

Agilent G2721AA/G2733AA Spectrum Mill MS Proteomics Workbench

Quick Start Guide

What is the Spectrum Mill MS Proteomics Workbench?	2
What's New in Version B.04.01?	3
Where to Find More Information	4
Setting Up Spectrum Mill Server and Client	6
Transferring Spectral Files to Spectrum Mill Server	9
Starting the Spectrum Mill Software	14
Using Basic and Iterative Workflows to Process Data	16
Running Workflow Automation or Interactive Automation	21
Quick Reference	24
Spectrum Mill Modules and Their Purposes	24
Interpreting Scores from MS/MS Search	26
Protein/Peptide Summary Modes	28

Use this guide for your first steps with the Spectrum Mill workbench.



Agilent Technologies

What is the Spectrum Mill MS Proteomics Workbench?

The Spectrum Mill software is a collection of tools for high-throughput processing of MS/MS and MS spectra to provide protein and peptide identifications and relative quantitation. The Spectrum Mill workbench can do in minutes what could take you hours or days to do manually.

This software includes bioinformatics tools for:

- Spectral preprocessing for MS/MS and MS-only data, including spectral quality filtering and precursor charge state assignment for MS/MS spectra
- Protein database search, including searches for both MS/MS and MS-only spectra
- Data review, validation, and comparison for large data sets, including those from two-dimensional LC/MS/MS analyses and differential expression studies
- *De novo* sequencing
- Identification of mutations, post-translational modifications, and chemical modifications
- Differential expression quantitation

The Spectrum Mill workbench accepts Agilent data, as well as data from other vendors' mass spectrometers. The software executes on a web server, and you access the program via a web browser window.

The software also includes process automation tools to automate the preprocessing, search, validation and review steps.

What's New in Version B.04.01?

- MS/MS Searches now support multiple CPU cores for faster searches.
- Data extractors for Agilent Q-TOF (.d) and Thermo (.raw) can extract to mzXML format, instead of just multiple pkl files.
- Peptide Selector support to build Agilent Q-TOF target lists based on search results.
- Improvements to Protein/Peptide Summary Peptide Mode, including more ways to filter to distinct Peptides, and ability to create an inclusion list of top N peptides per protein for Agilent Q-TOF.
- Protein-Protein Comparison Columns enhanced for VML reporting and Excel export.
- Export of Protein-Protein Comparison summary results to Mass Profiler Professional.
- Support for Thermo Q-Exactive instrument.
- Autovalidation Auto thresholds strategy supports a minimum sequence filter.
- Support for newer TMT6 reagents with modified 127 and 129 reporter ion masses.
- Tool Belt has enhanced support for quality metrics.
- New utility to remove duplicate entries in a Protein Database.
- More filtering options in the Build TIC form.
- Internet Explorer 8.0 and higher supported.
- Java Runtime Environment Version 7 Update 51 provided on installation disc.
- Windows XP is no longer supported.

For more details on these new features or the features that were new for Version B.04.00, see the online Help.

Where to Find More Information

Scientists: Check out these guides and online help

Where to Find More Information

Scientists: Check out these guides and online help

Familiarization Guide Follow step-by-step instructions to process example data.

Application Guide Learn details to use the software.

Online Help Consult the online help for in-depth information not given in the *Application Guide*. To access help, click links on the home page, **Help** buttons on Spectrum Mill pages, or blue dividing bars on Spectrum Mill pages.

System Administrators: Check out these guides and online help

Installation Guide Use this guide to install the software on the server.

Application Guide See the following chapters:

- **Chapter 9: System Administration**

Get an overview to install databases and do other system administration tasks.

- **Chapter 10: Files Created during Spectrum Mill Data Processing**

Refer to this chapter to troubleshoot data processing, to selectively remove parts of the processing, or to decide which files to archive.

Online Help From any **Help** page, click links under **For System Administrators**:

- **Protein Databases** (link to `millhtml\SM_instruct\faman.htm`)

Learn details to install databases, create indices, and create subset databases.

- **Server Administration** (link to `millhtml\SM_instruct\servadm.htm`)

Learn details to do other system administration tasks.

Contact Agilent Technical Support

Internet, phone and e-mail

Internet: In the upper right corner of the Spectrum Mill home page, click **Support**.

Phone: For Agilent worldwide sales and support center telephone numbers, go to <http://www.agilent.com/chem/contactus>. From the US and Canada, call 800-227-9770 (select options 3-2-2).

E-mail: For e-mail assistance, go to <http://www.agilent.com/chem/contactus>. Under **e-mail Assistance**, click **Support Related Inquiries**. Fill out the form and submit.

Software Status Bulletins and software patches

A list of known problems for the Spectrum Mill workbench, along with possible solutions, is found on the Agilent Web site.

- 1** Go to <http://www.chem.agilent.com/en-US/Technical-Support/Software-Informatics/Spectrum-Mill>. You can also:
 - a** Go to <http://www.agilent.com>.
 - b** Click **Technical Support**.
 - c** Under **Life Sciences & Chemical Analysis**, click **Software & Informatics**.
 - d** Under **Software & Informatics**, click **Data Systems And Software**.
 - e** Under **Products**, click **MassHunter Workstation Software**.
 - f** Under **Spectrum Mill - Technical Support**, click **DETAILS**.
- 2** For software status bulletin, click **Status Bulletins**. For software patches, click **Patches**.

