Compound Discoverer 3.2 Release Notes

These release notes briefly list the new features in the Thermo Compound Discoverer™ 3.2 application, a qualitative data-processing application for small molecule research. Also included are known issues in the 3.2 release of the application and defects that are fixed in this release.

Contents

For information on installing the Compound Discoverer software, refer to the instruction card in the full-version box. For information on licensing, configuring, and using the Compound Discoverer 3.2 application, refer to the user guides and tutorials available as PDF files or the Help.

Features

The Compound Discoverer 3.2 release contains the following new features and enhancements.

New features

- Processing capability for GC-Orbitrap data
  - NIST library searches for GC data.
  - Deconvolution node for GC EI data processing.
  - Deconvolution node for GC CI data processing.
  - Compatibility with most of the existing peak area refinement, identification, pathway mapping, and post-processing nodes—including statistical analysis, biological pathway mapping, QC-based normalization, formula prediction, FISh scoring, and ChemSpider searches.
- Neutral loss searches
  - New workflow node for automated neutral loss searches.
  - New table of neutral losses under Lists & Libraries with 60 predefined neutral losses. You can edit these entries and add your own custom entries.
- Custom tags for annotating and filtering result table entries. These tags are available as a new result table column in most of the result tables.
- Support for importing structures from SDF files into the Mass lists and Expected Compounds lists under Lists & Libraries.
- Support for importing the information for a study from an Excel spreadsheet, including the names and location of the input files, their sample types, and their study factor values.
- New Apply Missing Value Imputation node that is available as an alternative to the Gap Filling node. The Apply Missing Value Imputation node is faster than the Gap Filling node, and because it does not revisit the initial raw data stream to impute area values for missing chromatographic peaks, you can add it as a downstream node when reprocessing an analysis.

XCALI-98189, Rev A

thermoscientific

- New Apply SERRF QC Correction node for QC-based correction of batch effects that uses the Systematic Error Removal using Random Forests method described by Fan[1] and others. You can use this node to correct for batch effects across a single sequence run or multiple sequence runs that use the same chromatography and data-acquisition parameter settings.
- Independent generation of Xcalibur inclusion lists by using a dialog box that you can open from the Lists & Libraries > Expected Compounds view.
- New Search mzCloud™ node features:
  - The Mass Spectrum view displays the noise threshold that the search applies to the fragmentation spectra. It also displays the fragments that the search node does not use to calculate the match score in gray.
  - Reports confidence values for the spectral match in the main compounds tables and the related mzCloud Results table for a compound.
- Installation of the mzVault 2.3 application for creating and editing mass spectral libraries.

Note The mzVault 2.3 application supports importing NIST MSP and MassBank MB files.

Enhancements

- Improvements for Molecular Networking
  - The new interactive behavior for the molecular networks viewer lets you copy the information for selected nodes (compounds) in the molecular networks viewer to the Clipboard, and then use this information to mark compounds in the Compounds table of a result file with custom tags.
  - A new display option for the nodes displays each node (compound) as a pie chart of the relative peak areas for the input files or study groups.
  - A new display option for the node connections displays the node connections as arrows that indicate the direction of the transformation reactions.
  - Structures with charge are now supported in the Expected Compounds workflow.
- The new Annot. ΔMass [ppm], Annot. ΔMass [Da], and Annotation MW columns in the Compounds table and Expected Compounds table show the mass difference between the calculated neutral mass (Calc. MW) of the compound, which is based on the observed ions, and the structure annotation from a search node or mapping node.
- The new Gap Fill Status column—in the Compounds table, GC EI Compounds table, and GC CI Compounds table—describes how the analysis filled the area gaps caused by missing chromatographic peaks.
- The new MS Depth column in all the main compounds tables displays the depth of the fragmentation scans (MSn where n equals the fragmentation depth) in the input files for an analysis.
- The Compound Annotation Editor supports copying and pasting chemical structures as InChI strings.
- Additional compound databases under Lists & Libraries > Mass Lists:
  - GC Orbitrap Contaminants Library Compound Database
  - GC Orbitrap Flavor and Fragrances Compound Database
  - GC Orbitrap Metabolomics Library Compound Databases
  - Natural Product Atlas version 2020_06[2]
- New version of the mzCloud Offline Spectral Library (2020B) under Lists & Libraries > Spectral Libraries.
- The new Promote command on the Job Queue page lets you change the execution order of jobs that are waiting to be processed.
- The new Study Information table in a result file displays information about the input files for the analysis, including their file names and study factor values.
- The parameters previously provided by the Normalize Areas node have been separated into three nodes—Apply QC Correction, Normalize Areas, and Scale Areas.

