

*MS-FINDER tutorial:
Universal program for metabolite annotation*

15/04/2020

Justine Chervin and Guillaume Marti

justine.chervin@lrsv.uprs-tlse.fr

guillaume.marti@univ-tlse3.fr

<https://www.lrsv.uprs-tlse.fr/metatoul-en/>

Table of content

1. Prerequisite:.....	1
2. Importing data into MS-FINDER	1
3. Define search parameters.....	5
4. Annotation results	9
5. Export results.....	11

1. Prerequisite:

Software installation:

MS-DIAL last version: <http://prime.psc.riken.jp/compms/msdial/main.html>

MS-FINDER last version: <http://prime.psc.riken.jp/compms/msfinder/main.html>

MSFileReader for Thermo data: see the following link which explains how to process <http://fields.scripps.edu/rawconv/>

2. Importing data into MS-FINDER

a. From a folder containing .MAT files

On the MS-DIAL interface, right-click on one point, then on the “**Search formula and structure**” menu select “**Add component to search list**” and export « **all peaks** » as **.MAT files** for further identification purpose with MS-FINDER.

MS-DIAL ver. 4.18 C:\Users\ju_78\Desktop\Solstice\bois-T3-2019\abf-neg\boisT3-negv2.mtd

File Data processing Post processing Data visualization Search View Option Export Help

File navigator

- blanc-ext-neg1
- blanc-ext-neg2
- blanc-neg-2bis
- blanc-neg-3
- blanc-neg-4
- blanc-neg-4_200309080858
- blanc-neg-5
- blanc-neg-6
- blanc-neg-7
- blanc-neg-8
- blanc-neg-9
- blanc-neg-10
- blanc-neg-11
- blanc-neg-12
- blanc-neg-13
- blanc-neg-14
- blanc-neg-15
- blanc-neg-16
- blanc-neg-17
- blanc-neg-18
- blanc-neg-19
- blanc-neg-20

Alignment navigator

- alignmentResult_2020_3_18_14
- alignmentResult_2020_3_18_1E5
- alignmentResult_2020_3_18_8E5
- alignmentResult_2020_3_19_5E5
- alignmentResult_2020_3_20_5E5
- alignmentResult_2020_3_20_13_

Peak spot navigator

Label: None

Peak spots: 100% Num. 1879

Display filter

☐ Ref. matched ☐ Suggested

☐ CCS matched ☐ Unknown

☐ MS2 acquired ☐ Molecular ion

☐ Blank filter ☐ Unique ions

EIC of focused spot

UnknownRT[min]=6.57 m/z=865.1995

Bar chart of aligned spot (NH)

EIC of aligned spot

Basic peak property

Annotation: Unknown

RT[min]: 6.574

Adduct type: [M-H]⁻

m/z: 865.19952

Peak height(jares): 1280850 (height average in samples)

Formula/Ontology: NA/NA

InChIKey: NA

Comment:

Compound detail

Survey scan (MS1) spectrum

Peak spot viewer

Alignment spot viewer

Show ion table

MS1 spectra Max intensity: 4146642

Relative abundance

174.9557

320.8863

466.817

alignmentResult_2020_3_19_5E5

m/z

1608; RT[min]: 6.57 m/z: 865.1995

Save image as...

Copy image as...

Search formula and structure

Export as MRMPROBS reference format

MS2 spectra Precursor: 865.1994

Measurement

125.0242

abundance

No information

Store MS annotation tag (MAT)

Path: C:\Users\ju_78\Desktop\bois-T3-2019\script-boisT3\neg\peaks

Browse

Spectrum type:

Export option: All peaks

Export

Cancel

* All of monoisotopic ions is exported in isotope tracking project.

Then on the MS-FINDER interface, in the “**File**” menu, click on “**Import**” and select the folder containing the .MAT files previously exported from MS-DIAL. The files imported are then visible in the top left-hand corner rectangle of the interface.

