MaXLinker Manual

1. What is MaXLinker?

MaXLinker is a search engine designed to identify cross-links from cross-linking mass spectrometry (XL-MS) data. More specifically it can identify cross-links from samples cross-linked by any cleavable cross-linker using MS2-MS3 XL-MS fragmentation strategy.

2. Dependencies:

- Python 3.7
- Tkinter (Usually an integral part of python installation)
- Proteome Discoverer 2.1

3. Installation and Running

Linux (Tested on Ubuntu 16.04 LTS 64-bit):

(i) Download the Linux-specific folder “MaXLinker_Linux.zip” and unzip it

(ii) The applications can be launched by the following commands through terminal
Pre-processing: “python3 maxlinker_preprocess_gui.py”
Main search: “python3 maxlinker_gui.py”

Mac (Tested on Mac OS High Sierra version 10.13.4):

(i) Download the Mac-specific folder “MaXLinker_Mac.zip” and unzip it

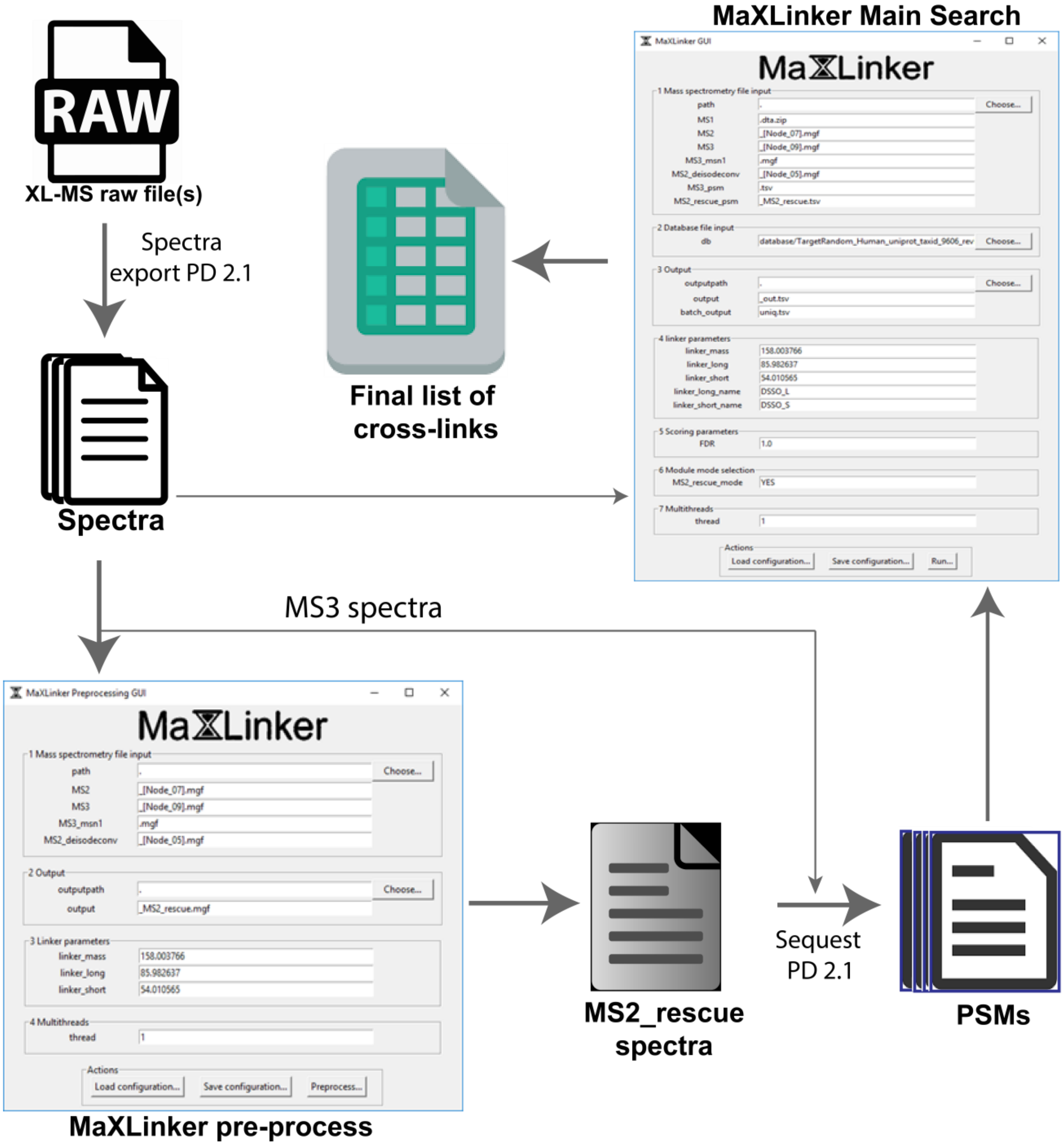
(ii) The applications can be launched by the following commands through terminal
Pre-processing: “python3 maxlinker_preprocess_gui.py”
Main search: “python3 maxlinker_gui.py”