Setting Up Spectrum Mill Server and Client

To set up the server

See the *Spectrum Mill MS Proteomics Workbench Installation Guide*. If you wish to update from a previous version of the Spectrum Mill workbench, see Chapter 2. Note that the server name cannot have an underscore.

To set up your client PC

Operating system

- ✓ Check that you have one of the following:
 - Windows Server 2008 R2
 - Windows 7
 - Windows Server 2003, SP2 or later

- 1 Select **Start >Run...**
- 2 In the **Open** box, type `winver`.

Both 32-bit and 64-bit versions of the above operating systems are supported.

Windows Server 2000 is no longer supported.

Browser

- ✓ Check that your web browser is Internet Explorer 8.0 or later. If you have Internet Explorer 9.0 or higher, check that the browser is running in Compatibility View.

- 1 Open Internet Explorer.

If your browser is IE 9.0 or higher, and you cannot see the menus you normally see in IE 8.0, right-click in the space next to the last tab and select Menu Bar. The menus now appear in the upper left of Internet Explorer.

Or

Press the **Alt** key to temporarily see the menus.

- 2 Select **Help > About Internet Explorer**.

Java Runtime Environment (JRE)

- ✓ Check that your computer is running Java Runtime Environment Version 7 Update 51 or higher.
 - 1 Select **Start > Control Panel > Java**.
 - 2 In the **General** tab, click **About**.

If needed, install or update Java from your Spectrum Mill installation disc. Multiple versions of Java can exist on your computer, but if you do not require a specific version for other software, remove the current version, then install the updated version.

Both 32-bit (i586) and 64-bit (x64) versions of Java are provided. For 32-bit operating systems, install the 32-bit version. For 64-bit operating systems, install both the 32-bit and the 64-bit versions to allow use of the proper Java with either version of Internet Explorer. (On 64-bit Windows, both 32-bit and 64-bit versions of Internet Explorer are installed.)

Be sure to install Java Runtime Environment on all clients.

Screen resolution

If supported on your monitor, use the **Control Panel** to set screen resolution to **1280 x 1024** or higher (preferably), with **True Color (32 bit)**.

- 1 In **Control Panel**, double-click **Display**.
- 2 Click the **Settings** tab.
- 3 Adjust the settings if necessary.

Stored pages

Disable caching.

- 1 Select **Tools > Internet Options...**
- 2 Click the **General** tab.
- 3 Under **Browsing history**, click **Settings...**
- 4 Click **Every time I visit the webpage**.

Setting Up Spectrum Mill Server and Client

To set up your client PC

Cookies

Enable cookies so you can save page settings.

- 1 Select **Tools > Internet Options...**
- 2 Click the **Privacy** tab and set to **Low**, or click the **Advanced** button and enable cookies.
- 3 For more detailed instructions to enable cookies, search Internet Explorer help for **cookies**.
- 4 If the drop-down menus in the Spectrum Mill workbench appear empty, you may need to add the Spectrum Mill server as a trusted site. Search Internet Explorer help for **trusted site**.

Active scripting

Enable JavaScript (Active Scripting).

- 1 Select **Tools > Internet Options...**
- 2 Click the **Security** tab.
- 3 Select **Local Intranet**.
- 4 Click **Custom Level...**
- 5 In the **Security Settings** dialog, scroll down until you see the **Scripting** section.
- 6 Under **Active Scripting**, click the **Enable** option.

Browser text size

For lower-resolution monitors (such as 1024 x 780), set a smaller text size to avoid excessive scrolling.

- 1 Select **Page > Text Size**.
- 2 Set to **Smaller**.

Transferring Spectral Files to Spectrum Mill Server

Before opening the Spectrum Mill Home Page, transfer your data files to the Spectrum Mill server. Most data files require no special preparation; you merely move or copy the files into the appropriate directory on the Spectrum Mill server, as described below. Some other types of data files must be preprocessed on the manufacturer's data system to generate generic peak list files. You then copy or move peak list files as described below.

To set up a useful directory structure

To make it easy to compare data sets, it is important that you set up the appropriate directory structure for your files on the Spectrum Mill server. Whenever you want to compare samples in a set, you need to set up a subdirectory for each sample. This sample directory may contain data files from multiple sample fractions or gel slices. Here are some examples:

- If you run a two-dimensional LC/MS/MS analysis of a single sample, transfer all the files to a single directory on the server.
- If you run the same two-dimensional LC/MS/MS analysis on a second sample, or if you repeat the run on the first sample, transfer all these files to a second directory.
- For differential expression quantitation experiments where you use isotopic labels, transfer all the files to a single directory on the server. This is because all of the versions of the peptide (light, medium, heavy) are found in the same file or sets of files.
- If you conduct a differential expression study where you use the absolute abundances to do the quantitation, transfer samples from one cell state into one directory and the second cell state into a second directory.

