



Crosslink identification with MS Annika 3.0

MS Annika 3.0 is a cross-linking search engine for use with cleavable and non-cleavable crosslinkers and MS2 or MS2-MS3 spectra. It can deal with a wide variety of crosslinkers and provides robust and transparent FDR control based on a target-decoy approach. MS Annika 3.0 is available as nodes for Proteome Discoverer and can be downloaded and used free of charge.

This tutorial covers the installation of Proteome Discoverer and MS Annika, how to setup a crosslink search with MS Annika, which results are provided by MS Annika, and finally an exemplary down-stream analysis of the MS Annika results.

This tutorial will use Proteome Discoverer version 3.0 and MS Annika 2.0 version 1.1.4 but should be applicable to any supported Proteome Discoverer and MS Annika version (some features might not be available in earlier MS Annika versions).

A video version of this tutorial can be found at:

<https://www.youtube.com/watch?v=L1IVt35PYv4>

References

More information about MS Annika is given in the MS Annika and MS Annika 2.0 publications:

- <https://doi.org/10.1021/acs.jproteome.0c01000>
- <https://doi.org/10.1021/acs.jproteome.3c00325>

As well as on the MS Annika webpage and GitHub repository:

- <https://ms.imp.ac.at/?action=ms-annika>
- <https://github.com/hgb-bin-proteomics/MSAnnika>

Contact

- Viktoria Dorfer (project head): viktoria.dorfer@fh-hagenberg.at
- Micha Birklbauer (developer): micha.birklbauer@fh-hagenberg.at
- Research group: <https://bioinformatics.fh-hagenberg.at/>

Table of Contents

Chapter	Page
Data	3
Downloading and installing Proteome Discoverer	4
Downloading and installing MS Annika	7
Importing the FASTA database in Proteome Discoverer	9
Creating a new study in Proteome Discoverer	10
Importing MS files in Proteome Discoverer	12
Working with MGF files	14
Working with timsTOF files	14
Creating an MS Annika workflow in Proteome Discoverer	15
Loading a template MS Annika analysis workflow	27
Inspecting MS Annika results	29
Down-stream analysis of MS Annika results	33
Example: Export to xiVIEW	34
Defining a crosslinker in Proteome Discoverer	41
Changing MS Annika settings and Additional Crosslink Doublet Distances	47
Additionally recommended Proteome Discoverer nodes	49
Getting help	50

Data

To run a crosslink search with MS Annika you will need at least two files:

- A file containing the recorded mass spectra, either in Thermo RAW format or in Mascot Generic Format (MGF).
- A database containing the sequences of the proteins that you expect in your sample in FASTA format.
- It's also necessary to know which crosslinker was used for the cross-linking experiment and which post-translational modifications you are expecting to see on your proteins.

If you don't have any of those files for yourself but still want to give MS Annika a try, you can find example files here:

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

Downloading and installing Proteome Discoverer

Because MS Annika is implemented as nodes within the Proteome Discoverer software suite, the first step is download and install Proteome Discoverer. Proteome Discoverer can be downloaded here:

<https://www.thermofisher.com/at/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html>



Click on “Free demo download” and you will be forwarded to login with your Thermo Scientific account. In case you don’t have an account, you can sign up for free. You will be presented with a list of available software, select “Proteome Discoverer” from that list.



Life Sciences Mass Spectrometry
Software Download and Licensing Portal

Software & Services

Product List

Product Search

Order History

Search Line Items

Files Not Downloaded

Recent Product Releases

Recent Files

Recent Emails

Licensing

Search Licenses

Offline Activation

Download Activation Codes

Administration

Change Password

Email Preferences

Product Preferences

Download Preferences

Your Profile

Information

FACs

ESDM User Manual

Support

Sessions

Logout

Product List

Welcome to the Thermo Fisher Scientific Life Sciences Mass Spectrometry and Chromatography Download Portal where you can manage your software from a secure, personalized Web site. To access an item, select a product below.

LSMS Products

To report an issue or concern you might have with any LSMS product, contact lra_issues.sl@thermofisher.com

Please use the following format when reporting an issue:

E-mail Subject: [Product Name] [Version] [Short Description of the software issue]

E-mail Body: [Detailed description of the problem] [Indicate steps to reproduce]

Chromleon CDS software

If you want to report an issue or concern you might have with Chromleon CDS software and have an active Chromleon Enterprise Contract, please contact support.informatics@thermofisher.com

If you do not have a Chromleon Enterprise Contract, please refer to the webpage <http://www.unifylabservices.com/en/contact-us.html>

[Application - STORBoard](#)

[Drivers for Chromleon](#)

[Instrument - Orbitrap Astral Series](#)

[Instrument - Orbitrap Tribrid Series](#)

[Instrument - Orbitrap Exploris Series](#)

[Instrument - Exactive Series](#)

[Instrument - TSO Series II](#)

[Instrument - TSO Series I](#)

[Instrument - LTQ Orbitrap Series](#)

[Instrument - LTQ Series](#)

[Instrument - LC Devices](#)

[Instrument - SII for Xcalibur](#)

[Instrument - Vanquish](#)

[Instrument - Aria MX](#)

[Instrument - GC & GC-MS](#)

[Application - Xcalibur](#)

[Application - Proteome Discoverer](#)

[Application - Compound Discoverer](#)

[Application - ProSightPD](#)

After selecting “Proteome Discoverer” you will be forwarded to the download page where you should select the version that you wish to install. For this tutorial we will go for version 3.0.

Product Information

Application - Proteome Discoverer

Select a version. To access older versions, click on the “[RELEASE_ARCHIVE]” tab

New Versions		Release Archive
Version	Description	
	Proteome Discoverer 3.0 SP1	
	Proteome Discoverer 3.0	
2.5	Proteome Discoverer 2.5 QF1	
2.5	Proteome Discoverer 2.5 QF2	
	Proteome Discoverer 2.5	
	Proteome Discoverer 2.4 SP1	
	Proteome Discoverer 2.4 QF1	
	Proteome Discoverer 2.3	
	Proteome Discoverer 2.1 SP1	

Download “Thermo Proteome Discoverer 3.0” and optionally (but recommended) “Thermo Proteome Discoverer ThirdParty 3.0”.

Product Download

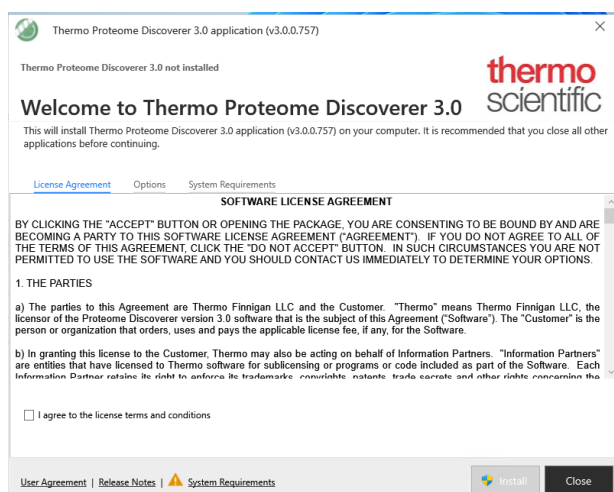
Proteome Discoverer 3.0

© Download Help

The software you are about to download is subject to export control laws and regulations. By downloading this software, you agree that you will not knowingly, without prior written authorization from the competent government authorities, export or reexport - directly or indirectly - any software downloaded from this website to any prohibited destination, end-user, or end-use.

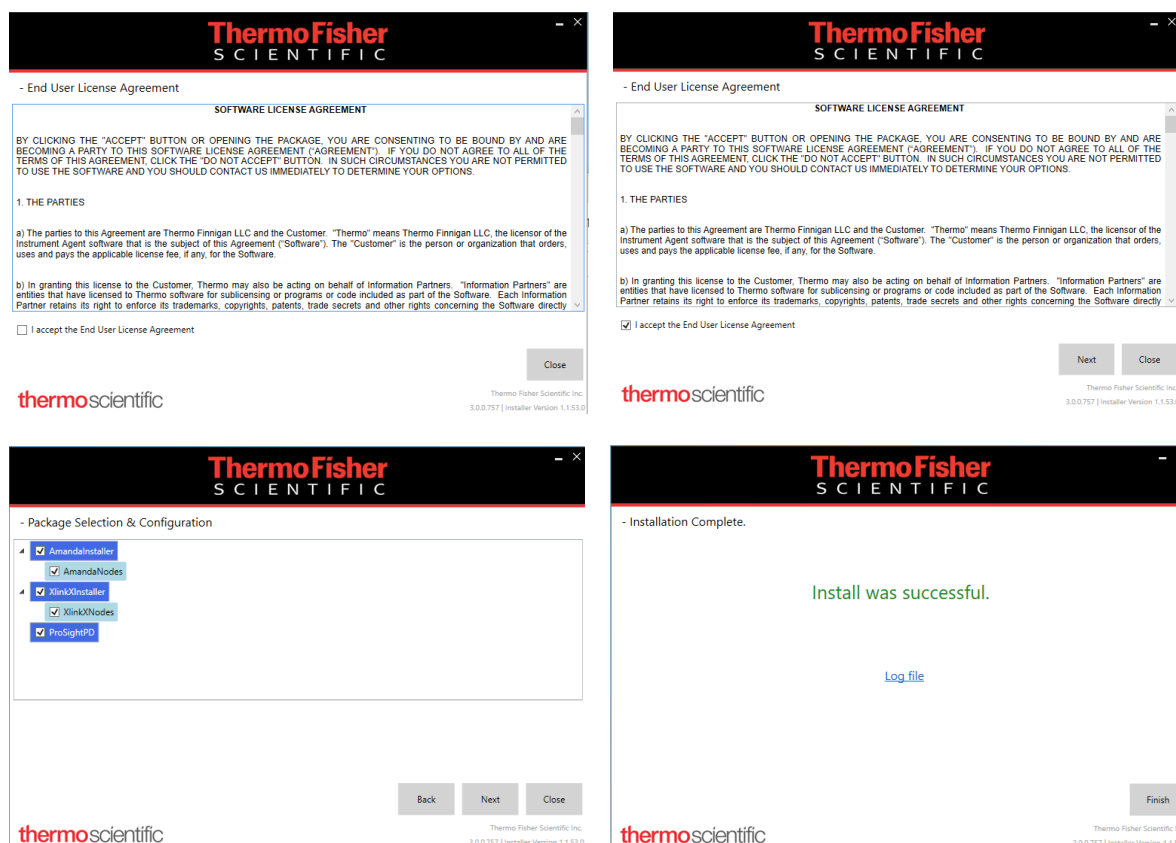
File Description	File Size	File Name
Proteome Discoverer 3.0 Release Notes	130.5 KB	Proteome Discoverer 3.0 Release Notes.pdf
Scripting Node Examples	2.9 GB	Scripting Node Examples.zip
Thermo Proteome Discoverer 3.0	539 MB	Thermo Proteome Discoverer 3.0.exe
Thermo Proteome Discoverer Daemon 3.0	232.4 MB	Thermo Proteome Discoverer Daemon 3.0.exe
Thermo Proteome Discoverer Familiarization Data Chapter 1	166.8 MB	Thermo Proteome Discoverer Familiarization Data Chapter 1.zip
Thermo Proteome Discoverer Familiarization Data Chapter 2	1.3 GB	Thermo Proteome Discoverer Familiarization Data Chapter 2.zip
Thermo Proteome Discoverer Familiarization Data Chapter 3	12.6 GB	Thermo Proteome Discoverer Familiarization Data Chapter 3.zip
Thermo Proteome Discoverer Familiarization Data Chapter 4	223.7 MB	Thermo Proteome Discoverer Familiarization Data Chapter 4.zip
Thermo Proteome Discoverer Familiarization Data Chapter 5	331.5 MB	Thermo Proteome Discoverer Familiarization Data Chapter 5.zip
Thermo Proteome Discoverer ThirdParty 3.0	89 MB	Thermo Proteome Discoverer ThirdParty 3.0.exe

Install Proteome Discoverer. The installation process requires administrative privileges.



Optional: Install the Proteome Discoverer Third Party nodes. In case you are not installing the Proteome Discoverer Third Party nodes it is highly recommended to install MS Amanda. Potentially, there is a newer version of MS Amanda available than provided in the Third Party Nodes, you can download the latest version here:

<https://ms.imp.ac.at/?goto=msamanda>



If the installation was successful, you should be able to start Proteome Discoverer now!
Make sure to exit Proteome Discoverer before installing MS Annika!

Downloading and installing MS Annika

MS Annika can be either downloaded from the MS Annika webpage at <https://ms.imp.ac.at/?action=ms-annika> which also includes some additional information, or alternatively from the MS Annika GitHub repository at <https://github.com/hgb-bin-proteomics/MSAnnika> which we will use for this tutorial. Clicking the download link for MS Annika 2.0 for Proteome Discoverer 3.0 will download the latest MS Annika 2.0 version and the user manual compressed as a ZIP folder to your computer. MS Annika 3.0 is available for Proteome Discoverer 3.1.

MS Annika

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

You can read more about MS Annika [here](#) and [here](#).

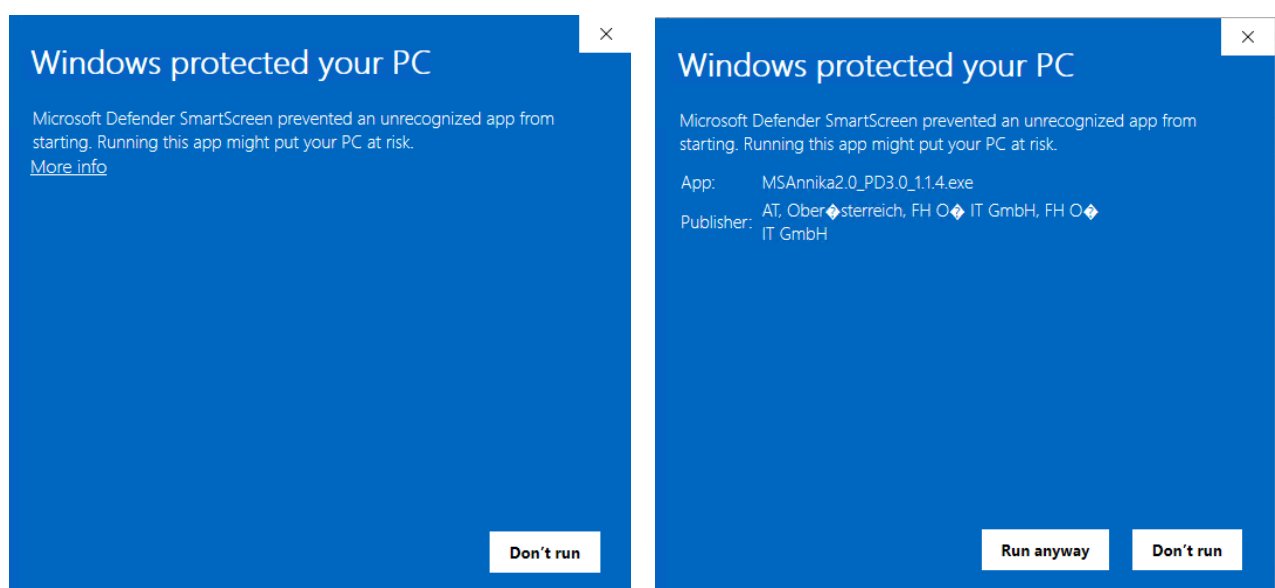
This repository contains the latest release versions of MS Annika.

- Latest MS Annika 2.0 [version](#) for Proteome Discoverer 3.1: [download](#)
- Latest MS Annika 2.0 [version](#) for Proteome Discoverer 3.0: [download](#)
- Latest MS Annika 2.0 [version](#) for Proteome Discoverer 2.5: [download](#)
- Latest MS Annika [version](#) for Proteome Discoverer 2.4: [download](#)
- Latest MS Annika [version](#) for Proteome Discoverer 2.3: [download](#)

A list of changes in every version can be found in [HISTORY.md](#).

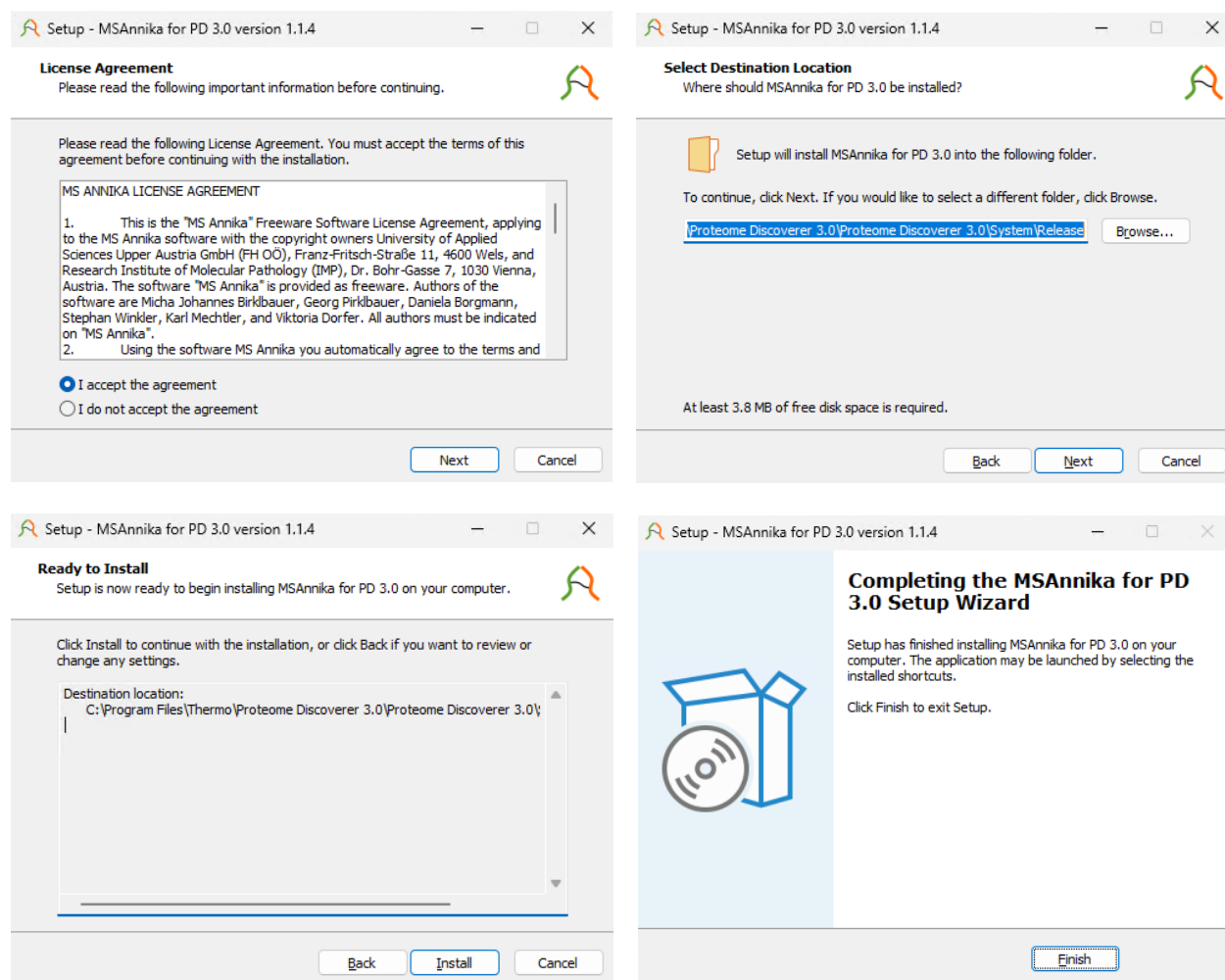
Make sure that Proteome Discoverer is not running during the installation of MS Annika!

Extract the ZIP file and double click the MS Annika 2.0 installer, in case a Microsoft Defender warning pops up, click on "More info" and "Run anyway".

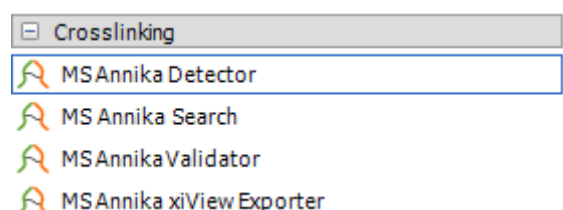


In case any other antivirus software is blocking the application, make sure to whitelist the MS Annika installer.

Installing MS Annika also requires administrative privileges, therefore a user-account-control prompt should pop up before the MS Annika installer launches. After that simply follow the instructions on screen. The installation path should be pre-set and no change should be needed.

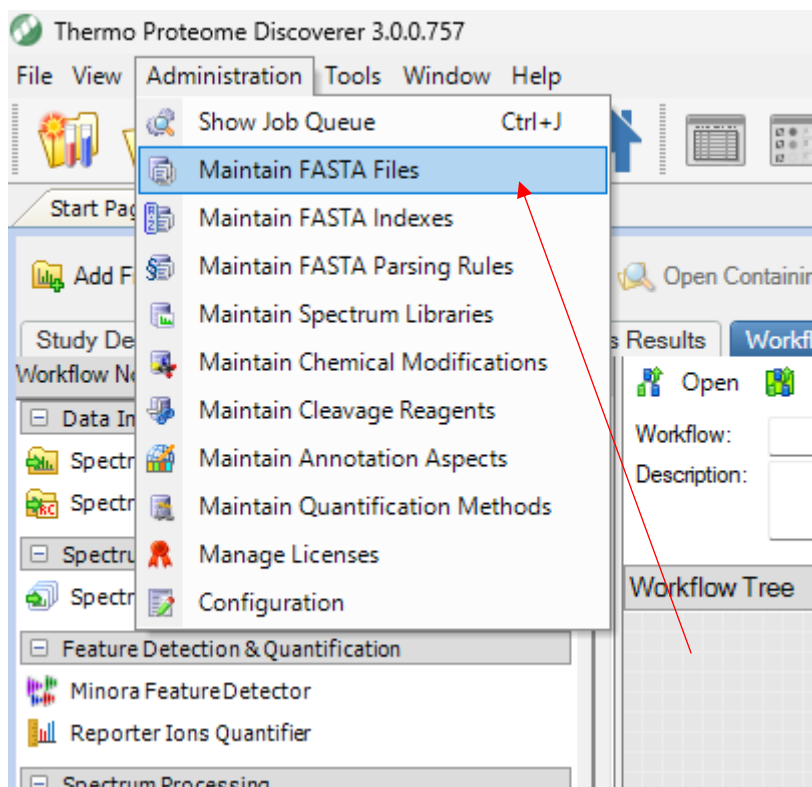


If the installation was successful, the MS Annika nodes should now be available in Proteome Discoverer.

