

Release Notes

ProteinPilot™ Software v. 4.5 Beta

Introduction

Thank you for trying out ProteinPilot™ Software v. 4.5 Beta. There are many new features and performance improvements in this version, however, remember that this is a Beta version! Please read the Release Notes carefully to make sure you are aware of the changes. New Beta features, resolved bugs, and remaining issues are noted. Please read all information carefully!

ProteinPilot Software allows you to identify proteins from database searches of mass spectrometry data. The software can search data collected using Analyst® Software, 4000 Series Explorer™ Software, and TOF/TOF™ Series Explorer™ Software, as well as data in Mascot Generic Format (MGF) to analyze data from other instrument vendors.

Two search algorithms are available:

- The *de facto* industry-standard Mascot algorithm.
- The Paragon™ Algorithm, a unique technology that enables simultaneous searching for amino acid substitutions, hundreds of modifications, and unexpected cleavages. For samples that have been labeled with the iTRAQ®, SILAC, mTRAQ®, Cleavable ICAT reagents, and other labeling schemes, the Paragon algorithm also allows you to determine relative protein levels for differential expression studies.

For searches using the Paragon algorithm, the results are further processed by the Pro Group™ algorithm to determine the minimal set of justifiable detected proteins. You can view the spectral evidence for the identifications as well as the sequence coverage. You can also export the results to text files for use in other programs.

Please note that if you want to perform searches with the Mascot algorithm, you need to purchase Mascot Software separately unless you already own a copy.

IMPORTANT! If you open a results file from versions of ProteinPilot Software prior to 4.0 (results files have the .group extension) and save it in ProteinPilot Software 4.0 or later, it cannot be opened again in any older version of ProteinPilot Software.

Requirements

Hardware Requirements

Single-core computers are best used for small-scale searches and experiments. For large-scale proteomics use, multi-core computers are strongly recommended.

Component	Single-core computers	Multi-core computers
RAM	1 GB minimum, with 2 GB strongly recommended for larger searches.	1.5 GB minimum per core, with 2 GB per core strongly recommended for larger searches. For a 64-bit system, the ProteinPilot Software components can use up to 2 GB per core. Any RAM totaling over 2 GB per core can still be used by the system overall (for example, by other searches being run simultaneously by the software).
Disk Space	5.3 GB, as follows: <ul style="list-style-type: none">• ProteinPilot Software – 300 MB for the software and the associated UniProt KB/Swiss-Prot FASTA file.• Mascot Software – 5 GB for the software and the associated FASTA files.	1 TB total storage or more recommended. Maintain a minimum of 200 GB free space for large scale work. This ensures room for temporary files during search. Solid state drives are not believed to offer any advantage in speed of search but will help improve the speed of opening and working with results.
Processor Speed	1 GHz, with 2 or more GHz recommended	1 GHz, with 2 or more GHz recommended

Note: If you install the software on an acquisition station, do not process data with ProteinPilot Software at the same time you are acquiring data. Searching while the instrument is acquiring data can cause problems with data acquisition as well as greatly slow the search speed.

Software Requirements

Operating System

ProteinPilot™ Software has been tested on Dell computers running Microsoft Windows XP with SP3 and Windows XP Professional with SP2 (64-bit) and Microsoft Windows 7 Professional with SP1 (64-bit). The following configurations were used for testing and similar configurations are recommended:

Dell Precision T5500
Windows 7 Professional, SP 1, 64 bit Operating System
Dual Hex Core Intel® Xeon® E5645@ 2.40 GHz CPU
12.0 GB of RAM

Dell Precision T5500
Windows 7 Professional, SP 1, 64 bit Operating System
Dual Quad Core Intel® Xeon® E5640@ 2.66 GHz CPU
12.0 GB of RAM

Dell Precision T7400
Windows XP Professional x64 Edition, Version 2003 SP 2
Dual Quad Core Intel® Xeon® CPU
E5410@ 2.33 GHz
16.0 GB of RAM

Note: If you install the software on an acquisition station, do not process data with ProteinPilot Software at the same time you are acquiring data. Searching while the instrument is acquiring data can cause problems with data acquisition as well as greatly slow the search speed.

Note: While ProteinPilot Software works on Windows 7, keep in mind that many other AB SCIEX software packages do not.

