Agilent G2721AA Spectrum Mill MS Proteomics Workbench Quick Start Guide

A guide to the Spectrum Mill workbench

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

What is the Spectrum Mill MS Proteomics Workbench?

The Spectrum Mill software is a collection of tools for high-throughput processing of MS and MS/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.

Agilent Technologies

New features in Spectrum Mill workbench version A.03.03

The following are key changes to the user interface and functionality:

Support for more instruments and data types

Now supported:

- Agilent Q-TOF
- Electron transfer dissociation (ETD) data from Agilent ion trap
- MALDI-TOF-TOF data (via *.mgf file, requires license for Generic Data Extractor)
- Thermo Fisher Scientific LTQ, LTQ FT, and LTQ Orbitrap, in various dissociation modes (requires a license for Thermo Data Extractor, as well as co-resident Thermo Fisher Scientific Xcalibur software that is equivalent to or newer than the version you used to acquire the data)

Data Extractor

Ability to disable similarity merging, to improve coverage and to detect more low-level peptides for simple samples (for example, a single protein)

MS/MS Search

Capability to search peptides that have a charge up to +7

New **Maximum ambiguous precursor charge**, which allows you to choose the maximum charge state that you want MS/MS Search to consider when the Data Extractor could not discern the charge state for the peptide

Capability to use protein pI to filter candidate proteins from the database

Improved scoring schemes, including **Dynamic peak thresholding**, which is a scoring enhancement that enables identification of more low-abundance and short-chain peptides

Explicit “mass gap” search that allows you to search for an unexpected or unknown modification

Ability to automatically create a file of previously validated hits from the MS/MS Search page, without using the Tool Belt page

MS/MS Autovalidation

Addition of +5 charge state to the autovalidation rules

Capability to perform autovalidation across data files in multiple subdirectories, with all peptides contributing to the protein score

Addition of peptide pI as a validation criterion for samples where the pI is known (for example, from offgel fractionation)

Protein/Peptide Summary	Spectrum Viewer support for c- and z-ions, which are prominent in ETD spectra
	Retention time in peptide summaries
	Fragmentation mode in peptide summaries, so you can ascertain the mode for each spectrum (for example, CID versus ETD)
	Enhanced protein grouping based upon shared peptides
	Ability to align sequences of proteins within a protein group
	Ability to sort and filter by peptide pI
	New capability in the Spectrum Viewer to type your own sequence and have it add to the list of available sequences that the software shows when you click the **Rank** arrow buttons
	Ability to sort peptides by the numerical position of their first amino acid in the protein sequence
	Capability to distinguish in results tables between true mass errors and mass differences produced by peptide modifications
Support for more amino acid modifications	Support for **Deamidated (N)** and **Deamidated (Q)** as variable modifications. The software displays the first by default; system administrators can add the second by copying a definition from **smconfig.misc.xml** to **smconfig.custom.xml**.
	Expanded iTRAQ support, to include data from the Generic Extractor and the *.raw (LCQ/LTQ) Extractor. The latter uses merged MS^3 spectra.
	Support for both original iTRAQ (which assumes complete labeling) and **iTRAQ Partial-mix**, which allows you to check for incomplete labeling
Miscellaneous	New Easy MS/MS Search form that simplifies use of the software and streamlines the data processing steps
	New Multiple Sequence Aligner utility to align sequences of proteins from a database and highlight amino acid differences
	Support for Web-based interaction with GeneSpring MS
	Easier, faster input of mass lists into Manual PMF, with MH^+/M selection that allows you to input either protonated or neutral species (directly compatible with output from a variety of instrument data systems)
	Faster selection of data folders
	Sherenga *de novo* sequencing support for the Agilent Q-TOF and ETD data from the Agilent ion trap

The ability to store iTRAQ correction factors for multiple iTRAQ batches and to apply them to the iTRAQ calculations for specific data sets

Expanded capabilities for Peptide Selector

Updated instrument selection lists

The default Spectrum Mill font size is now “smaller”, so you may use the default “medium” size in your browser and the Spectrum Mill software still uses “smaller.”

Spectrum Mill manuals and online help

Scientists

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 forms, or blue dividing bars on Spectrum Mill forms.

Quick Reference Card After you are familiar with the software, consult this card for an overview of the steps to process MS/MS data.

Thermo Ion Trap Data Extractor Quick Start Guide Get a quick overview of the optional data extractor for Thermo Fisher Scientific ion trap data files.

Generic Data Extractor Quick Start Guide Get a quick overview of the optional Generic Data Extractor for peak list files.

QSTAR Data Extractor Quick Start Guide Get a quick overview of the optional Applied Biosystems/MDS Sciex QSTAR Data Extractor.

System Administrators

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 perform 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\servadmn.htm**)
  Learn details to perform other system administration tasks.

Setting up the Spectrum Mill 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.