To access the server directory **msdataSM**

When you transfer data files you move them to the **msdataSM** directory in the Data folder on the Spectrum Mill server.

Access the folder from the server

- In your Spectrum Mill file system on the server, navigate to the directory **msdataSM**.

Transferring Spectral Files to Spectrum Mill Server

To transfer MS/MS data files

This directory is often on the Data drive under the Spectrum Mill folder. (If you don't know how to find your Spectrum Mill file system, ask the person who installed your software.)

Access the folder from a PC client

- 1 Make sure your administrator has shared the **msdataSM** folder and given you permissions. See *Installation Guide*.
- 2 Map the shared folder, \\servername\msdataSM, to a client PC. (Right-click **Computer** and select **Map Network Drive**. Follow the instructions.)

CAUTION

Make sure that the folders on the server are shared with the use of the Computer Management Console. If they are shared any other way, permission settings will be set incorrectly.

To transfer MS/MS data files

To transfer Agilent Q-TOF and ion trap, Thermo Fisher Scientific ion trap, or Applied Biosystems/MDS Sciex QSTAR data

- 1 In the directory **msdataSM**, create a new subdirectory for your sample files. Name it anything you want, but do not use spaces or any of these special characters in the name:

| () , ; % < > + ?

See “[To access the server directory msdataSM](#)” on page 9.

If you have multiple samples you wish to compare, remember to set up a subdirectory for each sample. Ensure that each subdirectory contains all fractions for a single sample.

You may add up to ten directory levels between **msdataSM** and your sample files, but shorter path lengths reduce memory usage and speed processing of large data sets.

- 2 Copy or move your *.d, *.raw, or *.wiff files to your new subdirectory or subdirectories.
- 3 After you transfer data files to the server, remove any spaces or parentheses in the data file names. Delete them or replace them with other characters, such as underscores.
- 4 Make sure you have both read and write permissions for each subdirectory that contains your sample files. (To check, right-click the folder and select

Properties.) Make sure all user groups have full permissions. (Check the **Security** tab.)

To prepare and transfer generic peak list files (such as Waters Micromass Q-ToF or MALDI-TOF-TOF data)

- 1 Prepare a generic peak list file.
 - For Waters Micromass Q-ToF, export centroided data with the Waters Micromass ProteinLynx package. The results are individual *.pkl or *.dta spectral files, or appended *.pkl files that contain multiple spectra.
 - For MALDI-TOF-TOF, export *.mgf files.
- 2 In the directory **msdataSM**, create a new subdirectory for your sample files. Name it anything you want, but do not use spaces or parentheses in the name. (See “[To access the server directory msdataSM](#)” on page 9.)

If you have multiple samples you wish to compare, remember to set up a subdirectory for each sample. Ensure that each subdirectory contains all fractions for a single sample.

You may add up to ten directory levels between **msdataSM** and your sample files, but shorter path lengths reduce memory usage and speed processing of large data sets.
- 3 If you have *.mgf files or appended *.pkl files, where each contain multiple spectra, copy or move them to your new subdirectory or subdirectories. Proceed to [step 5](#). If you have individual *.pkl or *.dta spectral files, proceed to [step 4](#).
- 4 *Under* your new subdirectory, create another subdirectory called **cpick_in** and copy or move your files there. Note that you *must* use this subdirectory name!
- 5 After you transfer data files to the server, remove any spaces or parentheses in the data file names. Delete them or replace them with other characters, such as underscores.
- 6 Make sure you have both read and write permissions for each subdirectory that contains your sample files. (To check, right-click the folder and select **Properties.**) Make sure all user groups have full permissions. (Check the **Security** tab.)

To transfer MS-only files

NOTE

For Agilent TOF .d files and Q-TOF data in MS-only mode, you do *not* follow the workflow described below for Agilent TOF .wiff files or for other instrument files. You first generate a mass list with the Molecular Feature Extraction algorithm within the MassHunter Workstation Qualitative Analysis software, then copy the mass list from the compound table into the Spectrum Mill Manual PMF page, described in the *Application Guide*. See also Chapter 5 of the *Familiarization Guide* for exercises to learn how to search MS-only data.

To transfer Agilent TOF, or Applied Biosystems/MDS Sciex QSTAR o-MALDI .wiff files

- 1 In the directory **msdataSM**, create a new subdirectory for your sample files. Name it anything you want, but do not use spaces or parentheses in the name. (See “[To access the server directory msdataSM](#)” on page 9.)

If you have multiple samples you wish to compare, remember to set up a subdirectory for each sample. Ensure that each subdirectory contains all fractions for a single sample.

You may add up to ten directory levels between **msdataSM** and your sample files, but shorter path lengths reduce memory usage and speed processing of large data sets.

- 2 Copy or move your *.wiff files to your new subdirectory or subdirectories. Spectrum Mill does not support Agilent TOF and Q-TOF .d files for use with PMF search.
- 3 After you transfer data files to the server, remove any spaces or parentheses in the data file names. Delete them or replace them with other characters, such as underscores.
- 4 Make sure you have both read and write permissions for each subdirectory that contains your sample files. (To check, right-click the folder and select **Properties**.) Make sure all user groups have full permissions. (Check the **Security** tab.)
- 5 Make sure that *.wiff files are *not* marked read-only.