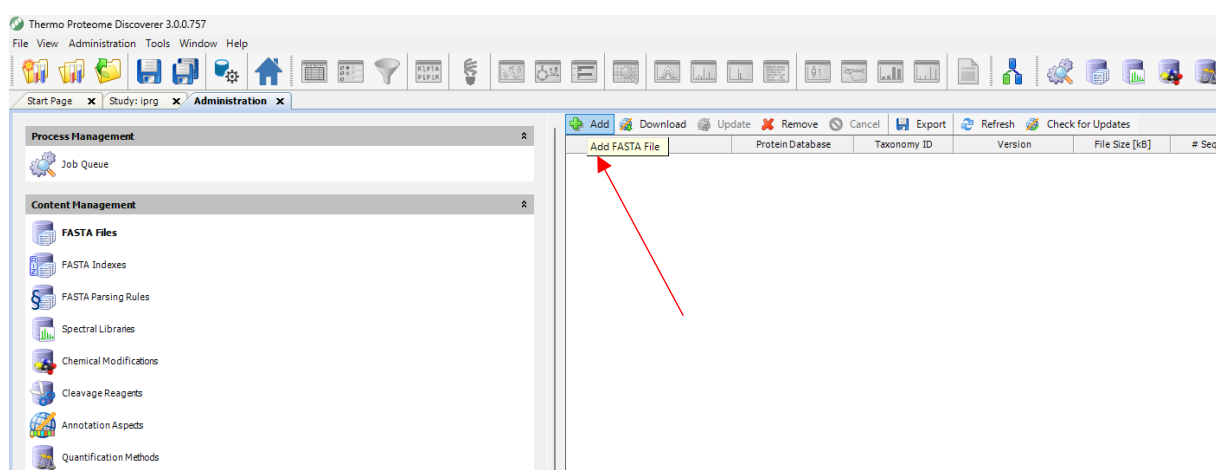


Importing the FASTA database in Proteome Discoverer

In Proteome Discoverer go to the “Administration” menu in the top left and select “Maintain FASTA Files”.



Click the “Add” button to add a FASTA file.

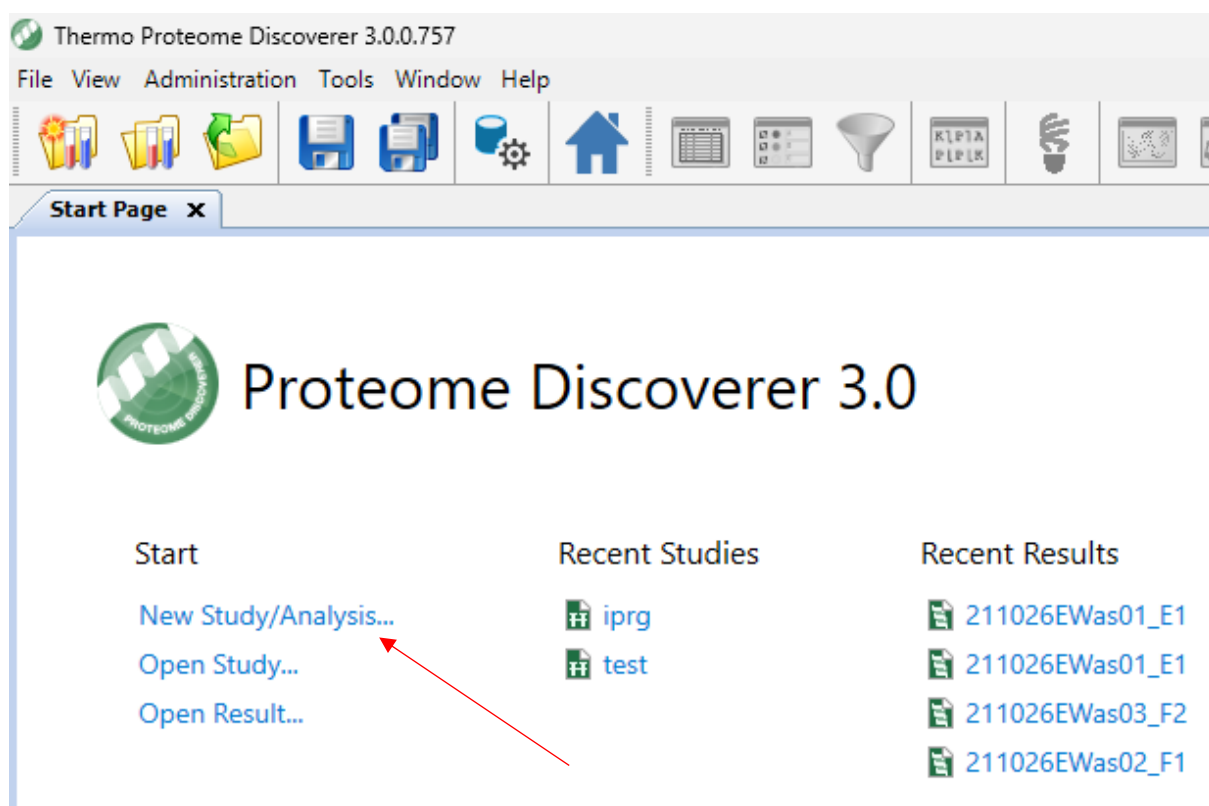


You will need to select the FASTA file in the explorer window that pops up. After clicking “Open”, Proteome Discoverer will switch to the Job Queue where an import job for the FASTA file will be running. If the job successfully completes the FASTA file will be available in Proteome Discoverer.

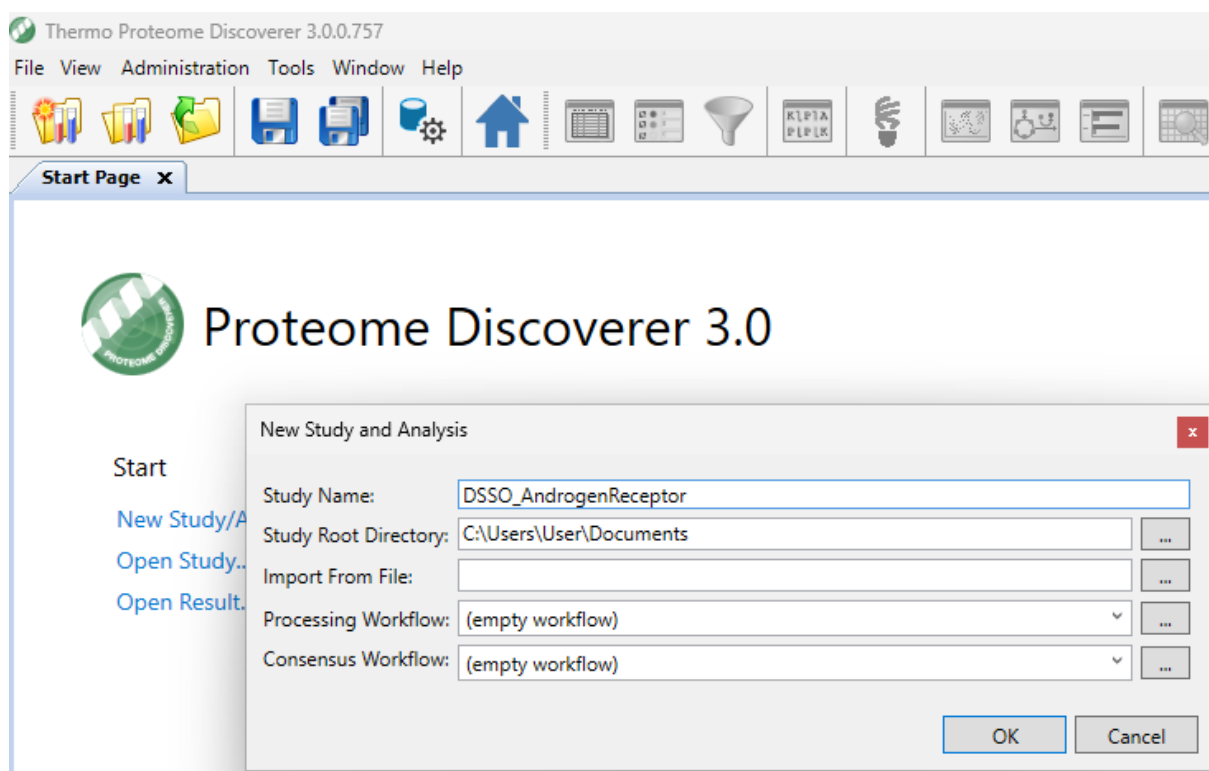
Completed	OK	100%	Processing	Import of HiRes.fasta	8/8/2023 6:27 AM
-----------	----	------	------------	-----------------------	------------------

Creating a new study in Proteome Discoverer

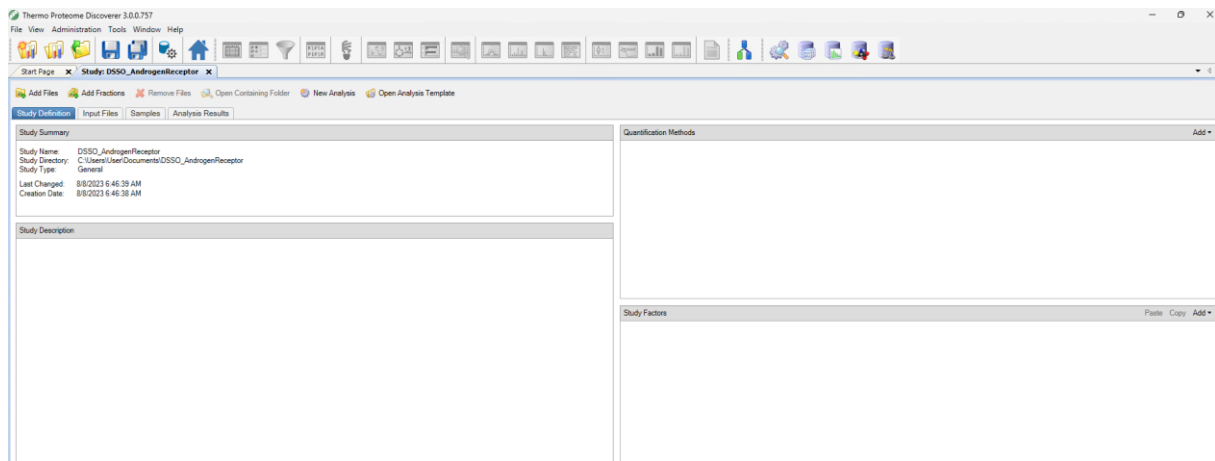
Go to the Start Page of Proteome Discoverer and select "New Study/Analysis..."



Choose a name and the path for your study and hit "OK".

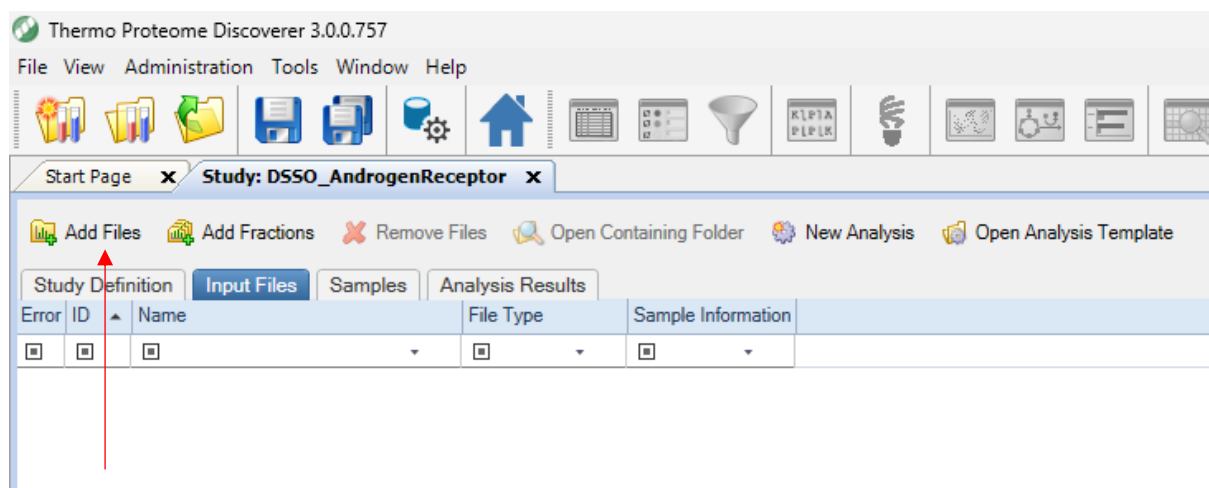


A new tab with your study should open that gives a basic overview of the study.

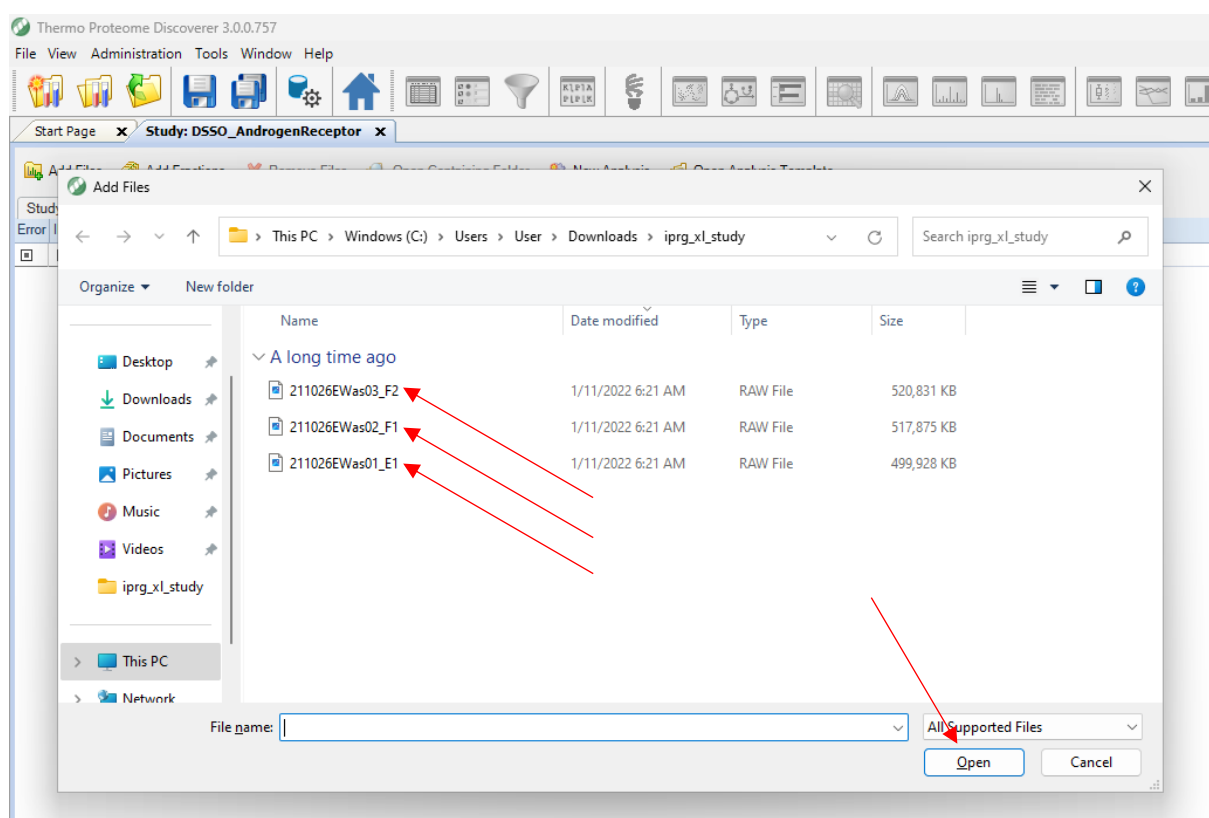


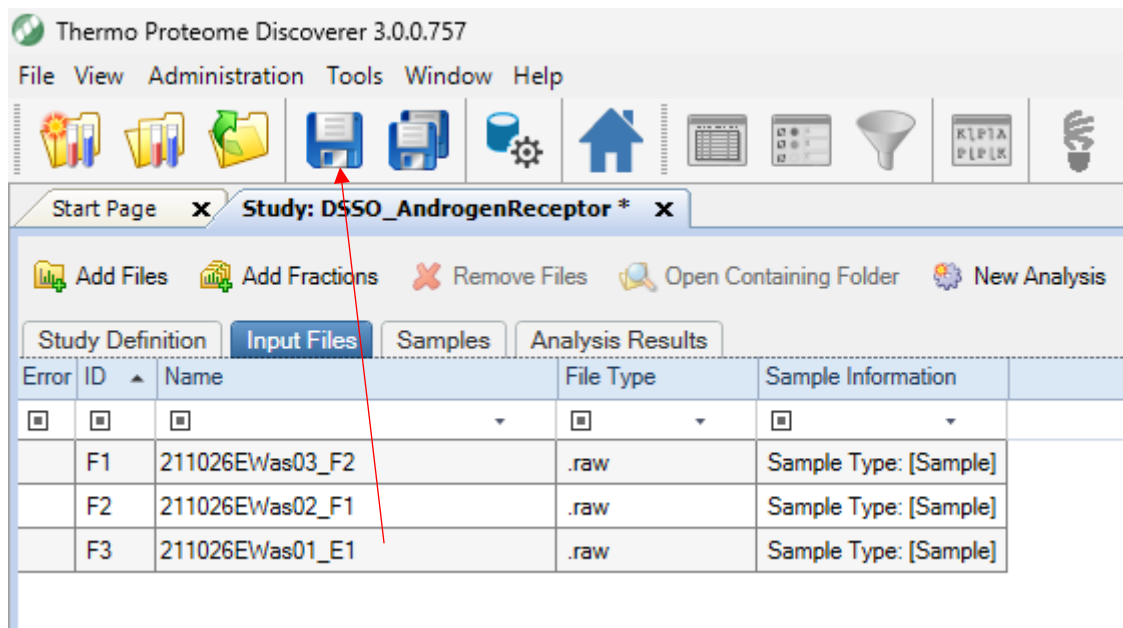
Importing MS files in Proteome Discoverer

Next you want to import your RAW files in Proteome Discoverer. For that go to the tab “Input Files” in your study and click “Add Files” in the top left corner of your study.



An explorer window will open where you can select all RAW files that you want to import. After confirming with “Open” they will appear in your “Input Files” tab.





Don't forget to also save your study every other step!

Working with MGF files

Importing an MGF file works the same way as importing a RAW file. Please do note, that not all MS Annika versions support MGF file input, for a full list of MS Annika versions that support processing of MGF files please refer to the according section of the MS Annika ReadMe file:

<https://github.com/hgb-bin-proteomics/MSAnnika#support-for-mgf-and-timstof-data>

Moreover, MGF files are only supported for MS2 searches in MS Annika. This is due to the fact that MGF lacks required information that is necessary to perform an MS3 search.

Working with timsTOF files

All MS Annika versions that support MGF file input also support processing of timsTOF ion mobility data. Again, the full list of MS Annika versions that support timsTOF data input can be found in the MS Annika ReadMe file here:

<https://github.com/hgb-bin-proteomics/MSAnnika#support-for-mgf-and-timstof-data>

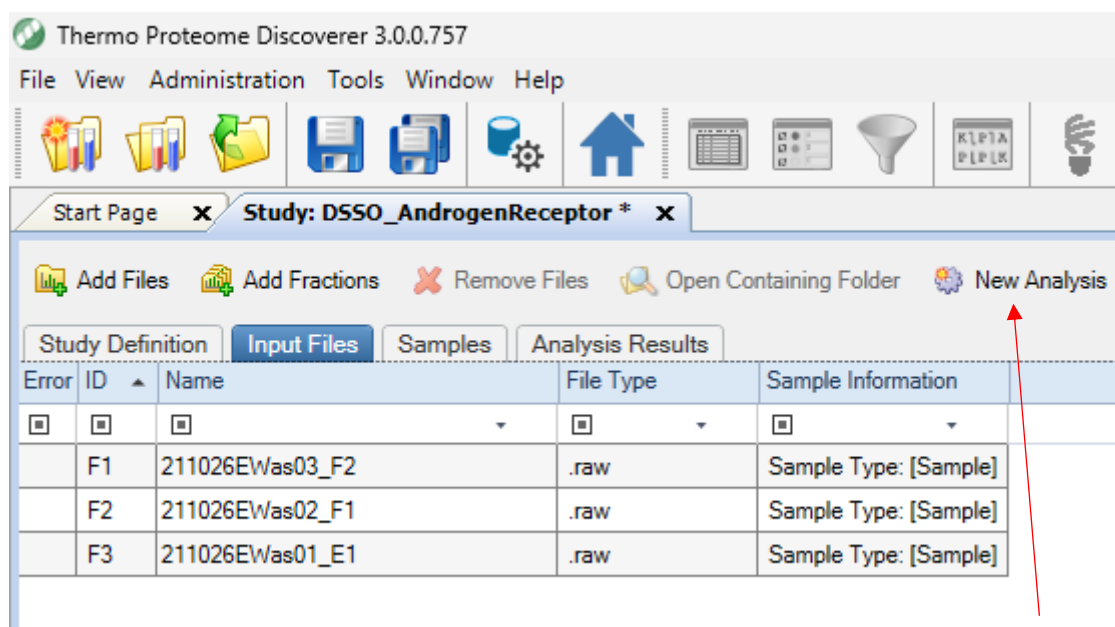
However, it should be noted that reading ion mobility data requires installation of the Bruker Ion Mobility Reader node in Proteome Discoverer which is not publicly available. Please contact us via

- micha.birklbauer@fh-hagenberg.at or
- viktor.dorfer@fh-hagenberg.at

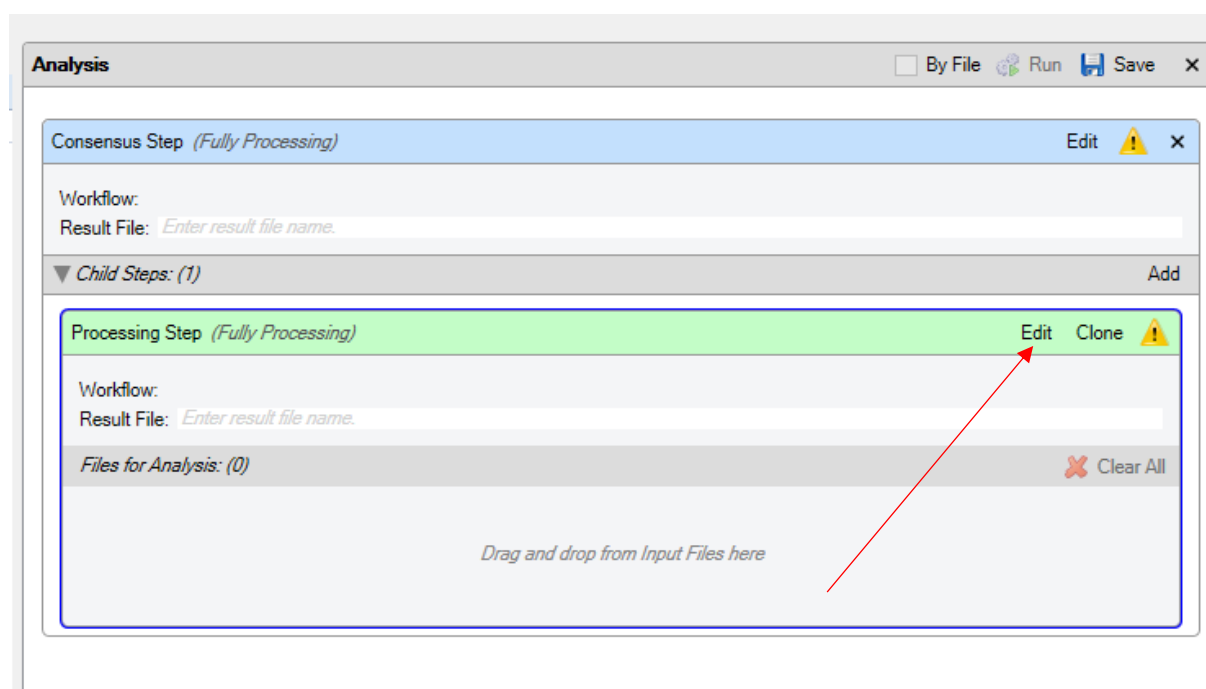
if you want to use MS Annika with timsTOF data. Furthermore, only MS2 search is supported for timsTOF data. Additionally, it should be noted that MS Annika does not make use of the ion mobilities for crosslink identification – the Bruker Ion Mobility Reader node is simply used for displaying ion mobilities in Proteome Discoverer but does not affect results of the crosslink search.

