

# SM finder 1.0

## USER GUIDE

## Contents

Minimum requirements.....	3
Installation .....	3
Menu.....	4
File.....	4
Settings.....	4
Process .....	4
Library .....	5
Untargeted Module .....	7
Target Module.....	9
C-Flux Tab.....	11

## Minimum requirements

N° of CPUs = 4

Ram = 8 Gb

OS = Windows 7, Windows 10

## Installation

Download SMfinder.zip from <https://www.ifom.eu/SMfinder/Download/> .

Unzip the folder and execute install.exe

The process will install Python, R, and Visual C++ Compiler for Python.

During the installation process Python 2.7.15 will be installed. Python can be installed everywhere in the machine, but the folder name cannot be modified (Python27).

During the installation a progress bar will appear. The progress may freeze during the process, however the installation is still running in background. At the end of the installation, a shortcut will appear on the desktop.

SMfinder is now ready to use.

Note: some function as peakpicker in untargeted tab require the installation of XCMS package in R. The installation will start automatically when peakpicker is used for the first time. This may slow down the first analysis. To avoid this issue, we recommend lunching for the first time the peakpicker with only one file selected.

# Menu

## File

**Open raw files:** Using this button, mzXML files can be selected for the analysis.

**Inspect Metafiles:** All generated metafiles are stored in Tables. This subfolder is created in the folder where mzXML files are located. This option allows the selection of one metafile which is displayed in SMfinder. Note that changing in the inspected metafiles are not saved as this option is for visual inspection only.

**Exit:** This button leads to the exit popup. In the popup the user can to overwrite setup parameters.

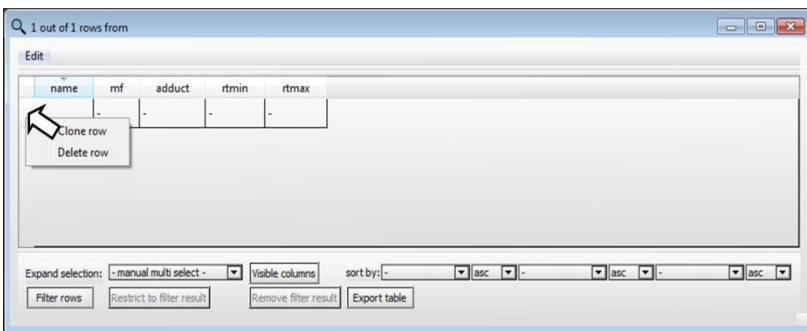
## Settings

**Import log:** This button opens a dialog where a log file in txt can be selected. The SMfinder settings are then updated accordingly.

**Export log:** SMfinder settings are exported in a text file.

## Process

**Generate MZ align table:** By clicking this option, a popup will open in SMfinder. A list of compound can be inserted providing name, molecular formula (mf), adduct and retention time window (rtmin and rtmax). Multiple compound can be added by right clicking on a row border and select Clone row. Adducts must be added



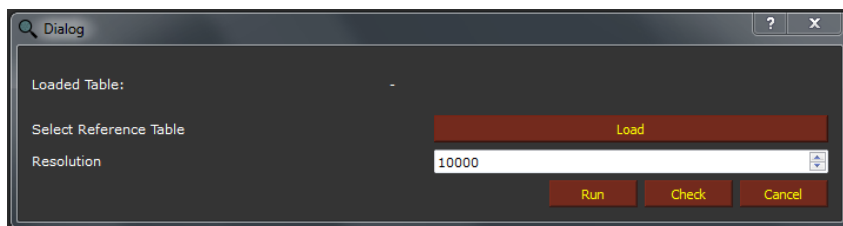
only for positive or negative mode i.e. mixed mode is not allowed and lead to error. Allowed adduct must be inserted with the following syntax with only one adduct per row:

Positive adducts: M+H, M+NH<sub>4</sub>, M+Na, M+H-2H<sub>2</sub>O, M+H-H<sub>2</sub>O, M+K, M+ACN+H, M+ACN+Na, M+2Na-H, M+Li, M+CH<sub>3</sub>OH+H

Negative adducts: M-H, M-H<sub>2</sub>O-H, M+Na-2H, M+Cl, M+K-2H, M+KCl-H, M+FA-H, M+CH<sub>3</sub>COO, M+F

Note: This table is required for MZ alignment which required a minimum of 3 compounds in order to compute the alignment.