MS-FINDER ver. 3.30 C:\Users\ju_78\Desktop\Solstice\bois-T3-2019\MAT-files-neg

File Analysis Setting Export Tool Help

Import... Select a folder including MAT or MSP files.

Create a query

AlignmentID 100_1.23_207.0508_2020_3_24_920_alignmentResu
AlignmentID 101_0.69_207.931_2020_3_24_920_alignmentResu
AlignmentID 102_1.2_209.0301_2020_3_24_920_alignmentResu
AlignmentID 103_1.35_209.0665_2020_3_24_920_alignmentResu
AlignmentID 104_1.21_209.0665_2020_3_24_920_alignmentResu
AlignmentID 105_12.17_211.0763_2020_3_24_920_alignmentResu
AlignmentID 106_14.65_211.0765_2020_3_24_920_alignmentResu
AlignmentID 107_1.4_212.0234_2020_3_24_920_alignmentResu
AlignmentID 108_1.06_212.9206_2020_3_24_920_alignmentResu
AlignmentID 109_1.19_215.0327_2020_3_24_920_alignmentResu
AlignmentID 110_1.46_115.0033_2020_3_24_920_alignmentResu
AlignmentID 111_1.16_216.9091_2020_3_24_920_alignmentResu
AlignmentID 112_1.18_217.048_2020_3_24_920_alignmentResu

File information

Name: Unknown
Scan number: 0
Retention time [min]: 1.17
Precursor m/z [Da]: 102.0558
Precursor type: [M-H]
Ion mode: Negative
Spectrum type: Centroid
Collision energy: 40
Formula:
SMILES:
InChIKey:
Intensity: 1878507
MS1 peak number: 3
MS2 peak number: 10

Molecular formula finder

Formula	Error [mDa]	Error [ppm]	Score	Resource	Select
---------	-------------	-------------	-------	----------	--------

Structure finder

Name	Score (max=10)	Ontology	InChIKey
------	----------------	----------	----------

MS1 spectrum

Relative abundance

m/z

MS/MS spectrum

Relative abundance

m/z

Spectrum Structure Meta data

Measurement vs. Reference

Actual MS/MS

Relative abundance

Fragment score

m/z

In silico MS/MS

Rechercher un dossier

Choisissez un dossier.

lefil

- Solstice
 - biblio-vigne
 - bois-2019
 - bois-T3-2019
 - abf-neg
 - abf-pos
 - BDD-vigne
 - DB-msdial03
 - DDB-msdial02
 - MAT-files-neg**
 - pos
 - raw-neg
 - script-boisT3
 - stat
 - bois-T6-2019