Windows (Tested on Windows 10 Enterprise 64-bit)

(i) Download the Windows-specific folder “MaXLinker_Windows.zip” and unzip it

(ii) The applications can be launched by double-clicking specific scripts
Pre-processing: Double-click “MaXLinker_Preprocess.bat”
Main search: Double-click “MaXLinker_Main.bat”

4. MaXLinker work-flow



MaXLinker GUI

MaXLinker

1 Mass spectrometry file input

path	.
MS1	.dta.zip
MS2	_[Node_07].mgf
MS3	_[Node_09].mgf
MS3_msn1	.mgf
MS2_deisodeconv	_[Node_05].mgf
MS3_psm	.tsv
MS2_rescue_psm	_MS2_rescue.tsv

2 Database file input

db	database/TargetRandom_Human_uniprot_taxid_9606_rev

3 Output

outputpath	.
output	_out.tsv
batch_output	uniq.tsv

4 linker parameters

linker_mass	158.003766
linker_long	85.982637
linker_short	54.010565
linker_long_name	DSSO_L
linker_short_name	DSSO_S

5 Scoring parameters

FDR	1.0

6 Module mode selection

MS2_rescue_mode	YES

7 Multithreads

thread	1

MaXLinker Preprocessing GUI

MaXLinker

1 Mass spectrometry file input

path	.
MS2	_[Node_07].mgf
MS3	_[Node_09].mgf
MS3_msn1	.mgf
MS2_deisodeconv	_[Node_05].mgf

