

MS Annika 3.0 User Manual

Micha Johannes Birklbauer, Georg Pirklbauer, Daniela Borgmann, Viktoria Dorfer, Stephan Winkler

MS Annika is a crosslink search engine based on MS Amanda, aimed at identifying crosslinks of non-cleavable and cleavable crosslinkers from MS2 and MS3 spectra.

For examples, templates, and help please refer to

- <https://github.com/hgb-bin-proteomics/MSAnnika>

Or contact us directly:

- micha.birklbauer@fh-hagenberg.at
- viktoria.dorfer@fh-hagenberg.at

Nodes

MS Annika consists of four nodes, one of which is optional:

- The **MS Annika Detector** is responsible for identifying spectra that are likely to contain crosslinks.
- The **MS Annika Search** is the actual database search engine.
- The **MS Annika Validator** calculates and applies CSM- and Crosslink-level false discovery rates and persists relevant CSMs.
- The **MS Annika XiView Exporter** can be used to export cross-links to the xiView-specific format (optional). **This node was removed in MS Annika 3.0 version 3.0.7**, it is recommended to use [MS Annika exporters](#) instead.

In the following, we describe the functionalities of these nodes as well as their parameters.

- Parameters in **purple**: Only affects cleavable crosslink MS2 search
- Parameters in **blue**: Only affects cleavable crosslink MS3 search
- Parameters in **orange**: Only affects non-cleavable crosslink search

All other parameters apply to all search approaches.

MS Annika Detector

The MS Annika Detector identifies cross-link spectra. This node splits input spectra into spectra that are likely to contain crosslinks and ones that are not. Spectra classified as unlikely to contain crosslinks can be searched in a normal search engine by attaching one to the Detector node (for example MS Amanda). All other spectra are subjected to crosslink search. Parameters for identifying crosslinks can be adjusted in the Detector node:

- **MS2 tolerance**: Tolerance used to identify if a pair of peaks (a *doublet*) is in range to be a crosslink doublet. Allowed doublet distances are calculated as the difference between heavy and light part, which can be set in the crosslink modification in the chemical modifications and the losses described in *Crosslink Modification Additions*. This tolerance is for identifying and validating Doublets.
- **Minimum Charge**: The minimum charge a peak must have to be considered in the identification of doublets. When setting the value to 0, all peaks will be used, even if the instrument or raw file reader attributed a 0 charge to the ion. We assume that these peaks carry a single charge.

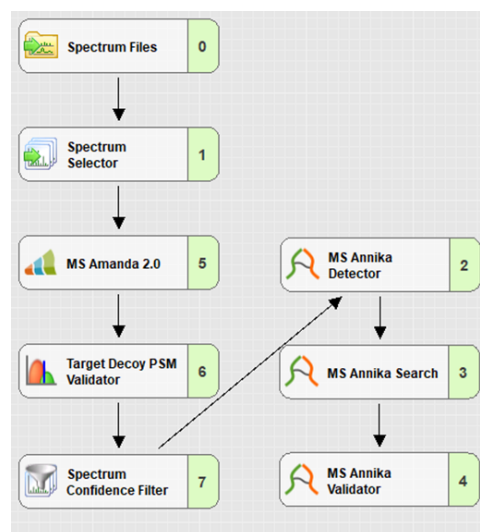


Figure 1: A typical MS Annika workflow. MS Amanda is not required but will identify linear peptides from spectra with no cross-link evidence.