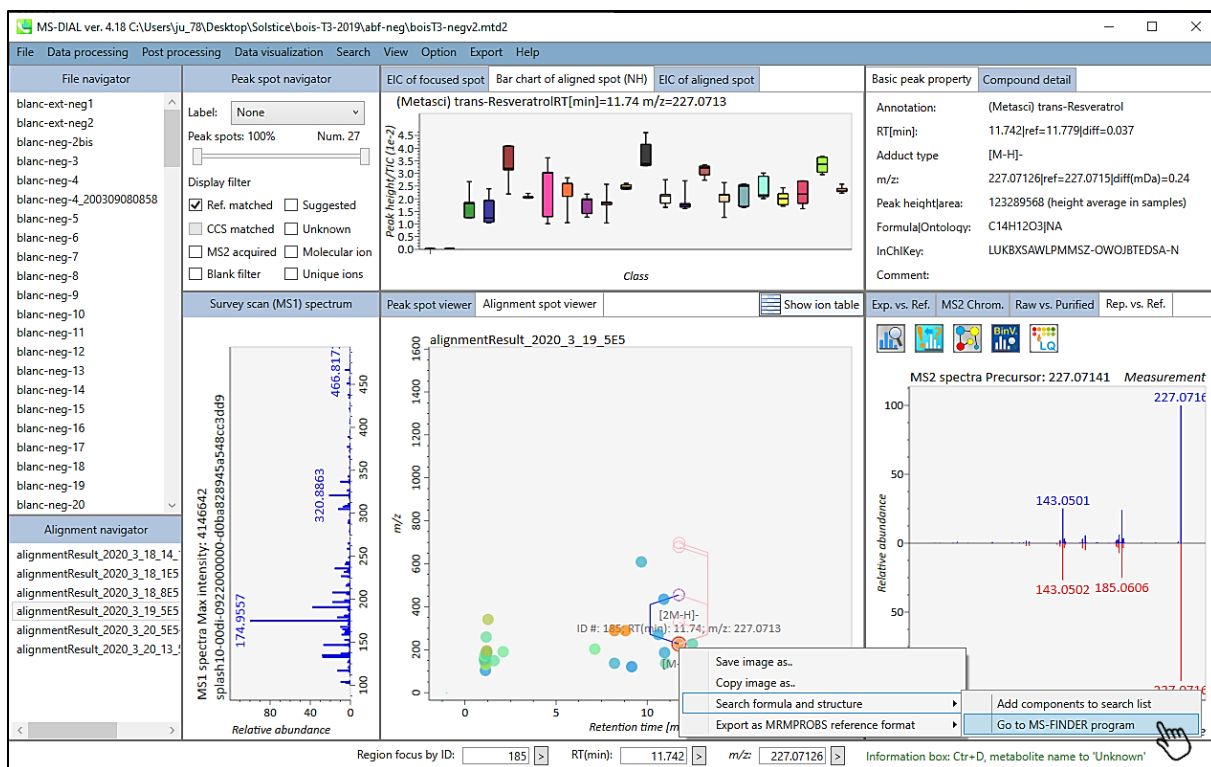
Créer un nouveau dossier OK Annuler

MAT-files-neg

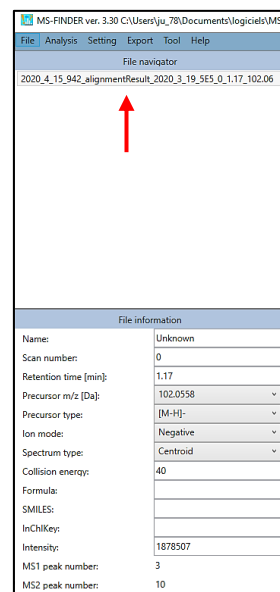
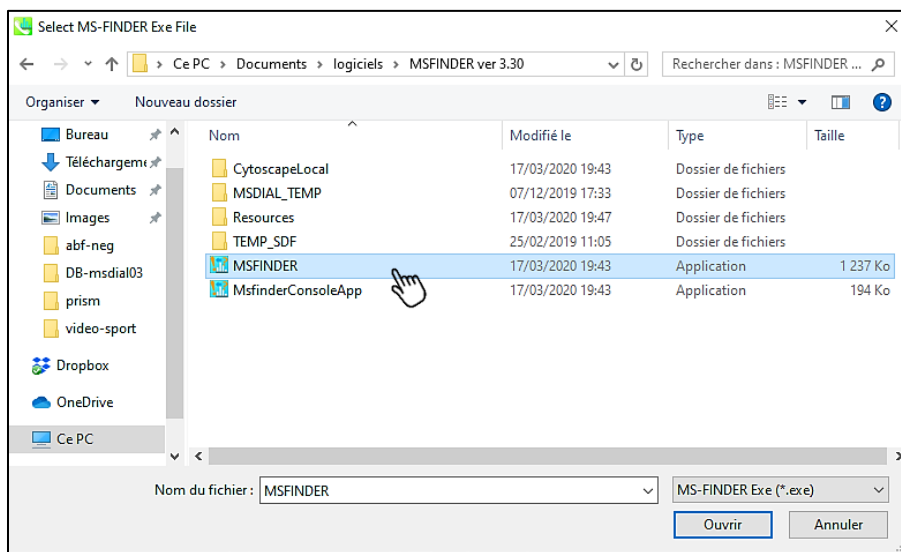
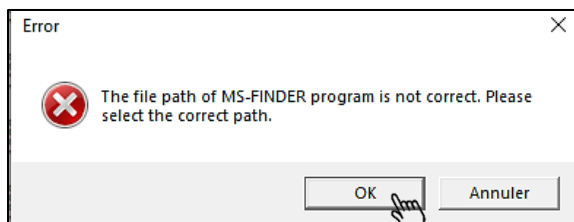
Noms	Modifié le	Type	Taille
AlignmentID 0_1.17_102.0558_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 1_1.12_104.0951_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 2_8.21_109.0206_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 3_8.81_109.0207_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 4_1.13_111.0083_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 5_1.05_112.9831_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 6_12.77_112.9832_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 7_1.14_113.0352_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 8_1.13_114.0162_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 9_0.95_114.0362_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 10_1.46_115.0033_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 11_0.56_115.8204_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 12_1.47_115.8205_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 13_0.56_116.0281_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 14_1.42_125.0231_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 15_8.21_125.0232_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 16_1.33_126.9540_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 17_1.14_127.051_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 18_1.33_128.033_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 19_1.13_128.033_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 20_1.06_128.9952_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 21_1.52_128.9952_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 22_0.95_130.9834_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 23_1.15_131.046_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 24_1.15_132.03_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 25_0.56_132.8677_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 26_1.46_133.0141_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 27_0.56_134.8851_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko
AlignmentID 28_1.18_217.048_2020_3_24_920_alignmentResu	24/03/2020 09:20	Microsoft Access Table Shortcut	1 Ko

