How to use Scannotation (v 3.5)

**Table of contents**

[1. Presenting Scannotation 2](#_Toc137737540)

[2. Step-by-step tutorial 4](#_Toc137737541)

[2.1. Installation 4](#_Toc137737542)

[2.2. Library 6](#_Toc137737543)

[2.2.1. Template 6](#_Toc137737544)

[2.2.2. Library tab 6](#_Toc137737545)

[2.2.3. Generating only m/z data 8](#_Toc137737546)

[2.2.4. Adding isotopic pattern data 10](#_Toc137737547)

[2.2.5. Adding Rt prediction from logP values 12](#_Toc137737548)

[2.2.6. Adding metabolites 14](#_Toc137737549)

[2.2.7. Adding data to an existing library 15](#_Toc137737550)

[2.3. Screening 16](#_Toc137737551)

[2.3.1. Templates 16](#_Toc137737552)

[2.3.2. Screening tab 17](#_Toc137737553)

[2.3.3. Dataset 17](#_Toc137737554)

[2.3.4. Chromatographic parameters 20](#_Toc137737555)

[2.3.5. Neutral loss 22](#_Toc137737556)

[2.3.6. Results 22](#_Toc137737557)

[3. FAQ 28](#_Toc137737558)

[3.1. Library 28](#_Toc137737559)

[3.2. Screening 28](#_Toc137737560)

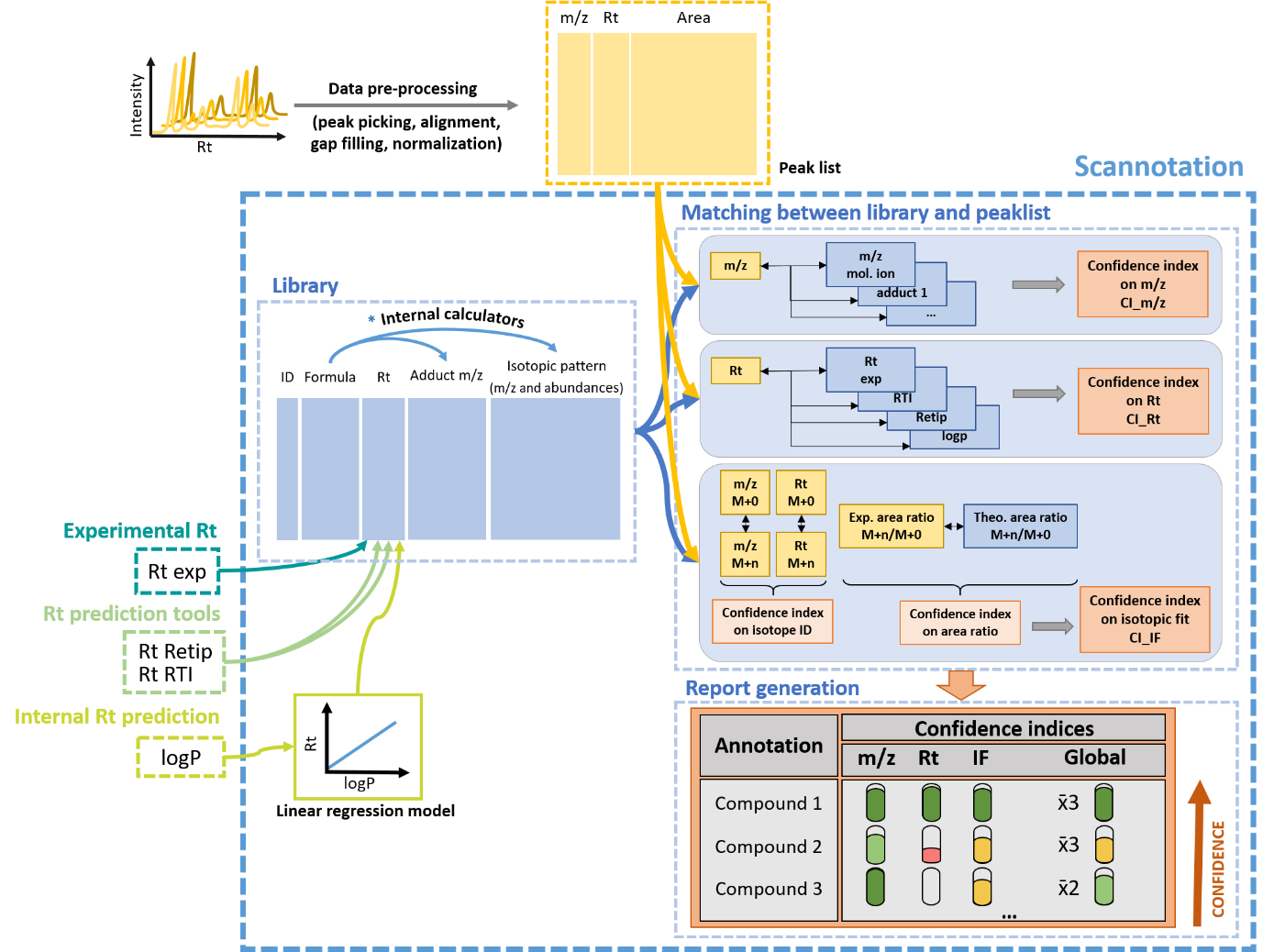
[3.2.1. Dataset 28](#_Toc137737561)

[3.2.2. Chromatographic parameters 28](#_Toc137737562)

[4. Screening results – Column names 29](#_Toc137737563)

# Presenting Scannotation

Scannotation is a Python-based program aimed at assisting LC-HRMS-based suspect screening approaches by establishing prioritized lists of scored pre-annotations from a user-provided library and a list of experimental features. Scores, called Confidence Indices (CI), are computed by comparing experimental features and suspects using MS1 chemical predictors: mass-to-charge ratios (m/z), retention time (Rt), and isotopic pattern. The detection of phase II metabolites and associated neutral loss is also accounted for in the final prioritization of scored pre-annotations. Its general operating principle is presented in the figure below.



Scannotation’s annotation workflow relies on comparing a user-built library to a list of features from a given dataset obtained from any vendor or open source/open access pre-processing software tools. Compounds’ identifiers (name and SMILES), molecular formula, experimental and predicted retention time (Rt), as well as logP values can be provided by the user. Scannotation then computes molecular ion and adduct masses, theoretical isotopic pattern, and a logP-predicted Rt. All data are compiled into an Excel file generated automatically, which means that Scannotation can also be used as a batch isotope calculator (up to several thousands in a few minutes depending on the PC performance). The software successively compares experimental features to the suspect library data for three predictors: m/z, Rt and isotopic fit. Neutral loss patterns for common phase II metabolites can also be investigated using Scannotation. When the user uploads a feature list, scores are generated for each predictor, and combined into a global score (Global CI). The final output generated automatically by Scannotation is a detailed Excel report that includes all scores, frequencies of detection of pre-annotated features, as explained in sections 2.3.5 and 4.

Scannotation’s operating principle is further described in the following publications:

* Chaker J, Gilles E, Léger T, et al. From Metabolomics to HRMS-Based Exposomics: Adapting Peak Picking and Developing Scoring for MS1 Suspect Screening. Analytical Chemistry 2021;93(3):1792–1800; doi: 10.1021/acs.analchem.0c04660.
* Chaker J, Gilles E, Monfort C, et al. Scannotation: a suspect screening tool for the rapid pre-annotation of the human LC-HRMS-based chemical exposome. Environmental Science and Technology 2023 [submitted].

# Step-by-step tutorial

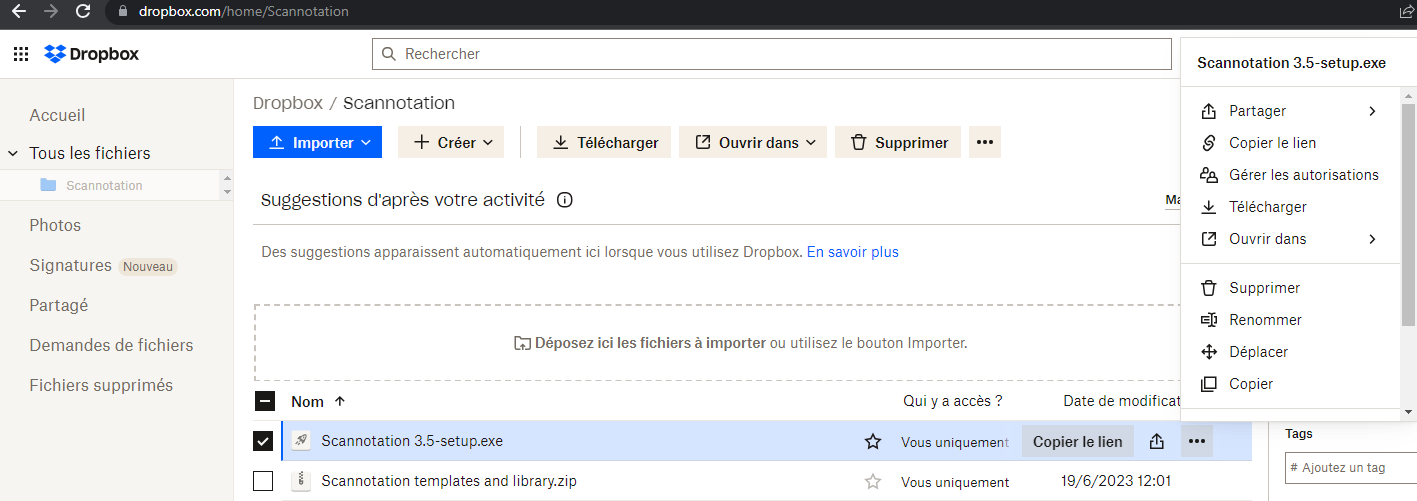
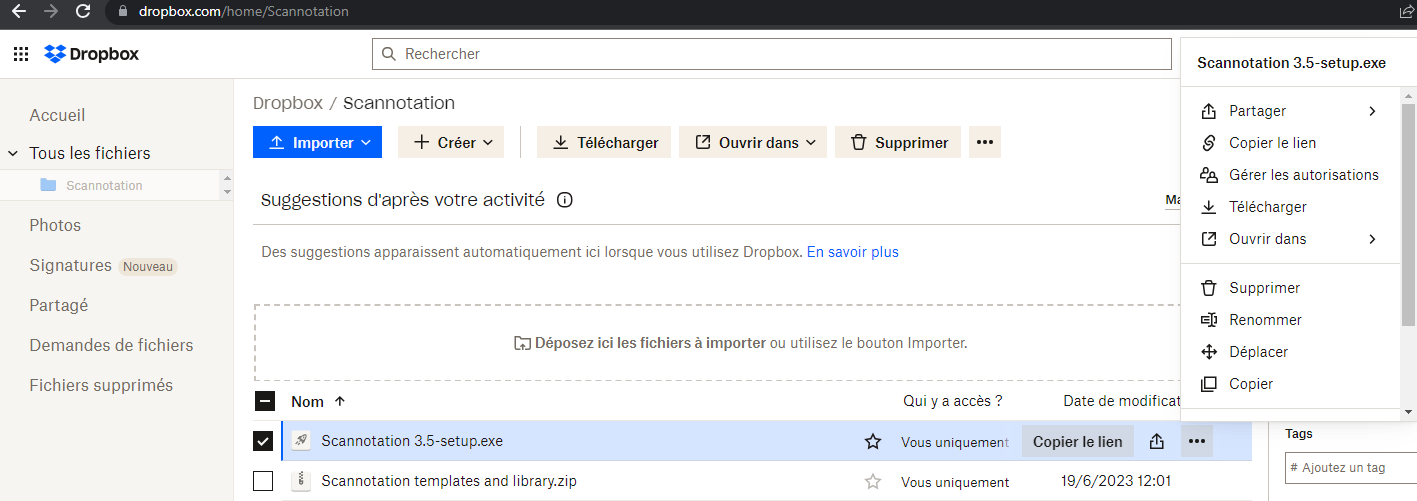
Are you ready? This is a step-by-step tutorial to get started with Scannotation. The document describes how to launch its main processes and presents the most common warning messages. Additional detailed explanations for all warning and error messages can be found in section 5. Similarly, responses to FAQs can be found in section 3.