2 Output

outputpath	.
output	_MS2_rescue.mgf

3 Linker parameters

linker_mass	158.003766
linker_long	85.982637
linker_short	54.010565

4 Multithreads

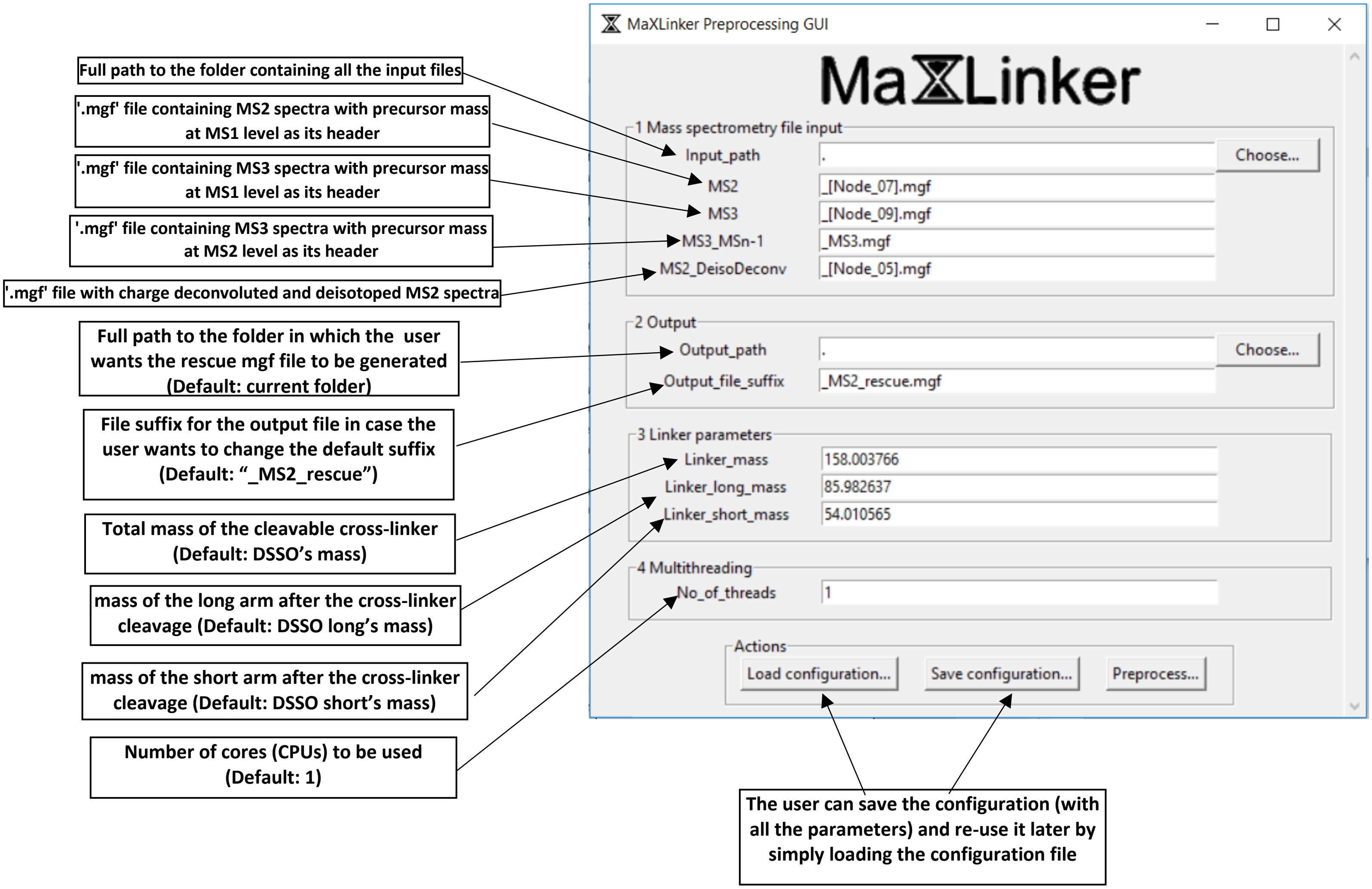
thread	1

5. Input / Output parameters

5.1 Pre-processing

5.1.1 Exporting spectra using Proteome Discoverer 2.1

The user need export spectra (at different levels such as MS1, MS2 and MS3) from the raw files using Proteome Discoverer 2.1. The templates that can be used for such export are provided in folder “Demo_files/PD_Templates/Spectra_export”. Additionally, one of the nodes in the spectral export templates require ‘MS2 Spectrum Processor’ node available at ‘http://ms.imp.ac.at/index.php?action=spectrum-processor’ that facilitates deisotoping and charge deconvolution of MS2 spectra. Ideally, the export should create five files (specifically four ‘mgf’ files and one ‘dta’ folder) per each raw file out of which the four mgf files are required as input in the Preprocessing step. Descriptions of those files are provided in the following annotation under “1 Mass spectrometry file input” option in the Preprocessing GUI image.



1 Mass spectrometry file input

Input_path: .

MS2: _[Node_07].mgf

MS3: _[Node_09].mgf

MS3_MSn-1: _MS3.mgf

MS2_DeisoDeconv: _[Node_05].mgf

2 Output

Output_path: .

Output_file_suffix: _MS2_rescue.mgf

3 Linker parameters

Linker_mass: 158.003766

Linker_long_mass: 85.982637

Linker_short_mass: 54.010565

4 Multithreading

No_of_threads: 1

Full path to the folder containing all the input files

'.mgf' file containing MS2 spectra with precursor mass at MS1 level as its header

'.mgf' file containing MS3 spectra with precursor mass at MS1 level as its header

'.mgf' file containing MS3 spectra with precursor mass at MS2 level as its header

'.mgf' file with charge deconvoluted and deisotoped MS2 spectra

Full path to the folder in which the user wants the rescue mgf file to be generated (Default: current folder)

File suffix for the output file in case the user wants to change the default suffix (Default: “_MS2_rescue”)

Total mass of the cleavable cross-linker (Default: DSSO’s mass)

mass of the long arm after the cross-linker cleavage (Default: DSSO long’s mass)

mass of the short arm after the cross-linker cleavage (Default: DSSO short’s mass)

