3.1.15 Using Spectral Libraries ................................................................. 21

3.2 Review Perspective .............................................................................. 21

3.2.1 Starting a directDI Analysis ............................................................. 22

3.2.2 Starting a Library base Analysis ....................................................... 22

3.2.3 Using an External Spectral Library for an Analysis ......................... 23

3.2.3.1 Spectral Library Columns .......................................................... 24

3.2.3.2 Modification Parsing ............................................................... 26

3.2.4 Configure Conditions ...................................................................... 27

3.2.5 Analysis Settings .............................................................................. 27

3.2.5.1 XIC Extraction ....................................................................... 27

3.2.5.2 Calibration ............................................................................. 27

3.2.5.3 Identification .......................................................................... 27

3.2.5.4 Quantification ......................................................................... 28

3.2.5.5 Workflow ............................................................................... 28

3.2.5.6 Protein Inference .................................................................... 29

3.2.5.7 Post Analysis ........................................................................... 30

3.2.5.8 Reporting ................................................................................ 30

3.2.6 Tree Filtering ................................................................................ 30

3.2.7 Tree Grouping ............................................................................... 30

3.2.8 Saving/Loading Experiments ......................................................... 30

3.2.9 Run Node Plots .............................................................................. 30

3.2.9.1 iRT Calibration Chart .............................................................. 31

3.2.9.2 XIC Extraction Width Chart ................................................... 31

3.2.9.3 TIC & TIC Overlay Chromatogram ........................................... 32

3.2.9.4 Analysis Log .......................................................................... 32

3.2.10 Peptide and Fragment Plots ........................................................... 32

3.2.10.1 MS2 XIC ............................................................................. 33

3.2.10.2 MS2 XIC Sum ...................................................................... 33

3.2.10.3 MS2 Intensity Correlation ....................................................... 34

3.2.10.4 MS1 Isotope Envelope XIC ..................................................... 34

3.2.10.5 MS2 Isotope Envelope Correlation ........................................... 35

3.2.10.6 MS2 XIC Alignment ............................................................... 35

3.2.10.7 iRT XIC Sum Overlay ............................................................ 36

3.2.10.8 MS2 Intensity Alignment ........................................................ 36

3.2.10.9 Cross Run RT Accuracy ........................................................ 37

3.2.10.10 MS1 Spectrum at Apex .......................................................... 37
3.2.10.11 MS2 Spectrum at Apex .......................................................... 38
3.2.10.12 Protein Coverage ................................................................. 38
3.2.11 Library Refinement ................................................................. 39
3.2.12 Export All ............................................................................... 41
3.2.13 Targeted Analysis of SONAR™ .................................................. 41
3.3 Post-Analysis Perspective ............................................................. 41
  3.3.1 Configure Conditions Editor .................................................... 41
  3.3.2 Post-Analysis Settings ............................................................... 42
  3.3.3 Fractionation ........................................................................... 43
  3.3.4 Analysis Overview ................................................................. 43
  3.3.5 Analysis Details ...................................................................... 47
  3.3.6 Differential Abundance ........................................................... 47
3.4 Report Perspective ........................................................................ 50
  3.4.1 Exporting Data Tables ............................................................. 51
    3.4.1.1 Experiment Report ............................................................ 51
    3.4.1.2 Run Report ........................................................................ 51
    3.4.1.3 Protein Group Report ........................................................ 52
    3.4.1.4 Elution Group Report .......................................................... 52
    3.4.1.5 Fragment Group Report ...................................................... 53
    3.4.1.6 Fragment Report ............................................................... 54
  3.4.2 Exporting Run Pivot Report ...................................................... 54
  3.4.3 Score Histograms .................................................................... 54
3.5 Quality Control Perspective .......................................................... 55
  3.5.1 QC Panels ............................................................................... 55
3.6 Pipeline Perspective ...................................................................... 55
  3.6.1 Pipeline Reporting Settings .................................................... 56
3.7 Databases Perspective ................................................................... 56
  3.7.1 Table Import .......................................................................... 56
  3.7.2 Modifications ......................................................................... 56
    3.7.2.1 Importing Modifications from Search Engine .................. 57
    3.7.2.2 Creating custom modifications ....................................... 57
  3.7.3 Protein Databases ................................................................. 57
  3.7.4 GO Databases ......................................................................... 58
    3.7.4.1 Gene Ontologies ............................................................... 58
    3.7.4.2 Gene Annotations ............................................................ 58
  3.7.5 Digest Rules ............................................................................ 59
1 General Information

1.1 Computer System Requirements

**Minimal**
- Operating System: Windows 7, x64
- CPU: Intel® Core™ CPU, 2.7 GHz (quad-core) or similar
- Hard drive: 200 GB free space
- Memory: 8 GB
- Software: .NET 4.6.1 or higher

**Recommended**
- Operating System: Windows 7 or higher, x64
- CPU: Intel Core i7 4770, 3.4 GHz (octa-core) or similar
- Hard drive: 500 GB free space, solid state drive (SSD)
- Memory: 16 GB or more
- Software: .NET 4.6.1 or higher

One can estimate the recommended amount of total system memory using the following equation:

\[
RAM_{GB} = 4 + \frac{0.85 \times (n + 5000) \times r}{1024^2}
\]

where \(n\) is the number of precursors in the library and \(r\) the number of runs in the experiment. 32 GB of RAM should be sufficient to process 500 runs with a library containing 50,000 precursors.

1.2 Scope of Spectronaut™ Software

Spectronaut™ is a software product for quantitative profiling of 100s to 1000s of proteins. Large experiments with several conditions and replicates consisting of dozens up to 1000s of LC-MS runs can be analyzed. The scope of Spectronaut™ is the targeted analysis of any data independent acquisition (DIA) experiments, especially Hyper Reaction Monitoring (HRM-MS™) data. For a description of HRM-MS™ please visit our webpage or have a look at our publication (Bruderer et al., 2015).

1.3 Spectronaut™ Pulsar

Spectronaut™ Pulsar is a commercial software package available here on our webpage. Spectronaut Pulsar can analyze DIA data sets without any calibration kit. However, the addition of an iRT Kit or the HRM Calibration Kit is still very much recommended as it ensures calibration on difficult matrices and allows for detailed quality control read-outs.
1.4 Spectronaut Release Features

1.4.1 Spectronaut Pulsar

- Support for direct DIA processing
- Integrated Protein-FDR control
- No more iRT-Kit requirement for DIA analysis
- Library generation from DDA/DIA using Pulsar
- Improved analysis processes for centroided DIA
- Support for library based Waters SONAR™ analysis
- Library generation from Mascot™
- GO Enrichment in post analysis
- Improved memory efficiency
- Un-supervised centroid algorithm in HTRMS converter
- Improved UI performance

1.5 Supported Mass Spectrometers

- Thermo Scientific™ Q Exactive™ Series
- Thermo Scientific™ Orbitrap Fusion™ Series
- Sciex TripleTOF® Series
- Bruker impact II™
- Waters Xevo® G2-XS QTof
1.6 Supported Data Acquisition Methods

Spectronaut™ analyzes a large variety of different DIA data. Minimal requirements are a reversed phase chromatography with a linear or nonlinear gradient that spans at least 10-35% Acetonitrile (most of the iRT or HRM kit peptides need to be detectable). Methods acquiring only MS2 scans are supported, as well as methods with both, MS1 and MS2 scans. The cycle time of the DIA method should be in the range of 2-3 seconds depending on the peak width of the chromatography used. MS1, as well as MS2, precursor ranges can be segmented. The MS2 scans should cover at least 500-900 m/z of precursor range. Gas phase fractionation is supported starting with Spectronaut™ 7 (Nimoy). More specifically Spectronaut™ supports HRM-MS™, WiSIM-DIA, AIF, SWATH™(Gillet et al., 2012) and SWATH™ 2.0. Multiplexed DIA is not supported.

In case you experience technical problems with the software or if you have feature suggestions please contact support@biognosys.com.
2 Getting Started

2.1 Getting Spectronaut™ Pulsar

Spectronaut™ Pulsar software licenses are available here on our webpage. Besides the software, additional reference material including raw data and the corresponding spectral libraries containing accurate iRT (Escher et al., 2012) can be downloaded.

2.2 Spectronaut™ Activation

When you start Spectronaut™ for the first time, you will be asked to activate your software. You should have received a license key with the download link for Spectronaut™ Pulsar after purchase. If you do not have a license key yet, contact us at order@biognosys.com.

If your computer has access to the internet, activation will be automatic once you have pasted your license key into the Spectronaut™ activation dialogue. In the event that your Spectronaut™ computer does not have an internet connection or the connection is blocked by a Firewall, you can also activate your software by email. The respective instructions will appear after a few seconds if online activation was not successful. Save the registration information file on your computer and send this file to support@biognosys.com. In general, you will receive a license file within one working day. To activate Spectronaut™ using a license file, click on the “Browse License File…” button in the Spectronaut™ Activation dialogue.

2.3 Calibration Kit

To enable fully automated and sensitive signal processing with Spectronaut™, we developed the calibration kit (iRT kit, HRM calibration kit) that is spiked into each sample before the measurement. The calibration kit contains a mix of synthetic peptides that do not occur in nature. Based on these peptides, Spectronaut™ makes a fast first pass analysis, calibrates important parameters for peak detection (iRT, mass) and can store quality control information available in the QC perspective. The integration of Spectronaut™ with the calibration kit translates into better results:

- **Smart in-run parameter calibration** guarantees quick and robust analysis
- **Increases sensitivity and specificity** by iRT and advanced handling of interfering signals
- **Higher quantitative accuracy** by supporting optimized ion current extraction
- **Automated quality control** by monitoring LC and MS performance

The use of calibration kit (iRT kit or HRM kit) is highly recommended to perform any analysis with Spectronaut™.

For more information about the calibration kit, please see here.
Figure 1. Quality control perspective of Spectronaut™ based on the peptides in the iRT kit or HRM calibration kit. Chromatography, mass spectrometer performance and analysis can be monitored in every run and over time using several performance indicators.

2.4 Spectral Library Guidelines

2.4.1 General

The generation of spectral libraries in the prepare perspective of Spectronaut™ from search results of MaxQuant, Proteome Discoverer or ProteinPilot is recommended. Alternatively, spectral libraries can also be imported when starting an analysis. If an external spectral library is used, please consider following recommendations to ensure best performance:

- Keep peptide precursors non-redundant in the spectral library
- Do not keep more than 15 most responsive fragment ions per peptide precursor
- Precursors with less than 3 fragment ions should not be considered
- Please see below for the recommended format of an external spectral library

2.4.2 iRT

We recommend spiking the iRT kit into the samples used for spectral library generation. However, for the most common sample types this is not always strictly required. If a spectral library is generated within the Spectronaut™ prepare perspective, Spectronaut™ will take care of calculating iRTs for all peptide detected in the database search. If an external spectral library is used, please consider following workflow:

- Spike the iRT kit or HRM calibration kit into your sample for shotgun analysis. The iRT kit is contained as a subset of the HRM calibration kit.
- Use the iRT peptide FASTA file in the shotgun search together with your FASTA file. The iRT peptide FASTA fusion file can be downloaded from the Biognosys webpage.
- Check whether the iRT peptides were identified in your search. If they are not this could be an indication of a low quality measurement. The iRT peptides are relatively high in concentration and most, if not all, should be easily detected.
• Generate a spectral library using SpectraST (Lam et al., 2007) or similar.

• iRTs can be calculated from retention times in the database search using a linear regression on iRT and retention time (RT) values of the iRT peptides $iRT_{iRT\text{-}pep} = f(RT_{iRT\text{-}pep})$. This linear transformation can then be applied to all peptides identified in the database search using their RT. The calculation can be performed in Excel, R or similar.
3 Spectronaut™ Usage

Use the test data provided on our webpage or generate your own DIA run with the calibration kit to start testing Spectronaut™. For the spectral library, you can use any of the provided teaser libraries or one of the freely and more extensive downloadable (DLC) spectral libraries in Spectronaut™. To generate your experiment specific spectral libraries, you can use the built-in functionality in the prepare perspective of Spectronaut™ (see chapters 2.4 and 3.1) or you can generate the necessary spectral library yourself and select it as an external spectral library when setting up an analysis.