## Installation

First, please download the Scannotation software and the .zip folder (containing templates) from its Dropbox repository, which you will find at:

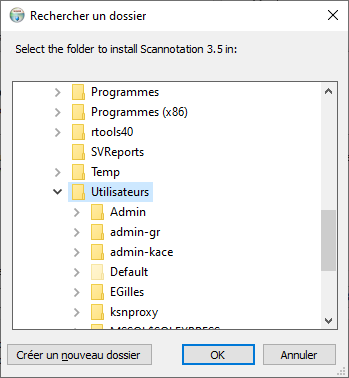
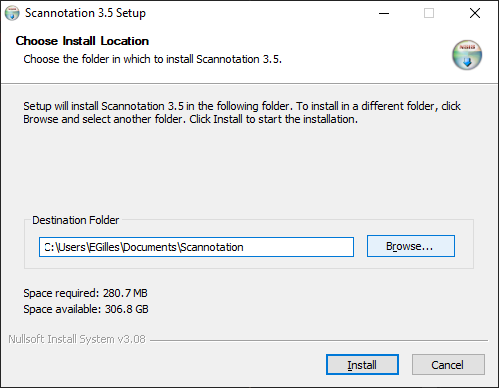
https://www.dropbox.com/sh/6ueh95xvfmaez47/AABTX2EaPV2rYS3Ck4rvu-MPa?dl=0

Mouse Cursor Arow Fixed | Free SVGMouse Cursor Arow Fixed | Free SVG

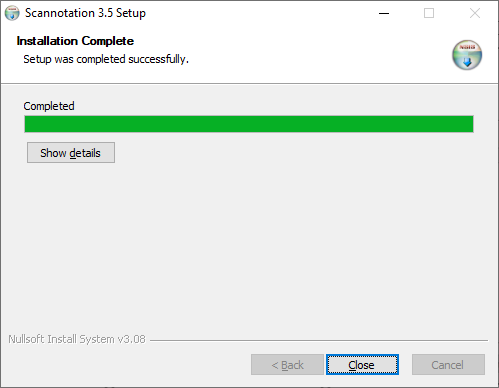
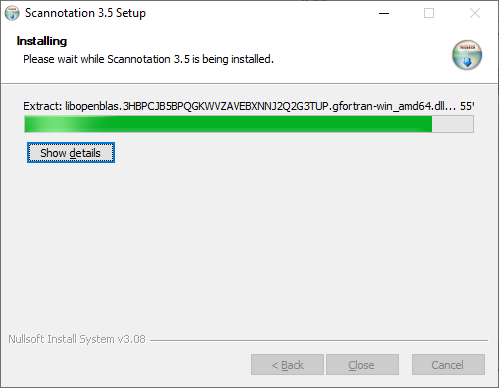


Scannotation’s development was done on Windows 10. There is no minimum requirement, as Scannotation is frequently run on average modern laptops. We recommend a more powerful desktop system to decrease computing time.

Then, choose the installation directory. Make sure you have administrative rights on the folder in which you are installing the software.



Follow the instructions until install is complete.



## Library

### Template

The template for the library file needed for Scannotation can be found in the zipped file (*Scannotation templates and library* > *Scannotation library\_template.xlsx*). It contains seven columns with the headers and contents described below.

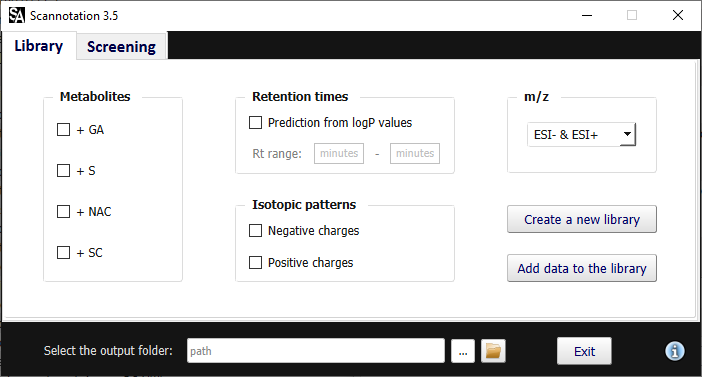
|  |  |
| --- | --- |
| **Column header** | **Detailed description** |
| **Molecule** | Compound name |
| **SMILES** | Canonical SMILES identifier (must be unique!) |
| **Formula** | Molecular formula |
| **Rt\_exp** | Experimental retention time (in minutes) |
| **Rt\_RTI** | Retention time predicted by RTI (in minutes)  **(Aalizadeh et al., 2016. DOI: 10.1021/acs.jcim.5b00752)** |
| **Rt\_Retip** | Retention time predicted by Retip (in minutes)  **(Bonini et al., 2020. DOI: 10.1021/acs.analchem.9b05765)** |
| **logP  OR  Rt\_logP** | logP value (preferably experimental, otherwise predicted) **OR** Retention time predicted by a logP-based regression, in case the user doesn't wish to use Scannotation's Rt prediction tool |

The columns “Molecule” and “Formula” **must be filled** for every compound of the library. We **strongly recommend** filling the other columns to limit redundancies in your suspect list (“SMILES” column) and to increase the confidence in the suggested annotations (Rt and logP-related columns).

### Library tab

The “Library” tab is where your suspect list’s data may be generated and formatted for the screening step. There are four main sections in this tab: “Metabolites”, “Retention times”, “Isotopic patterns”, and “m/z”. This tutorial will start with presenting the “m/z” section, which is the only mandatory step (although performing all steps is **strongly recommended** to increase confidence in your annotations). The tutorial will then explain how to generate the “Metabolites”, “Retention times”, “Isotopic patterns” data into the user’s library.

Overall, if all of the sections are active/used, and depending on the length of your suspect list, the library generation process may take anywhere from a few seconds to 10-20 minutes for approximately 10,000 compounds (the isotope calculation part being the most time consuming). Progress bars will appear along the process to let you know whether everything is working as intended.

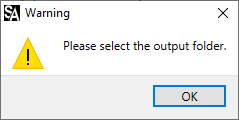
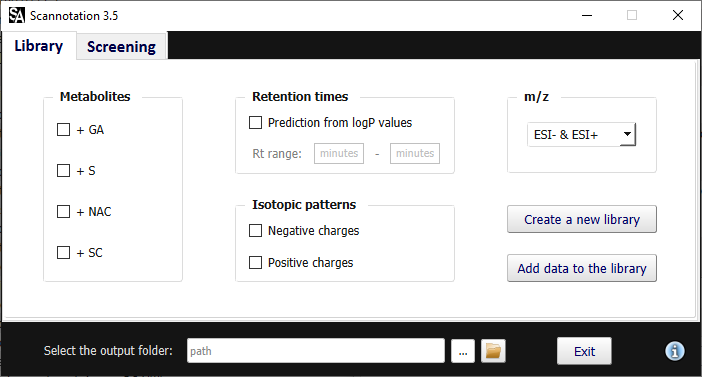


### Generating only m/z data

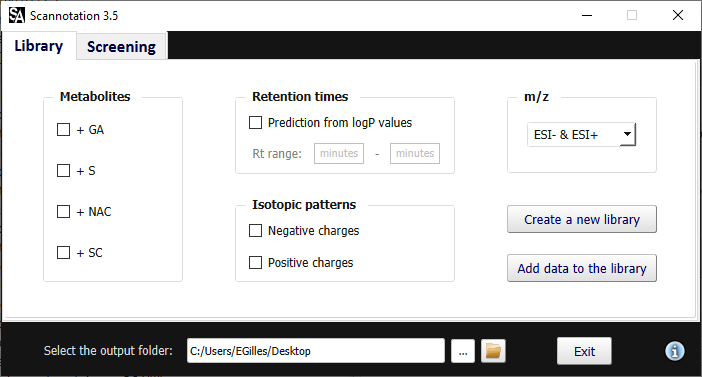
To generate m/z values in a given ionization mode for your library compounds, select it in the “m/z” scrolling menu as follows:



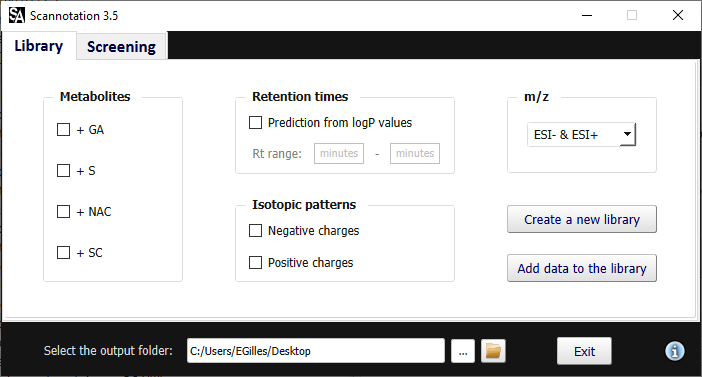
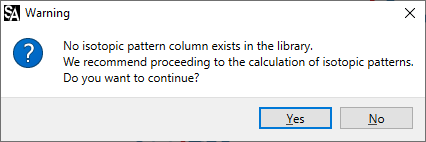
Then, select the output folder in the field at the bottom of your screen (or use Ctrl+O):



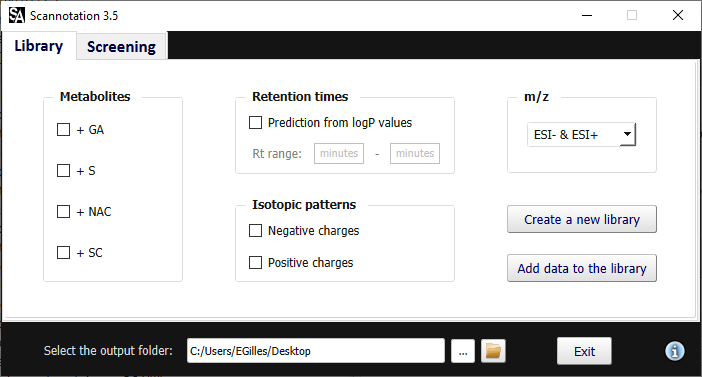
Then, select “Create a new library” to compute m/z values:

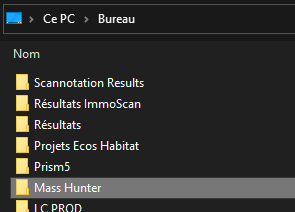
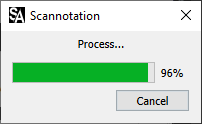
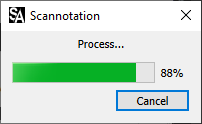


If no isotopic pattern-related columns are detected in your file, Scannotation will suggest aborting the process to select isotopic pattern calculation. You can ignore the message by agreeing to continue.

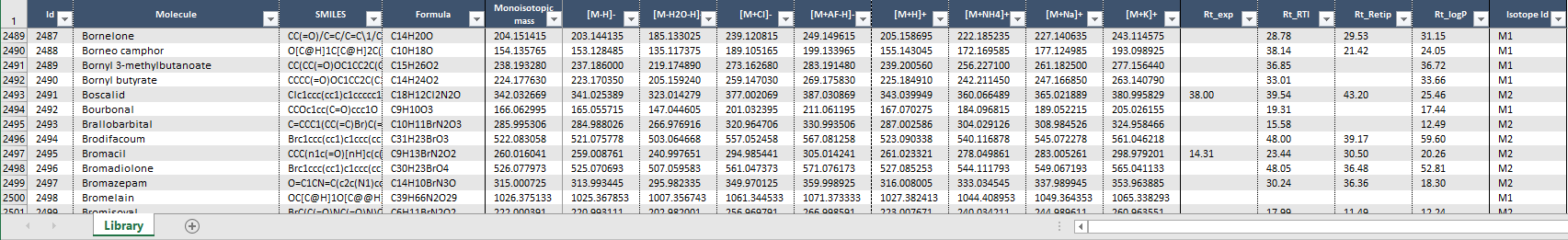


The m/z calculation will then be performed and may be accessible by clicking on the folder icon at the bottom of the graphical interface.