Other Required Software

- **Microsoft Visual C++ 2010 SP1 Redistributable Package (x86)** – Unlike releases prior to v. 4.5 Beta that required this package only for Windows 7, the C++ Redistributable Package is now required for all operating systems. The 2010 SP1 32-bit version must be installed prior to ProteinPilot Software v. 4.5 Beta installation. You can download it from the Microsoft Download Center at the location below:
<http://www.microsoft.com/en-us/download/details.aspx?id=8328>
- **Microsoft .NET Framework 4.0** – If this software is not already installed, you will need to download and install the framework from the Microsoft Download Center at the location below:
<http://www.microsoft.com/en-us/download/details.aspx?id=17718>
- **Microsoft Excel 2010 or 2007 with SP1** – Unlike releases prior to v. 4.2 Beta, you must have Excel 2007 or 2010 installed. Excel 2010 installed as a 64 bit application is preferred because it works better with other Excel template tools associated with ProteinPilot Software, however 2007 will work fine for the FDR analysis done during search.
- **Microsoft Internet Explorer 6.0, 7.0, or 8.0** – Internet Explorer must be set as your default browser.
- **Adobe Reader** – This is required to read documentation supplied with ProteinPilot Software. If it is not already installed, it is available at <http://www.adobe.com>. Click **Get Adobe Reader** and follow the instructions.
- **Mascot Software 2.2 or 2.3** – If you want to use the Mascot Search engine with ProteinPilot Software, you must install Mascot 2.2 or 2.3. Contact your AB SCIEX sales representative for information on purchasing Mascot Software.

Security

You must be an Administrator to install ProteinPilot Software. To install the license, you must be an Administrator, Power User, or Limited User with the ability to write files under the Program Files folder. Once the software is installed and registered, ProteinPilot Software users must have permission to write files to the computer's hard drive (for example, C:\AB SCIEX\ProteinPilot Data and its subfolders).

Regional Options

ProteinPilot Software requires the Regional and Language option for the Windows operating system be set to "English (United States)". To check your settings in Windows 7, from the Start button, click **Control Panel > Region and Language**. If necessary, change the **Format** in the **Format** tab to "English (United States)". Click **OK** to close the dialog box.

Installation Instructions

Note: If you want to use ProteinPilot Software with any of the MALDI TOF/TOF™ Systems, you must also install the Remote Data Access Software on the instrument workstation. This allows ProteinPilot Software to access the data stored in the instrument's database. Follow the instructions in the *ProteinPilot™ Software Installation Quickstart* to install this software.

About the Installation Location

For New ProteinPilot™ Software Installations

If you have never installed ProteinPilot Software before, it is installed by default to C:\Program Files\AB SCIEX\ProteinPilot. The search databases and results are located in C:\AB SCIEX\ProteinPilot Data.

To Upgrade From Installations of ProteinPilot™ Software Prior to v. 4.0

When you upgrade from a release prior to v. 4.0, the software is installed to the directories described above. The installation migrates the contents of the SearchDatabases subfolder to the new AB SCIEX location (for example, C:\AB SCIEX\ProteinPilot Data\SearchDatabases). All other files and folders under the previous company name remain in the same location but can be manually relocated.

Installation Instructions

1. **IMPORTANT!** If a previous version of ProteinPilot Software is already installed on your computer, you must uninstall it before installing ProteinPilot Software 4.5 Beta. Uninstalling does not remove the license file, any results files, or any FASTA files.

Click **Start > Control Panel > Add or Remove Programs**. In the Add or Remove Programs dialog box, click **ProteinPilot Software**, then click **Remove**. Follow the instructions in the dialog boxes.

2. Insert the installation DVD into your computer. Your Web browser should automatically launch and display the *ProteinPilot™ Software Installation Quickstart*.

Note: If the document does not appear automatically, locate the file "ProteinPilotInstallationQuickstart.htm" on the DVD and double-click it.

3. Follow the instructions to install ProteinPilot Software. You will need a license to use the program.

Note: Only licenses for ProteinPilot Software 4.0 will work with ProteinPilot Software 4.5 Beta. Be sure to back up any license files.

If you want to search data collected using 4000 Series Explorer™ or TOF/TOF™ Series Explorer™ software, you must install the Remote Data Access Software on the instrument computer. This software allows ProteinPilot Software to access the data stored in the instrument database. Follow the instructions in the *ProteinPilot™ Software Installation Quickstart* to install this software.

IMPORTANT: The Remote Data Access Software is in addition to any existing software that is already installed on your instrument workstation.

Note: Users of ProteinPilot Software using the Remote Data Access Software to access 4000 Series Explorer or TOF/TOF Series Explorer Software data in the instrument database are considered to be “named users” of the Oracle database under the licensing terms of the instrument. Make sure that you have the required number of “named user” licenses in the license agreement.

4. Optionally, you can install the UniProt KB/Swiss-Prot FASTA file. You can search this FASTA file with both the Paragon™ and Mascot algorithms. Follow the instructions in the *ProteinPilot™ Software Installation Quickstart* to install this file. During the installation, you are asked if you wish to append a file containing common contaminant proteins to the UniProt KB/Swiss-Prot FASTA. It is strongly recommended that you do this. Searching with the contaminant proteins present can eliminate false positive matches for species-specific searches.
Note that the version of UniProt KB/Swiss-Prot installed for Mascot is the standard version that includes only canonical sequences, while the version installed for use with the Paragon Algorithm includes both canonical and isoform sequences. The latter was not included with Mascot because the .dat file could not readily be created from the UniProt KB/Swiss-Prot web site.
5. Launch ProteinPilot Software and follow the instructions to obtain a license.

