Release Notes

ProteinPilot™ Software

Software for protein identification and quantitative analysis of mass spectrometry data used for protein characterization and proteomics

Version 5.0.1

Introduction

Thank you for using ProteinPilot™ software.

Read the *Release Notes* carefully for information on changes, new features, resolved and known issues. The information contained in this document is designed to help ensure successful installation and use of the ProteinPilot software.

The ProteinPilot software provides features for users to identify and quantify proteins and peptides from database searches of mass spectrometry data. The software can search data collected using the Analyst® software, the 4000 Series Explorer™ software, and the 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 search engine.
- 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 SCIEX iTRAQ®, SILAC, SCIEX mTRAQ®, Cleavable ICAT reagents, and other labeling schemes, the Paragon algorithm also allows users 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. Users can view the spectral evidence for the identifications as well as the sequence coverage. Results can also be exported to text files for use in other programs.

Mascot must be purchased separately.

New Enhancements in Version 5.0.1

This release addresses minor compatibility issues with the new recommended hardware configuration (Dell T7910, running Microsoft Windows 7 64-bit operating system) and provides support for Microsoft .NET 4.5.1 framework.

Requirements

Hardware Requirements

Dual-core computers are acceptable only for small-scale searches, such as the characterization of single proteins using older instruments. For large-scale proteomics use, multi-core computers (with 8 to 32 cores) are strongly recommended. Multi-core computers are required when working with data from TripleTOF® systems or other instruments producing high volumes of data.

Component	Dual-core computers	Multi-core computers
RAM	1 GB minimum, with 2 GB strongly recommended.	1.5 GB minimum per core, with 2 GB per core strongly recommended. As a 64-bit system, the ProteinPilot 5.0 software can use all available RAM.
Disk Space	305 GB, as follows: • ProteinPilot Software – 300 GB for the software, the associated UniProt KB/Swiss-Prot FASTA file, and storage of associated temporary files. • Mascot – 5 GB for the search engine and the associated FASTA files.	1 TB total storage or more recommended. Maintain a minimum of 300 GB free space for large scale work. This ensures room for temporary files during search. Solid state drives might not offer any advantage in search speed, but they will help to improve the speed of opening and working with the results.
Processor Speed	1 GHz, with 2 or more GHz recommended	1 GHz, with 2 or more GHz recommended

NOTE*:* It is strongly recommended that the software is not installed on an acquisition station. However, if the software must be installed on an acquisition station, make sure to not process data with the ProteinPilot software at the same time as acquiring data. Searching while the instrument is acquiring data can cause issues with data acquisition as well as greatly slow the search speed.

Software Requirements

Operating System

ProteinPilot 5.0.1 software has been tested on Dell computers running Microsoft Windows 7 Professional with SP1 (64-bit). ProteinPilot 5.0.1 software requires a 64-bit operating system.

The following configuration was used for testing. While other configurations might work, similar configurations are recommended.

Dell Precision T7910 or DTx Dual Xeon Workstation
Dual 8-core XEON @ 3.4 GHz
32 GB RAM
3 TB Hard Drive, 7200 RPM (primary)
3 TB Hard Drive, 7200 RPM (secondary)
100 Mb NIC/Network

While the ProteinPilot software works on 64-bit Windows 7 with SP1, some SCIEX software packages do not.

Other Required Software

The following Microsoft components are installed automatically during the ProteinPilot software installation.

- Microsoft Visual C++ 2010 SP1 Redistributable Package – 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.
- Microsoft .NET Framework 4.5.1 – If this is not already installed on the computer, it is installed as part of the ProteinPilot software installation.

The following software is also required for the ProteinPilot software use.

- Microsoft Excel 2010 or 2013 (64-bit) software – The default version of Microsoft Excel is 32-bit, even if the Windows operating system installed is 64-bit. Make sure to select and install the 64-bit version of Microsoft Excel 2010 or 2013.
- Microsoft Internet Explorer 6.0, 7.0, 8.0, or 9.0 – Internet Explorer must be set as the default browser.
- Adobe Reader – This is required to read documentation supplied with the 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 2.4 – To use the Mascot search engine with the ProteinPilot software, first install Mascot 2.4 and the associated FASTA files. Contact your SCIEX sales representative for information on purchasing Mascot. Mascot can be installed either on the same computer as the ProteinPilot software or, ideally, on a separate one.

Security

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

Regional Settings Requirements

The **Region and Language** setting for the Windows operating system must be set to **English (United States)**.

To check the settings in Windows 7, from the **Start** button, click **Control Panel > Region and Language**. If necessary, change the **Format** on the **Format** tab to **English (United States)**. Click **OK** to close the dialog.

Installation Instructions

NOTE*:* To use the ProteinPilot software with a 4700 Proteomics Analyzer™ system, a 4800 MALDI TOF/TOF™ Analyzer system, or a 5800 MALDI TOF/TOF™ system, first install the Remote Data Access Software on the instrument workstation. This allows the ProteinPilot software to access the data stored in the instrument's database. Follow the instructions in the *ProteinPilot™ Software Installation Quick Start* to install this software.

