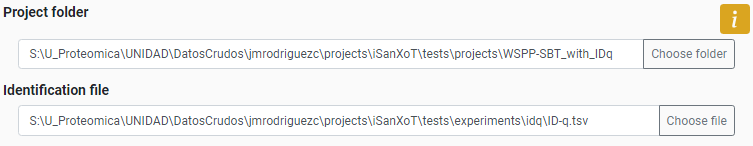
# Adaptors

### Main Input

The Main-Input adaptor provides the following parameters to your project:

* Project folder: describes the path to the folder where iSanXoT output files will be stored.
* Identification file: specifies the location of the file that contains the identification and quantification data.



**Figure 31. Panel of Main-Input adaptor with a worked-out example.**

The unique requirements for the Identification/quantification file are:

* Experiment column that contains the multiple names of your experiments.
* A column with identifiers that describe unequivocally a level. For example, if you want to use a WSPP-SBT “Complete module”, your identification file has to contain a column that describes the scans in a unique way.

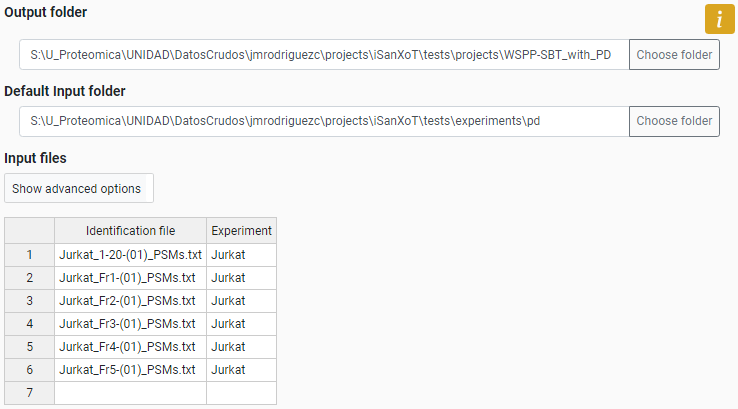
### Inputs from Proteome Discoverer

This adaptor allows you to transform the output of Proteome Discoverer (PD) into an Identification-Quantification file appropriate for iSanXoT workflows. iSanXoT has been tested for the outputs of PD 2.5.

The adaptor uses the following fields:

* Output folder: describes the path to the folder where iSanXoT output files will be stored.
* Default Input folder: specifies the location of the files containing PSM identification and quantification data (the PSMs.txt files).
* Input files: indicates which of the PSM.txt files stored in the input folder must be considered by iSanXoT. PSMs.txt file names are listed under Infile, while their experiment allocation is indicated in Experiment.

The following figure displays six PSMs.txt files originate from the same experiment (termed “Jurkat”), but larger projects may encompass several experiments (e.g. “TMT1”, “TMT2” and “TMT3”; or “Exp1”, “Exp2” and “Exp3”).



**Figure 32. Panel of "Input from PD" adaptor with a worked-out example.**

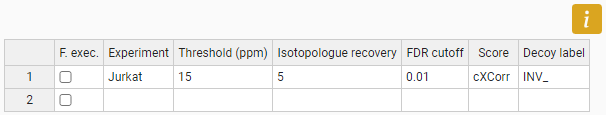
The PSM.txt files are plain text files that reveal information for every PSM obtained upon database searching; however, only a reduced subset of these data will be accessed by iSanXoT, as shown in the following Table.