Mascot Software

ProteinPilot Software 4.0 has been tested with Mascot 2.2 and Mascot 2.3 running on Windows XP with SP3 with IIS (Internet Information Services) using both “Anonymous access” and “Authenticated access”.

Although not supported, ProteinPilot Software might work with Mascot Software 2.1 on Red Hat Linux running Apache 2.0.40 with Web server security enabled on all of the Mascot Software directories. With Web server security enabled, the user account for running ProteinPilot Software must be able to read/execute files in the /mascot/x-cgi/, /mascot/cgi, and /mascot/html directories.

ProteinPilot Software does not support the role-based security features introduced in Mascot Software 2.1.

Note: Mascot Software 2.1 does not use the same nomenclature for reporting modifications as Paragon™ algorithm. Modifications identified by Paragon algorithm searches follow the HUPO-PSI Modification Nomenclature for Mass Spectrometry standard. Follow the instructions in “Updating Modification Nomenclature for Mascot (2.1 Only)” on page 14 to update the

modifications file for Mascot 2.1. The default modifications file for Mascot 2.2 uses the HUPO-PSI nomenclature, so no update is necessary.

Important Issues Remaining in v. 4.5 Beta

The “Save As” and “Save” Functions are not reliable for large datasets

As with the 4.2 Beta, these functions are still not reliable for large datasets. The functions will be very slow if they work at all. We recommend not using them. Because their only real function is capturing quantitation curations, and because curation of data on the scale where this saving will fail is not really tractable, this should not be a major problem. If you need to simply duplicate or rename a .group file, use normal Windows Explorer file manipulation functions as a workaround.

Exported Recalibrated MGF Must Not Be Fed Back to the Paragon Algorithm

The new MGF export function is currently designed to be used only to export for search with other tools after using the Paragon algorithm. It currently cannot be used as a way to avoid repeating the peak list generation stage for subsequent searches with the Paragon algorithm. If you do feed one of these exported peak lists back to Paragon, the spectrum identifiers will not be parsed to recognize the boundaries between original files, and this will cause collisions in spectrum identifiers. Therefore, you should use this export as input for Paragon only for searches of single input files.

Peptide sharing status does not work correctly with quantitation turned on

The new specificity analysis feature designed to provide peptide sharing information with SWATH™ acquisition was intended for use when the Sample Type in the Paragon method is Identification. The specificity analysis does not work correctly if the Quantitate checkbox is selected. If the Quantitate checkbox is not selected, the specificity analysis is reliable for other Sample Type settings, which is a workaround for generating a library for labeled peptides.

Issues and Bugs Resolved in v. 4.5 Beta

False extreme ratios with iTRAQ from MGF files issue is fixed

Multiple users had reported false extreme ratios that resulted from cases where real reporter signals were missed in iTRAQ reagent quantitation using MGF peak list input. ProteinPilot Software v. 4.2 Beta partially addressed this issue, but some false singletons remained. This issue has been further addressed in v. 4.5 Beta.

Failure to export Peptide Summary has been addressed

Previously, exporting the Peptide Summary occasionally failed, for results with more than 300K spectra or when a Save or Save As operation was triggered, resulting in the error message: "Object Reference not set to an instance of an object."

New Features in v. 4.5 Beta

Recently released AB SCIEX instruments are supported

ProteinPilot Software now supports the following additional instruments:

- TripleTOF™ 4600 System
- QTRAP 4500 System
- QTRAP 6500 System
- TripleTOF™ 5600+ System

Each instrument uses its own settings, except for the TripleTOF™ 5600+ system, which uses the same settings as the TripleTOF™ 5600 system.

Peptide Shared status is now indicated

The specificity of any spectrum (that is, the degree to which the associated Identification information is specific to one protein group or points to more than one detected protein group) is now formally recorded in the results. This is referred to as the Shared status of a peptide. The Pro Group algorithm determines this using a spectrum-centric approach, rather than a sequence-centric approach. This provides a more reliable determination of the specificity of a peptide, recognizing a lack of specificity even when peptide sequences are different. (For more information on sharing analysis, see the *Understanding the Pro Group Algorithm* PDF in the ProteinPilot™ Software Help.)

Designating a spectrum as specific or not enables proper selection of unshared peptides to extract SWATH data on proteins of interest when using MS/MS^{ALL} with SWATH™ Acquisition MicroApp 1.0 add-in for PeakView® Software (versions 1.2.0.3 or higher).

The specificity information is also included in the Peptide Summary export in a new column called Specific. If "Specific"=1, the spectrum points to only one confidently identified protein group (Unused >1.3). If more than one confidently identified protein group is indicated by a spectrum with a peptide match, the spectrum is considered Shared and Specific=0.

