	MS/MS spectra filtered F	MS/MS spectra valid V	Collection Yield V/C (%)	Validation Yield V/F (%)
yeast-1	1	19276	19044	17360	4451	23.1	25.6
yeast-2	1	22102	21663	19678	4912	22.2	25.0

Directory	Raw Files	MS/MS spectra collected C	MS/MS spectra merged	MS/MS spectra filtered F	MS/MS spectra valid V	Collection Yield V/C (%)	Validation Yield V/F (%)
ExampleData/Agilent/Q-TOF/Yeast/Y1	1	19276	19044	17360	4451	23.1	25.6
ExampleData/Agilent/Q-TOF/Yeast/Y2	1	22102	21663	19678	4912	22.2	25.0
all	2	41378	40707	37038	9363	22.6	25.3

Figure 52 View search statistics report

Note that under File/Directory the stats are given for the file first, then the sum of the files under the directory (in this case, there is only one file under each of two directories) and that under Directory appear the stats for each directory and then the sum of the directories.

See below for explanations of some of the headers.

MS/MS spectra collected	Number of MS/MS spectra in the raw data file
MS/MS spectra merged	Number of MS/MS spectra that result from merging by the Data Extractor
MS/MS spectra filtered	Number of MS/MS spectra exported by the Data Extractor program after filtering by spectral quality
Collection Yield (%)	Number of MS/MS spectra interpreted and validated divided by number of MS/MS spectra collected, expressed as percentage
Interpretation Yield (%)	Number of MS/MS spectra interpreted and validated divided by number of MS/MS spectra filtered, expressed as percentage

It is typical that not all spectra are interpreted. The obvious reason is that some spectra could not be matched to database entries, but there are other reasons as well. For example, to avoid missing any good spectra, you typically set Data Extractor to remove most, but not all, of the noisy spectra. The remaining noisy spectra are not peptides and are not interpreted.

If your **Collection Yield** seems particularly low, there may have been an unusually high number of noisy spectra in your analysis. Perhaps you used a low threshold for data acquisition, or maybe there was a high instrument background. In these cases, the relative number of spectra that are picked by the Data Extractor is low.

Both the **Collection Yield** and the **Interpretation Yield** reflect to some degree how much time you spent processing the data, via variable modifications searches, broader databases, etc. In general, processing is complete when sufficient information has been extracted from the data to meet the experimental goals.

To report FDR and search statistics

You can now report useful metrics in addition to the basic search statistics already calculated by the program. (For the basic search statistics, see “To create a summary table of MS/MS identification statistics” on page 103.)

NOTE

FDR results are not valid after manual validation. For this reason Agilent recommends you complete your final autovalidation and review, and check the FDR statistics before you do a manual validation.

- 1 From the Spectrum Mill home page, click **Quality Metrics & FDR**.
- 2 Mark each check box whose calculation or statistic you want the program to report.

Click the blue bars to see explanations for each option and why you would want to see that calculation or statistic.

Agilent Spectrum Mill - Quality Metrics and False Discovery Rates (FDR) - (none)

Spectrum Mill | Extractor | MS/MS Search | Protein/Peptide Summary | Spectrum Summary | Autovalidation | Workflows | Tool Belt | Help

Report ☐ Queue request ☐ Excel Export Save As... Load...

FDR Metrics (spectra, peptide, protein)

☒ FDR at the peptide & spectra level (from valid hits)

☐ FDR at the protein level

☐ Group proteins across all directories

Grouping method: 1 shared, expand subgroups

Precursor Ion Metrics

☒ Precursor mass error mean (ppm)

☒ Precursor charge count (from valid spectra)

☐ Precursor Isolation Purity & Average Chi²

☐ Precursor Acquisition Uncertainty: m/z and z

☐ Precursor Ion Fragmentation Table

MS/MS Interpretation Metrics

☐ Identification Scores

☐ Fragmentation Mode

☐ Variable modification site localization: silly

MS/MS Spectral Identifiability Metrics (define thresholds below)

☐ Sequence tag length: > 3

☐ Precursor isotope quality XIC's (Chi² vs. Average): > 0.7

☐ Precursor Isolation Purity: > 70 %

Peptide Separation Metrics

☒ Chromatography metrics for each run

Middle RT portion of validated spectra in LC run: 90 (%)

☐ Peptide pI median for each run

Sample Handling Metrics

☐ Isobaric label incorporation for each run: iTRAQ 4 Control: 115 116

☐ Digestion completeness

☐ Observed modifications by: Distinct Peptides

Distinct peptide comparison method: Case Insensitive(CI)

For fraction overlap, sample handling, or pI choose C S or CI. FDR calculations always use CI. Filtering to distinct peptides retains each highest scoring representative after C S or CI string comparison of sequences. Variable mods are lowercase.

Data Directories

Select ... ☒ ExampleData/Agilent/Q-TOF/Yeast/Y2

Figure 53 Report Quality Metrics and FDR - default check boxes marked

- 3 Select the **Data Directories** for which you want to report the FDR or statistics.

The page should now look like [Figure 53](#).

- 4 Click **Report**.

The first part of the report you see will be like that shown in [Figure 52](#). Added to that report are columns containing the statistics you chose to have the program report. Refer to [Table 6](#) to view the column headers that appear for each option.

Table 6 Results After Reporting FDR and Search Statistics

Report this statistic:	To view these report column headers:	Because you want to know:
FDR at the peptide & spectra level (from valid hits)	FDR Spectra (%) # Spectra False FDR Distinct Peptide (%) # Distinct Peptides # Distinct Peptides False	The FDR at both the spectra and peptide levels. The distinct peptide FDR can only be reported with the Tool Belt. The spectral level FDR may be seen in other places depending on the autovalidation strategy.
FDR at the protein level	All of the above, plus: FDR Distinct Protein (%) # Protein Groups # Distinct Proteins # Distinct Proteins False	The FDR at the spectra, peptide and proteins levels. This can only be reported with the Tool Belt.
Precursor mass error mean (ppm)	Mean Precursor Mass Error (ppm) Std Dev Precursor Mass Error (ppm) Mean is the average of all the values. Standard deviation measures the spread in the values away from the mean.	The ppm difference between the experimentally measured mass and the theoretical mass calculated from the amino acid sequence. A ppm error larger than expected can help identify instrumentation or data quality issues.
Precursor charge count (from valid spectra)	# spectra z2-z7 z2 (%) - z7 (%)	The number of spectra for each charge state from +2 to +7 and the percentage of spectra in each charge state from +2 to +7

2 Reviewing and Validating MS/MS Data

Identification Statistics and False Discovery Rate

Table 6 Results After Reporting FDR and Search Statistics (continued)

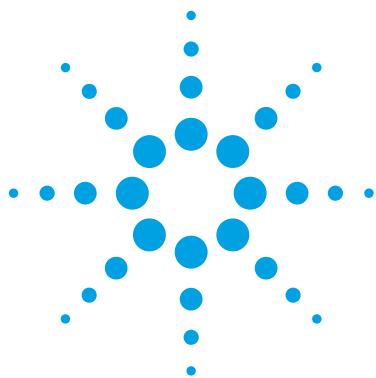
Report this statistic:	To view these report column headers:	Because you want to know:
Peptide pl median for each run	# Distinct Peptides pl Median Peptide pl Std Dev Peptide pl	Distribution of peptide pl in each run; typically used when pl-based fractionation has been done.
Chromatography stats for each run	Start time mid xx% matched spectra in run (min) End time mid xx% matched spectra in run (min) Time span mid xx% matched spectra in run (min) Median MS1 peak width mid xx% matched spectra (sec) Total precursor XIC mid xx% matched spectra in run Median MS1 intensity Trigger Apex mid 50% matched spectra (%)	How good the chromatography was For only the matched spectra between +/- one-half xx%, the program reports each of these statistics. For example, if mid xx% is specified as mid 60% and 10,000 spectra gave IDs in the run, then the statistics specified at mid 60% are reported for spectra #2000 to #8000 (+/- 30%). See the Online Help for descriptions of each of these statistics. To learn more about these metrics, refer to Rudnick PA, Clauser KR, Kilpatrick LE, et. al., "Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses", Mol Cell Proteomics. 2010 Feb;9(2):225-41 (http://www.ncbi.nlm.nih.gov/pubmed/19837981).
Distinct Peptide Fraction Overlap Table - Distinct	Distinct Peptide Fraction Overlap Table	The generated table shows for each peptide in each data file, the number of data files that peptide is found in. This is useful for seeing how good your pre-fractionation was. If everything is found in only one fraction and is also evenly distributed across the data files, your pre-fractionation worked perfectly. This is not likely, of course.
Distinct Peptide Fraction Overlap Table - Not Distinct	Distinct Peptide Fraction Overlap Table	The "best" peptide for each primary sequence, no matter the modification.

Table 6 Results After Reporting FDR and Search Statistics (continued)

Report this statistic:	To view these report column headers:	Because you want to know:
Precursor Ion Fragmentation Table	Precursor Ion Fragmentation Table	Numerous useful metrics about the spectra. You may need to scroll to see them all.
Precursor Acquisition Uncertainty: m/z and z	MS/MS spectra collected acquisition MI m/z = 0.0 MS/MS spectra collected acquisition z = 0 MS/MS spectra filtered acquisition MI m/z = 0.0 MS/MS spectra filtered acquisition z = 0 MS/MS spectra valid precursors adjusted	Spectral quality through these indicators; abnormally higher values for samples of the same complexity indicate poorer quality mass spectral data.
MS1 intensity stats for each run	# MS Scans Total TIC Total BPI Mean TIC Mean BPI	If something might be wrong, e.g., if the injection of the standard material shows an abnormally low value for these

2 Reviewing and Validating MS/MS Data

Identification Statistics and False Discovery Rate



3 Automating Workflows

Setting Up Automated Workflows 112

Executing and Monitoring Workflows 116

Workflow automation lets you place a series of tasks in either a normal workflow or iterative workflow into a queue; then Spectrum Mill workbench executes them in the proper order. For example, you can automate a typical normal workflow for MS/MS data files from protein digests:

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



Setting Up Automated Workflows

The first step to set up an automated workflow is to save the settings you intend to use in parameter files. The names of these parameter files become the “tasks” that make up the “workflow”.

You then place the workflow tasks in the order you intend. If you need to view a task before placing it in the sequence of the workflow, you can.

After creating the workflow you then execute it, monitoring the results to see the outcome of a task execution, viewing the queue to know which task is being executed at any time, and checking the completion log to view those tasks and workflows that have completed.

NOTE

For help on any section of a Workflows page, click the blue bar for that section.

To create parameter files or workflow tasks

Parameter files and workflow tasks are the same. You create parameter files, or edit them, in each individual Spectrum Mill page.

See [Chapter 1](#), “Processing MS/MS Data Interactively” and [Chapter 6](#), “Processing MS-Only Data Interactively” for details on how to create parameter files.

To view already existing workflows and their tasks

- 1 On the Spectrum Mill home page, click **Workflows**.
- 2 From Data Directories, click **Select**, then select a file of files and click **Open**.
- 3 Select a workflow in the list on the left to see a list of workflow tasks on the right that comprise the workflow ([Figure 54](#)).

Figure 54 Workflow page with selected workflow

- To make sure the parameters you want are in a task, click the task of interest to view a read-only version of the page (Figure 55).

Figure 55 View of parameter values for selected workflow task

To create/edit an automated workflow

- 1 From the Workflows page, select a workflow, and click **Edit Workflow**.
Note that the name of the selected workflow to edit is in the title bar in red.
(Figure 56)
If no workflows appear in the list, click **Edit Workflow** and create a new one with steps 2-6.
- 2 To filter the list of available tasks to show only a portion of them, select a category from the **Available tasks** list.
Click **Refresh** to make sure all the available tasks are in the list.
- 3 To make sure the parameter values for the selected task are the ones you want, select the task on the left and click **View** (Figure 55).
- 4 To add a task to a workflow, select the task in the list on the left, and click **Add**.
If the added task is in the wrong place in the order, select the task, then click **Up** or **Down** to move its order in the sequence.
To remove a task from the Workflow tasks list, click **Remove**.
To remove all the tasks, click **Clear All**.

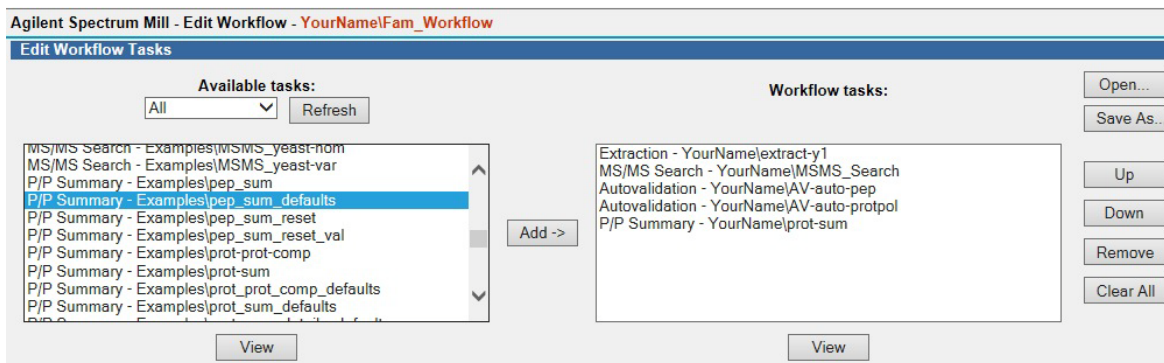


Figure 56 Edit Workflow page

- 5 To save the edited workflow to the same name or a different one, click **Save As**, enter the name of the workflow and click **Save** (Figure 57).

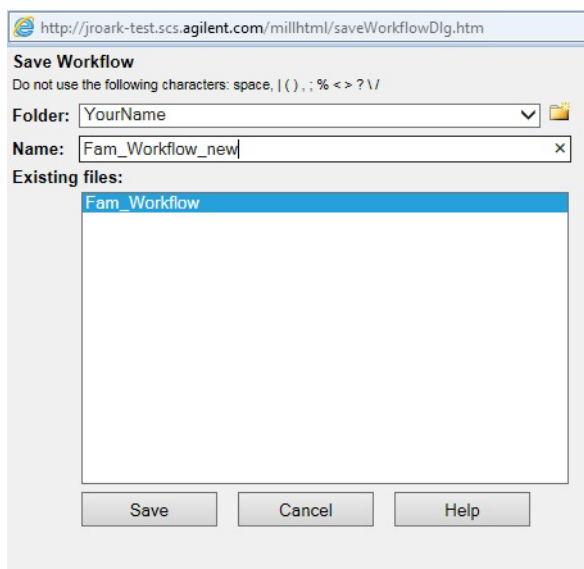


Figure 57 Save Workflow dialog box

6 Close the Edit Workflow window, and click **Refresh**.

The new workflow appears in the list on the left of the Workflows page.

To create/edit another workflow from within the Edit Workflow window

1 In the Edit Workflow window, click **Open**.

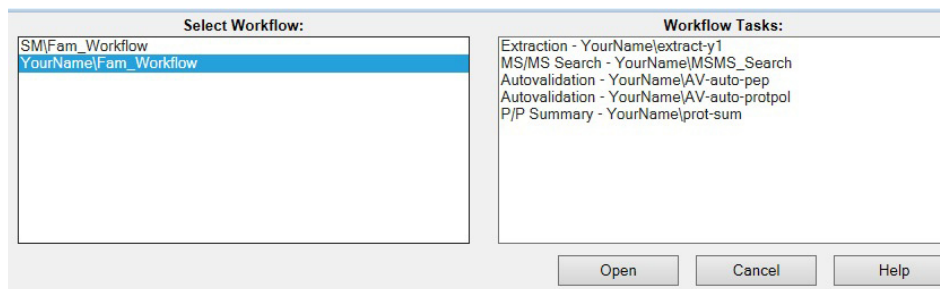


Figure 58 Open window accessed from Edit Workflow window

2 Select a workflow, then click **Open** and follow steps 4-6 above.

Executing and Monitoring Workflows

To execute an automated workflow

- 1 If you are not on the Workflows page, from the Spectrum Mill home page click **Workflows**.
- 2 Click **Select** to select one or more directories containing the data for executing the workflow.
- 3 For **Maximize CPUs**, mark the necessary check box(es) for **Extraction** and/or **Search**.

When you mark these check boxes for a workflow, the request queue shows two requests – the initial one to create the batch of files and the second one to show the progress and results of Extraction and/or Search.

- 4 Select a workflow from the list on the left, and click **Execute**.

A request submissions section for each task appears below the workflows list.

Workflow: millauto\YourName\workflow.Fam_Workflow.tsv	
requestScript	paramsFile
runXtractor.pl	YourName\extractor.extract-y1.params
batchTagPara.pl	YourName\mstag.MSMS_Search.params
validateTable.pl	YourName\autovalidation.AV-auto-pep.params
validateTable.pl	YourName\autovalidation.AV-auto-protpol.params
summaryPP.pl	YourName\ppsummary.prot-sum.params

Running runXtractor.pl

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

Running batchTagPara.pl

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

Figure 59 Request submissions list

5 Monitor results and view workflow progress. See below.

To monitor results

To monitor the results for each task, click one of three links from the request submissions list:

Link to Results

Click this link to see the results for the task in a separate window. The results in this window will not update until you close the window and click the link once more.

After you return to the request submissions list for all tasks, you can then link to the results of other tasks.

3 Automating Workflows

Executing and Monitoring Workflows

- Monitor Results** Click this link to see the results for the task on the Workflows page. With this choice you are able to view the updated execution messages for that task in real time but are unable to return to the request submissions list to see the links to other tasks' results.
- View Request Queue** Click this link to open the Request Queue viewer and select to monitor the results from that viewer. See [“To view workflow progress”](#) on page 119. With this choice you can monitor each task in real time and choose another task to monitor after closing the window.

```
Monitor output file: msdataSM\ExampleData\Agilent\Q-TOF\OGE\extractorAgilent.111026131207.50.htm
Stop Extraction PID: 3070

Processing Directory ExampleData\Agilent\Q-TOF\OGE
All prior extraction and search/summary files have been removed!

Processing File: ec-oge-2.d using millbin\extractorAgilent.cgi ...
Initialize extractor complete! D:\SpectrumMill\msdataSM\ExampleData\Agilent\Q-TOF\OGE\ec-oge-2.d\AcqData\oge-45.m method file is not found.
Ion Source Type: ESI D:\SpectrumMill\msdataSM\ExampleData\Agilent\Q-TOF\OGE\ec-oge-2.d\AcqData\oge-45.m method file is not found.

-----
Number of spectra: 5210
1000 scans read.
2000 scans read.
3000 scans read.
4000 scans read.
5000 scans read.
Read Scans: InitTime=0.31 secs : ReadScans=11.61 secs

-----
Data read into memory.
```

Figure 60 Results for the Extraction task (click **Monitor** in Request Queue)

To view workflow progress

- From the menu bar, click **Request Queue**,
or
From the request submissions list, click **View Request Queue**.

Request Queue Completion Log Help						
There are 5 requests in the queue. Available memory: 6.7 Gb of 12.9 Gb						
Remove						
X #	Task Id	Status	Task Type	Data Directory	Dependencies	
1	150821093216.141 Monitor	Running	Extraction	ExampleData/Agilent/Q-TOF/Yeast/Y1	none	
☐	2 150821093218.142P	Queued	MS/MS Search	ExampleData/Agilent/Q-TOF/Yeast/Y1	150606093439.470	
☐	3 150821093219.144	Queued	Autovalidation	ExampleData/Agilent/Q-TOF/Yeast/Y1	150606093439.470 ▼	
☐	4 150821093220.145	Queued	Autovalidation	ExampleData/Agilent/Q-TOF/Yeast/Y1	150606093439.470 ▼	

Figure 61 Request Queue while executing a workflow

- To monitor the individual task results from the Request Queue, click **Monitor** under the Task Id (Figure 60).
- To delete a task after the workflow starts running, mark the check box next to the task and click **Remove**.

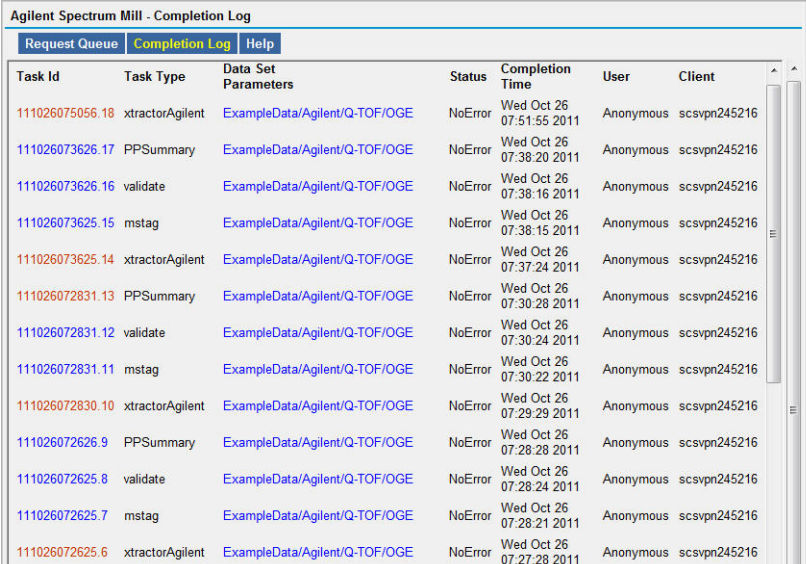
When a task is complete, it is automatically removed from the Request Queue list and appears in the Completion Log.

NOTE

You do not see any changes to the Request Queue Viewer unless you click Request Queue again to refresh the screen.

To check the Completion Log

1 In the top menu bar, click **Completion Log**.



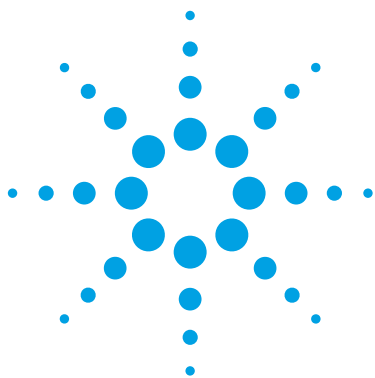
Task Id	Task Type	Data Set Parameters	Status	Completion Time	User	Client
111026075056.18	xtractorAgilent	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:51:55 2011	Anonymous	scsvpn245216
111026073626.17	PPSummary	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:38:20 2011	Anonymous	scsvpn245216
111026073626.16	validate	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:38:16 2011	Anonymous	scsvpn245216
111026073625.15	mstag	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:38:15 2011	Anonymous	scsvpn245216
111026073625.14	xtractorAgilent	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:37:24 2011	Anonymous	scsvpn245216
111026072831.13	PPSummary	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:30:28 2011	Anonymous	scsvpn245216
111026072831.12	validate	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:30:24 2011	Anonymous	scsvpn245216
111026072831.11	mstag	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:30:22 2011	Anonymous	scsvpn245216
111026072830.10	xtractorAgilent	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:29:29 2011	Anonymous	scsvpn245216
111026072626.9	PPSummary	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:28:28 2011	Anonymous	scsvpn245216
111026072625.8	validate	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:28:24 2011	Anonymous	scsvpn245216
111026072625.7	mstag	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:28:21 2011	Anonymous	scsvpn245216
111026072625.6	xtractorAgilent	ExampleData/Agilent/Q-TOF/OGE	NoError	Wed Oct 26 07:27:28 2011	Anonymous	scsvpn245216

Figure 62 Completion Log

2 To see the final result html file, click **Task Id**.

The Completion Log lets you view this information:

- Task Id - a number that identifies the workflow and the task within it.
- Task Type - extraction, search, autovalidation, summary, etc.
- Data Set Parameters - folder whose selected data files are submitted for workflow execution. Click to see a parameters table for the task.
- Status - no error, error and abort messages
- Completion Time - Date and time task completed
- User - name of user logged onto server
- Client - name of PC client being used



4 Sherenga *de novo* Sequencing

- To set parameters and run the *de novo* sequencing algorithm 122
- To generate a Sherenga report 125
- To view detailed Sherenga results 127
- To compare Sherenga results with MS/MS Search results 129

The Spectrum Mill workbench includes an advanced algorithm, called Sherenga, for *de novo* sequencing. This algorithm helps to sequence new peptides and to confirm results of database searches.