**Perform MZ align:** This option will open a popup asking for reference table and resolution. By clicking load, a reference table can be selected and



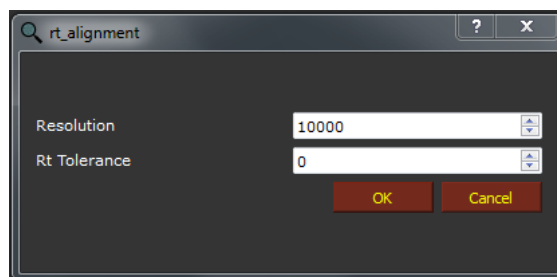
will be displayed on Loaded Table. The second value is the instrumental resolution which must be provided by the user according to the instrumental condition in which the chromatograms were acquired. By clicking on Run, the mass alignment will be computed on all the selected raw files (see open raw files). This process overwrites the original raw chromatogram with a corrected chromatogram.

With the check button, SMfinder will open a new window showing the extracted ion chromatogram generated from the given information. This option is allowing only inspection therefore changes will not be saved.

Note: The needed reference has a special format which is generated by Generate MZ alignment table. Tables generated with other formats are now allowed and lead to error.

**Perform RT align:** this option will open the following popup. The required information are the instrumental resolution, and the expected retention time shift in seconds.

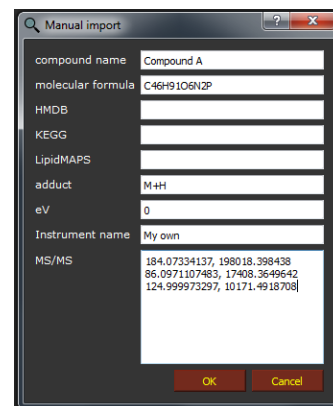
Note: As for MZ alignment, also this process overwrite the original raw file.



## Library

**Manual import:** This button will open a popup where all the information regarding a compound can be inserted. Compound name, molecular formula, adduct, eV, Instrument name and MS/MS are mandatory values.

MS/MS should contain at least three fragments with their intensities. MS/MS should be inserted by adding in rows fragments and intensities separated by comma starting from the most abundant fragment. By clicking OK, the inserted compound will be added in the in-house library.



**Import from txt file:** With this option compound can be added from a txt files which can be generated from target analysis or by exporting the in-house library.

Note: This option will delete identical entries, therefore importing multiple times the same library will not lead to changes in the in house database.

**Export inhouse library:** This option generate a txt file from the in-house library. It can be use for share the library with other users as well as importing the same in-house library from different computers.

**Edit inhouse library:** This option will open the in-house library in SMfinder. The library can be modified by deleting entries or editing values such as library IDs, name etc. After closing the library windows a popup will ask for overriding options.

**Update main library:** SMfinder has a second library which include the available MS/MS spectra from HMDB in SMfinder format. Updates of the HMDB will be released on the SMfinder webserver and can be used for update the main library. This options has no effect on the in-house library.

# Untargeted Module

**Peakpicker:** With this button, the feature detection for the selected chromatograms (see Open Raw files) will be performed based on the given resolution. If Deiso is checked, isotopes will be removed from the detected features.

**Unique ID:** This function requires metafiles which are generated by the peakpicker. Similar to Rt align, Unique ID performs a second alignment for all the features between samples based on the given retention time. Finally all the IDs between samples are overwritten based on the ID match. Note: Compare with RT align, this function does not modify detected retention times or original raw data.

**MS1:** With this function, SMfinder will assign to each feature with a chemical formula if any, matching the given parameters (ppm and adducts) and excluding halogenated matches if the option Exclude halogenated formulas is checked. The select adducts button gives the option of selecting one or multiple adducts types in positive or negative polarity. If no adduct is given, SMfinder will set as default M+H and M-H respectively for chromatogram acquired in positive or negative mode. MS1 requires the metafile generated by peakpicker. This function retains the IDs assigned in the original metafile, therefore although not mandatory, it is highly recommended to perform Unique ID before MS1.