Creating an MS Annika workflow in Proteome Discoverer

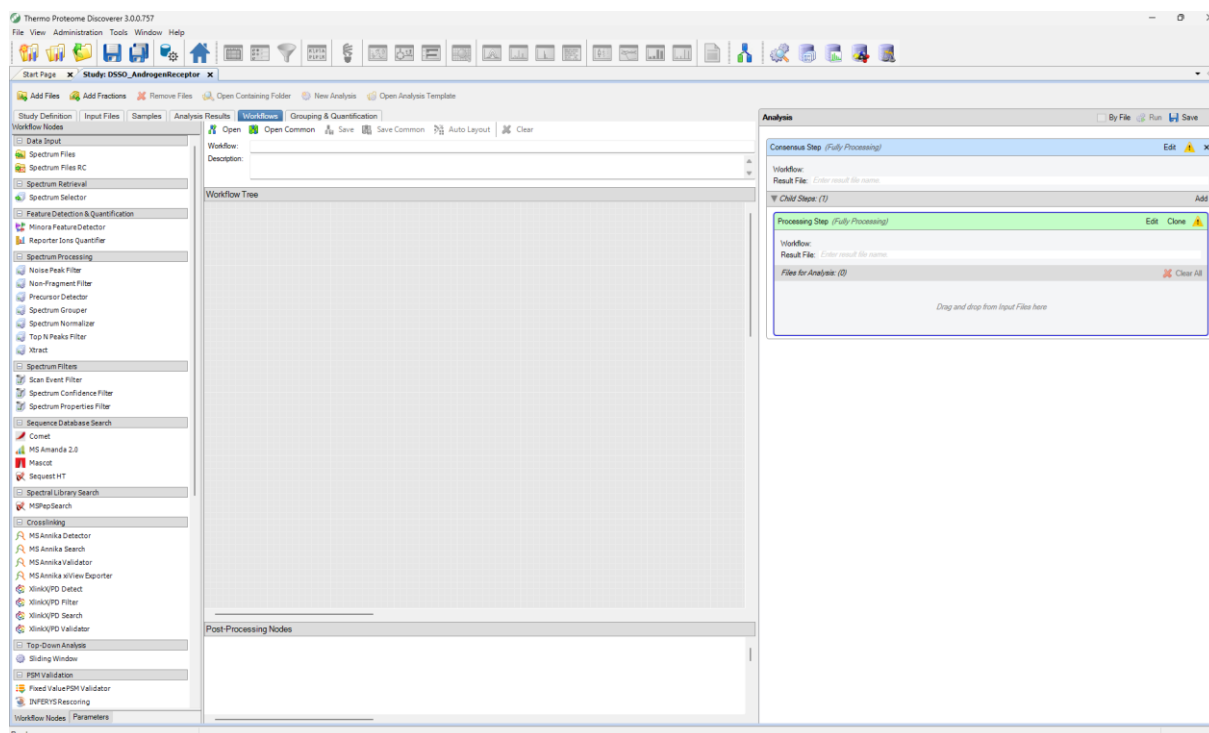
There are two ways to create an MS Annika workflow, either creating a new analysis from scratch or opening an analysis template. This section will cover the first approach. In your study click on "New Analysis".



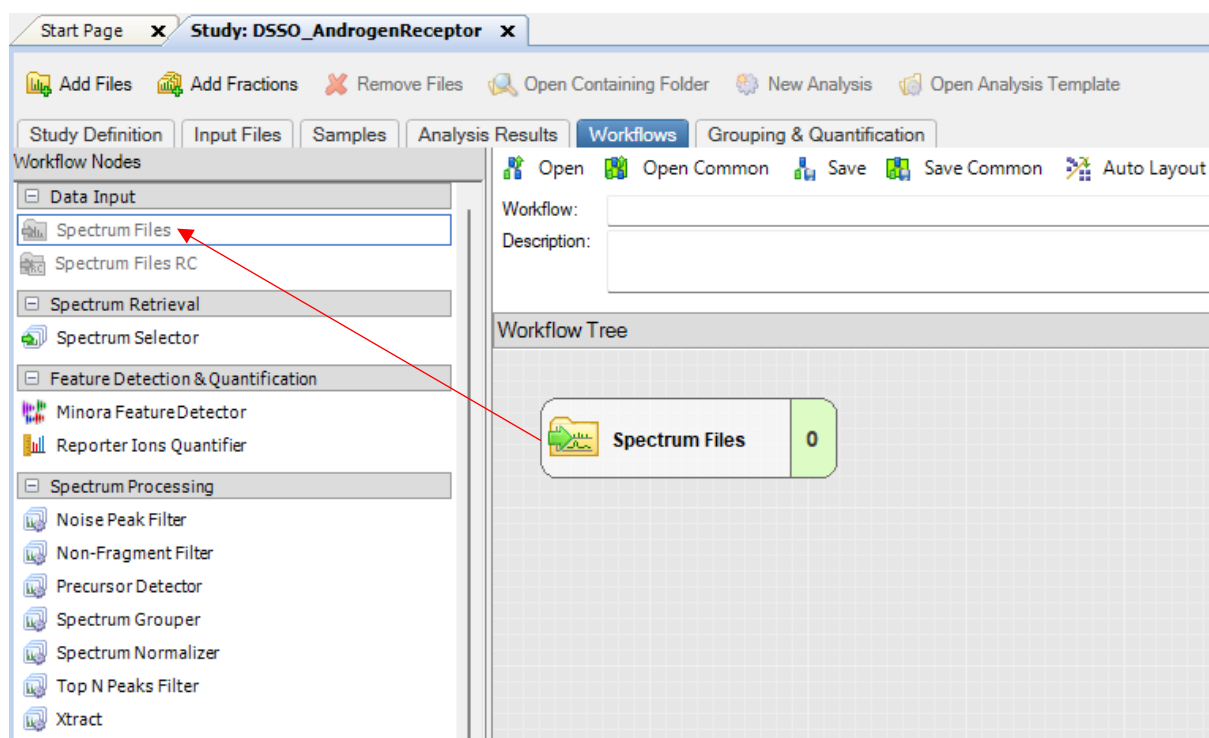
A new window will appear on the right side with two sub-windows called "Consensus Step" and "Processing Step". We will look at the "Processing Step" first, for that click on "Edit" at the top right of the "Processing Step" box.



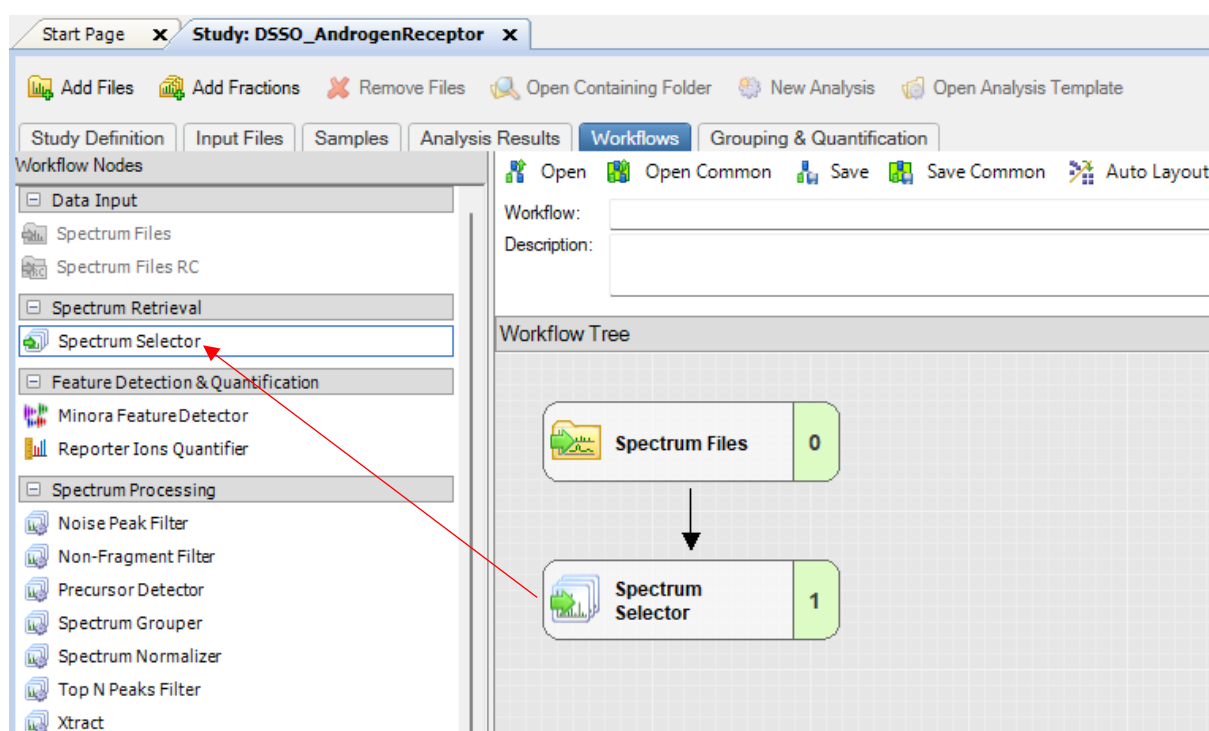
The “Workflow Tree” window should open where we are now able to create a workflow for the identification of crosslinks from mass spectra with MS Annika.



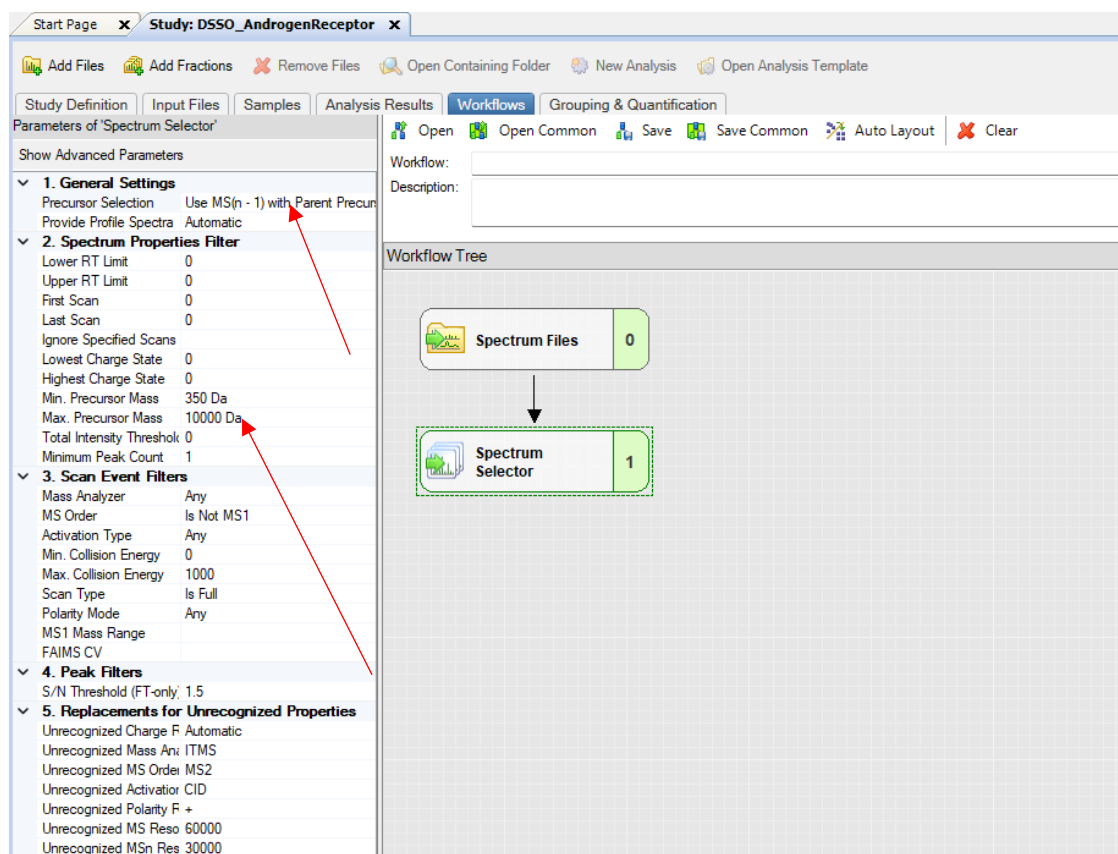
First, we are going to drag-and-drop the “Spectrum Files” node into the “Workflow Tree” window.



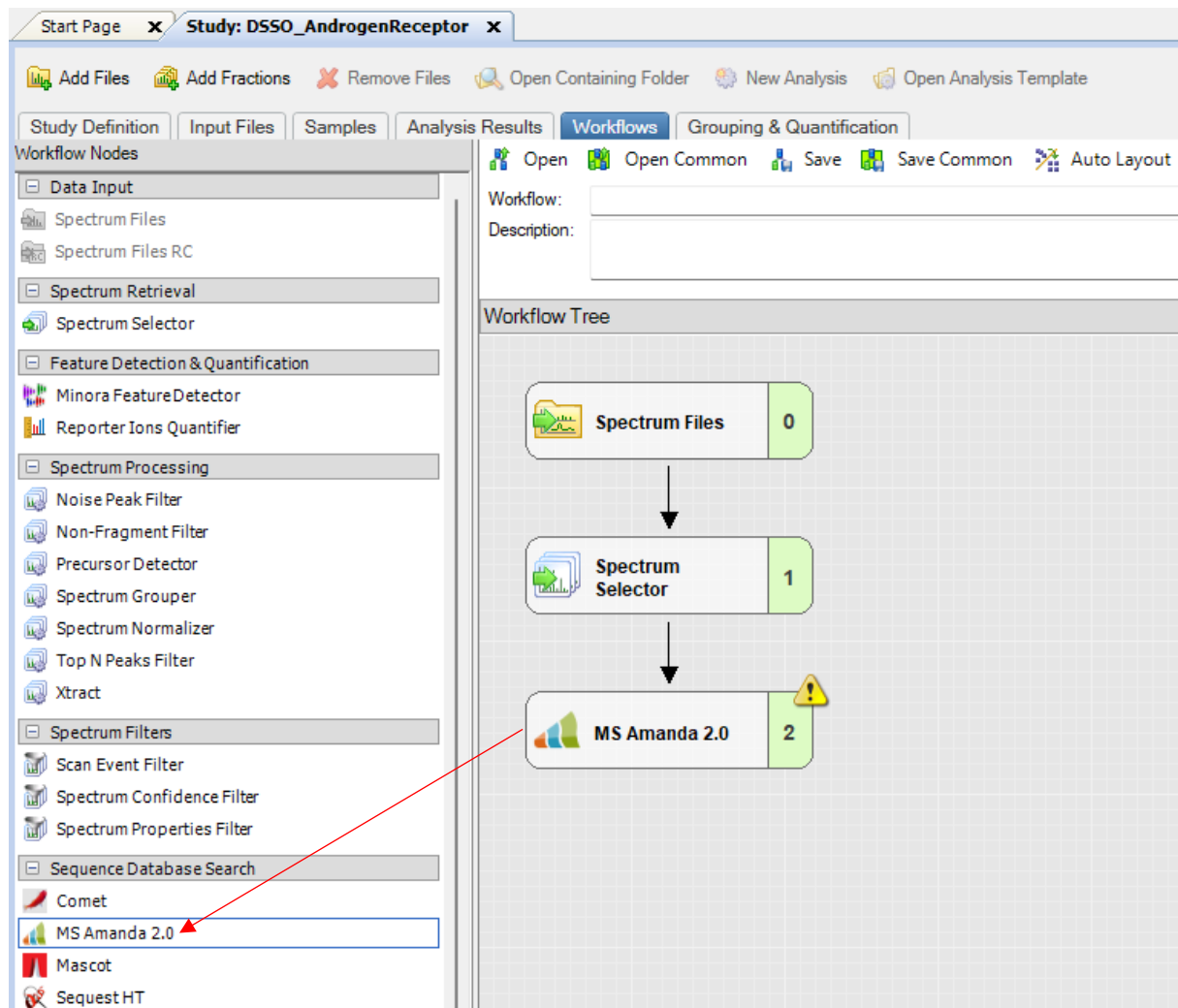
Secondly, drag-and-drop the "Spectrum Selector" node into the workflow window.



Then left click on the "Spectrum Selector" node to bring up the parameters of the node. Change "Precursor Selection" to "Use MS(n - 1) with Parent Precursors" and "Max. Precursor Mass" to "10000 Da".



The next node we need is the “MS Amanda” node: drag-and-drop it from the “Workflow Nodes” window into the workflow pane. Connect the “Spectrum Selector” node to the “MS Amanda 2.0” node by hovering over the “Spectrum Selector” node and dragging one of the appearing white rectangles onto the “MS Amanda 2.0” node.



Left click the “MS Amanda 2.0” node to bring up its parameters. Click on “Show Advanced Parameters” in the top left corner of the parameter window to display all available parameters. First you would want to set the “Protein Database” which should let you select your previously imported FASTA file from a drop-down menu. Usually, we also specify a database of common contaminants here, for example as can be found here: <https://www.thegpm.org/crap/>. For this workflow, the imported FASTA database already contains some contaminants, so we skip this step. Secondly, specify the enzyme used for digestion in the field “Enzyme Name”. Most cross-linking workflows use Trypsin, so the default is fine. The next parameter that usually needs changing is “Missed Cleavages”: in cross-linking workflows the maximum number of missed cleavages considered is usually higher, in our lab we mostly consider 3 or 4 to be the maximum. The parameters “MS1 tolerance” and “MS2 tolerance” depend on the used MS instrument. With most modern MS instruments values of 5 ppm “MS1 tolerance”

and 10 ppm “MS2 tolerance” are good choices, considering MS1 resolutions greater than/equal to 60K and MS2 resolutions greater than/equal to 30K. For older MS instruments (e.g., Q Exactive or older) or lower resolutions values of 10 ppm “MS1 tolerance” and 20 ppm “MS2 tolerance” are more suitable. Last but not least fixed and variable modifications need to be specified, this highly depends on your crosslinker and sample preparation. Every modification that is expected to be found on linear or monolinked peptides should be specified. For DSSO workflows the considered fixed modification is carbamidomethylation of Cysteine and the considered variable modifications are oxidation of Methionine as well as the monolink forms of DSSO.

The screenshot shows the Proteome Discoverer software interface for a study named "Study: DSSO_AndrogenReceptor". The left panel, titled "Parameters of MS Amanda 2.0", is expanded to show the "1. Input Data" section. The parameters are as follows:

Parameter	Value
Protein Database	HiRes.fasta
Enzyme Name	Trypsin (Full)
Missed Cleavages	4
MS1 tolerance	5 ppm
MS2 tolerance	10 ppm

The "2. Static Modifications" section is also expanded, showing:

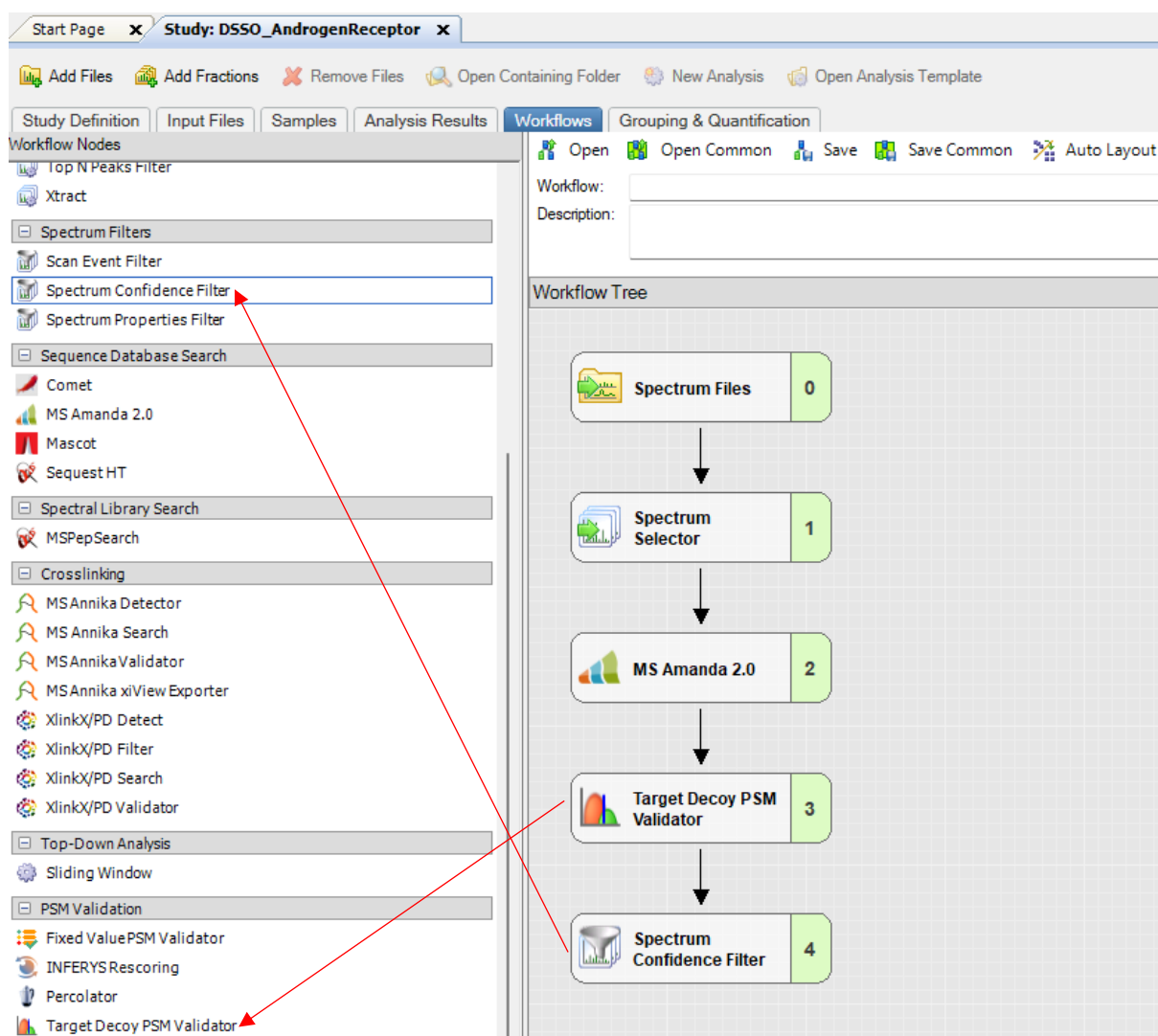
Modification	Value
1. Static Modification	Carbamidomethyl / +57.021 Da (C)
2. Static Modification	None
3. Static Modification	None
Static Peptide N-Terminal Modif	None
Static Peptide C-Terminal Modif	None
Static Protein N-Terminal Modif	None