---

[1] Fan et al., Systematic Error Removal using Random Forest (SERRF) for Normalizing Large-Scale Untargeted Lipidomics Data, Anal. Chem. 2019

[2] van Santen et. al. 'The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery', ACS Central Science, 2019, 5, 11, 1824-1833. 10.1021/acscentsci.9b00806

- In both QC correction nodes, you can now specify different thresholds for the pre- and post-correction RSD[%] values for the peak areas of particular compounds across the QC samples. When the RSD[%] of the peak areas for a compound fail either of these two thresholds, the node does not report the normalized area for the compound in the compounds table—that is, it leaves the Norm. Area column for the compound empty.
- The new ΔRT [min] column in the related Compounds table for the main mzVault Results table displays the difference in the measured retention time of the compound and the retention time of the matching mzVault library entry.

System requirements

The Compound Discoverer 3.2 application can process data files produced by high-resolution accurate-mass (HRAM) Thermo Scientific™ instruments, such as the Orbitrap ID-X™, Orbitrap Fusion™, Orbitrap™ hybrids, Q Exactive™, and Exactive™ mass spectrometers.

Table 1 lists the system requirements for the data processing computer.

Table 1. Minimum system requirements for the data processing computer

System	Requirements
Hardware	• 3.4 GHz dual-core processor • 16 GB RAM • 500 GB hard drive • DVD-ROM and USB drive • Display monitor resolution of 1920×1080
Software	• Microsoft™ Windows™ 7 Pro SP1 (64-bit) or Windows 10 (64-bit) operating system • Microsoft .NET Framework 4.7.2 and Microsoft .NET Core 2.1 Runtime • Microsoft Office 2013 or later • PDF reader

Table 2 lists the recommended hardware configurations for enhanced performance using the Compound Discoverer application.

Table 2. Recommended hardware configurations

System	Requirements
Hardware	• Dual 8-core processor (for example, 2x Intel™ Xeon™ Gold 6134 CPU @ 3.20 GHz) • 64 GB RAM • 1 TB SSD (solid-state disk) hard drive for operating system • 2nd 3 TB (conventional disk) hard drive for data storage • DVD-ROM and USB drive • Two 27 in. UHD monitors: Display monitor resolution of 3840 × 2160 pixels

Important information

Supported local language and date/time format for the data system computer

The Compound Discoverer 3.2 application was tested and is supported for US-English Only locale settings. To change the format on your computer, go to **Control Panel > Region and Language**. Or, type **Language** in the Windows 10 search box.

Upgrading from Compound Discoverer 1.0, 2.0, 2.1, 3.0, or 3.1

To install Compound Discoverer 3.2 software on a processing computer that has version 1.0, 2.0, 2.1, 3.0, or 3.1, you do NOT need to uninstall the earlier version. Compound Discoverer 1.0, 2.0, 2.1, 3.0, 3.1, and 3.2 can coexist on the same computer. For detailed installation instructions, refer to the Installation Instructions located on the software media.

Using the Search mzVault node

The Search mzVault node does NOT require the mzVault application, so installing the mzVault 2.3 software is optional. However, Thermo Fisher Scientific recommends installing the software to do any of the following:

- Create your own spectral libraries.
- Edit existing spectral libraries (in Library Manager format or mzVault format).
- Convert existing spectral libraries (in Library Manager format or mzVault 1.0 format) for use in the Compound Discoverer application.
- Import spectral libraries in the NIST MSP format (.msp) or MassBank records (.mb).

To use your own mzVault spectral libraries in the Compound Discoverer application, add it to the Lists & Libraries > Spectral Libraries list in the application. Refer to the User Guide or Help for additional information on using your own spectral libraries.