About the Installation Location

For New ProteinPilot Software Installations

If the ProteinPilot software has not been installed before, ProteinPilot 5.0.1 software 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

If upgrading 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.

About Mascot

The ProteinPilot 5.0 software has been tested with Mascot 2.4 running on Windows XP with SP3 with IIS (Internet Information Services) using both **Anonymous access** and **Authenticated access**.

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

Installing ProteinPilot Software

NOTE: If a previous version of the ProteinPilot software is already installed on the computer, the ProteinPilot 5.0.1 software removes it automatically. Removing the software does not remove the license file, any results files, or any FASTA files.

1. Insert the installation DVD into the computer.
   The Web browser should automatically launch and show the *ProteinPilot™ Software Installation Quick Start*.
   If the document does not appear automatically, locate the file **ProteinPilotInstallationQuickstart.htm** on the DVD and double-click it.

2. Follow the instructions to install the ProteinPilot software.

3. To search data collected using 4000 Series Explorer™ or TOF/TOF™ Series Explorer™ software, install the Remote Data Access Software on the instrument computer. This software allows the ProteinPilot software to access the data stored in the instrument database.

   IMPORTANT: The Remote Data Access Software is in addition to any existing software that is already installed on the instrument workstation. Follow the instructions in the *ProteinPilot™ Software Installation Quick Start* to install this software.

   Users of the ProteinPilot software that are also using the Remote Data Access Software to access the 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 the required number of “named user” licenses is available in the license agreement.

   Optionally, users can install the UniProt KB/Swiss-Prot FASTA file. This FASTA file can be searched with only the Paragon™ algorithm.

   The Paragon algorithm searches described in the *ProteinPilot™ Software Getting Started Guide* were performed using the June 22, 2010 version of the UniProtKB/Swiss-Prot FASTA file distributed with the software, which includes both canonical and isoform sequences and has had the contaminant protein FASTA file appended to it. This database can be used without the additional contaminants or another compatible FASTA file, however, the results will not exactly match those shown in the *Getting Started Guide*.

   The Mascot searches described in the *ProteinPilot™ Software Getting Started Guide* were performed using the Mascot 2.4.1 software and a Swiss-Prot FASTA file compiled on December 16, 2008. If a newer FASTA database is being used, the results might differ slightly.

4. Launch the ProteinPilot software and then follow the instructions to obtain a license.

   NOTE: A license is needed to run searches with the program. If installed without a license, the software functions as a viewer that can perform all of the tasks except running searches.

   ProteinPilot 5.0 software licenses must be used with ProteinPilot 5.0.1 software. Be sure to back up any license files.

New Features

New Features in v. 5.0

Support for new instruments

The AB SCIEX 6600 TripleTOF® system is now supported.

Speed and scale improvements

The speed of processing, opening results, browsing results, saving results, and exporting results have all been improved. Search speed is approximately twice as fast as that of ProteinPilot software 4.5 Beta. The upper limit of scale has been increased from approximately 1 to 2 million spectra in ProteinPilot software 4.5 Beta to more than 5 million spectra. With all of the components now converted to 64-bit, the limit to the scale of data that can be handled is based on the amount of RAM available.

ID improvements

The quality of identification has been improved. Signal processing improvements have yielded better detection of LCMS features, particularly in lower intensity situations where multiple peptide features are often blended together. This means that the search begins with an improved list of peptides to identify and quantify. The software can also now identify more than one peptide from a single MS/MS spectrum. Signal processing improvements also provide better mass accuracy and quantitation.

Redundant spectra can be suppressed from view

Advanced instruments often trigger on the same peptide multiple times. When inspecting results, redundant spectra can now be suppressed from view, leaving only the minimal information required to evaluate IDs. On the **Protein ID** tab and the **Protein Quant** tab, in the **Peptides in Group** and **Peptide Quantitation** tables, respectively, there is a new **Remove Redundancy** check box (selected by default). Select the **Remove Redundancy** check box to show each distinct peptide only once.

This uses the new peptide grouping structure, where multiple peptide hypotheses competing to explain the same physical peptide are formalized into peptide groups. Each group includes only the best evidence for each hypothesis from the spectrum where it was observed.

A new **Peptide Locus** column provides an index for peptide groups with the form x.00y, where x is an integer index for each peptide group and y is the rank of the hypothesis within each group. Clearing the **Remove Redundancy** check box shows multiple acquisitions of the same distinct peptide. In the redundant view, there will be more than one hypothesis with the same **Peptide Locus**. In the non-redundant view, each **Peptide Locus** is shown only once.

New signal processing

The ProteinPilot software has new signal 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.

Multiple Peptide IDs per spectrum Multiple Peptide IDs per spectrum (TripleTOF® systems and QSTAR® systems)

The ProteinPilot software now allows for the possibility of identifying more than one peptide per MS/MS spectrum, resulting in median identification gains of roughly 4 to 8 percent. Gains can be much more or less, depending on the complexity of the sample and the specific nature of the data. Multiple precursor peaks are detected in the Q1 selection window in the survey scan data and the search matches to these multiple observed precursor masses, not just one.