Historically, *de novo* sequencing algorithms have been difficult to develop. The major challenge in sequencing peptides from MS/MS spectra is that peptides do not usually fragment between each amino acid. While incomplete fragmentation information complicates any *de novo* program, Sherenga makes reasonable attempts to fill in the missing information.

The Sherenga algorithm uses graph theory for *de novo* peptide sequencing and sequence scoring.* Graph theory is a branch of mathematics which describes how networks (such as transportation networks) can be represented and their properties measured. A graph is an abstract symbolic representation of a network that consists of a set of vertices connected by edges. To simplify, one may use the analogy that the edges are like streets and the vertices are like intersections.

When graph theory is applied to spectra, the mass peaks are considered vertices and the differences in mass peaks are considered edges. These edges have lengths equal to the masses of the amino acids (or sometimes di- or tri-peptides). Each peak in a spectrum is assigned several possible fragment ion interpretations (vertices). Although peaks can arise from fragmentation at

* Dancik, V.; Clauser, K. R.; Addona, T.A.; Vath, J. E.; Pevzner, P.A. "De novo Peptide Sequencing via Tandem Mass Spectrometry", J. Comp. Biol. 1999, 6, 327-342.



either the N-terminus or the C-terminus, all peaks (vertices) are converted to N-terminal equivalents. Then the Sherenga algorithm attempts to find the longest possible path through as many interpretations (vertices) as possible.

To set parameters and run the *de novo* sequencing algorithm

- 1 Navigate from the Spectrum Mill home page to the Sherenga *de novo* Sequencing page, as shown in Figure 63.

Figure 63 Sherenga *de novo* Sequencing page

- 2 If you need to change the Data Directory, click **Select**, choose one, and click **OK**.
- 3 To use an existing parameter file to set parameter values, click **Load**.
- 4 Change parameters by using the instructions below and the Online Help. (Click the blue bar for each section to find explanations.)

In general, you should keep the default values, except for the data directory.

- Set the **Maximum reported hits** to the maximum number of Sherenga interpretations you would like for each spectrum.
 - Set the **Validation filter** in one of the following ways:
 If you took the time to designate your spectra as “good” or “bad,” as described in “[Step 2. Check for remaining high-quality spectra](#)” on page 37, set the **Validation filter** to **good-spectrum-sequence-not-validated**. This ensures that the software processes only the uninterpreted spectra you designated as “good”.
 Otherwise, set the **Validation filter** to **spectrum-not-marked-sequence-not-validated**.
 - For amino acids that have the same nominal mass, select how you would like to display the results under **Show Equivalent Masses**.
 For **K/Q**, choose **Both** if your instrument has sufficient mass accuracy to distinguish between these two.
 - Click the **Choose...** button to select the modifications that match the chemistry for your samples.
 For details, click the blue bar labeled **Modifications** to access the online help.
 - Select the instrument on which to base the **Scoring**.
 The scoring is based on learned parameters for each type of tandem mass spectrometer. The algorithm has “learned” rules regarding ion types and intensity thresholds from sample spectra of known amino acid sequence.
 - Mark **Sequence tag length** to prevent *de novo* sequencing of spectra that contain little sequence information. Do not set the tag length to less than 4.
 The sequence tag length represents the longest sequence of amino acids that can be located in the spectrum. Larger numbers filter out more spectra, keeping those of the best quality.
 - Click **Advanced Parameters** to display them.
 These are described in the online help.
- 5** To save any changed parameter values, click **Save As**.
 Once you save parameter files for this page, you can use them for workflow automation.
- 6** Click **Sequence**.

- 7 While Sherenga runs, check that you see a display like that shown in Figure 64.

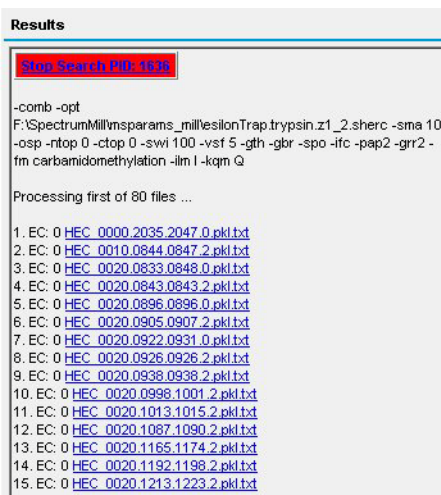


Figure 64 Display while Sherenga *de novo* Sequencing runs

- 8 If you want to view results while Sherenga runs, click a file name link. After you see the Sherenga report at the bottom of your screen, click an amino acid sequence link to display the Spectrum Viewer.

To generate a Sherenga report

- 1 When the Sherenga algorithm is complete, navigate from the Sherenga *de novo* Sequencing page to the Sherenga *de novo* Summary page.
- 2 Check that you see the page shown in [Figure 65](#).

Figure 65 Sherenga *de novo* Summary page

- 3 If it is not already selected, select the directory where your data resides.
- 4 If you want to use an already existing parameter file, click **Load**, select the file, and click **Load**.
- 5 Set the **Min. vertex score**.
If you do not want to include interpretations made on smaller peaks, raise this threshold. Amino acids interpreted based on smaller peaks are then replaced by their mass gaps.
In the detailed results that follow (see bottom blue highlight in [Figure 67](#) on page 128), **Orig. Score** represents the Sherenga score calculated with a minimum vertex score of 5, while **Score** is the Sherenga score calculated with the value you specify here.
- 6 Set other parameters. (The defaults are good choices.)
- 7 If you wish to consider only tryptic peptides in the results, mark the check box to **Force tryptic digest**.
- 8 To save the newly changed parameters, click **Save As**, name the parameter file and click **Save**.

Although parameter files can be saved, you cannot use them with workflow automation.

9 Click **Summarize**.

10 Check that you see the Sherenga top-level report, as shown in Figure 66. The components of the report are described below.

Results						
SS	TL	TS	MS/MS Search Delta MH ⁺	Filename	Sherenga Sequence Tag	MS/MS Search Result
307	9	5.0	0.7	HEC_0020.1556.1564.2.pkl.txt	[414.52]NINVVVDAGGAGQFISA[215.27]	MGLVIKAALGALVLLIGVLAK
292	10			HEC_0020.1766.1780.2.pkl.txt	[251.35]YHINVVVDAGGVASFEG[372.49]	
285	9			HEC_0020.2376.2385.2.pkl.txt	[414.41]NINVVSGGVAGQFI[373.45]	
283	9			HEC_0020.2242.2256.2.pkl.txt	V[292.61]IFSVAAGTAVGVGIN[396.32]	
280	10	5.1	0.6	HEC_0020.1537.1549.2.pkl.txt	[414.38]HGGHVVVDAGGEQFI[157.54][215.86]	MGLVIKAALGALVLLIGVLAK
278	10	6.5	0.7	HEC_0020.1387.1391.2.pkl.txt	[327.27]SGGIGGVVPGSMGQAATS[303.99]	
275	10			HEC_0020.1294.1298.2.pkl.txt	[414.45]HGGHVVVEGGGAGQFI[373.48]	MGLVIKAALGALVLLIGVLAK
274	9			HEC_0020.1510.1510.2.pkl.txt	[324.48]TIATGGTIAGGDI[158.35][229.18]	
269	10			HEC_0020.1595.1609.2.pkl.txt	[414.34]NIPASGVSGGAGEFI[157.93][215.57]	
264	9			HEC_0020.1429.1444.2.pkl.txt	[414.49]NINVVADGGEGAFI[373.31]	
264	8			HEC_0020.1962.1972.0.pkl.txt	[414.59]PGFGGAAPAGGAVPHN[373.40]	MGLVIKAALGALVLLIGVLAK
263	10			HEC_0040.1298.1311.2.pkl.txt	[212.87]IA[142.23]IQGDVAGEVGGHN[371.94]	
262	9			HEC_0020.1339.1349.2.pkl.txt	[396.35]GGNAGGVGGADVNNH[157.85][215.56]	
262	9	6.7	1.1	HEC_0020.1583.1588.0.pkl.txt	[414.39]GGHVVADNEQFI[373.44]	
257	10			HEC_0020.1708.1711.2.pkl.txt	[414.50]NIGGVVSGGVEQFI[373.44]	
257	9			HEC_0020.1924.1936.2.pkl.txt	[253.10]RQSPGGAGSGTVVNIHS[286.17]	
250	8			HEC_0020.2532.2547.2.pkl.txt	[358.21]AVINVVSTGAAYGPI[373.34]	
248	9			HEC_0020.2265.2268.2.pkl.txt	[303.48]SNPGPVVGGHGGGAPIH[158.19][215.25]	

Figure 66 Sherenga top-level report

SS Sherenga Score for the most likely interpretation

TL Tag Length - sequence tag length

TS Tag Score - peptide score for MS/MS Search results

**MS/MS Search
Delta MH⁺** Difference in mass between experimental MH⁺ and MH⁺ of database match

Filename Filename for detailed Sherenga results. These 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)

Sherenga Sequence Tag	Amino acid sequence for Sherenga interpretation with highest score. A bracketed number in the sequence indicates a mass segment that could not be reliably interpreted because there were insufficient cleavages to assign amino acids or because there was an unusual modification of an amino acid.
MS/MS Search Result	Amino acid sequence for best MS/MS Search interpretation

To view detailed Sherenga results

- 1 Click a link under the **Filename** header.
- 2 Check that you see the Spectrum Viewer, as shown at the top of [Figure 67](#), and the Sherenga results list, as shown at the bottom of [Figure 67](#).
The Sherenga results list is comprised of multiple Sherenga interpretations for the same spectrum. The most probable interpretation (**Sherenga 1**) is initially displayed in the Spectrum Viewer.
- 3 Click the **Rank** arrow buttons in the lower right-hand corner of the Spectrum Viewer to display the various interpretations in the Sherenga results list. See [“To use the Spectrum Viewer”](#) on page 83 of [Chapter 2](#) for more details.

NOTE

If you see an **MStag 1** sequence at the top of the Spectrum Viewer, first click the **Sherenga 1** sequence and make sure the background for the Sherenga sequence turns from gray to white. Then click the **Rank** arrow buttons.

- 4 Scroll below the Sherenga results list to view details on the interpretation.
- 5 If you have additional information about the sequence, you can evaluate your own sequence versus the spectrum.
 - a Be sure that you have selected the Sherenga sequence as described in the note just above. That way your sequence will append to the Sherenga list.
 - b Type your sequence in the box to the left of the **Go** button in the Spectrum Viewer. Where you have missing information, enter mass gaps in brackets, as illustrated with the default sequence shown in the box.

4 Sherenga *de novo* Sequencing

You can enter these mass gaps anywhere in the sequence, including the middle or ends.

- c Click the **Go** button.
- d Select **Bars** options to see how the theoretical fragment ions from your sequence line up with the spectrum.
- e Use the **Rank** arrow buttons (<- and ->) to go from the sequences that were identified by Sherenga *de novo* Sequencing to the sequence that you typed.

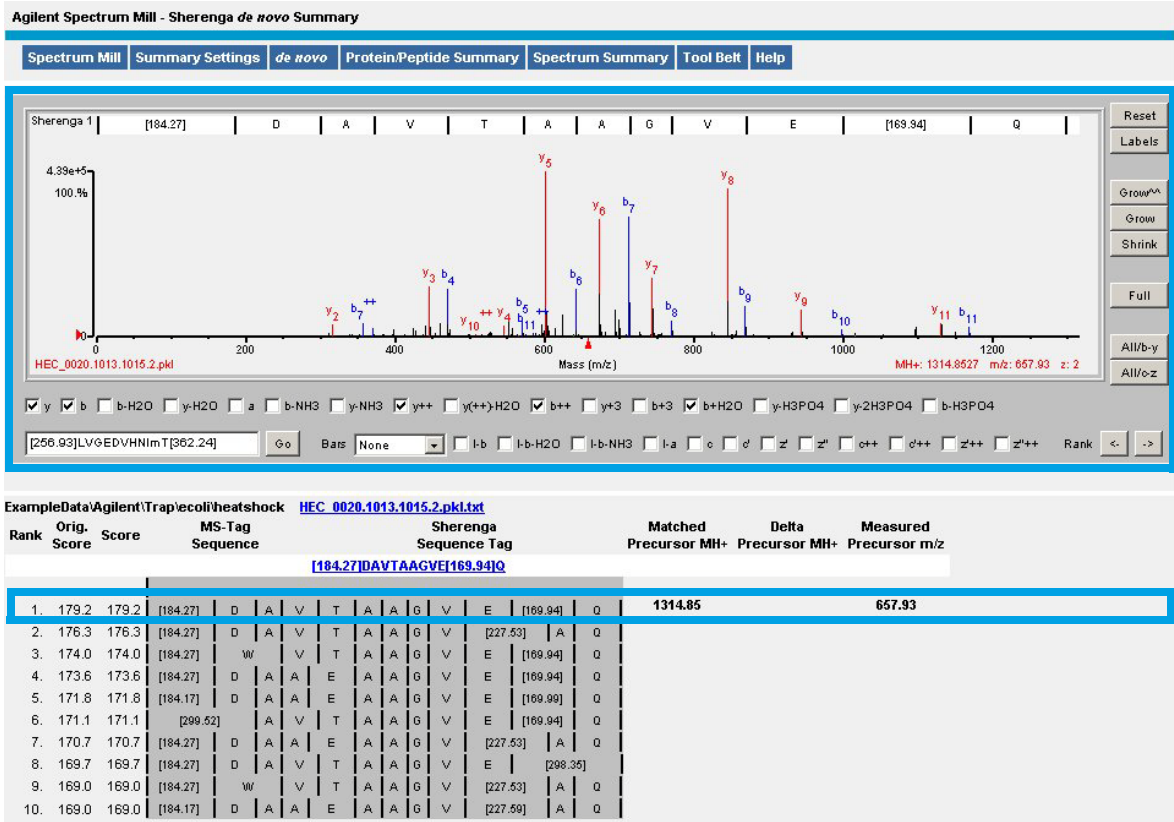
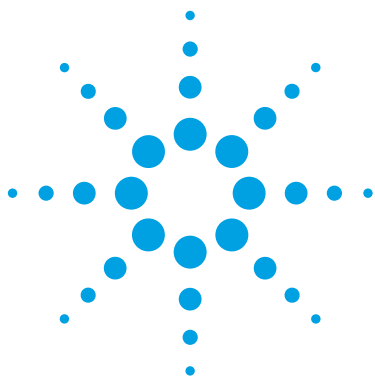


Figure 67 First Sherenga interpretation outlined in rectangles in both the Spectrum Viewer (top) and Sherenga results list (bottom)

To compare Sherenga results with MS/MS Search results

- In the detailed Sherenga results, compare the *de novo* sequencing results with any available MS/MS Search results.

When database search results are available, the software displays them along with the Sherenga interpretations. The database sequence (labeled **MSTag 1**) is displayed just above the Sherenga sequence shown at the top of [Figure 67](#). It is also displayed above the Sherenga results list shown at the bottom of [Figure 67](#).



5 Processing Data for Differential Expression Quantitation

- To use the Data Extractor for a differential profiling study 133
- To use MS/MS Search for a differential profiling study 135
- To calculate DEQ ratios for isotopic labels using the Protein/Peptide Summary page 136
- To interpret DEQ results for isotopic labels in Peptide - Spectrum Match mode 137
- To interpret DEQ results for isotopic labels in protein modes 139
- To calculate iTRAQ and TMT ratios using the Protein/Peptide Summary page 140
- To interpret results for iTRAQ labels in peptide mode 141
- To interpret results for iTRAQ labels in protein mode 142
- To view light/heavy results on the Spectrum Summary page 143

In this chapter, you learn in detail how to process data for differential expression quantitation (DEQ). You learn how to process stable isotope labeling by amino acids in cell culture (SILAC), isotope-coded affinity tag (ICAT) and other isotopically labeled data to compute light/heavy ratios. You also learn how to do quantitation with isobaric modifications, which use isobaric tags with low-mass MS/MS signature ions rather than isotopic labels (iTRAQ and TMT).

Labeled data is in general processed the same way as any other MS/MS data, so you follow the step-by-step procedures described in [Chapter 1](#), “Processing MS/MS Data Interactively”. As you process the data, you set specific parameters in the Data Extractor, MS/MS Search, Protein/Peptide Summary, and Spectrum Summary pages to enable differential expression quantitation. This chapter describes those settings.



At installation, the software supports the following modifications for differential expression quantitation:

- ICAT D₀/D₈
- cICAT ¹²C/¹³C
- C-terminal methyl ester D₀/D₃
- N-terminal propionyl D₀/D₅
- C-terminal ¹⁶O/¹⁸O
- iTRAQ and TMT reagents
- Metabolic labels:
 - SILAC 2 (Arg 0-6Da, Lys 0-8Da)-mix
 - SILAC 3 (Arg 0-6-10Da)-mix
 - ¹⁵N and ¹⁴N/¹⁵N mix

The software allows you to add custom modifications to support quantitation studies. See the online help for system administrators. Additional SILAC and other custom modifications are defined in **smconfig.misc.xml**.

Note that with the exception of iTRAQ and TMT reporter ion quantification, differential expression calculations are not supported for data that uses the generic Data Extractor. Because the generic extractor processes exported rather than raw files, it retains less functionality than the raw file extractors.

To use the Data Extractor for a differential profiling study

- 1 Click the **Choose...** button to set the appropriate **Fixed/Mix Modification**. [Table 7](#) lists the supported labels and the corresponding options you must select. If your sample contains a mixture of labels and you want the software to calculate the ratio, select one of the **mix** options.
- 2 Set other parameters as described in [Chapter 1](#).

Table 7 Reagents for differential expression quantitation and corresponding Spectrum Mill settings

Reagent	Fixed/Mix Modification listed under:	Option to select
ICAT D ₀ /D ₈ (original ICAT reagent)	Cysteine	ICAT-D0, ICAT-D8, or ICAT-mix
cICAT ¹² C/ ¹³ C (newer, cleavable ICAT reagent)	Cysteine	cICAT-C12, cICAT-C13, or cICAT-mix
N-terminal propionyl D ₀ /D ₅	N-terminus	Propionyl-D0, Propionyl-D5, or Propionyl-mix
iTRAQ/TMT	N-terminus and K	iTRAQ or iTRAQ Partial-mix (the latter if you suspect incomplete labeling) or TMT
C-terminal methyl ester D ₀ /D ₃	C-terminus, D, and E	Methyl_Ester, Methyl_Ester-D3, or Methyl_Ester-mix
C-terminal ¹⁶ O/ ¹⁸ O	C-terminus	O18 Free Acid or O16/O18 Free Acid-mix
¹⁵ N and ¹⁴ N/ ¹⁵ N mix	Metabolic	N15 or N14/N15 mix
SILAC	Metabolic	Any of the SILAC options (All are mixes.)

Extractor details for all -mix modifications except iTRAQ, TMT and ¹⁴N/¹⁵N mix

The **mix** options ensure that all the necessary chromatogram peak areas are calculated as the data are extracted. Data Extractor always calculates peak areas from the extracted ion chromatograms (EICs) of each precursor ion subjected to MS/MS.

When the modification is set to a mixture of light/heavy labels, Data Extractor calculates peak areas for a number of additional parallel EICs, one of which represents the other member of the light/heavy pair. More EICs are calculated than are actually necessary for the light/heavy computation. The reason for this in the case of ICAT quantitation is that Data Extractor does not know

whether it has extracted a cysteine-containing peptide, a light member of the pair, or a heavy member of the pair. It also does not know how many cysteines are present, nor the charge state of the precursor ions.

All the parallel EIC areas are stored until the Spectrum Mill workbench's interpretation is complete. Then the correct ones are retrieved for ICAT calculations. As a result, EICs are calculated for both the light and heavy labels, even if only one is subjected to MS/MS. A similar process occurs for SILAC, which can include light/medium/heavy labels.

The generic Data Extractor (the one that processes peak lists) does not have access to the raw mass spectral files, so it cannot calculate the EICs that are needed for ICAT-like quantitation.

Extractor details for iTRAQ and TMT

The iTRAQ and TMT intensity calculations do not require EICs. The abundances for the reporter ions are calculated from the MS/MS data.

You can calculate iTRAQ and TMT ratios for generic extractor data.

Extractor details for $^{14}\text{N}/^{15}\text{N}$ mix

For $^{14}\text{N}/^{15}\text{N}$ mix, quantitation occurs at the report level rather than at the extractor level. The Data Extractor does not calculate parallel EICs; so quantitation is possible only when MS/MS is run on both the ^{14}N and the ^{15}N members of a peptide pair, and the MS/MS Search results indicate the same sequence and charge. The calculations assume complete incorporation of ^{15}N .

The Data Extractor provides quantitative information for all ions for which MS/MS is run, and the ratios are calculated in Protein/Peptide Summary based on this information. You can calculate $^{14}\text{N}/^{15}\text{N}$ ratios with generic extractor data.

CAUTION

If you wish to do quantitation for labels with small mass differences (~4 Da) between the light and heavy versions, change **Merge scans with same precursor m/z** to **+/-1.0 m/z** or lower.

To use MS/MS Search for a differential profiling study

- 1 Click the **Choose...** button to set the appropriate **Fixed/Mix Modification**. See [Table 7](#) on page 133 for details.
- 2 Set other parameters as described in [Chapter 1](#), but only mark **Proton mobility scoring** if the MS/MS spectra are acquired on an ion trap mass spectrometer.

With the exception of iTRAQ and TMT, if you select one of the variations that ends in **mix**, each spectrum is searched multiple times – once for each possible label. The results are merged as a single output.

For iTRAQ and TMT, only a single search is necessary. Since the tags are isobaric, all versions of the iTRAQ or TMT label, whether iTRAQ4, iTRAQ8, TMT2 or TMT6, are simultaneously fragmented during MS/MS. Further, all iTRAQ labels produce the same MS/MS fragments for a given precursor peptide, as do TMT labels. Therefore, neither the iTRAQ nor TMT labels have to be searched as a mix. However, the tags produce different peaks depending on the mass associated with the tag, and the abundances of these peaks are used by the Spectrum Mill workbench for relative quantitation.

For convenience, **mix** definitions are provided by default, and can be used to investigate the labeling process. For example, you can use them to evaluate incomplete peptide modifications, or to search for a lysine modification with or without an accompanying N-term modification.

To calculate DEQ ratios for isotopic labels using the Protein/Peptide Summary page

Use the Protein/Peptide Summary page to calculate DEQ ratios.

- 1 Mark the **DEQ ratios** (differential expression quantitation ratios) check box.

NOTE

Only for protein modes: When you mark the DEQ ratios check box, you can select the Protein Quantitation Option to exclude outlier DEQ ratios (>2 std dev from mean).

- 2 If you want to see your modification in a report that shows peptides:
 - Mark the check box for **N-terminus**, **C-terminus**, or **Cysteines**, if your label reacts at one of those sites. (For example, mark **Cysteines** for ICAT reagents.)
 - Mark the check box for **Modification Names** if your label reacts at an amino acid other than cysteine.
- 3 For ICAT and similar labels, when you want to see the pairs together in a peptide report, set **Sort peptides by** to **Sequence**.
- 4 (Optional) Use the **Required AAs** box to filter results so that only the isotopically labeled peptides are shown. For example, select **C** in the **Required AAs** box if the isotopic label modifies cysteine.
- 5 Set other options.
- 6 Click **Summarize**.

To interpret DEQ results for isotopic labels in Peptide - Spectrum Match mode

- Check that your results resemble those in [Figure 68](#), which shows ICAT results displayed in Peptide - Spectrum Match mode on the Protein/Peptide Summary page.
- Note the columns labeled **Cysteines** and **L/H**. **L/H** means light/heavy. For SILAC reagents with three labels, you also see **L/M**, for light/medium.
- If **n/c** appears in the report, it means not calculated. If the Data Extractor cannot determine a charge for a peptide, it assumes a charge of +2 for determining the mass shifts for quantitation, and it looks for up to two modification sites in the peptide (e.g., two cysteines at most). When the actual charge is not +2, or when there are more than two modification sites in the peptide, the ratio is not calculated, and is reported as **n/c**.