![Image](spectronaut_hrm_lcms.png)

**Figure 2.** Powerful cross runs visualization of Spectronaut™, here for 9 HRM LC-MS runs.

3.1 Prepare Perspective

The prepare perspective in Spectronaut™ allows the generation, merging and management of spectral libraries that are needed for the analysis of DIA data in Spectronaut. Currently supported shotgun database search engines are MaxQuant(Cox et al., 2011), Proteome Discoverer™, ProteinPilot™, and a generic search result format as defined by Biognosys.

3.1.1 Spectral Library Generation, Calibration Kit and iRT

In order to generate a spectral library, typically shotgun runs of your samples of interest are acquired, searched with a database search engine and the search results are then condensed into a spectral library. To generate a spectral library with Spectronaut™ we recommend that the samples used in the shotgun proteomics runs are used in conjunction with the iRT kit. It is recommended to use the iRT kit. The iRT peptides in the kit are used to empirically determine accurate iRT (Escher et al., 2012; Bruderer et al., 2016) values for all the peptides in the spectral library. In order for the iRT peptides to be found in the database search simply add the iRT peptide fusion FASTA file to the searched FASTA files. Having accurate iRTs for all the peptides...
in your spectral library greatly increases sensitivity, reproducibility and accuracy of identification and quantitation (Bruderer et al., 2015; Bruderer et al., 2016).

3.1.2 Optimal Spectral Library Generation

In order to obtain comprehensive coverage of the samples analyzed we recommend measuring all samples that should be quantified in shotgun as well. For large experiments this is not reasonable and hence we recommend measuring pools of all samples. Pools of conditions will help to enrich for “biomarkers” in the spectral library. For instance, if the experiment consists of 120 samples to be quantified, we recommend measuring pools consisting of 20 samples. Another option would be to measure two pools with 3 fractions each (e.g. using high pH reverse phase fractionation). If the samples are very homogeneous measuring 3-6 technical LC-MSMS replicates is still recommended due to the semi-stochastic nature of shotgun proteomics. The optimal number of pools, fractions and replicates depends on the complexity of the sample analyzed. Overly large spectral libraries where only a fraction can be recovered from the data might negatively influence the sensitivity of your analysis.

3.1.3 Spectral Library Generation Settings

Since Spectronaut 8, spectral library generation schemas can be saved in the settings and reused when generating a spectral library. The most important settings are described below. For further information, there are also helpful text hovers directly in the software.

3.1.3.1 Identification

Specify how you would like to filter your search results for Mascot, MaxQuant, Proteome Discoverer, and ProteinPilot. For MaxQuant, you can filter by peptide spectrum match false discovery rate (FDR), for Proteome Discoverer the confidence level reported by it which can be applied to both peptide and protein FDR, and for ProteinPilot using its Peptide Conf. and Protein Unused scores. Only a more stringent filter as compared to the one used in the search engine settings will make a difference. For Proteome Discoverer the three confidence levels of High, Medium and Low correspond to what was set in Proteome Discoverer. By default, this is an FDR of 1% for High, 5% for Medium and everything else for Low. For ProteinPilot, we use a Peptide Conf. score of 99 by default. For Mascot, FDR thresholding is currently unavailable. Instead, the user specifies a “OneInXprobRnd” value which treats the probability of a match occurring just by chance as 1/X. In other words, the p value cutoff is set to 1/X. Note this value must be at least 2. The user may also filter PSMs based on Mascot’s assigned ions score, given as -10Log(P), where P is the calculated probability that the observed match between experimental data and theoretical is a random event. Please see www.matrixscience.com/help/interpretation_help.html#THRESHOLDS for more information.

3.1.3.2 Protein Inference

Since Spectronaut 8 it is possible to perform protein inference within Spectronaut. This means that Spectronaut can group peptides into protein groups. Protein inference in Spectronaut is performed using the ID Picker algorithm (Zhang, Chambers, & Tabb, 2007). Having Spectronaut perform the protein inference analysis has a number of advantages: Protein grouping is well defined and protein group counts will be comparable across search engines and spectral libraries. Further, on top of the ID Picker algorithm, Spectronaut also checks which peptides are proteotypic. Summary information for proteotypicity can be seen in the Library Summary, Analysis Summary and the proteotypicity status of peptides can be exported in the report perspective (EG.IsProteotypic). Further, Spectronaut can also report all the protein IDs a specific peptide points to (EG.AllProteinAccessions). Since Spectronaut 8 one can attach all necessary information for protein inference to a spectral library. To do so, check “Perform Protein
Inference” in Settings -> Library Generation -> Protein Inference. Further, specify the digest type and protease(s). When generating the spectral library, you will be further asked to provide the relevant FASTA protein databases. When using a protein-inference-enhanced spectral library Spectronaut will automatically perform protein inference when analyzing a new experiment (Settings -> Analysis -> Protein Inference -> Protein Inference Workflow -> Automatic).

3.1.3.3 Spectral Library Filters
You have a number of options to filter the search engine results for spectral library generation. For more details, please consult the help text hovers in the software.

3.1.3.4 Workflow
If generating a library for an isotopically labelled sample, you can choose to in-silico generate an assay for a missing label pair. For example, if you have a sample of only spiked-in heavy aqua peptides, selecting this option will make a light/heavy library where the label-free channel is generated in-silico based on the heavy channel and with a consistent fragment ion selection. This option works with n-channel search results as long as you are working with isotopic labels.

3.1.4 Spectral Library Generation
After selecting the search results, you can choose your library generation settings schema and adjust it in the Load Experiment form (see Figure 4). The linking of raw files to MaxQuant search results is highly recommended and for Proteome Discoverer it’s required. Runs that are not already linked are marked with a red cross (see Figure 4). Click “Assign Shotgun Files...” in order to link the missing files to your analysis. (Please note: For Bruker, you can simply select the top level folder which has all the raw file directories. If you need to specify each individually, then you need to select the analysis.baf file in the Bruker raw directory.)

You can automate the linking if you specify your usual storage path for shotgun raw files in the Spectronaut Settings (Settings -> Global -> Directories -> Shotgun Raw Repository). If no repository is specified Spectronaut will look for the raw files in the location where they were present at the time of the database search for Proteome Discoverer™.

For how to influence the filtering of the search results please see the section 3.1.3.1.

3.1.5 Spectral Library Generation from Pulsar
Pulsar is Biognosys’ search engine and is designed to facilitate your workflows even more and is integrated in Spectronaut. Pulsar is able to search data dependent acquisition (DDA), data independent acquisition (DIA) and parallel reaction monitoring (PRM) data. Vendors that are supported are Thermo Scientific (Orbitrap for both MS1 and MS2) and Sciex. Both centroid and profile mode can be processed. The main focus of Pulsar is spectral library generation. Keeping in mind that data sets are getting bigger and bigger, Pulsar is designed to be fast and to scale with data set size. Further, Pulsar is also able to identify co-fragmented peptides in multiple search rounds by subtracting previously identified fragment ions from the spectra. False identifications are controlled by an experiment-wide peptide-spectrum match (PSM) and protein group false discovery rate (FDR).

To generate a spectral library from Pulsar, go to the Prepare Perspective and click on the “Generate Library from Pulsar...” link in the lower left corner. A wizard will appear to help you to set up the experiment. Most of the pages in the wizard contain the following links: “Previous”, “Next” and “Finish”. When clicking “Previous”, you are redirected to the previous wizard page in the event that you want to change something. “Next” should be clicked when everything at the current page is set and “Finish” will finish the experiment set-up for you, setting the default
settings for all the pages where you did not select a settings schema. When you open the wizard, you can choose an experiment name and the raw files from which you want to create the spectral library. The runs can be DDA, DIA or a mixture of both. Once all the files and an experiment name are chosen, click “Next”. The second step is choosing the protein databases. Protein databases can be assigned on a run basis by selecting the run and clicking “FASTA File...”. It is also possible to assign protein databases for the whole experiment, this can be done by selecting the experiment node and then selecting the FASTA files. Multiple protein databases can be selected for a single run. FASTA files are usually added to Spectronaut via Databases – Protein Databases, however it is also possible by right-clicking in the FASTA file window. In the next wizard page, you are able to set the analysis schema (for more explanation about the settings, see section 3.1.5.1). Either you choose a saved settings schema or you can change a schema on the fly. Setting the analysis schema can also be done at the run level or for all the runs in the experiment, by selecting a run node or the experiment node. Only one settings schema can be set per run. When nothing is selected for a run, default settings will be applied. The following page contains experiment-wide settings for PSM and protein FDR and for generating the library. The last page shows an overview of the whole experiment set-up. Clicking “Finish” will start the experiment. With the “View Live Log...” it is possible to follow the progress of the experiment. As soon as the library is generated, it will appear in the library tree.

3.1.5.1 Pulsar Search Settings

In the Settings Perspective, the settings for generating spectral libraries from Pulsar can be found under the tab Pulsar Search. There are two sections in the settings: Sequences and Applied Modifications. While the first one is directed towards generating stripped sequences from databases, the second one is focused on the modifications applied to these stripped sequences. In the sequences tab, the settings found are:

- **Digest Rule**
  The proteases used to in silico digest the proteins from the protein database(s). Digest rules can be specified in Databases – Digest Rules.

- **Digest Type**
  Here you can specify the type of digestion: specific, where both N- and C-terminus follow the specified digest rules, semi-specific when only one of the termini follows the specified digest rules and unspecific when there are no specified digest rules.

- **Maximum Peptide Length**
  Maximum number of amino acids for a peptide.

- **Minimum Peptide Length**
  Minimum number of amino acids for a peptide.

- **Missed Cleavages**
  How many consecutive cleavage sites the protease could miss.

- **Special AAs for Decoy Generation**
  Amino acids that for decoy generation are swapped with their preceding amino acid in order to generate decoys that have the same peptide length distribution but slightly different masses than their target versions.

- **Toggle N-terminal M**
  Pre-processing of the protein database by removing the N-terminal M (when there is one).

In the modifications tab, the following settings can be assigned:

- **Maximum Variable Modifications**
  The maximum number of variable modifications on one peptide. The higher this number, the more combinations one can make with one peptide, so the bigger the search space and thus the longer the calculation time.
- Select Fixed Modifications
  Modifications that are always present on the peptide.

- Select Variable Modifications
  Modifications that might or might not be present. The more variable modifications, the more possible versions of one peptide, which increases the search space size and calculation time.

3.1.6 Spectral Library Generation from Mascot

It is important that, before a library can be generated from Mascot search results, the user attaches the correct modifications file. Mascot uses modifications directly taken from Unimod.

At the time of this writing, we do not attach the Unimod XML database automatically to Spectronaut™, so the user will need to download the latest Unimod XML database themselves from www.unimod.org/downloads.html. Once this is saved, you can link it to Spectronaut™ by going to Databases -> Modifications and selecting “Import Modifications from...” and “Unimod”. From this file dialog, you can navigate to the Unimod XML file you just downloaded.

Once you have connected the Unimod database, you are ready to generate your Mascot library. Go to “Generate Spectral Library...” in the Prepare perspective and select “Mascot”. From here, you can navigate to your search result file “.dat” format. You should also attach the raw data by selecting “Assign Shotgun Files...” once the Load Experiment form is shown. You can then navigate to the raw files used in for this search.

3.1.7 Spectral Library Generation from MaxQuant

Linking the raw files for MaxQuant is highly recommended in order to improve the fragment quality of your final spectral library. The reason is incomplete fragment ion charge assignment in the MaxQuant search results.