This feature was intended for use when the Sample Type in the Paragon method is Identification. The specificity analysis does not work correctly if the Quantitate

checkbox is selected. If the Quantitate checkbox is not selected, the specificity analysis is reliable for other Sample Type settings, which is a workaround for generating a library for labeled peptides.

Changed Features in v. 4.5 beta

New quantitation processing

ProteinPilot Software has new processing that results in a dramatic improvement in the quantitation of SILAC and in similar survey-level labeled quantitation workflows. This new quantitation processing also contributes to a general improvement in ID quality for any type of sample processed using .wiff data as input.

The new processing is on by default, but can be turned off by changing a setting in the configuration files, which can normally be found in either the C:\Program Files\AB SCIEX\ProteinPilot folder or C:\Program Files (x86)\AB SCIEX\ProteinPilot folder.

- ProteinPilot.exe.config
- ProteinPilotSignalProcessing.exe.config

You can open these files in any text editor to find the relevant key, which is:

- `<add key="UseAdvancedSignalProcessing" value="true" />`

The key turns on the new signal processing to extract MS1 features. If you change the settings, you need to change the key in both .config files.

Extreme peptide ratios are now excluded from the protein average

ProteinPilot Software now uses the following processing logic to exclude extreme peptide ratios from the protein average:

- (1) Determine quantitation values for each distinct peptide by combining multiple measurements on each distinct peptide as needed.
- (2) Compute the used status of each peptide.
- (3) Determine the proportion of distinct peptides that are extreme. Note that an individual spectrum from a set of repeated acquisitions of the same peptide can appear with an extreme ratios, but when combined with other spectra, result in a peptide level value that is not extreme. In this case, the peptide is not counted as extreme.
- (4) If the proportion of extreme peptides is greater than or equal to 0.15 (15%, the default setting), keep extreme ratios. Do not exclude these because they may actually be extreme ratio proteins.
- (5) If the proportion of extreme peptides is less than 15%, then exclude all peptides with extreme ratios.

Fewer blank ratio values using reduction in minimal S/N settings

Blank ratio cells appear for cases where the minimum signal-to-noise criterion is not met. While all of the data from individual spectra is summed together to compute the distinct peptide level ratio information, these blanks appeared too frequently in some cases. To reduce the number of blank ratios, the minimum S/N threshold has been lowered for some instruments. This threshold ("quant_threshold_sn") was previously set to 9 for TripleTOF™ 5600 and QSTAR® systems and has now been reduced to 7 (for the TripleTOF™ 4600 System also). You can change these settings in the ParameterTranslation file (C:\Program Files (x86)\AB SCIEX\ProteinPilot\WorkflowDirectory\ParameterTranslation.xml):

```
<INSTRUMENT xml:id="INSTRUMENT:60" name="TripleTOF 5600" ionization="ESI"
taglet_ef="EF_CURVE:1" fraglet_correction="1" quant_threshold_sn="7">
  <AUTO_CALIBRATION ms="true" msms="true" mstmstolsdratio="2.75"
mstolsdratio="7" type="dalton" squareroot="true"/>
  <MSTOLERANCE VALUE="0.05" TYPE="dalton"/>
</INSTRUMENT>
```

Notes on Use, Known Problems and Limitations

The following is a list of important notes on using ProteinPilot Software, as well as additional known issues and limitations.

License Issues

- You might see the message “The license for this feature is either invalid or expired.” This can happen when your computer’s network connection is disabled. The ProteinPilot Software license requires that your computer’s network adapter be enabled, though you do not have to be physically connected to the network. Some laptop computers automatically disable the network connection when the power cord is removed from the computer. To check this, click **Start > Connect To > Show all connections**. In the Network Connections window, verify that the Status field for all connections is “Enabled”. If any of the connections are disabled, right-click and choose **Enable**. Exit ProteinPilot Software, and then restart.
- For trial versions of ProteinPilot Software, you will not be able to perform additional searches after the license expires. However, you can still open results files. The same limitations apply to an expired time-limited license.
- ProteinPilot Software uses Log4Net, open-source software that is part of the Apache Logging Services Project, under the terms of the Apache License 2.0.

Memory and Performance Issues

A number of factors can affect the performance of ProteinPilot Software. To avoid or resolve performance issues, be aware of the following information and limitations.