Capture d’écranMouse Cursor Arow Fixed | Free SVG



The output can be found using Ctrl+N, or by navigating to the following directory: *Scannotation Results* > *[Date]* > *01\_Library* > *[Library name]\_Completed.xlsx*. The resulting file contains the following information (depending on the chosen ionization mode in the “m/z” scrolling menu).



ESI+ adducts m/z

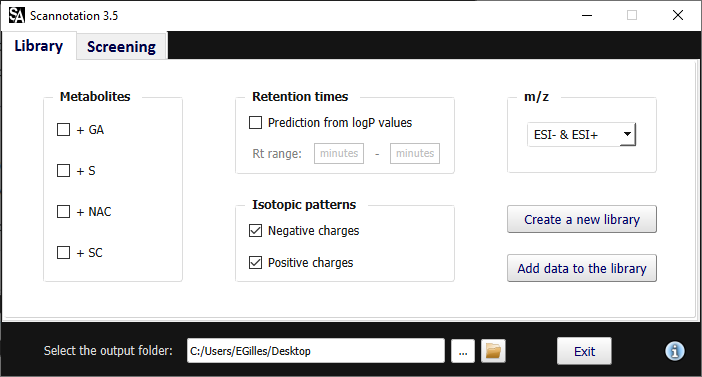
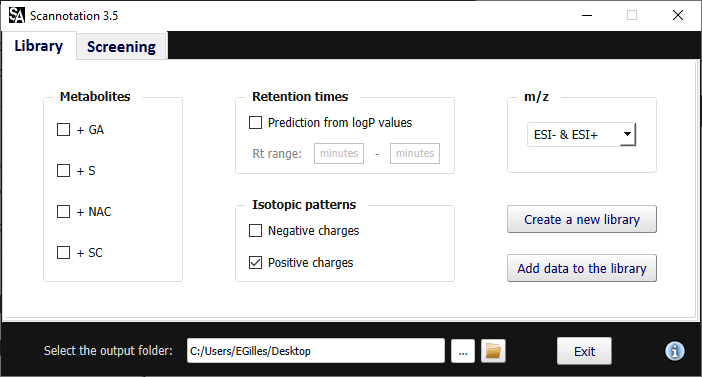
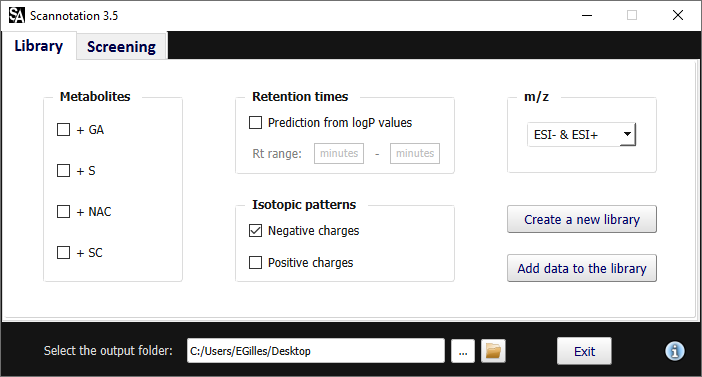
ESI- adducts m/z

monoisotopic

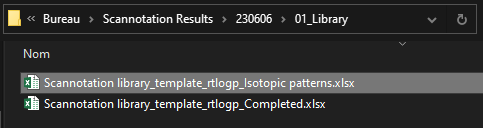
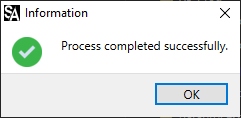
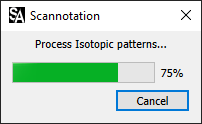
m/z

### Adding isotopic pattern data

To add isotopic pattern calculation to the output, tick one or both of the options in the “Isotopic patterns” section.

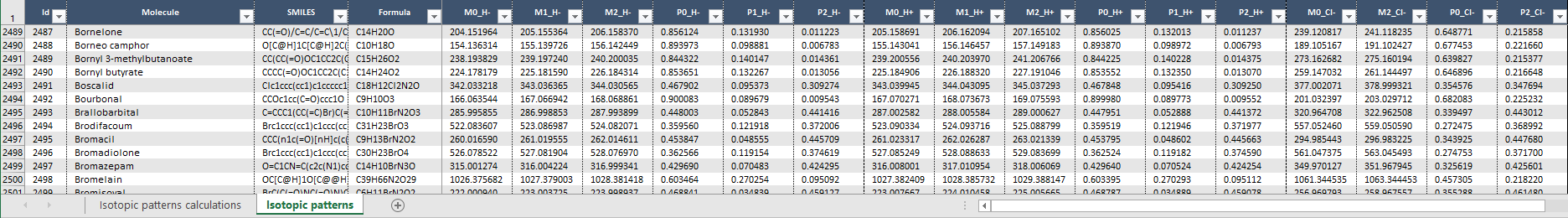
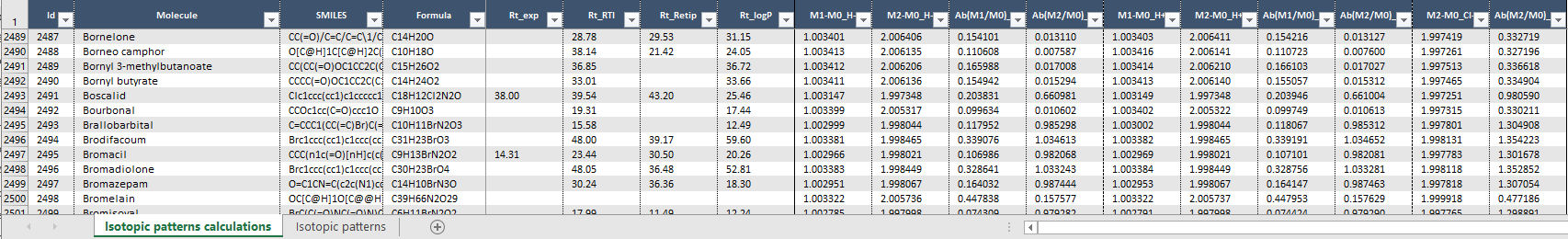


This will create two other output files titled “*[Library name]\_Isotopic patterns.xlsx*” and “*[Library name]\_Completed.xlsx*”.

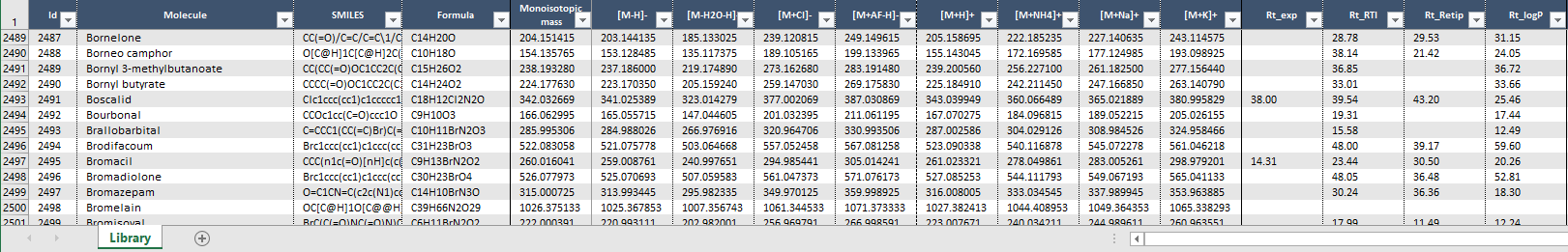


It should be noted that this first file can be reimported as is in the library (i.e., to perform metabolites and/or m/z generation) to avoid recalculating isotopic patterns, as it is the most time-consuming step of the library formatting.

This file has two tabs. The first tab contains all of the transformed data (i.e., mass differences and relative area ratios) used by Scannotation in the screening step, whereas the second tab contains raw data (i.e., theoretical masses and relative abundances of compounds’ isotopologues).



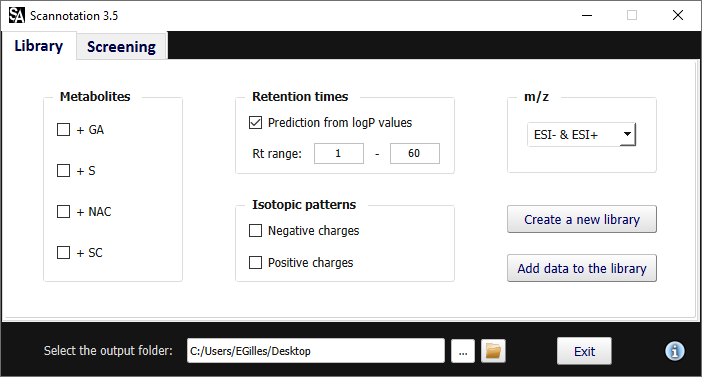
In the second file, the “Isotope Id” column refers to the isotopologue considered for the screening step. Compounds containing Cl, Br or S are scored based on their M+2 isotopologue (as a distinctive pattern is expected), whereas other compounds are scored based on their M+1 isotopologue.



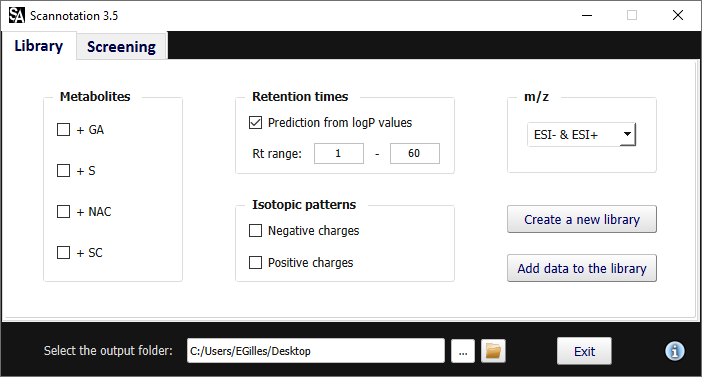
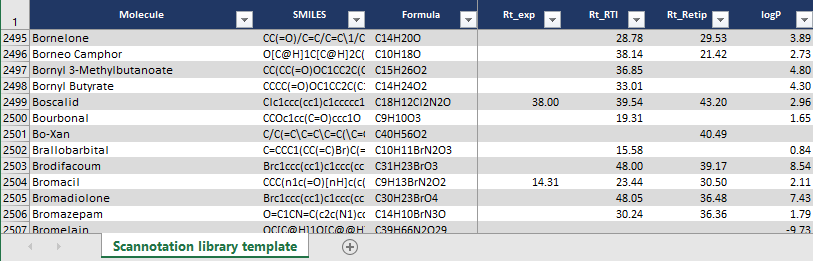
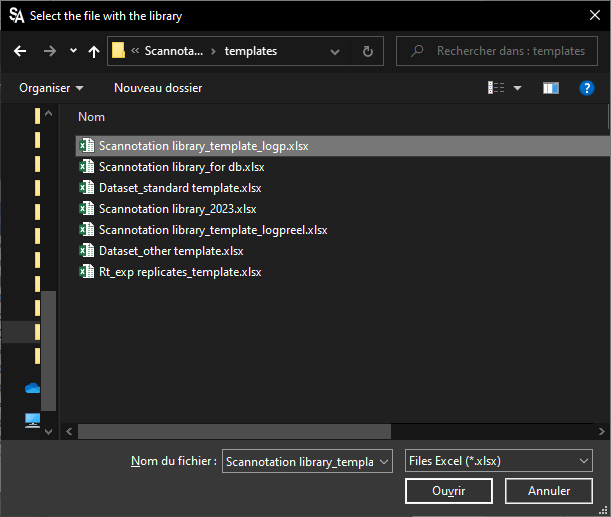
### Adding Rt prediction from logP values

Rt values can be predicted from logP values by checking the corresponding box in the “Retention times” section. The Rt range in which you wish to predict Rt values should also be specified.

Note that this range can be narrower or wider than your chromatographic method. However, the further you are from your training set (i.e., the experimental Rt values available in your library), the less accurate the prediction.