Compound Discoverer 3.2 ships with four mzVault spectral libraries:

- mzCloud Offline for mzVault Reference 2020B (Snapshot of the mzCloud Reference spectral library)
- mzCloud Offline for mzVault Endogenous 2020B (Snapshot of the mzCloud Reference spectral library with endogenous metabolites only)
- mzCloud Offline for mzVault Autoprocessed 2020B (Snapshot of the mzCloud Autoprocessed spectral library)
- mzCloud Offline for mzVault Endogenous-Autoprocessed 2020B (Snapshot of the mzCloud Autoprocessed spectral library with endogenous metabolites only)

These files are local copies of the mzCloud library (MS2 only) as of July 2020. Approximately every 6 months, Thermo Fisher Scientific provides an updated local mzCloud library as an mzVault database file, which you can download from the Flexera™ software download site (thermo.flexnetoperations.com). This local version of mzCloud is only compatible with mzVault 2.3 or later.

Multiple monitor support

Compound Discoverer data review supports multiple monitors. To connect the data system computer to two monitors, use the two display ports on the back of most computers.

Compound Discoverer permanent viewer

A licensed demonstration version of Compound Discoverer becomes a permanent viewer even after license expiration. With the Compound Discoverer viewer, you can open previously processed result files and generate reports. Nodes such as Input Files, Select Spectra, and Export Spectra are still available for use within the viewer, while other license-protected nodes will no longer be available for data processing.

Ensuring access to the online databases

A typical Compound Discoverer analysis that identifies unknown compounds searches mass spectrum databases on the Internet. To run these searches, the application must have unblocked Internet access to these databases. To test the application's access to the online mass spectrum databases, follow this procedure.

❖ To test and troubleshoot the application's access to the online databases

1. Run the communication tests (see "Running the communication tests" on page 5).

   If the communication tests succeed, the application has access to the online databases.
2. If a communication test fails, do the following as needed:
   - If only the mzCloud communication test fails, check the Date and Time settings on the processing computer (see "Setting the correct time and time zone on the processing computer" on page 6).
   - If the Check Subscription test for the BioCyc database fails, check the subscription information in the BioCyc User Login view of the Configuration page.

     IMPORTANT You must create a BioCyc user account and then enter, test, and save your account credentials in the BioCyc User Login view (see "Setting up a BioCyc account or subscription" on page 7).

- If any of the other communication tests also fail, check the access to the URLs for the online databases (see "Checking the URLs for the online databases in your browser" on page 6).

  If you can access the URLs for the online databases through your browser, but the communication tests still fail, the firewall or proxy setting for your company network is blocking the application's access to the online databases.

3. If the communication tests fail, but you can access the URLs for the online databases, do the following as needed:
   - If a firewall is blocking the application's access to the online databases, ask your IT department to make sure that the company firewall is not blocking "Compound Discoverer" or "Compound Discoverer Server" from accessing the URLs. The application uses the following protocol: http port 80 and https on port 443.
   - If a proxy setting is blocking access, see "Specifying the IP address of the proxy server" on page 6.

For details, see these topics:

- Running the communication tests
- Checking the URLs for the online databases in your browser
- Specifying the IP address of the proxy server
- Setting the correct time and time zone on the processing computer

Running the communication tests

Use the Communication Tests dialog box to test your processing computer's access to the online databases.