The "3. Dynamic Modifications" section is also expanded, showing:

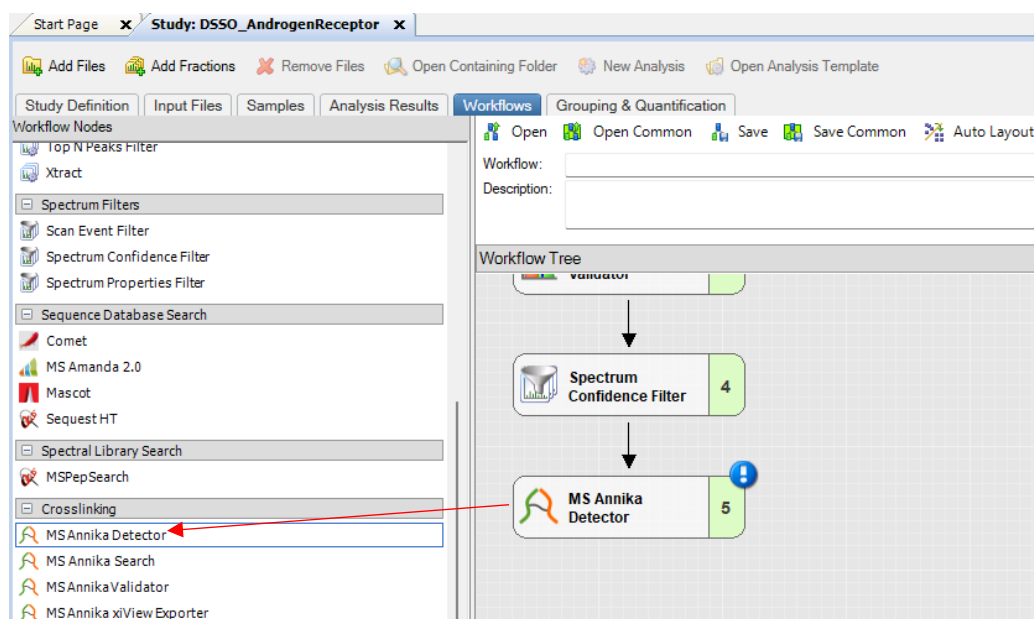
Modification	Value
1. Dynamic Modification	Oxidation / +15.995 Da (M)
2. Dynamic Modification	DSSO Amidated / +175.030 Da (K)
3. Dynamic Modification	DSSO Hydrolyzed / +176.014 Da (K)
4. Dynamic Modification	DSSO Tris / +279.078 Da (K)
5. Dynamic Modification	None
6. Dynamic Modification	None
7. Dynamic Modification	None
8. Dynamic Modification	None
9. Dynamic Modification	None
10. Dynamic Modification	None
11. Dynamic Modification	None
12. Dynamic Modification	None
13. Dynamic Modification	None
14. Dynamic Modification	None
15. Dynamic Modification	None
16. Dynamic Modification	None
17. Dynamic Modification	None
18. Dynamic Modification	None
19. Dynamic Modification	None

The right panel, titled "Workflow Tree", shows a sequence of nodes: "Spectrum Files" (0), "Spectrum Selector" (1), and "MS Amanda 2.0" (2). The "MS Amanda 2.0" node is highlighted with a dashed green border and a yellow warning icon. Red arrows point from the parameters in the left panel to the corresponding nodes in the workflow tree: "HiRes.fasta" to "Spectrum Files", "Trypsin (Full)" to "Spectrum Selector", "Missed Cleavages" to "Spectrum Selector", "MS1 tolerance" to "MS Amanda 2.0", "MS2 tolerance" to "MS Amanda 2.0", "Carbamidomethyl / +57.021 Da (C)" to "MS Amanda 2.0", "Oxidation / +15.995 Da (M)" to "MS Amanda 2.0", "DSSO Amidated / +175.030 Da (K)" to "MS Amanda 2.0", "DSSO Hydrolyzed / +176.014 Da (K)" to "MS Amanda 2.0", and "DSSO Tris / +279.078 Da (K)" to "MS Amanda 2.0".

After the identification of linear and monolinked peptides the resulting peptide spectrum matches (PSMs) need to be validated, for that the “Target Decoy PSM Validator” node is added to the workflow and connected to the “MS Amanda 2.0” node. Any spectrum that has a high-confidence PSM should be filtered out and not be considered for crosslink search and we can achieve that by adding the “Spectrum Confidence Filter” node to the workflow and connecting it to the “Target Decoy PSM Validator” node. The default parameters for both nodes are usually fine, there is no need to change them.



After that the components for the crosslink search are added, first and foremost the “MS Annika Detector” node which you will need to connect to the “Spectrum Confidence Filter” node.

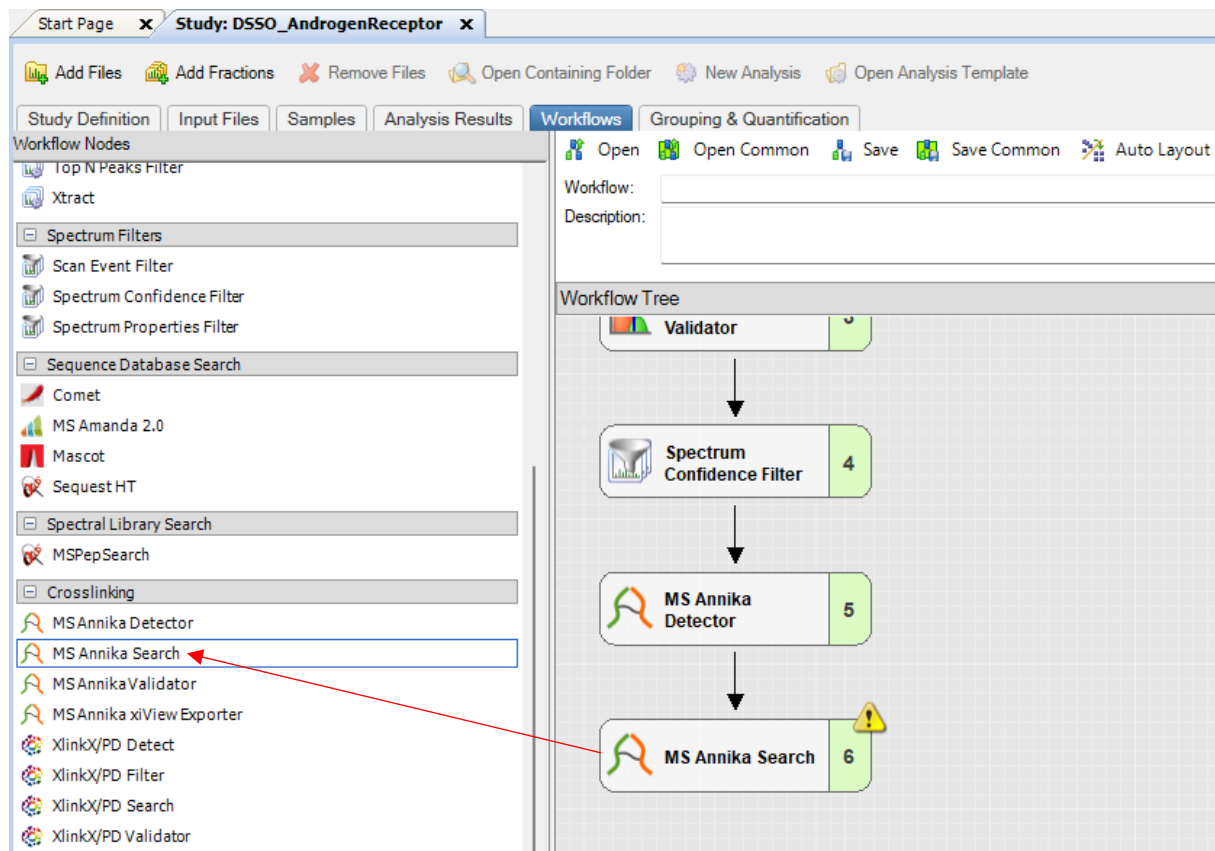


Left click on the "MS Annika Detector" node to access its parameters. The first parameter "MS2 tolerance" should be set to the value also chosen in the "MS Amanda 2.0" node, for this workflow we will use 10 ppm again. Optionally, "Minimum Charge" can be set to 0 which means that any ions with no assigned charge state will be treated as singly charged ions. Next the crosslinker needs to be specified: from the drop-down menu of the parameter "Crosslink Modification" select the crosslinker and target amino acid. If the crosslinker of your sample is not in the list you need to first define it in Proteome Discoverer, this will be covered in a later chapter of the tutorial. If the crosslinker cannot react with the protein's N-terminus make sure to set "N-Terminal Crosslink Modification" to "False". In case your crosslinker yields more than two possible fragments after cleavage you will also need to set the parameter "Additional Crosslink Doublet Distances". This only applies to a small number of crosslinkers, for example sulfoxy-based linkers like DSSO and DSBSO. DSSO primarily yields an alkene and a thiol fragment upon cleavage that exert a mass difference of ~32 Da, however there is also a chance that the crosslinker forms an alkene and a sulfenic acid fragment upon cleavage which exert a different delta mass of ~50 Da. This second case needs to be specified in the "Additional Crosslink Doublet Distances" parameter. If you are unsure how the crosslinker in your sample fragments, please consult the crosslinker vendor or the people synthesizing the crosslinker. It is also possible to use MS Annika for crosslinkers that yield more than three different fragments, please consult the section about defining your own crosslinker in this tutorial. The next important parameter is "Diagnostic Ions" where any masses of diagnostic ions should be set (if any are observed). For DSSO the default selection is fine. Optionally, you can increase the number of "top N most intense doublets" which controls how many of the doublets with highest intensity should be considered for search in cases where only doublets for one of the two peptides are found. In our lab we usually set this value to 3.

The screenshot displays the Proteome Discoverer software interface. On the left, the 'Parameters of MS Annika Detector' panel is open, showing three main sections: 1. Settings, 2. Crosslinker Settings, and 3. Doublet Selection. Red arrows point to specific values: 'MS2 tolerance' is 10 ppm, 'Minimum Charge' is 0, 'Crosslink Modification' is DSSO / +158.004 Da (K), 'Additional Crosslink Doublet Distances' is 49.982635, and 'top N most intense doublets' is 3. On the right, the 'Workflow Tree' shows a sequence of steps: 'validator' (step 3), 'Spectrum Confidence Filter' (step 4), and 'MS Annika Detector' (step 5). The 'MS Annika Detector' node is highlighted with a dashed blue border.

Section	Parameter	Value
1. Settings	MS2 tolerance	10 ppm
	Minimum Charge	0
	Use monoisotopic mass	True
	Precursor mass offset	5
	Use theoretic MS1 peaks	True
2. Crosslinker Settings	Crosslink Modification	DSSO / +158.004 Da (K)
	N-Terminal Crosslink Modification	True
	Additional Crosslink Doublet Distances	49.982635
	Diagnostic Ions	138.0911; 155.1179; 170.0634; 187.090
3. Doublet Selection	Doublet Pair Selection	combined mode
	top N most intense doublets	3
	Persist Doublets	False

Next you need to add the “MS Annika Search” node to the workflow. It should automatically connect to the “MS Annika Detector” node.



Left click on the “MS Annika Search” node to set its parameters. The “MS Annika Search” node is very similarly structured to the “MS Amanda 2.0” node. The first parameter to set is the “Protein Database”. This should be the imported FASTA file that contains the sequences of all cross-linked proteins in your sample. Additionally, one should also consider adding a database of common contaminants again. The digestion enzyme needs to be set under “Enzyme Name” which for most cross-linking workflows can be left at the default of Trypsin. “MS1 tolerance” and “MS2 tolerance” are again the same as in the “MS Amanda 2.0” and the “MS Annika Detector” node. The same goes for the maximum number of “Missed Cleavages”. The parameter “Search Approach” can be left at the default of “MS2” for any kind of MS2 workflows (MS2/MS2-MS2/stepped HCD MS2, CID/ETD/HCD). If MS2-MS3 acquisition was used the “Search Approach” should be set to “MS2/MS3”. The next parameter that usually needs to be adjusted is “Multiplicative Penalty for Crosslinker with Equal Sequences” which usually should be set to 1 which means that crosslinks where the two peptides have the same sequence are not penalized. Finally set the fixed and variable modifications again, for standard cross-linking workflows we consider carbamidomethylation of Cysteine as a fixed modification and oxidation of Methionine as a variable modification.

Parameters of 'MS Annika Search'

Hide Advanced Parameters

- 1. Settings**
 - Protein Database: HiRes.fasta
 - Enzyme Name: Trypsin (Full)
 - MS1 tolerance: 5 ppm
 - MS2 tolerance: 10 ppm
 - Missed Cleavages: 4
 - Search Approach: MS2
- 2. MS3 Search Settings**
 - Precursor Selection: Use MS1 Precursor
 - MS3 tolerance: 20 ppm
 - Create MS2 Spectra: True
 - Boost: 20
 - Isotopic Peak Tolerance: 0.01
 - Maximum Intensity Difference: 0.3
 - Maximum Considered Precursor: 4
- 3. Crosslink Search Settings**
 - Multiplicative Penalty for Crossli: 1
 - Perform decoy search: True
- 4. Static Modifications**
 - 1. Static Modification: Carbamidomethyl / +57.021 Da (C)
 - 2. Static Modification: None
 - 3. Static Modification: None
 - Static Peptide N-Terminal Modif: None
 - Static Peptide C-Terminal Modif: None
 - Static Protein N-Terminal Modif: None
 - Semi-Static Peptide N-Terminal: None
- 5. Dynamic Modifications**
 - 1. Dynamic Modification: Oxidation / +15.995 Da (M)
 - 2. Dynamic Modification: None
 - 3. Dynamic Modification: None

Workflow Tree

```

graph TD
    Validator[Validator] --> Spectrum[Spectrum Confidence Filter]
    Spectrum --> Detector[MS Annika Detector]
    Detector --> Search[MS Annika Search]
  
```

[Settings for MS2-MS3 workflows] If your cross-linking workflow used MS2-MS3 acquisition, some additional settings need to be considered: the parameter "Precursor Selection" should be set to the same value that was used in the "Spectrum Selector" node. It is recommended to use "Use MS(n - 1) with Parent Precursors" in both nodes. Secondly, the MS3 tolerance (fragment ion tolerance for MS3 spectra) should be adjusted to an appropriate value. For MS3 spectra recorded in the orbitrap this should be set to 10-20 ppm, for MS3 spectra recorded in the ion trap a higher tolerance like 0.5 Da is recommended.

Parameters of 'MS Annika Search'

Hide Advanced Parameters

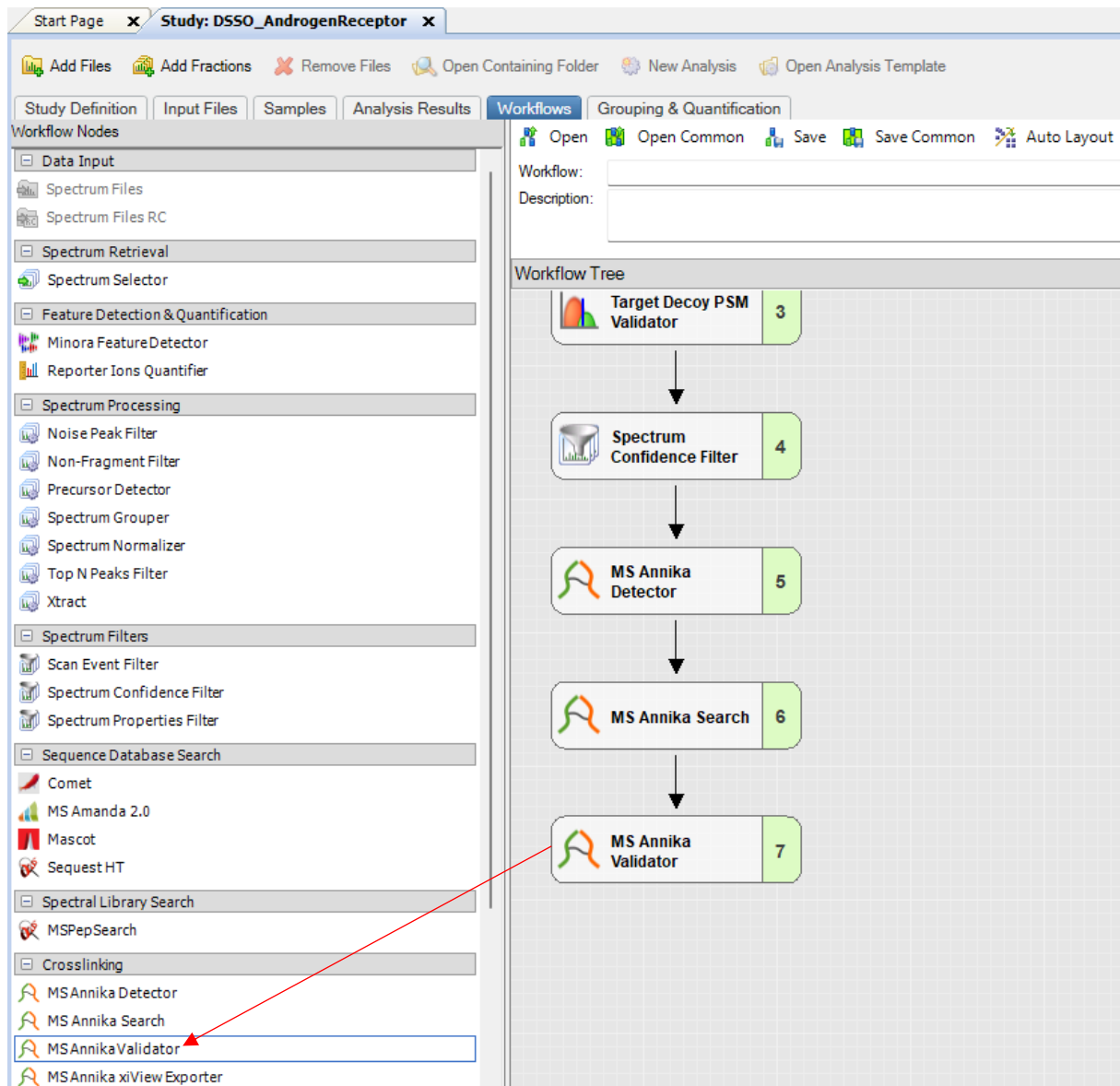
- 1. Settings**
 - Protein Database: HiRes.fasta
 - Enzyme Name: Trypsin (Full)
 - MS1 tolerance: 5 ppm
 - MS2 tolerance: 10 ppm
 - Missed Cleavages: 4
 - Search Approach: MS2/MS3
- 2. MS3 Search Settings**
 - Precursor Selection: Use MS(n - 1) with Parent Precursors
 - MS3 tolerance: 10 ppm
 - Create MS2 Spectra: True
 - Boost: 20
 - Isotopic Peak Tolerance: 0.01
 - Maximum Intensity Difference: 0.3
 - Maximum Considered Precursor: 4
- 3. Crosslink Search Settings**
 - Multiplicative Penalty for Crossli: 1
 - Perform decoy search: True
- 4. Static Modifications**
 - 1. Static Modification: Carbamidomethyl / +57.021 Da (C)
 - 2. Static Modification: None
 - 3. Static Modification: None
 - Static Peptide N-Terminal Modif: None
 - Static Peptide C-Terminal Modif: None
 - Static Protein N-Terminal Modif: None
 - Semi-Static Peptide N-Terminal: None

Workflow Tree

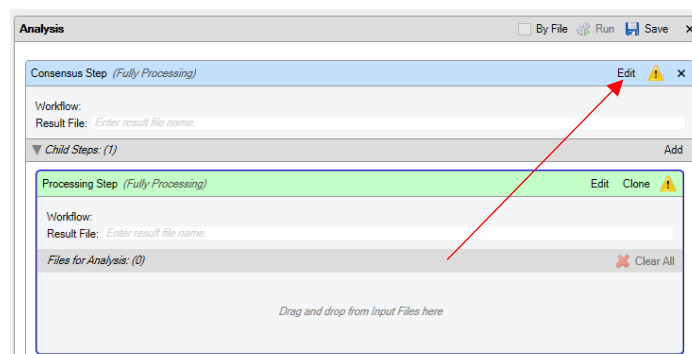
```

graph TD
    Validator[Validator] --> Spectrum[Spectrum Confidence Filter]
    Spectrum --> Detector[MS Annika Detector]
    Detector --> Search[MS Annika Search]
  
```

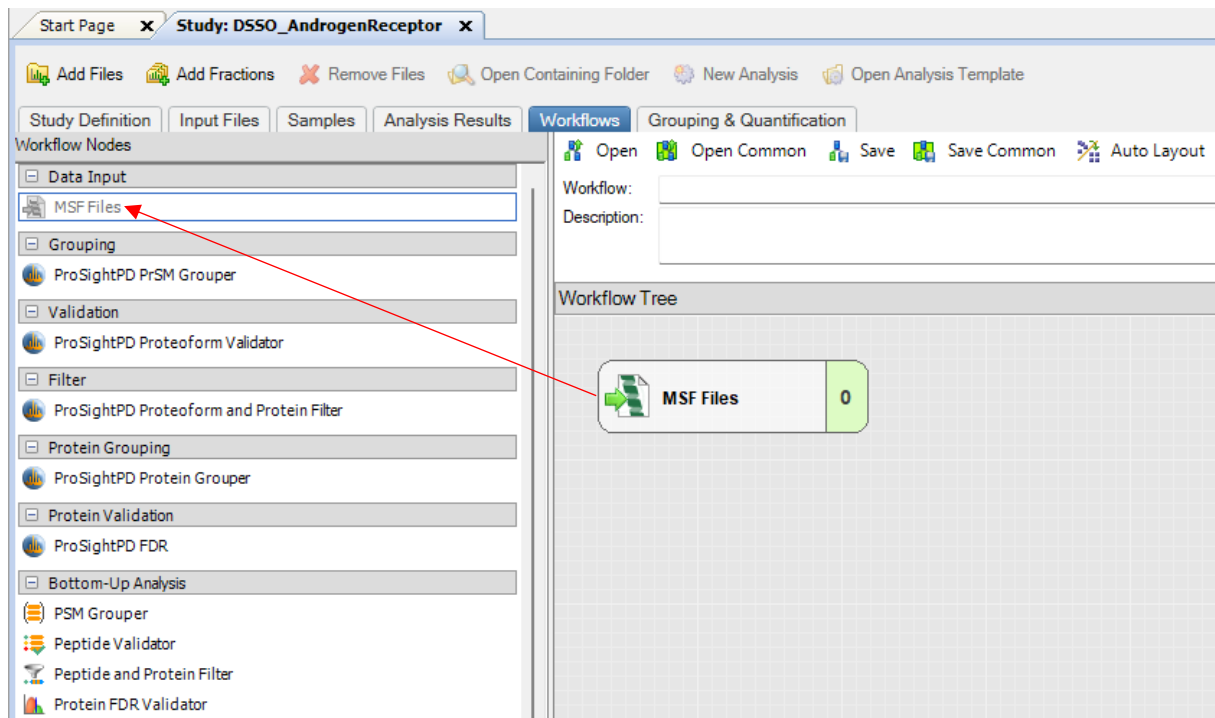
The last step in the processing workflow is to add a validator node for crosslinks and crosslink-spectrum-matches (CSMs). This is the task of the “MS Annika Validator” node. Add it to your workflow pane and it should automatically connect to the “MS Annika Search” node. No parameters need to be changed there.