Recalibrated MGF files can now be exported and reused by the Paragon™ algorithm

ProteinPilot software 4.2 Beta made it possible to generate an MGF peak list file after a search that could then be used with other search engines, gaining the benefit of the mass recalibration done during the Paragon algorithm. However, these MGF files could not be handled correctly if reused by the Paragon algorithm. In ProteinPilot software 5.0, this workflow can now be used to avoid repeating the peak list generation stage for subsequent searches. The spectrum identifiers are now parsed to recognize the boundaries between the original data files, which prevents collisions in spectrum identifiers. The identification levels are slightly lower when searching an MGF file versus searching the corresponding .wiff data because multiple-precursor ID cannot function on MGF peak lists.

Improved .group file size

Smaller .group files require less storage space and enable faster opening, saving, browsing, and exporting. The .group file sizes for searches of data from TripleTOF systems using the ProteinPilot software 5.0 have been reduced by approximately 30 to 40 percent of their size compared to their size when using the ProteinPilot software 4.5 Beta.

New Features tab

The purpose of the **Features** tab is to structure and present relevant evidence that a specific protein feature (for example, biological features such as post-translational modifications and sequence variants) has been detected on a specific protein at a specific site. The **Features** tab focuses exclusively on biological features. Chemical modifications (for example, cysteine alkylations and quantitation labels) and artifact modifications (for example, results of sample workup or instrumental analysis) are not reported, although they are included as counter-evidence, where relevant.

The **Features** tab is structured very similarly to the **Protein ID** tab, using a similar reporting paradigm as that used for reporting protein inference results from the Pro Group algorithm. This includes the formalization of a new feature group concept that captures all relevant sources of ambiguity in protein feature reporting.

As with the **Protein ID** tab, there are top, middle, and bottom sections on the tab. The top section shows a representative of each feature group (each group being a putative feature detection). The middle section shows the details of a given feature group, with the group members on the left and supporting and refuting peptide evidence on the right. The bottom section allows users to drill down to show more detail, either to a protein sequence view or to a fragmentation evidence view.

New ProteinPilot Report combines FDR and PDST Reports

The results from FDR analysis, a large amount of metadata, and the ProteinPilot Descriptive Statistics Template (PDST) have been combined into one report called the ProteinPilot Report, which provides a comprehensive summary of search results. At the conclusion of a search in which an FDR analysis is performed, the necessary ProteinPilot exports (Peptide Summary, Protein Summary, Distinct Peptide Summary, Metadata and Features Export) are automatically written to this new report. The report is automatically calculated and then saved in the same location as the .group file.

There are two versions of the ProteinPilot Report that differ only in whether they include the PDST content. By default, the software creates the light version of the report, which includes the metadata, FDR analysis, and export data, because it is faster to open and smaller in file size.

To access the additional information provided by PDST analysis, locate the full version of the report template **ProteinPilotReport_1.0 full report.xlsx** in the **Program Files\AB SCIEX\ProteinPilot\Help\Report Templates** folder and then follow these steps:

1. Copy the **ProteinPilotReport_1.0 full report.xlsx** full template file and then paste it to **Program Files\AB SCIEX\ProteinPilot\WorkflowDirectory\.**
2. Delete the light template file, **ProteinPilotReport.xlsx**, already present in this folder.
3. Rename the full template file to **ProteinPilotReport.xlsx** so that it is recognized by the software as the active template.

The next search will write to this template. The software does not need to be restarted for this change to be implemented.

New FDR Comparison Template

A new template is available that allows users to quantify and visualize the difference between multiple FDR analyses. The **FDR Comparison Template.xlsx** file is available in the ProteinPilot Help folder. Instructions for use are included in the template file.

Protein Modifications column

A new **Protein Modifications** column shows the position of a biological modification or amino acid substitution, relative to the protein sequence. This is different than the existing **Modifications** column, which shows the position relative to the peptide sequence.

This column can be shown on both the **Protein ID** tab in the **Peptides in Group** table and on the **Protein Quant** tab in the **Peptide Quantitation** table. It is hidden by default on the **Protein Quant** tab and is shown by default on the **Protein ID** tab. The column is also included in the **Spectrum Summary Export**, **Distinct Peptide**, and **Peptide Summary Export** tables.

The values in the **Protein Modifications** column include only modifications that are biological modifications or amino acid substitutions, which is why this column is typically empty in **Rapid** search mode, unless a special factor that involves a biological modification is used, such as phosphorylation emphasis.

Annotations from UniProt

Searches against UniProt databases can now retrieve UniProt annotations associated with the detected proteins. An Internet connection must be available at the time of the search and the uniprot.org server must be online and not blocked by any local firewall rules. UniProt annotations are retrieved during the final stages of the search. They are then written directly into the .group file so an Internet connection is not required to display the annotations when viewing the .group result afterwards in the ProteinPilot software.