The library generation of Spectronaut™ comes with the default peptide modifications used in MaxQuant. If you are using a different set of modifications you can import your MaxQuant modifications xml file into Spectronaut™ if you go to Settings -> Modification. The MaxQuant modifications xml file is usually stored in the MaxQuant installation folder under |bin|conf|modifications.xml. Modifications which are not properly imported into Spectronaut™ cannot be recognized.

Detailed information about how to setup a MaxQuant search containing the calibration kit for accurate iRTs in the spectral library can be found in the Biognosys Science Hub at www.biognosys.com/science-hub.

3.1.8 Spectral Library Generation from Proteome Discoverer™

Spectronaut also supports spectral library generation from search results processed with Proteome Discoverer™. Go to “Generate Spectral Library...” in the Prepare perspective and select “Proteome Discoverer”.

After selecting the search result file (PD 1.4 .msf, PD > 2.0 .pdResult) from your shotgun search results, the “Load Experiment Form” appears allowing you to select the processing schema and to link your raw files to the analysis in case they are not found. Proteome Discoverer™ does not report the fragment ion information. Linking the raw files is therefore mandatory in order to generate a spectral library from Proteome Discoverer™ results.

You do not need to import modifications into Spectronaut in order to generate libraries from Proteome Discoverer™ results.
3.1.9 Spectral Library Generation from ProteinPilot™

For generating spectral libraries from ProteinPilot search, you will need to use the ProteinPilot report that is automatically generated by ProteinPilot at the end of the search. The ProteinPilot report is of Microsoft Excel format and is named with the suffix “_FDR.” You will also need to import the modifications file of ProteinPilot into Spectronaut using Modifications Manager in the settings. The ProteinPilot modifications file is named as “Unified Modification Catalog.xlsx” and is located in the ProteinPilot/Help folder in the Program Files.

Once you select the shotgun search results, the “Load Experiment Form” appears allowing you to select the processing schema and to link your raw files to the analysis in case they are not automatically detected. Since ProteinPilot does not report the fragment ion information, it is mandatory to assign the shotgun raw files in order to generate a spectral library.

Spectronaut allows you to filter your ProteinPilot search results based on its Peptide Conf. Score (Spectronaut default set to 99) and Protein Unused Score (Spectronaut default set to 0). This can be done in the experiment setup form by clicking on the Identification node.

![Figure 3. Prepare perspective of Spectronaut for generating spectral libraries.](image-url)
Figure 4. Setup form for the spectral library generation. Runs that could not be matched to the original raw files are marked with a red cross. Use “Assign Shotgun Files...” in order to link the raw files to your analysis. The lower part of the form shows the setting options for the spectral library generation pipeline.

3.1.10 Spectral Library Generation from BGS Generic Format

Spectronaut supports generating spectral libraries from BGS Generic Format which is a generic and minimalistic search results format. The main motivation to include this was to allow the end-user to be able to use their favorite search engine with the aid of a basic script to convert their search engine format into BGS Generic format. This format is in a plain-text format where each row represents a PSM. The following columns are used in the format:

Raw File (Required)
The name of the DDA file in which this PSM was found. This column is used to map the DDA file to the PSM.

Stripped Sequence (Required)
The stripped sequence of the peptide that was found by the search engine for this PSM.

Precursor Charge (Required)
The charge that was associated with this PSM by the search engine.

Labeled Sequence (Required)
The sequence with encoded modification and label information. The only requirement is that the modification/label information should be enclosed in [modification/label] or (modification/label) brackets.
Retention Time (Required)
The retention time of the PSM. If available, retention time at apex intensity should be used.

Scan Number (Required)
The scan number of the PSM.

Scan Event (Required for Sciex instruments)
The scan event of the PSM. This is only relevant for .wiff files.

MS1 Intensity
The intensity of the PSM as reported by the search engine.

Protein Group Id
The protein group assigned by the search engine for this peptide. It is not necessary if using protein inference in Spectronaut.

3.1.11 Merging Spectral Libraries
Spectronaut allows merging of two or more spectral libraries in the prepare perspective. To do so, simply select the spectral libraries you would like to merge while holding the ctrl key and then, right-click to open the context menu and select the “Merge” option. This will open an experiment setup window similar to generating a spectral library from a database search.

Please note that if the spectral libraries used for merging have different types of Protein accessions (likely to happen if merging a library generated from MaxQuant with one generated from Proteome Discoverer™), then the protein counts in the merged library will be inflated as the same protein could be counted twice. Currently, in libraries generated from Proteome Discoverer™, the protein ID is specified by default as everything in the FASTA protein header after > and until the first space. You can avoid that issue when providing Spectronaut with the relevant information to do the protein inference (see section 3.1.3.2).

Please also be aware that merging many libraries can inflate the protein FDR. Performing a database search is always recommended, including all LC-MSMS runs at once and controlling the protein FDR at this step, which can be done in MaxQuant directly or using a tool such as Mayu (Reiter et al., 2009).

3.1.12 Spectral Library Overview
Spectronaut provides several different plots for obtaining an overview of your spectral library. You can access these plots by clicking on the spectral library node in the tree and then selecting an appropriate plot in the right panel.
3.1.13 Adding heavy labels to a library

By right-clicking on a spectral library in the prepare perspective, a user has the option to attach heavy labels to an existing library. Doing so will open the label editor form where you can select which labels should be applied to the existing library (see Figure 5). A selected library will be stripped of any pre-existing label. The selected workflow will decide how these new peptides will be treated during analysis. For detailed information about the supported workflows see section 3.2.5.5.

![Label Editor](image)

**Figure 5.** Example for applying a SILAC label to an existing library. The two isobaric modifications Arg6 and Lys4 are selected to be applied as label to all applicable peptides.
3.1.14 Enable Quality Control

When right-clicking on a spectral library in the prepare perspective you have the option to generate a new QC kit using this library. This will select 250 highly abundant peptides from the spectral library which will be added as a QC kit to the quality control perspective. The selection of peptides can also be altered manually in the dialogue. These peptides can then be tracked for quality control purposes in the quality control perspective whenever the corresponding spectral library was used.

3.1.15 Using Spectral Libraries

After the spectral library generation process is finished you will find a new spectral library in the prepare perspective (see Figure 3). To use a library from the prepare perspective for an HRM analysis in the review perspective, switch to “From Prepare Perspective” during the experiment setup. You can also select multiple spectral libraries sequentially. However, this is not recommended because duplicated peptides can complicate and bias data analysis.

3.2 Review Perspective

By default, Spectronaut™ starts up in the review perspective. This perspective allows a manual review of your DIA raw data during or after the data analysis is finished. Among many other visualizations you can cross check for instance TICCs of individual runs, signals of individual peptides aligned across all the runs in the experiment and annotated MS1/MS2 spectra.

![Figure 6. Review perspective of Spectronaut™. The lower panel shows the correlation of relative fragment ion intensities between measurement and the values provided in the spectral library.](image-url)
### 3.2.1 Starting a directDIA Analysis

Spectronaut Pulsar comes with a new analysis workflow for DIA that does not require the prior generation of a spectral library. This novel analysis allows you to directly search DIA files using nothing but FASTA files as input. To start a new directDIA analysis click on “Load directDIA Experiment...” when opening a new experiment tab. This will open the directDIA experiment setup wizard that will guide you through the setup step-by-step.

The directDIA analysis pipeline will not generate a spectral library but rather will directly apply a discovery analysis similar to shotgun proteomics. All identified signals are then processed using a targeted re-extraction, enabling all the benefits of a regular library-based analysis.

### 3.2.2 Starting a Library base Analysis

To load an experiment, click on the “Load Raw Data...” button. See section 1.6 for which DIA methods are supported. The experiment setup dialogue will appear. First, load one or several runs. You can then assign one or several spectral libraries either to each run individually, or to all the runs in the whole experiment. Spectral libraries generated in the prepare perspective or from external files can be used. For more information about importing spectral libraries into Spectronaut™ see section 3.1.4.

In the event that you are using multiple libraries per run in your analysis be aware that Spectronaut does not merge the libraries. In case a peptide is represented in both libraries, the final results will contain two instances of this peptide, one from the first and one from the second library source.

While setting up your experiment, you can select an analysis schema for your experiment. You can generate new customized schemas in the settings perspective. The analysis schema allows you to specify a certain set of parameters that you want to use for your analysis. In most cases the default schema will produce the best results. If you would like to have Spectronaut perform protein inference, you also have to specify the FASTA files in the experiment setup.
After you click “Start” Spectronaut™ will switch to the review perspective and start the analysis. First, in every run, Spectronaut™ will detect the respective calibration kit and calibrate analysis parameters for each run based on your specific data. The evaluation of signals from the calibration kits allows extraordinarily fast data processing and improves the sensitivity, specificity and accuracy of quantitation. Additionally, the peptides from the calibration kits are used for quality control monitoring, as summarized in the quality control perspective. If neither the HRM calibration kit nor the iRT Kit can be detected the data analysis will abort. Browsing your data will be possible a few seconds after the initial calibration process is finished. At the bottom left you can see a progress bar informing you of the overall status of the analysis. After the analysis has finished the number of unique precursors, peptides and protein groups identified will be shown in the bottom right for a qvalue cutoff of 0.01 (corresponding to a false discovery rate of 1% or an estimated 1% of false identifications among the accepted identifications).

3.2.3 Using an External Spectral Library for an Analysis

For HRM-MS™ a spectral library is necessary to analyze your data. There are two ways of importing a spectral library in Spectronaut. The first is by importing a compatible spreadsheet in a plain text format. The spectral library imported from a spreadsheet must follow two general formatting rules:

- Plain text, tab separated value format (tsv)
- A header needs to specify your columns

The “Spectral Library...” dialogue will try to auto-detect column names. If there are new column names, Spectronaut™ will ask you for every new column association whether or not you want to store it as a recognized synonym for this column. This allows Spectronaut™ to automatically
select these columns the next time you load a spectral library with a similar format. In case you accidentally save an association that is incorrect, you can remove the user-defined column synonyms in the settings perspective -> table import.

The minimal requirements of the spectral library are two columns; one containing the precursor ion m/z and one column containing the fragment ion m/z. Although an analysis would work with only these two minimally required columns in the spectral library, it is highly recommended that the information listed below is provided in order to achieve the best possible results.

In addition to spreadsheet import, spectral libraries generated by Spectronaut™ can be used. Upon setting up an experiment simply select “From Prepare Perspective” to use a spectral library from your collection in the prepare perspective.

![Figure 8. Import of a spectral library in spreadsheet format. Columns are auto detected. If a column name is unknown a new synonym can be saved to the settings -> table import.](image)

To view an example of a spectral library see our online material for download options or export a spectral library from the prepare perspective.

### 3.2.3.1 Spectral Library Columns

A Spectronaut™ spectral library strongly resembles a typical MRM/SRM transition list. Please use the following information to achieve the best possible results.

**PrecursorMz (Required)**

The \textit{in silico} calculated m/z of the peptide precursor ion. Please do not round this number.

**FragmentMz (Required)**

The \textit{in silico} calculated m/z of the peptide fragment ion. Please do not round this number.

**iRT (Highly Recommended)**

The peptide retention in the reverse phase chromatography as expressed with the normalized retention time iRT (Escher et al., 2012). If accurate iRTs are provided this will dramatically speed up the analysis and increase the quality of your results (sensitivity, specificity, accuracy). If no iRTs are available Spectronaut™ will predict an iRT for each peptide, which is less accurate than an empirically determined iRT value. In order to derive iRTs for your peptides, we recommend spiking the Biognosys’ iRT kit or HRM calibration kit into your shotgun proteomics runs. If the prepare perspective of Spectronaut™ is used, iRT values will be automatically determined for your spectral library.
**RelativeFragmentIntensity (Highly Recommended)**

The relative peptide fragment ion intensity expressed as a percentage of the most intense fragment ion. Please report the raw number without the percentage sign. This information improves the limit of detection.