- Use monoisotopic mass: Defines whether the monoisotopic or average amino acid mass should be used. Used at different stages in the search.
- Precursor mass offset: Charges between 0 and n (the value set here) are considered for the doublet search. For example: If set to 2 the MS1 precursor, the MS1 precursor minus one proton mass, and the MS1 precursor minus two proton masses are considered when deciding if a combination of doublets is a valid doublet.
- Use theoretic MS1 peaks: If set to false, only peaks found in the MS1 spectrum are used in the precursor mass offset calculation. If set to true, the precursor mass is used to calculate theoretical peaks with up to n isotope shifts, even if the peak is not in the MS1 spectrum.
- Crosslink Modification: The crosslink modification is created using the Chemical Modifications interface (*Administration > Chemical Modifications*). Create a chemical modification and fill out all the required fields. Then, mark the modification and click the *Extended Properties* button at the top of the list. Create the two fragments of the crosslinker. It is also crucial to provide the connected fragments via the *Extended Properties* window. **If more than one connection is provided, the program will notify the user about the used connection (first connection in the table). It's also important that the first connection is the one with the shortest doublet distance e.g., Alkene – Thiol instead of Alkene – Sulfenic Acid for DSSO.**
- N-Terminal Crosslink Modification: Set this value to true if you want to search for protein N-terminal crosslinking sites.
- Additional Crosslink Doublet Distances: Additional distances to be considered between the light and the heavy part of the linker. Use this if the linker can have additional losses, such as a water loss as for example exhibited by DSSO. The default doublet distance for DSSO is the distance between the alkene and the thiol fragment at 31.9721 Da – if you want to additionally consider doublets originating from the alkene and sulfenic acid fragments you would need to add their distance here, which is 49.982635 Da. Possible values can be set in the configuration page for the MS Annika Detector (*Administration > Configuration > MS Annika Detector > Additional Crosslink Doublet Distances*). Each line is a possible value. The values are parsed when the Detector node is run and will result in a warning if they cannot be parsed. Attention: If an empty line is included in the possible values, Proteome Discoverer shows an error message! In this case, check the settings (*Administration > Configuration > MS Annika Detector > Additional Crosslink Doublet Distances*) and make sure no empty lines are included.
- Diagnostic Ions: The m/z value for ions can be set here. These ions are often found in cross-link spectra and are different for each linker. The presence of diagnostic ions in a spectrum increases the likeliness for that spectrum to be selected for the cross-link search. MS Annika also reads diagnostic ions directly from the crosslinker modification definition.
- Use Diagnostic Ions for Spectrum Filtering: If true, only spectra containing diagnostic ions will be considered for search. Be sure that the correct diagnostic ions are set either in the above setting or in the crosslinker definition in Proteome Discoverer as otherwise the number of identifications is going to be very low as crosslink spectra can't be properly detected. In our experience, most non-cleavable crosslinkers don't yield enough diagnostic ions for proper detection of crosslink spectra, hence this option is turned off by default and all MS2 spectra are considered for search. (Only affects non-cleavable search).
- Doublet Pair Selection: One of evidence mode, indication mode or combined mode. Determines how cross-link spectra are searched for cross-link information.
- top N most intense doublets: Since indication mode identifies a substantial number of doublets, this parameter allows for the selection of only the most intense ones.
- Try infer missing charge states: If this option is enabled (by default) any peak in any given spectrum that does not have an associated charge state will have its charge calculated by its isotope distribution. This is especially important when dealing with MGF files, as MGF does not store fragment ion charge states. Disabling this parameter will increase search speed but MS Annika will not be able to deal with MGF data then.
- Maximum considered fragment ion charge: The maximum charge considered for calculating the associated charges of peaks without annotated charge states (used in the algorithm for the above parameter).
- Persist Doublets: A Boolean value determining whether found Doublets should be persisted. If set to true, the found doublets will be reported as a result table in Proteome Discoverer. This is false by default, since it is possible that MS Annika identifies a lot of doublets, resulting in huge result files.

Additional parameters in the Configuration Tab:

- **Additional Crosslink Doublet Distances:** Define any additional doublet distances between the light and one or more heavy crosslinker fragments here. Note that any given doublet must have the light fragment specified in the crosslinker definition set as the light fragment while the heavy fragment can vary. Values set here will be parsed and available for selection in the *Additional Crosslink Doublet Distances* setting for MS Annika Detector.
- **Diagnostic Ions:** See *Diagnostic ions* above. Set potential values for selection here as a semicolon separated list. After reloading open studies, these masses should be available for selection in the Detector node.

MS Annika Search

The MS Annika Search Engine node is very similar to the MS Amanda database search node. Usual search engine parameters can be set here. For more information see the [MS Amanda web page](#). An MS Annika CSM Validator is needed to calculate an FDR and persist CSMs and cross-links.