**MS2:** With this function, features with formulas assigned by the MS1 are evaluated at MS<sup>2</sup> level by comparing empirical MS<sup>2</sup> spectra from the chromatograms with the MS<sup>2</sup> spectra available in the two libraries (main and in-house). Different options are available and determine how the library is matched. Canonically, matches between samples and library will have the same chemical formula, adduct, and collision energy (eV). Moreover, the library will be filtered based on the selected instrument. List of all available instruments between the two libraries can be selected from the dropdown list. If Use blind library is checked, eV and instrument are excluded from the filtering, while force library association will exclude the type of adduct in the

The screenshot displays the 'Untargeted' module interface with the following settings and buttons:

- Resolution:** 30000
- ☒ **Deiso**
- Peakpicker** button
- Rt Tolerance:** 120
- Unique ID** button
- PPM:** 10
- ☒ **Exclude halogenated formulas**
- Select Adducts** button
- MS1** button
- Electronvolt:** 20
- Max Perm.Num:** 500
- Instrument:** LC-ESI-IT (LC/MSD Trap XC)
- ☐ **force library association** ☒ **Use blind library**
- MS2** button
- Hierarchy:** ms2 fdr + ppm
- Filter** button
- Inspect** button
- Minimum Count:** 3
- Filler** button
- ☒ **Overlay**
- Manual Integration** button
- Starting Point:** MS1
- Automate** button
- Export Results** button

matching criteria. The maximum number of permutations is used for calculating the FDR.

**Filter:** Data filtering can be performed from the metafiles generated either with MS1 or MS2. In case of MS1 filtering only ppm or iso are allowed while all the other filtering options apply on MS2 metafiles. Filtering is performed within the entire dataset and the best match for each feature will be selected for all the given samples. This allows the association and homogenization of features which were unassigned e.g. because MS2 scans were not present in one or more samples.

**Inspect:** With this option, samples are aligned based on their feature ID and plotted as overlay. Changes during inspection are not saved. If modifications are needed, Manual Integration can be used instead.

**Filler:** This option allows the search of missing features between batches and requires metadata generated from the function Filter. All the detected features are clustered and a minimum count filter is applied. The minimum count indicates the number in which the molecule was detected. All the compounds below the given minimum count are excluded from the analysis.

**Manual integration:** This option metadata generated by Filter function can be manually reintegrated. Modifications are automatically stored at the end of the editing. If Overlay is checked, all the samples in the selected batch are plotted in a unique table, otherwise each sample will appear after that the preview has been closed.

**Automate:** With this button, the core process in the untargeted module will be performed automatically in the following order: Peakpicker, Unique ID, MS1, MS2, Filter, Filler. The dropdown bar defines the starting point e.g. when MS2 is selected as a starting point, Automate will perform MS2, Filter and Filler option only.

**Export results:** By clicking on export results, all the samples are aligned based on their ID. Information regarding identification and peak area are then extracted and stored as excel file.

The screenshot displays a software interface with a dark grey background and a red border. At the top, there are three tabs: 'Untargeted' (selected), 'Target', and 'C-Flux'. The interface is organized into several sections, each with a label and a corresponding button or dropdown menu. The sections include: 'Resolution' (30000) with a 'Peakpicker' button; 'Rt Tolerance' (120) with a 'Unique ID' button; 'PPM' (10) with a 'Select Adducts' button and an 'MS1' button; 'Electronvolt' (20) and 'Max Perm.Num' (500) with an 'MS2' button; 'Instrument' (LC-ESI-IT (LC/MSD Trap XC)) with checkboxes for 'force library association' and 'Use blind library'; 'Hierarchy' (ms2 fdr + ppm) with 'Filter' and 'Inspect' buttons; 'Minimum Count' (3) with a 'Filler' button; 'Overlay' (checked) with a 'Manual Integration' button; 'Starting Point' (MS1) with an 'Automate' button; and an 'Export Results' button at the bottom. The buttons are red with yellow text, and the dropdown menus are white with black text.



# Target Module

**Load reference Table:** This button asks for the file containing the list of compound for target analysis.

**Edit:** Open the reference table for editing.

**Import from CSV:** generate a reference table from CSV file. In csv the first row must contains the following column names: name, mf, adduct, rt, eV, instrument and can have optional names as id, KEGG, HMDB and LipidMAPS. Information are filled based on the relative column.

**Generate:** Open an empty reference table. It is important to change the ID of each compound inserted. Compounds with the same ID will generate cloned columns.

**Check retention time:** This button generate an extracted ion chromatogram using the first selected file and the reference table. Changes in the retention time are saved in the reference table.

**Perform target analysis:** This button compute the extraction and integration of the selected compounds.

**Perform MS1 and MS2 evaluation:** This button compute the calculation for ppm error, isotopic similarity on MS level, and MS2 score and FDR on MS2 levels returning the best match in the results.