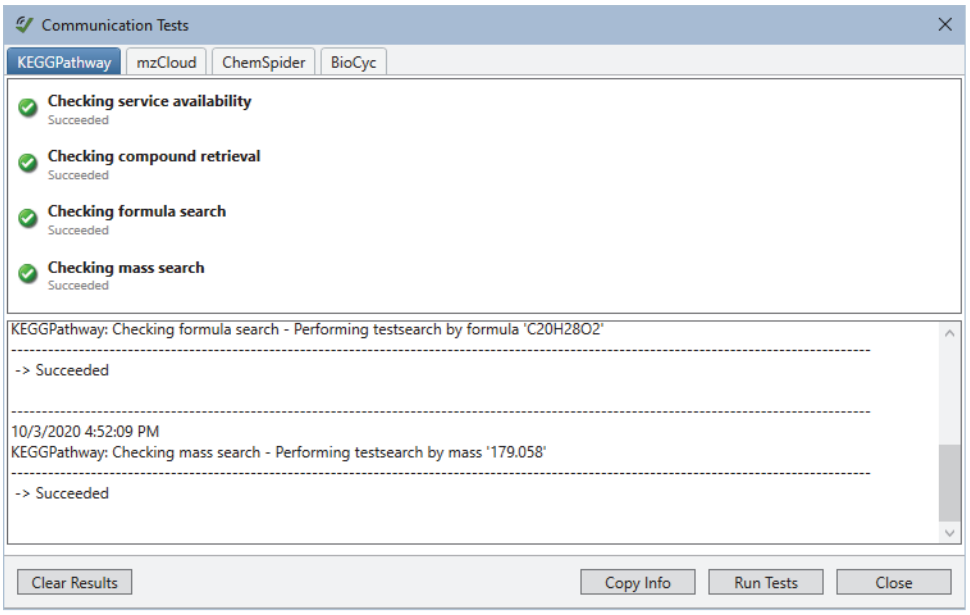
❖ To verify that your computer has access to the external databases

1. From the menu bar, choose **Help > Communication Tests**.

   The Communication Tests dialog box opens.
2. To open the page for the database that you want to access, click its tab.
3. Click **Run Tests**.

   Figure 1 shows the communication tests in progress.

   Figure 1. KEGG Pathway communication tests



4. When the tests are successful, your computer has access to the required databases on the Internet. If only the mzCloud test fails, check the Date and Time settings for the processing computer (see "Setting the correct time and time zone on the processing computer" on page 6). If any of the other tests also fail, check the access to the URLs in your browser (see Checking the URLs for the online databases in your browser).

Checking the URLs for the online databases in your browser

Table 3 lists the URLs for the online mass spectrum databases. If a communication test fails, test the URL for the affected database.

Table 3. URLs of online mass spectrum databases

Database	URL
mzCloud Identity	https://identity.mzcloud.org/
	https://www.mzcloud.org/Services/MzCloudApiV1.svc
	https://www.mzcloud.org/Services/MzCloudApiLightService.svc
ChemSpider	http://www.chemspider.com
	http://api.rsc.org/
KEGG™: Kyoto Encyclopedia of Genes and Genomes	http://www.kegg.jp/
	http://rest.kegg.jp
	https://proxy.online-licensing.net
BioCyc	https://biocyc.org/
	https://biocyc.org/web-services.shtml
	https://pwapi.arabidopsis.org

Specifying the IP address of the proxy server

If the communication tests fail but you can access the online databases through your browser, follow this procedure to specify the IP address of the proxy server.

❖ **To configure the IP address of the proxy server**

1. Go to *drive*:\Program Files\Thermo\Compound Discoverer 3.2\bin\Config.
2. Open the **Proxy.config** file in Notepad.
3. Remove the text that is highlighted in yellow in Figure 2—that is, remove the XML comment delimiters: <!-- and -->.

Figure 2. Proxy configuration setting with XML comment delimiters

```
1  <?xml version="1.0" encoding="utf-8" ?>
2  <defaultProxy enabled="true">
3    <!--
4     <proxy bypassonlocal="true"
      proxyaddress="http://127.0.0.1:8118/" />
5    -->
6  </defaultProxy>
```

