AB SCIEX

ProteinPilot™ Software 4.0

Release Notes

Please read the Release Notes prior to installation.

The information contained in this document is designed to help ensure a successful installation for new users.

Contents

Introduction

Thank you for purchasing ProteinPilot™ Software 4.0. There are many new features and performance improvements in this version. Please read the Release Notes carefully to make sure you are aware of the changes.

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).

Two search algorithms are available:

- The *de facto* industry-standard Mascot algorithm.

- The Paragon™ algorithm, a unique new 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®, or Cleavable ICAT reagents, 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 a text file 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 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.

New Features in ProteinPilot™ Software 4.0

SUPPORT FOR THE AB SCIEX TRIPLETOF™ 5600 SYSTEM

ProteinPilot Software 4.0 is compatible with data collected from the new AB SCIEX TripleTOF™ 5600 System.

COMBINED SEARCHING OF DATA FROM DIFFERENT INSTRUMENTS

You can now select data from different instruments as input for a single, combined search. The correct instrument type is automatically determined for all native AB SCIEX data types and one additional instrument type can be added by MGF file format submission. Peptide identification is performed based on data type, and then all peptide identifications from submitted instruments are combined in a single protein identification result by the Pro Group algorithm. This provides the full benefit of any complementarity of coverage (for example, ESI-MALDI complementarity). All input data must be of the same sample type (for example, you cannot combine 4plex and 8plex iTRAQ reagent data).

QUANTITATION OF ITRAQ REAGENT DATA FROM NON-AB SCIEX INSTRUMENTS

Quantation analysis on iTRAQ reagent data generated by other vendors' instruments can now be done by MGF file format submission. MS-based quantitation analysis (for example, SILAC, mTRAQ) cannot be done by MGF file format submission because MGF files do not typically support the amount of information needed for MS-based quantitation analysis.

PRECURSOR-LEVEL ANALYSES

The following three features comprise the new precursor-level analyses functionality provided in ProteinPilot 4.0 software. These features are available for native input data types, but not available for MGF input.

PRECURSOR INTENSITY INFORMATION

The peptide export now contains an estimate of the intensity of each acquired precursor, currently reported as the MS signal intensity at the apex retention time. This enables a wide range of label-free quantitation analyses. This feature is not available for QTRAP instruments.

PRECURSOR APEX RETENTION TIME

The peptide export now contains both the retention time where the product spectrum was acquired and the retention time of the apex of elution. The retention time of the apex of elution can be more useful for certain types of analysis, for example, Scheduled MRM™ analysis.

PRECURSOR-WIDE MASS DETERMINATION

The precursor mass is now estimated by using multiple measurements across the elution of a peptide, rather than by taking the measurement in the survey spectrum that triggered the product spectrum. This determination method results in improved mass accuracy, particularly for LC-ESI data where the acquisition often occurs at a point below the apex intensity of elution.

PEPTIDE CONFIDENCE RECALIBRATION YIELDS MORE ACCURATE PEPTIDE AND PROTEIN CONFIDENCES

When the confidences reported by a search engine do not correspond well to the error rate they indicate, we refer to this as *confidence inaccuracy*. If the degree to which peptide confidences are inaccurate vary across a set of fractions or across data coming from different instruments, the overly confident components can negatively affect the quality of the whole. In ProteinPilot Software 4.0, we prevent this problem by performing FDR analysis internally after the peptide ID stage; this recalibrates the results for each input file. For example, the 5% local FDR rate occurs at the 95% reported peptide confidence level in each input file. This results in more accurate peptide identifications and in better protein identification, particularly for multi-fraction data sets and combined data from multiple instruments.

NEW PROTEIN CONFIDENCE ALGORITHM

ProteinPilot Software 4.0 features a completely new approach to calculating protein confidences from peptide identifications. This approach recognizes that, with each iteration of the protein grouping process, the proportion of correct answers in any given peptide confidence bin is depleted as the grouping process proceeds. For example, a peptide given 95% confidence during peptide identification does not have the same reliability when computing the confidence for the 1000th protein in the list that it did for the first protein in the list. This is because the pool of

peptides that initially had 95% confidence at the beginning of grouping is attributed out to many multi-hit proteins as the grouping process proceeds. This depletes this pool of correct answers far more than incorrect answers, which therefore increases the chance of an incorrect answer.

The confidence bin was effectively getting enriched for wrong answers as peptides were assigned to proteins. The new approach to calculating protein confidences corrects for this effect, resulting in better discrimination in reported proteins, and it can yield more protein identifications at a given FDR.

IMPROVED IDENTIFICATION FOR QTRAP SYSTEMS

Because of improvements in signal processing to make peak lists from raw data, searches of QTRAP data now provide substantially improved identification results. Improvements are particularly noticeable for MIDAS data and other data with very low signal strength. Improvements as large as 80% additional identifications at the same FDR have been observed at the spectrum level.