New export formats available

All exports, with the exception of the Metadata export, can be performed from (1) the **File > Export** menu, (2) the **Workflow** task bar, or (3) the command line using the **GroupFileExtractor** tool.

NOTE: The **group2xml.exe** command line tool that shipped with previous versions of the ProteinPilot software has been renamed to **GroupFileExtractor.exe**. This tool has been enhanced to support additional export formats while maintaining the prior functionality and usage of **group2xml.exe**. The **GroupFileExtractor** tool is typically located in the **C:\Program Files\AB SCIEX\ProteinPilot** folder and can be used to convert a .group file to an .XML output file or to generate any of the following export formats programmatically.

For further information, refer to *Exporting and Working with ProteinPilot™ Software Results* in the Help folder.

Data can now be exported in the following formats.

(1) **Distinct Peptide Summary**

The **Distinct Peptide Summary** export provides a summary of all peptide groups for each group that has at least one hypotheses with at least 15% confidence. The purpose is to report a list of all distinct physical peptides that have been detected. Each physical peptide can have multiple precursor charge states observed as detected, separate LCMS map features. To attempt to identify the peptide, each of these LCMS features can have MS/MS spectra triggered.

The process of identification does not always result in a clear single answer, so ambiguity in identification can occur. The purpose of the peptide group is to group all competitive hypotheses for a given physical peptide arising from the following three levels:

- multiple precursor charge states,
- multiple MS/MS spectra of a given charge state, and
- multiple answers for a given MS/MS spectrum.

The **Distinct Peptide Summary** export is unlike the previous **Peptide Summary** export, which exported information only about the peptides claimed by reported proteins. The **Distinct Peptide Summary** export shows all hypotheses that are within the margin of error of being the right answer, regardless of whether they are attached to reported proteins, making this export useful for peptide-centric workflows.

(2) **Spectrum Summary** (as of ProteinPilot Software v. 4.5 Beta)

The spectrum export includes all of the top hits to each spectrum, making it spectrum-centric, rather than protein-peptide ID-centric. This gives the user a way to access all of the information, rather than just the protein-filtered information as is done with the existing **Peptide Summary** export.

(3) **mzIDentML**

The ProteinPilot software now exports search results in mzIDentML version 1.2 candidate format.* This is a standardized format developed by the Human Proteome Organizations (HUPO) Proteomics Standards Initiative (PSI) for reporting parameters and results from mass spectrometric database search engines. Users can most easily export to mzIDentML format from the export options available in the software, but exports can also be done programmatically.
* http://onlinelibrary.wiley.com/doi/10.1002/pmic.201400080/abstract Seymour, Sean *et al*. “A standardized framing for reporting protein identifications in mzIdentML 1.2.” *Proteomics*. doi: 10.1002/pmic.201400080.

(4) **MGF Peak list**

Recalibrated MGF (Mascot Generic File) peak lists can now be exported programmatically.

(5) **Metadata**

This is a new export option that provides metadata based on search criteria (for example, search date, time, other search parameters). Metadata can be exported in .txt or .xml file format.

(6) **Features**

The **Features** export contains all of the detected protein features (for example, post-translational modifications and sequence variants).

Confidence Interval (CI) columns added to Proteins Detected table

The ProteinPilot software now reports a confidence interval for SILAC and other MS-based quantitation workflows, such as SCIEX mTRAQ. Four new column options are hidden by default but can be shown on the **Protein Quant** tab in the **Proteins Detected** table: **Upper CI M:L, Lower CI M:L, Upper CI H:L,** and **Lower CI H:L**. Upper and Lower CI column headings reflect whether the user selected a Heavy (H) or Light (L) denominator.

N-15 universal labeling SILAC support

Quantitation of universally N-15 labeled SILAC duplex samples was introduced in ProteinPilot software 4.2 Beta but required some special configuration. This feature is now available by default and no longer requires any additional steps to perform this type of quantitation analysis.

Peptide Shared status is now indicated correctly when quantitation turned on

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) was introduced in ProteinPilot 4.5 Beta software. However, it was implemented only in workflows when quantitation was turned off. The feature now works for both identification and quantitation workflows.

The **Specific** status of a peptide in the ProteinPilot software is determined using the new peptide grouping analysis. A peptide group with confident peptide hypotheses can come from more than one confidently identified protein group. This means that the physical molecule associated with the peptide group cannot be assumed to report on only one protein because it is not specific to one protein.

The specificity information is also included in the **Distinct Peptide Summary** export in a column called **Specific**. If **Specific**=1, the peptide group points to only one confidently identified protein group (**Unused** >1.3). If more than one confidently identified protein group is indicated, the peptide is considered shared and **Specific**=0. In a peptide group having a specific hypothesis, the specific hypothesis will be marked as 1 and all other hypotheses will be marked as 0. In a peptide group that is not specific, all hypotheses in the group will be marked as **Specific**=0.

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

Can now reconnect raw data to group file

If the data files associated with a group file result have been moved from their original location, the ProteinPilot software shows a dialog when the group file is opened, allowing the user to browse to the new data file location. This allows users to resume viewing the raw data as spectral evidence within the group file.