b. Directly into the MS-DIAL interface

You can call MS-FINDER directly from Ms-DIAL. By clicking on one point, then on the “**Search formula and structure**” menu select “**Go to MS-FINDER program**” and export « **all peaks** » as **.MAT files** for further identification purpose with MS-FINDER.



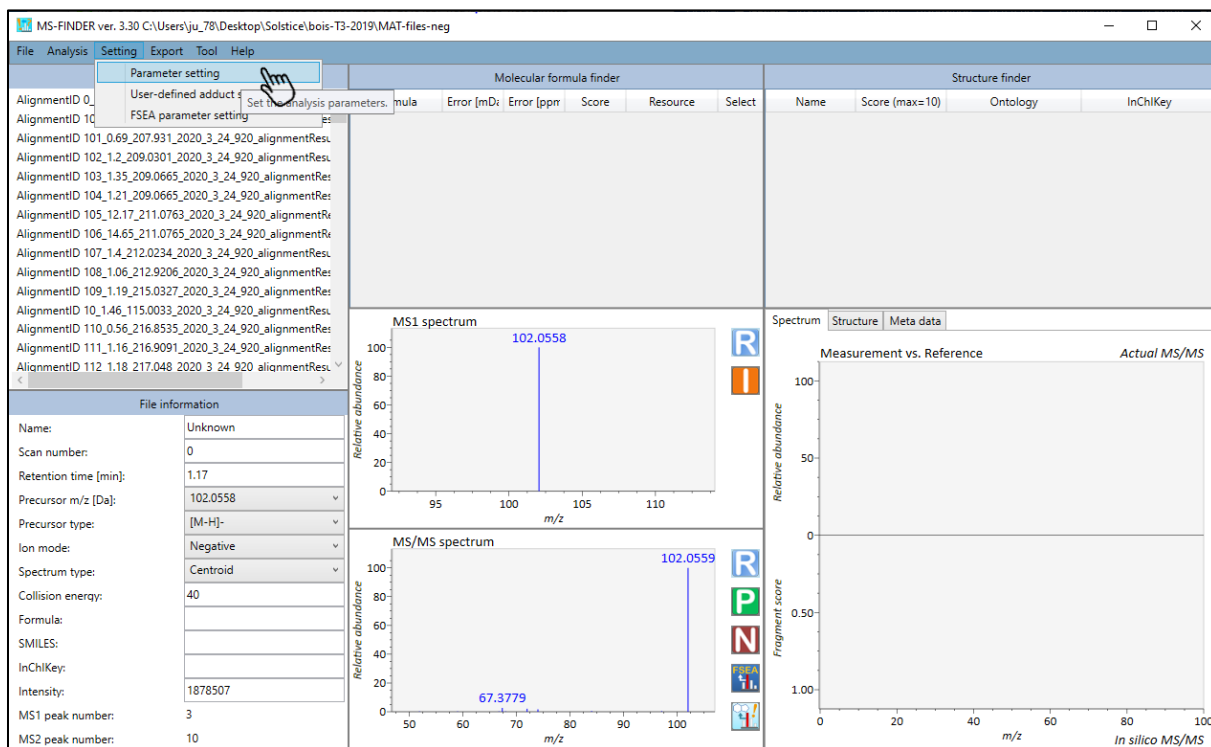
The following error window appears and you have to click on “OK” and go find the application MSFINDER.exe in your computer and open it.