**StrippedSequence (Recommended)**

The stripped amino acid sequence of the peptide excluding any modifications. Please only use the single letter code for the 20 standard proteinogenic amino acids. This information is used for labelling and scoring of your fragment ions in Spectronaut™. Further, it is included in the automatically generated unique ID for your precursor if necessary.

**PrecursorCharge (Recommended)**

The peptide precursor ion charge. This information is used to label your precursors in Spectronaut™ and to automatically generate a unique ID for your precursor if necessary.

**FragmentType (Recommended)**

The peptide fragment ion type. Usually this is “y” or “b”. This information is used for labelling and scoring of your fragment ions in Spectronaut™.

**FragmentNumber (Recommended)**

The peptide fragment ion number. This number should be between 1 and the length of your peptide in amino acids minus one. This information is used for the labelling and scoring of your fragment ions in Spectronaut™.

**FragmentCharge (Recommended)**

The peptide fragment ion charge formatted as number. This information is used for the labelling and scoring of your fragment ions in Spectronaut™.

**FragmentLossType (Recommended)**

Allows you to specify the fragment loss-type (e.g. NH3 or H2O). This is mainly used to label your fragment ions in plots.

**ExcludeFromQuantification**

This column can be specified optionally and is meant for specifying fragment ions that should only contribute to identification but not to quantification (TRUE or FALSE). TRUE means that Spectronaut will never consider this fragment for quantification and FALSE that Spectronaut will always use it for quantification. If you leave the column empty, you will let the automatic interference correction decide whether or not to use it (the latter is the default if the column is not specified).

**ModifiedSequence**

In the event that your peptide is modified use this column to specify the amino acid sequence including modifications. The modified sequence should be constant for one unique precursor. This information is used to label your precursors in Spectronaut™ and automatically generate a unique ID if necessary. Spectronaut will try to parse and map modifications from the provided sequences to the internal modification database. This field does not contain any label specific modifications (see LabeledSequence)

**LabeledSequence**

Similar to modified sequence. This sequence will specify the respective label for each channel in a labeled experiment. It will also contain any modifications. An example SILAClabeled peptide
could look like this: \_NAYVC[+57]WTLK\_ for the light channel and \_NAYVC[+57]WTLK[+8]\_ for the heavy channel of the same peptide.

**ProteinId**

The ID of the protein the peptide is derived from. This information is used to label your peptides. Spectronaut™ provides filtering capabilities in the review perspective including filtering for the protein ID.

**UserGroup**

A column for additional user information that can be used for grouping and filtering.

For an exemplary spectral library list please visit the Spectronaut™ webpage on [https://biognosys.com/shop/spectronaut-pulsar](https://biognosys.com/shop/spectronaut-pulsar) and download the test data set.

### 3.2.3.2 Modification Parsing

Once the library is imported, Spectronaut will try to parse all values imported from the “ModifiedSequence” and the “LabeledSequence” columns to assign modification specifications to them. This allows Spectronaut to have greater control over decoy generation. If possible, Spectronaut will automatically assign known modifications from its internal database. If a certain modification is unknown, you will be prompted to assign the modification specification from the database to the new keyword (see Figure 9). The only parsing requirement for external modification definitions is that the modification tag is specified within round or square brackets. Spectronaut will not parse modifications specified as single letter special amino acids (like ‘B’ for carbamidomethyl cysteine or ‘O’ for oxidized methionine). You can remove previously assigned parsing synonyms in the modification tab in the settings (see section Error! Reference source not found.).

![Figure 9. Modification assignment during import of an external library. Two new definitions where found that could not be automatically resolved.](image-url)
3.2.4 Configure Conditions

Here you can specify the conditions, fractions and replicates of your experiment which are required for the testing of differential protein abundance. Spectronaut will perform a pairwise comparison of all conditions specified in the Conditions Editor (If not set to “None” in Settings -> Analysis -> Post Analysis -> Differential Abundance Testing). The results are reported in the post-analysis perspective.

You can define a global rule to automatically parse file names such that conditions and replicates are defined (Settings -> Global -> General -> File Name Parsing Schema). This will allow one to avoid having to manually enter conditions and replicates for very large experiments. Please see the post-analysis perspective section 3.3 for more information on differential expression analysis.

You can also reach this editor on a finished analysis by right-clicking on the experiment tab in the review perspective, choosing settings and then “Configure Conditions...”.

3.2.5 Analysis Settings

Spectronaut offers analysis settings schemas for the DIA data analysis and quantification. The most important analysis settings are described below. For further information, there are also help text hovers directly in the software.

3.2.5.1 XIC Extraction

Parameters related to ion trace extraction and peak boundary determination for identification and quantification. XIC RT Extraction Window defines whether iRT should be used to predict the elution of a peptide. The dynamic window predictor will let Spectronaut determine the ideal extraction window dynamically depending on iRT calibration and gradient stability. Sections of the gradient that show higher variability during the calibration step will automatically be extracted using wider windows. A correction factor of 2.0 would mean that you want to use 2 times the window that Spectronaut suggests. The default settings are recommended.

If for some reason the ion traces should be extracted over the complete gradient chose the “Full” option.

3.2.5.2 Calibration

For htrms files Spectronaut stores the calibration. Automatic will leave the decision to Spectronaut whether an htrms file needs to be recalibrated or not.

The iRT Calibration Strategy defines whether a linear or local regression shall be used for the iRT calibration. With the non-linear calibration Spectronaut can correct for small variations within a linear gradient, as well as perform correct retention time prediction on completely non-linear gradients. Should the number of possible calibration data-points be insufficient to perform a correct non-linear calibration, Spectronaut will automatically switch to the linear calibration approach by default.

3.2.5.3 Identification

This is about the identification of peptide precursors in Spectronaut. The first option lets you decide whether you’d like to perform the identification process in each single LC-MS run individually or whether this should be done on the experiment as a whole.

The most important option is the “Qvalue Cutoff”. A qvalue cutoff of 0.01 corresponds to a false discovery rate of 1% or to an estimated amount of 1% of false peptide precursor identifications among the accepted identifications. For a detailed description of the algorithms used to
determine a qvalue in DIA data see (Reiter et al., 2011). If Spectronaut is set to perform protein inference it will perform it on the set of peptides as defined with this qvalue cutoff. Further, exports from the report perspective will be according to this selected qvalue cutoff. Also the differential regulation analysis in the post analysis perspective is applied to this qvalue cutoff for identification.

3.2.5.4 Quantification

Since Spectronaut 9, all quantitation aspects are regulated via the settings directly. You can specify not only the MS level that you want to quantify but also whether you prefer XIC peak area or height for your quantitation. Additionally, you can select what kind of data-filtering should be applied for post-analysis (Qvalue, Qvalue sparse or Qvalue complete). This will affect the post-analysis itself, as well as any plots related to post-analysis. A “Qvalue” filter corresponds to keeping all signals with a qvalue below the threshold (Settings -> Analysis -> Identification -> Qvalue Cutoff). A “Qvalue Sparse” filter means that a particular peptide precursor had to be identified once with a qvalue below the threshold in the complete experiment. Hence, this filtering is substantially more relaxed than the “Qvalue” filter. “Qvalue complete” means that a particular peptide precursor had to be identified with a qvalue below the threshold in all the runs in a particular experiment. “Qvalue complete” is much more stringent than the other two settings.

Interference correction allows you to automatically exclude fragments across several runs that likely interfere with other ions. Fragment ions detected to have interferences are consistently excluded across all loaded runs. The same applies to manual exclusion of fragment ions (right-click on fragment ion node). If interference correction is not selected, possible interferences will be highlighted in the plots but will still be part of the quantification. Interference correction requires at least 6 fragment ions per precursor to be present in the library. An interference score for each fragment ion can be exported from the report perspective. Please see this reference (Bilbao et al., 2015) for information on the interference correction algorithm.

The normalization process implemented in Spectronaut™ corrects for systematic variance in the LC-MS performance. The normalization is based on the assumption that on average a similar number of peptides are up- and down-regulated, that the samples used are similar and that the majority of the peptides within the sample are not regulated across runs and along retention time. This normalization is suited for most but not every situation and should therefore be selected carefully. The normalized quantitative information can be found as additional columns in the report. The normalization is performed in such a way that the final values are in the same range as the initial values. The algorithm is based on the algorithm published in (Callister et al., 2006). When using sample pre-fractionation for DIA, these can be annotated in the “Conditions Editor” when setting up an experiment. This will cause Spectronaut to perform normalization only within one fraction. See also the section 3.3.3 for more information.

Alternatively, one can also choose a global normalization instead to either normalize by median or the average intensity of all identified peptides. If you select global normalization for example, the normalized quantity for each peptide is the original quantity minus the run average quantity (average quantity of all peptides of that particular run) plus the global average quantity (average quantity of all peptides in all runs).

3.2.5.5 Workflow

The profiling workflow allows the user to carry over the measured, indexed retention times (iRT) of peptides that could be identified (qvalue ≤ 0.01) in certain runs, in order to fix integration boundaries in runs where an identification could not be achieved with sufficient confidence.

Usually it is recommended to only profile peptides that were identified at least once over all runs. To determine which peptides within a profile require alignment, Spectronaut performs an
outlier detection based on retention time shifts and differences in fragmentation patterns of peptides across runs. Further, the user has the option to only consider peptides for alignment that were not identified. The default setting for “profiling row selection” is “Minimum qvalue Row Selection” of 0.01. This is equivalent to a sparse profile filter which selects any profile that was identified with a qvalue cutoff of 0.01 in at least one of the runs.

You can also specify what the default workflow should be. The “default workflow” setting only applies to multi-channel experiments. Any precursors containing only one channel specification will automatically be treated as “label-free”.

Currently supported workflows in Spectronaut are:

- **Label-free**: Default workflow for all single channel experiments. Peak detection, scoring and identification are applied as usual.
- **Label**: Default workflow for all 2-channel workflows. Peak detection and scoring will be applied to all channels. Quantification in post-analysis will be performed on the light to heavy ratio.
- **Spike-in**: Peak detection will be performed on only the reference (heavy) channel. Scoring and identification will be performed on the target (light) channel. The heavy channel is expected to be easily detectable and considered a peak-picking aid in this experiment. Quantification in post-analysis will be performed on the target to reference ratio.
- **Inverted spike-in**: Similar to spike-in but the light channel is considered the reference.

### 3.2.5.6 Protein Inference

Since Spectronaut 8 it is possible to perform protein inference within Spectronaut. This means that Spectronaut can group peptides to protein groups. Protein inference in Spectronaut is performed using the ID Picker algorithm (Zhang et al., 2007). Having Spectronaut perform protein inference has a number of advantages: Protein grouping is well defined and protein group counts will be comparable across search engines and spectral libraries. Further, on top of the ID Picker algorithm, Spectronaut also checks which peptides are proteotypic. Summary information for proteotypicity can be seen in the Analysis Summary and the proteotypicity status of peptides can be exported in the report perspective (EG.IsProteotypic). Further, Spectronaut can also report all the protein IDs a specific peptide points to (EG.AllProteinAccessions). Since Spectronaut 8 one can attach all necessary information for protein inference to a spectral library (see section 3.1.3.2). When using a protein-inference-enhanced spectral library and selecting “Automatic” for the “Protein Inference Workflow”, Spectronaut will automatically perform protein inference. If the “Protein Inference Workflow” option is set to “From search engine” the protein inference of the respective database search engine is used. When setting this option to “From protein-db matching” new protein digest options can be specified. Spectronaut will then ask for the FASTA files when setting up an analysis.

The protein groups resulting from the ID Picker protein inference are used to perform the statistical analysis for differential protein expression in the post-analysis perspective, to group peptide precursors according to protein groups in the review perspective (right-click on experiment tab, see section 3.2.7), as well as in the report perspective for the protein group reportables.