|  |  |
| --- | --- |
| **Data accessed by iSanXoT from the Proteome Discoverer results (PSMs.txt files)** | |
| **Header** | **Description** |
| Spectrum File | Name of the raw LC-MS/MS file |
| First Scan | Scan number identifier |
| Sequence | Peptide amino acid sequence |
| Modifications | Unimod[[1]](#footnote-1) chemical or posttranslational modifications to peptide sequence |
| Charge | Peptide charge |
| Search Engine Rank | Rank of search engine |
| XCorr | Cross-correlation value as provided by SEQUEST[[2]](#footnote-2) algorithm |
| MHplus in Da | Measured monoisotopic protonated peptide mass in Da |
| Theo MHplus in Da | Theoretical monoisotopic protonated peptide mass in Da |
| Delta M in ppm | Difference between measured and theoretical monoisotopic mass in ppm |
| Protein Accessions | Accession codes for the proteins to which the peptide sequence is ascribed |
| Protein Descriptions | List of protein descriptions separated by comma |
| 113-121, 126-131, etc | Intensity of reporter ions |

#### Validating peptide identification with FDR

iSanXoT relies on the probability ratio (pRatio) method, an algorithm that calculates the probability of random peptide matching and provides the corresponding FDR for peptide identification. The FDR parameters displays the following fields:

* Forced execution: checkbox that determines to force the execution or not.
* Experiment: the aforementioned experiment allocation.
* Threshold (ppm): is the postscoring mass filtering cut-off to be applied after using wide mass windows in the database search, as was the case with these sample data. Threshold is actually equivalent to the relative deviation experimentally observed for precursor ions in a particular LC-MS/MS run.
* Isotopologue recovery: allows pRatio to recover precursor m/z values matching some 13C isotopologue when using wide mass windows in the LC-MS/MS acquisition that otherwise would remain unnoticed. For that, it must be indicated whether precursor m/z values should be tracked only around their experimental m/z (value = 1) or also ± 1 Th (value = 3) and ± 2 Th (value = 5, the one used here) away.
* FDR cutoff: establishes the FDR cut-off for PSM validation. The value used here (0.01, i.e. 1% FDR) implies that one in every 100 validated PSMs is incorrect.
* Score: determines the score used to calculate the FDR. These scores are “XCorr” (SEQUEST5 cross correlation score), or “cXCorr” (the corrected XCorr) will be used by pRatio for FDR calculation.
* Decoy label: is the tag attached to decoy protein identifiers in the concatenated protein database previously used for peptide identification (“INV\_” in this example).



**Figure 33. Task-Table of FDR module.**

### Inputs from MSFragger

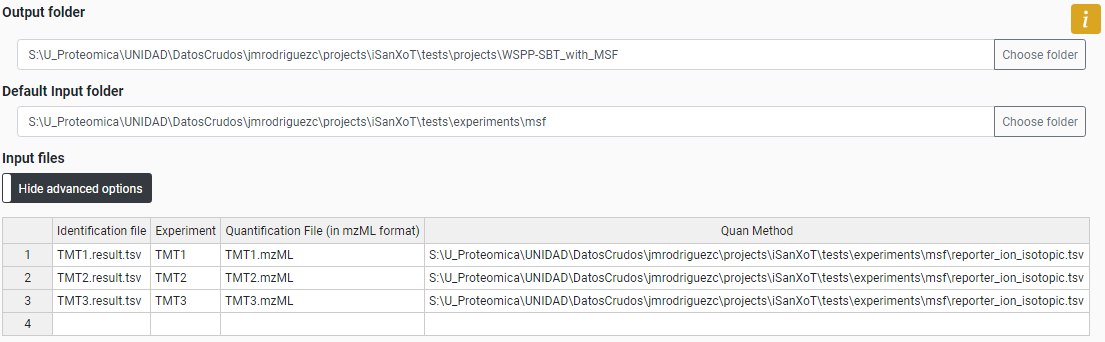
This adaptor allows you to transform the output of MSFragger [4] into an Identification-Quantification file appropriate for iSanXoT workflows. iSanXoT has been tested for the outputs of MSFragger version 3.1.1 (with FragPipe version 14.0).

The adaptor uses the following fields:

* Output folder: describes the path to the folder where iSanXoT output files will be stored.
* Default Input folder: specifies the location of the files containing results files.
* Input files: indicates which of the text plain files stored in the input folder must be considered by iSanXoT. The file names are listed under Identification file column, while their experiment is indicated in Experiment column.