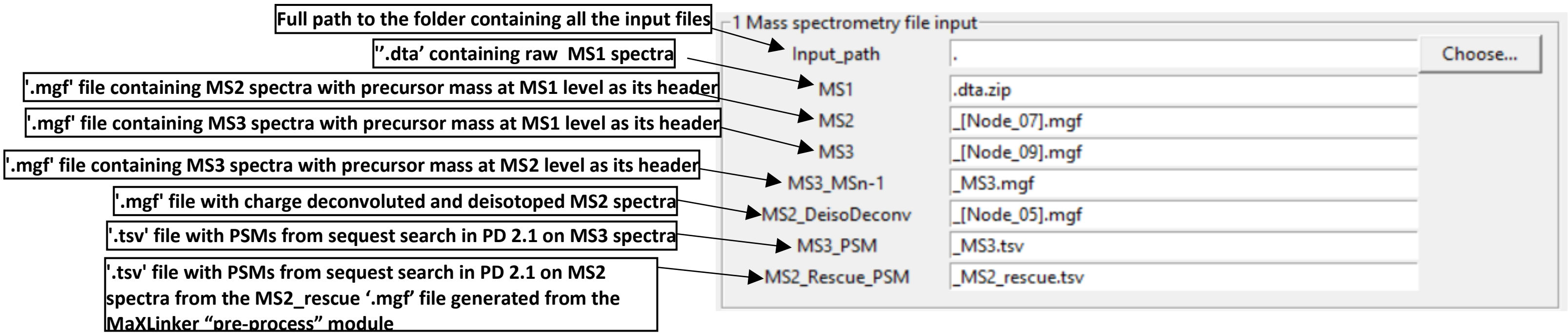
Number of cores (CPUs) to be used (Default: 1)

The user can save the configuration (with all the parameters) and re-use it later by simply loading the configuration file

5.2 Performing PSM search using Sequest available in Proteome Discoverer 2.1

The preprocessing step generates a .mgf file that consists of spectra with theoretical precursor masses for the potentially rescued spectra. This rescue file along with the original raw file of each fraction n is subjected to a sequest search using Proteome Discoverer 2.1. The templates that can be used for such search are provided in folder “Demo_files/PD_Templates/PSM_search”. Please note that the search parameters are different for the PSM search of MS3 spectra compared to that of MS2_rescue spectra. For a detailed description of search parameters, please see “Data Processing” under “EXPERIMENTAL PROCEDURES” in the manuscript.

5.3 Main Search



Full path to the folder containing all the input files

’’.dta’ containing raw MS1 spectra

'.mgf' file containing MS2 spectra with precursor mass at MS1 level as its header

'.mgf' file containing MS3 spectra with precursor mass at MS1 level as its header

'.mgf' file containing MS3 spectra with precursor mass at MS2 level as its header

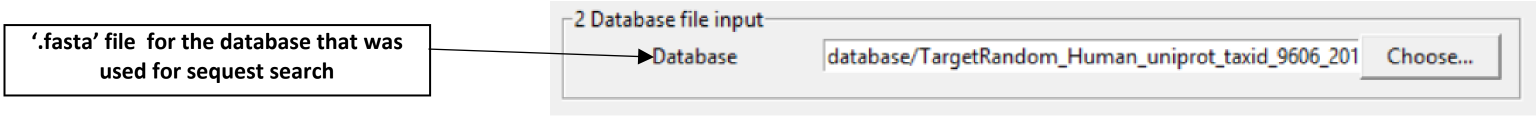
'.mgf' file with charge deconvoluted and deisotoped MS2 spectra

'.tsv' file with PSMs from sequest search in PD 2.1 on MS3 spectra

'.tsv' file with PSMs from sequest search in PD 2.1 on MS2 spectra from the MS2_rescue ‘.mgf’ file generated from the MaXLinker “pre-process” module