The MS-FINDER interface will then start with the importation of the peak you wanted to interrogate.

3. Define search parameters

On the menu "Setting", select "Parameter setting" to define all the parameters of formula and structure search.



a. Method parameter

MS-FINDER proposes two options to annotate compounds:

- By spectral databases (*.msp)
- By formula and structure finder programs based on *in silico* fragmenter

The screenshot shows the 'Analysis parameter setting' dialog box. The 'Method' tab is selected, and the 'Search option' section is expanded. Red arrows point to specific settings with annotations:

- ☐ Spectral database search: Check for spectral databases search
- ☒ Formula prediction and structure elucidation by in silico fragmenter: Check for *in silico* elucidation of unknown features
- ☐ Use internal experimental library (MassBank, GNPS, ReSpect): To search internal MSMS databases
- ☐ Use in silico spectra for lipids (LipidBlast): To search LipidBlast database
- ☐ User-defined DB: To search your own .msp format database
- ☒ Precursor oriented spectral search: Check to filter out spectral records by precursor m/z value

The 'Information' section at the bottom provides instructions for using the spectral database search option:

1. uncheck 'Precursor oriented spectral search'
2. check 'TMS-MEOX derivative compound' at 'Formula finder' tab.
3. set the same mass tolerance for both MS1 and MS2.

*The above thing is not the case for structure elucidations by accurate EI-MS spectra.

b. Mass spectrum parameter

Mass tolerance setting

Mass tolerance type: ☐ Da ☒ ppm

Mass tolerance (MS1): +-Da or ppm

Mass tolerance (MS2): +-Da or ppm

Abundance setting

Relative abundance cut off: %

Mass range setting

Mass range max: Da

Mass range min: Da

Finish Cancel

Set the mass tolerance for formula search

Set the mass tolerance matching experimental and reference fragments

Set an abundance cut off to remove noise

Indicate the range of MS1 mass

c. Formula finder parameter

Formula calculation setting

LEWIS and SENIOR check: ☒

Isotopic ratio tolerance: %

Element ratio check: covering

Element probability check: ☒

Element selection

☒ O ☒ N ☒ P ☒ S ☐ F ☐ Cl ☐ Br ☐ I ☐ Si

☐ TMS-MEOX derivative compound

Minimum TMS count:

Minimum MEOX count:

Options

Maximum report number: up to 100

Time out (-1 means infinite): min

Advanced settings for AIF: ☐ Setting

Finish Cancel

To respect the valence rules of elements

To calculate the isotopic score

To respect satisfactory element ratios

To respect heuristic rules (Seven Golden Rules)

Select the pertinent elements

Number of reported formula

Define a time-out to accelerate the search process

d. Structure finder parameter

Analysis parameter setting

Method | Mass spectrum | Formula finder | **Structure finder** | Data source | Retention time