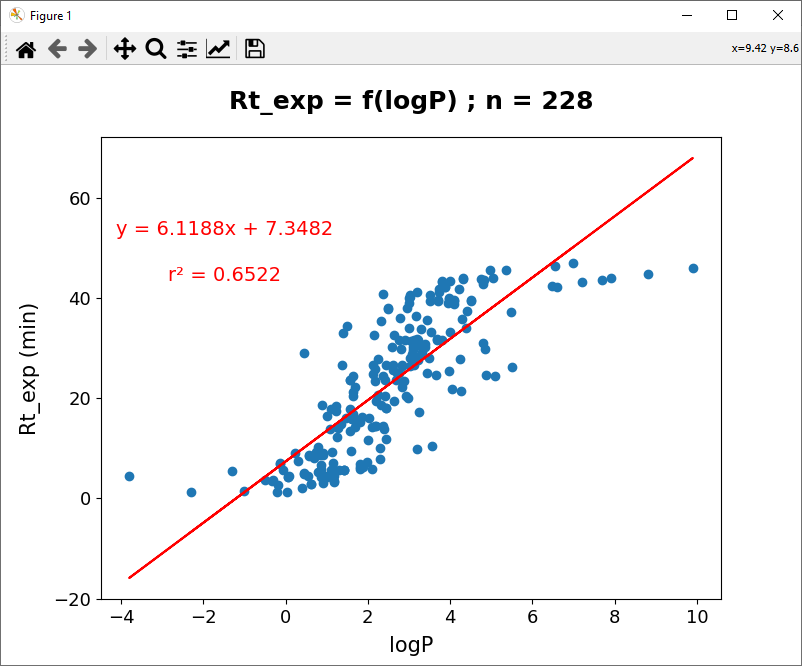


Please also note that if you wish to perform Rt prediction, you need to have a “logP” column in your file in place of an “Rt\_logP” column.

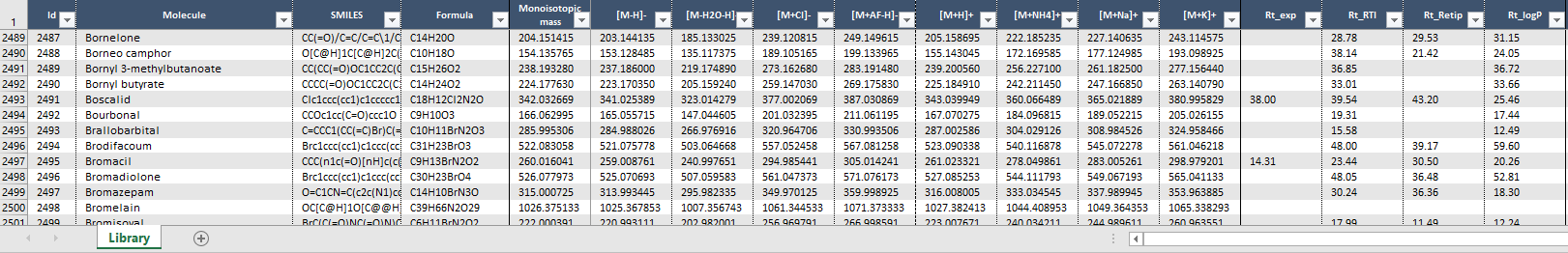
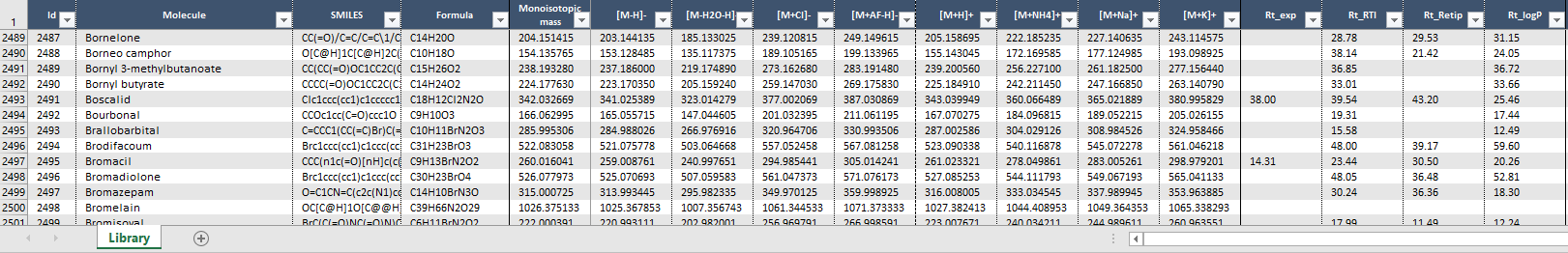


The graphical Rt prediction output can be found in *Scannotation Results* > *[Date]* > *01\_Library* > *[Library name]\_logp\_Rt\_exp = f(logP).png*. A linear regression is plotted, with its equation and r² value. It should be noted that at least **20 experimental Rt values are required** to create a model, although **we recommend more than 100 values** to increase the model accuracy. If the linear regression’s r² is under 0.6, no Rt prediction will be performed, as the model will be considered too inaccurate to help during the suspect screening step (this can be sometimes explained by the lack of coverage of the chromatographic space for the chemical selected).

Moreover, if a few or no logP values are entered for your library compounds, a logP value will be predicted using an elemental composition-based model. However, we recommend providing experimental logP values, if possible, as it will help increase the model’s accuracy.



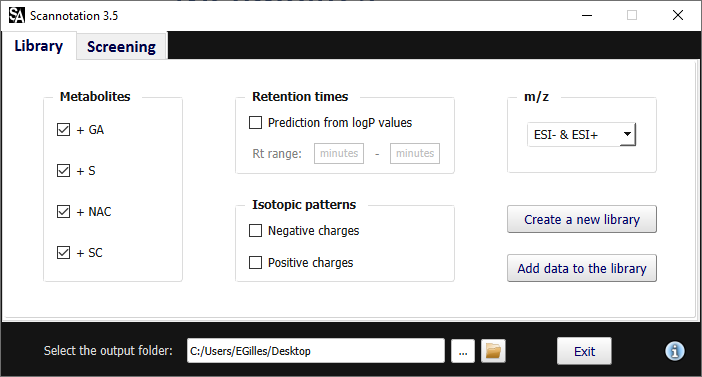
On the generated “*[Library name]\_Completed.xlsx*” file, Rt prediction is now reported in addition to isotopic and m/z data.



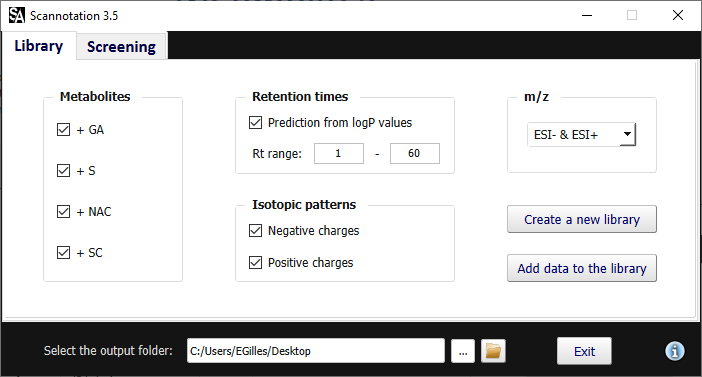
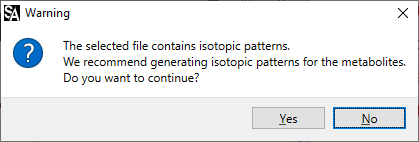
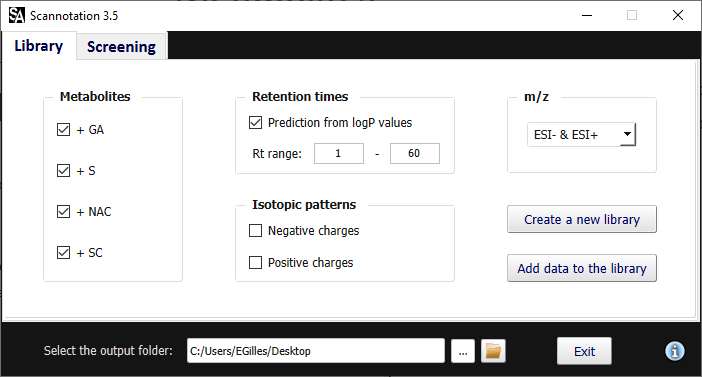
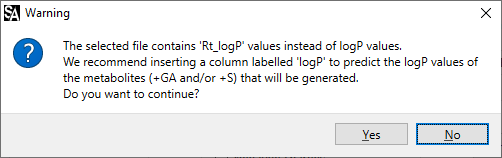
[…]

### Adding metabolites

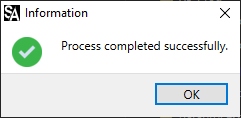
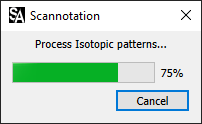
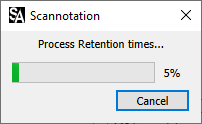
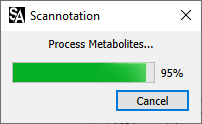
Scannotation may add suspects to your suspect list by generating metabolites of the compounds in your library. Four types of metabolites are currently supported: glucuronic acid (+GA, computed if presence of O or N in the molecular formula), sulfate (+S, computed if presence of O in the molecular formula), n-acetyl cysteine (+NAC, always computed), and cysteine (+SC, always computed). Choose one or more by ticking the corresponding boxes in the “Metabolites” section, depending on your research interest.



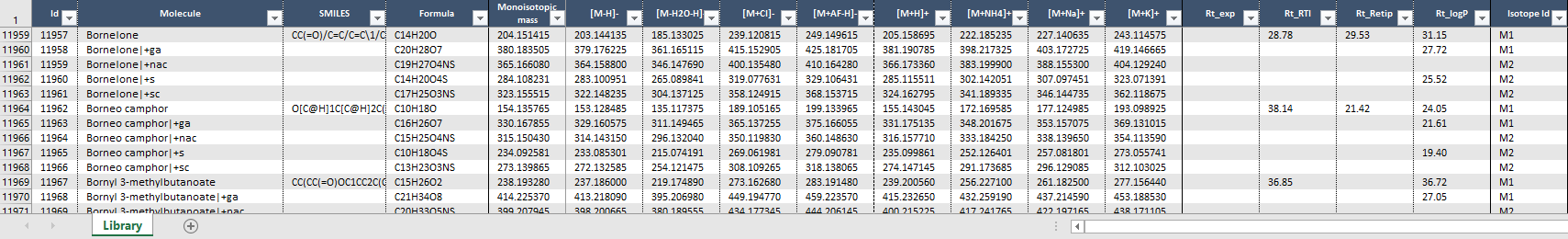
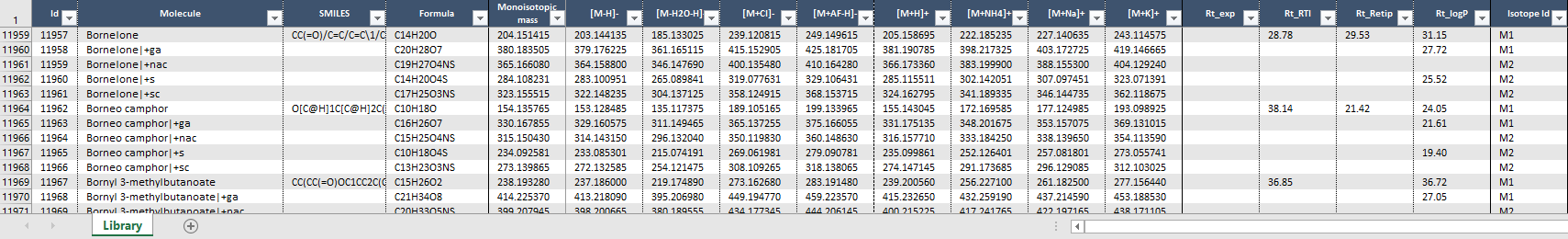
We recommend enabling the prediction of Rt from logP values, as Scannotation can predict logP values for your metabolites, and then a corresponding Rt value. If you choose to do so, please ensure that your file contains a “logP” column instead of an “Rt\_logP” column.