The protein inference calculation will not just allow you to re-calculate the protein grouping, but also allows you to map any meta-information parsed from the FASTA headers to your protein groups to be used in further analysis.
3.2.5.7 Post Analysis
See the post-analysis perspective section 3.3.2 for more information on post-analysis settings.

3.2.5.8 Reporting
See the pipeline perspective section 3.6.1 for more information on reporting settings.

3.2.6 Tree Filtering
Using the filtering system implemented in Spectronaut, one can apply one or several filters on the review data tree. These filters only influence what is shown in the review perspective but not for instance the post-analysis perspective. Select a filter from the dropdown menu and set the filter criteria. The filter is now marked as selected within the dropdown menu. To combine filters, select a different filter and define the value that should be applied. A precursor has to apply to all selected filters in order to be shown in the review tree.

A special filter criterion is the “User Group” filter. This value can be set during the library import by selecting a specific column as “User Group”.

By default, the Qvalue filter is set to what was chosen in the settings. The default is 0.01.

3.2.7 Tree Grouping
If you right-click on the experiment tab, you can apply different grouping strategies to the experiment tree. By default, precursors are grouped by their respective precursor window. Other grouping possibilities are:

- **Protein Group**: Enables grouping by protein group. If protein inference was performed by Spectronaut this corresponds to the ID Picker protein group. If the protein grouping from the search engine was taken over this corresponds to the protein group of the search engine used to create the spectral library. If an external spectral library was imported this corresponds to the field that was imported as “ProteinId”.
- **Run**: Shows all the precursors of one run as one group. This mode is most useful with small spectral libraries.
- **User defined Group**: Similar to the tree filtering by user group. Can be defined during the import of an external spectral library.

3.2.8 Saving/Loading Experiments
Since Spectronaut™ 5.0, saving and loading of a complete analysis is supported. Saving an experiment will also store manually integrated peaks, manually refined assays, comments, excluded fragment ions and whether a peak was manually accepted or rejected. To save an experiment right-click on the experiment tab and select “Save As...” in the context menu. The most recently loaded or saved experiments are shown in the “New Experiment” page of the Spectronaut™ review perspective. A saved experiment in Spectronaut™ will contain the full analysis information. If you save the iontraces the .sne file will be very large but Spectronaut no longer requires the raw file to be available (Settings -> Global -> Reporting -> Store Iontraces in SNE). If you do not save the iontraces to the .sne file, the file will be smaller and you have the option to reconnect the raw files to the analysis again once the .sne is loaded in Spectronaut.

3.2.9 Run Node Plots
On the run node level, one can access different plots that give you detailed information about the calibration status, DIA method used, TICC, run meta information and cross run performance.
3.2.9.1 iRT Calibration Chart

This chart shows the status of run calibration. Initially, only showing the peptides of the used calibration kit. Spectronaut now also supports non-linear gradients using a refined calibration based on the initial calibration and detailed information in the user library.

![iRT Calibration Chart](image)

**Figure 10.** iRT Calibration Chart showing the non-linear transformation from library iRT to actual predicted retention times. The chart is shown after extended non-linear calibration has been performed. The extended calibration allows you to correct even small gradient fine-structure fluctuations in order to get the most accurate retention time prediction for your library.

3.2.9.2 XIC Extraction Width Chart

Based on extensive calibration using the respective calibration Kit and information about the spectral library used, Spectronaut can determine the optimal XIC extraction width for your experiment. This extraction width is calculated dynamically over the whole gradient and allows Spectronaut to automatically adapt to areas with lower retention time prediction accuracy. The window width can be influenced with a correction factor in the Settings perspective (Settings -> Analysis -> Peak Detection -> Correction Factor).

![XIC Extraction Width Chart](image)

**Figure 11.** The XIC Extraction Width plot gives insight into your gradient stability and the overall accuracy of your library’s iRT values. The blue and the orange lines show the window selections as suggested by Spectronaut (blue) and as set by the user (orange -> correction factor in the settings). These extraction window widths (y-axis) change over time (x-axis) based on gradient stability and iRT accuracy. The red dot’s show your libraries iRT accuracy assuming a linear iRT to RT transformation. The green dots show the iRT accuracy using the extended non-linear iRT to RT transformation. The later one is used for the actual analysis.
3.2.9.3  **TIC & TIC Overlay Chromatogram**

The TIC Chromatogram plots show your run’s total ion current chromatogram. The TICC overlay combines the information of all runs of your experiment in one plot for better insight into instrument stability and amount of sample injected.

![TIC Overlay Chromatogram](image)

**Figure 12.** The TIC Overlay chromatogram shows the total ion current of all runs in your experiment in a single plot, giving insight into instrument stability and amount of sample injected.

3.2.9.4  **Analysis Log**

The analysis log contains all the information pertaining to the analysis of your whole experiment. In the event of errors one can consult the analysis log for detailed information of what went wrong.

![Analysis Log](image)

**Figure 13.** The analysis log with detailed information about the analysis processes in Spectronaut.

3.2.10  **Peptide and Fragment Plots**

There are several different visualization options for the different peptide and fragment node levels in your review tree. These range from XIC chromatogram visualization to more score-centric and cross-run profile visualizations. The latter ones are only available in multi-run experiments and disabled for experiments containing only one run or peptides that are only targeted in one run.
3.2.10.1 MS2 XIC

The default plot on Elution Group (EG or peptide), Fragment Group (FG or peptide precursor) and Fragment (F or fragment ion) level. This plot shows the extracted ion current chromatogram of the selected peptide. The plot contains the XICs for all fragments present in the library. Additionally, the expected retention time is marked (black dotted line) and the currently selected peak for this peptide (green area). On the Elution Group and Fragment Group level the coloring of the fragments is based on its expected relative intensity ranging from red (expected most intense fragment) to blue (expected least intense fragment). You can directly change integration boundaries in the plot.

On fragment level, this plot only shows the selected fragment in color and all others in gray.

There are a number of options available upon right-clicking in the plot such as switching the y-axis to log scale, toggling accept/reject the peak, changing x-axis scale to iRT, showing normalized intensities and showing the XIC chromatogram for the whole gradient.

![Figure 14. XIC chromatogram for the peptide MPENMNIK++. The color coding of the fragments indicate an overall good correlation to the expected fragment intensities. The dotted blue line indicates a potential interference that was detected by Spectronaut and automatically removed for relative quantitation.](image)

3.2.10.2 MS2 XIC Sum

The XIC Sum chromatogram chart shows the selected peptides quantitative information. The XIC shown is the sum of all fragment XICs that qualified for quantitation. All fragments that were excluded due to interfering signals are not used to calculate the sum XIC chromatogram.
3.2.10.3 MS2 Intensity Correlation

This plot shows detailed information about the correlation of predicted and measured relative fragment ion intensity. The predicted values in red correspond to the relative intensities provided by the spectral library. The black lines correspond to the relative measured intensity of each fragment ion. Fragment ions with potential interferences are displayed as dotted lines.

Figure 16. The MS2 Intensity Correlation plot for a given peptide precursor. The plot indicates a very good correlation between the expected relative intensities (red) and the measured intensities of the library fragments.

3.2.10.4 MS1 Isotope Envelope XIC

This plot shows the monoisotopic precursor plus its most abundant isotopic forms as an XIC chromatogram. The XIC chromatograms on MS1 and MS2 level are expected to have identical apex retention times and elution shapes. As with the MS2 XIC chromatogram plot, the MS1 XIC chromatogram is also color coded to reflect the predicted relative intensities. A color coding from red (highest) to blue (lowest) indicates high correlation with the predicted abundance.
3.2.10.5 MS2 Isotope Envelope Correlation

Similar to the MS2 Intensity Correlation plot, this plot highlights the correlation between the expected and the observed fragment isotope patterns. A high correlation between the measured (black) and the predicted (red) abundances signals high confidence in the identification and quantification.

Figure 18. The envelope correlation plot for two fragments with very high correlation with respect to the predicted abundance. The measured intensities (black) are almost perfect mirror images of the predicted (red) intensities.

3.2.10.6 MS2 XIC Alignment

The MS2 XIC Alignment plot allows you to visualize the extracted ion chromatograms of a single peptide across all of your runs. You can directly change integration boundaries in the plot.
3.2.10.7 iRT XIC Sum Overlay

This plot allows you to show all sum XIC chromatograms of your selected peptide from all runs as an overlay plot. The x-axis scale is by default in iRT but on right-click can be changed to actual retention time.

![Figure 19. The MS2 XIC Alignment of the peptide IILDLISESPIK++ across 4 runs. The x-axis is automatically changed to iRT to reduce chromatographic variance. The axis can be changed to retention time by right-clicking on the plot and un-selecting the “Use iRT Scale” option.](image)

3.2.10.8 MS2 Intensity Alignment

Similar to the MS2 XIC Alignment, this plot gives detailed information about the signal stability for one peptide across several runs. The different bars show the relative abundance of each fragment ion across multiple runs. Using this plot, one can quickly identify an inconsistent signal by the change in the color pattern. Right-click on the plot and un-selecting “Normalize” to show intensities on absolute scale.

![Figure 20. The iRT XIC Sum Overlay chart for the peptide IILDLISESPIK++. The 4 XICs correspond to the sum XIC of one peptide in the 4 different runs loaded for this experiment. The title of the plot additionally shows the peptides coefficient of variation, in this case 4.5%.](image)
Figure 21. The MS2 Intensity alignment for a peptide containing 8 fragment ions. Each fragment ion's relative intensity compared to the total peak height is indicated using a differently colored bar. The peptide was targeted in 24 runs. An inconsistent signal can be easily identified due to the sudden change in the color pattern.

3.2.10.9 Cross Run RT Accuracy

Similar to the MS2 Intensity Alignment chart, the Cross Run RT Accuracy plot allows one to quickly validate the peak picking across several LC-MS runs. The x-axis shows colored bars that correspond to the peptide in different runs. On the y-axis one can see the retention time in iRT.

The height of each box corresponds to the peak width at the start and the end iRT according to the y-axis. The line through the middle of the box shows the apex retention time in iRT while the blue colored boxes in the back show the total XIC extraction width. The colors of the bars again correspond to the relative intensities of the measured fragment ions. The bar with the green background is the currently selected node. The black, dotted, horizontal line corresponds to the expected retention time in iRT. You can hop to any other node by clicking on the colored bar.

Figure 22. The Cross Run RT Accuracy plot for a peptide measured in 4 different runs. The multi-colored bars correspond to the detected peak with the colors encoding the relative fragment intensities and the upper and lower boundary of the bar corresponding to the peaks start and end retention time in iRT.

3.2.10.10 MS1 Spectrum at Apex

This plot shows the MS1 signal at apex retention time for a given peptide. The chart automatically zooms in on the isotopic envelope and labels the different peaks accordingly. The
red bars indicate the expected relative intensities of the different isotopic forms of the precursor.

![MS1 spectrum at apex for the peptide IILDLISESPIK++ showing the monoisotopic precursor plus its first 3 isotopic forms.](image)

**Figure 23.** MS1 spectrum at apex for the peptide IILDLISESPIK++ showing the monoisotopic precursor plus its first 3 isotopic forms.

### 3.2.10.11 MS2 Spectrum at Apex

Similar to the MS1 Spectrum at Apex, this chart shows the full recorded MS2 spectrum corresponding to a given peptide’s LC-peak apex. By default, the plot only highlights the fragments as provided by the library. Additionally, other fragments can be calculated and annotated by right-clicking on the plot and un-selecting “Show Only Library Fragments”.

![MS2 Spectrum at Apex for the peptide ILLDLISESPIK++. The option “Show only Library Fragments” was turned off.](image)

**Figure 24.** MS2 Spectrum at Apex for the peptide ILLDLISESPIK++. The option “Show only Library Fragments” was turned off.