IMPROVED TREATMENT OF EXTREME RATIOS

In previous versions of ProteinPilot Software, extreme ratios were indicated by a convention where extreme low ratios were reported as zero and extreme high ratios were reported as 9999. Because calculation and graphing of ratio data is generally transformed to log space, this convention was problematic. At the low end log, zero was undefined and at the high end log, 9999 was only a convention, with no connection to how large the ratio might actually be. With this approach, extreme ratios at the peptide level could only be ignored in the calculation of a protein level average, unless all values were extreme, which would be the single scenario where an extreme ratio would be reported at the protein level. In ProteinPilot Software 4.0, extreme ratios are treated differently. The upper limit is set at 100-fold, with extreme low ratios reported as 0.01 and extreme high ratios reported as 100. The 100-fold limit is considered a realistic upper limit estimate of the most extreme

ratio that could be meaningfully reported by most label-based quantitation methods. This approach allows extreme values at the peptide level to play a role and facilitates other calculations and visualization of results in log space.

MORE ROBUST MASS ACCURACY RECALIBRATION

While ProteinPilot Software has always used an iterative search approach to recalibration masses of both precursors and fragments, ProteinPilot Software 4.0 has been substantially improved to be more robust in how it performs this recalibration. This results in better functional mass accuracy for any instrument and better data robustness for instruments that may not be calibrated accurately.

UPDATED FDR TEMPLATES

The Excel-based report produced when you invoke a false discovery rate (FDR) analysis with your search has been revised and includes additional information. New tables in ProteinPilot Software 4.0 show a reported confidence level that corresponds to an FDR level of interest. This allows you to threshold your peptide or protein results at the reported confidence level for ProteinPilot Software that corresponds to the FDR level you prefer. For example, in previous versions, if you identified 8,534 distinct peptides at a 5% local FDR, you did not know what peptide confidence level this corresponded to. The new table indicates the exact confidence threshold to use in order to include this set of peptides. A 5% local FDR should correspond to 95% confidence; however, in a specific result, the report might show that 5% local FDR corresponds to a confidence of 93.2 % or some other value close to 95%.

INCREASED SINGLE FILE SIZE LIMIT FOR WIFF-BASED DATA

Wiff-based data files are now automatically split into multiple segments as needed to process larger files. Multiple segments are then processed in parallel. Files with 60,000 MS/MS spectra have

been successfully processed. The limit on number of spectra in one file is expected to be much larger than 60,000.

Note: There is still a limit of approximately 20,000 spectra for MGF and TOF/TOF input.

NO SIZE LIMIT FOR FASTA FILES

There is no longer a 4 GB size limit for FASTA files. While this enables you to now search very large databases, including the most recent versions of NCBI NR and UniProt TrEMBL, there can be a performance impact and, for this reason,we recommend against regular searches of extremely large databases.

IMMEDIATE PARALLEL SEARCHES ON MULTI-CORE COMPUTERS

In previous versions of ProteinPilot Software, when a search was already in process, there was often a lag between the submission and the initiation of additional searches to the queue. Using an appropriately configured multiple-core system with adequate RAM, you can submit and process multiple searches in parallel immediately, if cores are available. This results in higher overall search throughput.

Note: This feature requires an eight-core license type and additional hardware as described in Hardware Requirements on page 10.

ADDITIONAL WORKFLOWS, SETTINGS, AND OPTIMIZATIONS

As with each release, ProteinPilot Software 4.0 includes expanded support using new settings options and improved optimization of existing settings in the Paragon Method screen. A substantial number of additional workflows are supported in version 4.0. For more information, see the *Paragon™ Method Settings Guide* in the Help folder.

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: • **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

Component	Single-core computers	Multi-core computers
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). The following configurations were used for testing and similar configurations are recommended:

Dell P390
Windows XP 2002 SP 3
Intel®Core™2 CPU
6700@ 2.66 GHz
2.00 GB of RAM

Dell T3400
Windows XP 2002 SP 3
Intel®Core™2 Duo CPU
E6750@ 2.66 GHz
2.00 GB of RAM

Dell T7400
Windows XP Professional x64 Edition, Version 2003 SP 2
64-bit Quad-Core Intel® Xeon®CPU
E5410@ 2.33 GHz
16.00 GB of RAM

Initial testing of the ProteinPilot software on the Windows 7 x64 Professional operating system was successful; however, neither Analyst® Software nor BioAnalyst® Software support Windows 7 and they are not currently known to work on Windows 7. This can result in failed searches with both .wiff and MALDI TOF/TOF files. If you want to use the ProteinPilot software on the Windows 7 x64 Professional operating system, you need to install the Microsoft *Visual C++ 2008 SP1 Redistributable Package (x86)*

download from the Microsoft Download Center at
http://www.microsoft.com/downloads. [22816]

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.