1 Mass spectrometry file input

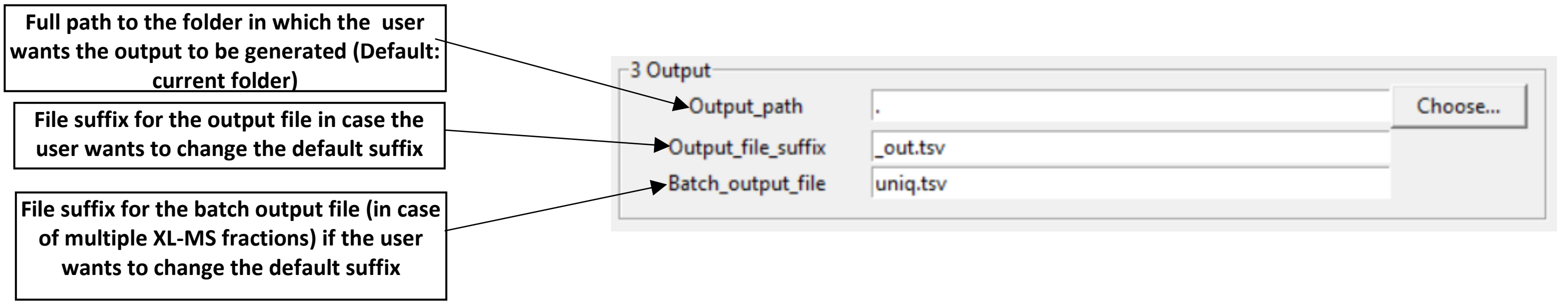
Input_path	.
MS1	.dta.zip
MS2	_[Node_07].mgf
MS3	_[Node_09].mgf
MS3_MSn-1	_MS3.mgf
MS2_DeisoDeconv	_[Node_05].mgf
MS3_PSM	_MS3.tsv
MS2_Rescue_PSM	_MS2_rescue.tsv



‘.fasta’ file for the database that was used for sequest search

2 Database file input

Database	database/TargetRandom_Human_uniprot_taxid_9606_201



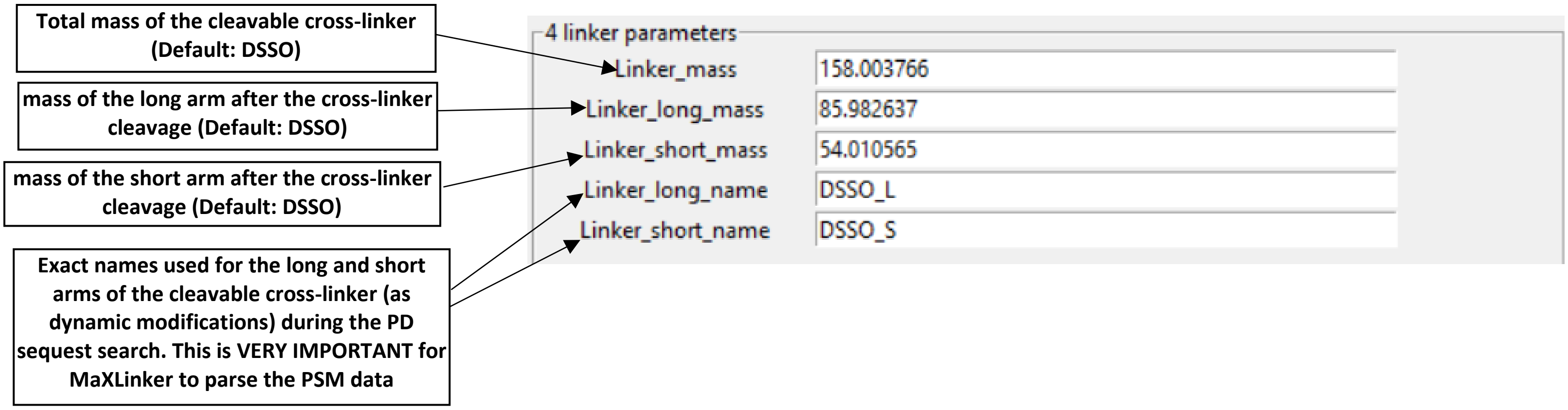
Full path to the folder in which the user wants the output to be generated (Default: current folder)

File suffix for the output file in case the user wants to change the default suffix

File suffix for the batch output file (in case of multiple XL-MS fractions) if the user wants to change the default suffix