### 3.2.10.12 Protein Coverage

Since Spectronaut 9 you can get a detailed overview of a protein’s coverage within your analysis. The protein coverage plot shows you all the peptides of a protein that were targeted within your current experiment. The color coding by default indicates the confidence level for each peptide.
This can be toggled on right-click to show the proteotypicity status for each peptide. This option is only available if protein inference was enabled during the DIA/SWATH analysis.

![Protein coverage plot for protein O00410. The protein sequence is covered by 53% by the peptides of the selected spectral library.](image)

**Figure 25.** Protein coverage plot for protein O00410. The protein sequence is covered by 53% by the peptides of the selected spectral library.

### 3.2.11 Library Refinement

Since Spectronaut 8, users have the option to refine the fragment ion selection of their spectral libraries. This way one can permanently remove fragments that show interferences, specifically select fragments to cover interesting modification sites or add additional fragments that were not detected in the original DDA analysis but are nicely visible in HRM.

In order to perform library refinement, the peptide assay must originate from a spectral library generated in Spectronaut’s prepare perspective.

To refine an assay, an HRM run and the library must be loaded in the review perspective of Spectronaut. To refine the fragment selection for a certain peptide, select the respective node in the review perspective. Right-click on the peptide node and select “Refine Fragment Selection (F)” or simply press the “F” key.
Figure 26. Fragment selection form showing a peptide with a prominent interference. The spectral library contains 8 different fragments for this peptide of which 6 are currently selected.

A new dialog appears that shows you the selected peak in detail. The list on the left will show you all fragments of this peptide that are present in the library and which of those are currently selected. Additionally, a list of theoretical fragments is generated. In order to change the selection of theoretical fragments, right-click on the “Theoretical Fragments” node and select “Set Fragment Filter (F)”. This will allow you to expand the set of theoretical fragments so that it will contain different ion types, as well as common loss types.

On the right you can see a live preview of how the fragment selection will affect both the XIC and the match between the predicted and measured fragmentation pattern. Please note that theoretical fragments will not contain a predicted intensity. To add or remove a fragment to and from the spectral library, simply check or un-check the tick-box in front of the corresponding fragment node.

After the new fragment selection is complete, click on “Apply Selection” to mark this peptide for refinement. Please note that the refinement is not performed immediately. In order to effectively change the peptide in your experiment, as well as in the spectral library, right-click on the experiment tab in the review perspective and click “Commit Library Changes...”.

This extra step is necessary since the update will require a re-extraction of this peptide from all currently loaded runs. Spectronaut will therefore perform this operation in batches once all manual fragment selection is done and not for each peptide individually. All changes made to any peptide’s fragment selection will only take effect once the library changes are committed.

After clicking “Commit Library Changes...” another popup window will appear asking you for a version name. Spectronaut features a version control for spectral libraries that allows you to switch between different versions of a certain library. This way any changes to the original library can always be reverted. In order to change the version of a spectral library, go to the prepare
3.2.12 Export All

When right-clicking on the experiment tab in the review perspective you can choose from a large number of different reports to be generated such as data reports, post-analysis reports, a graphical representation of the normalization and the Spectronaut experiment .sne file.

3.2.13 Targeted Analysis of SONAR™

As of Spectronaut 11/Spectronaut Pulsar, the targeted analysis of Waters SONAR™ is supported. There is however currently no direct loading pipeline implemented in Spectronaut as of yet. In order to analyze SONAR™ data it has to be converted to HTRMS first using the HTRMS converter installed together with Spectronaut.

3.3 Post-Analysis Perspective

The post-analysis perspective in Spectronaut reports information that is less raw data focused, as compared to the review perspective. It shows summary information on identification, quantitation, a statistical analysis of differential expression and the results of the hierarchical clustering.

3.3.1 Configure Conditions Editor

In order to perform statistical analysis of differential expression in Spectronaut it is necessary to assign condition/sample and replicate information to your runs. This can be done during analysis setup by clicking on the “Configure Conditions...” link to open the Conditions Editor (Figure 27). Spectronaut will perform a pairwise comparison of all conditions specified in the Conditions Editor (if not set to “None” in Settings -> Analysis -> Post Analysis -> Differential Abundance Testing). The results are reported in the “Differential Abundance” node in the post-analysis perspective.

To simplify assignment of conditions and replicates from run file names you can define a global rule to automatically parse file names such that conditions and replicates are defined (Settings -> Global -> General -> File Name Parsing Schema). This will avoid having to manually enter conditions and replicates for very large experiments. “Is Reference” is used only for plotting purposes. When right-clicking on one of the profile plots (Post-Analysis Perspective -> Differential Abundance -> Candidates) there is the option to choose “Normalize by Reference”. This will divide every condition by the reference condition selected in the configure conditions editor. The statistical testing for differential regulation will be done by pairwise comparison of all “Conditions”. Further, coefficients of variation will be calculated within conditions for some of the plots as indicated. The “Fraction” column can be used to indicate fractions within one sample (see section 3.3.3). Using this annotation causes normalization to be performed only within one fraction. The “Run Label” is what was extracted from the run file name. The “Replicates” column indicates biological, technical or other replicates within a condition. For label free quantitation we recommend at least 3 biological replicates depending on the
experiment performed. The “Label” column is used for the plots in the post-analysis perspective. The label is concatenated as <Label> + “R” + <Replicate>.

You also have the option to enter quantity correction factors per sample. Spectronaut will multiply the final quantities with this factor. An application example would be if quantities should expressed per initial sample volume (for instance in plasma).

You can also reach this editor on a finished analysis by right-clicking on the experiment tab in the review perspective, choosing settings and then “Configure Conditions...”. It’s also possible to change annotation after an analysis has finished. The differential expression analysis will be automatically retriggered.

Each condition in Spectronaut will get a color assigned during the setup which will later be used for certain post-analysis plots, as well as for the heatmap.

![Conditions Editor](image)

Figure 27. The condition editor allows you to further specify your experiment setup. Alternatively one can import predefined condition assignments from a text-file.

### 3.3.2 Post-Analysis Settings

If the option “Differential Abundance Testing” is set to Student’s t-test, Spectronaut will perform testing for differential abundance according to the settings in the conditions editor. Independent of the other settings Spectronaut will perform a one sample t-test on the log ratios of the quantitative data points between two conditions.

You can choose what the basis for quantitation should be with “Quantity” (MS2 peak area is recommended). If you have enabled the normalization and interference correction algorithms, the normalized and interference corrected values will be used for differential abundance testing.
With “Data Filtering” under the “Quantitation” settings node you can choose how the data should be filtered. For more information, see section Error! Reference source not found.. With the “Differential Abundance Grouping” you can choose which groups the differential abundance statistics shall be computed for. A typical setting is “By protein group” and means that all data points are grouped according to protein group. “By modified peptide” is a useful setting for Phosphoproteomics for instance. There, it makes particular sense to consider that each modified peptide can be regulated separately.

With the “Smallest Quantitative Unit” you can control on what quantity the log ratios are calculated. For large experiments performing the calculations on the fragment ion level will lead to much longer processing times.

If you also have high resolution full range MS1 scans in your DIA data you can include the MS1 data in addition to the MS2 data into your quantitation with the “Include MS1” option.

3.3.3 Fractionation

The “Fraction” column can be used to indicate fractions within one sample. This can be used if every sample was pre-fractionated into a fixed number of fractions. Using this annotation causes normalization to be performed only within one fraction. It is advisable to use specific spectral libraries for each fraction and set up the experiment accordingly. Sparse and full profile qvalues will always be calculated on every targeted instance of a peptide. Therefore it is important that peptides that are not expected to be present in a certain fraction, are also not targeted in this fraction. The following table indicates this behaviour showing three different scenarios of an experiment containing of 2 fractions with each 2 replicates.

<table>
<thead>
<tr>
<th>PEPTIDE</th>
<th>IDENTIFIED IN F1R1</th>
<th>IDENTIFIED IN F1R2</th>
<th>IDENTIFIED IN F2R1</th>
<th>IDENTIFIED IN F2R2</th>
<th>FULL PROFILE</th>
<th>SPARSE PROFILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
<tr>
<td>B</td>
<td>TRUE</td>
<td>FALSE</td>
<td>Not Targeted</td>
<td>Not Targeted</td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
<tr>
<td>C</td>
<td>TRUE</td>
<td>TRUE</td>
<td>Not Targeted</td>
<td>Not Targeted</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

3.3.4 Analysis Overview

Here some information about identification, coefficients of variation and normalization is provided for the whole experiment. In the data completeness plot the “Cumulative Sparse Profiles” describes the cumulative number of identifications when looking at 1 to n runs, where n corresponds to the complete experiment. The “Cumulative Full Profiles” corresponds to identifications that were consistent across 1 to n runs, i.e. identified in all the runs currently looked at. The Coefficients of Variation plot shows the distribution of CVs in the experiment. On right-click you can change the basis of quantitation between precursor, peptide, protein and protein group. In the normalization you can see boxplots of responses for the individual runs before and after normalization.
Figure 28. The upper plot under the “Data Completeness” node shows you the difference between cumulative sparse and full profiles. The green bars represent the growth of cumulative precursor identification. So for example, if you only consider the first 3 runs (x = 3) 26878 precursors were identified at least once in these 3 runs. The blue bars represent the decline of full profiles. For example, after the first 3 runs, only 22747 precursors were identified in all 3 previous runs. Adding more runs, the green bars can only ever go up or remain constant while the blue bars can only ever go down or remain constant. Like with many other plots you can change the context to display the data completeness on precursor, peptide, protein-group or protein level.

Figure 29. The lower plot under the “Data Completeness” node shows the histogram of missing values. In this plot, about 16500 precursors did not have any missing values and were identified in every run (this is equivalent to the last blue bar in Figure 28) while about 2200 precursors had one missing value and so forth. The last bar represents the number of single-hit precursors. These are the precursors that were only identified in one single run (hence they were only identified in 1 out of 18 runs in this case). Like with many other plots you can change the context to show you the missing values on precursor, peptide, protein-group or protein level.
Figure 30. The upper plot under the “Coefficients of Variation” node shows the %CV distribution for all conditions in your experiment. You can also change the context by right-clicking on the plot to show you the CVs for precursor, peptide, protein-group and protein quantities or to show the CV distribution across all conditions.

Figure 31. Similar to Figure 30, this plots shows the distribution of CVs for each condition. For experiments with large numbers of conditions, this view might be preferable. Like with the previous figures, you can change the context of the plot by right-clicking on it to select either precursor, peptide, protein-group or protein scope.
Figure 32. The CVs below X plot shows the number of precursors that were below either 20% or 10% CVs. The red bar shows the number of all identifications for each condition regardless of quantitative precision. The bar for ≤ 20% CVs also includes all counts from ≤ 10% CVs. Like with the previous figures, you can change the context of the plot by right-clicking on it to select either precursor, peptide, protein-group or protein scope.

Figure 33. The “Normalization” node shows details about the normalization status of your experiment. The left side shows boxplots of precursor responses before normalization for each run. The right side shows boxplots of the same precursor responses after normalization.
Figure 34. The sample correlation matrix shows correlation of precursor quantities between all samples. The rows and columns of the matrix are ordered by condition and replicate annotation. A high correlation between technical replicates of the same sample is to be expected while low correlation between different samples might indicate biological variance. By default, the coloring range is set from 0.75 to 1.0 but can be changed via right-click option.

3.3.5 Analysis Details

The binned identification plot in the analysis details shows you the number of identifications across conditions and binned according to the three variables: iRT, intensity and m/z. This provides valuable feedback on the performance of the measurements in the different conditions and according to technical criteria such as liquid chromatography and mass spectrometer performance. The binned coefficients of variation shows CVs within conditions and binned according to the same three variables: iRT, intensity and m/z. The coefficients of variation and normalization indicates the improvement of CVs after the normalization process.