Ratios are also reported as **n/c** when the peptide does not contain the amino acid that reacts with the labeling reagent.

Agilent Spectrum Mill - Protein/Peptide Summary														
Spectrum Mill		Summary Settings		Autovallation		Easy MS/MS		MS/MS Search		Spectrum Summary		Build TIC	Tool Belt	Help
Results Shown Filtered by Validation Category: valid														
Data Directory: msdataSM\ExampleData\ICAT														
hit table read - SpecFeatures read														
valid hits read from tagSummary file - Files: 41 Hits: 40														
#	Filename	z	Score	Fwd-Rev Score	SPI (%)	Spectrum Intensity	Sequence	Cysteines	L/H	MH+ Matched (Da)	Accession #	Protein Name		
1	ICAT_Mix.1064.1064.0	2	14.17	14.17	85.1	8.13e+006	(K)AGLCQTFVYGGCR(A)	ICAT-D0	0.891	2259.079	115114	Aprotinin		
2	ICAT_Mix.0837.0842.2	2	14.02	14.02	71.5	1.03e+008	(K)CDEWSVNSVCK(I)	ICAT-D8	0.717	1673.811	4557871	transferrin		
3	ICAT_Mix.0843.0847.2	2	14.13	14.13	75.2	8.62e+007	(K)CDEWSVNSVCK(I)	ICAT-D0	0.812	1665.761	4557871	transferrin		
4	ICAT_Mix.1083.1083.2	2	8.88	8.88	78.4	2.94e+007	(K)CSTSSLLLEACTFR(R)	ICAT-D8	0.266	2318.195	4557871	transferrin		
5	ICAT_Mix.1095.1098.2	2	11.08	11.08	75.7	1.47e+007	(K)CSTSSLLLEACTFR(R)	ICAT-D0	1.090	2302.095	4557871	transferrin		
6	ICAT_Mix.1067.1067.0	3	11.61	11.61	97.3	5.54e+007	(K)DAIPENLPPLTADFAEDKDVCK(N)	ICAT-D8	n/c	2851.434	1351907	bovine serum alk		
7	ICAT_Mix.1068.1073.0	3	19.67	19.67	94.4	5.00e+007	(K)DAIPENLPPLTADFAEDKDVCK(N)	ICAT-D0	n/c	2843.384	1351907	bovine serum alk		
8	ICAT_Mix.1072.1072.0	3	10.74	10.74	95.6	6.29e+007	(K)DAIPENLPPLTADFAEDKDVCK(N)	ICAT-D8	n/c	2851.434	1351907	bovine serum alk		
9	ICAT_Mix.0863.0863.0	2	7.91	7.19	89.2	1.72e+007	(K)DDPHACYSTVFDR(L)	ICAT-D0	1.559	1939.857	1351907	bovine serum alk		
10	ICAT_Mix.0997.0997.0	2	12.58	12.58	77.4	5.52e+007	(K)DYELLCLDGT(R)	ICAT-D8	0.492	1747.884	4557871	transferrin		
11	ICAT_Mix.1002.1008.0	2	21.43	21.43	98.2	3.80e+007	(K)DYELLCLDGT(R)	ICAT-D0	0.591	1739.834	4557871	transferrin		
12	ICAT_Mix.1003.1003.2	2	12.83	12.83	74.9	5.84e+007	(K)DYELLCLDGT(R)	ICAT-D8	0.497	1747.884	4557871	transferrin		
13	ICAT_Mix.0844.0844.0	3	11.28	11.28	73.6	1.77e+007	(R)EGTCPEAPTDECKPVK(W)	ICAT-D8	n/c	2604.312	4557871	transferrin		
14	ICAT_Mix.0922.0922.0	2	8.25	8.25	63.2	2.14e+007	(K)EYEATLECCAR(D)	ICAT-D8	0.339	2289.121	1351907	bovine serum alk		
15	ICAT_Mix.0967.0967.0	2	12.36	12.36	84.0	4.36e+007	(R)EDRRSRGCDGSR(K)	ICAT-D8	0.416	1970.911	4557871	transferrin		

Figure 68 ICAT results shown in Peptide - Spectrum Match mode

In the Peptide - Spectrum Match mode, you may see a calculation discrepancy in cases where both members of a D₈/D₀ ICAT pair have been subjected to MS/MS. This is because the heavy- and light-labeled peptides often do not quite co-elute, so the EICs used for the calculations are slightly offset. (The EIC windows are determined by the Data Extractor setting for the time window for merging scans.)

You do not see this discrepancy in the protein mode, *provided* that both the D₈- and D₀-labeled precursor ions were subjected to MS/MS and that these results were of sufficient quality to be interpreted and included in the final results summary. In the protein modes, the peptide ratios are combined to calculate a ratio for the protein. In these modes, the ratio for the ICAT pair is recalculated directly from the EICs of each precursor, rather than the parallel EICs from the calculated m/z shifts from the precursor.

To interpret DEQ results for isotopic labels in protein modes

- Check that your results resemble those in [Figure 69](#), which shows ICAT results displayed in one of the protein modes on the Protein/Peptide Summary page.
- Note the special light/heavy columns labeled **L/H (mean)**, **L/H (std dev)**, and **L/H (# pairs)**. For SILAC reagents with three labels, you also see columns for **L/M (mean)** and **L/M (std dev)**.

Agilent Spectrum Mill - Protein/Peptide Summary

Spectrum Mill

Summary Settings

Autovalidation

Easy MS/MS

MS/MS Search

Spectrum Summary

Build TIC

Tool Belt

Help

Results Shown Filtered by Validation Category: valid

Data Directory: msdataSM\ExampleData\ICAT

hit table read - SpecFeatures read

valid hits read from tagSummary file - Files: 41 Hits: 40

beginning to assemble proteins proteins assembled 0.005125 sec

proteins filtered by unique peptides 0.003677 sec

proteins filtered by score

calculated protein coverage maps 0.039664 sec

beginning to roll up proteins into groups ... proteins rolled up into groups 0.002366 sec

protein groups ready for display

proteinGroupingMethod: oneSharedPeptide 6 Proteins listed

Group (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Mean Peptide Spectral Intensity	L/H (mean)	L/H (std dev)	L/H (# pairs)	Protein MW (Da)	Database Accession #	Protein Name
1	24	10	148.36	19	3.15e+007	0.72	0.23	11	77050.4	4557871	transferrin
2	10	7	83.55	16	2.67e+007	0.91	0.47	5	69293.9	1351907	bovine serum albumin
3	2	1	17.20	2	4.40e+007	0.84	0.00	1	53354.2	71826	fibrinogen beta chain
4	2	1	16.34	3	4.52e+007	0.96	0.00	1	42881.5	129293	ovalbumin
5	1	1	14.17	13	8.13e+006	0.89	0.00	1	10902.8	115114	Aprotinin
6	1	1	11.20	1	1.00e+007	1.23	0.00	1	116483.5	114939	beta-galactosidase
Totals:	40	21									

Figure 69 ICAT results shown in Protein Summary mode

The protein modes all work similarly. First, the interpreted spectra for peptides are grouped because they correspond to a single protein. Then a light/heavy ratio for the protein is calculated from the mean of the values for the peptides. The protein modes report the mean, standard deviation, and number of values that contribute to the mean. Note that the standard deviations are larger for low-level, noisy data.

To calculate iTRAQ and TMT ratios using the Protein/Peptide Summary page

Use the Protein/Peptide Summary page to calculate iTRAQ and TMT ratios. In the iTRAQ calculations, the Spectrum Mill workbench incorporates the correction factors from your iTRAQ certificate of analysis.

iTRAQ reagent 8-plex does not require special correction factors. Default values that apply to all batches are included in the Spectrum Mill calculations.

You must first use the Tool Belt page to enter the correction factors and apply them to your sample.

- 1 Do the following on the Tool Belt page:
 - a Enter the correction factors from the iTRAQ certificate of analysis that you received for the reagents that you used to process your sample. See [“To create a file of reporter ion correction factors”](#) on page 163.
 - b Apply the correction factors from the iTRAQ certificate of analysis to the data directory for your sample. See [“To apply a file of reporter ion correction factors”](#) on page 165.
- 2 Do the following on the Protein/Peptide Summary page:
 - a Select the iTRAQ or TMT (Tandem Mass Tag) option you are using for your quantitation: iTRAQ4, iTRAQ8, TMT2, TMT6, or TMT10.
 - a Mark the **intensities** check box (for modes that show peptides).
 - b Mark the **Reporter Ratios** check box and select the reporter mass you wish to use as the control (denominator) for the ratio calculations.

If you have multiple control ions, use the **Ctrl** key to select them. Ratios will then be calculated and reported for each control ion.
 - c If you want to see the iTRAQ modification in a report that shows peptides, mark check boxes for both **N-terminus** and **Modification Names**, since the reagents react at both the N-terminus and lysines.
 - d Set other options.
 - e Click **Summarize**.

To interpret results for iTRAQ labels in peptide mode

- Check that your results resemble those in [Figure 70](#), which shows iTRAQ results displayed in Peptide - Spectrum Match mode on the Protein/Peptide Summary page.
- Confirm that you see a message near the top of the report that says the correction factors were found, and confirm that the batch number is correct.
- Note the columns that show the iTRAQ intensities and ratios.

Agilent Spectrum Mill - Protein/Peptide Summary																			
Spectrum Mill		Summary Settings		Autovaildation		Build TIC		MS/MS Search		Spectrum Summary		Tool Belt		Help					
Results Shown Filtered by Validation Category: valid																			
Data Directory: msdata\SMExampleData\iTRAQ																			
hit table read - SpecFeatures read																			
Correction factors for iTRAQ batch 0501017 found.																			
valid hits read from tagSummary file - Files: 17 Hits: 17																			
#	Filename	z	Score	SPI (%)	Spectrum Intensity	N-term	Sequence	Modifications	iTRAQ 114	iTRAQ 115	iTRAQ 116	iTRAQ 117	iTRAQ 118	iTRAQ 115/114	iTRAQ 116/114	iTRAQ 117/114	MH-Matched (Da)	Accession #	Protein Name
1	24prot_114_116.0229.1323.2	2	15.08	99.6	3.25e+004	ITRAQ	(K)GILADESTGSIAR(R)	K:iTRAQ	1074	88	968	15		0.000	0.898	0.000	1620.905	P04075	Fructose-bisphosphate aldolase
2	24prot_114_116.0200.1158.3	3	14.80	93.5	7.00e+004	ITRAQ	(K)SHGQDYLVGNR(L)		822	56	551	6		0.000	0.667	0.000	1369.700	P13745	Glutathione S-transferase Ya c
3	24prot_114_116.0213.1227.2	2	13.83	87.1	6.90e+004	ITRAQ	(R)ALQASALK(A)	K:iTRAQ	4051	431	2983	174		0.025	0.730	0.010	1089.688	P04075	Fructose-bisphosphate aldolase
4	24prot_114_116.0242.1396.4	4	13.62	92.9	6.99e+004	ITRAQ	(K)KTEREDLIATLK(K)	K:iTRAQ	623	37	298			0.000	0.476	0.000	1911.128	P00005	Cytochrome c
5	24prot_114_116.0204.1177.3	3	13.29	98.7	5.09e+004	ITRAQ	(K)SHGQDYLVGNR(L)	K:iTRAQ	780	63	483	13		0.003	0.615	0.000	1505.796	Q28035	Glutathione S-transferase alpha
6	24prot_114_116.0205.1182.3	3	12.92	71.8	3.86e+004	ITRAQ	(R)LQSIGYENTENRR(F)		2008	189	1292	74		0.016	0.638	0.008	1790.912	P04075	Fructose-bisphosphate aldolase
7	24prot_114_116.0251.1446.3	3	11.50	100.0	2.90e+004	ITRAQ	(K)YSHRIAMATVTLR(R)		191	6	104			0.000	0.543	0.000	1835.945	P00883	Fructose-bisphosphate aldolase
8	24prot_114_116.0227.1307.3	3	11.34	77.7	4.82e+004	ITRAQ	(K)ADDERPPFPQVIR(S)	K:iTRAQ	1417	167	856	30		0.041	0.598	0.000	1630.916	P04075	Fructose-bisphosphate aldolase
9	24prot_114_116.0207.1196.3	3	9.58	87.1	4.99e+004	ITRAQ	(K)KIFVQR(C)	K:iTRAQ	845	43	479			0.000	0.565	0.000	1194.794	P04657	Cytochrome c-1 (Cytochrome c
10	24prot_114_116.0215.1245.2	2	9.41	74.2	8.01e+004	ITRAQ	(K)YNLYQR(D)	K:iTRAQ	4808	371	3127	201		0.000	0.645	0.013	1045.593	Q28035	Glutathione S-transferase alpha
11	24prot_114_116.0240.1386.3	3	9.39	84.0	3.49e+004	ITRAQ	(R)KLFEELAR(S)	K:iTRAQ	421	31	231			0.000	0.545	0.000	1293.777	P11974	Pyruvate kinase, M1 isozyme (E
12	24prot_114_116.0254.1466.3	3	9.30	93.3	2.41e+004	ITRAQ	(K)TEREDLIATLK(K)	K:iTRAQ	364	8	165			0.000	0.452	0.000	1638.931	P00005	Cytochrome c
13	24prot_114_116.0210.1210.3	3	9.28	85.0	2.12e+004	ITRAQ	(R)EAEANFHR(K)		160	8	98			0.000	0.610	0.000	1205.586	P11979	Pyruvate kinase, M1 isozyme (E
14	24prot_114_116.0227.1311.4	4	8.74	83.3	3.70e+004	ITRAQ	(K)TGPNLHGLFGRR(T)	K:iTRAQ	426		196			0.000	0.460	0.000	1584.922	P00020	Cytochrome c
15	24prot_114_116.0206.1185.3	3	8.09	85.9	3.13e+004	ITRAQ	(K)ELSDIAHR(I)		167	14	91			0.008	0.541	0.000	1084.587	P04075	Fructose-bisphosphate aldolase
16	24prot_114_116.0244.1406.2	2	7.95	95.0	3.66e+004	ITRAQ	(R)GDLCIEIIPAER(V)	K:iTRAQ	951	52	711	9		0.000	0.745	0.000	1429.815	P11979	Pyruvate kinase, M1 isozyme (E
17	24prot_114_116.0202.1166.2	2	7.32	88.9	7.16e+004	ITRAQ	(R)IVAPGR(G)	K:iTRAQ	5972	527	4452	204		0.007	0.740	0.001	872.581	P04075	Fructose-bisphosphate aldolase
17 Files listed.																			

Figure 70 iTRAQ results shown in Peptide - Spectrum Match mode

To interpret results for iTRAQ labels in protein mode

- Check that your results resemble those in [Figure 71](#), which shows iTRAQ4 results displayed in one of the protein modes on the Protein/Peptide Summary page.
- Confirm that you see a message near the top of the report that says the correction factors were found, and confirm that the batch number is correct.
- Note the columns that show the means and standard deviations of iTRAQ ratios, as well as the number of iTRAQ pairs. For the proteins, the software calculates the means and standard deviations of the iTRAQ ratios from the component peptides.

Agilent Spectrum Mill - Protein/Peptide Summary														
Spectrum Mill Summary Settings Autovalidation Build TIC MS/MS Search Spectrum Summary Tool Belt Help														
Results Shown Filtered by Validation Category: valid Data Directory: msdata\SMExampleData\iTRAQ hit table read - SpecFeatures read Correction factors for iTRAQ batch 0501017 found. valid hits read from tagSummary file - Files: 17 Hits: 124 beginning to assemble proteins ... proteins assembled 0.012857 sec proteins filtered by unique peptides 0.019828 sec proteins filtered by score calculated protein coverage maps 0.108943 sec beginning to roll up proteins into groups ... proteins rolled up into groups 0.014742 sec protein groups ready for display proteinGroupingMethod: oneSharedPeptide 5 Proteins listed														
Group (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Mean Peptide Spectral Intensity	iTRAQ 115/114 (mean)	iTRAQ 115/114 (std dev)	iTRAQ 116/114 (mean)	iTRAQ 116/114 (std dev)	iTRAQ 117/114 (mean)	iTRAQ 117/114 (std dev)	iTRAQ (# spectra)	Database Accession #	Protein Name
1	7	7	80.08	21	4.57e+004	0.01	0.01	0.67	0.12	0.00	0.00	7	P00883	Fructose-bisphosphate aldolase A (EC 4.1.2.13) (Muscle-type aldolase)
2	4	4	41.24	28	4.52e+004	0.00	0.00	0.49	0.04	0.00	0.00	4	P00004	Cytochrome c
3	3	3	26.62	5	3.09e+004	0.00	0.00	0.63	0.08	0.00	0.00	3	P11974	Pyruvate kinase, M1 isozyme (EC 2.7.1.40) (Pyruvate kinase muscle isozyme 1)
4	2	2	24.21	7	7.50e+004	0.00	0.00	0.66	0.01	0.01	0.01	2	P30115	Glutathione S-transferase Yc (EC 2.5.1.18) (GST class-alpha)
5	2	2	22.70	7	6.55e+004	0.00	0.00	0.63	0.02	0.01	0.01	2	P09210	Glutathione S-transferase A2 (EC 2.5.1.18) (GTH2) (HA subunit 2) (GST-gamma)
Totals:	18	18												

Figure 71 iTRAQ results shown in Protein Summary mode

To view light/heavy results on the Spectrum Summary page

- To view ICAT or other light/heavy intensities on the Spectrum Summary page, mark the check box for **MS L/H EIC intensity**. Check that the results are similar to those in [Figure 72](#).

Agilent Spectrum Mill - Spectrum Summary

Spectrum Mill Summary Settings Protein/Peptide Summary Build TIC MS/MS Search Tool Belt Help

Down Up ICAT_Mix.1164.1164.2.pkl 1 of 105 Update

reset good-spectrum bad-spectrum Click to set category of current spectrum

Data Directory: msdata\SM\ExampleData\ICAT
Validation Preset: all Input files: 105 - 105 files (after filtering by maxSequenceTagLength > 3)

Validation category	#	Filename	z	Precursor m/z	Precursor m/z	# Detected Peaks	MS Precursor EIC Intensity	Light Intensity	Heavy Intensity	Light Intensity 2 sites	Heavy Intensity 2 sites	MS Order	2nd Precursor m/z	Max Tag Length	Longest Sequence Tag
reset	1	ICAT_Mix.1164.1164.2	2	2127.0300	1064.0200	25	1.01e+007	1.53e+005	1.73e+007	1.53e+006	4.18e+005	2	0.0000	9	[401.15]LLSLYDETIG732.63
reset	2	ICAT_Mix.0738.0738.2	2	1703.2500	852.1300	25	7.27e+006	2.02e+006	7.56e+006	4.30e+006	8.03e+005	2	0.0000	8	[447.21]HESVAEREI317.78
reset	3	ICAT_Mix.0843.0847.2	2	1666.3300	833.6700	25	8.62e+007	2.69e+006	1.06e+008	1.80e+006	6.40e+006	2	0.0000	8	[402.27]TVSWEDRMV202.78
reset	4	ICAT_Mix.1168.1173.2	2	1816.2100	908.6100	25	9.26e+006	2.66e+006	8.67e+005	3.90e+006	3.52e+006	2	0.0000	8	[389.16]NVLVGGQMMI371.64
reset	5	ICAT_Mix.0837.0842.2	2	1674.3500	837.6800	25	1.01e+008	7.40e+007	5.71e+006	2.27e+006	1.24e+007	2	0.0000	7	[503.22]IVSWFFNMJ221.06
reset	6	ICAT_Mix.0963.0963.0	2	1813.2300	907.1200	22	3.53e+007	1.56e+007	1.26e+007	2.15e+006	1.08e+007	2	0.0000	7	[696.29]NEDGFLT336.60
reset	7	ICAT_Mix.0989.0989.0	2	2319.9100	1160.4600	25	1.11e+007	6.63e+006	4.86e+006	6.75e+006	1.75e+007	2	0.0000	7	[923.27]IAAFNPAGT67.31
reset	8	ICAT_Mix.1064.1064.0	2	2261.1100	1131.0600	20	8.13e+006	1.93e+006	9.97e+005	3.15e+005	9.12e+006	2	0.0000	7	[719.60]GGYVFTG187.68
reset	9	ICAT_Mix.1083.1090.2	2	2320.3500	1160.6800	25	3.94e+007	1.06e+007	4.31e+006	2.76e+007	7.69e+006	2	0.0000	7	[975.61]AELLSSIT641.83
reset	10	ICAT_Mix.1104.1104.0	2	2324.5700	1162.7900	25	4.58e+006	5.89e+006	3.44e+006	1.82e+006	7.80e+005	2	0.0000	7	[834.98]VAGNVTI889.22
reset	11	ICAT_Mix.1283.1288.0	2	1711.5300	856.2700	25	9.24e+006	6.46e+006	4.42e+006	5.33e+005	2.23e+006	2	0.0000	7	[497.23]MPHQFVL271.95
reset	12	ICAT_Mix.0699.0703.0	2	1055.0700	528.0400	25	2.12e+007	1.47e+006	5.15e+006	4.55e+006	2.90e+006	2	0.0000	6	[224.80]TNHAEAI204.64

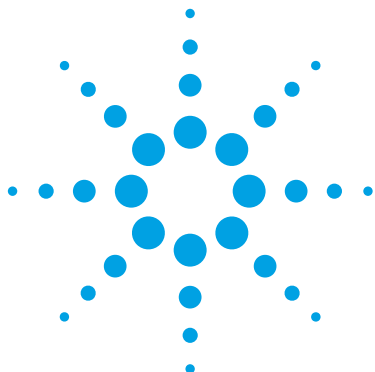
Figure 72 ICAT intensities shown on the Spectrum Summary page

NOTE

The Spectrum Summary page shows the intensities of the parallel light/heavy EICs calculated during data extraction. Since you typically use the Spectrum Summary page to display spectra that have not been interpreted, you do not see the light/heavy ratios calculated in this table, nor should you attempt to calculate them from these data. Instead, use the calculations from the Protein/Peptide Summary page, as described earlier in this chapter.

Note that the values are reported as 0's for "metabolic" modifications, such as SILAC and $^{14}\text{N}/^{15}\text{N}$ mixes.

5 Processing Data for Differential Expression Quantitation



6

Processing MS-Only Data Interactively

- Step 1. Find the compounds of interest in MassHunter Qualitative Analysis 146
- Step 2. Copy and paste compound masses into PMF Search 147
- Step 3. Search the MS-only data and display results 148
- Step 4. Examine the results 151

In this chapter you learn the basics to process MS-only data files with the Spectrum Mill workbench. This chapter guides you through a typical workflow with the software.

PMF Search is limited to the analysis of digests from simple protein mixtures (usually three to five proteins). When there are peptides from too many different proteins in your spectrum, you do not achieve statistically significant scores because the more complex the protein mixture, the more non-matching (noise) peptides for any given protein. Use of the **Mixture scoring** option helps to overcome this limitation.

For Agilent TOF and Q-TOF .d data, you must copy masses from MassHunter Qualitative Analysis and paste them into Manual PMF search to use Spectrum Mill for that data type. This statement is also true for Agilent Mass Profiler and Mass Profiler Professional.



Step 1. Find the compounds of interest in MassHunter Qualitative Analysis

The MFE settings in MassHunter Qualitative Analysis may need to be optimized for a given sample set or analytical condition.

- 1 Open **MassHunter Qualitative Analysis**.
- 2 Under the **Data** folder, select an Agilent TOF or Q-TOF .d file, and click **Open**.
- 3 In Method Explorer, click **Find by Molecular Feature**.
- 4 Under the **Extraction** tab and from the **Target data type** list, select **small molecules (chromatographic)**.
- 5 Click **Ion Species** and clear all check boxes except for **+H** and **-H**.
- 6 Click **Charge State** and from the **Isotope Model** list, select **peptides**.
- 7 Mark **Limit assigned charge states to a maximum of** and type 7.
- 8 Click **Compound Filters**, and clear **Relative height**, then specify an **Absolute height** estimated from an examination of the 5-10 compounds of lowest height/abundance.
- 9 Click **Results** and clear all check boxes in the tab except for the first one.
- 10 Click **Find Compounds by Molecular Feature**.