- **Protein Database:** The FASTA database to use. Provide it in *Administration > Maintain FASTA Files*.
- **Enzyme name:** The enzyme used for *in-silico* digestion of protein databases.
- **Enzyme Cleavage Pattern:** Regex pattern to be used to cleave protein sequences into peptides. 'Auto' uses the default cleavage pattern of the specified enzyme, e.g. `(?<=[KR])(?!P)` for Trypsin. For multi-digestion using Trypsin + AspN this would need to be set to `(?<=[KR])(?!P)|(?=D)`, for example. See discussion [here](#).
- **MS1 Tolerance:** The tolerance used to compare data at the MS1 level.
- **MS2 Tolerance:** The tolerance used to compare data at the MS2 level.
- **Missed Cleavages:** The number of missed cleavages considered in the digest of the protein database.
- **Search Approach:** Search for crosslinks in MS2 spectra or search for crosslinks in MS2 and MS3 spectra.
- **Precursor Selection:** This should match the setting selected in the 'Spectrum Selector Node'. It's recommended to select 'Use MS(n - 1) with Parent Precursors' in both the 'Spectrum Selector Node' and the 'MS Annika Search Node'. MS3 spectra are **not** visualized in the CSM result table if 'Use MS1 Precursor' is selected! (Only affects MS3 search)
- **MS3 tolerance:** Fragment ion tolerance used at the MS3 level. (Only affects MS3 search)
- **Create MS2 spectra:** If set to 'True' the peptide masses inferred from the MS3 spectra will be used to also search the precursor MS2 spectrum to identify the peptides. Otherwise, the MS2 spectra will only be searched for peptide masses directly found in the MS2 spectra. (Only affects MS3 search)
- **Multiplicative Penalty for Crosslinker with equal Sequences:** Scores of CSMs with two equal sequences will be multiplied by this value. To negate the effect of this parameter, set the value to 1. If set to a value greater than 1, CSMs containing two identical peptides are preferred.
- **Perform Decoy Search:** Whether CSM Decoy Search should be performed.

Non-cleavable crosslink search parameters:

- **Top N:** The number of highest scoring peptide candidates that should be considered for search. Higher numbers may favor more accurate results but increase search time exponentially!
- **Top N Filter:** The number of highest scoring peptide candidates to retain after exact search to calculate CSMs. The difference between this and the previous parameter is that at first only peptide candidates are selected based on an approximated correlation score. The filter is applied after calculating the exact MS Amanda score for each candidate peptidofrom arising from the candidate peptides. Generally, it is recommended to decrease the filter number with database size. E.g. for bigger databases it's favorable to go down to a Top N Filter of 2 instead of the default of 10. The search time also increases with higher Top N Filter numbers.
- **Consider Missing Isotope Peaks:** The number of isotope peaks to consider that are lighter than the reported precursor, in case the reported precursor is not the monoisotopic peak. This should be greater than zero for spectra files with lots of miss-annotated precursors.
- **Approximation Tolerance:** Tolerance in Dalton used for approximate search to identify potential peptide candidates.

- Core Usage: Percentage of CPU cores used by the approximate candidate search in MS Annika. By default uses all available resources.
- Perform Rescoring for non-cleavable CSMs: **[Removed in MS Annika 3.0 version 3.0.7]** If non-cleavable CSMs should be rescored based on additional ions calculated from both peptides. This is an experimental feature that might improve identifications at the given FDR, but only recommended when paired with comparing to non-rescored results. The safe choice is to leave this off!

Advanced parameters for non-cleavable crosslink search:

- Normalize Approximated Scores: Whether or not scores of candidate peptides should be normalized by candidate length. In most cases the search yields better results leaving this off.
- Use Gaussian Tolerance Modelling: Whether or not spectrum peaks should be modelled as Gaussian functions with $SD = Tolerance/3$. Also see the non-cleavable search publication for reference.
- Candidate Search Approach: Whether or not to use approximate search to find peptide candidates. If not, an exhaustive crosslink search is performed considering all possible combinations of peptides. This might take very long or not even complete at all, depending on the size of the protein database. An exhaustive search is only recommended for very small databases with less than 50 proteins. If set to "Auto" MS Annika will automatically choose an exhaustive search approach if the database is small enough to be feasible, and otherwise an approximate approach. We recommend leaving this setting on "Approximate" as the approximate search usually performs equally good on small databases while being a lot faster.
- Keep all non-unique CSMs: If true all non-unique CSMs will be kept in memory during search. For big datasets (large number of spectra and/or large protein databases) this means very high memory usage and possibly searches that slow down once no more memory is available. By default, this is off and only the top 5 CSMs per spectrum will be kept in memory.
- Candidate Search Backend: Which computational backend should perform candidate peptide search. By default, the CPU is used for search. The alternative option is to use the GPU, which requires an Nvidia GPU that is CUDA capable and CUDA version 12.2 or higher! We recommend using exactly CUDA version 12.2 as we can't guarantee correctness at newer CUDA versions! For more information, please refer to <https://github.com/hgb-bin-proteomics/CandidateVectorSearch/issues/32>
- Candidate Search Method: Which algorithm should be used for calculating peptide candidates. Please note that all algorithms yield the same results but not the same performance (speed). It's recommended to leave this at the default which is fast enough on most systems. The speed does not only depend on the available hardware but also on the sparsity patterns of the underlying matrices, a general overview of which algorithms perform the fastest can be found at:
 - <https://github.com/hgb-bin-proteomics/CandidateVectorSearch/blob/master/benchmarks.md>
 - <https://github.com/hgb-bin-proteomics/CandidateSearch/blob/master/benchmarks.md>
 The available algorithms are denoted as `compute_type` (e.g. `i32` = 32-bit integer) and `matrix_type` (e.g. `DV` = dense vector, `DM` = dense matrix, `SV` = sparse vector, `SM` = sparse matrix).
- Disable CUDA Check: By default, MS Annika checks if the correct CUDA version (=12.2) is installed or not and will raise an exception if not. This can be disabled, for example if the user has a newer CUDA version installed or there is an error with detecting the CUDA version on the target system. However, please be aware that we can not guarantee that MS Annika works with future CUDA versions.
- Rescoring Window Size: **[Removed in MS Annika 3.0 version 3.0.7]** The maximum number of amino acids considered in both directions of the crosslink modified residue for calculation of additional ions for rescoring. E.g. if two peptides PEPTKIDE and PEKPTIDE are cross-linked at the lysine and the window size is two, we will calculate the ions for the following fragments for the first peptide: PEPTKIDE-K, PEPTKIDE-EK, PEPTKIDE-KP, PEPTKIDE-EKP. And vice versa for the other peptide.
- Use Internal Ions for Rescoring: **[Removed in MS Annika 3.0 version 3.0.7]** If internal ions should be calculated for re-scoring. By default, re-scoring will only use n-terminal and c-terminal ions. In our experience, considering internal ions gives slightly better results. To stick with the same example as in the above parameter, not considering internal ions means we get the following fragments for the first peptide instead (with window size two): PEPTKIDE-PEK, PEPTKIDE-PEKP, PEPTKIDE-KPTIDE, PEPTKIDE-EKPTIDE.

Advanced parameters for MS3 search:

- **Boost:** If a peptide is identified in the MS2 precursor and one or more MS3 product ion spectra its score will be boosted by the number of percent specified here (see also publication for the exact scoring function). Usually, the default of 20 is sufficient and should not be changed outside of benchmarking when the ground truth is known! (Only affects MS3 search)
- **Retention Time Window for MS3 Spectra:** The estimated retention time window in which the MS3 spectra are recorded within after the initial MS2 scan with the corresponding doublet peaks is recorded. By default, MS Annika considers MS3 spectra recorded within 5 minutes after the MS2 Scan. The supplied value should be given in minutes. (Only affects MS3 search)
- **Isotopic Peak Tolerance:** Maximum allowed deviation of m/z for peaks of the same isotopic envelope e.g., if set to 0.01 a doubly charged peak will still be considered part of the isotopic envelope if it is between 0.49 and 0.51 m/z away from the next peak of the isotopic envelope. (Only affects MS3 search)
- **Maximum Intensity Difference:** Maximum relative intensity difference between peaks of the same isotopic envelope given as the fraction of the next, more intense peak e.g., if set to 0.3 a peak must have at least 30% as much intensity as the neighboring more intense peak to be considered to be of the same isotopic envelope. (Only affects MS3 search)
- **Maximum Considered Precursor Charge:** The maximum charge to be considered during search when the charge of the precursor is unknown e.g., if set to 4 the charges 1, 2, 3, 4 will be considered as possible charge states for the precursor. (Only affects MS3 search)