Setting up your client PC

Operating system

Check that you have one of the following:

- Windows® 2000 Server, SP4
- Windows Server 2003, SP1
- Windows XP, SP2

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

NOTE

At the time of release, Windows Vista is not yet supported. If you wish to use Vista, please contact Agilent to check the support status of Vista.

Browser

Check that your web browser is Internet Explorer 6.0 (IE 6.0) or Internet Explorer 7.0 (IE 7.0).

1. Open Internet Explorer.
2. Select **Help > About Internet Explorer**.

Java

Check that Internet Explorer on the client is configured for Sun JRE 1.5.x (also called J2SE 5.0) or Java SE 6. You can get a compatible version of JRE from your Spectrum Mill installation CD.

1. With IE 6.0 or IE 7.0, select **Tools > Internet Options...**.
2. Click the **Advanced** tab.
3. Scroll down to the **Java (Sun)** section to view the version.

Screen resolution

If supported on your monitor, use the **Control Panel** to set screen resolution to **1280 x 1024**, 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. With IE 6.0 or IE 7.0, select **Tools > Internet Options...**.
2. Click the **General** tab.
3. Do one of the following:
   - If you have IE 6.0, under **Temporary internet files**, click the **Settings...** button. Then click **Every visit to the page**.
   - If you have IE 7.0, under **Browsing history**, click the **Settings...** button. Then click **Every time I visit the webpage**.

Cookies

Enable cookies so you can save form settings.

1. With IE 6.0 or IE 7.0, 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. With IE 6.0 or IE 7.0, select **Tools > Internet Options...**.
2. Click the **Security** tab.
3. Select **Local Intranet**.
4. Click the **Custom Level...** button.
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

Especially for lower-resolution monitors (e.g., 1024 x 780), set a smaller text size to avoid excessive scrolling.

1. Do one of the following:
   - With IE 6.0, select **View > Text Size**.
   - With IE 7.0, select **Page > Text Size**.
2. Set to **Smaller**.

Running the Spectrum Mill workbench from the client

Open Internet Explorer and point to the URL for the Spectrum Mill server. This has the general format **server/millhome.htm**. If you are not sure of the URL, ask your system administrator.

Setting up Agilent TOF analyses for use with the Spectrum Mill workbench

When you set up data acquisition parameters on the Agilent TOF, it is critical that you establish analysis settings that are compatible with Spectrum Mill processing. For details, see the chapter on processing MS-only data in the *Spectrum Mill MS Proteomics Workbench Application Guide*.

Software Status Bulletin 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.agilent.com/chem/.
2. On the left, click **Technical Support**.
3. Under **Downloads and Utilities**, click **Status Bulletins and Patches**.
4. On the **Software Status Bulletins and Patches** page, click **Spectrum Mill**.
5. If you are prompted to log in, and you are registered on the Agilent Web site, type your login name and password, then click **Enter**. If you are not registered, click the link to register.
6. On the **Software Status Bulletins and Patches** page, click **Spectrum Mill**.
7. When prompted for your Spectrum Mill password, type `specmill` (all lower case) into the text box, then click **SUBMIT**.
8. View the Software Status Bulletin, or download a software update.