The compound list should resemble the one in [Figure 73](#).

Agilent MassHunter Qualitative Analysis 8.07.00 - Peptides.m

File Edit View Find Identify Spectra Chromatograms Method Sequence Wizards Actions Configuration Tools Help

Data Navigator

Sort by Data File

- peptide-ms-only.d
 - User Chromatograms
 - Λ + TIC Scan
 - User Spectra
 - Background Spectra
 - Compounds
 - Cpd 1: 2.185
 - Cpd 2: 2.186
 - Cpd 3: 2.212
 - Cpd 4: 2.227
 - Cpd 5: 2.228
 - Cpd 6: 2.332
 - Cpd 7: 2.425
 - Cpd 8: 2.454
 - Cpd 9: 2.473
 - Cpd 10: 2.559
 - Cpd 11: 2.629
 - Cpd 12: 2.648
 - Cpd 13: 2.653
 - Cpd 14: 2.670
 - Cpd 15: 2.691
 - Cpd 16: 2.716

Compound List

Automatically Show Columns

Show/Hide	Cpd	File	Saturated	Start	RT	End	m/z	Mass	Polarity	Ions	Min Z	Max Z
☒	1	peptide-ms-only.d		2.157	2.185	2.223	430.1703	858.3255	Positive	2	2	2
☒	2	peptide-ms-only.d		2.16	2.186	2.219	469.855	1406.5425	Positive	7	2	3
☒	3	peptide-ms-only.d		2.183	2.212	2.236	575.7291	1149.4434	Positive	4	2	2
☒	4	peptide-ms-only.d		2.196	2.227	2.269	489.1914	1464.5494	Positive	9	2	3
☒	5	peptide-ms-only.d		2.206	2.228	2.266	330.1933	658.3718	Positive	2	2	2
☒	6	peptide-ms-only.d		2.292	2.332	2.371	405.0148	404.0074	Positive	2	1	1
☒	7	peptide-ms-only.d		2.401	2.425	2.454	539.8553	1616.5428	Positive	7	2	3
☒	8	peptide-ms-only.d		2.427	2.454	2.49	424.2542	846.4948	Positive	3	2	2
☒	9	peptide-ms-only.d		2.441	2.473	2.51	358.5016	1072.4838	Positive	8	2	3
☒	10	peptide-ms-only.d		2.523	2.559	2.593	443.71	885.4067	Positive	4	1	2
☒	11	peptide-ms-only.d		2.596	2.629	2.665	1052.4007	1051.3915	Positive	8	1	2
☒	12	peptide-ms-only.d		2.612	2.648	2.685	537.28	536.273	Positive	2	1	1
☒	13	peptide-ms-only.d		2.622	2.653	2.688	325.4894	973.4476	Positive	7	2	3
☒	14	peptide-ms-only.d		2.642	2.67	2.698	421.2334	840.4525	Positive	2	2	2
☒	15	peptide-ms-only.d		2.652	2.691	2.748	345.1891	688.3638	Positive	7	1	2
☒	16	peptide-ms-only.d		2.672	2.716	2.758	722.8175	1443.6202	Positive	5	2	2
☒	17	peptide-ms-only.d		2.721	2.758	2.801	570.7362	1139.4579	Positive	9	1	3

Method Explorer: Peptides.m

- MS Target Compound Screening Workf...
- Chromatogram
- Spectrum
- General
- Reports
- Find Compounds
 - Find by Auto MS/MS
 - Find by Targeted MS/MS
 - Find by Molecular Feature
 - Find by Protein Deconvolution
 - Find by MRM
 - Find by Integration

Chromatogram Results

0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3 3.2 3.4 3.6 3.8 4 4.2 4.4 4.6 4.8 5 5.2 5.4 5.6 5.8 6 6.2 6.4 6.6 6.8

Counts vs. Acquisition Time (min)

MS Spectrum Results

Figure 73 MassHunter compound list

Step 2. Copy and paste compound masses into PMF Search

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

- 1 Right-click the **Masses** heading, and select **Copy Column to Clipboard > Using Newline separator**.
- 2 Open the **Spectrum Mill** software.
- 3 Navigate to the **PMF Search** page from the Spectrum Mill home page.

- 4 Click the **Manual PMF** button at the top of the page.
- 5 Paste your mass list into the **Peptide Masses** box.
See [Figure 74](#) on page 149.

Step 3. Search the MS-only data and display results

In this step you enter the remaining parameters and start a search.

- 1 If your masses represent protonated species, click **MH⁺**. If your masses represent neutral species, click **M**.
- 2 If necessary, type charges for the masses in the **Peptide Masses** box.
 - If you clicked **MH⁺**, the software assumes a charge of +1.
 - If you clicked **M**, the software assumes a charge of 0.
 - A charge that you type into the box overrides the **MH⁺/M** settings.
- 3 For **Masses are**:
 - You generally select **Monoisotopic**. With **Monoisotopic**, for H⁺, the mass calculation uses the mass of a proton rather than the mass of a hydrogen atom.
 - Select **Monoisotopic no e- correction** if your data is from an instrument such as MALDI QSTAR where the mass calibration is done in such a way that there is no correction for the loss of the electron for protonated species.
- 4 Set the other parameters. Click the blue divider bars for more information in the online help.
- 5 Click **Start Search**.
- 6 When the search is complete, scroll down and check that you see PMF Search Results, as shown in [Figure 75](#) on page 150.

Agilent Spectrum Mill - Manual PMF Search

Spectrum MillMS EdmanMS DigestMS IsotopeBatch PMF SearchHelp

Search

Start SearchSample ID: Enter_Comment

Maximum reported hits: 25☐ MOWSE scores - Pfactor: 0.4☐ Mixture scoring

Search Parameters

Min. # peptides required to match: 4

Database: SwissProt

Species: All

MW of protein: from 1000 Da to 100000 Da☐ All

Protein pI: from 3.0 to 10.0☒ All

Digest: Trypsin

Maximum # of missed cleavages: 1

Modifications

Choose...

Fixed: Carbamidomethylation (C)

Variable: Acetyl (ProtN-term)
Pyroglutamic acid (N-termQ)
Oxidized methionine (M)

Peptide Masses

Mass tolerance: +/- 10 ppm

Masses are: Monoisotopic

☐ MH+☒ M

Mass (m/z)	Charge (z)
676.3718	
925.4571	
1017.5370	
1026.5964	
1040.6121	
1105.4550	
1196.7132	
1218.5796	
1243.5887	
1255.6538	
1260.5928	
1266.6281	
1276.5836	
1288.6094	
1305.6310	
1388.6975	
1398.6805	
1413.7126	
1420.6741	
1482.8101	
1576.8138	
1932.7067	
2188.0850	

Figure 74 Manual PMF Search page

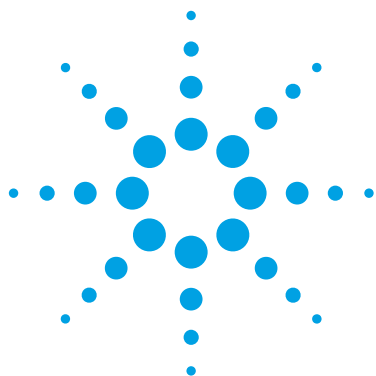
6 Processing MS-Only Data Interactively

PMF Search Results									
Sample ID (comment): Enter_Comment Database searched: NCBIInr.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)LVTDLT(K)		
789.4719	789.4716	0.3	1	1	257	263	(K)LVTDLT(K)		
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 75 Manual PMF Search results

Step 4. Examine the results

- To see more search details, scroll down or click links under the **Rank** heading shown in [Figure 75](#). Note that you see detailed results for only the top two hits.
- To view information in the database you searched, click a link under the **Accession #** header.
- To view amino acid coverage, click a link under the **Protein Coverage** header. In the coverage map:
 - Red = matched
 - Black = not matched
 - Blue = modified amino acid, e.g., cysteine
 - Green = consensus N-linked glycosylation site



7 Using the Tool Belt

- To terminate a process 154
- To create a saved results file so you can search previous hits 156
- To create an MS/MS Search summary file if search terminated abnormally 158
- To create a summary table of previously-used parameters 159
- To list details about amino acid modifications 161
- To create a file of reporter ion correction factors 163
- To apply a file of reporter ion correction factors 165
- To calculate discriminant scoring coefficients 166
- To copy files to your collections directory 168
- To export PepXML files 169
- To convert spectra 171

This chapter describes helpful tools that you can use when you process data with the Spectrum Mill workbench. These tools help you accomplish a range of tasks, from terminating processes to summarizing identification statistics. You access these tools from the Tool Belt page.

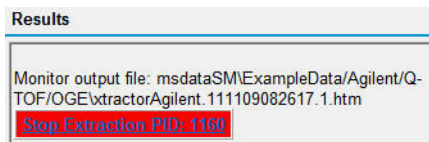


To terminate a process

The Spectrum Mill workbench performs functions like Data Extractor, MS/MS Search, and Sherenga *de novo* Sequencing by launching a parent script, which then repetitively launches child programs. The **Stop** button for a program aborts the parent script and prevents new child programs from starting. However, it does not abort the child program that is currently running. The child program runs until its normal conclusion unless it is aborted separately through the **Stop Process** button on the Tool Belt page.

To use the Stop button on the main program page

Processes (such as Data Extractor, MS/MS Search, and Sherenga *de novo* Sequencing) that take longer than a few seconds have a red **Stop** process link that is displayed in the **Results** window where the process runs.



- 1 To abort a process, click the **Stop** process link.
- 2 Wait a few seconds until a message is displayed to indicate that the process has stopped.
- 3 Stop child programs with the **Stop Process** button on the Tool Belt page. See the next section.

To use the Stop Process button on the Tool Belt page

The **Stop** button described above terminates the parent script, but not child programs. Go to the Tool Belt to terminate these child programs:

- 1 Navigate to the Tool Belt page and select the **Stop process** option, as shown in Figure 76. (This is the default.)
- 2 Click the **List Processes** button to list Spectrum Mill processes that are currently running on the server.
- 3 Select the processes you wish to stop.
- 4 Click the **Stop Process** button.
- 5 Click **List Processes** again to make sure the stopped processes no longer appear in the list.

Agilent Spectrum Mill - Tool Belt Utilities

Spectrum Mill

Extractor

MS/MS Search

Protein/Peptide Summary

Autovalidation

Quality Metrics

Select a Tool:

☒ Stop process

☐ List modifications details

☐ File collector

☐ Create saved results file

☐ Create Reporter Ion correction factors

☐ Export PepXML

☐ Create MS/MS search summary file

☐ Apply Reporter Ion correction factors

☐ Convert spectra

☐ View parameter table

☐ Calculate discriminant scoring coefficients

Stop Process

Terminate processes launched via the Spectrum Mill web server.
Click "List Processes" to list running processes related to Spectrum Mill.
Select one or more processes to stop, using the Ctrl key for multiple selection.

Stop Process

List Processes

Figure 76 Stop process

NOTE

It is also possible to terminate parent processes from the Tool Belt page. This may be more convenient if you have navigated to another Spectrum Mill page and you no longer see the red **Stop** process link. Use the same instructions as for child processes.

You can also stop processes in the Request Queue if they are still queued, or if they are running with **Maximize CPUs** enabled.

CAUTION

Avoid terminating processes initiated from other client PCs. You may be able to identify your processes by name. For example, the data extractor for Agilent Q-TOF files is called **xtractorAgilent.cgi**, while the data extractor for Agilent ion trap files is called **xtractorAgilentTrap.cgi**.

To create a saved results file so you can search previous hits

NOTE

If you mark **Search previous hits** in the MS/MS Search page, it automatically creates a saved results file if it doesn't exist.

The Spectrum Mill workbench allows you to save valid hits (protein identifications) and then search them again. For example, you can save a set of hits from an identity mode MS/MS Search and then search this set with a variable modifications or homology mode MS/MS Search. This saves time because the modified peptides you find in variable modifications or homology modes are often derived from the proteins you already identified in identity mode. The computation-intensive homology mode search is much faster when you search a smaller list.

When you create a saved results file, Tool Belt creates a ***.res** file (list of accession numbers) of valid hits. When you search this list, MS/MS Search reads this ***.res** file and gets entries from the database file. This requires you to both save and use the ***.res** file from the same data directory.

To save hits from MS/MS Search:

- 1 Make sure you already have a set of search results you wish to save. The saved results file will contain only the search results for which you designated the protein identifications as “valid.”
- 2 Navigate to the Tool Belt page and click the option to **Create saved results file**, as shown in [Figure 77](#).
- 3 Select the **Database** you previously searched.
- 4 Select the **Data Directory** where the search results reside. This is the same data directory where the data files reside.
- 5 Click the **Create File** button.

Agilent Spectrum Mill - Tool Belt Utilities

Spectrum Mill

Extractor

MS/MS Search

Protein/Peptide Summary

Autovalidation

Quality Metrics

Select a Tool:

☐ Stop process

☒ Create saved results file

☐ Create MS/MS search summary file

☐ View parameter table

☐ List modifications details

☐ Create Reporter Ion correction factors

☐ Apply Reporter Ion correction factors

☐ Calculate discriminant scoring coefficients

☐ File collector

☐ Export PepXML

☐ Convert spectra

Create Saved Results File

Create a saved results file from validated rank 1 hits (small subset to quickly search previous hits with MS/MS Sea

Create File

Database:

SwissProt.yeast

Data Directory

Select ...

ExampleData\Agilent\Q-TOF\Yeast\Y1

Figure 77 Create saved results file

To create an MS/MS Search summary file if search terminated abnormally

If your MS/MS Search terminated abnormally (e.g., because you aborted it), you can still create an MS/MS Search Summary File so that you can review partial results on the Protein/Peptide Summary page. To create this file:

- 1 Navigate to the Tool Belt page and click the option to **Create MS/MS search summary file**, as shown in [Figure 78](#).

Agilent Spectrum Mill - Tool Belt Utilities

Spectrum Mill	Extractor	MS/MS Search	Protein/Peptide Summary	Autovalidation	Quality Metrics
---------------	-----------	--------------	-------------------------	----------------	-----------------

Select a Tool:

- ☐ Stop process
- ☐ Create saved results file
- ☒ Create MS/MS search summary file
- ☐ View parameter table
- ☐ List modifications details
- ☐ Create Reporter Ion correction factors
- ☐ Apply Reporter Ion correction factors
- ☐ Calculate discriminant scoring coefficients
- ☐ File collector
- ☐ Export PepXML
- ☐ Convert spectra

Create MS/MS Search Summary File

Create MS/MS search summary file of rank 1 hits
(to use with Protein/Peptide Summary if MS/MS Search terminated abnormally)

Create File

Data Directory

Select ... ExampleData\Agilent\Q-TOF\Yeast\Y1

Figure 78 Create MS/MS search summary file

- 2 Select the **Data Directory** for which you want to create a summary file.
- 3 Click the **Create File** button.

To create a summary table of previously-used parameters

The Spectrum Mill workbench keeps a cumulative record of the parameters you set in programs such as Data Extractor, MS/MS Search, and Sherenga *de novo* Sequencing. This option can be helpful to diagnose problems because the settings are the internal Spectrum Mill variables, not those displayed in the Browser.

To view previously-used parameters:

- 1 Navigate to the Tool Belt page and click the **View parameter table** option, as shown in [Figure 79](#).

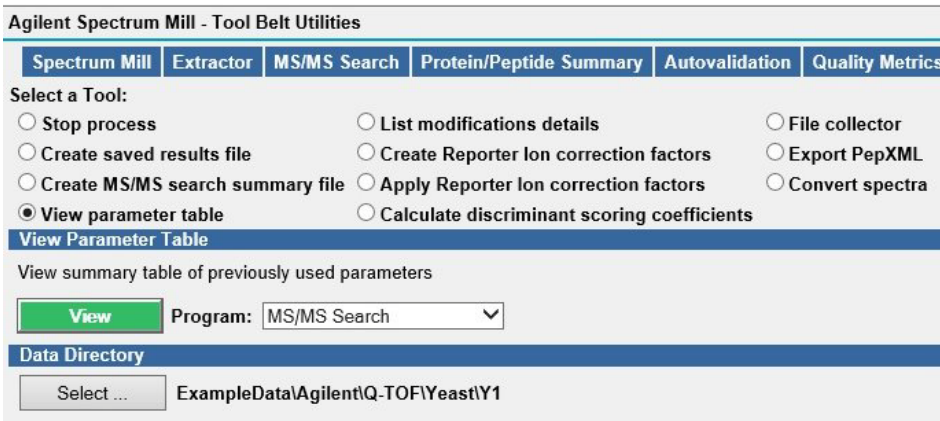


Figure 79 View parameter table

- 2 Select the **Program** for which you want to summarize parameters.
- 3 Select the **Data Directory** for which you want to summarize parameters.
- 4 Click the **View** button.
- 5 Check that you see a report like that shown in [Figure 80](#). In the table, a dash indicates that the setting was the same as used previously.

Spectrum Mill - Parameter Table		
	mstag.1	mstag.2
batchSize	81	-
composition_exclude	-	-
composition_include	-	-
dataDir	ExampleData\Agilent\Trap\ecoli\heatshock	-
database	NCBI\ncr.ecoli	-
dna_frame_translation	6	-
enable_multiple_mods	1	-
enzyme	Trypsin	-
filterByFeature	maxSequenceTagLength	-
fixedMods	carbamidomethylation	-
fragment_mass_tolerance	0.7	-
full_mw_range	1	-
full_pi_range	-	-
hide_html_links	-	-
high_pi	10.0	-
homology_gap_mode	-	-
input_filename	NCBI\ncr.ecoli	-
input_program_name	mstag	-
instrument_name	ESI-HN-TRAP-AGILENT	-
instrument_nameMSMS	-	-
low_pi	3.0	-
maxSequenceTagLengthFilter	3	-
maxSequenceTagLengthFilterDirection	>	-
max_ms_prod_charge	3	-
max_num_unused_ions	50	-
max_reported_hits	5	-
minMatchedPercent	50	-
min_ions_to_search	4	-
missed_cleavages	2	-
mparams_dir	mparams_mill/	-
mutation_matrix_off	-	-
parent_mass_convert	monoisotopic	-
parent_mass_tolerance	2.5	-
parent_shift	130.0	-
parent_shift_type	+/-	-
proton_mobility_scoring	1	-
requestScript	batchTagPara.pl	-
res_mstag_dir	F:\SpectrumMill\msdata\SM\ExampleData\Agilent\Trap\ecoli\heatshock\results_mstag	-
results_from_file	-	1
results_to_file	-	-
search_type	identity	variable

Figure 80 Parameter Table that shows settings used for an identity mode search and a subsequent variable modifications search

To list details about amino acid modifications

- 1 Navigate to the Tool Belt page and click the **List modifications details** option, as shown in [Figure 81](#).
- 2 Select an **smconfig** file:
 - **smconfig.xml** - includes all the modifications that are currently displayed in the Spectrum Mill pages. The file **smconfig.xml** is created automatically by merging **smconfig.std.xml** and **smconfig.custom.xml** (which is a configuration file your server administrator optionally creates). For details, click the Help link for Server Administration, then see the section on “To Add/Change Amino Acid Modifications.”
 - **smconfig.std.xml** - includes all of the modifications that are available when you first install the Spectrum Mill workbench
 - **smconfig.misc.xml** - includes additional modification definitions that Agilent provides “as is”. They are provided as a convenience and can be copied into **smconfig.custom.xml**, but they have not been rigorously tested at Agilent.
- 3 Click the **List** button.

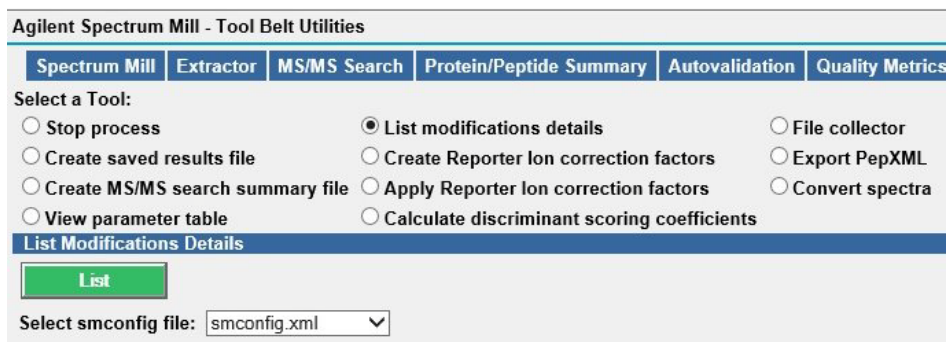


Figure 81 List modifications details

The listed details include:

- **Modification** - the name you see in the Spectrum Mill pages
- **Internal ID** - the name used by the Spectrum Mill programs
- **Type** - the modification type (fixed, variable, or cyclic)
- **Site(s)** - the amino acid or peptide-terminal location(s) for the modification

- **Delta formula** - the difference in chemical composition between the modified and unmodified site
- **Delta mass (Da)** - the monoisotopic mass difference between the modified and unmodified site

To create a file of reporter ion correction factors

Proteomics calculations often use mass tags, such as iTRAQ (isobaric tags for relative and absolute quantitation) or TMT (tandem mass tags). The Tool Belt provides the capability to incorporate the correction factors from a certificate of analysis into a file within the Spectrum Mill workbench. iTRAQ corrections factors apply only to iTRAQ 4-plex quantitation, while you can have TMT correction factors for 2-plex, 6-plex, and 10-plex calculations. iTRAQ 8-plex correction factors do not vary by batch lot, so you do not need to enter them.

Once you have created the files with the correction factors, you can apply them to iTRAQ or TMT calculations for multiple samples. The files can store correction factors from multiple batches of iTRAQ or TMT reagents and are appended every time you type correction factors for a new batch number.

To access the capability to store correction factors:

- 1 Navigate to the Tool Belt page.
- 2 Click **Create Reporter Ion correction factors**.
- 3 Refer to your certificate of analysis to fill in the page. See [Figure 82](#) as an example.
- 4 Click the **Create** button.
- 5 To apply the correction factors from a particular iTRAQ or TMT batch to the calculations for your sample, see [page 165](#).

Agilent Spectrum Mill - Tool Belt Utilities

Spectrum Mill	Extractor	MS/MS Search	Protein/Peptide Summary	Autovalidation	Quality Metrics
---------------	-----------	--------------	-------------------------	----------------	-----------------

Select a Tool:

- ☐ Stop process
- ☐ Create saved results file
- ☐ Create MS/MS search summary file
- ☐ View parameter table
- ☐ List modifications details
- ☒ Create Reporter Ion correction factors
- ☐ Apply Reporter Ion correction factors
- ☐ Calculate discriminant scoring coefficients
- ☐ File collector
- ☐ Export PepXML
- ☐ Convert spectra

Create Reporter Ion correction factors

(iTRAQ 8-plex corrections do not vary by batch lot)

Create iTRAQ 4-plex ▼

Batch: 123456

Mass	% of -2	% of -1	% of +1	% of +2
114	0.0	1.0	5.9	0.2
115	0.0	2.0	5.6	0.1
116	0.0	3.0	4.5	0.1
117	0.0	4.0	3.5	0.1

Figure 82 Create Reporter Ion correction factors, shown with example iTRAQ correction factors

To apply a file of reporter ion correction factors

After you have stored iTRAQ or TMT correction factors in a file (see [page 163](#)), then you can apply them to your data file so that the Spectrum Mill workbench can do all the final iTRAQ or TMT calculations. (You do not need to export results to Excel to finalize the calculations.)

- 1 Navigate to the Tool Belt page.
- 2 Click **Apply Reporter Ion correction factors**.
- 3 For **Batch**, select the batch number for the reagents you used with the samples.
- 4 Select the **Data Directory** that contains your data files.
- 5 Click the **Apply** button.