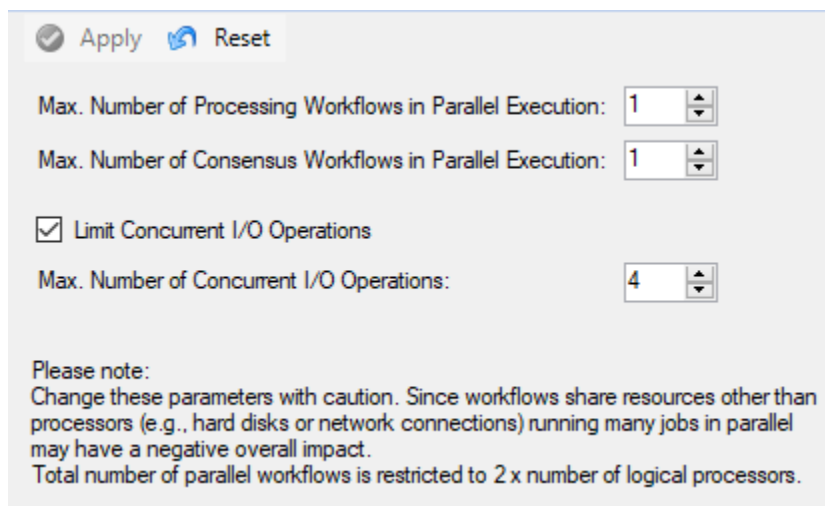
MS Annika Validator

The MS Annika Validator node allows for the calculation CSM as well as cross-link-level FDR. Furthermore, the Validator is responsible for persisting the CSMs and cross-links to Proteome Discoverer. Found CSMs are attributed with a confidence value according to the settings described below. Then all CSMs are reported in Proteome Discoverer with their respective confidence.

- **Medium confidence FDR cutoff:** Percentage for FDR calculation, usually 5%. CSMs are considered as **decoy** as soon as at least one of the crosslinker peptides is a decoy. The FDR is calculated at the CSM level.
- **High confidence FDR cutoff:** Percentage for FDR calculation, usually 1%.
- **Include Decoy CSMs/Cross-links:** Whether decoy CSMs or cross-links are included in the output. This is included here since there is no Decoy CSM or Decoy Crosslink tab in Proteome Discoverer (yet). The respective items are marked as decoy or target in the output.
- **Separate Intra/Inter-link FDR:** Determines if Intra- and Interlink CSMs are separated before FDR calculation. If set to true, an FDR calculation is done for each subset of CSMs and those two subsets are combined to yield the result.
- **Group Crosslinks by:** Choose how crosslinks are grouped in the output. They can either be grouped by the position of the cross-linker in the protein sequence or the peptide sequence.
- **DeltaCn Filter:** Retain only CSMs with a DeltaScore less than the specified value. This is only relevant when persisting multiple CSMs for one scan number. The delta score is calculated on CSMs sorted by score. The formula is $1 - \text{score}(\text{csm}_n) / \text{score}(\text{csm}_c)$ where csm_n is the CSM with the lower score, and csm_c the CSM with the higher score. This score is a measure of similarity ranging between zero and one. The more similar two results, the lower the score.
- **Individual peptide score filter:** A score cutoff. Both CSM scores (for the alpha and beta peptide) must clear this threshold, or the CSM is discarded.
- **Top N CSMs:** The number of CSMs displayed in the output for each scan number. This allows for inspection of multiple CSMs identified from a spectrum.