4. Replace the text that is highlighted in yellow in Figure 3 with your company proxy address.

Figure 3. Default proxy address highlighted in yellow

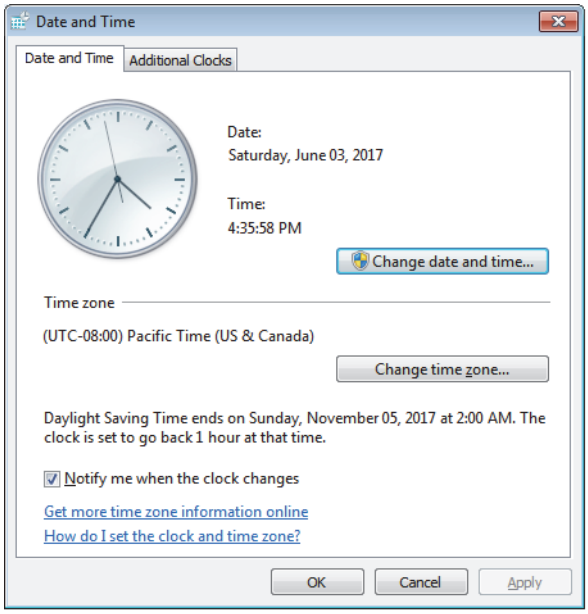
```
1  <?xml version="1.0" encoding="utf-8" ?>
2  <defaultProxy enabled="true">
3    <proxy bypassonlocal="true" proxyaddress=
     "http://127.0.0.1:8118/" />
4  </defaultProxy>
```

Setting the correct time and time zone on the processing computer

The mzCloud communication test includes a validation of the date and time settings on the processing computer. If the mzCloud communication test fails, but the other communication tests succeed, check the date and time settings for the processing computer.

❖ To check the time and time zone settings

1. Open the Date and Time dialog box as follows.
   - For the Windows 7 operating system, open the Control Panel. In the View By list, select **Category**. Then choose **Clock, Language, and Region > Date and Time**.
   - For the Windows 10 operating system, type **Date** in the search box.



2. Make sure that the date, time, and time zone settings are correct.
3. If the Internet Time tab is available, synchronize the Internet time.

> **Note** If your computer is not part of a network domain that synchronizes the computer's clock to the network server, you can use an Internet server to synchronize the computer's clock.

Setting up a BioCyc account or subscription

Follow the instructions in the BioCyc User Login view to set up your BioCyc subscription or user account. You must have an organization subscription or an individual subscription to access all BioCyc databases. A user account allows you to access only the EcoCyc database.

To set up a BioCyc account or individual subscription and enter your credentials in the BioCyc User Login view, follow these procedures as needed:

- To open the BioCyc User Login page
- To set up a BioCyc account or individual subscription
- To enter, test, and save your account information

> **Note** To set up a user account or an individual user subscription, you must have Internet access.