Agilent Spectrum Mill - Tool Belt Utilities

Spectrum Mill	Extractor	MS/MS Search	Protein/Peptide Summary	Autovalidation	Quality Metric
Select a Tool:					
☐ Stop process	☐ List modifications details	☐ File collector			
☐ Create saved results file	☐ Create Reporter Ion correction factors	☐ Export PepXML			
☐ Create MS/MS search summary file	☒ Apply Reporter Ion correction factors	☐ Convert spectra			
☐ View parameter table	☐ Calculate discriminant scoring coefficients				
Apply Reporter Ion correction factors					
Apply					
Batch: DiLeu4_DiLeu4-B1 ▼					
Data Directory					
Select ...		ExampleData\Agilent\Q-TOF\IOGE			

Figure 83 Apply iTRAQ correction factors

When the software does iTRAQ or TMT calculations in Protein/Peptide Summary, it automatically applies the correction factors. Be sure to check the report for a message about the batch number of the correction factors that the software used.

NOTE

If you copy or move your data files to a new directory, copy **reporterIons.corrections.txt** as well. This file must be present in the data directory so that future Spectrum Mill calculations can use the iTRAQ or TMT correction factors that it specifies.

To calculate discriminant scoring coefficients

To have MS/MS Search calculate discriminant scores, you select among sets of standard Agilent discriminant scoring coefficients, either Q-TOF or ion trap. With this Tool Belt option you can create and update your own set of coefficients, but you must remember to use only validated hits.

To create your own set of coefficients:

- 1 Navigate to the Tool Belt page and click **Calculate discriminant scoring coefficients**.

The screenshot shows the 'Agilent Spectrum Mill - Tool Belt Utilities' window. It has a tabbed interface with tabs for 'Spectrum Mill', 'Extractor', 'MS/MS Search', 'Protein/Peptide Summary', 'Autovallation', and 'Quality Metrics'. The 'MS/MS Search' tab is active. Under 'Select a Tool:', there are several radio button options: 'Stop process', 'Create saved results file', 'Create MS/MS search summary file', 'View parameter table', 'List modifications details', 'Create Reporter Ion correction factors', 'Apply Reporter Ion correction factors', 'Calculate discriminant scoring coefficients' (which is selected), 'File collector', 'Export PepXML', and 'Convert spectra'. Below this is a section titled 'Calculate Discriminant Scoring Coefficients' with a green 'Create' button. Underneath is a 'Name:' text field, followed by 'Existing coefficient sets:' with a dropdown menu set to 'Off' and an 'Update' checkbox. At the bottom is a 'Data Directory' section with a 'Select ...' button and the text 'ExampleData\Agilent\Q-TOF\Yeast\Y1'.

Figure 84 Calculate discriminant scoring coefficients fields

- 2 In the **Name** field, type a name for the new set of coefficients.
The name appears in the **Name** field.
- 3 Under **Data Directory**, click **Select** to choose the directory containing valid data results.
- 4 Click **Create**.

At the bottom of the page you see the calculations as they are done. After they have completed. The name of the new set of coefficients appears in the Discriminant Scoring selection list on the MS/MS Search page.

NOTE

The calculations can take some time. Wait for the calculations to complete before you leave or refresh the page. The **Create** button will be re-enabled when the calculations are done, and the new set will be listed in the **Existing coefficients sets**.

To update your coefficients:

- 1** Navigate to the **Tool Belt** page and click **Calculate discriminant scoring coefficients**.
- 2** In the **Existing Coefficient Sets** list, select the coefficients you intend to update.
- 3** Mark the **Update** check box.
- 4** Under **Data Directory**, click **Select** to choose the directory containing valid data results.
- 5** Click **Create**.

To copy files to your collections directory

Use this tool to copy files from multiple directories to a single collection directory. It is typically used to consolidate exported summary files (.ssv) to one folder.

- 1 Navigate to the Tool Belt page and click the **File collector** option, as shown in Figure 85.

The screenshot shows the 'Agilent Spectrum Mill - Tool Belt Utilities' window. At the top is a navigation bar with tabs: 'Spectrum Mill', 'Extractor', 'MS/MS Search', 'Protein/Peptide Summary', 'Autovalidation', and 'Quality Metrics'. Below the tabs, under 'Select a Tool:', there are several radio button options. The 'File collector' option is selected. Below this, the 'File Collector' section is active, showing a 'Copy Files' button and a description: 'Copy files (* wildcard filenames allowed) from multiple directories to a single collection'. Under 'Data Directories', there is a 'Select ...' button and a list of directories. Two directories are selected: 'ExampleData\Agilent\Q-TOF\Yeast\Y1' and 'ExampleData\Agilent\Q-TOF\Yeast\Y2'. Below this, there is another 'Select ...' button and a 'To directory:' field containing 'ExampleData\Agilent\Q-TOF\Yeast\Y1Y2'. At the bottom, under 'Files:', there is a text box containing 'PSMExport*.ssv'.

Figure 85 File collector

- 2 Select the **From directories**.
- 3 Select the **To directory**. Note that you cannot select a directory unless it contains a data file.
- 4 Under **Files**, modify the list to specify the files to copy in the **From directories**. Wildcards (*) are supported.
- 5 Click the **Copy Files** button.

When the files are copied to the **To directory**, they are renamed with a prefix that includes the name of the **From directory**.

To export PepXML files

PepXML format is a results exchange format that is supported by the Trans-Proteomic Pipeline from the Institute for Systems Biology. This format lets you import Spectrum Mill results into other software that supports the pepXML format, such as Skyline, an MRM data analysis package from the University of Washington.

To convert search results to this format:

- 1 Navigate to the Tool Belt page and click **Export PepXML**.

The screenshot shows the 'Agilent Spectrum Mill - Tool Belt Utilities' window. It has a tabbed interface with tabs for 'Spectrum Mill', 'Extractor', 'MS/MS Search', 'Protein/Peptide Summary', 'Autovalidation', and 'Quality Metrics'. The 'MS/MS Search' tab is active. Under 'Select a Tool:', there are nine radio button options: 'Stop process', 'List modifications details', 'File collector', 'Create saved results file', 'Create Reporter Ion correction factors', 'Export PepXML' (which is selected), 'Create MS/MS search summary file', 'Apply Reporter Ion correction factors', and 'Convert spectra'. Below this is a section titled 'Export Spectrum Mill MS/MS Search Results to PepXML' containing a green 'Export' button and the text 'Convert MS/MS search summary file to multiple PepXML files (runName.pepXML)'. The 'Validation filter:' is set to 'all'. 'Required AAs:' is set to 'any' and 'Integrate rank 1 reversed hits as pseudo <search_hit> entries' is checked. 'Disallowed AAs:' is set to 'none' and 'Omit protein names' is checked. There is a 'Data Directories' section with a 'Select ...' button and the text 'Data directories to export:'. Below this, a list shows 'ExampleData\Agilent\Q-TOF\OGE' with a checked checkbox. At the bottom, another 'Select ...' button is next to the text 'Export directory: ExampleData\Agilent\Q-TOF\OGE'.

Figure 86 Export PepXML fields

- 2 For **Validation filter**, select **Valid** or **All**.
- 3 If you want to include the top-ranked reversed hits, mark the check box for **Integrate rank 1 reversed hits as pseudo <search hit> entries**.
- 4 If you want to exclude the protein names, mark the check box for **Omit protein names**.

- 5 Under **Data Directories** click **Select** to choose the data directory that contains the summary files you want to export. You may select one or more data directories.
- 6 Click **Select** to choose the Export directory to which you intend to export the PepXML-formatted data. If you do not select an Export directory, each data directory will contain its own PepXML files.

- 7 Click **Export**.

Note that the program converts the search summary file(s) to PepXML format, and puts the output file(s) in the selected Export directory for which they apply. The output files are of the form DataFileName.pep.xml.

- 8 To access these files, do one of the following:
 - Click the link(s) from the output frame, under **Spectrum Mill - Convert to PepXML**.
 - Navigate to the data folder and click to open the pep.xml file.

To convert spectra

You use this tool to convert mzXML files to pkl files and pkl files to mzXML files.

As of Spectrum Mill B.04.01, the data extractors create mzXML files instead of pkl files. You can use this tool to convert the mzXML files to the pkl files required by other software packages.

If you have files that were extracted to pkl using prior versions of the Spectrum Mill workbench, you can convert pkl files to mzXML files with this tool.

- 1 Navigate to the Tool Belt page and click **Convert Spectra**.

Agilent Spectrum Mill - Tool Belt Utilities

Spectrum Mill	Extractor	MS/MS Search	Protein/Peptide Summary	Autovalidation	Quality Metrics
---------------	-----------	--------------	-------------------------	----------------	-----------------

Select a Tool:

☐ Stop process

☐ List modifications details

☐ File collector

☐ Create saved results file

☐ Create Reporter Ion correction factors

☐ Export PepXML

☐ Create MS/MS search summary file

☐ Apply Reporter Ion correction factors

☒ Convert spectra

☐ View parameter table

☐ Calculate discriminant scoring coefficients

Convert Spectra

Convert

Inter-convert MS/MS spectral files ./cpick_in/runName.*.pkl <- -> ./runName.mzXML

Convert pkl to mzXML Validation filter: all

Data Directories

Select ...

Data directories to convert:

☒ ExampleData\Agilent\Q-TOF\OGE

Select ...

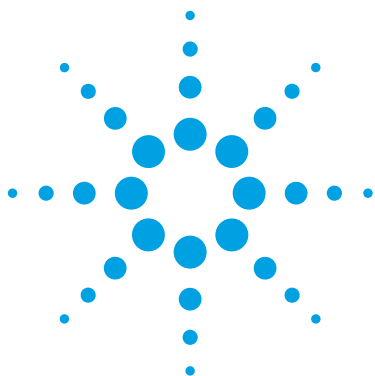
Export directory: ExampleData\Agilent\Q-TOF\OGE

Figure 87 Convert spectra option

- 2 Choose either **Convert pkl to mzXML** or **Convert mzXML to pkl**.
- 3 For the former conversion, select the **Validation filter** (either **all** or **valid**).
- 4 Under **Data Directories**, click the top **Select** button to choose the directory that contains the spectra you intend to convert.
- 5 If you want to convert pkl files to mzXML files, click the bottom **Select** button to choose the **Export directory**. (If you want to convert mzXML files

to pkl files, you do not need to specify a directory. The pkl files go automatically to **cpick_in** under the corresponding data directory.)

6 Click **Convert**.



8 Using Spectrum Mill Utilities

- To archive data 174
- To identify digest peptides likely to meet specific experimental goals (Peptide Selector) 176
- To build MRM lists for triple quadrupole instruments 179
- To align sequences (Multiple Sequence Aligner) 181
- To list peptides that correspond to a theoretical protein digest (MS Digest) 182
- To retrieve database entries using text searches (MS Edman) 184
- To list theoretical fragment ion masses for peptides (MS Product) 186
- To list amino acid compositions that fit precursor mass and partial composition (MS Comp) 188
- To show isotope patterns of peptides (MS Isotope) 190
- To search a database with a list of peptide sequences (Peptide String Match) 191
- To calculate the masses and formulas for peptides 193

This chapter describes Spectrum Mill workbench peptide and protein utilities. These utilities help you accomplish a range of tasks, from retrieving database entries using text searches to simulating protein digestions. You access these utilities from the Spectrum Mill home page.



To archive data

This option lets you archive instrument-created files, search result files, spectral files and data directories, and lets you unzip search result files and spectral files. You archive data to create more available disk space. Once results are archived (zipped), you cannot summarize or view them until you click the **Unzip** option on the Spectrum Mill server.

Starting with Spectrum Mill workbench B.06.00, searches create the **spo.zip** file automatically.

NOTE

You cannot unzip an entire data directory that has been zipped using the Archive tool. You must unzip it on the Spectrum Mill server.

You can include Archive as a step in a workflow.

- 1 In the Utilities section of the Spectrum Mill home page, click **Archive Data**.
- 2 Under **Data Directories**, click **Select** and choose the directory whose files and/or folder you intend to zip, move or unzip.

You can select only one at a time.

- 3 Select the general categories of files or directory you intend to archive and how you want to archive them. (Figure 88)

Below the options in each category, please read the description of what happens when you select an option.

- 4 Click **Archive**.

Archive Data	
Archive ☐ Queue request Save As... Load...	
Archive Parameters	
Instrument created files ☒ Ignore instrument data files ☐ Delete data files after making placeholder Creates a placeholder file for each raw data file, then deletes the raw data files. Afterwards, the previously extracted data may be searched and summarized, but the raw data files cannot be re-extracted.	Spectrum Mill search result files ☒ Ignore ☐ Zip results_mstag/*.spo to spo.zip ☐ Unzip spo.zip to results_mstag/*.spo After zipping to create spo.zip deletes results_mstag/*.spo. After unzipping to create results_mstag/*.spo deletes spo.zip.
Spectrum Mill spectral files ☒ Ignore ☐ Zip cpick_in/*.pkl to pkl.zip ☐ Unzip pkl.zip to cpick_in/*.pkl Zippping prevents further searching and spectral visualization. After zipping to create pkl.zip deletes cpick_in/*.pkl. After unzipping to create cpick_in/*.pkl deletes pkl.zip.	Spectrum Mill data directories ☒ Ignore ☐ Zip dataDir/*. * to dataDir.zip Zippping a data folder prevents further use in SM. After zipping to create dataDir.zip deletes dataDir/*. *.

Figure 88 Archive Data selections

To identify digest peptides likely to meet specific experimental goals (Peptide Selector)

Peptide Selector runs theoretical digestions and then automatically selects from the theoretical peptides those that fit specific criteria. The most common uses are:

- List only those peptides suited to be synthesized with a stable isotope label and used for quantitation via a multiple reaction monitoring (MRM) experiment.
- List only those peptides suited for incorporation into an accurate mass inclusion list to be used in a data-dependent MS/MS experiment.
- List only those peptides expected to give doubly-charged ESI spectra.
- List only the likely detectable forms of all peptides that contain a possible phosphorylation site.
- List only those peptides that contain cysteines.
- List peptides according to their predicted pI range for inclusion in target lists specific to a fraction from isoelectric focusing (IEF). This option can reduce the number of fractions that you must analyze to get the target proteins.

In many ways, Peptide Selector is similar to MS Digest, described on [page 182](#). Both Peptide Selector and MS Digest run automatic protein digestions. The difference is that Peptide Selector additionally creates limited lists based on specific criteria.

- 1 On the Spectrum Mill home page, click the **Peptide Selector** link in the **Utilities** section.
- 2 Check that you see the page shown in [Figure 89](#).
- 3 To use an existing parameter file, click **Load**.
- 4 Set **Digest Parameters** and **Product Ion Parameters**. See the explanations in the online help.
- 5 Set **Saved File Parameters**. To see all the options for this section as shown in [Figure 89](#) on page 178, mark the **Generate inclusion or MRM list file** check box. See the explanations in the online help.
- 6 Set **Criteria for Excluding Peptides**. These criteria allow you to set the rules to discard peptides that you do not want to include in the list. The excluded peptides do not appear in the final report.

- 7 Click the **Choose...** button to select the modifications that match the chemistry for your samples.
- 8 Under **Protein(s) to Select From**, set **Database** options as follows:
 - To use a sequence from a database, select the appropriate database name.
 - To supply your own sequence, select **User Protein** and type or paste a sequence into the **User-supplied sequence** box. If you have multiple sequences, you can use a variety of delimiters, as found in HTML- and Excel-formatted Spectrum Mill reports.
- 9 Set **Search Mode** parameters as described in the online help.
- 10 Set the **Scoring** parameters.

If you do not get enough peptides to pass the filter, you can apply a penalty rather than excluding the peptides. When you mark the **Penalize rather than exclude** check box, the settings for which you can apply penalties change to green font. (Figure 89) These settings no longer cause peptides to be excluded. See the *Online Help* for details.

- 11 To save the changed parameters, click **Save As**.
- 12 Click **Select**.

See the *Online Help* to learn how to transfer the inclusion list to the Agilent MassHunter Data Acquisition software.

Agilent Spectrum Mill - Peptide Selector - (none)

Spectrum Mill | MS Edman | Multiple Sequence Aligner | Databases | MS Digest | Help

Selection

Select Save As... Load ... ☐ Hide HTML links (better Excel cut/paste)

Saved File Parameters

☒ Generate inclusion or MRM list file MassHunter QQQ MRM list
☐ IEF

Filename: lastPeptideSelectorResult.txt Run time(min): 89.9 List Type: m/z
 Precursor Charge Min.: 2 Max.: 4 Precursor m/z Min.: 300.0 Max.: 2000.0

Digest Parameters **Product Ion Parameters**

Digest: Trypsin Maximum # missed cleavages: 0 ☒ Show Product Ion Masses

Criteria for Excluding Peptides

Max. # basic residues (RHK): 2
 Peptide MH⁺ Min.: 900.0 Max.: 2800.0
AA Composition Filtering:
 AAs Required : KR Disallowed: CM

Peptide exclusion criteria:

- ☒ Has nearby cleavage site within 3 residues
- ☒ Contains peptide N-terminal Gln to pyroGlu
- ☐ Contains protein N-terminus Acetylatable
- ☐ Contains consensus N-linked glycosylation site
- ☐ Contains no variable modification

Protein Position Filtering:
 Each peptide must contain at least 1 of the following positions in the protein (enter 1 position/line)

Modifications

Choose... Fixed: Variable:

Protein(s) to Select From

Database: SwissProt
 Select Peptides from only the Database Entries with the accession numbers below
 Allowed delimiters (!:,)
 P07288
 P02741
 P00533
 P06702
 P16860
 P03372

Search Mode

Count Peptide Uniqueness in Database by: Sequence
 Species: All

Scoring

☐ Penalize rather than exclude

Figure 89 Peptide Selector

To build MRM lists for triple quadrupole instruments

The MRM Selector uses Spectrum Mill MS/MS Search results to create multiple reaction monitoring (MRM) lists for triple quadrupole instruments. You select data that has been searched, then filter the results to include only those peptides that meet certain requirements. The filters are similar to those in Protein/Peptide Summary.

- 1 On the Spectrum Mill home page, click the **MRM Selector** link in the **Utilities** section.
- 2 Check that you see the page shown in [Figure 90](#).
- 3 Set the **Format** and choose the **Data Directories**.
- 4 Set the **Validation and Sorting** parameters. They are very similar to the ones in Protein/Peptide Summary.
- 5 Set the **MRM Parameters**, which determine the contents of the MRM list. These specific settings depend on the **Format** you selected.
 - The Agilent Q-TOF Data Extractor outputs the collision energy (CE), peak apex, and chromatographic peak width in the **specFeatures.tsv** file in order to use these values in MRM and dynamic MRM (DMRM) generation.
 - You may instead choose to let the program calculate the collision energies based on an equation.
 - You have the option to type a chromatographic peak width that will apply for all peaks in the DMRM analysis. This peak width, which corresponds to the **delta RT** setting in the MassHunter Data Acquisition software, is the retention time window for which the MRM transitions are monitored. For example, if you have a peak apex at 2.5 min and a delta RT of 1.0 min., the MassHunter Data Acquisition software monitors the MRM transitions for that peak from 2.0 min until 3.0 min.
- 6 Click **Select MRMs**.
- 7 See the online help to learn how to transfer the MRM list into MassHunter Data Acquisition software.

Select Results for MRM Selection	Validation and Sorting	MRM Parameters									
Select MRMs Save As... Load... Format: MassHunter QQQ DMRM Filter to distinct peptides: Off Data directories: Select ... ☒ ExampleData/Agilent/Q-TOF/OGF Search result files: *.spo	Filter results by: sequence-not-validated Protein grouping method: 1 shared, expand subgroups Sort proteins by: Score Filter by protein score: > 13.0 Top n peptides for MRM: Limit to 5 Rank peptides by: Score Sort MRM List by: Precursor m/z Filter peptides by: Score: > 10 % SPI: > 70.0 Required AAs: any Disallowed AAs: none Peptide pI: from 3.0 to 10.0 ☐ All Accession #'s:	☐ Text file export Top n Transitions: 5 ≥ 100 % of precursor ☐ y-ions only Trigger threshold: 500 for 0 secondary transitions Z options: Observed precursor/fragment charge only ☐ Use peak width of: 1 min Collision Energy: (CE= slope*m/z + intercept) ☒ Use actual CE if available z-state: slope: intercept: <table border="1"><tbody><tr><td>2</td><td>3.6</td><td>-4.8</td></tr><tr><td>3</td><td>3.6</td><td>-4.8</td></tr><tr><td>4</td><td>3.6</td><td>-4.8</td></tr></tbody></table>	2	3.6	-4.8	3	3.6	-4.8	4	3.6	-4.8
2	3.6	-4.8									
3	3.6	-4.8									
4	3.6	-4.8									

Figure 90 MRM Selector

To align sequences (Multiple Sequence Aligner)

The Multiple Sequence Aligner enables alignment and comparison of the amino acid sequences of proteins that are present in a database. This stand-alone utility within the Spectrum Mill software highlights the amino acids that differ among the sequences. (When you access multiple sequence alignment from Protein/Peptide Summary, the software highlights the identified peptides within the sequence.)

- 1 On the Spectrum Mill home page, click the **Multiple Sequence Aligner** link in the **Utilities** section.
- 2 Check that you see the page shown in [Figure 91](#).
- 3 Select the name of the database that contains the proteins you wish to align.
- 4 Type the accession numbers of the proteins that you wish to compare.
- 5 Click the **Align** button.

Figure 91 Multiple Sequence Aligner

NOTE

If the database is too large, the alignment does not work properly. In that case, create a subset database before you do the alignment.

To list peptides that correspond to a theoretical protein digest (MS Digest)

MS Digest runs theoretical digestions and calculates masses of peptides that result. The program accepts both user proteins and sequences from databases.

- 1 On the Spectrum Mill home page, click the **MS Digest** link in the **Utilities** section.
- 2 Check that you see the page shown in [Figure 92](#).
- 3 Click the **Choose...** button to select the modifications that match the chemistry for your samples. For details, click the blue bar labeled **Modifications** to access the online help.
- 4 Fill in the other options. See the explanations below and in the online help. In general, you should keep the default values, except for settings that are highlighted in red.
- 5 If you want to include a **User-specified amino acid**, indicate it with a lower-case **u** in the **Protein sequence** box, and then specify an elemental composition.
- 6 Click **Digest**.

Agilent Spectrum Mill - MS Digest

Spectrum Mill | MS Product | Peptide Selector | Tool Belt | Help

Digest

☐ Report multiple charges
☐ Hide protein sequence
☐ Show only protein sequence
☒ Hide HTML links

Protein

Digest:

Maximum # missed cleavages: Calculate masses as:

Modifications

Fixed: Variable:

Database

Database: User-specified amino acid (u): (Default is Gly)

C: H: N: O: S: P:

Protein sequence:

Paste or type sequence in this field. Delete these instructions first.
 Tabs, returns, and spaces are ignored.
 USE CAPITAL LETTERS for unmodified amino acids or those with fixed modifications.
 Use lower case letters to indicate VARIABLE modifications. Only those Variable modifications that have been selected will be considered.

To enable the user-specified amino acid, use the lower case letter u:
 u - user specified amino acid (provide elemental composition above)

Do NOT use the letters B, J, O, U, X, or Z.
 Do NOT use 3-letter code amino acid symbols.

Figure 92 MS Digest

Database To select a sequence from a database, select the appropriate database name. To supply your own sequence, select **User Protein** and supply a sequence in the **Protein sequence** box.

Retrieve database entry by: When you select a database name, you then have two choices:

- Select **Accession Number** and supply the correct **Database accession number** in the box to the right.
- Select **Index Number** and supply the correct **MS Digest index number** in the box to the right.

The **Accession Number** and **Index Number** entries apply only if you have selected a database name as the **Database** option. If you have selected **User Protein**, these entries are ignored.

To retrieve database entries using text searches (MS Edman)

MS Edman allows you to search text fields (such as sequence, name, accession number or species) in protein databases.

MS Edman can help identify a protein if you know only the molecular weight of a tryptic fragment and some of the sequence. For example, if you know a peptide mass and some sequence information from *de novo* sequencing, then MS Edman is a useful search tool.

MS Edman is also the first step when you want to create a specialized subset database of entries that match your text search criteria. For example, you could use MS Edman to find all proteins that contain a certain amino acid sequence. If you marked the **Save hits to file** check box on the MS Edman page, you could then use Protein Databases to create a subset database from these saved results. See [“To create a subset database from saved hits”](#) on page 204.