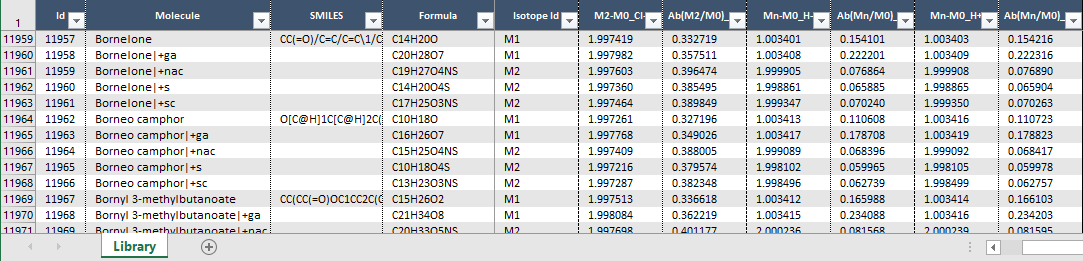
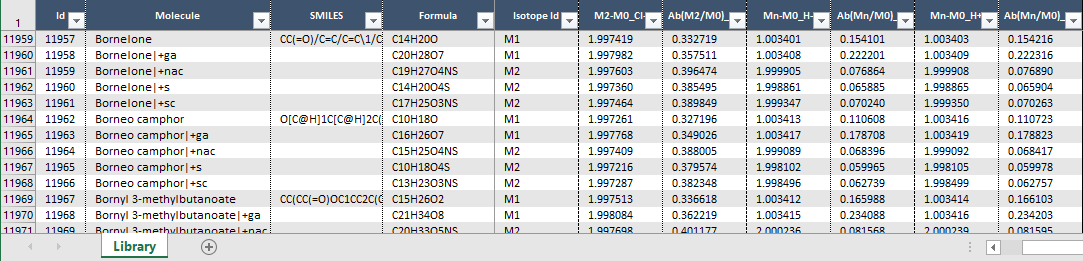
Combined with isotopic pattern and m/z generation, a complete library can be created. Progress bars will appear along the formatting and calculation steps to verify that everything is proceeding as expected.



In the generated “*[Library name]\_Completed.xlsx*” file, metabolites are now reported in addition to Rt prediction, isotopic and m/z data.

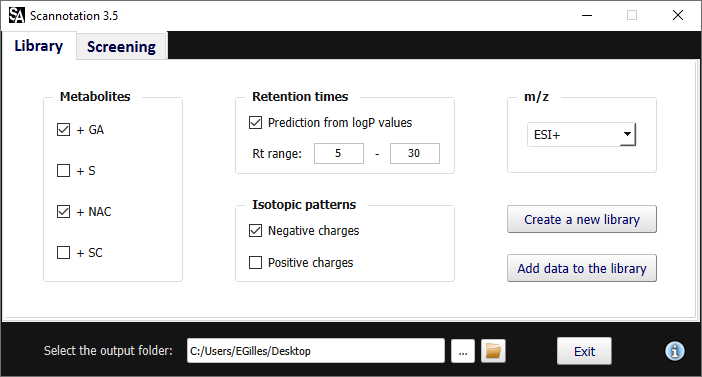
 

[…]



### Adding data to an existing library

If you have a processed library to which you wish to add a few compounds, you can choose to select “Add data to the library” instead of “Create a new library”. You can process these new compounds differently to the data existing in the library (e.g., different ionization modes for isotopic patterns and/or m/z, etc.).



However, if you choose to do so, please beware of the following:

* Redundant compounds will not be added to the library
* If you choose to perform Rt prediction from logP, the regression will be updated using the new compounds if they have an experimental Rt available (unless you provide a file with a “Rt\_logP” column instead of a “logP” column\* and in that case, existing Rt prediction are kept as is).

*\*e.g., if you reimport the output file titled “[Library name]\_Isotopic patterns.xlsx” as is in the library to avoid recalculating isotopic patterns*

## Screening

### Templates

#### Dataset templates

The dataset may come from any pre-processing software, although it is **strongly recommended** to avoid any data filtering/data reduction steps (i.e., no isotope filtering, no adduct filtering, etc.) since it can prevent Scannotation from calculating isotope scoring for instance. Templates for two types of datasets supported by Scannotation are available in the *Scannotation templates and library* folder.

Standard datasets (i.e., obtained from XCMS for instance) must contain columns labelled “Row”, “mzmed”, “rtmed”, as well as an empty column with the header “Samples” right before the first sample column as shown in the template file (*Scannotation templates and library* > *Dataset\_standard template.xlsx*).

For the other template (*Scannotation templates and library* > *Dataset\_other template.xlsx*), columns “Row”, “m/z”, “Ret. Time”, and “Use” must be present.

The following table summarizes the column headers and contents needed for both dataset templates.

|  |  |  |
| --- | --- | --- |
| **Column header  Standard dataset (XCMS)** | **Column header  Other template dataset** | **Detailed description** |
| **Row** | **Row** | Row identifier |
| **mzmed** | **m/z** | m/z reference value for the feature |
| **rtmed** | **Ret. Time** | Rt reference value for the feature |
| **Samples** | **Use** | Content does not matter, indicates the beginning of sample columns |
| **[Sample name 1]** | **[Sample name 1]** | Area (or height) values for each feature in each sample |
| **[Sample name 2]** | **[Sample name 2]** |
| **[Sample name i]** | **[Sample name i]** |

Please note that other columns can be present in the dataset, as long as these headers are used for the correct columns.

#### Experimental Rt replicates template

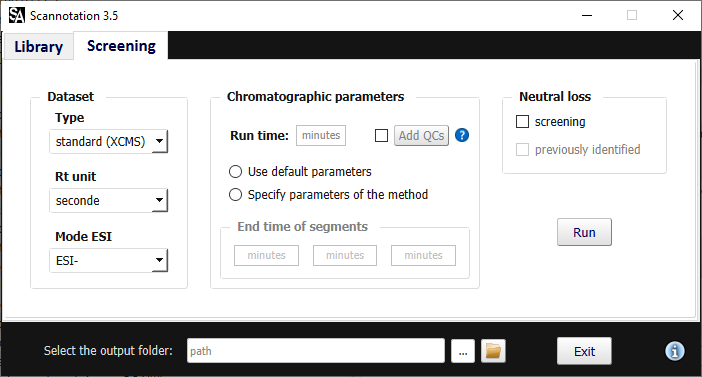
To adapt experimental Rt tolerance values to each user’s method, we recommend that users provide a file with replicate experimental Rt values for any set of compounds. A template reference file (*Scannotation templates and library* > *Rt\_exp replicates\_template.xlsx*) is available to help you enter any available Rt replication data you may have on your chromatographic method. Any set of compounds may be used, with at least four **recommended** replicated Rt values per compound to calculate means and standard deviations. You can add as many columns as you want to increase accuracy, provided the column name format is respected (i.e., “QC[i]”). Please make sure the chosen compounds adequately cover your chromatographic range

|  |  |
| --- | --- |
| **Column name** | **Detailed description** |
| **Compound** | Compound name |
| **QC1** | Rt values for the compound in each replicate QC sample |
| **QC2** |
| **QC3** |
| **QC4** |
| **QCi** |

### Screening tab

The “Screening” tab is where parameters for the suspect screening step can be declared. There are three main sections in this tab: “Dataset”, “Chromatographic parameters”, and “Neutral loss”.

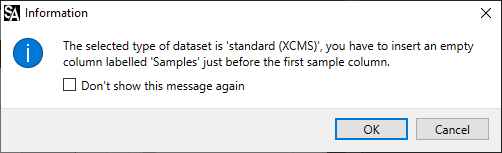
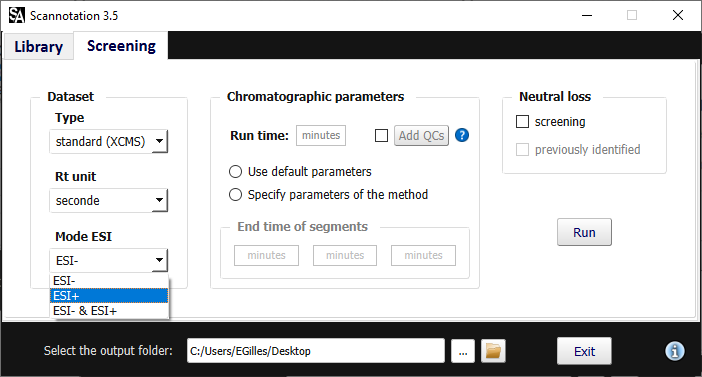
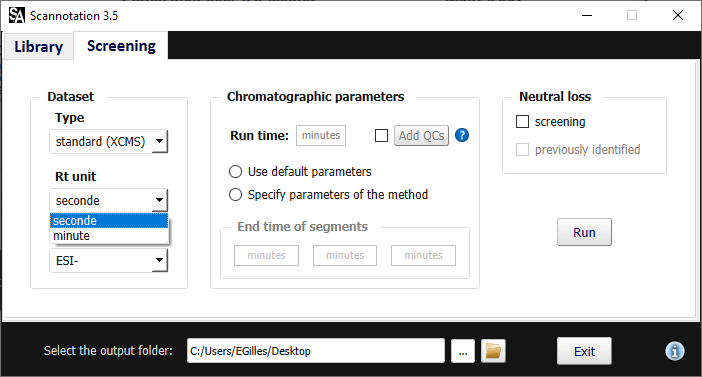
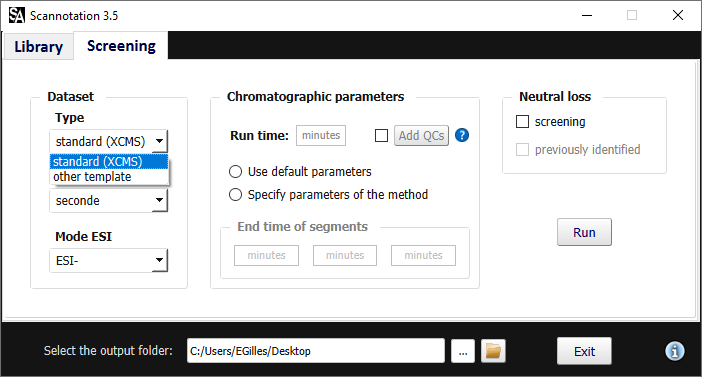
Overall, if all of the sections are active/used, and depending on the length of your suspect list, the suspect screening process may take anywhere from a few seconds up to a few minutes. Progress bars will appear along the process to let you know whether everything is working as intended.