To prepare and transfer other MS-only data types

- 1 Use your instrument data system to export calibrated, centroided, de-isotoped data as peak list files. Next you will transfer the peak lists to the Spectrum Mill server.
- 2 In the directory **msdataSM**, create a new subdirectory for your sample files. Name it anything you want, but do not use spaces or parentheses in the name. (See “[To access the server directory msdataSM](#)” on page 9)

You may add up to ten directory levels between **msdataSM** and your sample files, but shorter path lengths reduce memory usage and speed processing of large data sets.
- 3 *Under* your new subdirectory, create another subdirectory called **fit_batch_in** and copy or move your files there. Note that you *must* use this subdirectory name!
- 4 After you transfer data files to the server, remove any spaces or parentheses in the data file names. Delete them or replace them with other characters, such as underscores.
- 5 Make sure you have both read and write permissions for each subdirectory that contains your sample files. (To check, right-click the folder and select **Properties**.) Make sure all user groups have full permissions. (Check the **Security** tab.)

Starting the Spectrum Mill Software

Start the software from the server

- 1 Double-click the desktop icon to launch the Spectrum Mill workbench.

The icon is created during installation.

- 2 Ensure that you see the Spectrum Mill home page (Figure 1).



Start the software from the PC client

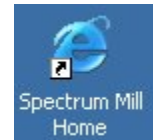
- 1 Double-click the desktop icon to launch the Spectrum Mill workbench.

- a If you do not see this icon, start your Internet Explorer web browser on the client PC.

- b In your browser window, type the address of the Spectrum Mill home page.

This has the general format **server/millhome.htm**. (If you don't know the URL, ask the person who installed the software or your system administrator.)

- 2 Ensure that you see the Spectrum Mill home page (Figure 1).



NOTE

If the Spectrum Mill Home shortcut is not on a client and you want to create it, do this:

- a Right-click the title bar on the home page, and select **Create Shortcut**.
- b Rename the Desktop icon to **Spectrum Mill Home**.

Or follow the instructions in the Tips and Techniques section of the Online Help entitled "To Create a Shortcut to the Spectrum Mill Workbench on Your Desktop".

Agilent Technologies **Spectrum Mill MS Proteomics Workbench** Rev B.04.01.141 Agilent Proteomics Support

Process Automation Tools

- Workflows** Automate data extraction, search, validation, and summary.
- Completion Log** View log of completed extractions, searches, validations, and summaries
- Request Queue** View status of request queue for submitted requests, monitor results output

Mass Spectral Interpretation Tools

- Data Extractor** Prepare MS/MS data files for Spectrum Mill processing
- MS/MS Search** Search database with MS/MS spectra
- Autovalidation** Autovalidate MS/MS spectra, calculate false-positive rates
- Spectrum Matcher** Match MS/MS spectra against other MS/MS spectra
- de novo Sequencing** Perform Sherenga *de novo* MS/MS spectral interpretation
- PMF Search** Search database with MS spectra from Peptide Mass Fingerprint data

Result Summary Tools

- Protein/Peptide Summary** Summarize results from MS/MS searches
- Spectrum Summary** Summarize characteristics of MS/MS spectra
- de novo Summary** Summarize results of Sherenga *de novo* MS/MS spectral interpretations
- PMF Summary** Summarize results from PMF searches

Utilities

- Tool Belt** Collection of Spectrum Mill utility tools
- Protein Databases** Manipulate FASTA sequence databases for use with Spectrum Mill software
- Build TIC** Display a TIC or neutral loss chromatogram
- Peptide Selector** Select peptides from protein digest likely to produce high quality MS/MS spectra
- MRM Selector** Build/Export MRM transition lists based on experimental data in Spectrum Mill
- Multiple Sequence Aligner** Run Clustal W and highlight non-identical AA's across the alignment
- MS Digest** Predict peptide masses for enzymatic digestion of protein
- MS Edman** Search database with text or partial peptide sequence
- MS Product** Predict product ion masses from peptide sequence
- MS Comp** Calculate AA compositions fitting parent mass and immonium ions
- MS Isotope** Display isotope patterns of peptides
- Peptide List to Masses** Convert a list of peptides to masses and formulas

Help:

- Spectrum Mill Basics
- MS/MS Search
- Workflow Automation
- PMF Search
- Protein Databases
- Server Administration
- Tips and Tricks
- FAQs
- What's New

Useful Tables:

- Mutation Mass Shifts
- Dipeptide Masses

Manuals:

- Quick Start Guide
- Application Guide
- Familiarization Guide

View End User License Agreement
 © Copyright 2003-2014 Agilent Technologies, Inc.
 © Copyright 1997-2004 Millennium Pharmaceuticals, Inc.
 © Copyright 1995-1999 The Regents of the University of California
 © Copyright 2006-2014 Broad Institute of MIT and Harvard
 All rights reserved.

