MSnPy

Release 0.1.0

Ralf Weber, Thomas Lawson

Aug 04, 2021

CONTENTS:

CHAPTER

ONE

INTRODUCTION

The latest Documentation was generated on: Aug 04, 2021

1.1 Installation

- MSnPy requires Python 3.7.+.
- MSnPy is published under GNU General Public License v3.0.

Get the latest version via github

https://github.com/computational-metabolomics/msnpy

or the latest package at pypi or conda

https://pypi.python.org/pypi/msnpy
hhttps://anaconda.org/bioconda/msnpy

1.2 Developers & Contact Information

Developers:

Ralf J. M. Weber - r.j.weber@bham.ac.uk
Thomas Lawson - t.n.lawson@bham.ac.uk

Contact information:

Ralf J. M. Weber

Phenome Centre Birmingham
School of Biosciences, University of Birmingham
Edgbaston, Birmingham, B15 2TT, U.K.
Tel: +44(0)121 414 3388
Email: r.j.weber@bham.ac.uk

https://github.com/computational-metabolomics

1.3 Citation

Please cite us when using MSnPy in your work.

ref

CHAPTER

TWO

MSNPY PACKAGE

2.1 Subpackages

2.2 Submodules

2.3 msnpy.msnpy module

2.4 msnpy.processing module

msnpy.processing.**assign_precursor**(*peaklist: dimspy.models.peaklist.PeakList*, *header_frag: str*, *tolerance: float = 0.5*)

Parameters

- **peaklist** –
- **header_frag** –
- **tolerance** –

Returns

Return type

msnpy.processing.**create_graphs_from_scan_ids**(*scan_dependents: list*, *scan_events: dict*, *ion_injection_times: dict*)

Create Directed Graph from scan dependent relationships

Parameters

- **scan_dependents** –
- **scan_events** –
- **ion_injection_times** –

Returns

Return type

msnpy.processing.**create_spectral_trees**(*trees: Sequence[networkx.classes.ordered.OrderedDiGraph]*, *peaklists: Sequence[dimspy.models.peaklist.PeakList]*)

Parameters

- **trees** – list of NetworkX graphs
- **peaklists** – list of PeakList objects

Returns

Return type Sequence[nx.OrderedDiGraph]

msnpy.processing.**create_templates**(*graphs: Sequence[networkx.classes.ordered.OrderedDiGraph]*, *nh: int*)
Create a 'master' graph that include all the experimental trees Loop through all the subgraphs/graphs

Parameters

- **graphs** –
- **nh** –

Returns

Return type

msnpy.processing.**group_by_template**(*graphs: Sequence[networkx.classes.ordered.OrderedDiGraph]*, *templates: list*)

Parameters

- **graphs** –
- **templates** –

Returns

Return type

msnpy.processing.**group_scans**(*filename: str*, *nh: int = 2*, *min_replicates: int = 1*, *report: Optional[str] = None*, *max_injection_time: Optional[float] = None*, *merge_ms1: bool = False*, *split: bool = False*, *remove: bool = True*)

Parameters

- **filename** – Path to a .raw or .mzML file

  When using .mzML files generated using the Proteowizard tool, SIM-type scans will only be treated as spectra if the 'simAsSpectra' filter was set to true during the conversion process: *msconvert.exe example.raw* **–simAsSpectra** *–64 –zlib –filter "peakPicking true 1-"*
- **nh** – Number of overlapping or matching scan events that should be considered for grouping.
- **min_replicates** – Minimum number of replicate trees required for each group.
- **report** – Path to a tab-delimited text file to which to write a summary of the groups (e.g. scan events, replicates, etc).
- **max_injection_time** –
- **merge_ms1** –
- **split** –
- **remove** –

Returns

msnpy.processing.**hdf5_peaklists_to_txt**(*filename: str*, *path_out: str*, *delimiter: str = '\t'*)