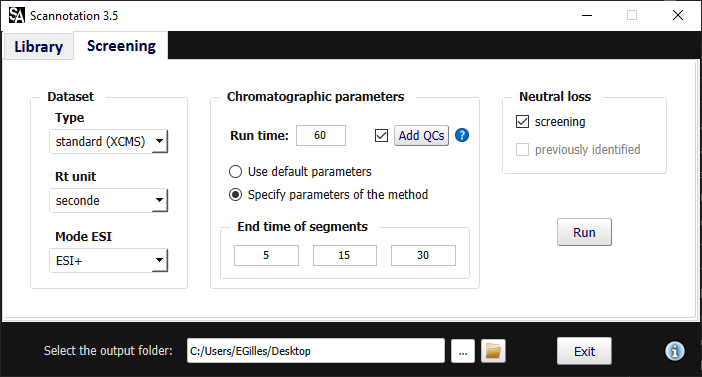
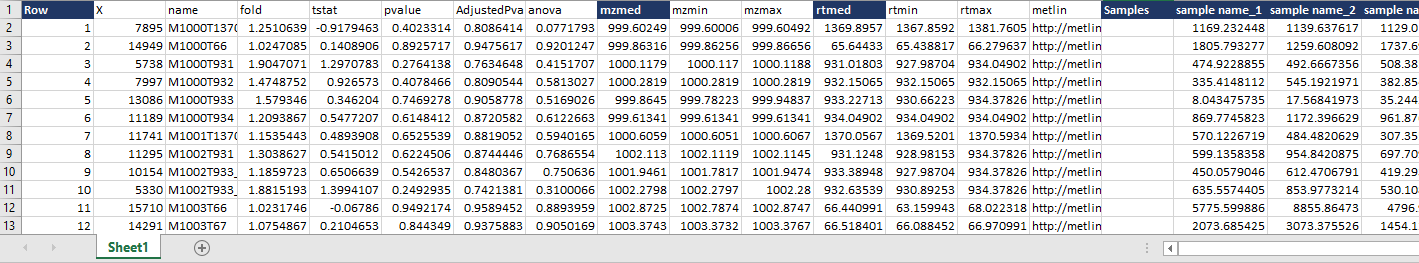
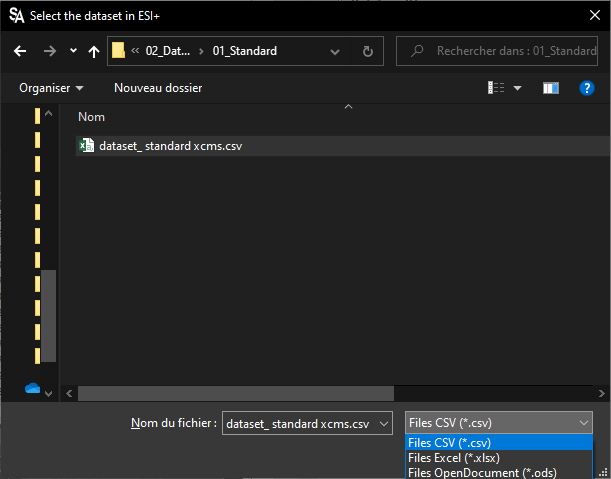
### Dataset

The dataset may come from any pre-processing software, although it is **strongly recommended** to avoid any data filtering/data reduction steps (i.e., no isotope filtering, no adduct filtering, etc.) since it can prevent Scannotation from calculating isotope scoring for instance. Please check paragraph 2.3.1.1. to ensure that your dataset is correctly formatted for Scannotation.

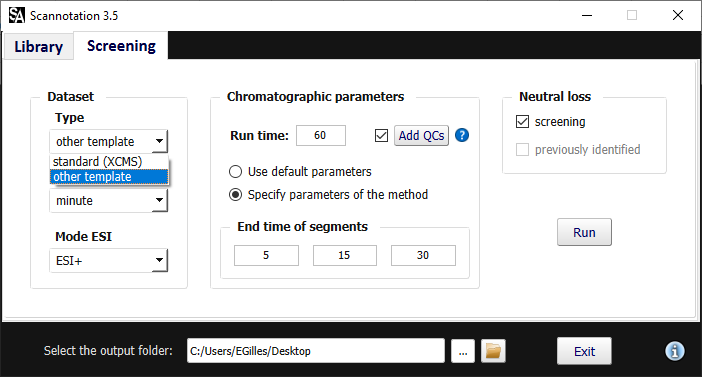
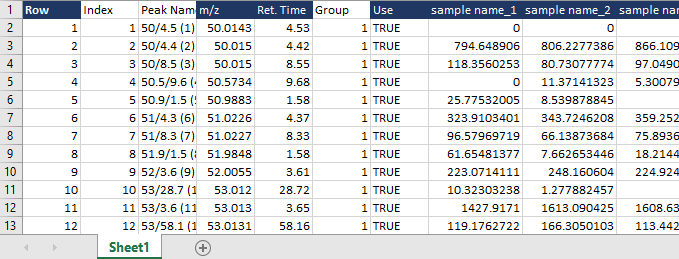
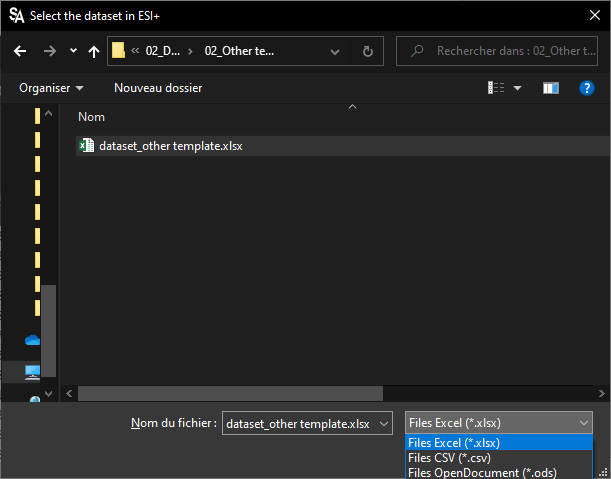
Depending on your data, make sure to choose the correct Rt unit in the second scrolling menu.



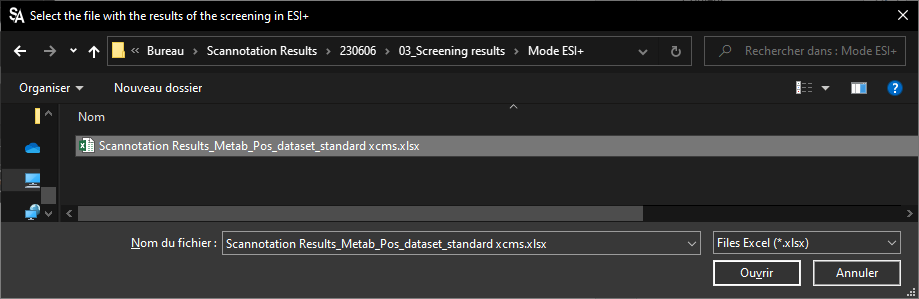
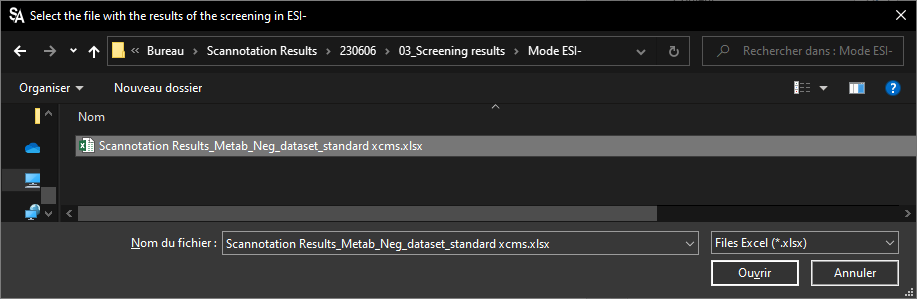
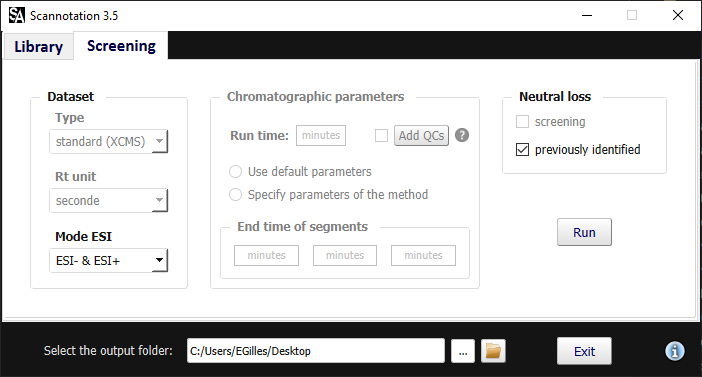
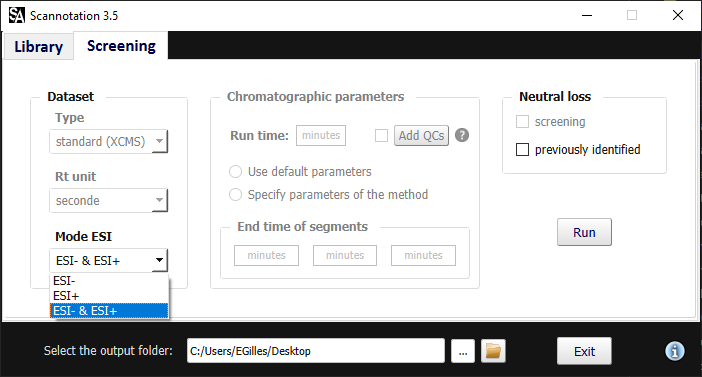
For standard datasets (i.e., obtained from XCMS for instance), it is important that you check for the presence of columns labelled “Row”, “mzmed”, “rtmed”, and that you insert an empty column labelled “Samples” right before the first sample column as shown in the template file.



For the other template, columns “Row”, “m/z”, “Ret. Time”, and “Use” must be present.

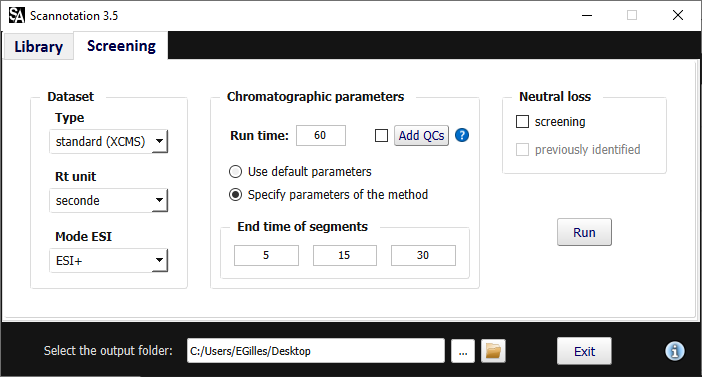
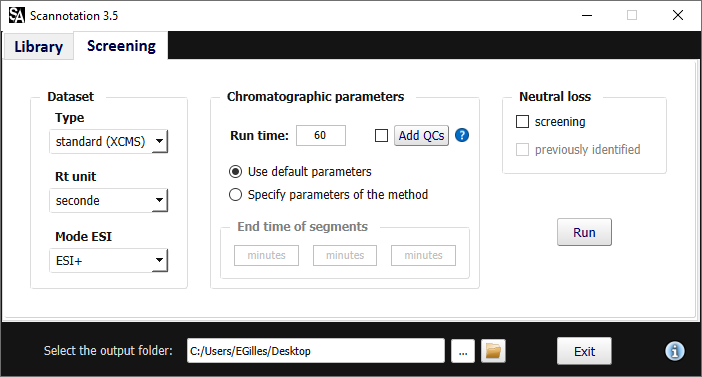


The suspect screening step should be performed on individual ESI modes first (i.e. “ESI-“ or “ESI+” in the third scrolling menu of the “Dataset” section). If you have screening results for the same samples in both ionization modes, you can combine the results using the “ESI- & ESI+” option in this scrolling menu.