Panther columns no longer displayed

Panther information (**Panther ID, Biological Processes, Molecular Functions** columns) is no longer available on the **Protein ID** tab, the **Protein Quant** tab, nor in the **Protein Summary** export. This has been replaced by the UniProt annotation information, which is more extensive.

Improved Mascot signal processing

The ProteinPilot software now uses parallel signal processing for Mascot, resulting in significantly faster searches.

PEFF (PSI Extended Fasta Format) database searching

The ProteinPilot software now supports the PEFF database format. PEFF format provides improved functionality over traditional FASTA searching. PEFF databases contain a file header that clearly describes the file (including the database name, description, source, version, date) and includes a standardized definition line that eliminates FASTA-related parsing problems.

New licensing options

Additional licensing options allow the flexibility to select a 2, 8, 16, or 32-core license, or to upgrade and convert an existing 2 or 8 core license to a 16 core license.

Column display preferences are now retained

Changes to column order, column width, and **Show/Hide** column preferences are preserved between user sessions. Refer to the Help for updated procedures, including how to change and reset column preferences.

New Features as of 4.x Beta Releases

Additional instruments supported

In v. 5.0

- AB SCIEX 6600 TripleTOF® System

As of v. 4.5 Beta

- 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.

Improvements in speed and scale

The ProteinPilot 4.2 Beta software release focused on improving search speed and results manipulations, as well as the upper limits of the scale of datasets. The following speed and scale improvements were made:

- Thorough search speed has been improved. For large searches, the gain approaches the upper limit of 2-fold improvement.
- The Pro Group™ algorithm was partially parallelized to take advantage of multiple cores. Previously, the protein inference stage ran on one core only, regardless of license or available hardware. Enabling at least part of this stage to run on multiple cores reduced the time cost.
- The Pro Group™ algorithm is now used twice in a Thorough search: at the end and now also after the initial fraglet stage-search. This allows for better determination of the protein probabilities used to influence the taglet stage search. This also yields speed gains on larger databases.
- The speed of the false discovery rate analysis stage has increased substantially. Previously, the time to write results to Excel caused this stage to be slower. The native writing speed of Microsoft Excel is faster now. In **Rapid** searches, this improvement can have a dramatic impact on overall search time.
- Results file opening, browsing, and exporting results have been improved and the upper limits of scale have been extended.

- The size of temp files generated during a search has been reduced several fold. Previously, the amount of free hard disk space on the drive where the temp files were written created limitations as to how much data could be successfully processed. This caused failed searches or results that appeared complete, but did not actually include all spectra submitted for analysis. The efficiency of storing these files has been improved, making this much less of a risk. AB SCIEX still recommends having 200 GB of free space for any large scale work to be safe.
- The peptide export is now filtered to >15% confidence by default. This reduces the size of this export by as much as 10-fold in some cases and has little impact on loss of real ID information. The main benefits of this change are that the export will be faster, and most importantly, larger datasets can be analyzed with the ProteinPilot Descriptive Statistics Template. To alter this 15% cutoff, open **ProteinPilot.exe.config**, which is typically in C:\Program Files (x86)\AB SCIEX\ProteinPilot\ and then change this setting:
  <add key="PeptConfThreshOnPeptExport" value="0.15"></add>
  in the config file from 0.15 to another value. Save the file and then restart the software.

Many Additional Workflows Supported

ProteinPilot 4.2 Beta software introduced numerous additional workflows that are directly supported. For further information, refer to the *Paragon Method Settings Guide* in the Help folder. This document lists the current testing status and degree of optimization for any setting. For a complete list of all changes, refer to the header of **Parameter Translation.xml**.

- Performance improved for Phos-Tyr affinity special factor

- Triplex SILAC (Lys+4, Arg+6),(Lys+8, Arg+10)

- Triplex SILAC (Lys+4, Arg+4),(Lys+8, Arg+10)

- Triplex SILAC (Lys+4D, Arg+4),(Lys+8, Arg+10) (4xD for +4)

- Triplex SILAC (Lys+4, Arg+6),(Lys+8, Arg+10)

- Triplex SILAC (Lys+4D, Arg+6),(Lys+8, Arg+10) (4xD for +4)

- Triplex SILAC Arg +0,+5,+10

- Duplex SILAC Lys+4, Arg+10

- Duplex SILAC Lys +8, Arg +10, Glu+6, Asp+5

- Dimethyl +0,+4 (Peptide Labeled)

- Dimethyl +0,+6 (Peptide Labeled)

- Dimethyl +0,+8 (Peptide Labeled)

- Dimethyl triplex +0,+4,+8 (Peptide Labeled)

- Duplex RABA +0, +6 (Reductive Alkylation By Acetone)

- Steen lab 5plex SILAC per ASMS2011 presentations added as "TestCHB_1"

- Trypsin digestion definition to support MSIPI database's use of J as a delimiter

- Digestion with thermolysin

- Iodo-thyroxine modification enrichment set

- TMT cold, 2plex, and 6plex (ID only – not quant)

- Protein-level TMT labeling (ID only – not quant)

Extreme peptide ratios are now excluded from the protein average

The 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 might be extreme ratio proteins.