Figure 1 Spectrum Mill home page

NOTE

For help on the home page, click a topic on the left side of the page.

Using Basic and Iterative Workflows to Process Data

In this section you learn about the basic and iterative workflows to process multiple MS/MS data files with the Spectrum Mill MS Proteomics Workbench. Both workflows use tasks set up in Spectrum Mill modules ([Figure 1](#)). For an overview of the purpose of each module and where more information can be found for each one, see “[Spectrum Mill Modules and Their Purposes](#)” on page 24.

[Figure 2](#) shows two road maps for processing MS/MS data. For some samples and studies, it is sufficient to process only through a basic workflow, which reliably identifies most proteins in a sample with one round of processing. Detailed steps are covered in “[Basic Workflow – Setting Up Parameter Files and Processing Tasks Interactively](#)” in Chapter 1 of the *Application Guide*.

For others, you take advantage of an iterative workflow, which iterates through multiple rounds of database search and results validation, with the goal of gathering as much information as possible to better characterize the identified proteins.

The Spectrum Mill workbench provides a means to segregate search results that contain a valid interpretation of an MS/MS spectrum from those that do not. Results that are not validated can then be subjected to subsequent rounds of searches (against other databases or in variable modifications mode, for example). Results that are validated can be summarized in a results table. These alternative steps are covered in “[Iterative Workflow – Alternative Interactive Processing](#)” in Chapter 1 of the *Application Guide*.

Both workflows are summarized in “[MS/MS Workflows Summary](#)” on page 18. While the summary steps described here are representative of what many users want to do, there are certainly other paths through the software. The manual set and the online help give you the details you need to customize the workflows.

To use Spectrum Mill workbench for an MS-only workflow (.wiff files only) with a PMF search, see Chapter 6 of the *Application Guide*.

CAUTION

The default settings for the Spectrum Mill workbench have been carefully optimized. In general, you should change settings only for the settings that are highlighted in red text. If you change other settings, be sure you are confident in your choices. If you wish to restore default settings, see the online help for the Spectrum Mill page of interest.