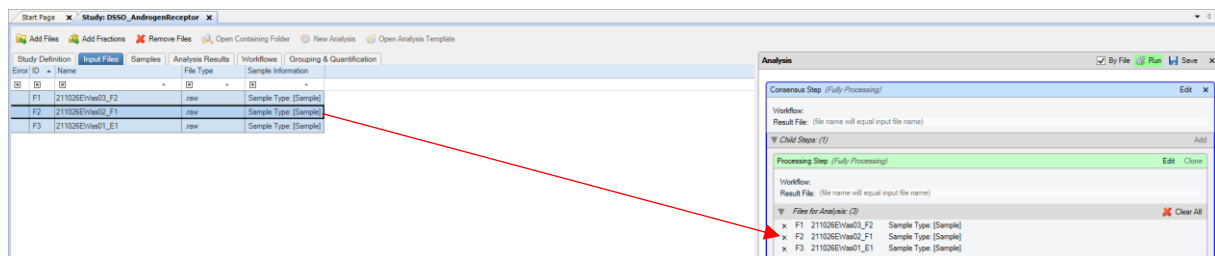
Now open the “Consensus Step” by clicking “Edit” in the top right corner of the blue box.



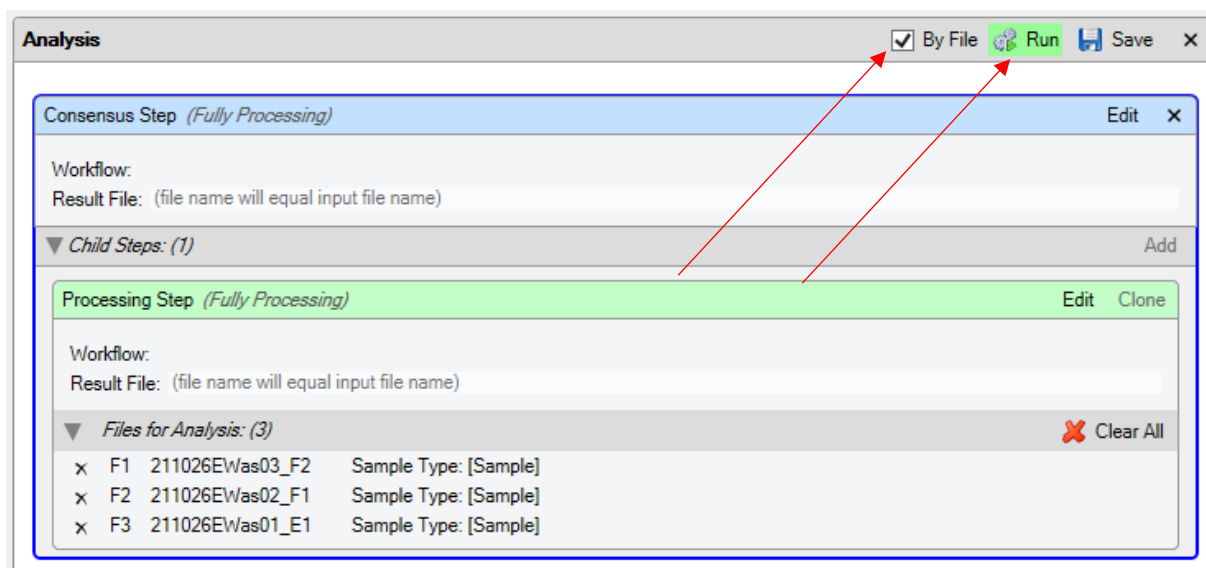
Drag-and-drop the “MSF Files” node into the “Workflow Tree” of the “Consensus Step”. That’s the minimal setup to get MS Annika result files. Adjust the consensus workflow to your need, e.g., by opening a common consensus workflow.



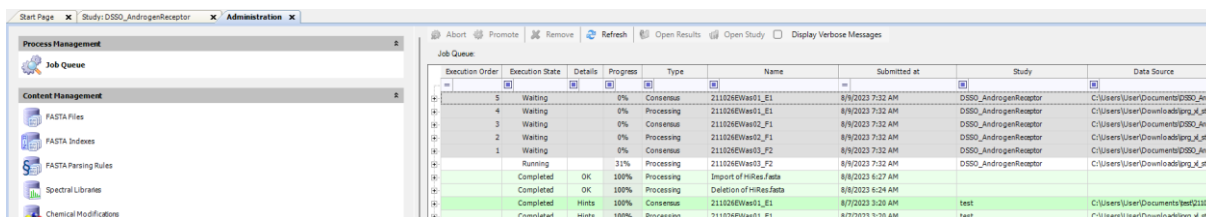
Now all that’s left to do is to add the imported MS files to the workflow. Go to the “Input Files” tab and select all files you want to analyze, then drag-and-drop them into the “Files for Analysis” area of the “Processing Step” (green box).



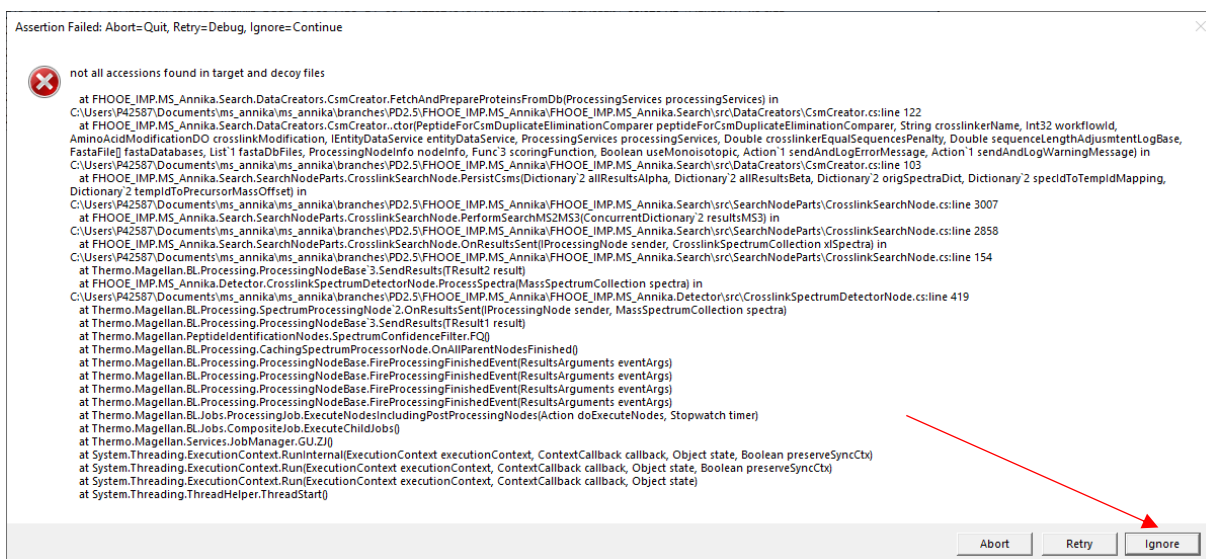
If you are analyzing technical replicates, make sure to tick the “By File” checkbox in the top right corner of the “Analysis” window. If your files are not replicates but e.g., different fractions of your chromatography you should not check this box! For this tutorial our files are technical triplicates, so we tick the “By File” checkbox. After that press “Run” in the top right corner of the “Analysis” window.



The analysis should be started and you should be redirected to the "Job Queue" view of the "Administration" window. The job queue should display the current process of the analysis. If everything works correctly the job will be highlighted in green and get the status "Completed" after it is finished.



In some rare cases (usually when the FASTA headers are not in a standard format) the job might be interrupted by the following assertion warning. If it pops up just press "Ignore" and the job will complete successfully. This issue only happens in Proteome Discoverer version <= 3.0.



Loading a template MS Annika analysis workflow

Analysis templates for MS Annika can be found in the MS Annika GitHub repository at <https://github.com/hgb-bin-proteomics/MSAnnika>.

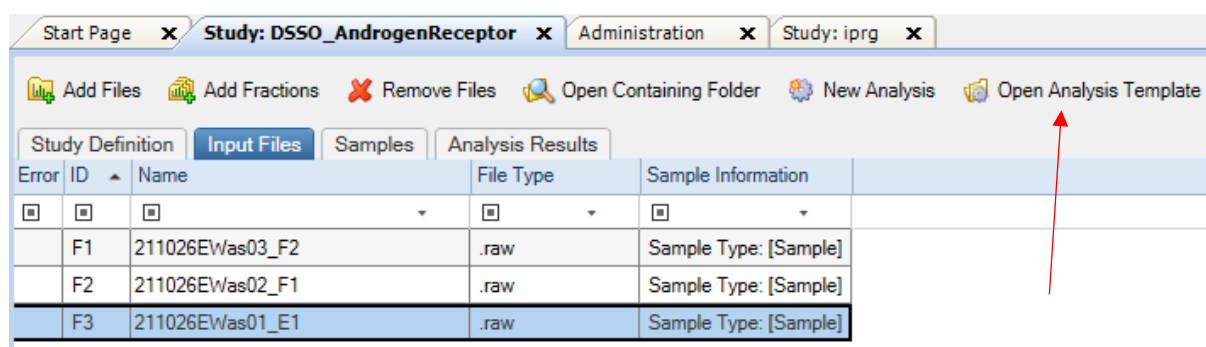
Example Workflows

Example workflows that can be used in Proteome Discoverer*:

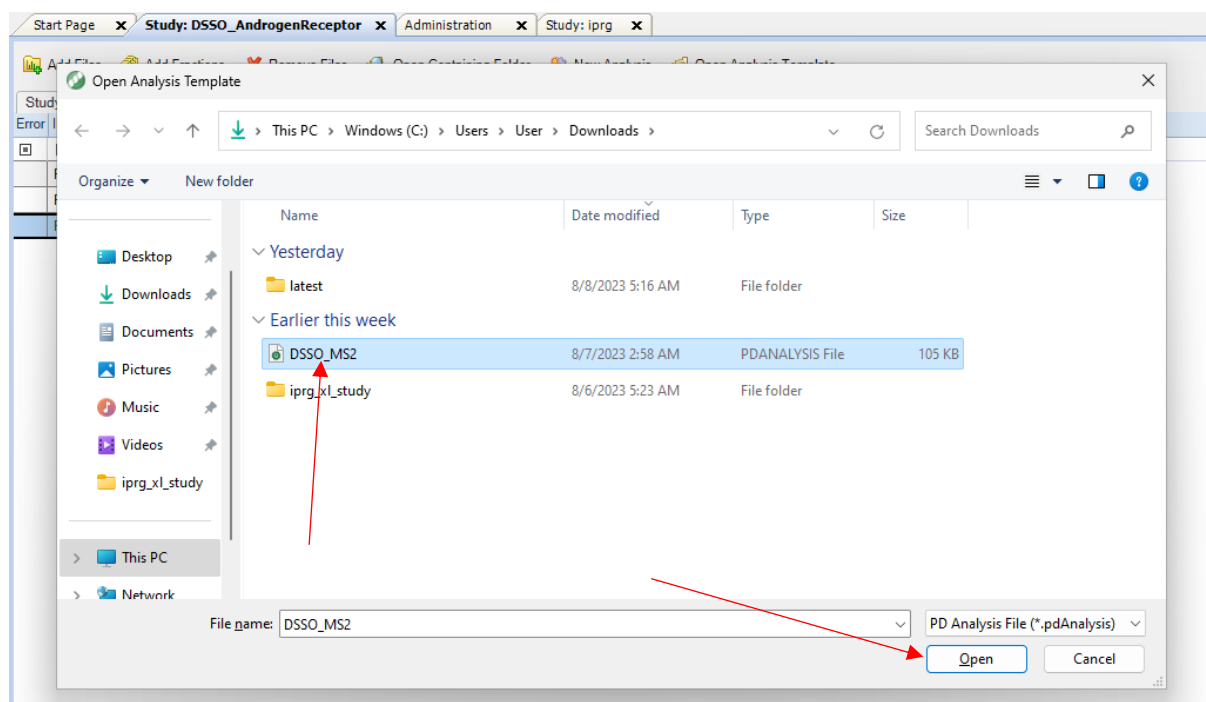
- DSSO MS2 search (CID, ETD, HCD, stepped HCD) ←
- DSSO MS2-MS3 search (MS3 recorded in the orbitrap)
- DSSO MS2-MS3 search (MS3 recorded in the ion trap)

*The provided workflows also require the installation of [MS Amanda](#) which can be downloaded [here](#).

They can be optionally used ***instead*** of creating your own workflow. Download the workflow of your choice to your computer, then select "Open Analysis Template" in Proteome Discoverer. **This will close any currently open analysis!**



Selected the analysis template that you want to load and then click "Open".



The "Analysis" window should open. You will still need to set the parameter "Protein Database" in both the "MS Amanda 2.0" and the "MS Annika Search" node. You should also check if all other parameters are suitable for you!

Lastly you need to add your MS files and start the analysis, exactly as described in the previous section.

Inspecting MS Annika results

In the "Analysis Results" tab double click on the result that you want to open.

Start Page

Study: DSSO_AndrogenReceptor

Administration

211026EWas01_E1

Add Files

Add Fractions

Remove Files

Open Containing Folder

New Analysis

Open Analysis Template

Study Definition

Input Files

Samples

Analysis Results

Workflows

Grouping & Quantification

Open Result

Reprocess

Search:

Error	ID	Execution State	Creation Date	File Name	File Type	File Size	Description
	3	Completed	8/9/2023 7:32 AM	211026EWas01_E1.pdResult	.pdResult	116,624 KB	
	2	Completed	8/9/2023 7:32 AM	211026EWas02_F1.pdResult	.pdResult	127,464 KB	
	1	Completed	8/9/2023 7:32 AM	211026EWas03_F2.pdResult	.pdResult	117,776 KB	

A new tab with the file name should open. By default, it will display the "Proteins" tab which shows protein coverage.

Start Page		Study: DSSO_AndrogenReceptor		Administration		211026EWas01_E1						
Proteins		PSMs	MS/MS Spectrum Info	Input Files	Specialized Traces	Study Information	Crosslinks	CSMs	Crosslink Summary			
	Checked	Accession	Description		Coverage [%]	# Peptides	# PSMs	# Unique Peptides	# AAs	MW [kDa]	calc. pI	Sequence Length
1	<input type="checkbox"/>	EWas01	Androgen receptor (530-899)(AR), mouse, NP_038504.1, G		46%	0	3408	0	372	42.6	8.59	372
2	<input type="checkbox"/>	P0A6Y8	Chaperone protein DnaK OS=Escherichia coli (strain K12) C		25%	0	916	0	638	69.1	4.97	638
3	<input type="checkbox"/>	P0AEU7	Chaperone protein Skp OS=Escherichia coli (strain K12) G		22%	0	95	0	161	17.7	9.70	161
4	<input type="checkbox"/>	EWas03	FOXA1 Hepatocyte nuclear factor 3-alpha, human, NP_004		21%	0	466	0	474	49.3	8.82	474
5	<input type="checkbox"/>	ANDR_RAT	Androgen receptor OS=Rattus norvegicus OX=10116 GN=I		17%	0	3357	0	902	98.2	6.71	902
6	<input type="checkbox"/>	P0A746	Peptide methionine sulfoxide reductase MsrB OS=Escherid		15%	0	8	0	137	15.4	5.94	137
7	<input type="checkbox"/>	P09372	Protein GrpE OS=Escherichia coli (strain K12) GN=grpE PE		13%	0	12	0	197	21.8	4.75	197
8	<input type="checkbox"/>	P02413	50S ribosomal protein L15 OS=Escherichia coli (strain K12)		13%	0	32	0	144	15.0	11.18	144
9	<input type="checkbox"/>	P0A951	Spermidine N(1)-acetyltransferase OS=Escherichia coli (str		12%	0	8	0	186	21.9	6.68	186
10	<input type="checkbox"/>	P21645	UDP-3-O-(3-hydroxymyristoyl)glucosamine N-acyltransferas		11%	0	270	0	341	36.0	6.54	341
11	<input type="checkbox"/>	P0A850	Trigger factor OS=Escherichia coli (strain K12) GN=tig PE=		9%	0	348	0	432	48.2	4.88	432
12	<input type="checkbox"/>	P0A7M2	50S ribosomal protein L28 OS=Escherichia coli (strain K12)		9%	0	19	0	78	9.0	11.41	78
13	<input type="checkbox"/>	P0A6T5	GTP cyclohydrolase 1 OS=Escherichia coli (strain K12) GN=		9%	0	37	0	222	24.8	7.33	222
14	<input type="checkbox"/>	P0AG59	30S ribosomal protein S14 OS=Escherichia coli (strain K12)		8%	0	79	0	101	11.6	11.17	101
15	<input type="checkbox"/>	P0A825	Serine hydroxymethyltransferase OS=Escherichia coli (strai		7%	0	48	0	417	45.3	6.48	417
16	<input type="checkbox"/>	P0A836	Succinate-CoA ligase [ADP-forming] subunit beta OS=Escl		6%	0	41	0	388	41.4	5.52	388
17	<input type="checkbox"/>	P0A8M3	Threonine-tRNA ligase OS=Escherichia coli (strain K12) GI		6%	0	81	0	642	74.0	6.19	642
18	<input type="checkbox"/>	P08622	Chaperone protein DnaJ OS=Escherichia coli (strain K12) G		6%	0	65	0	376	41.1	7.84	376
19	<input type="checkbox"/>	P62399	50S ribosomal protein L5 OS=Escherichia coli (strain K12) C		6%	0	39	0	179	20.3	9.48	179

The PSMs tab gives an overview of PSMs of both the linear, monolink and crosslink search.

Start Page	Study: DSSO_AndrogenReceptor	Administration	211026EWas01_E1						
Proteins	PSMs	MS/MS Spectrum Info	Input Files	Specialized Traces	Study Information	Crosslinks	CSMs	Crosslink Summary	
	Checked	Confidence	Identifying Node	PSM Ambiguity	Annotated Sequence	Modifications	# Proteins	Protein Accessions	# Missed Cleavages
1	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	TPLIISGPAEDSSEmYKR	M15(Oxidation)	1	P10408	1
2	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	LGmKPGEAIEHPwVTK	M3(Oxidation)	1	P10408	0
3	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	TPLIISGPAEDSSEmYKR	M15(Oxidation)	1	P10408	1
4	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HDAVLEAGGLHIIGTER		1	P10408	0
5	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HDAVLEAGGLHIIGTER		1	P10408	0
6	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	TPLIISGPAEDSSEmYKR	M15(Oxidation)	1	P10408	1
7	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HLSEQFGW/LQITPQEF ^{Lcm}	C18(Carbamidomethyl); M19	2	EWas01; ANDR_RAT	0
8	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HFDVQLLGmVLNER	M10(Oxidation)	1	P10408	0
9	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HDAVLEAGGLHIIGTER		1	P10408	0
10	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HDAVLEAGGLHIIGTER		1	P10408	0
11	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HFDVQLLGmVLNER	M10(Oxidation)	1	P10408	0
12	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HDAVLEAGGLHIIGTER		1	P10408	0
13	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HSLSFNDcFVKVAR	C8(Carbamidomethyl)	1	EWas03	1
14	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	GGVSLRPGDGVHISwLNR		1	P36683	0
15	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HFDVQLLGmVLNER	M10(Oxidation)	1	P10408	0
16	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	HLSEQFGW/LQITPQEF ^{Lcm}	C18(Carbamidomethyl); M19	2	EWas01; ANDR_RAT	0
17	<input type="checkbox"/>	<div></div>	MS Amanda 2.0 (A2)	Unambiguous	FFDEI ^{Rm} NYIKELDR	M7(Oxidation)	2	EWas01; ANDR_RAT	2

The “Crosslink Summary” gives an overview of how many crosslinks were identified, how many of them fall within the 1% estimated FDR threshold (High Confidence) and how many of them fall within the 5% estimated FDR threshold (Medium Confidence). It also displays these numbers for CSMs. Additionally, the number of crosslink spectra and the number of identified doublets is shown.

Start Page Study: DSSO_AndrogenReceptor Administration 211026EWas01_E1													
	Proteins	PSMs	MS/MS Spectrum Info	Input Files	Specialized Traces	Study Information	Crosslinks	CSMs	Crosslink Summary				
	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
1	<input checked="" type="checkbox"/>	Total	268	108	101	1261	373	201	175	218	155	11246	264316
2	<input checked="" type="checkbox"/>	211026EWas01_E1.raw	268	108	101	1261	373	201	175	218	155	11246	264316

The most important crosslink results are in the “Crosslinks” and “CSMs” tab. The crosslinks tab features a list of all identified crosslinks that can be filtered for high and medium confidence crosslinks.

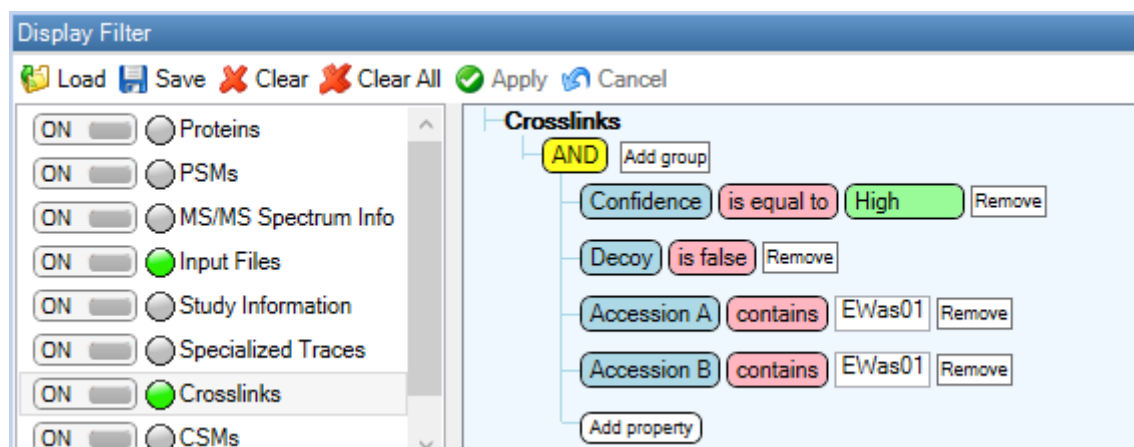
Start Page Study: DSSO_AndrogenReceptor Administration 211026EWas01_E1															
Proteins		PSMs	MS/MS Spectrum Info		Input Files	Specialized Traces	Study Information		Crosslinks	CSMs	Crosslink Summary				
#	Checked	Crosslinker	Crosslink Type	# CSMs	# Proteins	Sequence A	Accession A	Position A	Sequence B	Accession B	Position B	Protein Descriptions A		Protein Descriptions B	Best CSM Score
1	<input checked="" type="checkbox"/>	DSSO	Inter	7	3	MhKQSR	Evla01.ANDR_RAT	3	SPDQPGK	Evla03	4	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	FOXA1 Hepatocyte nuclear factor 3-alpha, human, NP_004	223.68	
2	<input checked="" type="checkbox"/>	DSSO	Inter	1	1	DEENVK	Evla01.ANDR_RAT	-1	FKQCEK	Evla03	2	FOXA1 Hepatocyte nuclear factor 3-alpha, human, NP_004	FOXA1 Hepatocyte nuclear factor 3-alpha, human, NP_004	7.01	
3	<input checked="" type="checkbox"/>	DSSO	Intra	1	2	MhKQSR	Evla01.ANDR_RAT	3	KQNPITSCSR	Evla01.ANDR_RAT	1	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	83.68	
4	<input checked="" type="checkbox"/>	DSSO	Inter	2	2	MhKQSR	Evla01.ANDR_RAT	3	KHGSQSLK	Evla03	-1	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	32.59	
5	<input checked="" type="checkbox"/>	DSSO	Inter	2	3	MhKQSR	Evla01.ANDR_RAT	3	QPGAGGGGGGSGGGGAQKGGPESR	Evla03	18	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	FOXA1 Hepatocyte nuclear factor 3-alpha, human, NP_004	191.47	
6	<input checked="" type="checkbox"/>	DSSO	Inter	2	3	MhKQSR	Evla01.ANDR_RAT	3	AGGGGGGAQNTTKR	Evla03	8	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	FOXA1 Hepatocyte nuclear factor 3-alpha, human, NP_004	141.45	
7	<input checked="" type="checkbox"/>	DSSO	Inter	1	3	KQNPITSCSR	Evla01.ANDR_RAT	1	AGGGGGGAQNTTKR	Evla03	8	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	FOXA1 Hepatocyte nuclear factor 3-alpha, human, NP_004	128.59	
8	<input checked="" type="checkbox"/>	DSSO	Intra	1	2	IACQKR	Evla01.ANDR_RAT	5	KQNPITSCSR	Evla01.ANDR_RAT	1	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	Androgen receptor (530-899)(AR), mouse, NP_038504.1.G	210.94	

A filter can be applied by selecting the cone item at the top of Proteome Discoverer which should open the “Display Filter” window.