(5) If the proportion of extreme peptides is less than 15%, then exclude all of the peptides with extreme ratios.

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

Blank ratio cells appear in the **Peptide Quantitation** table of the **Protein Quant** tab, 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). Users can change these settings in the ParameterTranslation file (in the C:\Program Files\AB SCIEX\ProteinPilot\WorkflowDirectory):

```
<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" msmstolsdratio="2.75"
mstolsdratio="7" type="dalton" squareroot="true"/>
          <MSTOLERANCE VALUE="0.05" TYPE="dalton"/>
</INSTRUMENT>
```

MGF Support Improvements

Support of workflows using MGF peak lists was improved in ProteinPilot 4.2 Beta software:

- Very large MGF files can now be handled and large MGF files are searched faster. As with TripleTOF 5600 wiff files, large MGF files are now automatically split into sub-file components for parallel processing on multi-core computers (if allowed by license).

- False singleton ratios issue with SCIEX iTRAQ data solved. False singleton ratios when analyzing SCIEX iTRAQ data from MGF peak lists have been reduced (as of 4.2 Beta). A bug was reported by multiple users where real reporter signal was missed, resulting in false extreme ratios.

- MGF peak lists created for submission to Mascot now convert all fragments to their 1+ equivalent *m/z*, that yields slightly better results. This is also how MGF peak lists are exported with the new export button. The actual observed charge state is noted in these peak lists and currently ignored by Mascot but used by the ProteinPilot software to show fragmentation evidence correctly at the observed m/z, not the 1+ equivalent.

- The retention time is now read from the MGF, if written using the standard RTINSECONDS line.

Improvements to Survey-Level Quantitation Analysis

ProteinPilot™ 4.1 Beta software improved quantitative workflows that use the survey level information like SILAC in several ways:

- Support for triplex MS-based quantitation schemes, included triplex SILAC and triplex mTRAQ schemes. Only identification is possible with MGF input data.

- Apex intensity is used for MS-based quantitation, which reduces variance. Previously, the closest survey spectrum to an identified product spectrum was used, which tended to use data removed from the apex of elution where the signal is not as good as at the apex.

Improvements to Handling Large-Scale Data

In the ProteinPilot 4.0 software, there were many observations of .group files that could not be opened. This occurred for result files greater than approximately 4 GB to 10 GB. Version 4.1 was improved to be able to open all previously observed failing cases.

Issues Resolved

Resolved in version 5.0

Peptide Summary Export reflects new export ratios, area values, and background correction status

ProteinPilot 4.5 Beta software users reported that changing the **Background Correction** status did not affect the **Peptide Summary Export** values and that original values remained. Correct information (export ratios, area values, and background correction status) is now shown.

MS data file name used in Mascot search results

The **MS Data File** field lists the data files used in Mascot searches. Data file names indicate the raw data file format (for example, mrf, wiff, mfg, toftof). If multiple files are used, then the data files are listed in the queue search order.

Resolved as of 4.x Beta Releases

False extreme ratios with SCIEX 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 SCIEX iTRAQ reagent quantitation using MGF peak list input. ProteinPilot 4.2 Beta software 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 begun, resulting in the error message: “Object Reference not set to an instance of an object.”

Dimethyl quantitation workflows supported

Dimethyl quantitation workflows were not possible without manually swapping in additional files. These sample types now work directly with no changes needed.

Database issues resolved

Database support fixes were provided for ENSEMBL and MSIPI databases.

Notes on Use and Limitations in version 5.0

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

Software is not backward compatible

ProteinPilot 5.0 software is not backward compatible. If users want to view .group files from previous versions of the ProteinPilot software, then the files must be viewed on a computer running a viewer (an unlicensed installation of the software) or a previous licensed version of the ProteinPilot software. It is recommended that previous searches still required by the user for active work are run again using the ProteinPilot 5.0 software.

Working with different versions of the ProteinPilot software

Two different versions of the ProteinPilot software cannot be run at the same time on one computer. However, while it is not recommended, different versions of the ProteinPilot software can be installed and be operational on the same computer. Use the following procedure to install two versions of the software on the same computer.

1. On the computer that has an older version of the ProteinPilot software, navigate to the **Program Files** folder.
2. Locate and copy the **ProteinPilot** program folder.
3. Name the copied ProteinPilot program folder (for example, ProteinPilot4).
4. Use **Add/Remove Programs** to uninstall the older version of the ProteinPilot software.
   NOTE: The copied program folder remains on the computer.
5. Install the new version of the ProteinPilot software, following installation instructions.
6. Technically, two versions of the ProteinPilot software are now installed and operational on the same computer. The older version of the software can be launched directly from the folder or a desktop shortcut can be created to launch it.
   NOTE: Two different versions of the ProteinPilot software cannot be run at the same time on one computer.
7. When the older version of the ProteinPilot software is no longer required, make sure to delete the copied ProteinPilot software program folder manually.