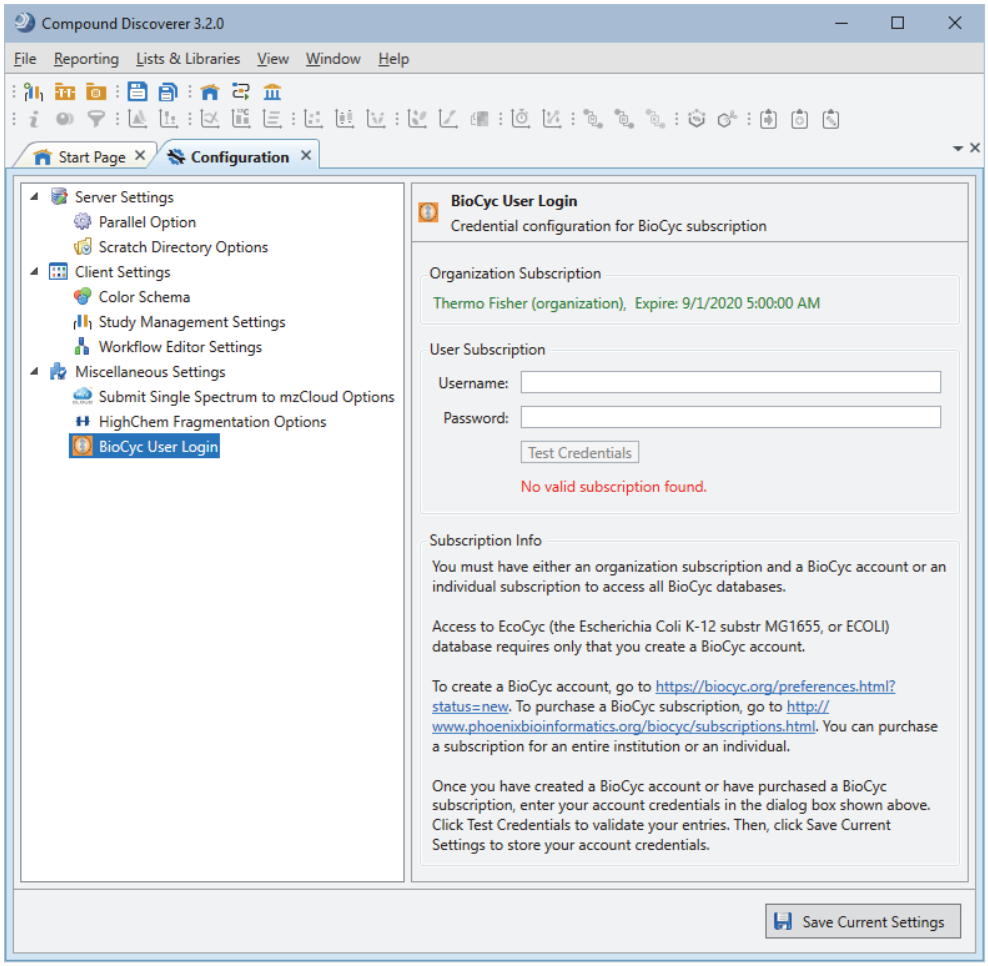
❖ To open the BioCyc User Login page

1. From the application window, choose **Help > Configuration**.

   The Configuration page opens.
2. In the left pane, under Miscellaneous settings, select **BioCyc User Login**.
   - In the Organization Subscription area, if you have a subscription, the organization name appears in green. Otherwise, this area displays, "No Valid Subscription Found" in red.
   - In the User Subscription area, if you already created a BioCyc user account, entered and tested your credentials, and saved the settings, the User Name box displays your email address. Otherwise, the following text appears in red below the Test Credentials button: No Valid Subscription Found.

   Figure 4 shows a BioCyc User Login page for a user without a subscription.

Figure 4. BioCyc User Login page for a user without a subscription



❖ To set up a BioCyc account or individual subscription

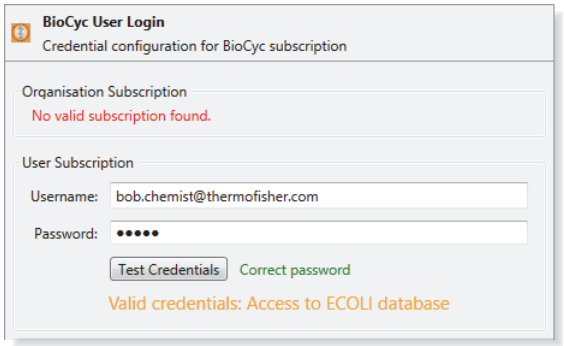
Follow the instructions in the BioCyc User Login view.

❖ To enter, test, and save your account information

1. In the User Subscription area, type your email address in the User Name box and your BioCyc password in the Password box.
2. Click **Test Credentials**.
3. Click **Save Current Settings**.

Figure 5 shows the subscription information if you have a user account but do not have an organization subscription or an individual subscription. When running analyses that map compounds to the BioCyc pathways, you can access only the ECOLI database.

Figure 5. Settings for a user account only



Obtaining a KEGG license

To use KEGG Pathways in the Compound Discoverer application, you must install a valid KEGG license on your processing computer. To license the KEGG Pathways module, contact Pathway Solutions for the activation key. After you obtain the activation key from Pathway Solutions, install it on your processing computer by using the Compound Discoverer License Manager.

Note You must contact Pathway Solutions for a KEGG license. Kanehisa Laboratories does not provide KEGG licenses for the Compound Discoverer application.