3 Output

Output_path	.
Output_file_suffix	_out.tsv
Batch_output_file	uniq.tsv



4 linker parameters

Linker_mass 158.003766

Linker_long_mass 85.982637

Linker_short_mass 54.010565

Linker_long_name DSSO_L

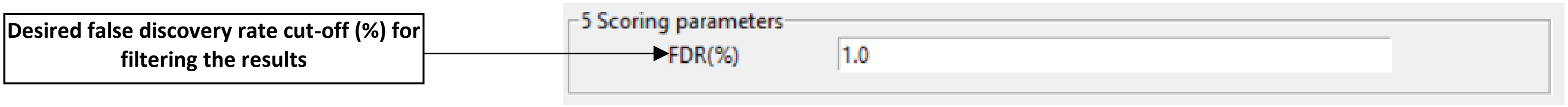
Linker_short_name DSSO_S

Total mass of the cleavable cross-linker (Default: DSSO)

mass of the long arm after the cross-linker cleavage (Default: DSSO)

mass of the short arm after the cross-linker cleavage (Default: DSSO)

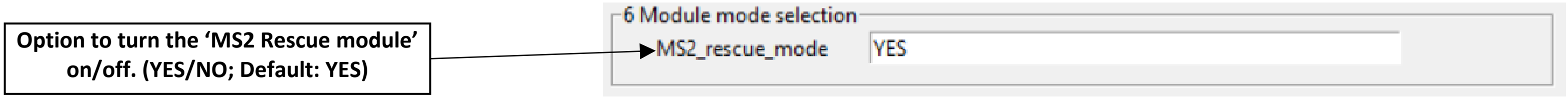
Exact names used for the long and short arms of the cleavable cross-linker (as dynamic modifications) during the PD sequest search. This is VERY IMPORTANT for MaXLinker to parse the PSM data



5 Scoring parameters

FDR(%) 1.0

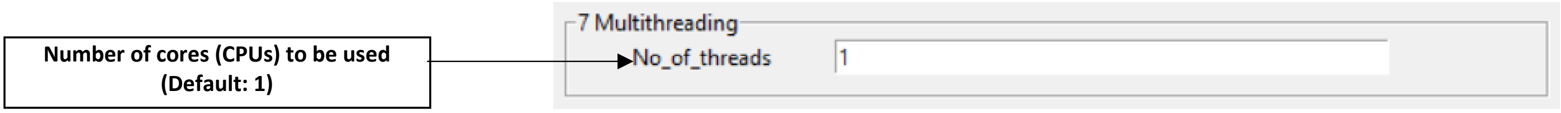
Desired false discovery rate cut-off (%) for filtering the results



6 Module mode selection

MS2_rescue_mode YES

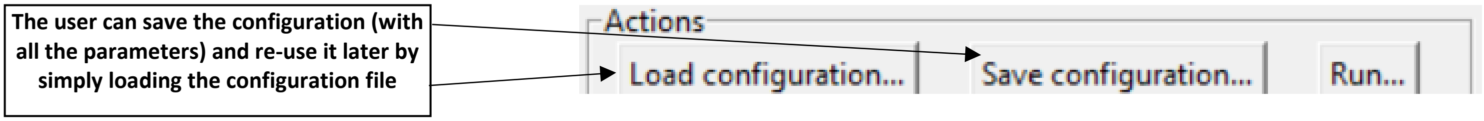
Option to turn the ‘MS2 Rescue module’ on/off. (YES/NO; Default: YES)