3.3.6 Differential Abundance

By default, Spectronaut will try to parse the condition and replicate the setup from the file names. In order to enable this, one must specify the name parsing schema in the settings (Settings -> General -> File Name Parsing Schema).

Parsing of the file names requires the file annotation to be separated by the "." character. By defining the meaning of the different blocks in the file name, Spectronaut is able to obtain this information automatically. There are three different types of blocks currently specified for Spectronaut. [CONDITION] which specifies the name of the sample condition, [REPLICATENR] which specifies the replicate and [OTHER] which is used to annotate blocks that are currently not relevant for parsing.
An example parsing schema for the file names as shown in Figure 27 would be
[OTHER]_[OTHER]_[CONDITION]_[REPLICATENR].

Once the experiment is set up and the analysis is done, one can inspect the candidates list in the Post-Analysis perspective to get an overview of any potentially regulated proteins in the sample (see Figure 35).

New since Spectronaut 9 is hierarchical clustering in the post-analysis. The clustering and subsequently the heatmap are generated from the same data-points that are also used for the t-test during the regulation analysis. This means that, if the t-test was performed on stripped sequences instead of protein groups, the clustering will also be calculated that way. The heatmap is one of the plots that make use of the color assignment for conditions during the setup. On the bottom of the heatmap you can see a color bar that corresponds to this color assignment (see Figure 36). The raw data of the heatmap can be exported via right-click on the plot.

![Figure 35](https://example.com/figure_35.png)

**Figure 35.** Candidates list in the Post-Analysis perspective showing all the candidates of an experiment containing 3 conditions to each 3 replicates. By default, this list is filtered by >= 1.5 absolute fold change and <= 0.01 p-value.
Figure 36. Heatmap showing the clustering of 9 runs from 3 conditions. Runs within the same condition cluster nicely as illustrated by the condition based color code in the bottom of the heatmap and the x-axis dendrogram.

Figure 37. Volcano plot showing the potential candidates of an experiment containing 3 conditions to each 3 replicates. By default, the filters are set to $\geq 1.5$ absolute fold change and $\leq 0.01$ p-value.

Spectronaut 11 offers gene ontology (GO) enrichment analysis as a new feature (Figure 38).
Figure 38. GO Enrichment analysis. Analyses are conducted independently for each comparison and namespace that is reflected in the default result grouping. To change grouping drag and drop the column labels to the blue area at the top of the page.

GO enrichment analyzes which processes, functions or compartments the regulated candidate proteins belong to. This allows one to determine areas that are affected by the experimental conditions.

During the analysis, the first step is to determine how frequently a GO term occurs in the background proteome. Based on this information the term is expected to occur a certain number of times in a candidates set of a given size. If the GO term occurs more frequently than expected it is overrepresented, if it occurs less it is underrepresented (Mi, Muruganujan, Casagrande, & Thomas, 2013). P-values corrected for multiple testing are displayed (Benjamini & Hochberg, 1995).

All filtered proteins in the candidates node or Volcano plot are candidates for the GO enrichment analysis. The background proteome consists of all unfiltered candidates (= cumulatively identified proteins).

Prerequisites for GO enrichment analysis in Spectronaut:

1) Gene Ontology: Spectronaut provides the go-basic.obo ontology.

2) Gene Annotation:
   Download Gene Annotation, e.g. from geneontology.org and load it via Databases perspective -> GO Databases -> Import Gene Annotation
   Select the Gene Annotation for your experiment via Review perspective -> Right-click on experiment -> Settings -> Experiment Setup -> Gene Annotation.

3.4 Report Perspective

After your experiment was successfully loaded in the review perspective you can switch to the Report perspective to export detailed information in spreadsheet format. Further, you can export some statistics on the scoring and identification process. The Report perspective provides you with great flexibility, allowing you to specify and save your own report schemas. In the
schemas you can specify which data should be exported and define column names such that they fit your needs.

![Figure 39. Report perspective of Spectronaut™. Experiment, LC-MS run, precursor (Elution Group) and labeled precursor (Fragment Group) centric reports can be generated. A Run report for example will have one row per run in the experiment, a Fragment report will have one row per fragment ion measured in an LC-MS run.](image)

### 3.4.1 Exporting Data Tables

Normal reports can be made on the following levels: Experiment, Run, Protein Group, Elution Group, Fragment Group and Fragment report. Lower level reports can include information from higher level reports but not vice versa. The lowest level of a report will influence how many rows a report has with a fragment report having the most rows. The most important reportable values are described below.

#### 3.4.1.1 Experiment Report

Experiment centric report. The experiment name can be reported.

#### 3.4.1.2 Run Report

Run centric report. The run label, as well as annotation used for the differential expression analysis can be reported, such as condition, fraction and replicate. Further, a number of benchmarking numbers from the quality control perspective can be reported, such as peak capacity, data points per peak and median FWHM.
### 3.4.1.3 Protein Group Report

This is a protein group centric report. The protein groups can originate from the Spectronaut protein inference (ID Picker algorithm, (Zhang et al., 2007)) or from the search engine used to generate the spectral library. For MaxQuant, as well as Proteome Discoverer, a single peptide can belong to multiple protein groups. Those peptides will be reported in multiple rows if protein group fields, as well as elution group fields, are reported at the same time.

**PG.ProteinGroups**

One or several protein groups separated with a “|”. Protein ids within protein groups are separated with a “;”. The protein groups can either originate from the Spectronaut ID Picker protein grouping or from the search engine used to generate the spectral library.

**PG.ProteinAccessions**

The protein accessions in the same order as the protein ids in “PG.ProteinGroups”. The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported.

**PG.Genes**

The genes in the same order as the protein ids in “PG.ProteinGroups”. The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported.

**PG.ProteinDescriptions**

The protein descriptions in the same order as the protein ids in “PG.ProteinGroups”. The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported.

**PG.UniProtIds**

The UniProt ids in the same order as the protein ids in “PG.ProteinGroups”. The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported.

**PG.ProteinNames**

The protein names in the same order as the protein ids in “PG.ProteinGroups”. The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported.

**PG.FastaHeaders**

The FASTA headers in the same order as the protein ids in “PG.ProteinGroups”. The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported.

### 3.4.1.4 Elution Group Report

This is the peptide centric report. According to the definition used here, a peptide in an isotopically light and heavy-labelled form would be the same peptide. It includes the charge state of the peptide but excludes any isotopic labelling. The best signal is determined automatically by the scoring algorithm. If you disagree with the automatic scoring you can manually integrate your peak in the Review perspective.

**EG.Id**

A unique ID for a peptide precursor. Can be used, for example, for downstream peptide profiling across runs.

**EG.Label**
A label for a peptide precursor compiled of the protein id, stripped sequence and the precursor ion charge. The label is not necessarily unique and therefore not intended for downstream peptide profiling across runs.

**EG.HeightCV**

In a multi-run experiment, this value reports the CV of the peptide based on the summed apex height of the detected peak.

**EG.AreaCV**

In a multi-run experiment, this value reports the CV of the peptide based on the summed area under the curve of the detected peak.

**EG.iRTEmpirical**

The iRT (Escher et al., 2012) as determined in this specific analytical run.

**EG.MeanApexRT**

The average retention time of the peak apexes across all fragment ions of this peptide.

**EG.MeanTailingFactor**

The average tailing factor of the elution group across all the fragment ions determined at the FWHM.

**EG.DeltaiRT**

The difference between the predicted and empirical iRT. A measure of the reproducibility of chromatography.

**EG.IsVerified**

Reports manual accepting or rejecting of peaks assigned in the review perspective.

**EG.Cscore**

The Spectronaut™ identification score, which is based on an advanced mProphet (Reiter et al., 2011) scoring. A high score indicates high quality identifications.

**EG.Qvalue**

If your experiment data is filtered with a qvalue cutoff of \(\leq x\), your data set will have an estimated false discovery rate (FDR) of \(x\). This is recommended to be the primary filtering option for your data. Typically, a qvalue cutoff of 0.01 is chosen. As of Spectronaut 8 by default the report is filtered by the qvalue selected in the settings.

**EG.IsUserPeak**

Specifies whether the peak was integrated automatically or manually in the review perspective.

**EG.AllProteinAccessions**

All protein accessions this peptide points to. This field is only reported when the Spectronaut protein inference was used. It represents the input to the ID Picker protein grouping algorithm.

### 3.4.1.5 Fragment Group Report

Peptide precursor centric report. According to the definition used here, a peptide in an isotopically light and heavy-labelled form would be considered different peptide precursors. It includes the charge state of the peptide, as well as the isotopic labelling. Quantitative information is reported here. If the experiment is a label free experiment one “Elution Group” corresponds to one “Fragment Group”. Hence, for a label free experiment the same number of rows is reported as for the “Elution Group Report”.

---

**BIOGNOSYS**

**NEXT GENERATION PROTEOMICS**
**FG.Id**
A unique ID of the peptide precursor. Corresponds to the EG.Id if the experiment is label free.

**FG.Label**
A label of the fragment ion group. The label is not necessarily unique and therefore not intended for structuring data.

**FG.Charge**
The charge state of the peptide precursor.

**FG.PrecMz**
The peptide precursor m/z.

**FG.PrecWindowNumber**
The precursor window number in which the precursor was measured.

**FG.TotalPeakArea**
The summed up peak area of all fragment ions for the corresponding peptide precursor.

**FG.TotalPeakHeight**
The summed up apex peak height of all fragment ions for the corresponding peptide precursor.

**FG.SignalToNoise**
The signal-to-noise of this peptide precursor.

3.4.1.6 **Fragment Report**
Fragment ion centric report. One row is reported for each measured fragment ion (this type of report can get very large). Fragment ion interferences are reported here, as well as quantitative information on the fragment ions.

**F.InterferenceScore**
Fragment ion interference score as determined by Spectronaut™.

**F.PossibleInterference**
Highlights fragment ions that Spectronaut™ considers to have an interfering signal based on within peptide interference statistics.

3.4.2 **Exporting Run Pivot Report**
The run pivot report is wide-formatted instead of long-formatted. This means that for one row a single value (e.g. intensity) is reported across runs in as many columns as there are runs. Each column header will specify the parent run, as well as the value reported. The row labels, representing the meta information, will be only present once in the report. If a selected cell value does not pass a selected filter criteria then it will be reported as “Filtered”. Peptide centric, as well as protein group centric reports, can be made.

Using pivot report one can generate a spreadsheet reporting protein (protein group) quantities with one row per protein and one column per run.

3.4.3 **Score Histograms**
Score histograms for target and decoy peptides. Sensitivity and qvalue plot.
3.5 Quality Control Perspective

The quality control perspective of Spectronaut™ is based on the peptides in the iRT kit/HRM calibration kit. Chromatography, mass spectrometer performance and analysis can be monitored over time using several performance indicators. Every successful analysis is stored in the quality control perspective. Spectronaut™ automatically detects various instruments and will create a separate quality control history for each of them. In the event that you have more than one instrument of the same type it might be useful to rename them manually. Additional folder structures can be made according to the established quality control testing performed in a specific laboratory.

Only as many runs as specified in Settings -> Global -> General -> QC Plot History Length are shown in the plots.

3.5.1 QC Panels

In addition to the calibration kit, sample specific QC panels can be created in the prepare perspective. Right-click on a spectral library and select “Enable QC”. Whenever this spectral library is used in an analysis a respective QC file is written. When selecting this novel QC panel all the corresponding QC files are filtered and can be visualized in the plots.

3.6 Pipeline Perspective

The pipeline perspective allows one to batch process a large number of runs grouped into experiments using different settings. Spectronaut works most efficiently when several experiments are processed sequentially rather than in parallel (because of disc IO). If you are not interested in manual evaluation of your peaks, the pipeline perspective might be your preferred choice. The setup of an experiment works similarly to the setup in the Review perspective. Spectronaut™ will automatically generate the report according to the settings in the chosen schema.
3.6.1 Pipeline Reporting Settings

Here you can choose which reports should be written, whether run-based or experiment-based analysis should be performed, whether scoring histograms should be reported and if the whole experiment should be saved to an SNE file.