- It is strongly recommended that you do not install ProteinPilot Software on a computer that is also an acquisition station. If you choose to do this anyway, make sure that you do not process data with ProteinPilot Software at the same time you are acquiring data. Searching while the instrument is acquiring data can cause problems with data acquisition as well as greatly slow the search speed.
- If you regularly need to search large datasets (for example, large sets of cation exchange fractions) or if you are working with data from TripleTOF™ systems, make sure that your hardware configuration exceeds the minimum hardware requirements. In these cases, a multi-core system is recommended (for example, an 8-core system with a minimum 1 TB of disk space and 12 GB of RAM). A multi-core system requires an eight-core license type and additional hardware as described in Hardware Requirements on page 2.
- Searching on a computer with multiple cores can result in faster creation of peak lists. On a multi-core system, you can submit multi-file or multi-fraction searches in parallel and then peak list generation and peptide identification are done in parallel. This is faster and more efficient than using a single-core system to submit searches individually and create peak lists serially.
- Searching very large input files (for example, over 8,000 MS/MS spectra) on a computer with multiple cores can also result in faster searches of single files because, when using wiff-based data or MGF files, large files are automatically split into segments and processed in parallel on separate cores. This feature is not available with TOF/TOF input data.
- ProteinPilot Software requires a minimum of 0.5 GB installed RAM per core. If sufficient RAM is not available, the software will limit the number of cores used to meet this RAM requirement, thereby affecting search speed. If your system is below the minimum 0.5 GB installed RAM per core, adding more RAM can increase search speed on multi-core computers.
- ProteinPilot Software disables sorting on the Spectra tab at 300,000 or more spectra, to enable you to load larger data sets and to improve performance. This is by design, not a bug or error.
- Open results files and other applications use RAM. Before starting a large search, in order to make sure enough memory is available, close any open results files and other applications using any significant amount of RAM and restart the ProteinPilot Software.
- Running many searches or opening many results files at the same time might cause your computer to run out of memory and the application to crash. Other symptoms can include very slow performance, “Out of Memory” error messages, and a large red “X” in place of spectral data in graph panes. If any of these problems occurs, make sure you open fewer results files at the same time and do not open results files while you are running a search.

- If you are working with a TripleTOF™ system or other very large data sets with ProteinPilot Software, be aware that Paragon™ Algorithm searches can create extremely large intermediate results files (larger than the eventual .group results files). Regularly check available disk space and delete unused files to make sure sufficient storage space is available when performing new searches. Having less than the recommended amount of free disk space, especially when searching a large set of cation-exchange fractions, might result in the search finishing without error, but with fewer protein identifications. Before starting a large search, make sure you have 200 GB of free disk space on the hard disk used for ProteinPilot Software data (typically C:\AB SCIEX\ProteinPilotData). This is where the Temp folder is stored but does not need to be where you save the final result file.
- If ProteinPilot Software crashes, please restart your computer before running any additional searches.

If you perform a search with data from the TOF/TOF™ Series Explorer™ software or the 4000 Series Explorer™ software and the Analysis Log window shows “Generating peaks...” for a very long time, the network might have been disconnected or the AB SCIEX Remote Data Access Service on the server might have stopped. In most cases, ProteinPilot Software shows an error message, but sometimes it does not. For information on how to resolve this problem, see Troubleshooting search errors with TOF/TOF data on page 11.

Troubleshooting search errors with TOF/TOF data

If you perform a search with data from the TOF/TOF™ Series Explorer™ software or the 4000 Series Explorer™ software and the Analysis Log window shows “Generating peaks...” for a very long time, the network might have been disconnected or the AB SCIEX Remote Data Access Service on the server might have stopped. In most cases, ProteinPilot Software displays an error message, but sometimes it does not. Use the following procedure to fix this problem.

1. Check the network connection and restore the connection, if necessary. Contact your local network administrator if you require assistance.
2. To check the status of the AB SCIEX Remote Data Access, in ProteinPilot Software, click **Configure > Remote Data Servers**.
3. In the **Remote Data Servers** dialog box, click the row for the appropriate server and then click **Test**.

If no error is displayed, go to Step 6. If the test returns an error message, go to Step 4.

4. To check the service on the instrument computer, click **Start > Control Panel > Administrative Tools > Services**.

If the AB SCIEX Remote Data Access service status is “Started,” go to Step 6. If not, go to step 5.

5. Start the service.
6. Restart ProteinPilot Software.
7. Repeat the search.

Issues with Mascot Searches

Searches occasionally fail and display the message “The operation has timed out.” in the Analysis Log window. This can happen when you have several searches in the queue. If you repeat the search when the queue is empty it can complete successfully. [12081]

TOF/TOF Issue with Mascot and Paragon Algorithm Searches

You might notice an MS spectrum that is different from the one you see in the TOF/TOF Series Explorer console. This can occur if an MS job run contains multiple acquisitions on the same spot. If this happens, ProteinPilot Software always submits the spectrum from the most recent acquisition of the spot in the job run to the search. [12882]

Errors in Paragon™ Algorithm Methods

If you try to edit a Paragon algorithm and see the message “One or more errors have been found in this method,” followed by a list of errors. This can happen if ProteinPilot Software cannot locate a valid license or if a setting that is stored in the method is not available (for example, a database cannot be found). Follow the steps in License Issues on page 9 to resolve the problem. [1980]

Searching Nucleotide Databases

The Paragon Algorithm cannot search a nucleotide FASTA file directly (for example, a DNA or an EST file). If you attempt to search a nucleotide FASTA file, the program will crash. If you want to search a nucleotide database, you must first translate the database to a protein sequence. For more information, go to <http://www.ebi.ac.uk/Tools/emboss/transeq/index.html> and use the Upload a file option.