7 Multithreading

No_of_threads 1

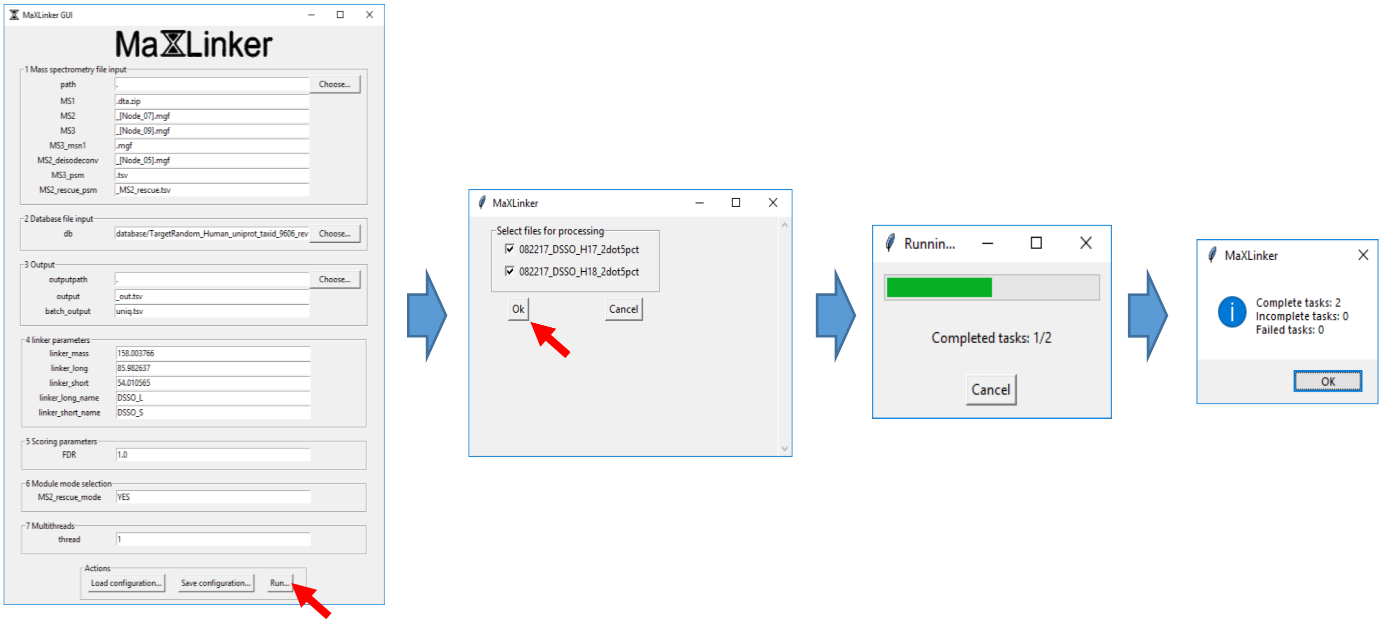
Number of cores (CPUs) to be used (Default: 1)



The user can save the configuration (with all the parameters) and re-use it later by simply loading the configuration file

Please refer to the sample input files provided in /Demo_files folder for a better understanding of the input file format. Also, please note that the PSM input files **MUST** contain the following columns (not necessarily in the same order) exported from the sequest search performed in Proteome Discoverer 2.1: “**First Scan”**, “**Confidence”**, “**PSM Ambiguity”**, “**Annotated Sequence**”, “**Protein Accessions**”, “**Protein Descriptions**”, “**MH+ [Da]**”, “**Modifications**”, “**Percolator q-Value**”. Please refer the sample input files for more information.

6. Quick glance of work-flow for MaXLinker's main search



MaXLinker GUI

1 Mass spectrometry file input

path	.
MS1	.dta.zip
MS2	_[Node_07].mgf
MS3	_[Node_09].mgf
MS3_msn1	.mgf
MS2_deisodeconv	_[Node_05].mgf
MS3_psm	.tsv
MS2_rescue_psm	_MS2_rescue.tsv

2 Database file input

db	database/TargetRandom_Human_uniprot_taxid_9606_rev

3 Output

outputpath	.
output	_out.tsv
batch_output	uniq.tsv

4 linker parameters

linker_mass	158.003766
linker_long	85.982637
linker_short	54.010565
linker_long_name	DSSO_L
linker_short_name	DSSO_S

5 Scoring parameters

FDR	1.0

6 Module mode selection

MS2_rescue_mode	YES

7 Multithreads

thread	1

MaXLinker

Select files for processing

☑ 082217_DSSO_H17_2dot5pct

☑ 082217_DSSO_H18_2dot5pct

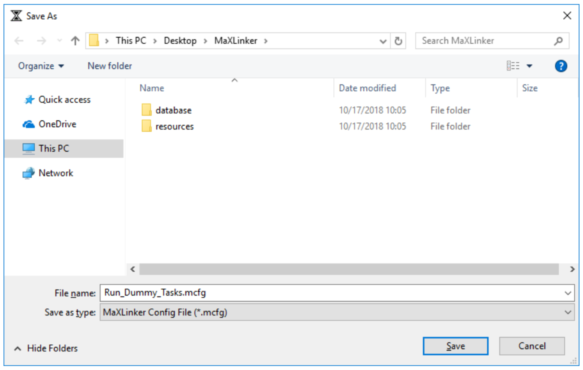
Completed tasks: 1/2

Complete tasks: 2
Incomplete tasks: 0
Failed tasks: 0

It is highly recommended that the users perform a 'test search' on the sample input files provided to make sure they can reproduce the results from the sample output files to make sure the software is setup and working properly. Please reach out to us for support, if you notice any inconsistencies.