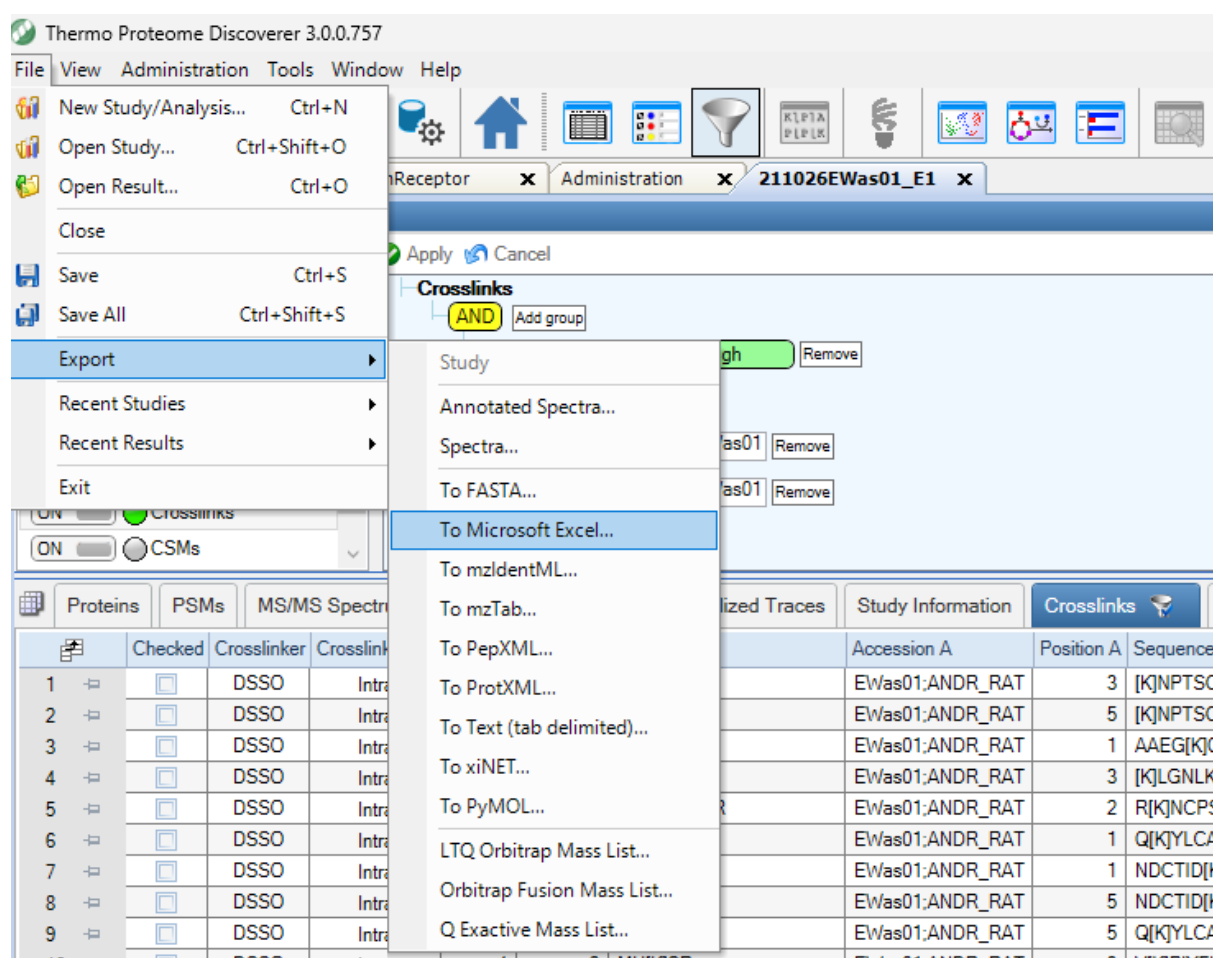
We are going to filter for high confidence crosslinks without decoys by applying “Confidence is equal to High” and “Decoy is false”.

The resulting list of crosslinks contains only high confidence (1% estimated FDR) target-target hits. These can be exported for further analysis.

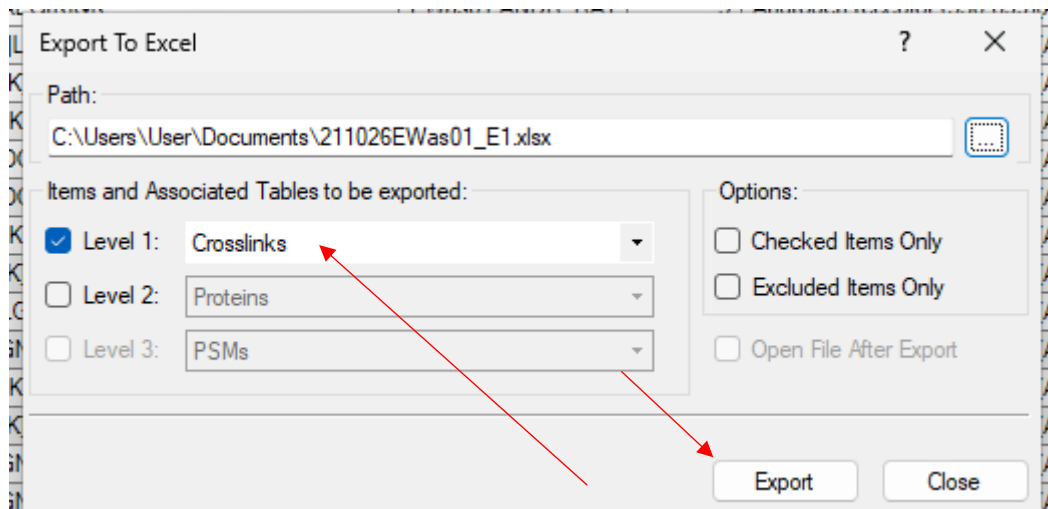
For this tutorial we will also filter out any crosslinks that are not within the mouse androgen receptor because we want to analyze those further later. Applying the filter option “Accession A contains EWas01” and “Accession B contains EWas01” will achieve this.



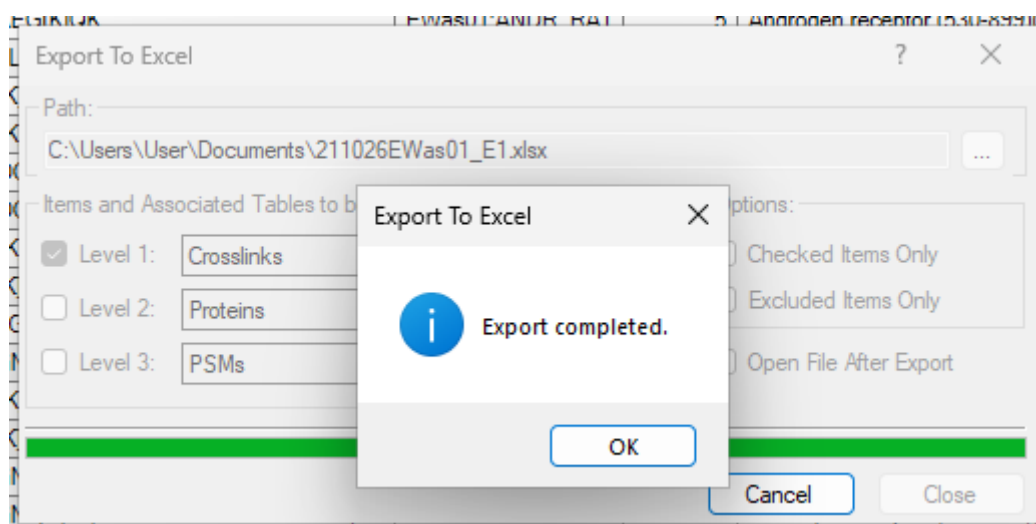
We can export the remaining crosslinks by clicking “File” > “Export” in the top left corner of Proteome Discoverer. We will export to Microsoft Excel format.



Select the path and which table to export. By default the "Proteins" table is pre-selected, make sure to export the "Crosslinks" table and click "Export".



If everything works correctly an "Export completed." should pop up.



Another important table is the "CSMs" table that shows which crosslink is associated to which spectrum.

Start Page	Study: DSSO_AndrogenReceptor	Administration	X	211026Wex01_E1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
------------	------------------------------	----------------	---	----------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Down-stream analysis of MS Annika results

We provide a variety of scripts that aid in the down-stream analysis of MS Annika results. Most of them are based on the exported Microsoft Excel file from the previous section. An overview and download options for all scripts is given in the GitHub repository https://github.com/hgb-bin-proteomics/MSAnnika_exporters.

Currently scripts for the export to the following tools are provided:

- xiNET: <https://crosslinkviewer.org/>
- xiVIEW: https://xiview.org/xiNET_website/index.php
- PyXlinkViewer (pyMOL plugin): <https://github.com/BobSchiffrin/PyXlinkViewer>
- XMAS (ChimeraX plugin): https://github.com/ScheltemaLab/ChimeraX_bundle
- PAE Viewer: <http://www.subtiwiki.uni-goettingen.de/v4/paeViewerDemo>

Quick start

- **Exporting to xiNET**

Files needed:

- result.xlsx - MS Annika result file(s) exported to .xlsx
- seq.fasta - FASTA file containing sequences of the crosslinked proteins

```
python xiNetExporter_msannika.py result.xlsx -fasta seq.fasta
```



- **Exporting to xiVIEW**

Files needed:

- result.xlsx - MS Annika result file(s) exported to .xlsx
- seq.fasta - FASTA file containing sequences of the crosslinked proteins

```
python xiViewExporter_msannika.py result.xlsx -fasta seq.fasta
```



- **Exporting to pyXlinkViewer (pyMOL)**

Files needed:

- result.xlsx - MS Annika result file(s) exported to .xlsx
- structure.pdb - 3D structure of the protein (complex) that crosslinks should be mapped to, alternatively you can also just provide the 4-letter code from the [PDB](#) and the script will fetch the structure from internet

```
python pyXlinkViewerExporter_msannika.py result.xlsx -pdb structure.pdb
```



- **Exporting to XMAS (ChimeraX)**

Visualization of MS Annika results works out of the box with .xlsx files exported from Proteome Discoverer.

- **Exporting to PAE Viewer**

Files needed:

- pyXlinkViewer_export.csv - Crosslinks exported from pyXlinkViewer as .csv

```
python PAEViewerExporter_msannika.py pyXlinkViewer_export.csv
```



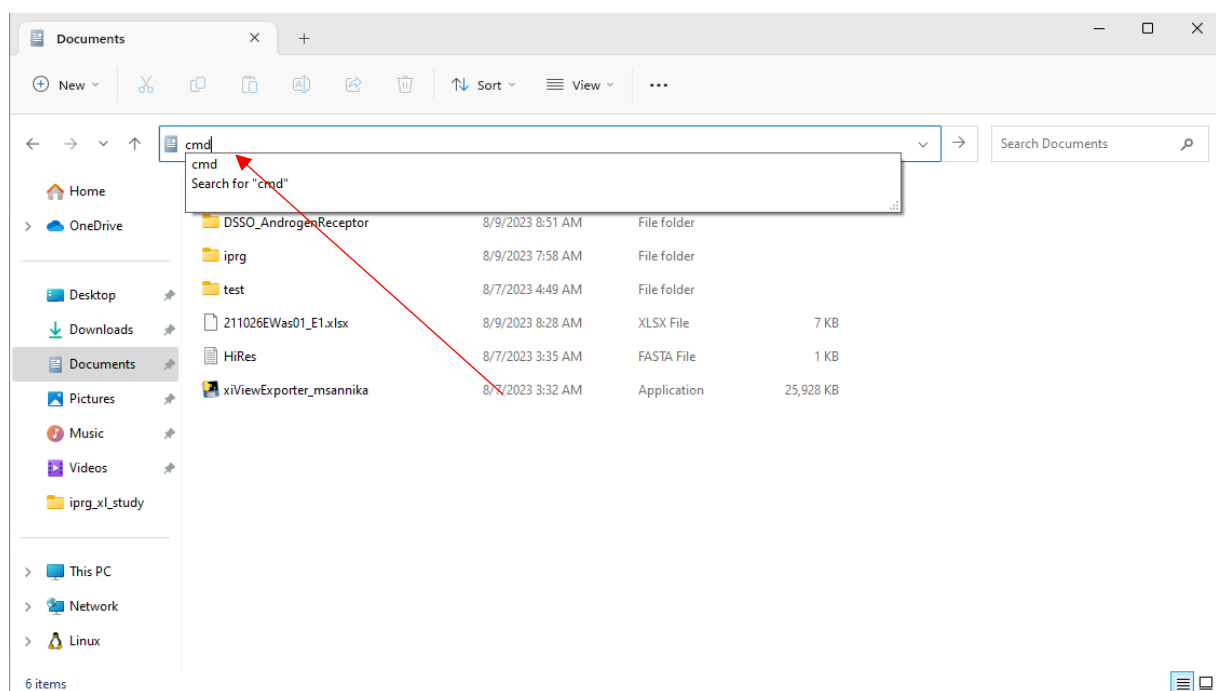
Example: Export to xiVIEW

This section describes how the previously created Microsoft Excel file of crosslinks identified by MS Annika can be used to visualize crosslinks in xiVIEW. The export process is similar for all exporter scripts and should be easily transferable.

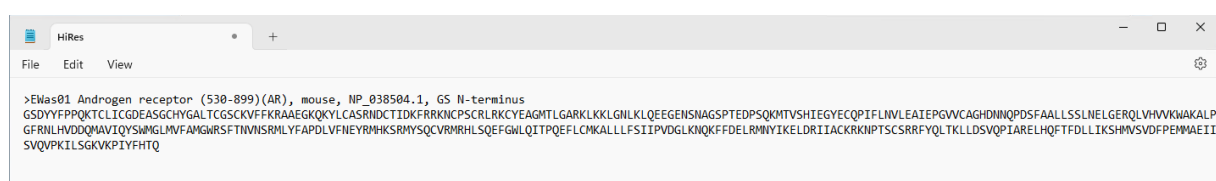
First download the xiVIEW exporter (*xiViewExporter_msannika.exe*) from the GitHub repository https://github.com/hgb-bin-proteomics/MSAnnika_exporters/releases.

Alternatively, you can also download the exporter as python script, the process is the same but obviously knowledge about running python scripts is required. The tutorial will only cover running the executable!

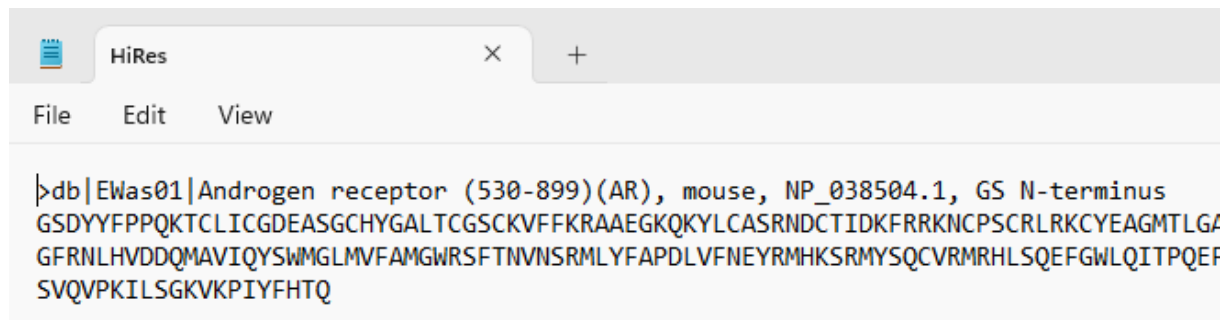
After downloading the exporter, make sure to put the exporter, the used FASTA file and the exported Microsoft Excel file into the same folder. Subsequently open a command line window in that folder. In Windows 10/11 that is for example possible by typing "cmd" into the path.



Now before we run the exporter from the command line window, we need to modify the FASTA file. We only want to analyze crosslinks of the mouse androgen receptor, so we are going to delete all other sequences from the FASTA file. After the other sequences are deleted, the FASTA file should look like this:

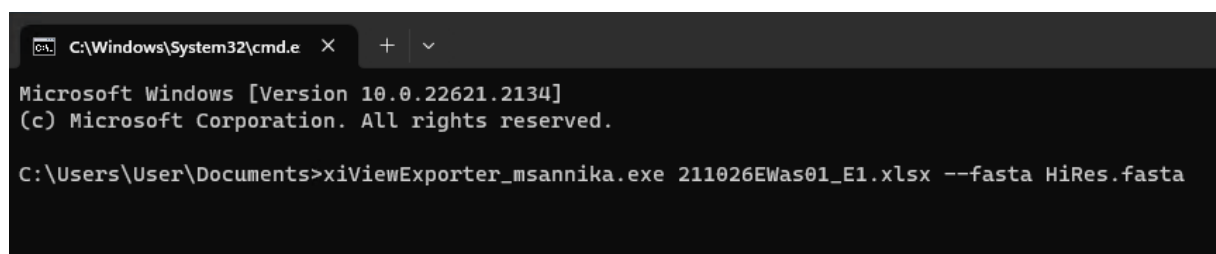


As mentioned in the documentation of the xiVIEW exporter, the FASTA needs to conform to UniProtKB standard formatting, so we still need to change the FASTA header to the following:



```
>db|EWas01|Androgen receptor (530-899)(AR), mouse, NP_038504.1, GS N-terminus
GSDYYFPPQKTCLICGDEASGCHYGALTCGSKVFFKRAAEGKQKYLCA SRNDCTIDKFRRKNCPSCRLRKCYEAGMTLGA
GFRNLHVDDQMAVIQYSWMGLMVFAMGWR SFTNVNSRMLYFAPDLVFNEYRMHKSRMYSQCVRMRHLSQEFGLQITPQEF
SVQVPKILSGKVKPIYFHTQ
```

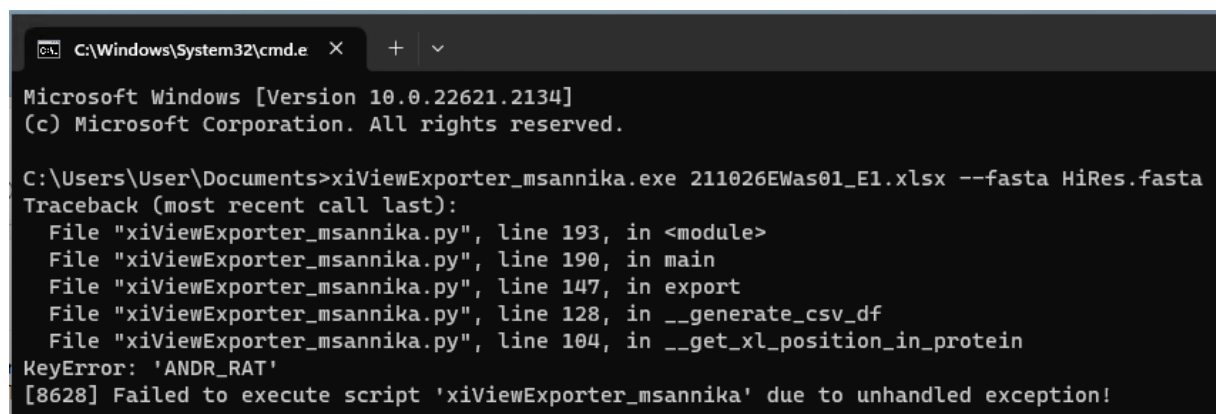
Save the FASTA file and close it. Next you want to open the terminal and type "xiViewExporter_msannika.exe CROSSLINKS_FILE --fasta FASTA_FILE". In our case the crosslinks file is named "211026EWas01_E1.xlsx" and the FASTA file is called "HiRes.fasta" so we will enter the following command:



```
C:\Windows\System32\cmd.e X + v
Microsoft Windows [Version 10.0.22621.2134]
(c) Microsoft Corporation. All rights reserved.

C:\Users\User\Documents>xiViewExporter_msannika.exe 211026EWas01_E1.xlsx --fasta HiRes.fasta
```

After hitting enter you will get the following error:



```
C:\Windows\System32\cmd.e X + v
Microsoft Windows [Version 10.0.22621.2134]
(c) Microsoft Corporation. All rights reserved.

C:\Users\User\Documents>xiViewExporter_msannika.exe 211026EWas01_E1.xlsx --fasta HiRes.fasta
Traceback (most recent call last):
  File "xiViewExporter_msannika.py", line 193, in <module>
  File "xiViewExporter_msannika.py", line 190, in main
  File "xiViewExporter_msannika.py", line 147, in export
  File "xiViewExporter_msannika.py", line 128, in __generate_csv_df
  File "xiViewExporter_msannika.py", line 104, in __get_xl_position_in_protein
KeyError: 'ANDR_RAT'
[8628] Failed to execute script 'xiViewExporter_msannika' due to unhandled exception!
```