Updating FASTA Files for Use by Mascot

If you update an existing FASTA file for the Mascot server, the new FASTA file will not be listed in the Database field in the Mascot Method dialog box until the Mascot server reports the status of the FASTA file as “in use.” To verify the status, click the **Mascot Database Status** link on the Mascot web page.[1604]

Updating Modification Nomenclature for Mascot (2.1 Only)

If you are using Mascot 2.1, follow the instructions below to update the modifications file on the Mascot server with a version that uses the HUPO-PSI Modification Nomenclature for Mass Spectrometry standard. (Mascot 2.2 and later use this modifications file by default.)

1. Go to the download page at the Unimod website:
<http://www.unimod.org/downloads.html>.
2. Right-click **mod_file** and choose **Save Target As** to save the file to your local computer.
3. When the download is complete, edit the file name to remove the .txt extension.
4. In the ..\mascot\config folder on the Mascot server, rename the current mod_file so you have a backup copy.
5. Copy the downloaded mod_file to the mascot\config folder.

6. If you added modifications using the script from the *ProteinPilot™ Software Installation Quickstart*, follow the “Configure the Mascot FASTA database and modifications” step in the *ProteinPilot™ Software Installation Quickstart* to add those modifications again.
7. If you added any modifications by hand previously, edit the mod_file to add them again.

Searching with Multi-Period .wiff Files

ProteinPilot Software cannot search multi-period .wiff files. If you choose to process a multi-period .wiff file, you will receive an error when the program begins the search.

Searching Infusion Data

ProteinPilot Software cannot search infusion data.

Searching with Files from oMALDI™ Xpert Software

Searches with QSTAR® System LC MALDI data collected with oMALDI™ Xpert Software 2.0 or higher are between 4 to 10 times slower than a comparable search with an electrospray .wiff file. [1402]

Searching with QSTAR® System Data

If you process QSTAR System data and see many errors in peak detection and precursor charge determination, the Bins to Sum parameter during acquisition might have been higher than recommended.

QSTAR System data that is excessively binned during acquisition will not produce good results with ProteinPilot Software. You can tell that the data is excessively binned if peaks have just 1 or 2 points above half height. This excessive binning causes errors in the software’s precursor charge determination.

Binning is controlled by the Bins to Sum acquisition parameter. On a QSTAR System equipped with a four channel time to digital converter (TDCx4), Bins to Sum should be set to 1. On a QSTAR System equipped with an eight channel time to digital converter (TDCx8), Bins to Sum should be set to 4 or 6. If Bins to Sum was set too high, you might want to reacquire your data using the recommended value for Bins to Sum.

Searching .wiff Data on a Remote Computer (or Saving Results to a Remote Computer)

We recommend copying all .wiff data to the local computer before using it with ProteinPilot Software. If you attempt to process .wiff data or save a result to a network drive (rather than your computer’s local hard drive), the processing will take much longer or it can fail (for example, the processing time was 5 minutes for a local file versus 30 minutes for the same file on the network). [410]

Species-Specific Searches Unreliable with Some FASTA Types

Some FASTA files have more than one species listed for a given sequence entry, including the version of SwissProt available from NCBI. If you perform a species-specific search with the Paragon algorithm against one of these FASTA files, only the first species in the sequence entry is considered. (If the species of interest is not listed first, the search will not report a match to it.) If you search the UniProt KB/Swiss-Prot FASTA file, species-specific searches will correctly report all matches. [1649]

Saving .group files can be slow or fail

Saving a .group file after you make changes is slow. It can take a few minutes for small .group files to save, over an hour for large .group files (over 100 MB), and may fail for .group files larger than 1 GB. To avoid this problem, if global settings can be made easily the next time you open a result (for example, turning off bias correction or changing the denominator), then you should either avoid saving global settings after each change, or set all of the global settings and then save them. If you manually curate your quant results, you should also save your curations to avoid having to repeat them.

Opening Results Files from Versions Prior to 4.0

IMPORTANT! If you open a results file from a previous version of ProteinPilot Software (results files have the .group extension) and save it in ProteinPilot Software 4.0, it cannot be opened again in any older version of ProteinPilot Software.