- 1 On the Spectrum Mill home page, click the **MS Edman** link in the **Utilities** section.
- 2 Check that you see the page shown in [Figure 93](#).
- 3 Set options for **Search**, **Search Parameters**, and **Modifications**. See explanations in the online help.
- 4 Select a **Search Mode**. Your choice determines which additional options appear. There are three types of search modes:
 - **Sequence Only**: Search for an amino acid sequence. Type the sequence as a **Regular expression**. (See the next step.)
 - **Sequence and Mass**: This is the same as the **Sequence Only** mode, with the addition that search results are filtered by peptide mass. Type a sequence and mass(es).
 - **Name, Accession Number or Species**: Type one of these in the box below.
- 5 To type the sequence in the box under **Regular expression**, use regular expressions that the UNIX® **grep** command understands. Think of regular expressions as very intelligent wild cards. Use the following guidelines:
 - WVTF (no brackets) means the sequence is exactly WVTF. Use this type of designation for parts of the sequence that are known with certainty.
 - [AK] means the amino acid is *either* A or K. The program searches for either one.

- [^EF] means the amino acid is *neither* E nor F. The program searches for all other possibilities.
- . (period) means a single amino acid is unknown. The program searches for all possibilities.
- .* (dot star) means a section of the sequence is unknown. The length of that section is not specified. For example, FMQ.*K means that the sequence begins with FMQ and ends with K, but there is a section of unknown sequence of unspecified length in the middle.

In the example shown in [Figure 93](#), the sequence is: either A or K, followed by either A or K, followed by PV, followed by either I or L, followed by ED, followed by either I or L, followed by R.

For more information on the grep command, go to a UNIX® system and type **man grep**, or search the internet for **grep**.

6 Click the **Start Search** button.

Agilent Spectrum Mill - MS Edman

Spectrum Mill | Manual PMF | MS Digest | Databases | *de novo* | Help

Search

Start Search Maximum reported hits: 200

☐ Display combinatorial peptide output

Sample ID: Enter_Comment

Search Parameters

Database: NCBInr.bovine

Species: All

Digest: No enzyme

☐ Search hits from file: lastres

☐ Save hits to file: lastres

Maximum # of mismatched AA's: 2

MW of protein: from 1000 Da to 100000 Da ☒ All

Modifications

Choose... Fixed: Carbamidomethylation (C)

Search Mode

Search mode: Sequence Only

Regular expression (Use CAPITALS):

[AK][AK]PV[IL]ED[IL]R

Figure 93 MS Edman

To list theoretical fragment ion masses for peptides (MS Product)

MS Product calculates theoretical ion masses from peptides that undergo dissociation via post-source decay or high- or low-energy collision-induced dissociation.

- 1 On the Spectrum Mill home page, click the **MS Product** link in the **Utilities** section.
- 2 Check that you see the page shown in [Figure 94](#).
- 3 Fill in the options. See the online help for explanations. Note that under **Product Ion Types**, you need to select the ion types that are appropriate for your instrument.
- 4 If you want to include a user-specified amino acid, indicate it with a lower-case **u** in the **Sequence** box, and then specify an elemental composition at the bottom of the page.
- 5 Click the **Fragment** button.

Agilent Spectrum Mill - MS Product

Spectrum Mill **MS Digest** **Help**

Fragmentation

Fragment Calculate masses as: **Monoisotopic**

Maximum reported charge: **1**

Peptide Sequence

Use **CAPITAL** letters for unmodified amino acids or those with Fixed Modifications.
 Use **lower case** letters for amino acids to indicate a Variable Modification (eg. KCmqYQG).
 Use 'u' for the user-specified amino acid (see below).
 Use 'o' for Homoserine Lactone.

Enter sequence: **SAMPLER**

Modifications

Choose... Fixed: Carbamidomethylation (C) Variable:

Product Ion Types
 (some types require presence of specific AA's in ion)

AA composition ions	N-terminus sequence ions	C-terminus sequence ions	Internal product ions	Ladder sequencing ions
☒ i ☐ m	☒ a ☒ b ☐ c ☐ c'	☐ x ☒ y ☐ Y ☐ z ☐ z'	☒ internal	☐ N-terminus ☐ C-terminus

Satellite sequence ions (side-chain loss)	Neutral-loss sequence ions	Peeling sequence ions
☐ d ☐ v ☐ w	☒ -H ₂ O S, T ☒ -NH ₂ R, K, Q ☐ -H ₃ PO ₄ s, t (S, T + PO ₄) ☐ -SOCH ₃ m (M + Ox)	☒ b+H ₂ O R, H, K

User-specified AA elemental composition (u):

C: **2** H: **3** N: **1** O: **1** S: **0** P: **0** (Default is Gly)

Figure 94 MS Product

To list amino acid compositions that fit precursor mass and partial composition (MS Comp)

MS Comp fills in possible amino acid compositions for a peptide, given a peptide mass and partial composition determined from immonium ions present in MS/MS spectra.

- 1 On the Spectrum Mill home page, click the **MS Comp** link in the **Utilities** section.
- 2 Check that you see the page shown in [Figure 95](#).
- 3 Select the **Combination type**. You have three choices:
 - Select **Amino Acid** to get a list of amino acids.
 - Select **Peptide Elemental** to get a list of elemental compositions corresponding to these amino acids.
 - Select **Elemental** to get a list of elemental compositions without regard to whether or not they could be peptides. This option speeds searches at higher masses.
- 4 In the **Peptide** section, type an **m/z** value and select possible **Ion types** that correspond to this *m/z* value.
- 5 Fill in the other options. See the explanations in the online help.
- 6 Click the **Composition** button.

Agilent Spectrum Mill - MS Comp

Spectrum Mill | MS Edman | Spectrum Summary | *de novo* | Help

Compositions

Composition Maximum reported compositions: 1000 Combination type: Amino Acid

Note: If the mass tolerance is large and few ions below are checked then the program will take a long time to run.

Peptide

m/z: 1260.5992
 Da +/-: 10 ppm
 Charge (z): 1
 m/z is: Monoisotopic

Ion types:

MH+	y-NH3	y-H3PO4
y	b-NH3	b-H3PO4
b	a-NH3	a-H3PO4
a	y-H2O	y-SOCH4
c	b-H2O	b-SOCH4
b+H2O	a-H2O	

AA Composition

Based on immonium and related ions:

☐ 60 S	☐ 70 R,P	☒ 72 V	☐ 74 T	☐ 86 L,I	☐ 87 N,R	☐ 88 D	☐ 84,101,129 K,Q	☒ 102 E
☐ 104 M	☒ 110 H	☐ 120 F	☐ 126 P	☒ 136 Y	☒ 117,130,159,170 W	☒ 70,73,87,100,112,185 R		

Based on loss from precursor ion:

☐ -SRH,-R C (select modified Cys above)	☐ -80,-98 PO ₄ - S,T (check s and t below)	☐ -64 SOCH ₄ - M (check m below)	☐ -17 R
---	---	---	-----------------------------------

Note: For ions with multiple residues (i.e. 86: means I or L).
 Note: Requires presence of at least one AA for each composition ion selected.

AminoAcids

Absent amino acids (Check to prevent possible inclusion)

☐ A	☐ C	☐ D	☐ E	☐ F	☐ G	☐ H	☐ I	☐ K	☐ L	☐ M	☐ N	☐ P	☐ Q	☐ R	☐ S	☐ T	☐ V	☐ W	☐ Y
----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------

Modifications

☐ User defined amino acid: C: 2 H: 3 N: 1 O: 1 S: 0 P: 0 (Default is Gly)

Choose... Fixed: Carbamidomethylation (C) Variable:

Figure 95 MS Comp

To show isotope patterns of peptides (MS Isotope)

MS Isotope calculates and displays isotope patterns of peptides.

- 1 On the Spectrum Mill home page, click the **MS Isotope** link in the **Utilities** section.
- 2 Check that you see the page shown in [Figure 96](#).
- 3 Select **Peptide sequence** to display the isotope pattern for a peptide sequence, or select **Elemental composition** to display the isotope pattern for an elemental composition. The other options change accordingly.
- 4 Fill in the options. For explanations, see the online help.
- 5 Click **Calculate**.

Agilent Spectrum Mill - MS Isotope

[Spectrum Mill](#) [Tool Belt](#) [Help](#)

Isotope Distribution

Calculate Calculate masses as: Monoisotopic ☐ Show detailed report

☒ Peptide sequence
☐ Elemental composition

Peptide Sequence

Use **CAPITAL** letters for unmodified amino acids.
 Use lower case letters for variable modified amino acids to indicate modification (eg. KCmqYQ).
 Use 'u' for the user-specified amino acid.

Sequence

User-specified AA elemental composition (u):

C: H: N: O: S: P: (Default is Gly)

Modifications

Fixed: **Variable:**

Figure 96 MS Isotope

To search a database with a list of peptide sequences (Peptide String Match)

Peptide String Match finds peptides that contain a specific sequence of amino acids.

- 1 On the Spectrum Mill home page, click the **Peptide String Match** link in the **Utilities** section.
- 2 Under **Database**, select the name of the database for which you want to find peptides that match a given peptide sequence
- 3 For **Subset of Accession #'s to search**, type or paste the accession numbers. Leave the box empty if you want to find all proteins in the database that match your criteria.
- 4 Mark check boxes for the options you wish to use:
 - **All gene symbols** - Supported only for UniProt, SwissProt and IPI databases
 - **All Matched Peptides** - Select to show the peptide sequences that were matched.
 - **Species**
 - **All accession numbers** - Select to show the protein accession numbers for the peptides that were matched.
 - **All Start AA** - Select to show the locations of the peptide sequence in the proteins.
 - **Link to ClustalW Multi-Aligner** - Provides a link to align the sequences of the matched proteins with the input peptide highlighted.
 - **Link to Database Website**
 - **Show flanking sequence around variable modification sites** - Select this option to indicate the variable modification site motif within the number of residues indicated, and for the selected modifications.
- 5 For **Enter peptide sequences**, type or paste the sequences you wish to match
- 6 Click **Find Peptides**.

8 Using Spectrum Mill Utilities

Agilent Spectrum Mill - Peptide String Match

Spectrum Mill	MS Edman	Peptide Selector	Peptide List to Masses	MS Digest	MS Product	Help
---------------	----------	------------------	------------------------	-----------	------------	------

Find Peptides Using case-insensitive (CI) string comparison

Database: SwissProt.ecoli

Filters	Output Features
☒ Require Preceding Tryptic Site Subset of Accession #'s to search: 	☐ All gene symbols ☐ All Matched Peptides ☐ Species ☐ All accession numbers ☐ All Start AA
	☐ Link to ClustalW Multi-Aligner ☐ Link to Database Website ☐ Show flanking sequence around variable modification sites 5 residues +/- s t y

Enter peptide sequences: (regular expressions allowed)

FIVIQLVGMLR
VIGVGEGEVCSGR
TYIDPHtyEDPNQAVHEFAK
TYVDPFTYEDPNQAVR
NILVNSNLVCK
TYIDPHTYEDPNQAVHEFAK
TYVDPFTYEDPNQAVR
LIEDNEyTAR
SVLEDFFTATEGQyQPQP
NLLSVAYK
EMQPTHTPIR
EDLVYQAK

Figure 97 Peptide String Match utility

To calculate the masses and formulas for peptides

With the Peptide List to Masses utility, you can have the program calculate the masses and formulas for a set of peptides you specify.

- 1 On the Spectrum Mill Home page, under Utilities, click **Peptide List to Masses**.

Agilent Spectrum Mill - Peptide List to Masses

Spectrum Mill MS Digest Peptide Selector MS Product Tool Belt Help

Calculate Masses

Calculate masses as: **Monoisotopic**

Calculate

Reported precursor charge Min: **2** Max: **3**

Limit actual charge by: ☒ RKH present ☐ RKHQN present ☐ Above min/max

Modifications

Choose... Fixed: Carbamidomethylation (C) Variable:

Peptide Sequences

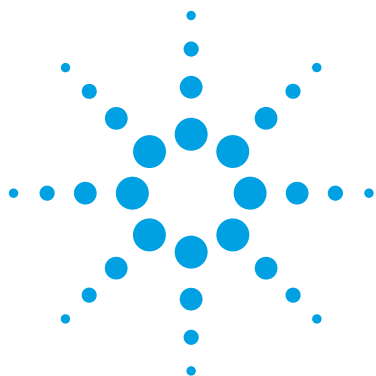
Use **CAPITAL** letters for unmodified amino acids or those with Fixed Modifications.
Use **lower case** letters for amino acids to indicate a Variable Modification (eg. KCmqYQ).

Enter peptide sequences:

SAMPLER
SAMcFLER
SAMsPLEK
SAMyPLEK
SQAMyPLEK
SNAMyPLEK

Figure 98 Peptide List to Masses utility

- 2 From the **Calculate masses as:** list, select how to calculate masses.
- 3 Type a minimum and maximum charge for the **Reported precursor charge**.
- 4 Next to **Limit actual charge by:**, select to limit the charge by the arginine, histidine and lysine present (**RKH present**), by those three amino acids plus glutamine and asparagine (**RKHQN present**), or by the Min and Max specified in the previous line (**Above min/max**).
- 5 Click **Choose** to select the modifications.
- 6 Type the **Peptide Sequences** whose masses and formulas you want to know.
- 7 Click **Calculate**.



9 System Administration

Manipulating sequence databases	196
Other system administration tasks	211

This chapter describes system administration tasks to install, maintain, and customize sequence databases, and to customize various aspects of the Spectrum Mill workbench. This chapter covers the highlights; you can find additional information in the online help. See **millhtml\SM_instruct\faman.htm** and **millhtml\SM_instruct\servadmn.htm**. You can link to these from any Spectrum Mill help page. The **Protein Databases** link takes you to **faman.htm**, while the **Server Administration** link takes you to **servadmn.htm**.



Manipulating sequence databases

To install or update databases

A number of public protein and DNA sequence databases are available via the Internet. Here is how you install them:

- 1 Establish an internet connection to a site that has a database available for download.
 - The easiest way to connect is to click the appropriate link in the “Updating Databases” section of the online help file **faman.htm** (help for the Protein Databases page). If the links have become obsolete, check the *Agilent Software Status Bulletin and Patches* Web site for a file update. See the *Quick Start Guide* for details to connect to this Web site.
 - If you want to automate this process, you may want to use an ftp utility.
- 2 To download the database file, drag the file (zip format) from the browser window to an open directory window on your Spectrum Mill server. For performance reasons, it is best to:
 - Locate the databases in a separate drive from the rest of the Spectrum Mill files.
 - Locate the databases on the Spectrum Mill server PC rather than on a remote computer.
- 3 Use a ZIP file utility program to extract the database file into the **seqDB** directory in your Spectrum Mill workbench installation.
- 4 Rename the database file with the required name prefix, as described in the online help. Use capital letters *exactly* as described in the online help. This prefix is critical for the Spectrum Mill workbench to parse the header for each protein to retrieve protein name and species information. Use of a file extension (such as **.fas**) is optional.

NOTE

When you update a database, use the same file name. Do not append a new suffix. Otherwise, you will have problems when you review older data.

You *must* recreate database indices each time you update a database.

- 5 If you download a second database of the same FASTA format, and you want to link to different URLs from the search results, you rename the second database with a similar lower case database prefix. See the “Updating Databases” section of the online help file **faman.htm**. See also “[To change the URLs of HTML links in the search results](#)” on page 212.
- 6 If you want to search a user-created database, it is important to know the FASTA format of the database and to follow Spectrum Mill naming conventions. See the “The Spectrum Mill Workbench File Naming Conventions for Proprietary/Generic FASTA Databases” in the online help file **faman.htm**.
- 7 Follow directions in the next section to create database indices.

To create database indices

- 1 If you have not done this before, read the explanation at the end of the task.
- 2 If you wish to use the command line version of **Protein Databases** for automation purposes, skip the rest of this section and see the online Help for **Protein Databases**. Read the section entitled “The Command Line Version of Protein Databases.” Otherwise, continue to the next step.
- 3 To use a browser window to create indices, click the link to the **Protein Databases** page from the **Utilities** section of the Spectrum Mill home page.
- 4 Select the **Create indices for new database** option.
- 5 Check that you see the page shown in [Figure 99](#).
- 6 Type the *exact* name of the new database. Be sure to use capital and lower-case letters *exactly* as in the database name.
- 7 Click the **Create Indices** button.
- 8 As the indices are created, check that you see a display like the one in [Figure 100](#).
- 9 Click the **Back** arrow to return to the **Protein Databases** page.
- 10 Click **Update Database List** and verify that the database you just indexed appears in the list of databases.

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill MS/MS Search PMF Search Tool Belt Help

☒ Create indices for new database
 ☐ Concatenate FASTA files
 ☐ Compare two databases

☐ Re-index existing database
 ☐ Create subset with indices from saved hits
 ☐ Calculate statistics

☐ Create species subset database
 ☐ Make subset FASTA file from Accession Numbers
 ☐ Make Non-redundant

☐ Database summary report
 ☐ Create or append user database
 ☐ Create category file

Update Database List (After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button.)

Create Indices

Newly downloaded database:

Existing databases: ▼

Figure 99 Create indices for new database

Protein Database
Counting entries in database and creating index file (.idx). Processing entries to find species, accession number, and unknown sequence characters. Calculating Protein MW and pI and creating unknown sequence characters file (.unk). 18 entries in the database. 18 entries processed. 0 entries contained unreadable species. Creating mole_wt file (.rmw). Creating pI file (.pI). Creating species file (.sp). Creating species list file (.sl). Creating accession_number file (.acn). Indexed Database: NCBInr.stdmix

Figure 100 Display while database indices are created

Indices serve a number of useful purposes in the Spectrum Mill workbench, including:

- Faster output generation when you review results
- Accelerated searches when you use molecular weight or species filters. See [“To create a species and protein molecular weight subset database”](#) on page 201.
- Searches of results saved from other search programs. See [“To create a saved results file so you can search previous hits”](#) on page 156 and [“To create a subset database from saved hits”](#) on page 204.
- Creation of custom databases. See [“To create a user \(proprietary\) database or add sequences to one”](#) on page 206.

NOTE

Spectrum Mill indices bear no relationship to indices used in another popular protein database search program.

To re-index an existing database

Do this when you download a new version of a database.

- 1 Click the link to the **Protein Databases** page from the **Utilities** section of the Spectrum Mill home page.
- 2 Select the **Re-index existing database** option.
- 3 Check that you see the page shown in [Figure 101](#).
- 4 For **Existing database to re-index**, select a name.
- 5 Click the **Re-index** button.
- 6 Click **Update Database List** and verify that the database you just indexed appears in the list of databases.

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill MS/MS Search PMF Search Tool Belt Help

☐ Create indices for new database ☐ Concatenate FASTA files ☐ Compare two databases

☒ Re-index existing database ☐ Create subset with indices from saved hits ☐ Calculate statistics

☐ Create species subset database ☐ Make subset FASTA file from Accession Numbers ☐ Make Non-redundant

☐ Database summary report ☐ Create or append user database ☐ Create category file

Update Database List (After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button)

Re-index

Existing database to re-index: SwissProt

Figure 101 Re-index an existing database

To create a species and protein molecular weight subset database

For faster searches, you create subset databases based on species and molecular weight filters.

- 1 To create a subset database, click the link to the **Protein Databases** page from the **Utilities** section of the Spectrum Mill home page.
- 2 Select the **Create species subset database** option.
- 3 Check that you see the page shown in [Figure 102](#).
- 4 Fill in the page. Choose a descriptive suffix for the database, such as “human” or “yeast.”
- 5 Click the **Create** button.
- 6 In this example, the created subset database is **SwissProt.mammals**.

Figure 102 Create species and protein molecular weight subset database

NOTE

Because of inconsistencies in the way species information is organized in different databases, the Spectrum Mill workbench cannot read about 1 to 2% of the species designations in NCBIInr, and may not read all the species information in trEMBL or SwissProt. To create a molecular weight subset database from these databases, set the **Species** filter to **All**.

The **species.txt** file can be edited to improve the species filtering

To generate a database summary report

You use a database summary report to view a range of database entries. This report is useful for troubleshooting.

- 1 To create a summary report, click the link to the **Protein Databases** page from the **Utilities** section of the Spectrum Mill home page.
- 2 Select the **Database summary report** option.
- 3 Check that you see the page shown in [Figure 103](#).
- 4 Fill in the page. A typical **Index number range** is 1 to 100. The summary report will then allow you to see the next 100 (or your range), and so on.
- 5 Click the **Summarize** button.

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill MS/MS Search PMF Search Tool Belt Help

☐ Create indices for new database
 ☐ Concatenate FASTA files
 ☐ Compare two databases

☐ Re-index existing database
 ☐ Create subset with indices from saved hits
 ☐ Calculate statistics

☐ Create species subset database
 ☐ Make subset FASTA file from Accession Numbers
 ☐ Make Non-redundant

☒ Database summary report
 ☐ Create or append user database
 ☐ Create category file

Update Database List (After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button)

Summarize

Database: SwissProt

Reading frame: 1 (for DNA only)

Index number range: 1 to 100 ☐ All

☐ Hide protein sequences

Figure 103 Database summary report

To concatenate databases (FASTA files)

You use concatenate FASTA files to combine databases. The files you concatenate must reside under the SeqDB folder. If you concatenate all files in a folder, the folder must reside under SeqDB.

- 1 To combine databases, click the link to the **Protein Databases** page from the **Utilities** section of the Spectrum Mill home page.
- 2 Select the **Concatenate FASTA files** option.
- 3 Check that you see the page shown in [Figure 104](#).
- 4 Click either **Select files to concatenate** or **Concatenate files in folder**.
- 5 Select the existing databases to concatenate, or (if you clicked **Concatenate files in folder**) enter paths to FASTA files.
- 6 Click the **Concatenate** button. The new database is created in the SeqDB folder.
- 7 After concatenating, you can use the **Make Non-redundant** tool to remove redundant entries.

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill

MS/MS Search

PMF Search

Tool Belt

Help

☐ Create indices for new database
 ☒ Concatenate FASTA files
 ☐ Compare two databases

☐ Re-index existing database
 ☐ Create subset with indices from saved hits
 ☐ Calculate statistics

☐ Create species subset database
 ☐ Make subset FASTA file from Accession Numbers
 ☐ Make Non-redundant

☐ Database summary report
 ☐ Create or append user database
 ☐ Create category file

(After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" butt

☒ Select files to concatenate
 ☐ Concatenate files in folder

New database (relative to SeqDB):
 (The name must begin with a supported database name prefix.)

Existing databases:

SwissProt.ecoli
SwissProt.mammals

Figure 104 Concatenate FASTA files

To create a subset database from saved hits

The Spectrum Mill workbench allows you to create a subset database from a *.res file (list of accession numbers) you save in MS Edman. This is useful because MS Edman allows you to do a text search on a database and then save the results. The text search could find, for example, all proteins with a given name, or all proteins that contain a certain partial amino acid sequence. You use Protein Databases to read the *.res file and create a new subset database. You can then use any of the Spectrum Mill database search programs to search this database from any data directory.

- 1 Make sure you have already run MS Edman and saved hits to a file.
- 2 To create a subset database, click the link to the **Protein Databases** page from the **Utilities** section of the Spectrum Mill home page.
- 3 Select the **Create subset with indices from saved hits** option.
- 4 Check that you see the page shown in [Figure 105](#).
- 5 Fill in the page. Be sure to use a unique suffix so you don't overwrite previously-created databases.
- 6 Click the **Create** button.

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill | MS/MS Search | PMF Search | Tool Belt | Help

☐ Create indices for new database
 ☐ Concatenate FASTA files
 ☐ Compare two databases
☐ Re-index existing database
 ☒ Create subset with indices from saved hits
 ☐ Calculate statistics
☐ Create species subset database
☐ Make subset FASTA file from Accession Numbers
☐ Make Non-redundant
☐ Database summary report
☐ Create or append user database
☐ Create category file

Update Database List (After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button.)

Create

Suffix for subset database: lastres

Database used in search to save hits: SwissProt

From MS Edman results file: lastres

Figure 105 Create subset database with indices from saved hits

To make a subset FASTA file from Accession Numbers

This option creates a subset FASTA file from accession numbers that you type. It is useful for limiting searches to the set of proteins of particular interest.