3.7 Databases Perspective

The database perspective was first introduced in Spectronaut 11 to enable better organization and to separate settings schemas from a variety of managed databases for analysis such as FASTA files and gene annotation records.

3.7.1 Table Import

Spectronaut can remember column names in user spectral libraries. Once you import a new spectral library format into Spectronaut when setting up an experiment Spectronaut will ask you whether it should store novel synonyms. You can manage those synonyms in the column recognition settings tab.

3.7.2 Modifications

By default, modifications of MaxQuant are loaded here. These are needed for spectral library generation from MaxQuant. If you use special modifications in MaxQuant please import the corresponding modifications file into Spectronaut in order for Spectronaut to support those modifications.

The Spectronaut modifications panel provides a central location to manage modifications. Modifications specified here are used to automatically detect modified sequences in a search engine output when creating a spectral library from MaxQuant or ProteinPilot. They are also used to detect modified sequences encoded in an external user spectral library by means of...
synonyms. We recommend that modifications in a spectral library are mapped to Spectronaut modifications for optimal downstream analysis. If a spectral library is generated using Spectronaut this is taken care of automatically.

### 3.7.2.1 Importing Modifications from Search Engine

To import modifications into Spectronaut, you can batch import modifications from MaxQuant (using the modifications.xml file in its bin folder) or ProteinPilot (using “Unified Modification Catalog.xlsx”, located in the ProteinPilot/Help folder in the Program Files). It is mandatory that one imports modifications specified in a search engine to create a spectral library (except for Proteome Discoverer). When possible, Spectronaut will merge identical modification from multiple sources and save only the necessary search engine specific mapping information. However, if it is not able to unambiguously merge two or more modifications, you will be asked to resolve any conflicts at the time of import. You can tell if a modification has been mapped to multiple search engines by looking at the “Mapped to” data grid in the panel.

### 3.7.2.2 Creating custom modifications

It is also possible to specify a brand new modification. You would first save an existing modification as a new modification and then edit its content. A custom modification cannot be used for mapping modifications for spectral library generation as it would not have the search engine specific information. However, a useful workflow would be to add a custom label modification to generate a heavy library from an existing light library in the Prepare Perspective.

### 3.7.3 Protein Databases

This section of the settings perspective lets you import and manage your protein databases. Since Spectronaut 8 protein databases have been used to perform protein inference within Spectronaut and to annotate your proteomics data with fields from your FASTA database. The protein databases are imported from FASTA files and contain all of the sequences, as well as Meta information extracted from the FASTA protein headers, using the specified parsing rule.

In order to import a new proteome database from FASTA click on “Import…” at the bottom left under the file tree.

While importing a new protein database from FASTA, Spectronaut will try to find the appropriate parsing rule for this file format from the already specified rules. Should there be no matching parsing rule found, you will be asked to specify a new one. You can specify a new parsing rule by clicking “New Rule” next to the parsing rule dropdown menu.
Figure 41. Importing a new FASTA file into Spectronaut. 1. Click on “Import...” and select the FASTA file from your file system. 2. Pick an appropriate parsing rule. 3. If necessary, specify a new rule (For help generating a parsing rule please contact spectronaut@biognosys.ch). 4. Check whether the FASTA header was parsed correctly. 5. Specify a name under which you want to save this protein database. 6. Finalize your import by clicking on “Import”.

3.7.4 GO Databases

Similar to the Protein Databases, GO Databases in Spectronaut are used to further annotate your data. The GO databases section manages two different data structures.

3.7.4.1 Gene Ontologies

These allow you to import complex gene ontology graphs. These graphs are used for hierarchical grouping of functions, components and processes. The ontology itself is organism independent. Currently, Spectronaut supports the *.obo file format from the GO Consortium. The go-basic.obo is already part of the Spectronaut installation. Information from a gene ontology tree can only be used in combination with an organism specific gene annotation file.

3.7.4.2 Gene Annotations

The gene annotation file functions as a link between the protein identifier (Uniprot accession number) and the gene ontology tree. In its most basic form, the gene annotation file must feature two columns. The protein id that specifies the Uniprot accession number and the GO-ID that specifies the unique identifier of each GO term. Using this format, Spectronaut will connect the protein entries of your analysis via the GO-ID with the respective entries in the Gene Ontology to annotate your data further.

Again, the official GO Consortium annotation file (*.gaf) is recommended but you can also specify a custom annotation file. A custom annotation file mostly has two different import formats. Either the above described protein-ID and GO-ID columns, or you can also specify the GO-terms directly in the annotation file. This way, a GO-ID is not required but the columns for GO term and namespace (function, component or process) are.
3.7.5 Digest Rules

This lets you define the rule to in silico digest your proteins from the protein database(s). Digest rules have been used since Spectronaut 8 to perform protein inference within Spectronaut. A typical digest rule is the endoprotease trypsin.

3.8 Settings Perspective

Using the settings perspective of Spectronaut the user can specify analysis settings schemas and alter global settings of Spectronaut. Detailed information regarding each setting option can be obtained by hovering the mouse over the label of a specific settings variable. The analysis settings are distributed according to perspective in this manual.

3.8.1 Global Settings

The “Global” settings tab in the Settings perspective will allow you to change parameters that can be considered analysis unspecific. Here you will find options regarding plotting, working directories, as well as some general settings.

3.8.1.1 General

This section contains rather miscellaneous settings options that allow you to modify the default behavior of Spectronaut. For more information about these options use the tool-tip hover for each individual entry.

3.8.1.2 Directories

Here you can setup the different storage paths for data managed by Spectronaut. Should you have a central storage location for all your DDA raw/wiff files, you can specify this location here. This will allow Spectronaut to automatically map the correct shotgun acquisitions during the setup of the library generation pipeline. Please note that all changes within the “Directories” section will require a restart of Spectronaut in order to take effect.

3.8.1.3 Plotting

The plotting section allows you to customize the look and feel of most of the plotting options used in Spectronaut. You can specify whether XIC plots should show the integration boundaries, as well as the expected elution time. Additionally, you can also apply smoothing to your plots. For more information about these options use the tool-tip hover for each individual entry.

3.8.1.4 Reporting

This section contains parameters regarding the reports generated by Spectronaut. Here you can specify the default location for the pipeline and command line analysis of Spectronaut, as well as the naming schema that should be used. The option “Default Report Schema” allows you to specify which report you want to have automatically selected when switching from the Review to the Report perspective. The option “Store Iontraces in SNE” specifies whether the full XIC data should be stored within a saved experiment. Disabling this option will drastically decrease the size of the saved experiment file but will require the original raw files in order to visualize the XICs in review again.
3.8.2 Spectronaut™ Command Line Mode

In addition to the visual pipeline mode, Spectronaut™ is also capable of running the pipeline from command line. To run Spectronaut™ in command line mode you simply call the Spectronaut.exe file using the following parameter.

- r adds a raw file to the experiment
- d adds all raw files of a specified directory to the experiment
- a assigns a spectral library to every run in the experiment
- ar assigns a spectral library to the last run added to this experiment
- s [OPTIONAL] selects a settings schema (the schema must already exist)
  If this schema is not present, the default settings are used
- o [OPTIONAL] specifies an output directory for the reports
  If this parameter is not present, reports will be generated in %Appdata%/Spectronaut/Results
- n [OPTIONAL] specifies a name for this experiment
- f [OPTIONAL] A regex file filter to select only certain runs while using the -d parameter for selecting whole directories.
  Default parameter is “.*\.htrms”

An example:
Spectronaut.exe –d “C:\data\My Experiment” –a “C:\data\My Experiment\library.txt”
–s “my_settings” –o “C:\data\My Experiment\Results” –n “My Experiment 1” –f “.*\.wiff”

If you encounter problems with the automatic parsing of your spectral library, please first try to load the spectral library using Spectronaut’s graphical user interface and make sure that all necessary columns are recognized automatically.

4 HTRMS Converter

Spectronaut™ can directly load instrument raw data or analyze a vendor independent data format that is optimized for a fast analysis called HTRMS. If data will be analyzed several times it is beneficial to first convert the data into HTRMS as overall analysis time will be much faster. To generate HTRMS files use the converters which are installed together with Spectronaut™.

The HTRMS converter is free to use and can be run on multiple computers without the requirement of a license key.
The new HTRMS converter with multiple tasks added to the task-list and monitoring a local directory for new MS/MS files.

The HTRMS converter can be run in 3 different modes using the drop-down list menu in the Converter perspective.

4.1.1 File Conversion

In order to select one or more files from your hard-drive to be converted, click on the dropdown arrow in the Converter perspective and select “Add Files...”. After selecting one or more MS/MS files from the hard-drive an input form will show that allows you to specify the conversion parameter. Click on “OK” to add all selected files to the main task list in the Converter perspective.

You can add more tasks to the list at any moment.

4.1.2 Folder Conversion

Using the dropdown arrow and selecting “Add Folder...” will ask you to specify an input folder. The folder conversion will automatically convert all valid MS/MS files within the target folder that meet some basic filter criterion.

Using the folder conversion, you now have access to the “Batch Conversion” settings, which allow you to specify filter criteria like vendor or file age. You can also specify monitoring of the folder in order to automatically convert every new file that is added to the input folder.
4.1.3 Raw File Merging

Despite improved support for gas phase fractionated runs since Spectronaut 9, some extreme forms of gas phase fractionation might not cover enough iRT/HRM Kit peptides to ensure a successful calibration. For such cases we offer the merging of runs using the converter. For best performance we recommend that gas phase fractions of the same sample that require merging, be acquired after each other to minimize gradient fluctuations across these runs. To merge such runs, simply select the “Merge Runs” from the dropdown menu and select all the runs you want to merge. Please make sure that the runs all cover distinct precursor selection windows. Full range MS1 scans will be cut and assigned to match the respective selection windows.

5 FAQ

Why is the number of precursors smaller in the “Analysis Summary” when compared to my spectral library?

Because in the Analysis Summary precursors outside of the acquisition range are removed from the statistical calculations.

What is Spectronaut’s definition of a proteotypic peptide?

A peptide that only points to a single protein ID as defined in the parsing rule of the FASTA file and the selected protein inference algorithm. Such peptides are particularly well suited to be followed up by PRM or MRM.

Why does my spectral library have slightly less peptides than my database search?

When generating a spectral library Spectronaut will filter certain peptide spectrum matches (PSM). The most likely reason is that a PSM did not contain enough fragment ions in the MS2 spectrum to qualify for an assay in the spectrum library. You can change your filters accordingly in the library generation settings to filter less stringently. Another possible reason is that you did not use the exact same FASTA files and digest options as in your database search.

Why does the protein count in the Analysis Summary differ to my protein count?

We parse the ProteinGroupID using a “;” separator to separate individual protein IDs and a “|” to separate several protein groups. If the format in the ProteinGroupID column of your external spectral library is different this can lead to inconsistent protein counts.

Can I install several Spectronaut releases in parallel?

Yes, upon installation simply choose a different installation directory. If you already have an activated version of Spectronaut you will not have to reactivate a second installation.

The log window says “Calibration Failed” and the analysis has stopped. What is wrong?
This means that neither the iRT kit nor the HRM calibration kit could be detected. Spectronaut™ requires one of these kits to be present in order to optimize parameters to analyze your data. If no calibration kit is detected, the analysis will abort. Make sure you use the iRT kit or the HRM calibration kit in the recommended quantity for each run. For further information, please contact Biognosys at support@biognosys.com. Please also make sure to read the section entitled Supported Methods.

**How do I import my FASTA database into Spectronaut in order to use it for protein inference?**

You can find instructions how to import a FASTA file in section Error! Reference source not found..

**My problem is not listed here, what should I do now?**

Please contact Biognosys at support@biognosys.com.

### 6 References