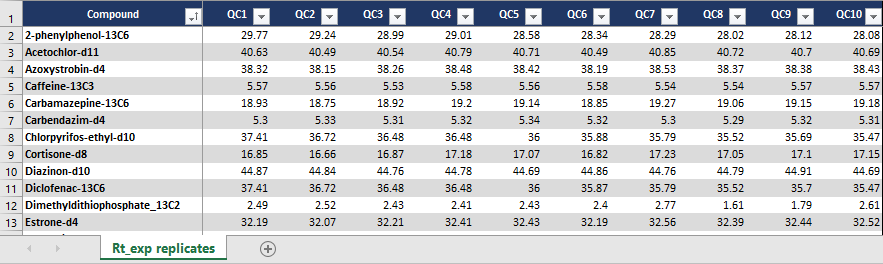
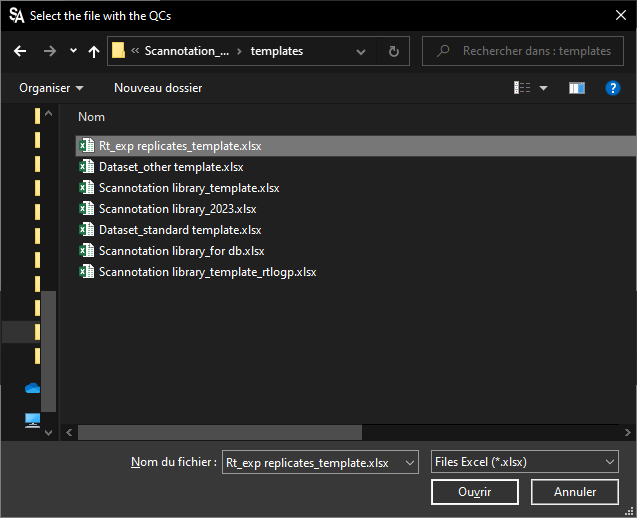
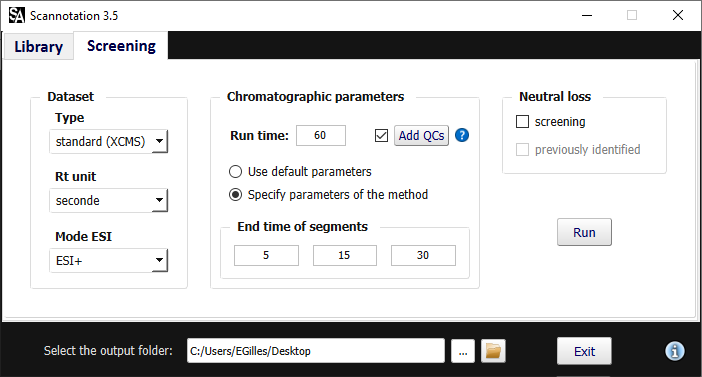
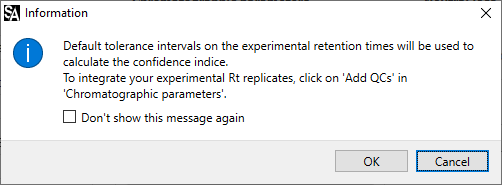


### Chromatographic parameters

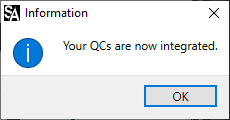
Regarding chromatographic parameters, the run time must first be specified in the dedicated input window. You will then have to choose between “Use default parameters” or “Specify parameters of the method” by creating four segments in your chromatographic run (informed by three end times). These parameters are used to create non-linear tolerance for the Rt tolerance. This is particularly relevant if you observe different Rt shifts along your chromatographic run that is matrix-dependent (e.g., lysophospholipids can cause higher Rt variations towards the end of the run for plasma/serum samples). Thus, segments are implemented to account for varying expected uncertainty on Rt along the run time.

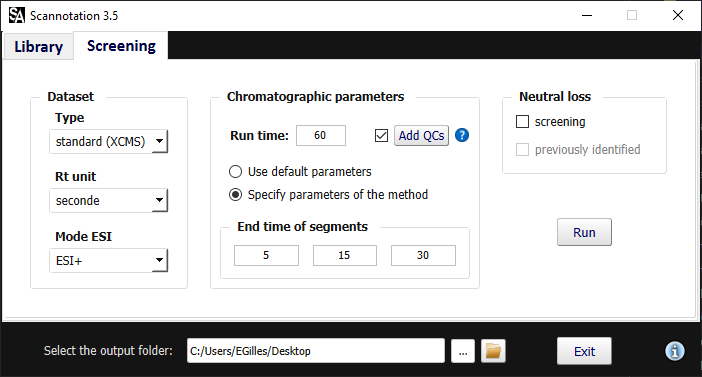


By default, the run time is separated in four equal segments, and tolerance values for one of our methods are used. However, for more accurate results, we **strongly recommend** to tailor results to your methods by specifying parameters for your method. End time of segments should be chosen as the critical points of the gradient, e.g., where the gradient line breaks. Moreover, to compute custom tolerance values, you will need to check the box next to “Add QCs”. Please check paragraph 2.3.1.2. to ensure that your data is correctly formatted.



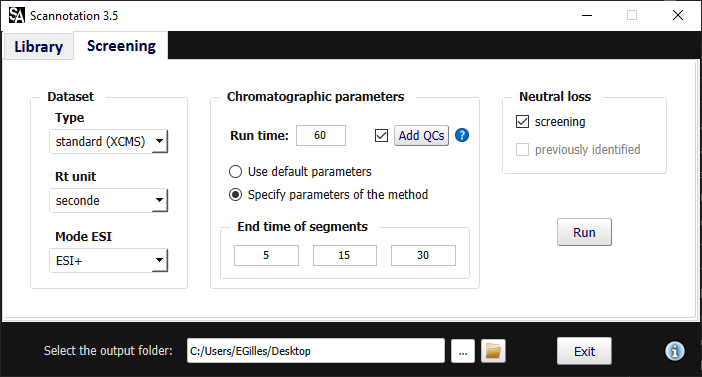
If you have provided a correctly formatted file, a message will appear to confirm that your data was taken into account.





### Neutral loss

If you wish to check for the presence of neutral loss patterns associated with the conjugated metabolites generated through the “Library” tab in your dataset, you can check the “Screening” box in the “Neutral loss” section.

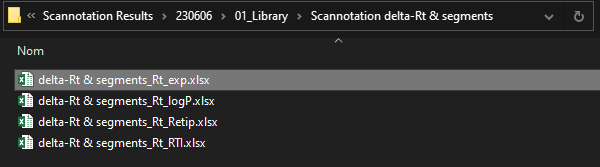
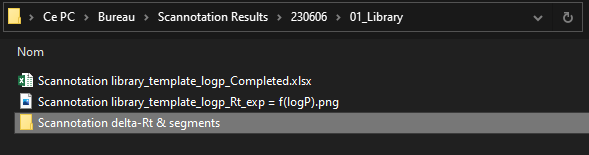


### 

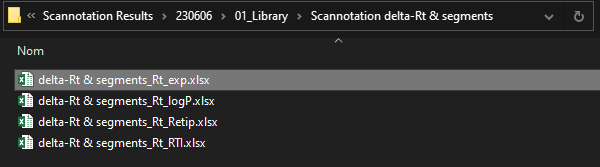
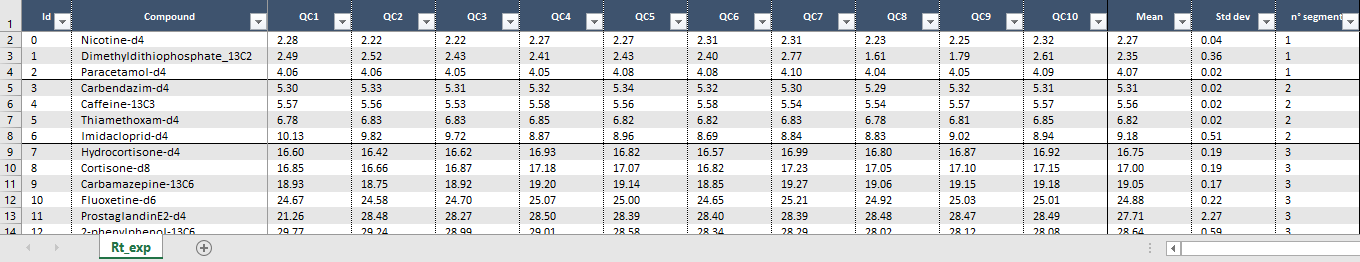
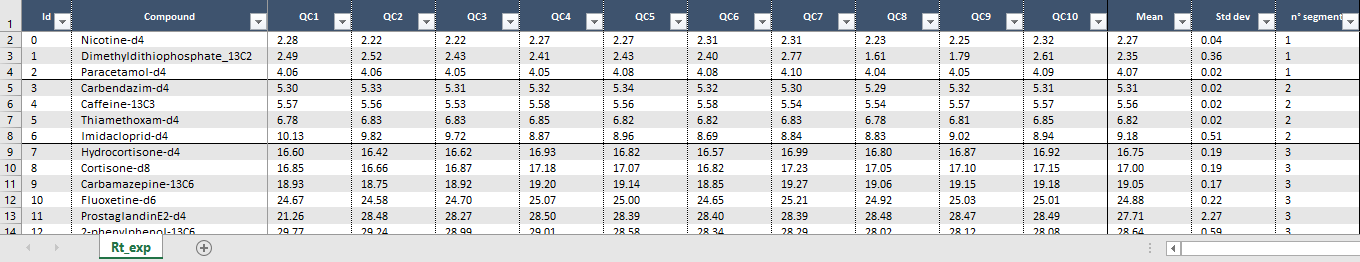
### Results

#### Rt tolerance results

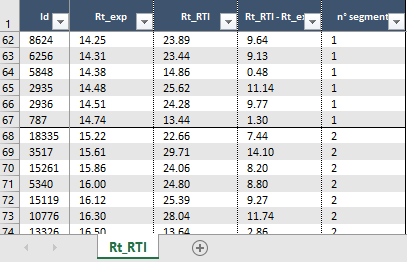
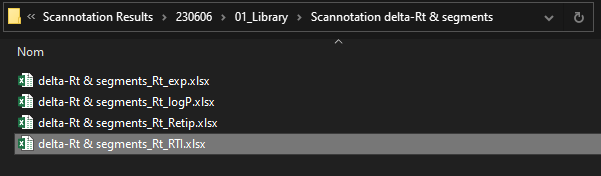
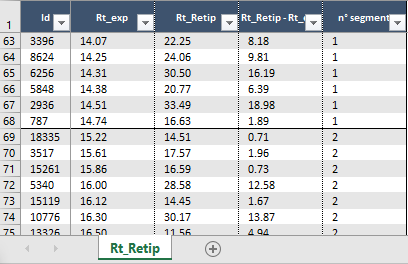
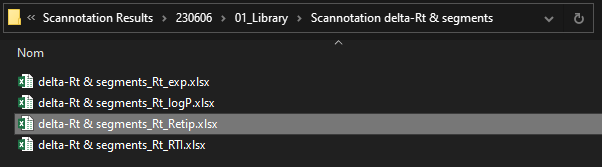
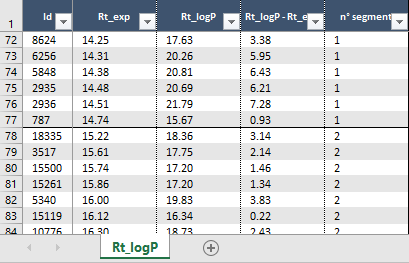
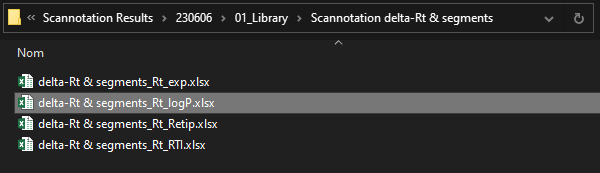
Several output files are generated to illustrate the tolerance values (“delta-Rt”) associated with each combination of Rt type and chromatographic segment.



For experimental Rt, results are computed if you provided repeatability data for a set of compounds using the “Add QCs” option.

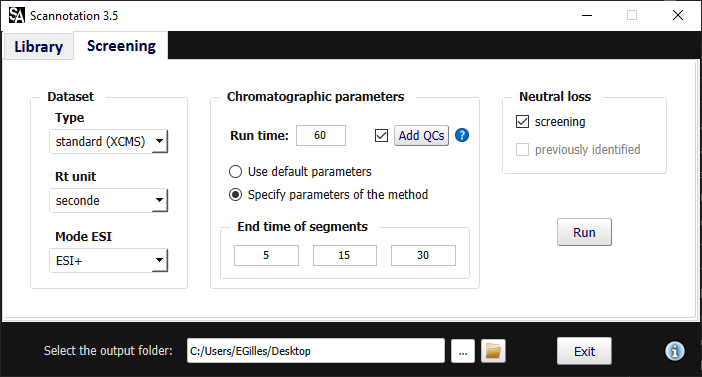


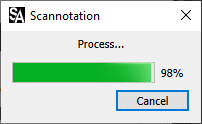
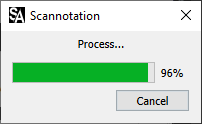
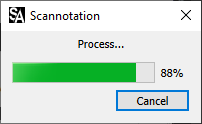
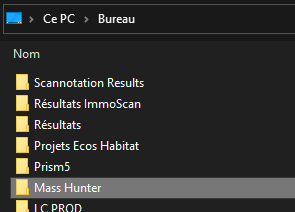
For all predicted logP values, differences between predicted and experimental Rt (when available) are computed to illustrate the variation in prediction accuracy within each chromatographic segment.