- 1 Under the Utilities section of the Spectrum Mill Home page, click **Protein Databases**.
- 2 Click the **Make subset FASTA file from Accession Numbers** option.

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill | MS/MS Search | PMF Search | Tool Belt | Help

☐ Create indices for new database
 ☐ Concatenate FASTA files
 ☐ Compare two databases
☐ Re-index existing database
 ☐ Create subset with indices from saved hits
 ☐ Calculate statistics
☐ Create species subset database
☒ **Make subset FASTA file from Accession Numbers**
☐ Make Non-redundant
☐ Database summary report
☐ Create or append user database
☐ Create category file

(After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button)

Make Subset

Suffix for subset database:

Existing database: SwissProt ▼

Accession #'s: ▲▼

Figure 106 Make subset FASTA file from Accession Numbers option

- 3 In the **Suffix for subset database** field, type the name of the suffix for your database.
- 4 From the **Database** list, choose a database.
- 5 Type the accession numbers you want to include, separated by a semi-colon (;).
- 6 Click Make Subset.
- 7 On the Spectrum Mill server, navigate to the folder where your databases are stored (for example, D:\seqdb).
Note the new files created there.
- 8 Create indices for the new subset database. See [“To create database indices”](#) on page 198.

To create a user (proprietary) database or add sequences to one

The Spectrum Mill workbench allows you to create and search your own FASTA-format databases.

- 1 To create a user database, click the link to the **Protein Databases** page from the **Utilities** section of the Spectrum Mill home page.
- 2 Select the **Create or append user database** option.
- 3 Check that you see the page shown in [Figure 107](#).
- 4 Follow procedures in the online help document **faman.htm**. It is best if you read the entire document. As a minimum, read the following two sections:
 - “Creating or Appending to a Database Containing User Supplied Protein or DNA Sequences”
 - “The Spectrum Mill Workbench File Naming Conventions for Proprietary/Generic FASTA Databases.” (The database will not work if you fail to name it correctly.)

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill	MS/MS Search	PMF Search	Tool Belt	Help
☐ Create indices for new database ☐ Concatenate FASTA files ☐ Compare two databases ☐ Re-index existing database ☐ Create subset with indices from saved hits ☐ Calculate statistics ☐ Create species subset database ☐ Make subset FASTA file from Accession Numbers ☐ Make Non-redundant ☐ Database summary report ☒ Create or append user database ☐ Create category file				
Update Database List (After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button)				

Create

Database name:

Description:

Species: Accession number:

Paste or type sequence in this field. Delete these instructions first.

Tabs, returns, and spaces are ignored.

USE CAPITAL LETTERS.

Do NOT use the letters B, J, O, U, X, or Z.

Do NOT use 3-letter code amino acid symbols.

DNA sequences can be added to DNA databases.

Figure 107 Create or append to user database

To compare two databases

With this utility you can compare the listings of two databases to view their similarities and differences.

- Longest and shortest proteins in the two selected databases
- # of entries of exactly matching sequences
- # of matching entries with different accession #s
- # of entries in one database with no matching sequences in the other
- Sequence Length Histograms for both databases

- 1 From the Utilities section of the Spectrum Mill Home page, click **Protein Databases**.
- 2 Click **Compare two databases**.
- 3 Select a database from the **Database 1** list.
- 4 Select a database from the **Database 2** list.

The screenshot shows the 'Agilent Spectrum Mill - Protein Database Utilities' window. At the top, there are tabs for 'Spectrum Mill', 'MS/MS Search', 'PMF Search', 'Tool Belt', and 'Help'. Below the tabs, there are several radio button options. The 'Compare two databases' option is selected. Other options include 'Create indices for new database', 'Re-index existing database', 'Create species subset database', 'Database summary report', 'Concatenate FASTA files', 'Create subset with indices from saved hits', 'Make subset FASTA file from Accession Numbers', 'Create or append user database', 'Calculate statistics', 'Make Non-redundant', and 'Create category file'. Below the radio buttons, there is a button labeled 'Update Database List' and a note: '(After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button)'. At the bottom, there is a green 'Compare' button. Below the 'Compare' button, there are two dropdown menus. The first dropdown menu is labeled 'Database 1:' and has 'NCBIInr.stdmix' selected. The second dropdown menu is labeled 'Database 2:' and has 'NCBIInr.yeast' selected.

Figure 108 Compare two databases option

- 5 Click **Compare**.

To calculate statistics

If you choose this option, the program calculates these statistics:

- Number of entries with leading methionine
- Number of distinct peptides of length 8-40
- Number of non-distinct peptides of length 8-40
- Most repeated peptide
- Distinct Peptide Redundancy Histogram
- Sequence Length Histogram

- 1 From the Utilities section of the Spectrum Mill Home page, click **Protein Databases**.
- 2 Click **Calculate Statistics**.

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill MS/MS Search PMF Search Tool Belt Help

☐ Create indices for new database
 ☐ Concatenate FASTA files
 ☐ Compare two databases

☐ Re-index existing database
 ☐ Create subset with indices from saved hits
 ☒ Calculate statistics

☐ Create species subset database
 ☐ Make subset FASTA file from Accession Numbers
 ☐ Make Non-redundant

☐ Database summary report
 ☐ Create or append user database
 ☐ Create category file

Update Database List (After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button.)

Calculate Stats

Database: SwissProt

☐ Generate table of number of observable tryptic peptides per protein

Figure 109 Calculate statistics option

- 3 From the **Database 1** list, select the database for which you want the program to calculate statistics.
- 4 Click **Calculate Stats**.
- 5 Click **Update Database List**.

To remove redundant database entries

- 1 From the Utilities section of the Spectrum Mill Home page, click **Protein Databases**.
- 2 Click **Make Non-redundant**.

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill	MS/MS Search	PMF Search	Tool Belt	Help
☐ Create indices for new database ☐ Concatenate FASTA files ☐ Compare two databases ☐ Re-index existing database ☐ Create subset with indices from saved hits ☐ Calculate statistics ☐ Create species subset database ☐ Make subset FASTA file from Accession Numbers ☒ Make Non-redundant ☐ Database summary report ☐ Create or append user database ☐ Create category file				
Update Database List (After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button)				

Make Non-redundant

When keeping/removing redundant entries (identical sequence), only the one nearest the top of the FASTA file is kept. However, for UniProt databases SwissProt entries are preferentially kept over TrEMBL entries regardless of order in the FASTA file.

Database: SwissProt

Figure 110 Make Non-redundant option

- 3 From the **Database** list, select the database for which you want the program to remove redundant entries.
- 4 Click **Make Non-redundant**.

The “non-redundant” file will be named the same as the selected database, but with “nr” appended. For example, specifying **NCBI_{nr}.yeast** will create the non-redundant file named **NCBI_{nr}.yeast_{nr}**.

To create a category file

This option creates a subset FASTA file from accession numbers that you type. A category file is used in Protein/Peptide Summary, where you can create your own report fields to associate with a protein. When you mark the check box for **Categories** in the Review Fields within Protein/Peptide Summary, the program displays the user-defined category for each protein. For more information, see the online Help for Server Administration.

- 1 Under the Utilities section of the Spectrum Mill home page, click **Protein Databases**.
- 2 Click the **Create category file** option.

Agilent Spectrum Mill - Protein Database Utilities

Spectrum Mill MS/MS Search PMF Search Tool Belt Help

☐ Create indices for new database
 ☐ Concatenate FASTA files
 ☐ Compare two databases

☐ Re-index existing database
 ☐ Create subset with indices from saved hits
 ☐ Calculate statistics

☐ Create species subset database
 ☐ Make subset FASTA file from Accession Numbers
 ☐ Make Non-redundant

☐ Database summary report
 ☐ Create or append user database
 ☒ Create category file

(After a re-install or indexing a database, if you don't see the database listed, click the "Update Database List" button)

Database: NCBIInr.ecoli

Figure 111 Create a category file

- 3 From the **Database** list, choose a database.
- 4 Click **Create File**.
- 5 On the Spectrum Mill server, navigate to the folder where your databases are stored (for example, D:\seqdb).
Note the new file created there.
- 6 Create indices for the new category file. See [“To create database indices”](#) on page 198.

Other system administration tasks

To add custom amino acid modifications

To set up custom amino acid modifications, start with the file `\msparams_mill\smconfig.custom.empty`.

CAUTION

Do *not* modify **smconfig.std.xml** or **smconfig.xml**.

The file **smconfig.misc.xml** contains definitions that are provided as a convenience but that have not been thoroughly tested at Agilent. Please verify the definitions before you use them.

- 1 Copy **smconfig.custom.empty** to **smconfig.custom.xml**.
- 2 Create definitions for modifications and/or mappings for homology search modes.
 - To make the task easier, examine the files **smconfig.std.xml** or **smconfig.misc.xml** for modification definitions that are similar to the ones you need. You may copy and paste these definitions into your **smconfig.custom.xml** file, and then modify the definitions.
 - See the online help document **SMCustomModifications.pdf** for details and syntax for adding modifications.
- 3 To make your custom modifications available in the software, run the perl script `\millscripts\updateModsFile.pl` or click the **Choose...** button in one of the Spectrum Mill pages.
 - The Spectrum Mill workbench automatically combines the **smconfig.custom.xml** file and the **smconfig.std.xml** file to produce the **smconfig.xml** file that is used by the Spectrum Mill programs.
 - When the software creates the **smconfig.xml** file, definitions in **smconfig.custom.xml** override any definitions in **smconfig.std.xml**.

NOTE

The **smconfig.custom.xml** file is *not* overwritten during a re-install of the Spectrum Mill workbench.

To change the URLs of HTML links in the search results

By default, search results use the HTML links shown in [Table 8](#). You edit text files listed in the last column of [Table 8](#) to change these links.

- 1 Make a backup of each file before you modify it.
- 2 Edit the file. See the online help document **servadmn.htm** for details.

Table 8 HTML links in search results

Click this	To go to this	Text file to edit link
Accession number	Remote database entry	mparams_mill\urlsAccessionNum.txt
MS Digest index number	MS Digest results for database sequence	mparams_mill\urlsIndexNum.txt
Amino acid sequence (in individual MS/MS or PMF search results)	MS Product listing for that sequence	mparams_mill\urlsSequence.txt

To enable the HTML link to BLAST search

If you wish to have the amino acid sequence in Protein/Peptide Summary link to BLAST (Basic Local Alignment Search Tool), edit the file **mllscripts\libSpecMill\lsmGlobals.pm**.

- 1 Make a backup of the **mllscripts\libSpecMill\lsmGlobals.pm** file.
- 2 Open the **mllscripts\libSpecMill\lsmGlobals.pm** file in a plain text editor (e.g., Notepad) to edit it.
- 3 Change the following line (near line 36):
 - From: **\$GLOBAL_ADD_BLAST_LINK = 0;**
 - To: **\$GLOBAL_ADD_BLAST_LINK = 1;**
- 4 Save the file.

If you enable the BLAST link, be aware that it will take longer to load large data sets in Protein/Peptide Summary.

To add/change options related to biology/chemistry

The Spectrum Mill workbench allows you to modify a number of parameters related to the biology or chemistry of your study. To customize these options, you modify text files in the folder **mparams_mill**.

- 1 Archive each file before you modify it.
- 2 Edit the file(s). See the online help document **servadm.html** for details. For edits that involve **smconfig.custom.xml**, see also “[To add custom amino acid modifications](#)” on page 211.

The biology/chemistry options that you can change include:

- Species filter (**species.txt**)
- Cysteine modification options (**smconfig.custom.xml**)
- N- or C-terminal groups (**smconfig.custom.xml**)
- Other custom amino acid modifications, including those used for differential expression quantitation (**smconfig.custom.xml**)
- Elements (**smconfig.custom.xml**)
- Enzyme cleavage rules (**enzyme.txt**)
- Categories of proteins (**categories.#.tsv**)

To add instrument types

You can add instrument types and configure MS/MS Search scoring and peak detection parameters for these instruments.

See the online help document **servadm.html** for details.

- 1 Archive these files before you modify them:
 - **mparams_mill\instrument.txt**
 - **millhtml\SM_js\instrument.js**
- 2 Edit the files.

To configure IEF Selector defaults

For Peptide Selector, you can configure these options for isoelectric focusing (IEF):

- pH range choices
- Fraction choices

Both are configured by editing JavaScript in the file **protChem.js**, which is located under **\SpectrumMill\millhtml\sm_js** on your server. Save a backup copy of **protChem.js** prior to editing. Once edits are complete, save a backup copy of the updated file so you can restore it if you reinstall the Spectrum Mill workbench. To edit **protChem.js**, you must have knowledge of JavaScript syntax.

To customize the list of pH ranges to select:

- 1 Edit **protChem.js**.
- 2 Look for the **displayIEFpHSelections()** function near the bottom of the file.
- 3 Customize the function to add the appropriate "<option>" elements to the selection lists.
- 4 Save the file, refresh the Peptide Selector page, and verify that the new selections are present and no JavaScript errors occur.
- 5 Save a backup copy of the updated **protChem.js**.

To customize the list of fractions to select:

- 1 Edit **protChem.js**.
- 2 Look for the **displayIEF_fractions()** function near the bottom of the file.
- 3 Customize the function to add the appropriate "<option>" elements to the selection lists.
- 4 Save the file, refresh the Peptide Selector page, and verify that the new selections are present and no JavaScript errors occur.
- 5 Save a backup copy of the updated **protChem.js**.

To maintain server performance

To maintain server performance:

- Periodically archive data that is no longer needed; remove it from the server and defragment the data drive.

Use the Tool Belt feature for archiving data to zip up files in folders that are not necessary for your current projects. You can unzip them if you need to look at or reprocess the data later.

- Always defragment the database drive after installing a new database.

Spectrum Mill data directories contain many very small files. Over time, disk fragmentation from accumulated files causes newly-added raw data files to become heavily fragmented. If you fail to defragment the data drive, Data Extractor speed declines.

If you fail to defragment the database drive, search speeds decline.

To use server administration scripts

The **milladmin** folder provides scripts that system administrators may use to configure the Spectrum Mill installation. These scripts:

- Configure IIS for the Spectrum Mill workbench
- Set file permissions for critical Spectrum Mill files
- Set file access permissions for **msdataSM** and subfolders
- Create data file placeholders to allow you to remove raw data files

For details, click the Help link for Server Administration, then see the section on “Server Administration Scripts (milladmin folder).”

To avoid problems with connection time-outs

The default Internet Information Server (IIS) settings applied by the Spectrum Mill installation may not be suitable when you process large data sets with long extraction and search times. To avoid having your connections time out, change IIS settings on the Spectrum Mill server. To change settings, click the Help link for Server Administration, then see the section on “Resolving Connection Timeout Problems.”

To configure server with drives other than the default configuration

- 1** Read the discussion below.
- 2** If the hard drive configuration changes (e.g. additional hard drive installed), then you need to uninstall and reinstall the Spectrum Mill software to see new partitions. If you need to move the collected data to a new data storage path, you must do this manually.

The Spectrum Mill installation program examines your current drive configuration and allows you to select from the partitions available where to install the Spectrum Mill program and example data. The default choice is the largest partition found. This path is highly recommended since all data used with the Spectrum Mill program must reside in the same partition as the program, and because the data will require the most space for storage.

The software installation also allows you to designate the location for the database(s). By default, the second largest partition is selected, but as with the program and data, you can change this designation. The second largest partition is recommended for installation of the databases because they are typically very large in size. (For example, NCBI in 2015 required 28 GBytes.) You should assign the Windows® Paging File location to the same partition since few disk writes will occur that could lead to fragmentation of the Paging File.

To remove and reinstall the Spectrum Mill workbench

If you wish to remove and reinstall the Spectrum Mill workbench, see the *Spectrum Mill MS Proteomics Workbench Installation Guide*.

If you remove the Spectrum Mill workbench and intend to reinstall it, note the following:

- Software removal does not delete the data folders (**msdataSM**), the protein databases (**seqDB**), or the workflow-related files (**millauto**).
- Be sure to archive and restore any configuration files that have been customized for your site. These might include:
 - **millhtml\sm_js** (**instrument.js**, **massOpt.js**, **protChem.js**, **species.js**)
 - **millscripts\libSpecMill\lsmGlobals.pm**
 - **mparams_mill** (various *.txt files, including contaminant mass list files)

To limit the number of parallel workflow processes to less than the CPU count

Before you modify any files, be sure to archive the originals so they can be restored if necessary!

When you execute a workflow, its tasks are added to the Spectrum Mill Request Queue, and executed in order for a particular data folder. If you select multiple data folders, the data folders are processed in parallel (extraction and search tasks) if you have more than one CPU.

The Spectrum Mill installation configures the number of parallel processes to one less than the number of (logical) CPUs. If you disabled hyperthreading after you installed the Spectrum Mill workbench, you should reduce the default value to one less than the number of processors viewed in Windows Task Manager.

To modify the maximum number of CPUs, first stop the Windows Service Request Manager (SRM). Then edit the file:

E:\SpectrumMill\millsrm\SMSRM.config

- 1 Find the line that contains **maxConcurrentTasks**. The line of text is similar to this example: `<provider hostname="localhost" available="true" maxConcurrentTasks="3" minRequiredTaskMemoryGb="2">`

For multicore processing, the new setting **minRequiredTaskMemoryGb="2"** is the minimal amount of free memory available for Spectrum Mill to run more than one concurrent task. (The **minRequiredTaskMemoryGb** setting prevents you from running out of memory if you use Maximize CPUs to queue multiple extractions.)

- 2 Change the number after **maxConcurrentTasks** and resave the file.
- 3 Restart the SRM service for the new setting to take effect.

See [“To start and stop the Spectrum Mill Workflow Manager Service”](#) below.

To start and stop the Spectrum Mill Workflow Manager Service

Spectrum Mill uses a Windows Service Request Manager (SRM) service to process workflows. The “Spectrum Mill Workflow Manager Service” is configured to automatically start when Windows starts.

To stop the service:

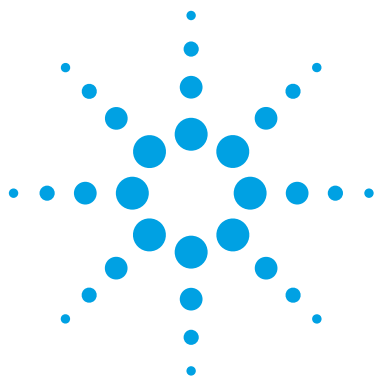
- 1 Wait until the Request Queue is empty. (Anything in the queue will be lost.)
- 2 Navigate to **E:\SpectrumMill\millsrc**.
- 3 Right-click **stopSRM.bat** and select **Run as administrator**.
- 4 If the service continues to run, as a last resort, stop it with Windows Task Manager. In the **Processes** tab of Task Manager, click **SRMHostSvc.exe**, and then click **End Process**.

To start the service:

- 1 Navigate to **E:\SpectrumMill\millsrc**.
- 2 Right-click **startSRM.bat** and select **Run as administrator**.

9 System Administration

Other system administration tasks



10 Files Created during Spectrum Mill Data Processing

Architecture Overview	222
Data Extractor (MS/MS raw data)	223
Data Extractor (generic data)	225
Data Extractor (MS-only raw data)	227
MS/MS Search	228
Protein/Peptide Summary, Spectrum Summary, and Autovalidation	230
Tool Belt	232
Sherenga de novo Sequencing	234
PMF Search	235
MS Edman	236
Protein Databases	237

This chapter begins with an architecture overview and then lists the files the Spectrum Mill workbench creates as you process results from analyses of protein digests. The file lists are organized by data processing module. Since the Spectrum Mill modules automatically create and call these files as needed, you don't need to know about them to use the software. They are listed as an aid to troubleshoot data processing, to selectively remove parts of the processing, and to decide which files to archive.



Architecture Overview

The following figure shows an overview of the Spectrum Mill architecture, including programs, scripts, and some of the files that are produced. For more details on the files, see the tables that follow.

Note that the *.spo files are temporary, but the program creates an spo.zip file to store them permanently.

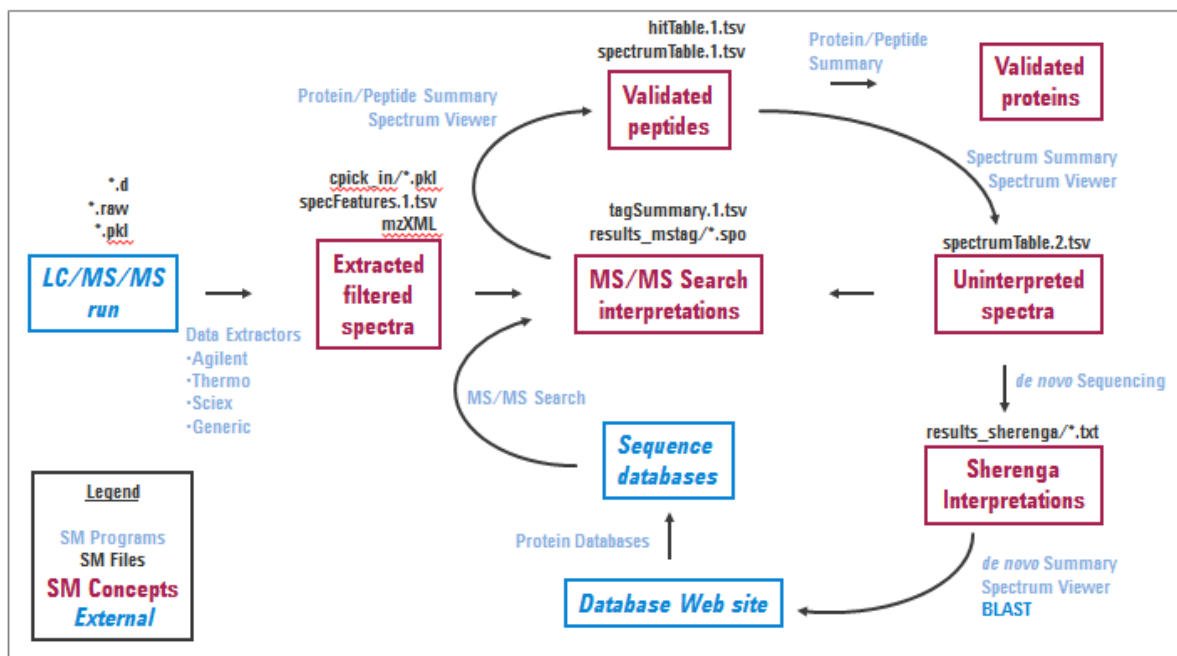


Figure 112 Architecture for MS/MS workflow

Data Extractor (MS/MS raw data)

Table 9 lists the files created when you run the MS/MS raw data extractor. These are all created under the **msdataSM** directory, in the subdirectory that contains your raw data files. All are text files; you can open them with Notepad.

Table 9 Files created by MS/MS raw data extractor

Files and Folders Created	Contents	Notes
parentChargeCheck.txt	- Scan number in format Data_File_Name.####, where #### is the scan number - Charge assignment for precursor ion - Precursor ion - Number of b/y pairs found	
xtractorRejectReasons.txt	- Scan number in format Data_File_Name.#### for single scan or Data_File_Name.aaaa.bbbb for merged scans, where - aaaa = first merged scan - bbbb = last merged scan - Reject reasons. Examples are: - Scans merged based on precursor ion and time window tolerances - Minimum sequence tag length requirement not met - Precursor ion out of range specified for MH⁺ <4 peaks after picking	Useful to troubleshoot data extraction problems.
runStats.txt	Various statistics for the extractor run, including numOutputFiles, which is the number of extracted spectral files	
SpecFeatures.1.tsv	Spectral features calculated for each spectrum	
xtractor.1.params	Record of data extractor parameters	

10 Files Created during Spectrum Mill Data Processing

Table 9 Files created by MS/MS raw data extractor (continued)

Files and Folders Created	Contents	Notes
cpick_in*.pkl	- Series of files - one for each spectrum extracted - File name in 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)	- If you rerun the Data Extractor using a new set of extraction parameters, mark the check box to Remove all prior results. - Agilent Q-TOF and Thermo .raw data extractors generate mzXML files rather than *.pkl files.
Series of folders in the subdirectory that contains your data file(s). Folders are listed in next column.	- cpick_in - fit_batch_in - results_msedman - results_msfit - results_mstag - results_sherenga	
*.aph for .d (Agilent) files *.rph for .raw (Thermo) files *.gph for .mgf (generic extractor) *.lnk for a linked archive file	These are placeholder files that correlate extracted spectra with original raw data file.	File name in format: Data_File_Name.xxx, where "xxx" is the file suffix.