- When you open a results file with quantitation generated by a previous version of ProteinPilot Software, the quantitation results are initially identical in both versions of the software. When a recalculation occurs with the old results in ProteinPilot Software 4.0, the previous results are altered based on version 4.0 logic. While this can result in some difference and improvement in results, the results may not be the same as if you ran the search again in ProteinPilot Software 4.0. This is because, when you simply open an older results file in a newer version of software, only the aspects of the results that are calculated dynamically in the user interface are recalculated; the different identifications and the different integration method from raw data do not change. To make sure new results are accurate, run the search again in ProteinPilot Software 4.0.
- When you open a results file generated by ProteinPilot Software 2.0.1 in ProteinPilot Software 4.0 and the file contains quantitation results, two new columns for each average ratio are available (but hidden by default) in the Proteins Detected table in the Protein Quant tab. The new columns are “Upper CI” and “Lower CI”, the 95% confidence interval bounds for the average ratio. These columns are also included in the Protein Summary export.
- If you open a results file from a CD, do not remove the CD before closing the file. If you remove the CD before you close the file, ProteinPilot Software will hang and you will have to restart your computer to remedy the problem.

- When you open a results file generated by ProteinPilot Software 1.0, you might see the message “Incorrect mod mass from prior version. Error: X” in the Modifications column.

In ProteinPilot Software 1.0, there was a problem searching sequences that contained an X (representing an unknown amino acid) in the FASTA file. The correct peptide is found, but sometimes an incorrect sequence is reported. This problem was fixed in ProteinPilot Software 2.0. The message appears when you open an old results file for a search where this problem with X was encountered.

This message is intended to indicate a change to one or more modifications, but it will also appear in this case (for a sequence with X). For a list of the modifications with changes, see *Modifications Catalog and Translations*, located in the ProteinPilot™ Software Help folder.

If possible, you should repeat the search with the most current version of ProteinPilot Software. [2014]

- Some of the modification names change between versions of ProteinPilot Software. For example, if you open a results file generated by ProteinPilot Software 1.0, the modification Ubiquitination GG is shown as GlyGly. For all modification changes between ProteinPilot Software versions, please see *Modifications Catalog and Translations* in the ProteinPilot™ Software Help folder.

Problems with Automatic Updates to Windows

Your computer might be set up to automatically download and install updates to the Windows operating system. Some of these updates require that the computer be restarted and these restarts could happen automatically during a search.

If a ProteinPilot Software search is in process when the computer is restarted, the search is terminated and no results are saved. You will have to run the search again. You can eliminate this problem in one of two ways:

- Do not run searches when your system is scheduled for updating.
- or*
- Change your update options so the computer is not automatically restarted. Click **Start > Control Panel > Automatic Updates** and choose either: **Download updates for me, but let me choose when to install them** or **Notify me but don't automatically download or install them**.

Converting existing .group files to .xml format

In order to facilitate the conversion of existing .group files to the .xml format, the following batch file can be created on your desktop. The batch file enables you to drag and drop any .group file from Windows Explorer to the desktop icon and the associated .xml file will be created in the same folder as the original .group file.

Prerequisites

The following software is required to use the conversion batch file:

- ProteinPilot Software must be installed
- Windows Scripting must be enabled

To create the batch file

1. Using the text editor of your choice, create a batch file named “convert group to XML.bat” on your desktop.
2. Edit the batch file and insert the following:

```
@ECHO OFF
echo set WshShell = WScript.CreateObject("WScript.Shell") > temp.vbs
echo DesktopFolder = WshShell.SpecialFolders("AllUsersDesktop") >> temp.vbs
echo set Lnk = WshShell.CreateShortcut(DesktopFolder ^& "\ProteinPilot.lnk") >> temp.vbs
echo wscript.Echo Lnk.WorkingDirectory >> temp.vbs
set vbscript=cscript //nologo temp.vbs
For /f "delims=" %%T in ( ' %vbscript% ' ) do set folder=%%T
del temp.vbs
:Loop
IF "%~1"==" " GOTO Continue
set inputFile=%1
for /f "useback tokens=*" %%a in ('%inputFile%') do set inputFile=%%~a
set outputFile=%inputFile%.xml
ECHO %outputFile%
IF EXIST "%folder%\group2xml.exe" "%folder%\group2xml.exe" XML "%inputFile%"
"%outputFile%"
SHIFT
GOTO Loop
:Continue
pause
```

3. Save the changes.

To use the batch script

1. Using Windows Explorer, navigate to the location of the .group files to be converted.
2. Drag and drop each .group file you want to convert from Windows Explorer to the “convert group to XML.bat” icon on your desktop.
A window appears while the conversion takes place.
3. When conversion is complete, press any key to close the window.
For each original .group file dropped on the “convert group to XML.bat” icon, an equivalent .xml file is now available in the same folder as the original .group file.

NOTE: If the conversion does not take place, make sure that the batch file is named “convert group to XML.bat” and that .txt has not been automatically added as an extension. If you use Notepad as your text editor, this can happen automatically.