In silico MS/MS or EI-MS fragmenter setting

Tree depth: [1-3]

☐ Use the fragmentation library for electron ionization (EI)

☐ Use the fragmentation library for low energy CID

Options

Maximum report number: up to 100

Time out (-1 means infinite): min

Finish Cancel

To restrict *in silico* cleavages. With 2, you generate fragments until product ions of a product ion

Number of reported formula

Define a time-out to accelerate the search process

e. Data source parameter

Analysis parameter setting

Method | Mass spectrum | Formula finder | Structure finder | **Data source** | Retention time

Local Databases

☐ HMDB (Human) ☐ Urine (Human) ☐ Saliva (Human) ☐ Feces (Human)

☐ Serum (Human) ☐ CSF (Human) ☐ SMPDB (Human) ☐ LipidMAPS (Lipids)

☐ YMDB (Yeast) ☐ ECMDB (E.coli) ☐ BMD8 (Bovine) ☐ DrugBank (Drug)

☐ FoodB (Food) ☒ PlantCyc (Plant) ☒ ChEBI (Biomolecules) ☐ T3DB (Toxin)

☐ STOFF (Environment) ☐ BLEX (blood exposome)

☒ KnapSack (Natural product) ☒ NNPDB (Natural product)

☐ PubChem (Biomolecules) ☒ UNPD (Natural product)

☒ User-defined DB

MINEs (Metabolic In silico Network Expansions) setting

☒ Never use it. ☐ Only use when there is no query in local DBs. ☐ Always use it.

PubChem Online setting

☒ Never use it. ☐ Only use when there is no query in local DBs. ☐ Always use it.

Finish Cancel

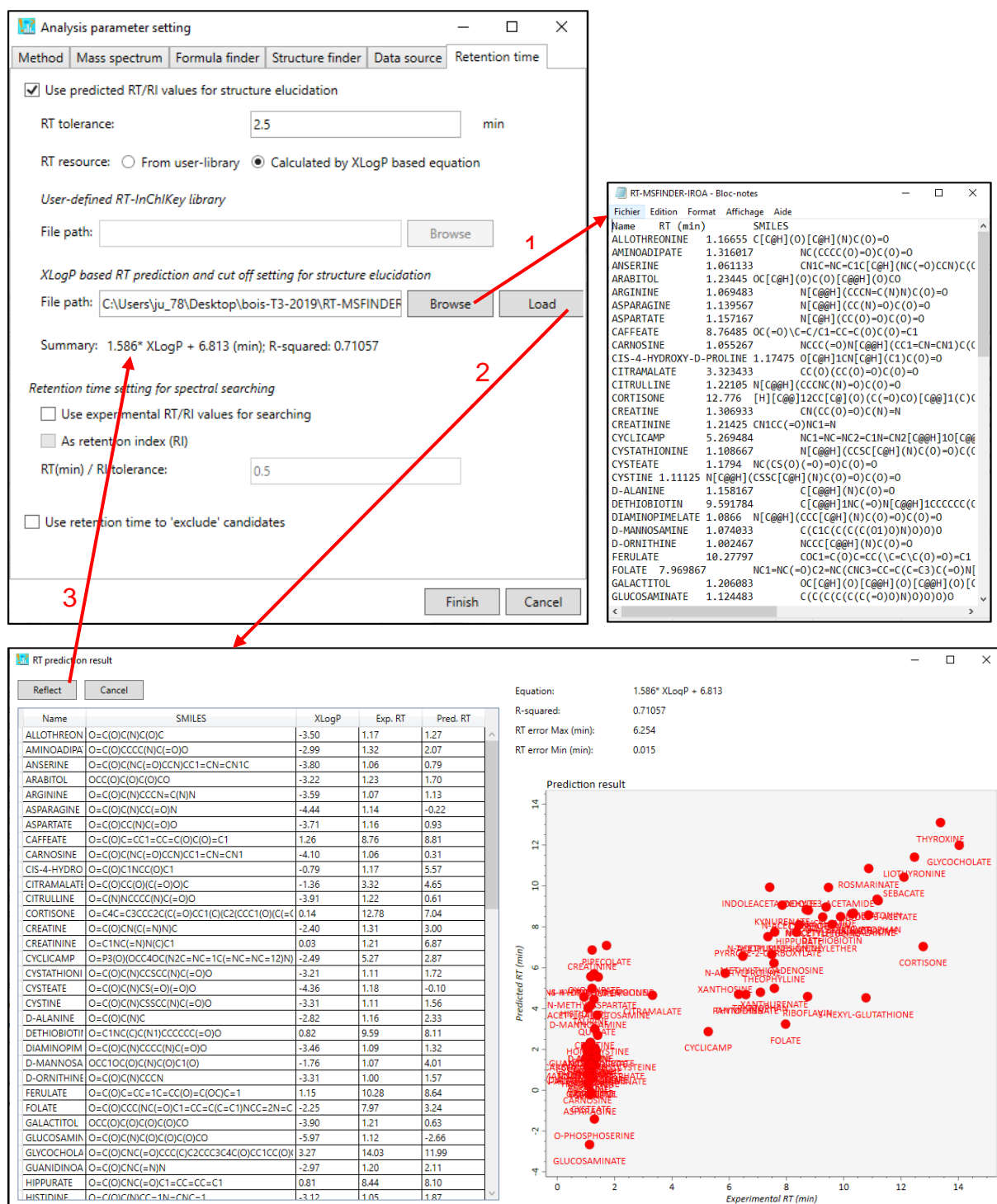
Select suitable databases among the 22 proposed local databases