License Issues

- The following message might appear: “The license for this feature is either invalid or expired.” This can happen when the network connection of the computer is disabled. The ProteinPilot software license requires that the computer’s network adapter be enabled, although a physically connection to the network is not required. 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 then click **Enable**. Close the ProteinPilot software and then restart.
- For trial versions of the ProteinPilot software, users are not able to perform additional searches after the license expires. However, results files can still be opened. The same limitations apply to an expired time-limited license. In this way, the software is functioning as a viewer only.

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 users do not install the ProteinPilot software on a computer that is also an acquisition station. If the software is installed on an acquisition station, do not process data with the ProteinPilot software at the same time as data acquiring data. Searching while the instrument is acquiring data can cause problems with data acquisition as well as greatly slow the search speed.
- If there is a need to search large datasets regularly (for example, large sets of cation exchange fractions) or work with data from TripleTOF® systems, make sure that the 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 8-core or a 16-core license type and additional hardware as described in Hardware Requirements on page 2.
- Searching input files that are over 8,000 MS/MS spectra on a computer with multiple cores can also result in faster searches of single files when using .wiff or MGF input because large files are automatically split into segments and processed in parallel on separate cores. This feature is not available with TOF/TOF input data.
- The ProteinPilot software requires a minimum of 1.5 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 the system is below the minimum 1.5 GB installed RAM per core, then adding more RAM can increase search speed on multi-core computers.
- The ProteinPilot software disables sorting on the Spectra tab at 300000 or more spectra, to enable larger data sets to be loaded and to improve performance. This is by design.

- Open results files and other applications use RAM. Before starting a large search, to make sure enough memory is available, close any open results files and other applications using any significant amount of RAM and then restart the ProteinPilot software.
- Running many searches or opening many results files at the same time might cause the computer to run out of memory and the application to stop responding. If this occurs, make sure to open fewer results files at the same time and do not open results files while running a search. Restart the computer before running any additional searches
- If working with a TripleTOF® system or other very large data sets with the 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 to have 200 GB of free disk space on the hard disk used for the ProteinPilot software data (typically C:\AB SCIEX\ProteinPilotData). This is where the Temp folder is stored but does not need to be where the final result file is saved.
- If performing a search with data from the TOF/TOF™ Series Explorer™ software or the 4000 Series Explorer™ software and the Analysis Log window shows the message “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, the ProteinPilot software shows an error message, but sometimes it does not. For information on how to resolve this issue, refer to Troubleshooting search errors with TOF/TOF data.

MGF files without correct retention times result in incorrect peptide groups

When searching MGF files that do not include retention times or that include retention times in an unexpected location (based on the MGF format specification), the peptide grouping functionality does not work correctly. Because the FDR analysis at the distinct peptide level operates at the peptide group level, the yields can be inflated and the FDR measures in this situation should not be considered reliable. It is recommended that an MGF creation tool is used that writes the RT in the expected RTINSECONDS location.

Troubleshooting search errors with TOF/TOF data

If performing 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, the ProteinPilot software shows 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 for 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, click the row for the appropriate server and then click **Test**.

   If no error is shown, 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.

Searches can stop running if computer is in Sleep mode

To prevent ProteinPilot software searches from being interrupted by Sleep mode, change the power settings of the operating system.

Issues with Mascot Searches

Searches occasionally fail and show the message “The operation has timed out.” in the Analysis Log window. This can happen when several searches are in the queue. Repeat the search when the queue is empty. (12081)

TOF/TOF™ System Issue with Mascot and Paragon™ Algorithm Searches

A different MS spectrum might be shown from the one shown 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. (12882)

TOF/TOF mzIdentML exports result in partial database submissions

MzIdentML exports from searches containing TOF/TOF data do not contain the spotLabel value in the spectrumID attribute of the SpectrumIdentificationResult element. Because the spotLabel value is excluded, any TOF/TOF mzIdentML submissions to the PRIDE (PRoteomics IDEntifications database, European Bioinformatics Institute) database are considered partial submissions.

Errors in Paragon™ Algorithm Methods

The following error message might appear when a Paragon algorithm is being edited: “One or more errors have been found in this method,” followed by a list of errors. This can happen if the 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 18.

Searching Nucleotide Databases

The Paragon algorithm cannot search a nucleotide FASTA file directly (for example, a DNA or an EST file). If a user attempts to search a nucleotide FASTA file, the program will stop responding. To search a nucleotide database, 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. (194, 373)

Updating FASTA Files for Use by Mascot

If updating 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 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]

Searching with Multi-Period .wiff Files

The ProteinPilot software cannot search multi-period .wiff files. If processing a multi-period .wiff file, an error message is shown when the program begins the search.

Searching Infusion Data

The 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 processing QSTAR System data and many errors in peak detection and precursor charge determination are occurring, the **Bins to Sum** parameter during acquisition might have been higher than recommended.

QSTAR System data that is excessively binned during acquisition does not produce good results with the ProteinPilot software. The data is excessively binned if peaks have just 1 or 2 points above half height. This excessive binning causes errors in the precursor charge determination of the software.