#### Screening results

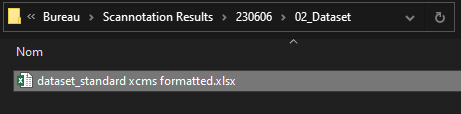
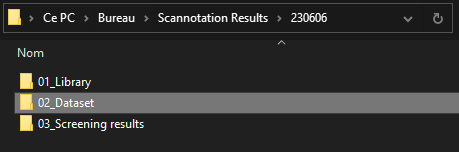
Screening results will be computed by clicking on the “Run” button. Progress bars will appear along the formatting and calculation steps to verify that everything is proceeding as expected.



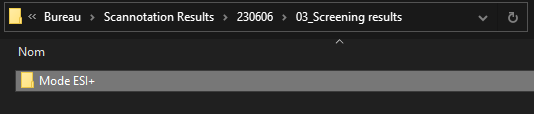
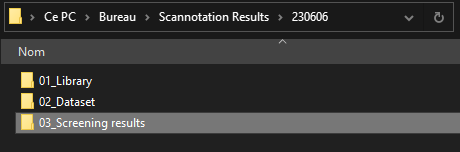


Your formatted dataset will be available using Ctrl+N, or in *Scannotation Results* > *[Date]* > *02\_Dataset* > *[Dataset Name] formatted.xlsx*, and screening results will be available in *Scannotation Results* > *[Date]* > *03\_Screening results* > *Mode ESI[polarity]* > *Scannotation Results[\_Metab]\_[Mode]\_[Dataset Name].xlsx*.





Screening results for a single mode are presented as a table following the template below. For a full description of each column’s purpose, see Section 4.



Identifiers

m/z data



Rt data



Isotopic fit data



Global CI

Ion detection



In this file, some cells may be empty (e.g., CI values for Rt or isotopic fit). This is either because no related data was found in the library (i.e., no Rt values or no isotopic data calculated), or because no related data was found in the feature list (i.e., no feature corresponding to an M+1 or M+2 isotopologue was found in the feature list).

**Codification of global confidence index (CI):**

|  |  |  |  |
| --- | --- | --- | --- |
| **CI** | **Neutral loss not detected (1)** | **Neutral loss detected (2)** | **AreaMetabolite ≥ AreaParent compound (3)** |
| m/z + Rt + Is\_M2 | G3b | G**4c**b | G4**d**b |
| m/z + Rt + Is\_M1 | G3a | G**4c**a | G4**d**a |
| m/z + Is\_M2 | G2b | G**3c**b | G3**d**b |
| m/z + Is\_M1 | G2a | G**3c**a | G3**d**a |
| m/z + Rt | G2 | G**3c** | G3**d** |
| m/z | G1 | G**2c** | G2**d** |

*(1) 1, 2, 3 refer to number of predictors used and "a" or "b" refer to the isotopologue considered*

*(2) Number of predictors used raised by 1 and "c" indicates one of the predictors is neutral loss*

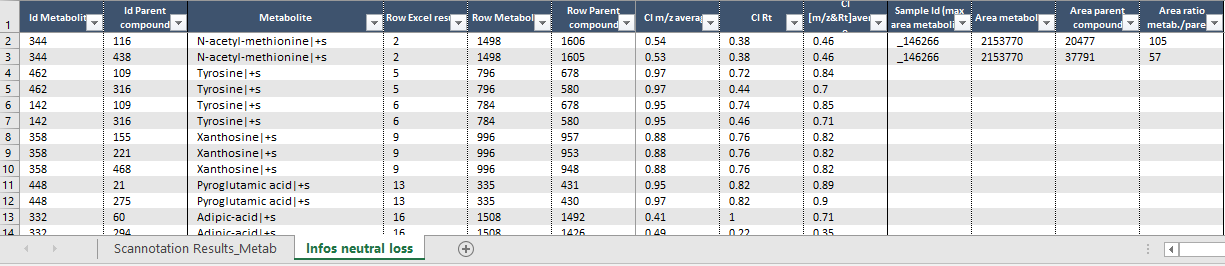
*(3) "d" indicates one of the predictors is neutral loss and metabolite is more abundant than parent compound*

**Report conditional formatting:**

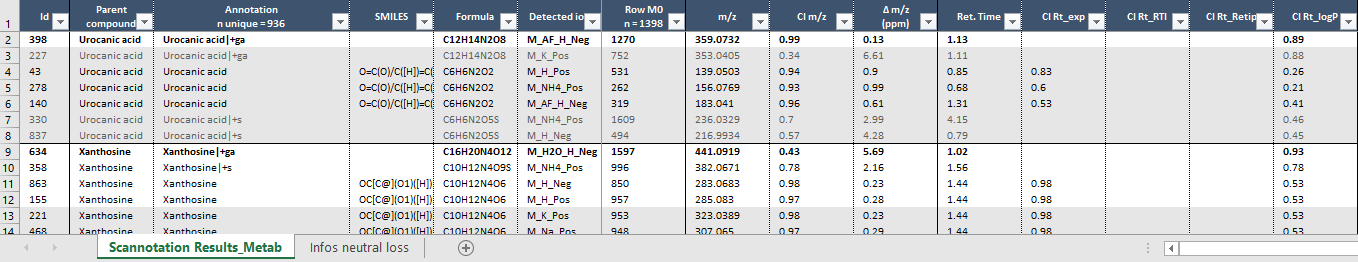
|  |  |
| --- | --- |
| **Condition** | **Fill color and font color** |
| new annotation | white fill, black and bold font, continuous top border side |
| CI Global G4, G3 ≥ 0.7 | white fill, black font |
| CI Global G2 ≥ 0.7 | grey fill, black font |
| CI Global G4, G3, G2 < 0.7 | grey fill, grey font |
| CI Global G1 | grey fill, grey font |

If a neutral loss screening was performed, a second tab named “Infos neutral loss” is also available, which links dataset rows identified as potential parent-metabolite pairs. Data regarding relative areas are also provided to help evaluate the plausibility of the pairing. Empty cells in the last four columns of this tab (i.e., “Sample Id”, “Area metabolite”, “Area parent compound”, and “Area ratio metab/parent”) indicate that although features coherent with a neutral loss pattern were found, one of the following abundance criteria was not met:

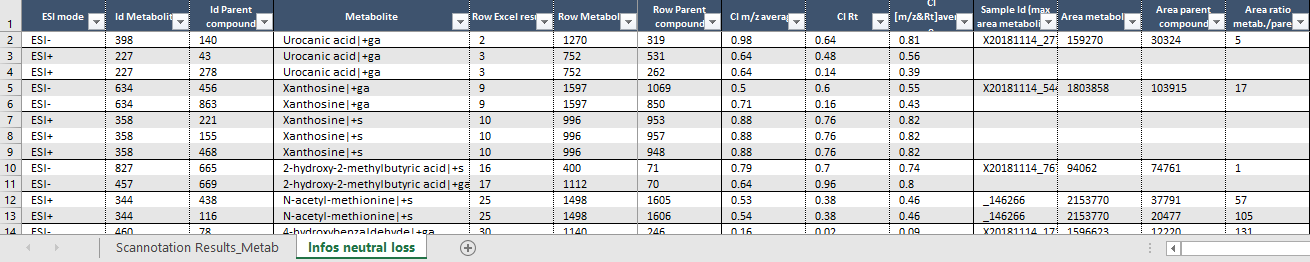
Areametab ≥ 100, Areaparent ≥ 20, and Areametab/Areaparent ≥ 1.



Screening results for both ionization modes are presented in the same way; all the detected ions (even from different modes) for the same parent compounds are combined in the same group of features.



Regarding neutral loss, results from both modes are also combined.



# FAQ

## Library

* **How to avoid reprocessing the isotopic pattern calculation for every screening?**

If you keep using the same library and wish to cut down on the library processing time, you can select a previously computed “*[Library name]\_Isotopic patterns.xlsx*” file to process the other parameters, as the isotopic pattern generation is the most time-consuming step. Please refer to paragraph 2.2.3 for more information.

## Screening

### Dataset

* **What is the minimum number of samples required to perform the screening step?**

The screening step may be performed using one sample only.

* **How do I choose the correct time unit for my dataset?**

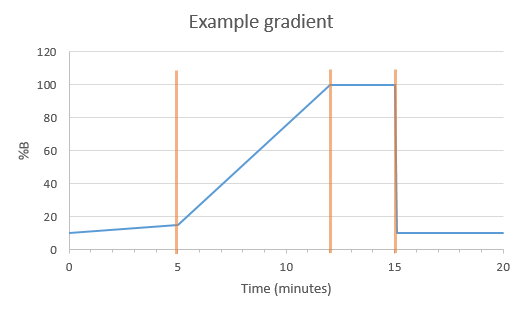
By default, the Rt unit is set in seconds for XCMS datasets and in minutes for the template provided in Scannotation. However, you are free to change it if needed using the “Rt unit” scrolling menu.

* **How to get a screening in both ESI- and ESI+ modes?**

Scannotation cannot perform both screenings at the same time. If you wish to have a screening in both ionization modes, you must first perform the screenings individually using the “ESI-“ and “ESI+” options in the “ESI mode” scrolling menu, and then combine the results using the “ESI- & ESI+” option of this same menu. Combining results from both modes allow grouping redundant information together, such as compounds detected in both modes or parent compounds detected in one mode and metabolites in the other (only if the neutral loss screening was performed), to increase confidence in annotations.

### Chromatographic parameters

* **What is the best way to choose end times of segments?**

Segments were created to account for varying expected uncertainty on Rt along the run time. Therefore, end times of segments should be chosen as the critical points of the gradient. For instance, in the following example, end times of segments should be placed at 5, 12 and 15 minutes.

* **Why hasn’t Scannotation computed all the available metabolites for all of my library compounds?**

Sulfate conjugates (+SO3) are only added for compounds presenting an oxygen atom in their molecular formula, and glucuronide conjugates (+C6H8O6) are only added for compounds presenting an oxygen or a nitrogen atom in their molecular formula.

# Screening results – Column names

Below is the detailed description of the content of each column of the screening results file.

|  |  |  |
| --- | --- | --- |
|  | **Column name** | **Detailed description** |
| **Identifiers** | **Id** | Scannotation identifier of the dataset row |
| **Parent compound** | Parent compound of suggested annotation (in the case of metabolites) |
| **Annotation** | Suggested annotation |
| **SMILES** | SMILES identifier provided in the library |
| **Formula** | Molecular formula provided in the library |
| **Detected ion** | Detected adduct |
| **Row M0** | Dataset identifier of the annotated feature |
| **m/z data** | **m/z** | Experimental m/z value |
| **CI m/z** | Confidence index computed for m/z |
| **Δ m/z (ppm)** | Relative deviation of m/z between the experimental feature and the suspect (in ppm) |
| **Rt data** | **Ret. Time** | Experimental Rt value |
| **CI Rt\_exp** | Confidence index computed for experimental Rt (if available) |
| **CI Rt\_RTI** | Confidence index computed for RTI-predicted Rt (if available) |
| **CI Rt\_Retip** | Confidence index computed for Retip-predicted Rt (if available) |
| **CI Rt\_logP** | Confidence index computed for logP-predicted Rt (if available) |
| **Isotopic fit data** | **Isotope Id** | Isotopologue considered to compare the experimental feature and the suspect |
| **Row Mn** | Dataset identifier of the M+n feature considered |
| **CI Is\_m/z** | Confidence index computed for isotopologue m/z |
| **CI Is\_Rt** | Confidence index computed for isotopologue Rt |
| **CI Is\_Average [m/z & Rt]** | Average of two previous CI values |
| **CI Is\_max area ratio Mn/M0** | Confidence index computed for isotopologue area ratios |
| **CI Is\_Total** | Confidence index computed for isotopic fit overall, combining the two previous CI values |
| **Area Mn** | Maximal area of M+n feature considered |
| **Area M0** | Area of M+0 feature considered in the sample with maximal area of M+n feature considered |
| **Sample Id** | Sample in which maximal area of M+n feature considered is found |
| **Global CI** | **CI Global** | Global Confidence index scoring the overall proximity between feature and suspect |
| **Ion detection** | **Detection Frequency (%)** | Detection frequency of feature in the dataset |
| **Area max M0** | Sample in which maximal area of feature is found |
| **Sample Id area max** | Sample in which maximal area of M+0 feature considered is found |