Parameters

- **filename** – Path to an existing HDF5 file
- **path_out** – Path to a new text file
- **delimiter** –

msnpy.processing.**merge_ms1_scans**(*graphs: Sequence[networkx.classes.ordered.OrderedDiGraph]*)

Parameters **graphs** –

Returns

Return type

msnpy.processing.**mz_pair_diff_tol**(*lower_mz: float*, *upper_mz: float*, *tol: float*, *unit: str = 'ppm'*)

msnpy.processing.**mz_tol**(*mz: float*, *tol: float*, *unit: str = 'ppm'*)

Parameters

- **mz** – mz value
- **tol** – tolerance
- **unit** – ppm or da

Returns

Return type float

msnpy.processing.**process_scans**(*filename: str*, *groups: list*, *function_noise: str*, *snr_thres: float*, *ppm: float*, *min_fraction: Optional[float] = None*, *rsd_thres: Optional[float] = None*, *normalise: bool = False*, *ringing_thres: Optional[float] = None*, *exclusion_list: dict = {}*, *report: Optional[str] = None*, *block_size: int = 5000*, *ncpus: Optional[int] = None*)

Parameters

- **filename** – Path to a .raw or .mzML file

  When using .mzML files generated using the Proteowizard tool, SIM-type scans will only be treated as spectra if the 'simAsSpectra' filter was set to true during the conversion process: *msconvert.exe example.raw* **–simAsSpectra** *–64 –zlib –filter "peakPicking true 1-"*
- **groups** –
- **function_noise** – Function to calculate the noise from each scan. The following options are available:
  - **median** - the median of all peak intensities within a given scan is used as the noise value.
  - **mean** - the unweighted mean average of all peak intensities within a given scan is used as the noise value.
  - **mad (Mean Absolute Deviation)** - the noise value is set as the mean of the absolute differences between peak intensities and the mean peak intensity (calculated across all peak intensities within a given scan).
  - **noise_packets** - the noise value is calculated using the proprietary algorithms contained in Thermo Fisher Scientific's msFileReader library. This option should only be applied when you are processing .RAW files.

- **snr_thres** – Peaks with a signal-to-noise ratio (SNR) less-than or equal-to this value will be removed from the output peaklist.
- **ppm** – Maximum tolerated m/z deviation in parts per million.
- **min_fraction** – A numerical value from 0 to 1 that specifies the minimum proportion of scans a given mass spectral peak must be detected in, in order for it to be kept in the output peaklist. Here, scans refers to replicates of the same scan event type, i.e. if set to 0.33, then a peak would need to be detected in at least 1 of the 3 replicates of a given scan event type.
- **rsd_thres** – Relative standard deviation threshold - A numerical value equal-to or greater-than 0. If greater than 0, then peaks whose intensity values have a percent relative standard deviation (otherwise termed the percent coefficient of variation) greater-than this value are excluded from the output peaklist.
- **normalise** – Normalise by Total Ion Current (TIC). Default = False
- **ringing_thres** – Fourier transform-based mass spectra often contain peaks (ringing artefacts) around spectral features (i.e. 1.0 Da) that require removal. This threshold is a positive float indicating the required relative intensity a peak must exceed (with reference to the largest peak in a cluster of peaks) in order to be retained.
- **exclusion_list** – This option allows for specific m/z valuess to be removed, this option may be useful for removing peaks that correspond to artifact and/or system noise peaks.
- **report** – Path to a tab-delimited text file to which to write measures of quality (e.g. RSD, number of peaks, etc) for peaks within each scan/fragmentation event processed in each .RAW or .mzML files.
- **block_size** – number of peaks in each clustering block.
- **ncpus** – number of CPUs for parallel clustering. Default = None, indicating using as many as possible

Returns List of (average) PeakList objects (DIMSpy)

Return type Sequence[PeakList]

2.5 Module contents

CHAPTER

THREE

INDICES AND TABLES

- genindex
- modindex
- search

PYTHON MODULE INDEX

m

INDEX