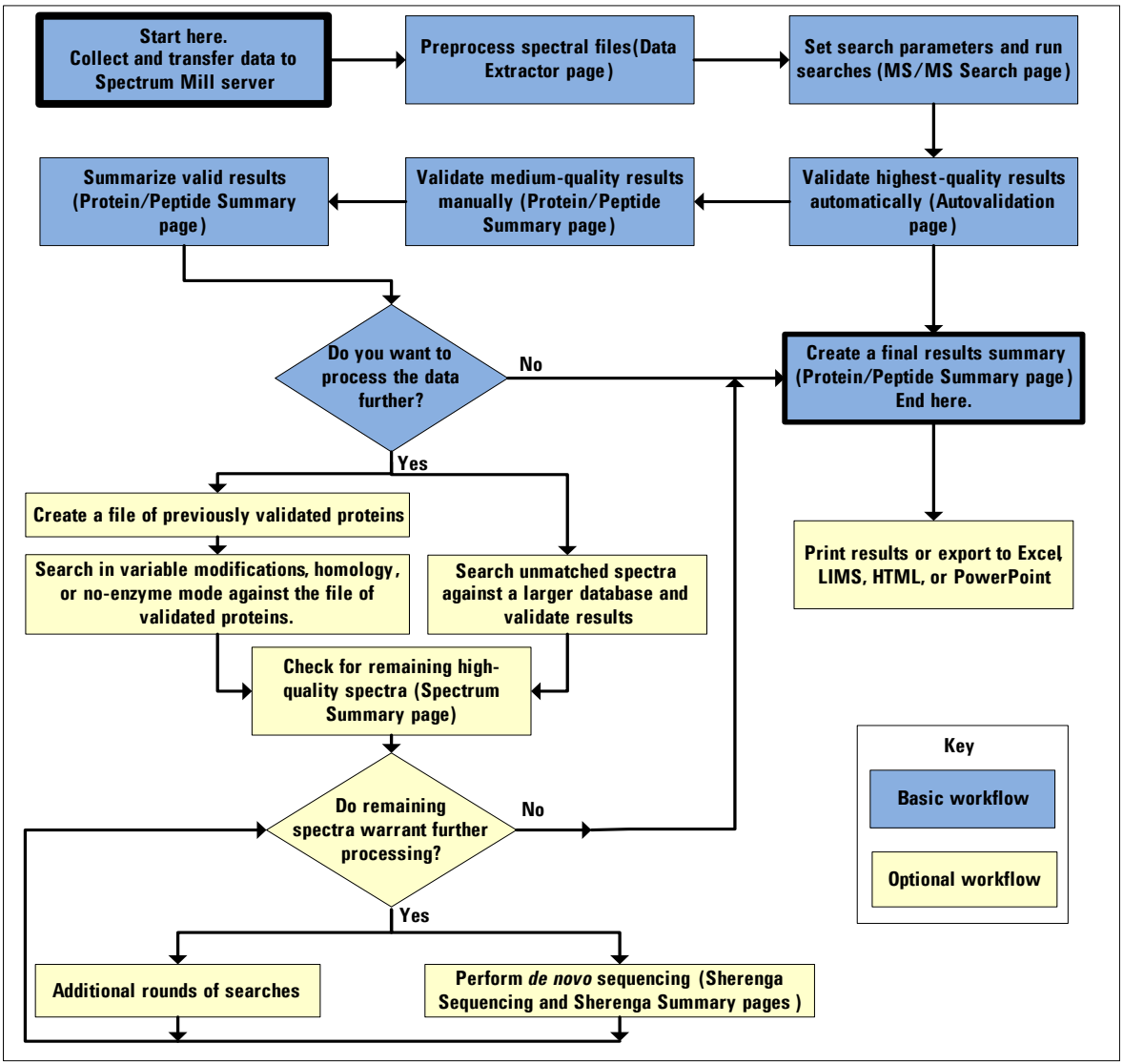
Spectrum Mill modules

Tool type	Module	Function	Application Guide Chapter(s)
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
	Spectrum Matcher	Match MS/MS spectra against other MS/MS spectra.	1
	de novo Sequencing	Use Sherenga *de novo* sequencing to interpret MS/MS spectra.	1, 3
	PMF Search	Search MS-only spectra against databases.	6
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
	de novo Summary	Summarize and review *de novo* interpretations.	1, 3
	PMF Summary	Summarize and review results from MS-only searches.	6

Tool type	Module	Function	Application Guide Chapter(s)
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	Perform theoretical digestion and then select peptides that meet certain criteria (e.g., likely to produce high-quality MS/MS spectra).	8
	Multiple Sequence Aligner	Align amino acid sequences of proteins from a database and highlight the amino acids that differ	8
	MS Digest	Perform 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

Roadmap for MS/MS data processing

See Chapter 1 in the *Application Guide* for details.