7. Other features

7.1 Saving / Loading parameters



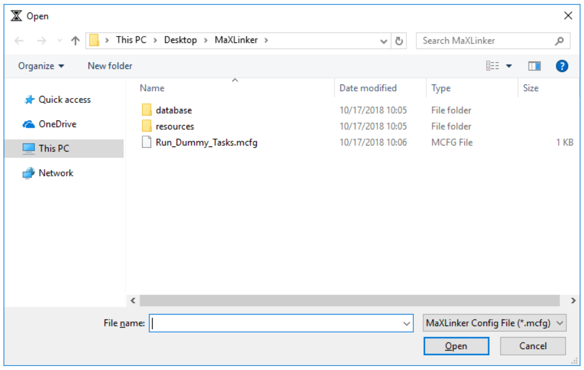
Save As

This PC › Desktop › MaXLinker ›

Name	Date modified	Type	Size
database	10/17/2018 10:05	File folder	
resources	10/17/2018 10:05	File folder	

File name: Run_Dummy_Tasks.mcfg

Save as type: MaXLinker Config File (*.mcfg)



Open

This PC › Desktop › MaXLinker ›

Name	Date modified	Type	Size
database	10/17/2018 10:05	File folder	
resources	10/17/2018 10:05	File folder	
Run_Dummy_Tasks.mcfg	10/17/2018 10:06	MCFG File	1 KB

File name:

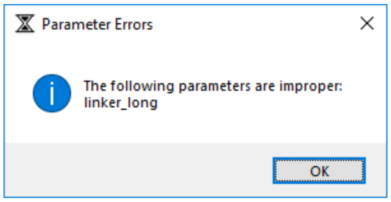
MaXLinker Config File (*.mcfg)

Users can save and load the parameters by using the "Load configuration…" and "Save configuration…" buttons.

7.2 Running the main modules

When user clicks "Run…", the program validates the parameters and searches for files according to the provided suffices if no parameter error is detected.

7.3 Improper parameters

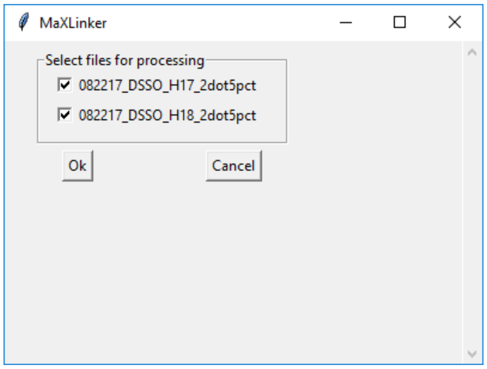


The following parameters are improper:
linker_long

The program attempts to detect improper parameters and gives warning. Users need to correct these parameters before they could successfully run the tasks.

7.4 File searching

The program looks for files in the input directory (and its sub-directories) and collect all file names with suffices matching those defined by users. Users can select the file prefixes they want to process.



Select files for processing

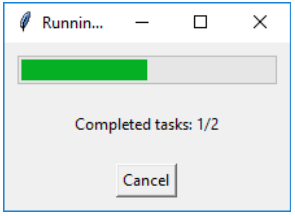
☑ 082217_DSSO_H17_2dot5pct

☑ 082217_DSSO_H18_2dot5pct

Some file prefixes are “greyed out” and not selectable because:

- Some files are missing (e.g. only 4 of 7 suffices could be matched)
- Duplicated file names are found.

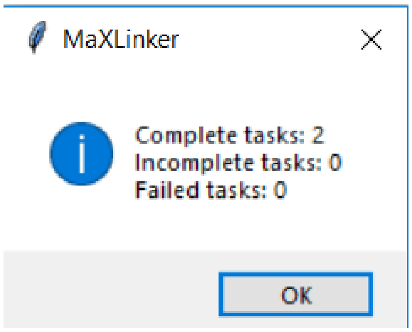
7.5 Progress



Completed tasks: 1/2

The program runs on the selected prefixes and shows the progress of completion.

7.6 Summary of completion status



Complete tasks: 2
Incomplete tasks: 0
Failed tasks: 0

Upon completion or cancellation, the program generates a brief summary on the completion status

- Complete: The tasks that complete successfully
- Incomplete: The tasks that are running or pending when users cancel the run
- Failed: The tasks with problems arise during the run