So far, the MSFragger version used did not contain the quantification values. Thus, this adaptor allows you to extract the quantification values from the “mzML” files and then, pair them with the identifications. For that, the “Input files” table has the “advanced options”:

* Quantification file (in mzML format): indicating the mzML names if the files are located in the “Default Input folder”, or giving the absolute path of mzML files.
* Quan method: table files that describes the ion isotopic.



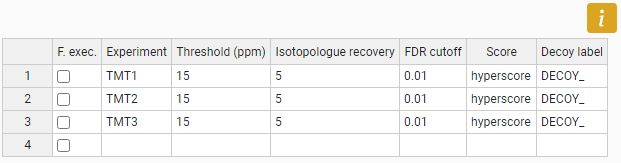
**Figure 34. Panel of "Input from MSF" adaptor with a worked-out example.**

The results files are plain text files that reveal information for every PSM; however, only a reduced subset of these data will be accessed by iSanXoT, as shown in the following Table.

|  |  |
| --- | --- |
| Table. **Data accessed by iSanXoT in the results from MSFragger** | |
| **Header** | **Description** |
| scannum | Scan number identifier |
| peptide | Peptide amino acid sequence |
| modification\_info | Unimod chemical or posttranslational modifications to peptide sequence |
| charge | Peptide charge |
| hit\_rank | Rank of search engine |
| calc\_neutral\_pep\_mass | Theoretical mass of the identified peptide ion in Da |
| massdiff | Difference between measured and theoretical precursor neutral mass |
| protein | Accession codes for the proteins to which the peptide sequence is ascribed |

#### Validating peptide identification with FDR

This adaptor also contains the probability ratio method that calculates the probability of random peptide matching and provides the corresponding FDR for peptide identification. This module contains the same parameters than the “Input from PD” adaptor. It should be pointed out that score to use has to be the “hyperscore” from MSFragger.



**Figure 35. Task-Table of FDR module for a work-out example with MSF input.**

### Inputs from Comet

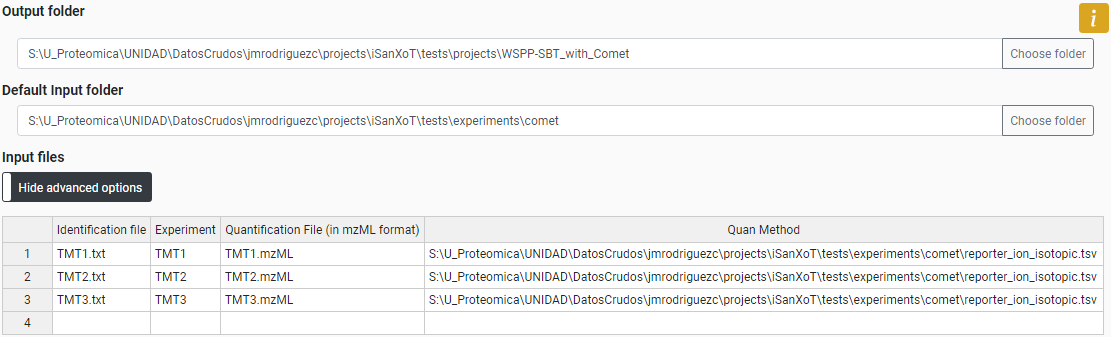
This adaptor transforms the output of Comet [5] into an Identification-Quantification file appropriate for iSanXoT workflows. iSanXoT has been tested for the outputs of Comet version 2017.01.

The adaptor uses the following fields:

* Output folder: describes the path to the folder where iSanXoT output files will be stored.
* Default Input folder: specifies the location of the files containing results files.
* Input files: indicates which of the text plain files stored in the input folder must be considered by iSanXoT. The file names are listed under Identification file column, while their experiment is indicated in Experiment column.