Define our own text format database for *in silico* elucidation of unknown features

You can choose to widen your search by selecting MINEs and PubChem databases

f. Retention time parameter

You can add a tab-delimited text format file containing retention time of standards to bring a supplementary filter from structure elucidation. The three required columns should contain in the order the “metabolite name”, the “retention time” in min and the “smiles”. Check **“Use predicted RT/RI values”** and **“Calculated by XLogP based equation”**. Then load the text file and click on **“Reflect”** and the equation will appear.



4. Annotation results

a. Single analysis

[1] For formula prediction, double-click on one file in the “**File navigator**”. Check the metadata well first especially the “**Precursor type**”.

[2] For structure prediction, select the formula you want to determine the structure, right-click on the formula result table and select “**Search the structures**”







[3] You obtain a ranking of structure candidates. Select one to see the experimental spectrum confronted to the *in silico* one of the reference (A), the structure (B) or some meta data (C).

The screenshot displays the MS-FINDER software interface with three main panels: File navigator, Molecular formula finder, and Structure finder.

- File navigator:** Lists alignment files. A red box labeled '1' highlights a file, and a red arrow points to the 'Molecular formula finder' panel.
- Molecular formula finder:** A table of formula candidates. A red box labeled '2' highlights the 'Search the structures...' button, and a red arrow points to the 'Structure finder' panel.
- Structure finder:** A table of structure candidates. A red box labeled '3' highlights a candidate, and a red arrow points to the 'Experimental spectrum vs. In silico spectrum' plot.
- Meta data:** A panel on the left showing sample information: Name: (Metasci) (-)-Epicatechin, Scan number: 210, Retention time [min]: 8.81, Precursor m/z [Da]: 289.0718, Precursor type: [M-H]⁺, Ion mode: Negative, Spectrum type: Centroid, Collision energy: 40, Formula: C15H14O6, SMILES: OC3=CC(O)=C2C(OC(C=1C=C...), InChIKey: PFTAWBLQPZVEMU-UKRRQH, Intensity: -1, MS1 peak number: 3, MS2 peak number: 117.
- MS1 spectrum:** A plot of relative abundance vs. m/z showing a base peak at 289.0717.
- MS2 spectrum:** A plot of relative abundance vs. m/z showing peaks at 109.029, 203.0712, 245.0817, and 289.072.
- Experimental spectrum vs. In silico spectrum:** A plot comparing the experimental spectrum (red) with the in silico spectrum (blue) for the selected candidate.