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 Algorithm Searches (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.

Other Required Software

- **Microsoft Internet Explorer 6.0, 7.0, or 8.0** – Internet Explorer must be set as your default browser.

- **Microsoft .NET Framework 3.5** – If this software is not already installed, it will be automatically installed when you install ProteinPilot Software.

- **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 the ProteinPilot Software, you must install Mascot Software 2.2 or 2.3. Contact your AB SCIEX sales representative for information on purchasing Mascot Software.

- **Microsoft Excel 2007 with SP1** – If you want to open and view the results of the false discovery rate (FDR) analysis, you must use Excel 2007 with SP1 (or later). See *False Discovery Rate Analysis with PSPEP* in the ProteinPilot Help folder, for information on obtaining the required 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 Options for the Windows operating system be set to "English (United States)". To check your settings, from the Start menu, click **Settings > Control Panel**. Double-click the Regional Options icon. If necessary, change the location in the **Your locale (location):** list to "English (United States)". Click **OK** to close the dialog box.

Installing ProteinPilot™ Software

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™ 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 PREVIOUS INSTALLATIONS OF PROTEINPILOT™ SOFTWARE

When you upgrade from a previous installation, 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 the ProteinPilot Software 4.0. 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™ 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 the ProteinPilot Software. You will need a license to use the program.

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

    In addition to ProteinPilot Software, the following items are installed:

    –   **Microsoft .NET Framework 3.5** – This software is required for ProteinPilot Software. If it is not present on the computer, it is automatically installed.

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™ 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™ 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.

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 Getting Started Guide* – A quick introduction to the program, with sample data files included.

- *ProteinPilot 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.

Notes on Use, Known Problems and Limitations

The following is a list of known issues, limitations, and important notes on using the software.

If you have a technical question or a request for assistance, please contact AB SCIEX support:

- email: support@absciex.com

- Web site: http://www.absciex.com

Support for Additional Workflows

If you don't see support for a workflow that is important to you, contact us. We are always looking 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.

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 the 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 the AB SCIEX TripleTOF™ 5600 instrument, 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 10.

- 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, large files are automatically split into segments and processed in parallel on separate cores. This feature is not available with MGF or TOF/TOF input files.

- If sufficient RAM is not available, the software will not use the additional cores for processing. Adding more RAM can increase search speed on multi-core computers.

- The 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.

- If a large (> 2 GB) results file is open when you begin to run a search, the search might not complete successfully. Before starting a large search, to make sure enough memory is available, close any open results files 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 an AB SCIEX TripleTOF™ 5600 instrument and ProteinPilot software, be aware that Paragon™ Algorithm searches can create extremely large intermediate results files (much 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.

- If the 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 26.

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 the 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 the ProteinPilot software.

7. Repeat the search.

ISSUES WITH MASCOT ALGORITHM SEARCHES

You might see searches fail with 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.

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, the ProteinPilot Software always submits the spectrum from the most recent acquisition of the spot in the job run to the search.

ERRORS IN PARAGON™ ALGORITHM METHODS

If you try to edit a Paragon algorithm method, you might 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 algorithm method is not available (for example, a database cannot be found). Follow the steps in License Issues on page 22 to resolve the problem.

ADDING ADDITIONAL FASTA FILES FOR USE BY THE PARAGON™ ALGORITHM

ProteinPilot Software cannot search a nucleotide FASTA file (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 THE MASCOT ALGORITHM

If you update an existing FASTA file for the Mascot server, you might not see the new FASTA file 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.

UPDATING MODIFICATION NOMENCLATURE FOR MASCOT ALGORITHM SEARCHES (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™ Installation Quickstart*, follow the "Configure the Mascot FASTA database and modifications" step in the *ProteinPilot™ 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

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

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.

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. [1990]

SEARCHING.WIFF DATA ON A REMOTE COMPUTER (OR SAVING RESULTS TO A REMOTE COMPUTER)

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). We recommend copying all .wiff data to the local computer before using it with ProteinPilot Software.

SPECIES-SPECIFIC SEARCHES

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.

BLANK PEPTIDE QUANT RATIOS SOMETIMES AFFECT QUANTITATION RESULTS

If a very low S/N value for the pair of reporter ions exists and no specific ratio is dispayed in the peptide quant table (the S/N value is "blank"), the actual observed ratio can be used in the protein quant calculation. While the effect from this spectra on the protein ratio is expected to be very minor, you can choose to not use the observed ratio in the calculation by clearing the "Use in quant" check box. [026193]

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. [025757, 025759]

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 Help folder.

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

- 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 Help folder. [2073]

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**.

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 Pro Group algorithm results from a search with a large number of results, the file might be too large to be opened in Excel. (Excel 2003 has a limit of 65,536 rows in one file; Excel 2007 has a limit of 1,000,000 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.)