Binning is controlled by the **Bins to Sum** acquisition parameter. On a QSTAR system equipped with a four channel time to digital converter (TDCx4), it is recommended that **Bins to Sum** be set to 1. On a QSTAR system equipped with an eight channel time to digital converter (TDCx8), it is recommended that **Bins to Sum** be set to 4 or 6. If **Bins to Sum** was set too high, consider reacquiring 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 of the .wiff data to the local computer before using it with the ProteinPilot software. If users attempt to process .wiff data or save a result to a network drive rather than the local hard drive, then the processing takes much longer or it can fail. (410)

Species-Specific Searches Unreliable with Some FASTA and ENSEMBL Types

Some FASTA files have more than one species listed for a given sequence entry, including the version of SwissProt available from NCBI. If users 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 the UniProt KB/Swiss-Prot FASTA file is searched, species-specific searches will correctly report all matches. [1649]

Another species-specific failure can occur with ENSEMBL databases. ENSEMBL does not have a species filter in the protein description, therefore Paragon cannot parse it. If users specify a species in a Paragon method and then search against a species-specific ENSEMBL database, the results are invalid. (TT35158). To avoid this issue, search using **None** as the Paragon method species.

UniProt KB/Swiss-Prot FASTA issues can occur because of firewall settings

Firewall settings can prevent UniProt annotations from being acquired if communication with UniProt.org is blocked. If this occurs, change the Windows Firewall settings to allow the ProteinPilot software through the firewall.

UniProt annotation retrieval can fail

Changes to how UniProt provides annotations or unexpected information in existing annotation retrieval mechanisms might cause annotation retrieval to fail.

Working with the Pro Group™ Algorithm Results

- If zooming in on the graph showing spectral evidence in the Fragmentation Evidence pane, then the highlight bars corresponding to the m/z for theoretical ions might appear slightly offset from the peaks. This is a drawing artifact and can be ignored. The matching is performed correctly.
- If zooming in repeatedly on any graph showing spectral evidence, eventually an error message might appear and a red X might replace the spectrum. If this occurs, close the result and then reopen it to show the spectral data again.

Problems with Automatic Updates to Windows

The 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, then the search is terminated and no results are saved. The search must be run again. This issue can be prevented as follows:

– Do not run searches when the system is scheduled for updating.

or

– Change the 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**.

Excel Limitations for Exported Results

If exporting the Peptide Summary from a search with a large number of results, the file might be too large to view completely in Excel. Excel 2010 has a limit of 1,048,576 rows, so no results are shown beyond these rows. If this happens, open the file in a text editing program like Notepad and then divide it into smaller files.

Tool Tips in the Add TOF/TOF Data Dialog

- In the **Add TOF/TOF Data** dialog, users might notice 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 shown 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, "0" might be shown 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, the ProteinPilot software identifies and then processes the correct parent job run during the search. (The number "0" might also be shown as the number of Completed Items when no data was collected.)

Support for Additional Workflows

If there is a workflow that is not supported, then contact us. AB SCIEX is always working to improve the ProteinPilot software. A solution might be able to be provided prior to the next release of the software.

Where to Get Help

Customer Training and Documentation

Use the ProteinPilot™ Software Help to help learn about the ProteinPilot software. Online help is available from the **Help** menu in the program. Pressing **F1** also opens the Help.

The following documents provide additional information about the 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.
- *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 ProteinPilot™ Software* – 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 protein grouping.
- *Unified Modification Catalog* – Detailed information about the modifications used in Paragon™ algorithm searches.
- *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 or a batch file. Files for an example search are included on the installation DVD.
- *Exporting and Working with ProteinPilot™ Software Results* – Instructions for converting the .group file containing the Pro Group results to XML, mzIdentML, or MGF peak list format. Example files are included on the installation DVD.

Contact Us

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

Document Revision History

Revision	Description of Change	Date
A	First release	September 2014
B	Added two Release Notes items to Known Issues and Limitations—TOF/TOF mzIdentML exports result in partial database submissions and UniProt annotation retrieval can fail	October 2014
C	Updated branding, updated computer information, changed Microsoft .NET Framework 4.0 to Microsoft .NET Framework 4.5	November 2015

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 use is intended only to designate those manufacturers' products as supplied by 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.

SCIEX warranties are limited to those express warranties provided at the time of sale or license of its products, and are SCIEX's sole and exclusive representations, warranties, and obligations. SCIEX makes no other warranty of any kind whatsoever, expressed or implied, including without limitation, warranties of merchantability or fitness for a particular purpose, whether arising from a statute or otherwise in law or from a course of dealing or usage of trade, all of which are expressly disclaimed, and assumes no responsibility or contingent liability, including indirect or consequential damages, for any use by the purchaser, or for any adverse circumstances arising therefrom.

For research use only. Not for use in diagnostics procedures.

AB Sciex is doing business as SCIEX.

AB Sciex Pte. Ltd.
Blk 33, #04-06
Marsiling Ind Estate Road 3
Woodlands Central Indus. Estate
SINGAPORE 739256