Recommended Settings for Proteome Discoverer

As of Proteome Discoverer version 3.1, the default setting for parallel job execution is set to 2 – we recommend lowering this back to 1. MS Annika will use all the available CPU/GPU cores if allowed and therefore running more than one job does not make sense and might actually slow MS Annika down as two instances would fight over the same resources. To change this go to *Administration > Configuration > Parallel Job Execution* and set the following:



The screenshot shows a configuration window with a light gray background. At the top left, there are two buttons: 'Apply' with a checkmark icon and 'Reset' with a circular arrow icon. Below these are three settings, each with a label and a numeric spinner control:

- 'Max. Number of Processing Workflows in Parallel Execution:' with a value of 1.
- 'Max. Number of Consensus Workflows in Parallel Execution:' with a value of 1.
- A checked checkbox labeled 'Limit Concurrent I/O Operations' followed by 'Max. Number of Concurrent I/O Operations:' with a value of 4.

At the bottom, there is a 'Please note:' section with the following text:

Change these parameters with caution. Since workflows share resources other than processors (e.g., hard disks or network connections) running many jobs in parallel may have a negative overall impact.
Total number of parallel workflows is restricted to 2 x number of logical processors.

Don't forget to hit apply after setting the values.

General Remarks

to work, MS Annika needs the right crosslinker masses. The predefined DSSO linker is in the right format: the light part and the heavy part are supplied. In the case of DSSO, a water loss in the linker is also possible. The additional doublet resulting from this must be specified in the *Additional Crosslink Doublet Distances* parameter in the Detector, for DSSO this value is already defined and can be readily selected, any other additional doublets for other crosslinkers need to be specified in *Administration > Configuration > MS Annika Detector > Additional Crosslink Doublet Distances* (see [MS Annika Detector](#)). Additional crosslinking sites can be defined in the *Maintain Chemical Modifications* tab. For more information on how to define masses, visit [the MS Annika home page](#).

Result Files

In general, there are three major tables in Proteome Discoverer that contain the results of the MS Annika searches:

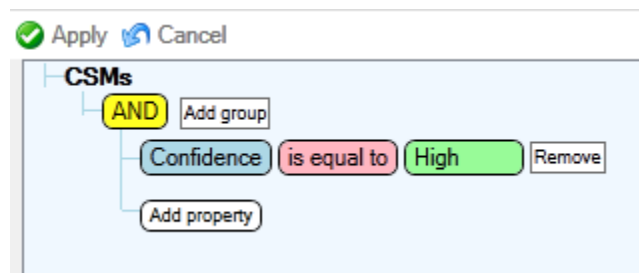
Crosslink Summary

The crosslink summary table contains a summary of the identified crosslinks in total and at specific FDRs for all input files. The columns should all be self-explanatory.

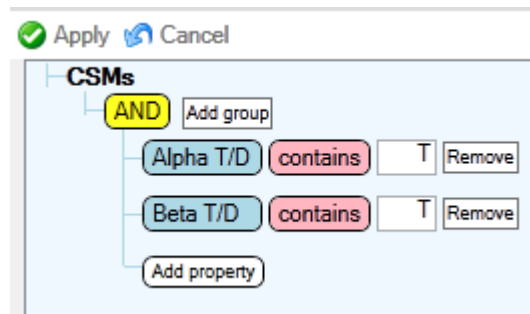
#	Checked	File name	# Cross-links	# Medium Confidence Cross-links	# High Confidence Cross-links	# All CSMs	# CSMs	# Medium Confidence CSMs	# High Confidence CSMs	# Interlink CSMs	# Intralinks CSMs	# Crosslink Spectra	# Found Doublets	File path
1	<input type="checkbox"/>	Total	305	257	225	2493	843	814	729	24	819	5152	0	
2	<input type="checkbox"/>	XLpeptide_Beveridge_GEx-HFX_DSS_R1.raw	305	257	225	2493	843	814	729	24	819	5152	0	C:\Users\P42587\Documents\data_ms\XLpeptide_Beveridge_GEx-HFX_DSS_R1

CSMs

The CSMs table shows information of all identified unique CSMs – that includes decoys and CSMs above any of the FDR thresholds. To display only high-confidence CSMs you can filter using the in-built Proteome Discoverer filter like this:



If you only want target CSMs you can filter like this:



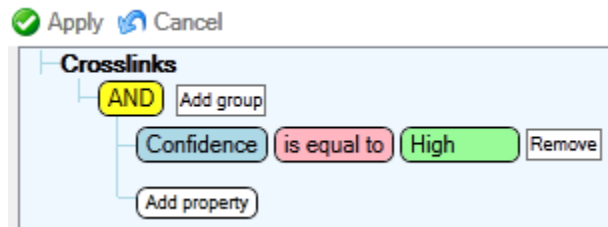
The specific columns in this table are:

- **Sequence:** The sequences of the two cross-linked peptides, delimited by “-”.
- **Crosslinker:** Name of the used crosslinker.
- **Crosslink Type:** *Intra* if it's a crosslink within the same protein or *Inter* if the crosslink is between different proteins.
- **Crosslink Strategy:** Either *MS Annika* if MS2 search or *MS Annika (MS2/MS3)* if MS3 search.
- **m/z [Da]:** The mass to charge ratio of the precursor ion in Dalton.
- **Charge:** The charge of the complete cross-linked precursor ion.
- **MH+ [Da]:** The measured protonated monoisotopic mass of the cross-linked peptides in Dalton.
- **First Scan:** The scan number of the corresponding CSM.
- **RT [min]:** The retention time of the corresponding spectrum of the CSM.
- **Delta-M [ppm]:** The relative difference in ppm between the theoretical mass of the entire crosslink and the experimental mass of the precursor ion.
- **Sequence A:** Amino acid sequence of the alpha peptide.

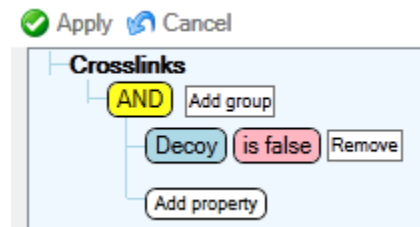
- Crosslinker Position A: The position of the crosslinker in the alpha peptide (1-based).
- Charge A: The charge of the alpha peptide (if it could be determined, this mostly only works for cleavable crosslinkers).
- m/z A [Da]: The mass to charge ratio of the alpha peptide.
- Delta-M A [ppm]: The relative difference in ppm between the theoretical mass of the alpha peptide and the experimental mass of the (doublet) precursor ion (mostly only relevant for cleavable crosslinkers).
- Sequence B: Amino acid sequence of the beta peptide.
- Crosslinker Position B: The position of the crosslinker in the beta peptide (1-based).
- Charge B: The charge of the beta peptide (if it could be determined, this mostly only works for cleavable crosslinkers).
- m/z B [Da]: The mass to charge ratio of the beta peptide.
- Delta-M B [ppm]: The relative difference in ppm between the theoretical mass of the beta peptide and the experimental mass of the (doublet) precursor ion (mostly only relevant for cleavable crosslinkers).
- Spectrum File: The name of the input file from which the fragment spectrum was retrieved.
- File ID: The short identifier of the input file from which the fragment spectrum was retrieved.
- Delta Cn: Normalized difference of the CSM score to the highest CSM score.
- Accession A: Accession(s) of the protein(s) the alpha peptide originates from. We recommend your fasta file headers to conform to the UniProt standard, otherwise Proteome Discoverer might not parse them correctly.
- Accession B: Accession(s) of the protein(s) the beta peptide originates from. We recommend your fasta file headers to conform to the UniProt standard, otherwise Proteome Discoverer might not parse them correctly.
- Uncharged Mass A: Calculated uncharged mass of the alpha peptide.
- Modifications A: Identified post-translational modifications of the alpha peptide.
- Modifications B: Identified post-translational modifications of the beta peptide.
- Uncharged Mass B: Calculated uncharged mass of the beta peptide.
- Compensation Voltage: FAIMS CV of the corresponding spectrum (if available).
- Activation Type: The activation type for acquisition of the fragment spectrum of the CSM, e.g. CID or HCD.
- NCE [%]: Normalized collision energy in percent.
- Stepped Collision Energies: If a stepped collision approach was used, all used stepped collision energies are reported here.
- Number of found Doublets: For cleavable crosslink search, the number of found crosslink doublets is reported in this column.
- Original (non-rescored) Score A: The original score of the alpha peptide before re-scoring (re-scoring is off by default and only possible with non-cleavable crosslink searches).
- Original (non-rescored) Score B: The original score of the beta peptide before re-scoring (re-scoring is off by default and only possible with non-cleavable crosslink searches).
- A in protein: Starting position(s) of the alpha peptide in the corresponding protein(s) (0-based).
- B in protein: Starting position(s) of the beta peptide in the corresponding protein(s) (0-based).
- Score Alpha: The score of the alpha peptide.
- Score Beta: The score of the beta peptide.
- Combined Score: The score of the CSM (minimum score of the two peptides).
- Alpha T/D: If the alpha peptide is a target (T) or decoy (D) hit.
- Beta T/D: If the beta peptide is a target (T) or decoy (D) hit.
- Matched Ions A: The number of matched ions for the alpha peptide.
- Search Space A: The search space of the alpha peptide.
- Matched Ions B: The number of matched ions for the beta peptide.
- Search Space B: The search space of the beta peptide.