To obtain the activation key for the KEGG Pathways module, contact Pathway Solutions at the following email address:

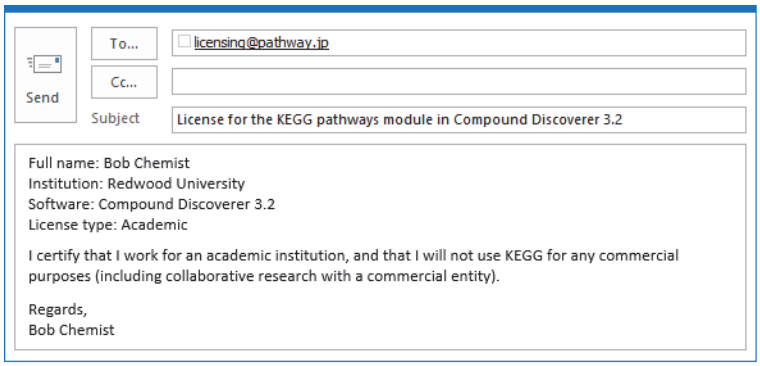
licensing@pathway.jp

Include the following information in your email:

- Your full name: First and last name
- Institution: Institution name
- Software: Compound Discoverer
- Version number: 3.2
- License type: Commercial or academic*

*If you are applying for an academic license, you must include the following statement in your email:

I certify that I work for an academic institution, and that I will not use KEGG for any commercial purposes (including collaborative research with a commercial entity).



Known issues

Suggested recovery actions

- For some issues, restarting the application or data system computer is the appropriate recovery action.
- As a fix we generally do not recommend reinstalling the software or the operating system, which more commonly occurs after you install a new hard drive.

Feature requests and other removed items

- We do not include issues where there is insufficient information logged to successfully reproduce the reported problem.
- We do not list feature requests as software issues, regardless of the reported significance or severity of the request. Product managers evaluate logged feature requests for future releases.
- We report only discrepancies in the documented software as known issues.

Terminology tables

Severity	Interpretation
Critical	A problem that renders the system unusable because either an entire function is unusable and no workaround exists, or use of the current system compromises data integrity or results in data loss. Catastrophic problems also include significant and non-obvious quantitative errors, and all human and instrument safety issues.
High	A serious issue that does not affect data integrity (meaning data loss, corruption of data, or the wrong answer), but affects the customer's ability to use the product as designed. It can be a failure, design issue, or documentation error or omission. A workaround might or might not exist.
Medium	A minor error or poor behavior of a product feature. There is probably a workaround.
Low	An issue that has a limited effect on customer usage of the product; for defects with visibility so low that a customer might never see it; or for ease of use issues or other items not causing any performance degradation.

Risk	Interpretation
High	Occurrence is likely to happen and can compromise operation.
Medium	Occurrence is uncommon, but if it occurs, can compromise operation.
Low	Issue is minor; however, the software might operate differently from a user's expectations. A workaround is often available.
No Risk	This issue causes no problems but is commonly an inconsistency or a cosmetic issue.

Known defects

Table 4 contains known defects in the software, categorized by software section, with a brief abstract and information related to each defect's severity and risk. The Item ID is the internal number assigned to each issue. Product management assesses risk, which can differ significantly from the reported severity.

Table 4. Software defects (Sheet 1 of 2)

Software section	Severity	Abstract	Risk	Item ID
Alignment	Medium	Inadequate alignment can cause reported peak areas being too low.	Medium	100374
Alignment	Medium	Insufficient alignment for samples with low feature density.	Medium	100391
Compound Editor	Low	Atom Properties dialog box | Periodic Table | Terrestrial Isotopes: The mass numbers displayed as superscript are not aligned for some elements.	No Risk	101420
Find Expected Compounds	Medium	In chromatograms with many adjacent peaks, some smaller peaks are not detected.	Medium	100416
Job Queue	Medium	When expanding the details on a job a red X is displayed and an error message is shown (does not affect data processing).	Medium	100544
mzCloud search	Medium	Some compounds are not identified when DIA scans are used for the search.	Medium	100449
mzVault Library	Low	Spectral library—Bamba Lab 538 polar metabolite library has all ID=1 for mzCloud.	Low	100529
Reporting	Medium	Displays the MS2 spectrum closest to the peak apex but does not support displaying the entire spectral tree for a compound.	Medium	101431

Table 4. Software defects (Sheet 2 of 2)