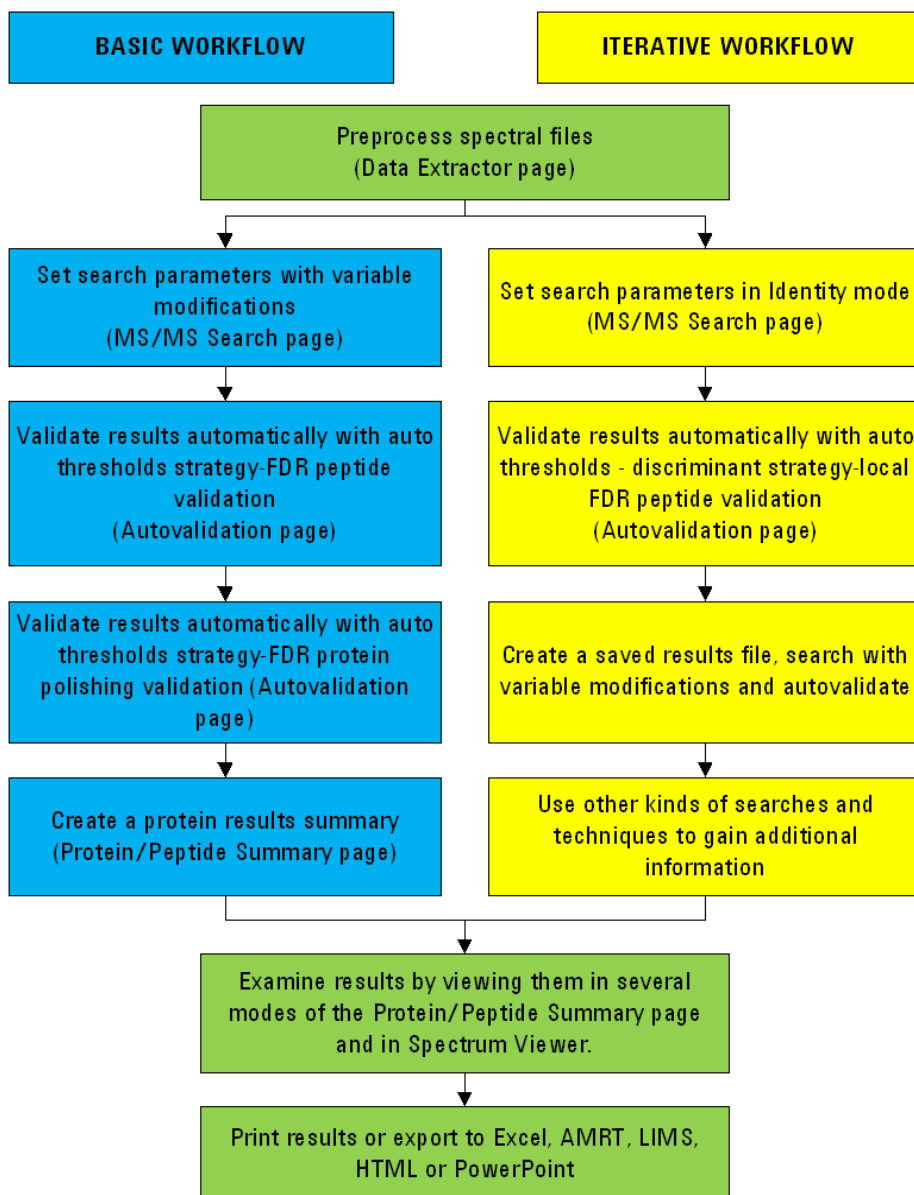


Figure 2 FDR-based workflows to process MS/MS data

MS/MS Workflows Summary

Figure 2 and the summary below describe the basic and iterative workflows based on False Discovery Rate (FDR), recommended for use with the Spectrum Mill workbench.

While the following instructions summarize the overall workflows, the step-by-step procedures are given in Chapter 1 of the *Application Guide*.

Before you begin processing make sure you have set up a directory structure on the server that makes it easy to summarize and compare your results and you have copied or moved the raw MS/MS data files to the Spectrum Mill server. See “[Transferring Spectral Files to Spectrum Mill Server](#)” on page 9.

Basic Workflow

The basic workflow assumes that the goal is to reliably identify most proteins in the sample in only one round. To do this, you search first with variable modifications you think are present in your sample. Yet, if you search with too many modifications, the search runs too slowly. For more details on choosing variable modifications, see “[Basic Workflow – Setting Up Parameter Files and Processing Tasks Interactively](#)” in Chapter 1 of the *Application Guide*.

- 1 Extract the data. (Data Extractor page)
 - Set the correct cysteine modification.
 - For **Merge nearby MSⁿ scans with same precursor m/z**, if necessary change the time range to be compatible with your chromatographic data. A good starting point is 30 seconds with a +/- 1.4 *m/z* window for Agilent Q-TOF spectra, if the General MS/MS Merging Constraint selected includes Spectral Similarity. If not, a .05 *m/z* window is better.
- 2 Search a database in **Variable Modifications** mode, preferably a species subset database for the first search. Add modifications you think are present in your sample. (MS/MS Search page)
- 3 Autovalidate the matches at a specified FDR, first in **Auto thresholds** strategy, **Peptide** mode, then (optionally) in **Auto thresholds** strategy, **Protein Polishing** mode. Use default settings or those in Chapter 1 of the *Familiarization Guide*. (Autovalidation page)
 - Validation means that you accept the proposed database match as the correct match for the MS/MS spectrum.
- 4 Summarize and view the results (Protein/Peptide Summary page).

NOTE

If you are unsure of which modifications to search, a mass gap search can help. This search shows mass shifts of the modified peptides from the expected unmodified peptides to help you identify modifications. Use a mass gap search in an iterative workflow after an identity search and autovalidation. See “[Iterative Workflow](#)” on page 19.

Once you identify the modifications you can use them as variable modifications in a basic workflow.

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

Iterative Workflow

If your study requires you to characterize the identified proteins more extensively, use an iterative workflow. Which 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. See Chapter 1 of the *Application Guide* and the instructions below for the possible steps you can use in an iterative workflow.

- 1 Extract the data. (Data Extractor page)
See “[Extract the data. \(Data Extractor page\)](#)” on page 18 for details.
- 2 Search a database in **Identity** mode, preferably a species subset database for the first search. Enable Dynamic peak thresholding and Discriminant scoring.
- 3 Autovalidate the results with the highest scores and account for false positives in **Auto thresholds-discriminant** strategy, **Peptide** mode. Select Local FDR. (Autovalidation page)
- 4 Use Tool Belt to create a saved results file of validated protein hits (*.res file).
 - Be sure to indicate the database you searched since this maps to accession numbers.
- 5 Search the spectra that have not been validated in **Variable Modifications** mode against the saved results file. Combine the following modifications into a single search: oxidized methionine and pyroglutamic acid (N-terminal); phosphorylated S, phosphorylated T, phosphorylated Y; and any other modifications you suspect for your sample.

To determine which variable modifications to use if you're not sure, do a mass gap search in step 5 of the iterative workflow to identify unknown modifications, then use the identified variable modifications in either kind of workflow. See Chapter 3 of the *Familiarization Guide* for exercises to help you do mass gap searches.

- 6 Autovalidate once more as in step 3, and review results if you choose.
- 7 Use other kinds of searches and techniques to gain additional information.
 - a Search in no enzyme (or half-enzyme) mode against previous validated results to find non-specific cleavages for proteins you have already identified. (Set **Digest:** to **No enzyme.**) Repeat autovalidation.
 - b When you think you are done, list sequence-not-validated spectra in Protein Summary Details mode and look for proteins with multiple peptides. These may represent legitimate proteins at low levels. Re-examine the spectra to confirm.
 - c Optional: Search again using a larger database (entire database or larger subset). This is most important when the original subspecies is not well-represented in the database. Autovalidate again.
 - d Check statistics in Tool Belt. If there is a significant number of unmatched filtered spectra, continue searching.