Crosslinks

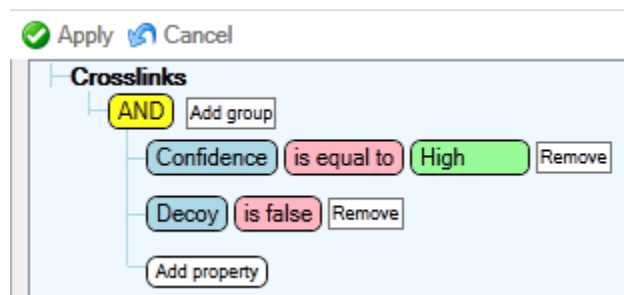
The Crosslinks table displays all identified unique crosslinks – that includes decoys and crosslinks above any of the FDR thresholds. To display only high-confidence crosslinks you can filter using the in-built Proteome Discoverer filter like this:



If you only want target CSMs you can filter like this:



Usually for downstream analysis you want a combination of both filters:



The specific columns in this table are:

- Crosslinker: Name of the used crosslinker.
- Crosslink Type: *Intra* if it's a crosslink within the same protein or *Inter* if the crosslink is between different proteins.
- # CSMs: The number of supporting CSMs for the crosslink.
- # Proteins: The number of proteins which contain the peptide sequences of the crosslink.
- Sequence A: Amino acid sequence of the alpha peptide.
- Accession A: Accession(s) of the protein(s) the alpha peptide originates from. We recommend your fasta file headers to conform to the UniProt standard, otherwise Proteome Discoverer might not parse them correctly.
- Position A: Position of the crosslinker in the alpha peptide.
- Sequence B: Amino acid sequence of the beta peptide.
- Accession B: Accession(s) of the protein(s) the beta peptide originates from. We recommend your fasta file headers to conform to the UniProt standard, otherwise Proteome Discoverer might not parse them correctly.
- Position B: Position of the crosslinker in the beta peptide.
- Protein Descriptions A: Protein description(s) of the associated protein(s) of the alpha peptide as parsed from the header in the fasta file.

- Protein Descriptions B: Protein description(s) of the associated protein(s) of the beta peptide as parsed from the header in the fasta file.
- Best CSM Score: The score of the crosslink (which is the maximum score of all associated CSM scores).
- In protein A: Position(s) of the crosslinker in the associated protein(s) of the alpha peptide (1-based).
- In protein B: Position(s) of the crosslinker in the associated protein(s) of the beta peptide (1-based).
- Decoy: Whether the crosslink is a decoy crosslink (at least one of the peptides is a decoy hit) or a target crosslink (both of the peptides are target hits).
- Modifications A: Identified post-translational modifications of the alpha peptide.
- Modifications B: Identified post-translational modifications of the beta peptide.
- Confidence: Confidence according to the parameters chosen in the MS Annika Validator node. Usually high-confidence (green) corresponds to 1% FDR, medium-confidence (yellow) corresponds to 5% FDR and low-confidence (red) corresponds to above 5% FDR.