Working with the Pro Group™ Algorithm Results

- If you zoom in on the graph showing spectral evidence in the Fragmentation Evidence pane, you might notice that the highlight bars corresponding to the m/z for theoretical ions are slightly offset from the peaks. This is a drawing artifact and can be ignored; the matching is performed correctly.
- If you zoom in repeatedly on any graph showing spectral evidence, eventually you might see an error message and a big red X replacing the spectrum. If this occurs, close the result and reopen it in order to see the spectral data again.
- If you use the Copy All command in a table immediately followed by Copy Selected, the Copy Selected command doesn't work correctly. Make a different selection in the table, select the row(s) you want to copy and then choose **Copy Selected**.

Excel Limitations for Exported Results

If you export the Peptide Summary from a search with a large number of results, the file might be too large to view completely in Excel. Excel 2007 and 2010 have a limit of 1,048,576 rows, so you will not see any results beyond these rows. If this happens, you can open the file in a text editing program like Notepad and divide it into smaller files.

Tool Tips in the Add TOF/TOF Data Dialog Box

- In the Add TOF/TOF Data dialog box, you might see a difference between the parent job run shown in the tool tip in the Available Data list (hold the cursor over a MS/MS job run) and the parent job run displayed in the MS Job Run list. The parent job run in the MS Job Run list is the correct job run and is processed during the search.
- In the tool tips in the MS Job Run list you might see “0” as the number of Completed Items when Manual Determination is selected. This can occur when the MS job run is created to perform MS/MS interpretation of another MS job run. In this case, ProteinPilot Software identifies and processes the correct parent job run during the search. (You may also see “0” as the number of Completed Items when no data was collected.)

Support for Additional Workflows

If you don't see support for a workflow that is important to you, contact us. We are always working to extend ProteinPilot Software, whether it is with something simple like a different digest agent, or something more complex like a new workflow. It might even be possible to give you a solution prior to the next release of the software.

Email your requests and suggestions to support@absciex.com.

Where to Get Help

Learning ProteinPilot™ Software

Use the ProteinPilot™ Software Help to help you learn about ProteinPilot Software. The Help is available from the Help menu in the program.

The following documents provide additional information about ProteinPilot Software. To view any of these documents, click **Start > All Programs > AB SCIEX> ProteinPilot > Help**.

- *ProteinPilot™ Software Getting Started Guide* – A quick introduction to the program, with sample data files included.
- *ProteinPilot™ Software Online Help* – The Online Help in PDF format, made available for printing.
- *Paragon™ Method Settings Guide* – Detailed information on each setting available for use with Paragon Algorithm searches, including guidance on when each setting is appropriate and a description of how well optimized each setting is.
- *False Discovery Rate Analysis with PSPEP* – An explanation of the false discovery rate analysis available in ProteinPilot Software.
- *Understanding the Pro Group™ Algorithm* – A conceptual explanation of the Pro Group™ algorithm and the protein grouping problem.
- *Modifications Catalog and Translations* – Detailed information about the modifications used by the Paragon™ algorithm.
- *Advanced Configuration of ProteinPilot™ Software* – Instructions for adding custom modifications digest agents, species filters, and other information for use by the Paragon algorithm.
- *Scripted Searching with ProteinPilot™ Software* – Instructions for running Paragon algorithm searches from the command line. Files for an example search are included on the installation DVD.
- *Working with ProteinPilot™ Software Results in XML* – Instructions for converting the .group file containing the Pro Group results to XML and the schema of the resulting file. Example files are included on the installation DVD.

Support

For support, contact AB SCIEX.

- Email: support@absciex.com
- Web: www.absciex.com

This document is provided to customers who have purchased AB SCIEX equipment to use in the operation of such AB SCIEX equipment. This document is copyright protected and any reproduction of this document or any part of this document is strictly prohibited, except as AB SCIEX may authorize in writing.

Software that may be described in this document is furnished under a license agreement. It is against the law to copy, modify, or distribute the software on any medium, except as specifically allowed in the license agreement. Furthermore, the license agreement may prohibit the software from being disassembled, reverse engineered, or decompiled for any purpose.

Portions of this document may make reference to other manufacturers and/or their products, which may contain parts whose names are registered as trademarks and/or function as trademarks of their respective owners. Any such usage is intended only to designate those manufacturers' products as supplied by AB SCIEX for incorporation into its equipment and does not imply any right and/or license to use or permit others to use such manufacturers' and/or their product names as trademarks.

AB SCIEX makes no warranties or representations as to the fitness of this equipment for any particular purpose and assumes no responsibility or contingent liability, including indirect or consequential damages, for any use to which the purchaser may put the equipment described herein, or for any adverse circumstances arising therefrom.

For Research Use Only. Not for use in diagnostic procedures.

The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners.

AB SCIEX™ is being used under license.

AB SCIEX

71 Four Valley Dr., Concord, Ontario, Canada. L4K 4V8.

AB SCIEX is ISO 9001 registered.

© 2012 AB SCIEX.