So far, the Comet version used did not contain the quantification values. Thus, this adaptor allows you to extract the quantification values from the “mzML” files and then, pair them with the identifications. For that, the “Input files” table has the “advanced options”:

* Quantification file (in mzML format): indicating the mzML names if the files are located in the “Default Input folder”, or giving the absolute path of mzML files.
* Quan method: table files that describes the ion isotopic.



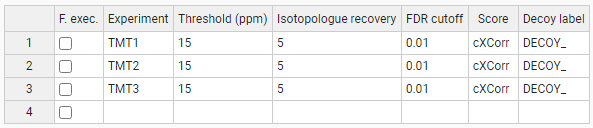
**Figure 36. Panel of "Input from Comet" adaptor with a worked-out example.**

The results files are plain text files that reveal information for every PSM; however, only a reduced subset of these data will be accessed by iSanXoT, as shown in the following Table.

|  |  |
| --- | --- |
| Table. **Data accessed by iSanXoT in the results from Comet** | |
| **Header** | **Description** |
| scan | Scan number identifier |
| plain\_peptide | Peptide amino acid sequence |
| modifications | Unimod chemical or posttranslational modifications to peptide sequence |
| charge | Peptide charge |
| num | Rank of search engine |
| xcorr | Cross-correlation value as provided by SEQUEST algorithm |
| calc\_neutral\_mass | Theoretical mass of the identified peptide ion in Da |
| exp\_neutral\_mass | Theoretical mass of the identified peptide ion in Da |
| protein | Accession codes for the proteins to which the peptide sequence is ascribed |

#### Validating peptide identification with FDR

This adaptor also contains the probability ratio method that calculates the probability of random peptide matching and provides the corresponding FDR for peptide identification. This module contains the same parameters than the “Input from PD” and “Input from MSFragger” adaptors.



**Figure 37. Task-Table of FDR module for a work-out example with Comet input.**

### Inputs from MaxQuant

This adaptor allows you to transform the output of MaxQuant into an Identification-Quantification file appropriate for iSanXoT workflows. iSanXoT has been tested for the outputs of MaxQuant 1.6.5.0.

The adaptor uses the following fields:

* Output folder: describes the path to the folder where iSanXoT output files will be stored.
* Default Input folder: specifies the location of the files containing results files.
* Input files: indicates which of the text plain files stored in the input folder. The “modificationSpecificPeptides.txt” is the file considered by iSanXoT. The file names are listed under Identification file column, while their experiment is indicated in Experiment column.

Identifier of the associated modification summary stored in the file “modificationSpecificPeptides.txt”. The following figure displays originate file from the experiment (called “PME12”), but larger projects may encompass several experiments.



**Figure 38. Panel of "Input from MaxQuant" adaptor with a worked-out example.**

The results files are plain text files that reveal information for every PSM; however, only a reduced subset of these data will be accessed by iSanXoT, as shown in the following Table.

|  |  |
| --- | --- |
| **Data accessed by iSanXoT from the MaxQuant (modifiedSpecificPeptides.txt)** | |
| **Header** | **Description** |
| Sequence | Peptide amino acid sequence |
| Modifications | Post-translational modifications contained within the sequence |
| Score | Andromeda identification score for the MS/MS spectrum |
| Proteins | The IPI identifiers of the proteins the identified peptide is associated with. |
| Protein Names | List of protein descriptions separated by comma |
| Intensities … | The intensities of the peaks in the fragmentation spectrum after top-N filtering |

The analysis of Label-Free starts with the quantification data from the peptide level and therefore, the WPP-SBT and WPPG-SBT “Complete modules” that start the integration at peptide level have been thought to these statistical analyses.

1. https://www.unimod.org/ [↑](#footnote-ref-1)
2. Eng J.K. *et al.* (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom*. **5**, 976–89. [↑](#footnote-ref-2)