Software section	Severity	Abstract	Risk	Item ID
Reporting	Low	Report Template View: When you delete multiple items, part of the warning message is cut off.	No Risk	100858
Reporting	Low	The structure annotations on MS/MS spectra are not clear in the PDF report.	No Risk	100859
Reporting	Medium	HTML file is not generated for a report that contains 1000 or more entries.	Low	101435
Reporting	Medium	Color-coding can be inconsistent between XIC overlays in the Compound Discoverer result view.	Low	101441
Reporting	Medium	Unable to report the expected compounds associated with radio trace in RT order.	Low	101021
Scatter Chart	Low	The tooltip for the data point is not displayed after docking the scatter chart in the data review window.	No Risk	100860
Detect Compounds node	Medium	Ambiguities during component assembly can cause erroneous assignment of M+NH4 adduct.	Medium	100216
Unknown Detector	Medium	D-labeled compounds are not detected in certain cases.	Medium	101467
Unknown Detector	Medium	Two M+H adducts are grouped into one hit.	Low	101425
Unknown Detector	Medium	Unknown Detector occasionally assigns A1 isotope peaks as A0.	Medium	101430
Unknown Detector	Medium	Adduct grouping issue causes duplicated compounds in the Compounds result table.	Medium	101442
Unknown Detector	Medium	Unknown Detector does not correctly assemble the isotopic pattern for Boron-containing compounds.	Low	101439
Unknown Detector	Medium	[M+H]+ is assigned as [M+NH4]+ for amine compounds, for example, amino acids.	Medium	101450

Resolved defects

These defects were resolved between the Compound Discoverer 3.1 SP1 and 3.2 applications. Both an engineering fix and follow-up testing have resolved each of these issues:

Table 5. Resolved defects (Sheet 1 of 2)

Item ID	Abstract	Severity
101422	The Compound Editor does not support charged compounds.	3 - Medium
100198	Loading a result filter file causes an unhandled exception.	3 - Medium
100287	When reprocessing an analysis with the Search mzVault node, the analysis validation sometimes fails when the user adds new mzVault libraries to the search.	3 - Medium
100837	The Compound Editor and the Compound Annotation Editor do not support structures that are energetically unstable, such as isocyanate.	3 - Medium
108272	In the mzCloud Search | Mirror plot, some matching spectral peaks are shown in red even when the query and library spectral peaks are within the specified fragment *m/z* tolerance.	3 - Medium
108471	The Export to Excel feature fails to export certain subtables correctly.	3 - Medium

Table 5. Resolved defects (Sheet 2 of 2)

Item ID	Abstract	Severity
108527	The Export to Excel feature fails to export some of the columns for the second-level related tables, for example, the mzCloud Results table.	4 - Low
109463	The BioCyc pathway view does not update when you select an Omics data overlay. The data reloads, but the view still displays the same unfiltered information from the compounds table.	4 - Low
114055	When you reprocess an analysis that includes the Differential Analysis node, the application ignores any changes you made on the Grouping & Ratios page of the analysis.	3 - Medium
118840	The Export to Excel feature fails when the compound name includes special characters assigned by a mass list search.	3 - Medium
120927	The KEGG pathways view does not work—that is, when you select a pathway in the related KEGG Pathways table for a compound, the view remains empty.	2 - High
124838	The KEGG Pathways view remains empty when you select a pathway if the URL of the pathway is too long due to the high number of matched compounds in the pathway. The URL of a KEGG pathway is based on the number of matching compounds it contains for the current selection in the compounds table.	2 - High
130907	When you open a result file, the result page is empty when the Windows performance counter is corrupted.	3 - Medium
133858	The formula check in the Compound Editor does not work for manual entries.	3 - Medium
135262	The Export to Excel feature creates invalid MOL strings when exporting compounds with structures to an Excel spreadsheet.	3 - Medium
135778	Compound detection accepts isotopic peak assignments that do not make sense. For example, it accepts compounds with isotope patterns that contain only A0 and A6 ions instead of rejecting these compounds as false positives.	3 - Medium
147764	The Metabolika search fails with an exception when the application is parallel processing more than one analysis that includes a Metabolika search.	3 - Medium
163482	When the application is parallel processing two jobs that contain mass list searches, the second job fails when the two jobs are simultaneously processing the same mass list.	3 - Medium
151228	For columns with multiple subcolumns, the subcolumn headings do not properly align with the column entries on very high resolution(4k) monitors.	3 - Medium

Trademarks