**Inspect:** With this options, samples are aligned based on their feature id and plotted as overlay. Changes during inspection are not saved. If modifications are needed, Manual Integration can be used instead.

**Manual Integration:** This option metadata generated by Filter function can be manually reintegrated. Modifications are automatically stored at the end of the editing. If Overlay is checked, all the samples in the selected batch are plotted in a unique table, otherwise each sample will appear after that the preview has been closed.

The screenshot displays the 'Target' module interface of a software application. At the top, there are three tabs: 'Untargeted', 'Target' (which is selected), and 'C-Flux'. Below the tabs, the interface is organized into several sections. The first section, labeled 'Loaded File:', contains a 'Load reference table' button and two smaller buttons, 'Edit' and 'Import from CSV', followed by a 'Generate' button. The second section features a 'Resolution' input field set to '20000' and a 'Check retention time' button. The third section has a 'Peak Width' input field set to '120' and a 'Perform target analysis' button. The fourth section includes three input fields: 'Electronvolt' set to '0', 'Max Perm. Num' set to '1000', and 'Instrument' set to 'EI-B (HITACHI RMU-6L)'. Below these fields are two checkboxes: 'force library association' (unchecked) and 'Use blind library' (checked). This section also contains a 'Perform MS1 and MS2 evaluation' button. The fifth section has an 'Inspect' button. The sixth section includes an 'Overlay' checkbox (unchecked) and a 'Manual Integration' button. The final section at the bottom contains two buttons: 'Export Results' and 'Export txt for library'. The entire interface is set against a dark grey background with red borders around the main content area.

**Export Results:** By clicking on export results, all the samples are aligned based on their ID. Information regarding identification and peak area are then extracted and stored as excel file.

**Export txt for library:** With this option, SMfinder search in the entire batch for the most abundant MS<sup>2</sup> spectra for each compound. After filtering a text file is generated containing with an header for each compound starting with the symbol ">" and a list of found MS<sup>2</sup> masses order by intensity. This txt file is already in a suitable format for been imported in the in-house library.

UntargetedTargetC-Flux

Loaded File:

Load reference table

EditImport from CSVGenerate

Resolution20000

Check retention time

Peak Width120

Perform target analysis

Electronvolt0

Max Perm. Num1000

InstrumentEI-B (HITACHI RMU-6L)

☐ force library association☒ Use blind library

Perform MS1 and MS2 evaluation

Inspect

☐ Overlay

Manual Integration

Export Results

Export txt for library

## C-Flux Tab

**Load reference table:** This button asks for the file containing the list of compound for <sup>13</sup>C-flux analysis. Reference table can be generated with the dropdown menu by pushing the button Go. Three different inputs are accepted in the mode Target, Untargeted and Reference Table accepting respectively, raw mzXML files analyzed with target or untargeted mode or a table generated for target analysis. After clicking, a list of compound is extracted from the source and isotopologues are automatically generated.

**Check retention time:** This button generate an extracted ion chromatogram using the first selected file and the reference table. Changes in the retention time are saved in the reference table.

**Mine <sup>13</sup>C Loading:** This button compute the extraction and integration of the selected compounds and their isotopologues and calculated the loading as percentage for each compound.

**Inspect:** With this options, samples are aligned based on their feature id and plotted as overlay. Changes during inspection are not saved. If modifications are needed, Manual Integration can be used instead.

**Manual integration:** This option metadata generated by Filter function can be manually reintegrated. Modifications are automatically stored at the end of the editing. If Overlay is checked, all the samples in the selected batch are plotted in a unique table, otherwise each sample will appear after that the preview has been closed.

**Export Results:** By clicking on export results, all the samples are aligned based on their ID and isotopologues. The percentage of loading is extracted for all the compounds.

The screenshot displays the 'C-Flux' software interface with the 'Untargeted' tab selected. The interface features a dark grey background with red borders. At the top, there are three tabs: 'Untargeted', 'Target', and 'C-Flux'. Below the tabs, the 'Loaded File:' section is empty. A 'Load reference table' button is present. The 'Generate from:' dropdown menu is set to 'Target', with a 'Go' button next to it. The 'Resolution' is set to '30000'. Below these, there are buttons for 'Check retention time', 'Mine <sup>13</sup>C Loading', 'Inspect', and 'Manual Integration'. An 'Overlay' checkbox is checked. At the bottom, there is an 'Export Results' button.