Data Extractor (generic data)

Table 10 lists the files created when you run the generic data extractor. These are all created under the **msdataSM** directory, in the subdirectory where you placed your appended *.pkl spectral files, (or one level above the **cpick_in** subdirectory where you placed your individual *.pkl or *.dta spectral files). All are text or html files.

Table 10 Files created by generic data extractor

Files and Folders Created	Contents	Notes
parentChargeCheck.txt	• Headers for scan number and charge assignment • No other data	
xtractorRejectReasons.txt	• Spectral file name • Reject reasons	• Useful to troubleshoot data extraction problems
runStats.txt, runStats.htm	• Empty files	
SpecFeatures.1.tsv	• Spectral features calculated for each spectrum in the subdirectory cpick_in	
qtofXtractor.1.params	• Record of data extractor parameters	

10 Files Created during Spectrum Mill Data Processing

Table 10 Files created by generic data extractor (continued)

Files and Folders Created	Contents	Notes
<pkIFileName>.mzXML	- Series of files - one for each spectrum extracted - created when you extract an appended *.pkl file that contains multiple spectra - File name in format: Data_File_Name.Scan_Number.0. Precursor_Charge.pkl, where - Data_File_Name = prefix of appended *.pkl file - Scan_Number = consecutive order of the scan in the appended *.pkl file 0 = placeholder where function number would be if created by Micromass ProteinLynx software - Precursor_Charge = charge of the precursor ion	If you rerun the Data Extractor using a new set of extraction parameters, mark the check box to Remove all prior results.
Series of folders in the subdirectory that contains your data file(s). Folders are listed in next column.	- cpick_in - fit_batch_in - results_msedman - results_msfit - results_mstag - results_sherenga	

Data Extractor (MS-only raw data)

Table 11 lists the files created when you run the MS-only raw data extractor. These are all created under the **msdataSM** directory, in the subdirectory that contains your raw data files. All are text files; you can open them with Notepad.

Table 11 Files created by MS-only raw data extractor

Files and Folders Created	Contents	Notes
parentChargeCheck.txt	• Headers for scan number and charge assignment • No other data	
xtractorRejectReasons.txt	• Headers for spectrumName, numScansAfterParent, and RejectReason • No other data	
runStats.txt	• Information such as fileName and sampleName	
SpecFeatures.1.tsv	• Various headers • No other data	
xtractor.1.params	• Record of data extractor parameters	
fit_batch_in*.mi or fit_batch_in*.mic	• Extracted file • File name in format: Data_File_Name.*mi or Data_File_Name.mic	• If you rerun the Data Extractor using a new set of extraction parameters, mark the check box to Remove all prior results.
Series of folders in the subdirectory that contains your data file(s). Folders are listed in next column.	• cpick_in • fit_batch_in • results_msedman • results_msfit • results_mstag • results_sherenga	

MS/MS Search

Table 12 lists the files created when you run MS/MS Search. These are all created under the **msdataSM** directory, in the subdirectory that contains your raw data files. All are either text or html files.

If you accidentally run a search with incorrect search settings, you should remove the associated files from your system, especially the files in the **results_mstag** subdirectory. Otherwise, they will be included in future results reviews, validations, and summaries. If you have run multiple searches, be sure to check the time and date stamps on the files so that you do not mistakenly remove good search results.

Table 12 Files created by MS/MS Search

Files Created	Contents	Notes
tagSummary.#.tsv	Summarizes all search results	- New file created each time you run MS/MS Search. Files are numbered tagSummary.1.tsv, tagSummary.2.tsv, etc. - You also create this file when you click the Create MS/MS search summary file button on the Tool Belt page.
mstag.#.params	Record of MS/MS Search parameters	- New file created each time you run MS/MS Search. Files are numbered mstag.1.params, mstag.2.params, etc. - When you search an ion trap data file that contains both CID and ETD spectra, then two additional files are created because there are two sets of search conditions. The additional files are of the formats mstag.1a.params and mstag1b.params.

Table 12 Files created by MS/MS Search (continued)

Files Created	Contents	Notes
results_mstag*.pkl.spo	- Series of text search results files - one for each extracted spectrum searched. (Some extracted spectra do not meet quality criteria for searching.) - File name (for spo files) in format: Data_File_Name.aaaa.bbbb.c.pkl.spo, where - aaaa = first merged scan - bbbb = last merged scan - c = assigned precursor charge (0 means charge was ambiguous) - Starting with Spectrum Mill B.0 6.00, spo files are automatically added to an spo.zip file and then removed when a search completes. This significantly reduces the load on the file system, and speeds archiving and copying results.	Contents of results_mstag subdirectory are updated each time you run MS/MS Search on the same file set.
results_mstag\junk.#.htm	Scratch file of html output before redirection to browser	New file created each time you run MS/MS Search. Files are numbered junk.1.htm, junk.2.htm, etc.
fdr"type".#.tsv, where "type" is the name extension for the .tsv file; e.g., fdrSummary.1.tsv	A series of FDR files created in MS/MS search and used by discriminant-based FDR	These files are not created when the Discriminant mode is set to Off.

Protein/Peptide Summary, Spectrum Summary, and Autovalidation

Table 13 lists the files created when you validate data with the Protein/Peptide Summary page, the Spectrum Summary page, or the Autovalidation page. These are all created under the **msdataSM** directory, in the subdirectory that contains your raw data files. All are text files.

Note that correlated files from a given validation step have the same time and data stamp, but not necessarily the same number in the file sequence. For example, **autovalidation.1.params** may be correlated with **spectrumTable.3.tsv** and **hitTable.2.tsv**. This is because multiple modules write to subsets of these files.

Table 13 Files created by Protein/Peptide Summary, Spectrum Summary, and Autovalidation

Files Created	Contents	Notes
hitTable.#.tsv	Record of MS/MS Search hit validation states	- Created when you click the Perform Validation button on the Protein/Peptide summary page or the Validate Files button on the Autovalidation page - New file created each time you click the button. Files are numbered hitTable.1.tsv, hitTable.2.tsv, etc. - Results are cumulative; each new file starts with the previous file and adds information.

Table 13 Files created by Protein/Peptide Summary, Spectrum Summary, and Autovalidation (continued)

Files Created	Contents	Notes
spectrumTable.#.tsv	- Record of spectrum validation states, including: - whether the spectrum has been designated good or bad - whether the spectrum has a validated database hit	- Created when you click the Update button on the Spectrum Summary page or the Perform Validation button on the Protein/Peptide summary page or the Validate Files button on the Autovalidation page - New file created each time you click the button. Files are numbered spectrumTable.1.tsv, spectrumTable.2.tsv, etc. - Results are cumulative; each new file starts with the previous file and adds information.
autovalidation.#.params	Record of autovalidation parameters	- Created when you click the Validate Files button on the Autovalidation page - New file created each time you click the button. Files are numbered autovalidation.1.params, autovalidation.2.params, etc.

Tool Belt

Table 14 lists the files created with the Tool Belt page. All are text files.

Table 14 File created by Tool Belt

File Created	Contents	Notes
*.res	• Results file of validated rank one hits • File name in format: Database_Name_with_any_extensions.res	• Created when you select the Create saved results file option and click the Create File button • Created in the results_mstag folder within the subdirectory that contains your raw data files • Overwritten each time you create a saved results file for that database and data directory
tagSummary.#.tsv	• Summarizes all search results	• Created when you select the Create MS/MS search summary file option and click the Create File button • Created in the subdirectory that contains your raw data files • The MS/MS Search module also writes to this file. See Table 12 on page 228.
ReporterIon.corrections.txt	• Reporter ion (iTRAQ/TMT) correction factors and batch numbers	• Created when you select the Create iTRAQ/TMT correction factors option, type a batch number and correction factors, and click the Create button • Created under the mparams_mill directory • File is appended every time you type correction factors for a new batch number • Software will not allow you to overwrite existing correction factors for a batch

Table 14 File created by Tool Belt (continued)

File Created	Contents	Notes
Discriminant Coefficients.txt	Includes predefined sets for Agilent discriminant coefficients and is updated when new coefficients are calculated.	
reporterion.corrections.txt	Reporter ion (iTRAQ/TMT) correction factors that apply to a particular sample or set of samples	- Created when you select the Apply iTRAQ/TMT correction factors option, set up the page, and click the Apply button - Created under the msdataSM directory, in the subdirectory that contains your raw data file(s)

Sherenga *de novo* Sequencing

Table 15 lists the files created when you run Sherenga *de novo* sequencing. These are all created under the **msdataSM** directory, in the subdirectory that contains your raw data files. All are text files.

Table 15 Files created by MS/MS Search

Files Created	Contents	Notes
sherenga.#.params	Record of Sherenga <i>de novo</i> Sequencing search parameters	New file created each time you run Sherenga <i>de novo</i> Sequencing. Files are numbered sherenga.1.params, sherenga.2.params, etc.
results_sherenga\#.pkl.txt	- Series of text Sherenga results files - one for each extracted spectrum processed. - File name in 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)	Contents of results_sherenga subdirectory are updated each time you run <i>de novo</i> sequencing on the same file set.

PMF Search

Table 16 lists the files created when you run PMF Searches. These are all created in the subdirectory that contains your spectral files, which is under the **msdataSM** directory. All are either text or html files.

Table 16 Files created by PMF Search

Files and Folders Created	Contents	Notes
msfit.#.params	Record of PMF Search parameters	New file created each time you run PMF Search. Files are numbered msfit.1.params, msfit.2.params, etc.
results_msfit*.MI.htm or results_msfit*.MIC.htm	- Series of html search results files - one for each extracted spectrum searched - File name formats: - Peak_File_Name.*MI.htm - Data_File_Name.*MI.htm - Data_File_Name.MIC.htm	Contents of results_msfit subdirectory are updated each time you run PMF Search on the same file set.
results_msfit*.MI.spo or results_msfit*.MIC.spo	- Series of text search results files - one for each extracted spectrum searched - File name formats: - Peak_File_Name.*MI.spo - Data_File_Name.*MI.spo - Data_File_Name.MIC.spo	Contents of results_msfit subdirectory are updated each time you run PMF Search on the same file set.
Series of folders in the subdirectory that contains the data file. Folders are listed in next column.	- cpick_in - fit_batch_in - results_msedman - results_msfit - results_mstag - results_sherenga	
results_msfit\junk.#.htm	Diagnostic html file created when PMF Search runs	New file created each time you run PMF Search. Files are numbered junk.1, junk.2, etc.

MS Edman

Table 17 lists the files created when you run MS Edman *and* you mark the check box to **Save hits to file**. These files are created under the **results_msedman** subdirectory directly under your Spectrum Mill installation. (They are not under a data file subdirectory.)

Table 17 Files created by MS Edman

Files Created	Contents	Notes
lastres.htm	• Search results in html format • File name is lastres or whatever you have chosen in the MS Edman page.	• Contents are overwritten unless you change the file name in the MS Edman page.
lastres.spo	• Empty file • File name is lastres or whatever you have chosen in the MS Edman page.	
lastres.par	• Search parameters in text format • File name is lastres or whatever you have chosen in the MS Edman page.	• Contents are overwritten unless you change the file name in the MS Edman page.
lastres.res	• List of accession numbers that correspond with MS Edman search hits • File name is lastres or whatever you have chosen in the MS Edman page.	• Contents are overwritten unless you change the file name in the MS Edman page.

Protein Databases

Protein Databases creates files in the directory where you store your sequence databases (usually **X:/seqDB**, where X is the disc drive letter).

Table 18 Files created by Protein Databases

Files Created	Contents	Notes
*.sub	- Subset database created from species and/or molecular weight information contained in the original database - File name includes sub or whatever suffix you have chosen in the Protein Databases page.	Created when you select the Create species subset database option and click the Create button
*.lastres	- Subset database created from hits saved in MS Edman - File name includes lastres or whatever extension you have chosen in the Protein Databases page.	Created when you select the Create subset with indices from saved hits option and then click the Create button
*.user	- User database - File name includes user or whatever extension you have chosen in the Protein Databases page.	Created when you select the Create or append user database option and then click the Create button
*.idx, idx2 (for databases > 4.2 Gb)	Index files	- Created when you click the Create button after you select any of these options on the Protein Databases page: Create indices for new database Create species subset database Create subset with indices from saved hits Create or append user database - See the online help document faman.htm for details regarding these files.
*.unk		
*.mw		
*.pi		
*.sp		
*.sl		
*.usp		
*.acc or *.acn		

10 Files Created during Spectrum Mill Data Processing

Index

A

- abort process, 154
- Agilent Q-TOF
 - Data Extractor, 13 to 17
 - manual data validation, 71, 78
- amino acid modifications
 - customizing, 211
 - list details, 161
- AMRT export, 75
- architecture overview, Spectrum Mill, 222
- archive data, 174
- automated workflow
 - completion log, 120
 - create/edit, 114
 - execute, 116
 - view request queue, 119
- automated workflows, 111 to 120
- automation
 - interactive, 13
- Autovalidation, 25 to 29, 61 to 68
 - auto thresholds strategy, 62
 - auto thresholds-discriminant strategy, 63
 - false discovery rate (FDR), 27
 - fixed thresholds strategy, 64
 - strategy/mode table, 27

B

- basic workflow, 13 to 33

C

- Choose button, 15, 19, 21, 51, 123, 133, 135, 177, 182, 211
- collection of spectra and interpretations, 168
- comparing two samples, 44
- completion log, 120
- concatenate
 - FASTA files, 203

- connection time-outs, resolving, 216
- contaminant mass list files, 217
- convert spectra
 - pkl files into mzXML files, 171
- copy spectra to collections directory, 168
- create discriminant scoring coefficients, 166
- create library for Spectrum Matcher, 168
- create MS/MS search summary file, 158
- create saved results file, 156
- create subset database from saved hits, 204
- create/edit an automated workflow, 114
- custom amino acid modifications, 211

D

- data display mode, 69, 90 to 98
- Data Extractor
 - generic peak list files, 17 to 20
 - MS/MS raw files, 13 to 17
- data file placeholders, 215
- database
 - previous hits, 156
 - saved results, 156
- database search
 - homology mode, 35 to 36, 48
 - manual PMF, 145 to 151
 - unknown modification, 36, 48
 - variable modifications mode, 20 to 25, 35 to 36
- database search modes, 36
- database summary report, 202
- databases
 - creating indices, 198, 200
 - installing, 196
 - molecular weight, 201
 - species, 201
 - subset, 204
 - updating, 196
 - user-created, 197, 206

- de novo* sequencing, 50 to 55, 121 to 129

- DEQ data

- Data Extractor, 133
 - MS/MS Search, 135
 - peptide mode, 137
 - protein mode, 139
 - Protein/Peptide Summary, 136
- differential expression quantitation, 73, 131 to 143, 213
- discriminant scoring
 - coefficients, 166
- dynamic peak thresholding, 22, 78

E

- evaluating results
 - MS/MS Search, 76 to 89
- Excel, 45, 98, 165
- export
 - AMRT, 75
 - Excel, 75
 - MPP, 75
 - pepXML files, 169

F

- false discovery rate (FDR), 62
 - autovalidation, 27
 - Tool Belt search statistics, 106 to 109
- FASTA files
 - concatenate, 203
- file
 - permissions, 215
 - placeholders, 215

Index

files created

- Autovalidation, 230
- Data Extractor (generic data), 225
- Data Extractor (MS/MS raw data), 223
- Data Extractor (MS-only data), 227
- MS Edman, 236
- MS/MS Search, 228
- PMF Search, 235
- Protein Databases, 237
- Protein/Peptide Summary, 230
- Sherenga *de novo* sequencing, 234
- Spectrum Summary, 230
- Tool Belt, 232

filtering search results, 70 to 72

filtering spectra, 37 to 40

folder selection, 14, 18, 20, 30, 38, 51, 76

G

global files, 217

graph theory, 121

guidelines

- manual results validation, 76 to 89

H

homology mode searches, 35 to 36, 48

HTML link to BLAST search, 212

HTML links to search results, 212

I

ICAT data

- Data Extractor, 133
- general discussion, 131 to 143
- MS/MS Search, 135
- peptide mode, 137
- protein mode, 139
- Protein/Peptide Summary, 136
- Spectrum Summary, 143

identification statistics

- summary table, 103

IIS configuration, 215

indices, 198, 200

installation, 217

installing databases, 196

interactive automation, 13

isotope-coded affinity tag data. *See* ICAT data.

isotopically labeled data, 131 to 143

iterative MS/MS data processing, 34 to 55

iterative workflow, 34 to 55

iTRAQ data

- correction factors, 140 to 142, 163 to 165

Data Extractor, 133

MS/MS Search, 135

overview, 131 to 132

peptide mode, 141

protein mode, 142

Protein/Peptide Summary, 140

J

journal submission, 101

L

light/heavy isotope ratios, 131 to 143

LIMS, 45

link to results, 117

list details of amino acid modifications, 161

M

MALDI-TOF-TOF

- Data Extractor, 17

manual PMF searches, 145 to 151

manual results validation

- Agilent Q-TOF, 71, 78
- ion trap CID data, 76 to 89

manual validation, 76 to 80

manuscript for journal, 101

mass gap search, 22, 35, 36, 48

Maximize CPUs, 14, 18, 25, 35

maximum sequence tag length, 37, 38

mode

Peptide - Spectrum Match, 91

Peptide Distinct, 92

Protein - Single Peptide ID, 101

Protein Summary Details, 93

Protein-Peptide Comparison, 98

Protein-Peptide Distribution Columns, 102

Protein-Protein Comparison, 95

Protein-Protein Comparison Redundant, 101

modifications, customizing, 211

molecular weight subset databases, 201

monitor results, 19, 117

MPP export, 75

MRM lists for triple quadrupole instruments, 179

MRM Selector, 179

MS Comp, 188

MS Digest, 182

MS Edman, 184

MS Isotope, 190

MS Product, 186

MS/MS data processing

- iterative, 34 to 55

MS/MS search

- homology mode, 35 to 36, 48

larger database, 47

no enzyme mode, 47

search modes, 36

summary file, 158

unknown modification, 36, 48

variable modifications mode, 20 to 25, 35 to 36

MS-only data processing

- Agilent TOF and Q-TOF .d files, 145 to 151

MS-only search

- manual, 145 to 151

Multiple Sequence Aligner, 181

mzXML files, 171

N

neutral loss, 78

no enzyme, 41
no enzyme mode search, 47

O

offgel fractionation, 79

P

parameter files, 37, 112
 MS/MS, 13
parameters
 summary table, 159
Peptide - Spectrum Match mode, 91
peptide charge, 87, 89
Peptide Distinct mode, 92
peptide list to masses utility, 193
peptide pl filtering, 79
Peptide Selector, 176
pepXML files, 169
phosphorylation, 22
pl, 30, 65, 79
placeholders, data files, 215
PowerPoint, 46
print results, 46
proprietary databases, 197, 206
Protein - Single Peptide ID mode, 101
protein and peptide utilities, 176 to 193
protein grouping, 71
protein quantitation options, 74
Protein Summary Details mode, 93
Protein/Peptide Summary, 69 to 75
 protein quantitation options, 74
Protein/Peptide Summary displays, 90 to 98
Protein-Peptide Comparison mode, 98
Protein-Peptide Distribution Columns mode, 102
Protein-Protein Comparison mode, 95
Protein-Protein Comparison Redundant mode, 101

Q

Q-TOF
 Data Extractor, 13 to 17
 manual data validation, 71, 78

R

rank1 - rank 2 scores, 78
report FDR and search statistics, 106 to 109
request queue, 19
results
 monitor, 19
results validation
 guidelines, 76 to 89
 parameters, 70 to 72
reversed database scores, 22, 78
review fields
 Protein/Peptide Summary, 72
 Spectrum Summary, 37

S

scripts, server administration, 215
search modes, 36, 184
search previous hits, 156, 204
search results
 changing HTML links, 212
 enabling HTML link to BLAST, 212
Select button, 14, 18, 20, 30, 38, 51, 76
sequence alignment, 181
sequence tag length, 37, 38
sequencing - *de novo*, 50 to 55, 121 to 129
server
 administration scripts, 215
 drives, 216
 performance, 215
Sherenga *de novo* sequencing, 50 to 55, 121 to 129
SILAC, 16, 132 to 139
smconfig.xml, 211
sorting search results, 70 to 72
sorting spectra, 37 to 40

species subset databases, 201
spectral validation parameters, 37 to 40
Spectrum Matcher, 55
 library, 168
Spectrum Mill server
 configuring drives, 216
 maintaining performance, 215
Spectrum Summary, 37 to 40
Spectrum Viewer, 83 to 86
stop process, 154
subset database for variable modifications
 or homology search, 156, 204
Summarize valid results, 30
summarize valid results, 41 to 46
summary
 database, 202
 identification statistics, 103
 parameters used, 159
summary mode, 69, 90 to 98

system administration

- adding instrument types, [213](#)
- calculating statistics, [208](#)
- changing biology/chemistry options, [213](#)
- changing categories of proteins, [213](#)
- changing cysteine modification options, [213](#)
- changing elements, [213](#)
- changing enzyme cleavage rules, [213](#)
- changing HTML links from search results, [212](#)
- changing N- or C-terminal groups, [213](#)
- changing species filter, [213](#)
- comparing two databases, [207](#)
- concatenating FASTA files, [203](#)
- configuring server drives, [216](#)
- connection time-outs, [216](#)
- creating a category file, [210](#)
- creating indices for databases, [198](#), [200](#)
- creating species or molecular weight subset databases, [201](#)
- creating subset databases from saved hits, [204](#)
- creating user (proprietary) databases, [197](#), [206](#)
- customizing amino acid modifications, [211](#)
- enabling HTML link from search results to BLAST, [212](#)
- generating database summary report, [202](#)
- installing or updating databases, [196](#)
- limiting number of parallel workflow processes, [218](#)
- maintaining server performance, [215](#)
- making a subset FASTA file from Accession Numbers, [205](#)
- reinstalling Spectrum Mill workbench, [217](#)
- removing redundant database entries, [209](#)
- server administration scripts, [215](#)
- settings for large data sets, [216](#)
- starting and stopping the Spectrum Mill

Workflow Manager Service, [219](#)

T

terminate process, [154](#)
time-outs, connection, [216](#)
TMT data
 Data Extractor, [133](#)
 MS/MS Search, [135](#)
 overview, [131 to 132](#)
 Protein/Peptide Summary, [140](#)
Tool Belt utilities, [153 to 174](#)

U

unknown modification search, [36](#)
unzip files, [174](#)
updating databases, [196](#)
user-created databases, [197](#), [206](#)
utilities
 protein and peptide, [176 to 193](#)
 Tool Belt, [153 to 174](#)

V

valid results
 print, [46](#)
 summarize, [41 to 46](#)
validation
 automatic, [25 to 29](#), [61 to 68](#)
 manual, [76 to 80](#)
 spectra, [37 to 40](#)
variable modifications search, [20 to 25](#), [35 to 36](#)
Variable modifications sites, [32](#)
view request queue, [19](#), [117](#), [119](#)

W

workflow automation, [111 to 120](#)
 completion log, [120](#)
 link to results, [117](#)
 monitor results, [117](#)
 view request queue, [117](#), [119](#)
 Workflow Manager Service, [219](#)
workflow tasks, [112](#)

workflows

- automation, [111 to 120](#)
- basic, [13 to 33](#)
- iterative, [34 to 55](#)
- parameter files, [112](#)
- tasks, [112](#)

Z

zip files, [174](#)

In this Book

The Application Guide presents basic instructions to use the Spectrum Mill MS Proteomics Workbench. In this guide you learn:

- How to process MS/MS and MS-only data sets
- How to automate Spectrum Mill workflows
- How to perform Sherenga *de novo* sequencing
- How to process data for differential expression quantitation
- How to use the software's protein/peptide analysis tools and utilities
- System administration details

© Agilent Technologies, Inc. 2016

Revision B, October 2016



G2721-90049



Agilent Technologies