NOTE

If the Tool Belt statistics show that only a small percentage of your collected MS/MS spectra were filtered, check your data file to see if you collected a large number of MS/MS spectra on background. If so, your MS/MS data acquisition threshold may be set too low.

- e Use Spectrum Summary to check for sequence-not-validated spectra with sequence tags greater than 6 or 7. Mark each **Good Spectrum** as appropriate.
- f Subject the good spectra to *de novo* sequencing.
- g When you have gained enough information from your data, summarize the results.

You can also manually interpret, review and validate the data if you wish to mine the data more deeply or to confirm autovalidation results.

NOTE

FDR results are not valid after manual validation. For this reason Agilent recommends you complete your final autovalidation and review, and check the Tool Belt FDR statistics before doing a manual validation. See Chapter 2 of the *Application Guide* for instructions on manual validation.

Running Workflow Automation or Interactive Automation

This section gives introductory background for setting up and executing workflow automation and for running interactive automation.

Setting up and executing workflow automation

Spectrum Mill B.04.00 introduced workflow automation, which lets you place a series of tasks into a queue; then the Spectrum Mill workbench executes them in the proper order. For example, you can automate a typical workflow for MS/MS data files from protein digests:

- Data Extractor
- MS/MS Search
- Autovalidation
- Protein/Peptide Summary

Or you can automate a typical MS-only workflow:

- Data Extractor
- PMF Search
- PMF Summary

Workflow automation uses a Workflows page, an Edit Workflow page, and a Request Queue/Completion Log viewer. You access these pages through the Process Automation Tools section of the Spectrum Mill home page (Figure 1). See Chapter 3 of the *Application Guide* for instructions on how to set up and run automated workflows.

Set up workflow automation with parameter files

To create a workflow within the Edit Workflow page, you select task parameter files in the order you want executed during the workflow. A *parameter file* specifies the settings for each task. The first step to set up a workflow is to save the settings you want to use in parameter files.

Most Spectrum Mill pages let you save and load parameter files associated with the page. For example, Data Extractor, MS/MS Search (and PMF Search for MS data), Autovalidation, and Protein/Peptide Summary (and PMF

Summary for MS data) include buttons to save settings in parameter files, and to load settings from parameter files. See Chapters 1 and 6 of the *Application Guide* for instructions on how to set up parameter files.

For nearly all the pages, you use the parameter files, whose names are called tasks in the Workflows page, to build and execute workflows. Peptide Selector, MRM Selector, and Sherenga **de novo** Summary are exceptions. For these programs, it is convenient to save parameter files, but you do not use them in an automated workflow.

The Workflows and Edit Workflow pages let you view a task's page in read-only mode, with the parameters shown for the parameter file you select.

Execute multiple data folders in parallel

Workflow automation lets the Spectrum Mill workbench execute an ordered list of tasks for one or more data folders. All data folders must contain the same type of data (for example, all .raw, all .d, etc.). When Spectrum Mill workbench executes a workflow, it adds the tasks to a Request Queue, and executes them in order for a particular data folder.

After you select multiple data file folders and start the workflow, the Spectrum Mill workbench first does the Spectral Extraction and MS/MS Search on each of them independently, or if you have a dual-core or quad-core CPU, it processes them in parallel. Then the program does the Autovalidation and Protein/Peptide Summary independently, or on the combination of the directories if you grouped proteins across multiple directories in the Autovalidation page.

Use the Request Queue/Completion Log Viewer

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

The Completion Log Viewer shows a list of all tasks that have completed, with most recent shown at the top. This log includes all queued requests, whether you queued them interactively or via a workflow. Two of the columns let you view additional information.

Running interactive automation

In the *Application Guide* see Chapter 1 for instructions on processing MS/MS data interactively, and Chapter 6 for MS data.

Interactively place tasks in the queue

The request queue concept extends to execution from a particular task's page. For example, in the Data Extractor page, the Extract button submits the extraction to the Request Queue, as does the Search button in the MS/MS and PMF Search pages.

Pages where interactive execution is helpful – Autovalidation and P/P Summary for MS/MS data and PMF Summary for MS-only data – have a Queue request check box that is cleared by default. If you do not mark the check box, you see the results in the current window, as in earlier Spectrum Mill versions. If you mark the Queue request check box, the program puts the request into the queue, and saves the results output to an HTML file that you can view later.

Running interactive automation

You can execute portions of an MS/MS or MS workflow one step at a time, by executing requests from each task's page. For example, you may:

- 1 Use the Data Extractor page to select one or more data folders to be processed with a parameter file, and submit the request.
- 2 Go to the MS/MS Search page, where the current data folders are already selected, load a parameter file, then start the search.
- 3 Go to the Autovalidation page, select a parameter file and click the Queue request check box to queue the validation request for one or both autovalidation modes.
- 4 Go to Protein/Peptide Summary and queue one or more summary requests.

Quick Reference

Spectrum Mill Modules and Their Purposes

The table below gives the purpose for each of the Spectrum Mill modules available on the Home page.