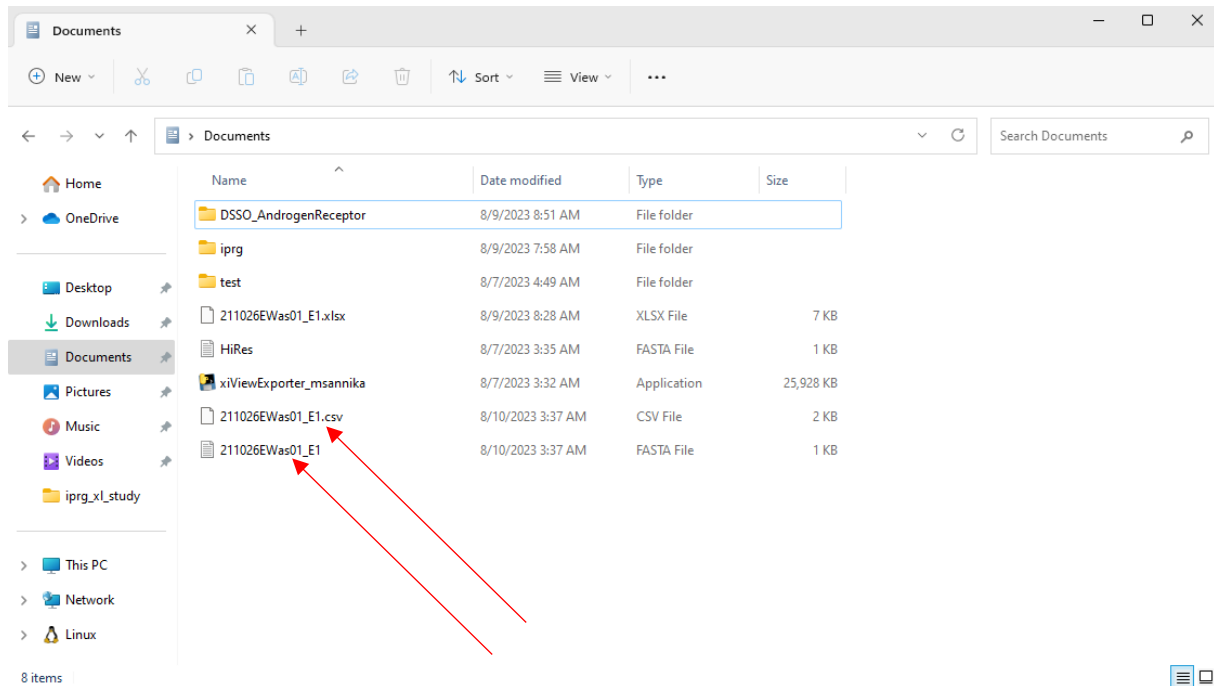
This error appears because some of the cross-linked peptides are both part of the mouse androgen receptor and the rat androgen receptor, however we only supplied a FASTA with the mouse androgen receptor. We need to add the flag "--ignore ANDR_RAT" to the command so the exporter knows it should not look for that sequence. With the following command the exporter now runs through successfully.

```
C:\Windows\System32\cmd.e x + v
Microsoft Windows [Version 10.0.22621.2134]
(c) Microsoft Corporation. All rights reserved.

C:\Users\User\Documents>xiViewExporter_msannika.exe 211026EWas01_E1.xlsx --fasta HiRes.fasta --ignore ANDR_RAT

C:\Users\User\Documents>
```

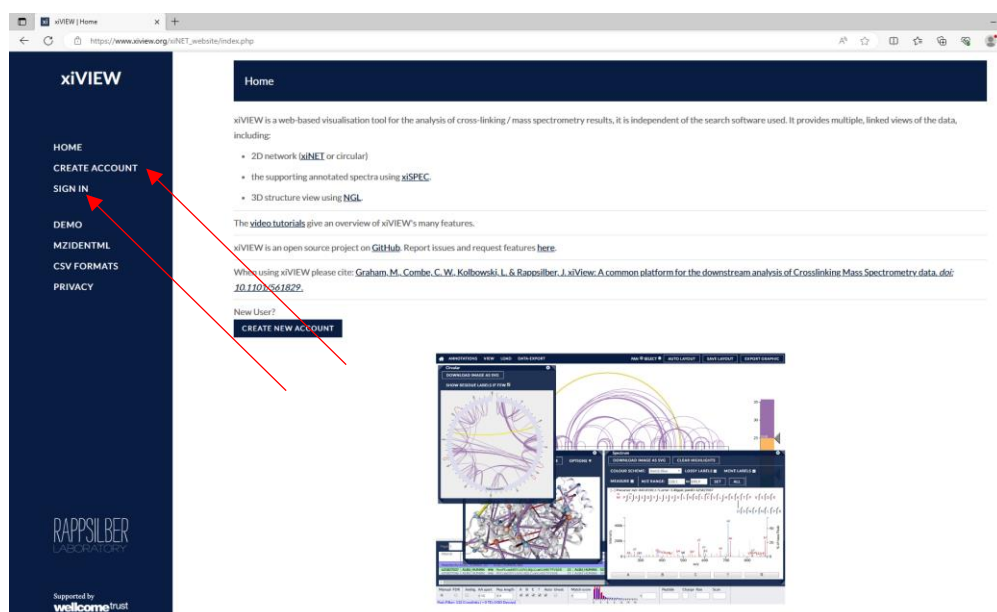
You now should have two files for upload to xiVIEW: "211026EWas01_E1.csv" and "211026EWas01_E1.fasta".



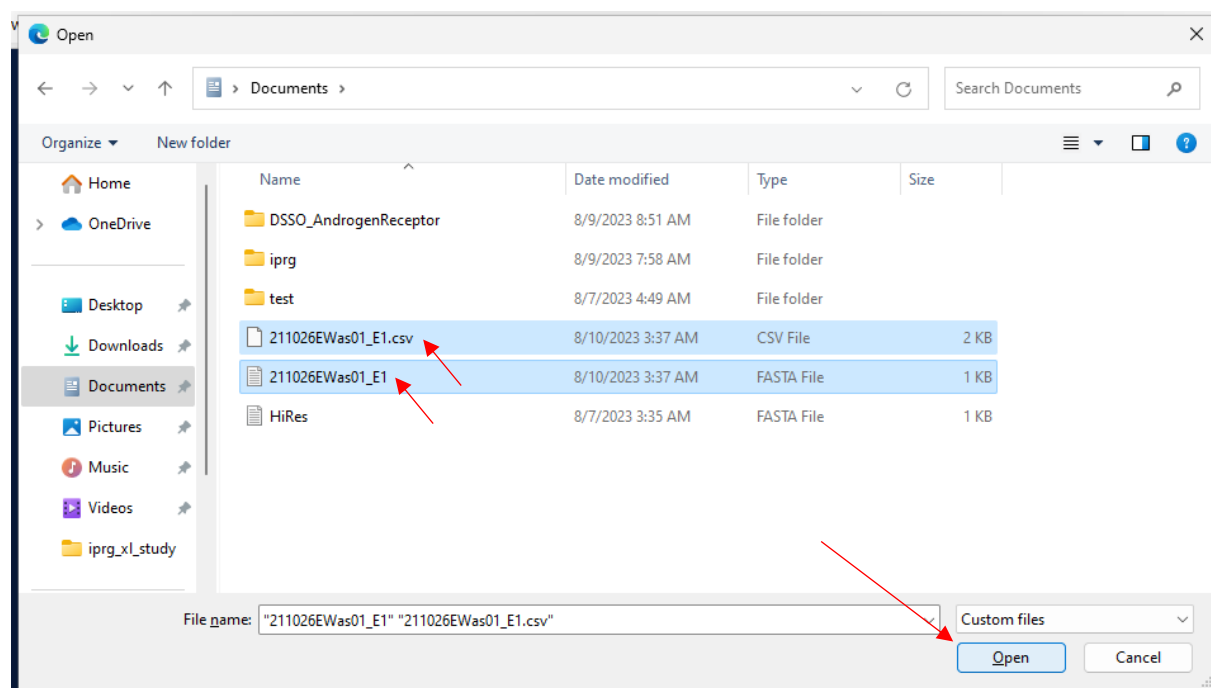
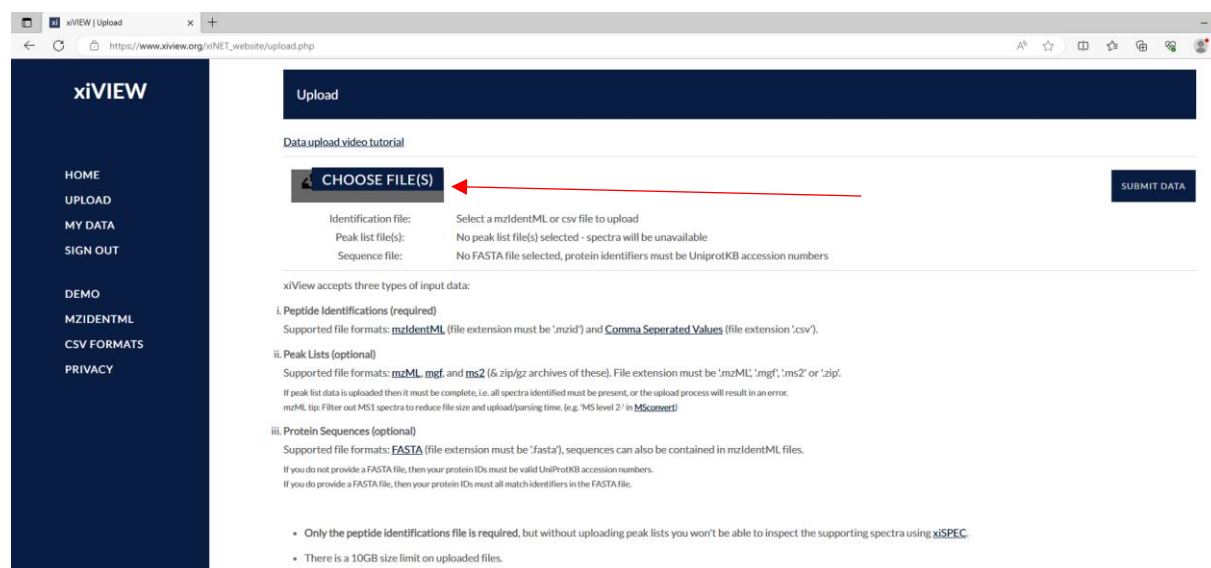
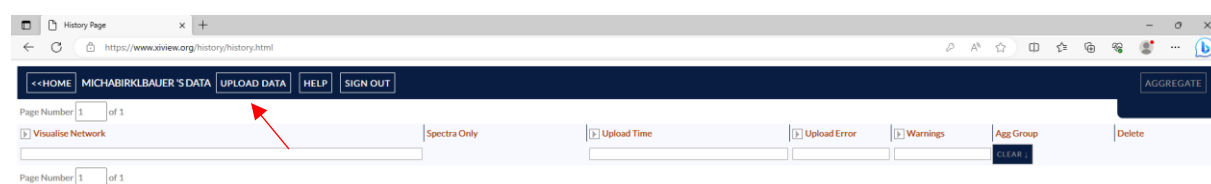
Next open a browser and head over to xiVIEW:

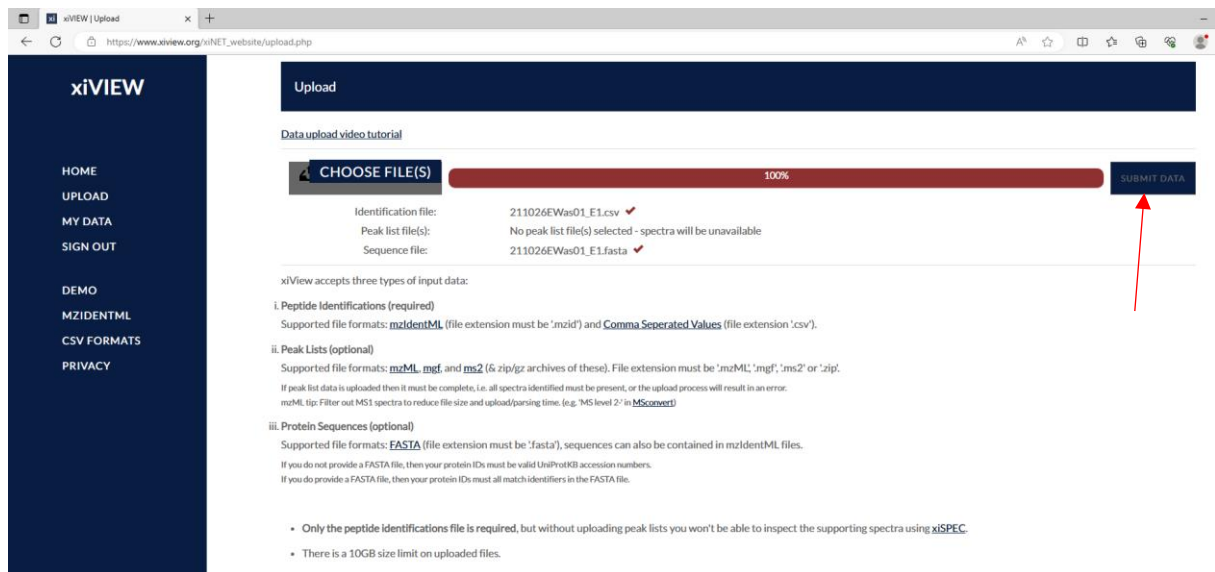
https://www.xiview.org/xiNET_website/index.php

If you don't have an account yet, create an account and login (it's free), if you already have an account -> sign in.

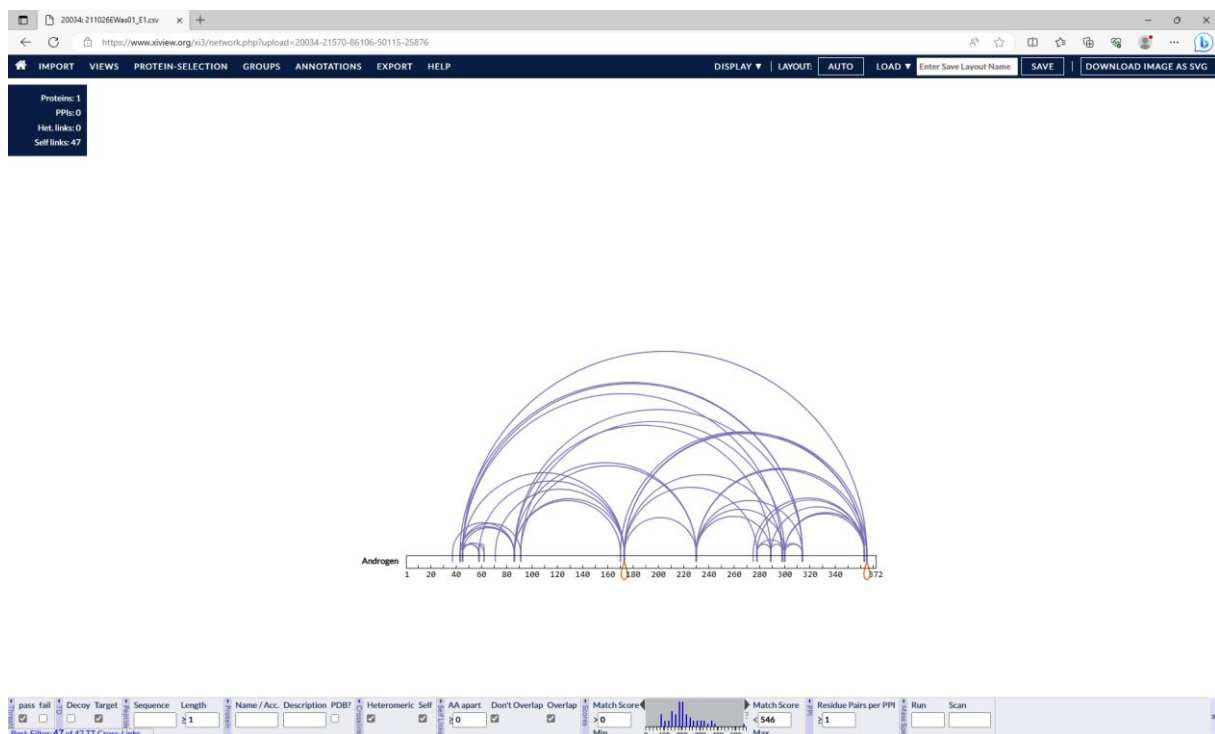


Select "UPLOAD DATA" and upload the two files generated by the exporter.

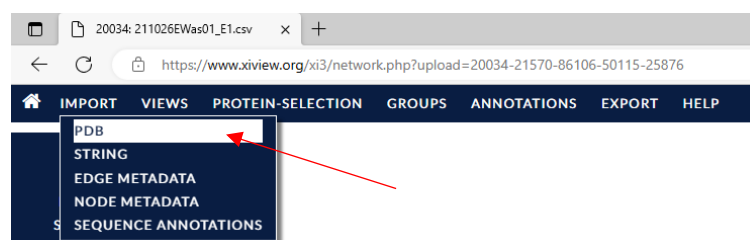




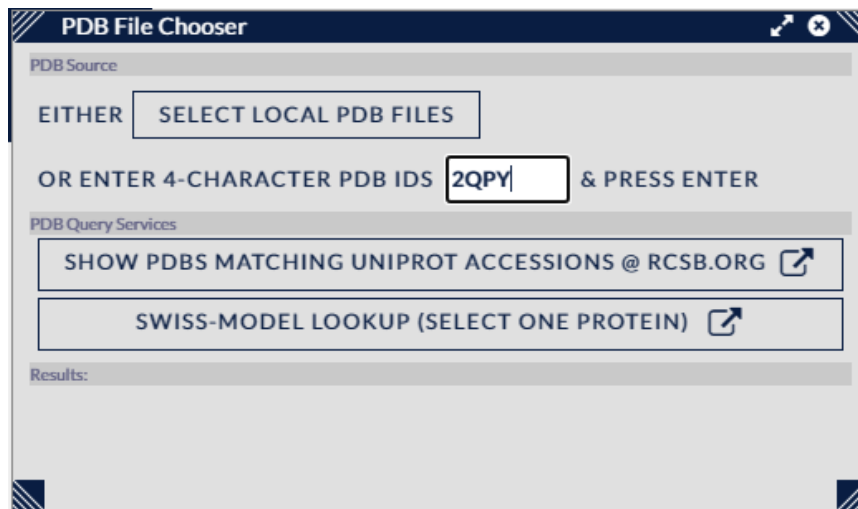
You should be forwarded to the result window that should look like this:



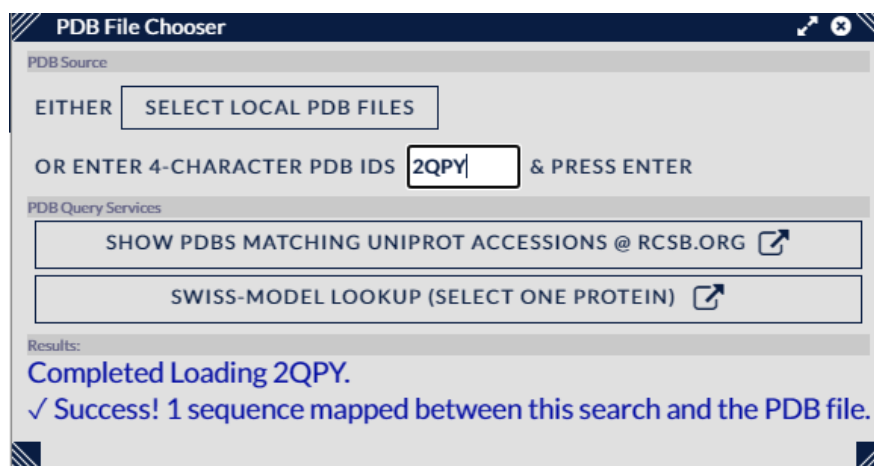
Depicted is the sequence of the mouse androgen receptor and which residues are cross-linked with each other. We can also upload a 3D protein structure to display the crosslinks. For that click on "IMPORT" and "PDB" in the top left corner.



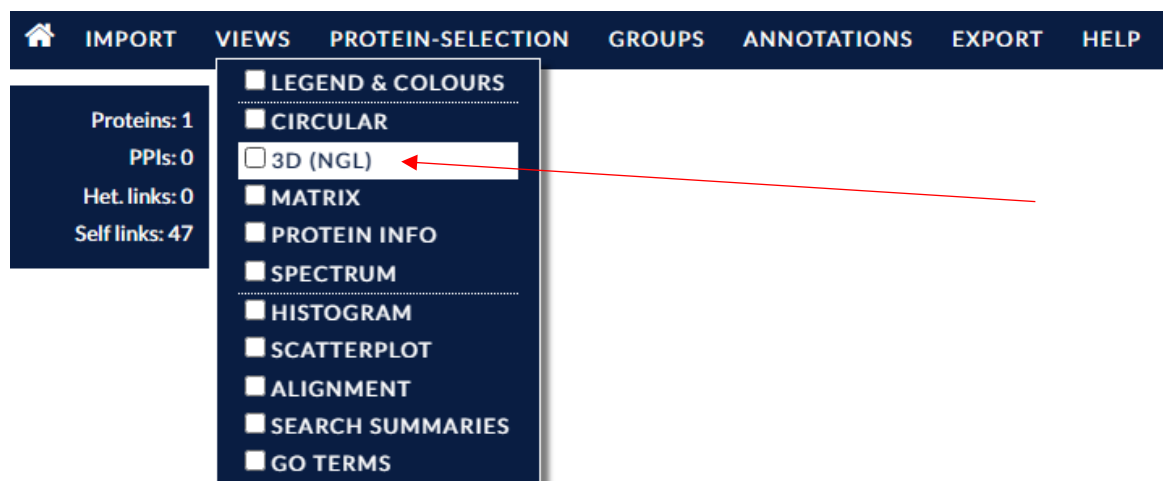
You can either select a local PDB file or enter a 4-character PDB code. We are going to do the later. The PDB short code of the androgen receptor of the mouse is "2QPY", we enter that into the field and press the enter key.



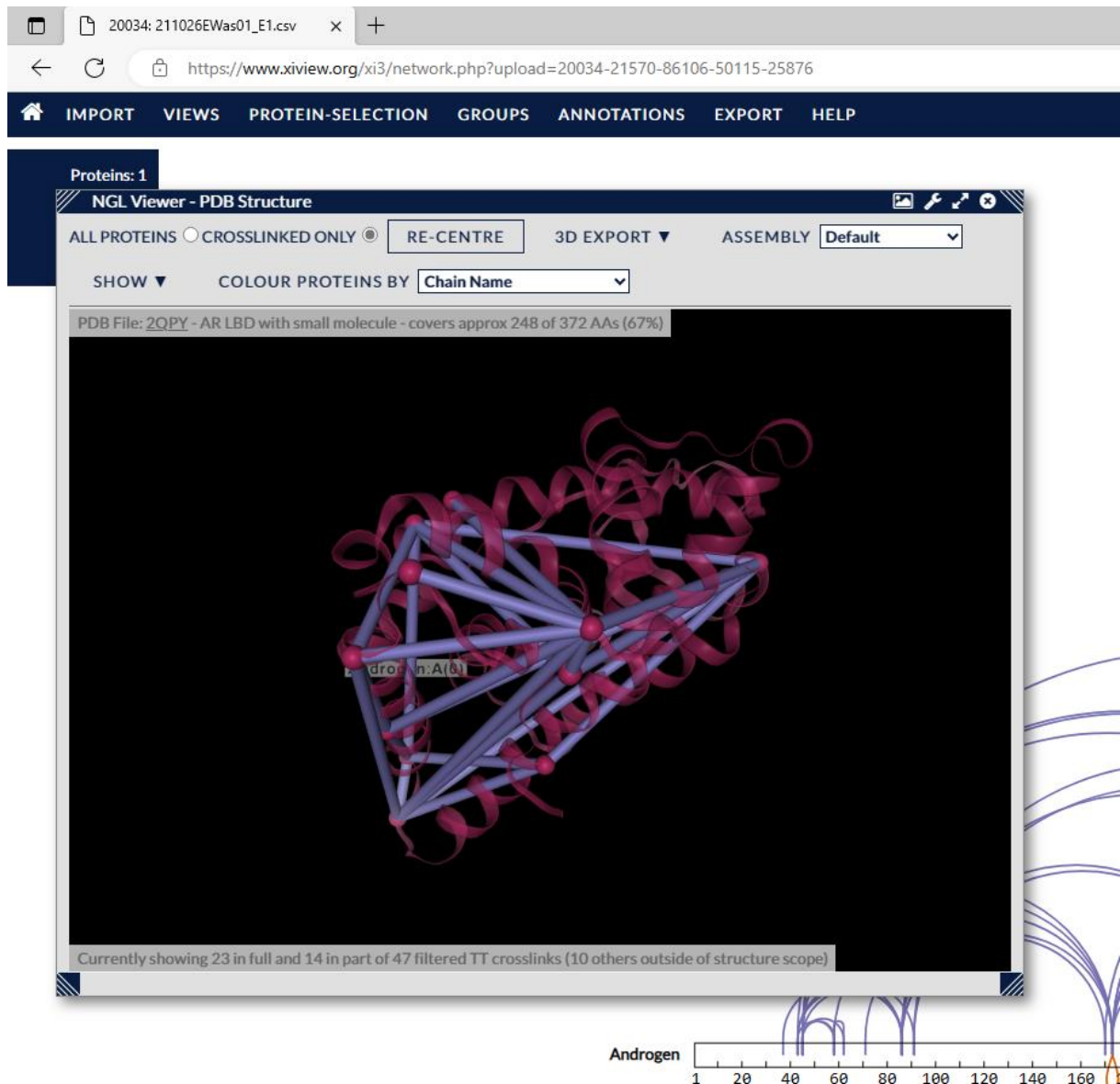
You should get a success message.