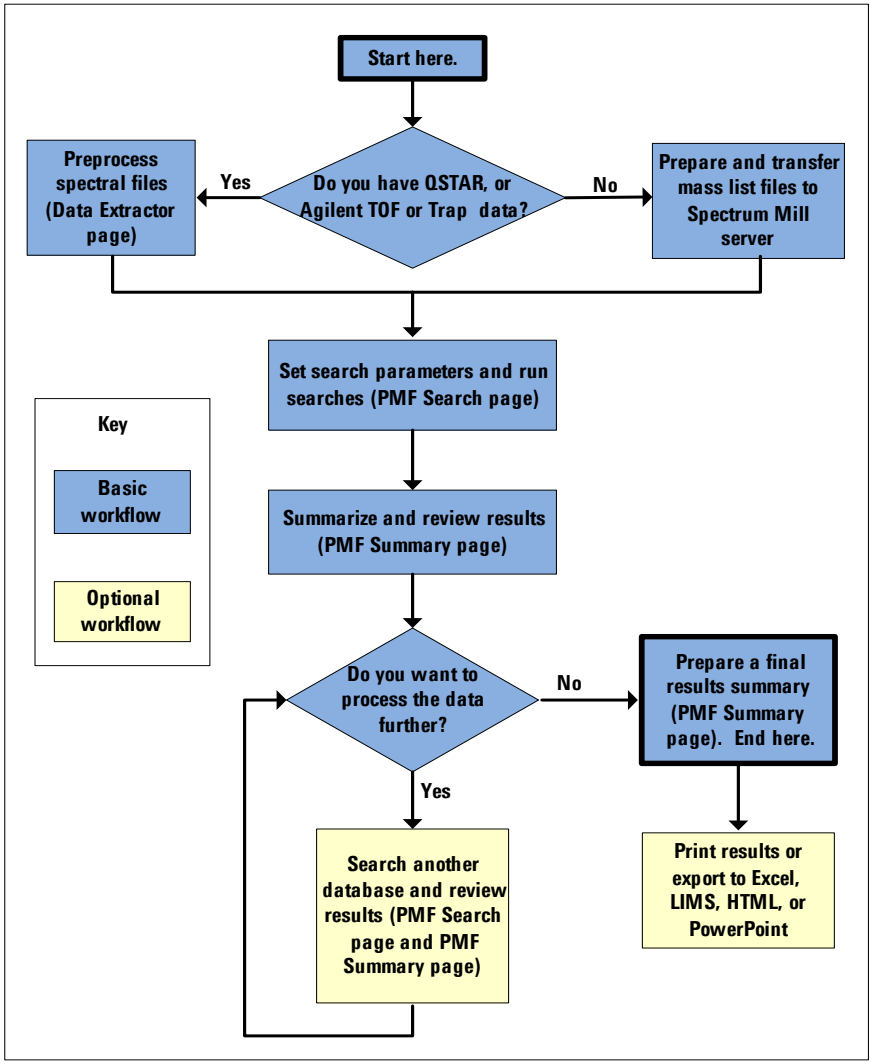


The following iterative search strategy for the Spectrum Mill workbench assumes that the goal is to identify as many proteins as possible. If your study does not require you to identify so many proteins, you may omit some steps (those between step 5 and step 13).

1. Copy or move the raw MS/MS data files to the Spectrum Mill server. Be sure to set up a directory structure on the server that makes it easy to summarize and compare your results.
2. Extract the data.
3. Search a database in identity mode, preferably a species subset database.
4. Autovalidate the results with the highest scores, first in **Protein details** mode, then (optionally) in **Peptide** mode. Use default settings.
5. Manually review the medium-quality spectra and validate additional results (for ion trap data down to score of 6, SPI 60 and for Agilent Q-TOF data down to score of 5, SPI 60).
6. Use Tool Belt to create a saved results file of validated protein hits (***.res** file).
7. Search the spectra that have not been validated. Search them in variable modifications mode against previous validated results. Search as follows with autovalidation and manual validation between each set of conditions:
   a. Combine the following common modifications into a single search: oxidized methionine (methionine sulfoxide) and pyroglutamic acid (N-terminal).
   b. If you suspect phosphorylation, combine the following modifications into a single search: phosphorylated S, phosphorylated T, and phosphorylated Y.
   c. Search for any other modifications you suspect for your sample.
8. Search in no enzyme mode against the previous validated results to find semi-tryptic or non-specific cleavages for proteins you have already identified. (Set **Digest:** to **No enzyme**.) Repeat autovalidation and manual validation.
9. 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.
10. 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 and manually validate.
11. Check statistics in Tool Belt. If there is a significant number of unmatched filtered spectra, continue searching.
12. Use Spectrum Summary to check for sequence-not-validated spectra with sequence tags greater than 6 or 7. Review these and mark as **Good Spectrum** as appropriate.
13. Subject the good spectra to *de novo* sequencing.
14. When you have gained enough information from your data, summarize the results.

Roadmap for MS-only data processing

See Chapter 6 in the *Application Guide* for details.



Modes for Protein/Peptide Summary

See Chapter 2 in the *Application Guide* for details.

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 (with all fractions) 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
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

Guide to common buttons

Any green button	Initiates a process. Also saves form settings (except for data directory) until you close your web browser window.
Save Settings	Saves form settings (except for data directory) for future web browser sessions.
Select	Selects a data directory for this and subsequent forms. Mark the **Make Default** check box to retain this directory for future web browser sessions.
Choose	Selects peptide modifications. If you wish to view details about the modifications that are currently displayed on your server, click the **Details** button on the lower right of the **Choose Modifications** dialog. For more details about choosing amino acid modifications, see the online help.
Reset	Resets settings to those last saved with a green button or **Save Settings**.

Interpreting scores from MS/MS Search

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

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 25 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 *ion trap* results based on MS/MS Search scores

Peptide score	Quality	Peptide fragmentation	Likelihood of valid interpretation
13-25	Excellent	Extensive	When combined with **% SPI** of 70 or greater, very likely to be valid
9-13	Good	Substantial	When combined with **% SPI** of 70 or greater, likely to be valid
6-9	Mediocre	Moderate	Review results to determine whether interpretation is valid
3-6	Weak	Sparse	Not likely to be valid

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.

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

- Peptide score > 5
- % SPI > 60

www.agilent.com

In this Book

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

Printed in USA
Fifth edition, June 2007

G2721-90028

Agilent Technologies