Table 1 Spectrum Mill Modules and Their Purposes

Tool type	Module	Function	Application Guide Chapter(s)
Workflow automation	Workflows	Create, edit and execute workflows, which are a sequence of tasks (parameter files) for extractions, searches, validations and summaries.	3
	Completion Log	View log of completed extractions, searches, validations and summaries.	3
	Request Queue	View status of request queue for submitted requests; monitor results output.	3
Spectral preprocessing	Data Extractor	Extract spectra from data files. Processing depends on data type and may include averaging, centroiding, filtering by quality, assigning precursor charge, and calculating spectral features.	1, 6
Interpretation	MS/MS Search	Search MS/MS spectra against databases.	1, 6
	Autovalidation	Validate search hits based on various threshold strategies and specified false discovery rate cutoffs	1, 2
	Spectrum Matcher	Match MS/MS spectra against other MS/MS spectra.	1
	<i>de novo</i> Sequencing	Use Sherenga <i>de novo</i> sequencing to interpret MS/MS spectra.	1, 4
	PMF Search	Search MS-only spectra against databases.	6

Table 1 Spectrum Mill Modules and Their Purposes (continued)

Tool type	Module	Function	Application Guide Chapter(s)
Result summary	Protein/Peptide Summary	Summarize, review, filter, and validate results from MS/MS searches.	1, 2
	Spectrum Summary	Sort and classify MS/MS spectra based on spectral characteristics.	1
	<i>de novo</i> Summary	Summarize and review <i>de novo</i> interpretations.	1, 4
	PMF Summary	Summarize and review MS-only results.	6
Utilities	Tool Belt	Use tools to speed processing, examine parameters and search statistics, stop processes, etc.	7
	Protein Databases	Create indices for databases so searches run faster, create subset databases, etc.	9
	Build TIC	Visualize how much of your sample has been interpreted. Locate spectra that indicate potential peptide phosphorylation sites.	1
	Peptide Selector	Run theoretical digestion and then select peptides that meet certain criteria.	8
	MRM Selector	Build/Export MRM transition lists based on experimental data in Spectrum Mill.	8
	Multiple Sequence Aligner	Align amino acid sequences of proteins from a database and highlight the amino acids that differ	8
	MS Digest	Run theoretical enzymatic digestion and calculate masses of peptides that result.	8
	MS Edman	Search text fields in protein databases, with or without peptide mass filtering.	8
	MS Product	Calculate theoretical fragment ion masses from peptides.	8
	MS Comp	Calculate amino acid composition for a peptide, given peptide mass and immonium ions.	8
	MS Isotope	Calculate and view isotope patterns of peptides.	8
	Protein List to Masses	Calculates masses and formulas for a set of specified peptides	8

Interpreting Scores from MS/MS Search

The following definitions and guidelines will help you evaluate your MS/MS search results.

Score definitions

- Peptide score** Score of an individual peptide. This score reflects the information content (amount of useful fragmentation) in the MS/MS spectrum.
- Protein score** Score of the overall protein, calculated by adding scores of all the peptides detected for the protein
- Scores greater than 15 for Agilent Q-TOF and ion trap instruments almost always represent valid results.
- % SPI** Percentage of the extracted MS/MS ion current explained by theoretical fragmentation of the database hit

Guidelines for validating *Agilent Q-TOF* results based on MS/MS Search scores

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.

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

When you manually validate, apply the following filters in Protein/Peptide Summary:

- Peptide score > 5
- % SPI > 60

Guidelines for validating *ion trap* results based on MS/MS Search scores

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

Protein/Peptide Summary Modes

For more details, see “Protein/Peptide Summary Displays” in the *Application Guide*.

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	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
	Samples, then proteins, then peptides	Protein-Sample Centric Rows Details	Manual review and validation of MS/MS search results, organized by sample
Summarize results by proteins	Proteins only	Protein Summary	List of all proteins identified in the data
	Proteins, then samples	Protein-Protein Comparison Columns	Compare two or more samples, each of which may contain multiple fractions. Each sample is organized in a separate directory.
	Proteins, then redundant hits	Protein-Protein Comparison Redundant	Same as immediately above, with additional detail on isoforms of proteins
	Proteins, then peptides	Protein Summary Details	View proteins, with supporting peptide details
Summarize results by samples	Samples, then proteins	Protein-Sample Centric Rows	View proteins from multiple 2D gel spots organized in a single directory
	Samples, then proteins, then peptides	Protein-Sample Centric Rows Details	Same as immediately above, with supporting peptide details

Table 2 Protein/Peptide Summary Modes (continued)

If you want to:	And you want results organized by:	Then use this mode:	Example application
Summarize results by peptides	Peptides only	Peptide	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)
	Peptides, then samples	Protein-Peptide Comparison Columns	Evaluation of fractionation scheme (provides more information and easier export to Excel)
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

www.agilent.com

In This Book

The *Quick Start Guide* presents first steps to use the Spectrum Mill MS Proteomics Workbench.

Windows[®], Excel[®], PowerPoint[®], and Internet Explorer[®] are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

© Agilent Technologies, Inc. 2014

Printed in USA
Revision A, February 2014



G2721-90047



Agilent Technologies