Now close the "PDB File Chooser" and select "VIEWS" and "3D (NGL)" in the top left corner.

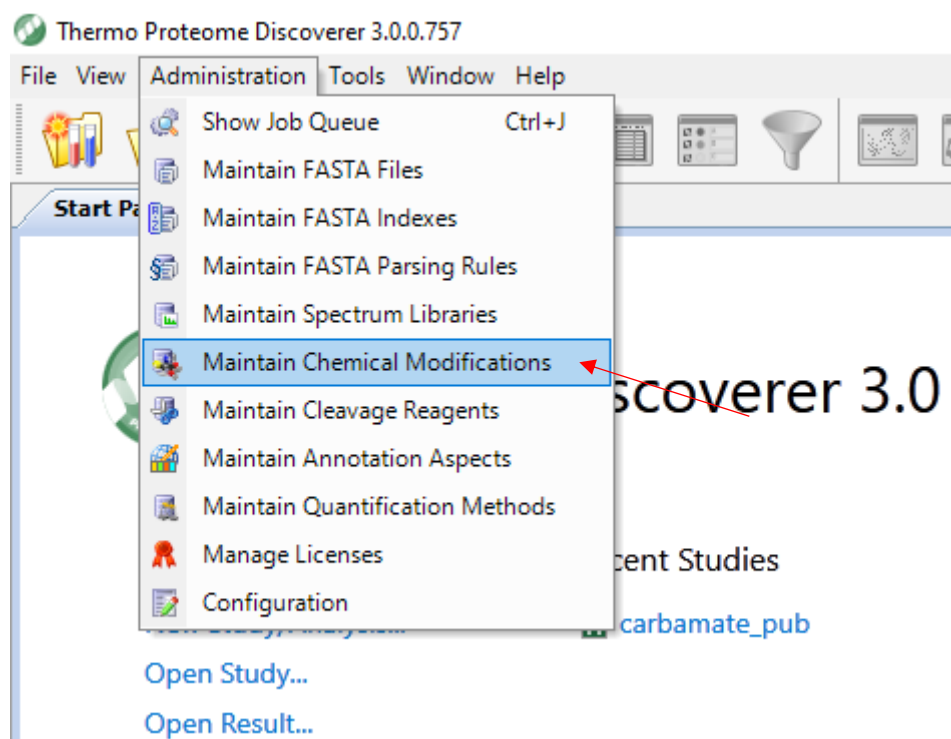


You should be presented with the 3D structure of the protein and the mapped crosslinks.

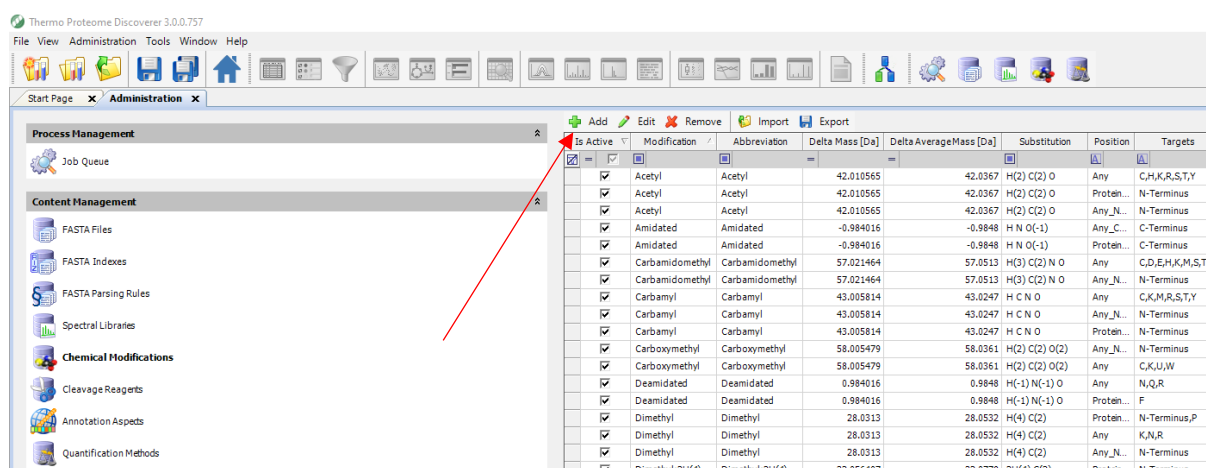


Defining a crosslinker in Proteome Discoverer

If you need to define your own crosslinker in Proteome Discoverer, open the “Maintain Chemical Modifications” window from the “Administration” menu.



Click on “Add” in the top middle of the screen to add a new modification.



First you have to enter a name, abbreviation, delta mass, delta average mass and the atomic composition in the “General” tab of the “Add Chemical Modification” window that pops up.

Add Chemical Modification

Name: Abbreviation:

General Neutral Losses Diagnostic Ions Crosslinking

Position: Unimod Accesion:

Delta Mass [Da]: Delta Average Mass [Da]:

Substitution: Leaving Group:

Amino Acid Site(s): Double click on a row to activate, press [ENTER] to add or [DEL]/Delete button to remove rows.

Target Amino Acid	Classification

Help OK Cancel

For this example, we are going to define a new DSSO variant that reacts with K, S, T, and Y.

Add Chemical Modification

Name: Abbreviation:

General Neutral Losses Diagnostic Ions Crosslinking

Position: Unimod Accesion:

Delta Mass [Da]: Delta Average Mass [Da]:

Substitution: Leaving Group:

Amino Acid Site(s): Double click on a row to activate, press [ENTER] to add or [DEL]/Delete button to remove rows.

Target Amino Acid	Classification

Help OK Cancel

Next, we are going to define the target amino acids in the "Amino Acid Site(s)" box.

The screenshot shows the 'Add Chemical Modification' dialog box with the 'General' tab selected. The 'Name' field contains 'DSSO K S T Y' and the 'Abbreviation' field contains 'DSSO:KSTY'. The 'Position' dropdown is set to 'Any' and 'Unimod Accesion' is '0'. 'Delta Mass [Da]' is '158.00376' and 'Delta Average Mass [Da]' is '158.17636'. 'Substitution' is 'H(6) C(6) O(3) S' and 'Leaving Group' is empty. The 'Amino Acid Site(s)' section has a yellow instruction bar: 'Double click on a row to activate, press [ENTER] to add or [DEL]/Delete button to remove rows.' Below this is a table with two columns: 'Target Amino Acid' and 'Classification'. The table contains four rows for K, S, T, and Y, all classified as 'CID cleavable crosslink'. Each row has a red 'X' icon in the right margin. A new empty row is highlighted in blue at the bottom of the table.

Target Amino Acid	Classification
K	CID cleavable crosslink
S	CID cleavable crosslink
T	CID cleavable crosslink
Y	CID cleavable crosslink

Change to the "Crosslinking" tab and select "Cleavable Crosslink".

The screenshot shows the 'Add Chemical Modification' dialog box with the 'Crosslinking' tab selected. The 'Name' and 'Abbreviation' fields remain the same. The 'Crosslinking' dropdown menu is open, showing three options: 'None', 'Cleavable Crosslink', and 'Non-cleavable Crosslink'. A red arrow points to the 'Cleavable Crosslink' option. The background of the dialog box is dimmed, and a yellow message bar at the top says 'Not a crosslink modification.'

Now enter the two crosslink fragments that can form during cleavage. You only need to enter "Name", "Abbreviation", "Substitution" and "Target(s)", the rest is automatically calculated by Proteome Discoverer.

Add Chemical Modification

Name: Abbreviation:

General Neutral Losses Diagnostic Ions **Crosslinking**

Cleavable Crosslink Double click on a row to activate, press [ENTER] to add or [DEL]/Delete button to remove rows.

Crosslink Fragments:

Name	Abbreviation	Substitution	Delta Mass	Delta Average Mass	Target(s)	
alkene	alkene	C3H2O	54.01056	54.04749	K,S,T,Y	<input type="button" value="x"/>
thiol	thiol	C3H2OS	85.98264	86.11358	K,S,T,Y	<input type="button" value="x"/>
						<input type="button" value="x"/>

Connected Fragments:

Left Fragment Right Fragment

Now add the fragments to the "Connected Fragments" box by selecting first the lighter fragment and "Add Selected Fragment" and then the heavier one and "Add Selected Fragment".

Add Chemical Modification

Name: Abbreviation:

General Neutral Losses Diagnostic Ions **Crosslinking**

Cleavable Crosslink Double click on a row to activate, press [ENTER] to add or [DEL]/Delete button to remove rows.

Crosslink Fragments:

Name	Abbreviation	Substitution	Delta Mass	Delta Average Mass	Target(s)	
alkene	alkene	C3H2O	54.01056	54.04749	K,S,T,Y	<input type="button" value="x"/>
thiol	thiol	C3H2OS	85.98264	86.11358	K,S,T,Y	<input type="button" value="x"/>
						<input type="button" value="x"/>

Connected Fragments:

Left Fragment Right Fragment

The result should look like this:

Add Chemical Modification

Name: Abbreviation:

General Neutral Losses Diagnostic Ions **Crosslinking**

Cleavable Crosslink Double click on a row to activate, press [ENTER] to add or [DEL]/Delete button to remove rows.

Crosslink Fragments:

Name	Abbreviation	Substitution	Delta Mass	Delta Average Mass	Target(s)	
alkene	alkene	C3H2O	54.01056	54.04749	K,S,T,Y	<input type="button" value="x"/>
thiol	thiol	C3H2OS	85.98264	86.11358	K,S,T,Y	<input type="button" value="x"/>
						<input type="button" value="x"/>

Connected Fragments:

Left Fragment	Right Fragment	
alkene	thiol	<input type="button" value="x"/>

In the "Diagnostic Ions" tab any crosslinker specific diagnostic ions should be added. MS Annika will use these to identify cross-linked spectra if no doublets are found. Alternatively, the diagnostic ions can also be set directly in the "MS Annika Detector" node.

Add Chemical Modification


Name: Abbreviation:

General Neutral Losses **Diagnostic Ions** Crosslinking

Double click on a row to activate, press [ENTER] to add or [DEL]/Delete button to remove rows.

Name	Formula	Monoisotopic Mass	Average Mass	Target(s)	
					<input type="button" value="x"/>


If your crosslinker yields more than two fragments during cleavage, these should be specified in the "Neutral Losses" tab. Any additional fragment not specified in the "Crosslinking" tab should be specified here (if they are expected to be observed in the cross-linked spectra). Additionally, you also need to adjust the "Additional Crosslink Doublet Distances" setting to cover all possible doublet distances that might appear due to the additional fragments. MS Annika can deal with any number of possible doublets as long as the light fragment stays constant. If you are synthesizing a new crosslinker that yields more than two fragments, please don't hesitate to reach out to us to make sure all the settings are set correctly.

 Add Chemical Modification — □ ×

Name: Abbreviation:

General **Neutral Losses** Diagnostic Ions Crosslinking

Double click on a row to activate, press [ENTER] to add or [DEL]/Delete button to remove rows.

Chemical Formula	Monoisotopic Mass	Average Mass	Target(s)	
<input type="text"/>				

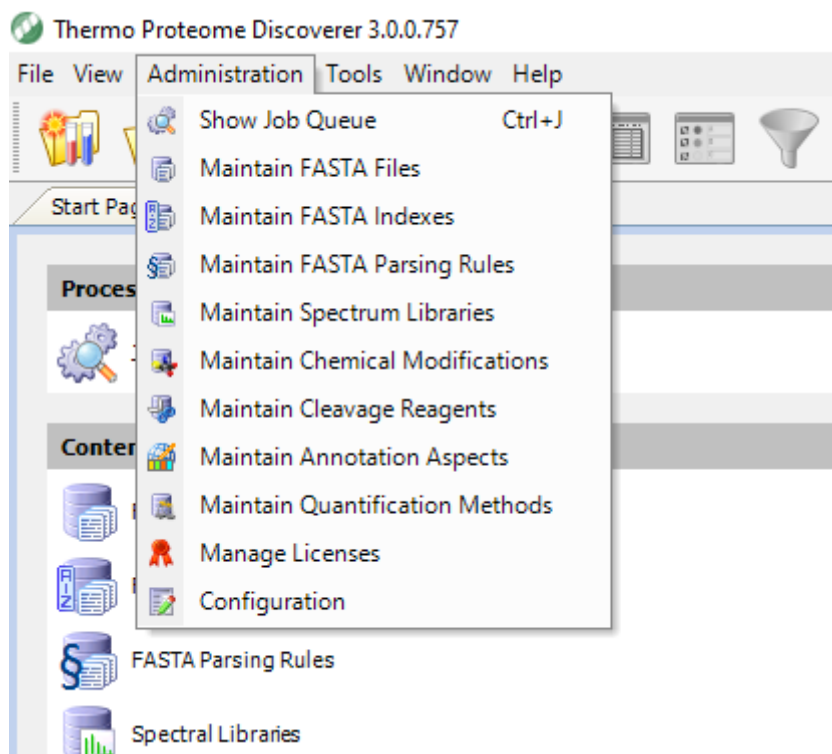
Help OK Cancel

When everything is set, click "OK" in the bottom right and the new crosslinker will be added to the list of available crosslinkers in Proteome Discoverer.

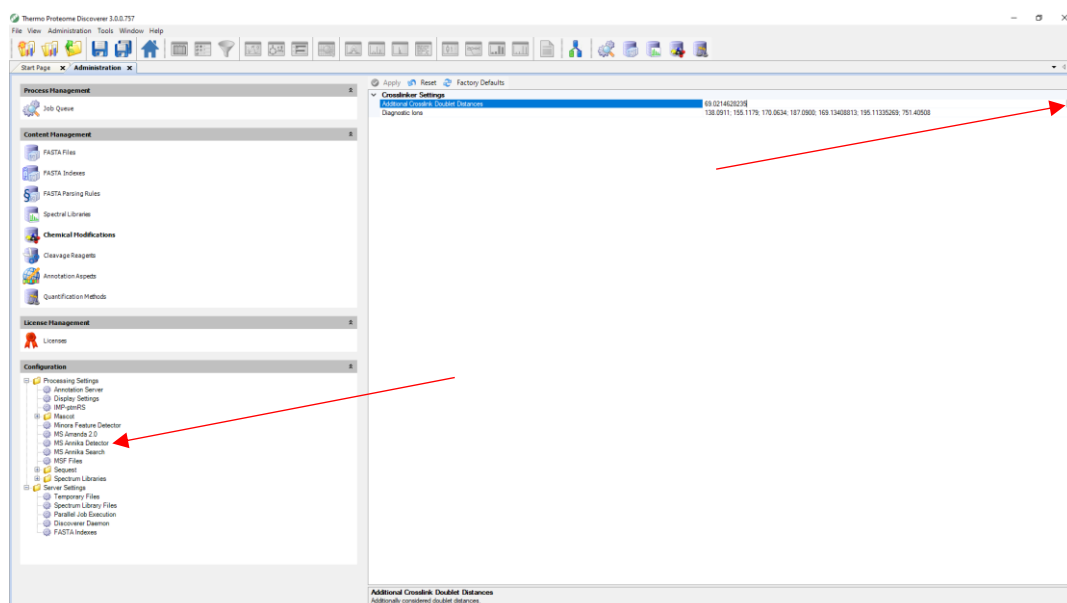
<input checked="" type="checkbox"/>	DSSO	DSSO	158.00376	158.17636	H(6) C(6) O(3) S	Any	K
<input checked="" type="checkbox"/>	DSSO Amidated	DSSO:NH3	175.03031	175.20692	H(9) C(6) O(3) NS	Any	K
<input checked="" type="checkbox"/>	DSSO Hydrolyzed	DSSO:H2O	176.01433	176.19165	H(8) C(6) O(4) S	Any	K
<input checked="" type="checkbox"/>	DSSO K S T Y	DSSO:KSTY	158.00376	158.17636	H(6) C(6) O(3) S	Any	K,S,T,Y

Changing MS Annika settings and Additional Crosslink Doublet Distances

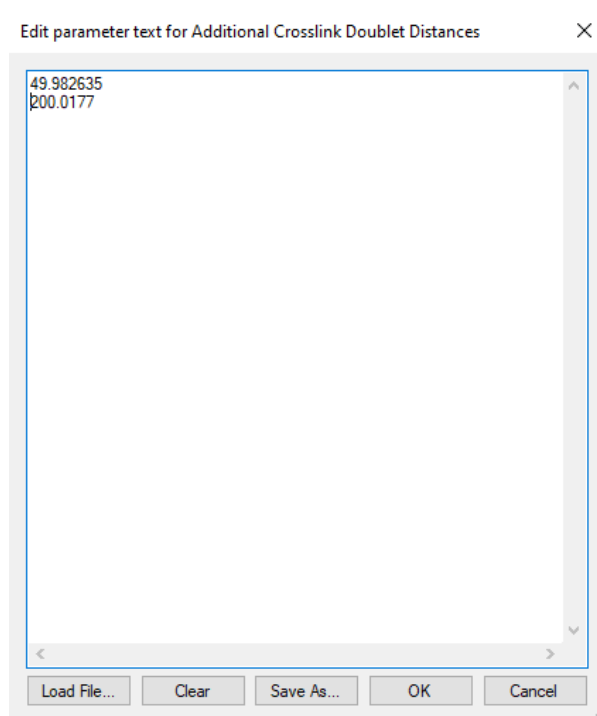
Certain menu items of MS Annika like “Additional Crosslink Doublet Distances” and “Diagnostic Ions” can only be modified in the MS Annika settings. For that go to “Administration” and “Configuration”.



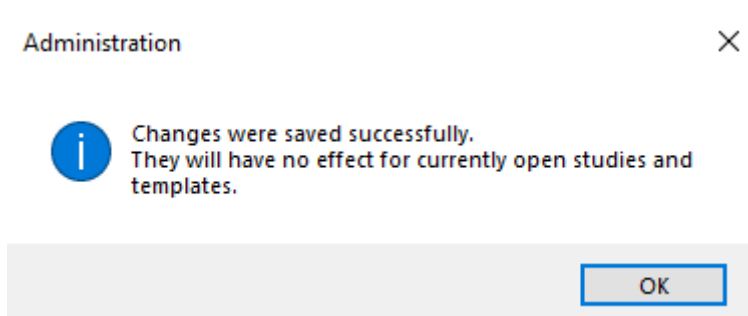
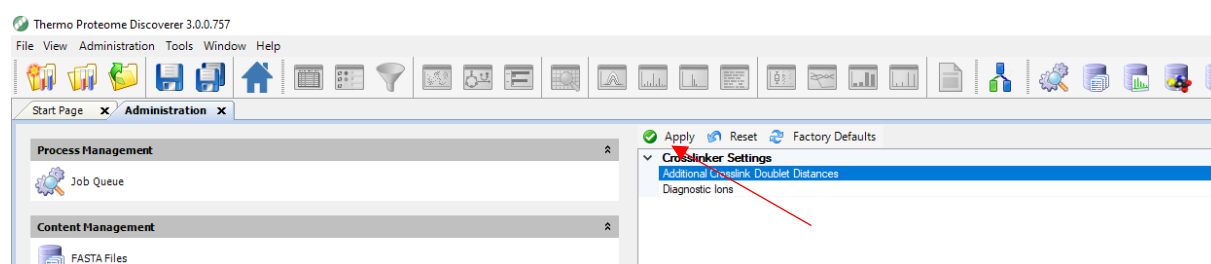
Then select the node that you want to adjust the settings for. For this example, we will add another crosslink doublet distance, more specifically that of DSBSO. Select the “MS Annika Detector” entry and then select “Additional Crosslink Doublet Distances” and click on the three ... that appear in the right corner.



The "Edit parameter text for Additional Crosslink Doublet Distances" window should pop up. By default, a single entry is pre-defined: 49.982635 – the mass difference of the alkene/sulfenic acid doublet of DSSO. We are now going to add the mass difference of the alkene/sulfenic acid doublet of DSBSO which is 200.0177 Da. Begin a new line and enter "200.0177", then click "OK" to confirm the change.



To apply the made change we additionally need to hit "Apply" in the top left corner of the settings window. The changes will not take effect until re-opening any open study.



Additionally recommended Proteome Discoverer nodes

The MS2 search of MS Annika can heavily benefit from the additional pre-processing steps carried out by the IMP MS2 Spectrum Processor node. This applies to all input formats whether it's Thermo RAW files, MGF files or Bruker timsTOF files and to all fragmentation methods whether it's CID, ETD, HCD or stepped HCD.

More information about the IMP MS2 Spectrum Processor node can be found here:

<https://ms.imp.ac.at/?action=spectrum-processor>

The appropriate version for your Proteome Discoverer version can be found in the table displayed here:

<https://ms.imp.ac.at/?action=home>

The version for Proteome Discoverer 3.0 can be directly downloaded from here:

https://ms.imp.ac.at/?file=spectrum-processor/ms2spectrumprocessor_3.0.zip

An exemplary workflow using the IMP MS2 Spectrum Processor node with MS Annika can be found here:

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

Important note: The IMP MS2 Spectrum Processor cannot be used in combination with the MS3 search of MS Annika as the MS3 search algorithm relies on the isotope distribution found in the MS2 spectra. Using the IMP MS2 Spectrum Processor node removes this information from the spectra and breaks the MS3 search.

Getting help

In case you run into problems, something isn't clear, you found a bug, or you have a suggestion, please reach out to us!

You can either open an issue on GitHub at <https://github.com/hgb-bin-proteomics/MSAnnika/issues> or contact us directly at:

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

More from our research group:

- MS Amanda: <https://ms.imp.ac.at/?goto=msamanda>
- MS Ana: <https://ms.imp.ac.at/?action=ms-ana>
- <https://bioinformatics.fh-hagenberg.at/>