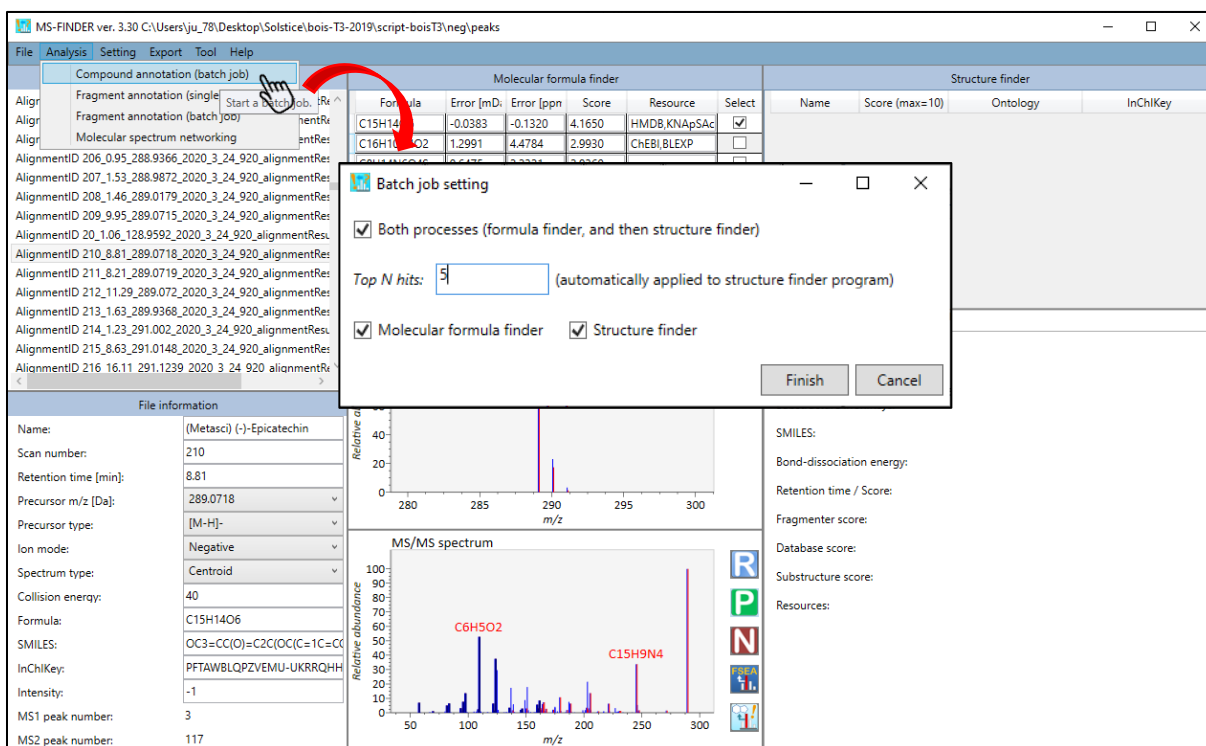
The results are presented in three panels:

- (A) Experimental spectrum of feature:** A plot of relative abundance vs. m/z showing the experimental spectrum (red) and the in silico spectrum (blue) for the selected candidate. The x-axis ranges from 50 to 300 m/z, and the y-axis ranges from 0 to 100 relative abundance. The plot shows a base peak at 289.072 and other significant peaks at 109.029, 203.0712, and 245.0817.
- (B) Structure:** A chemical structure diagram of the compound, showing a flavonoid structure with a central ring and two phenolic rings.
- (C) Meta data:** A table of sample information: Name: (-)-epicatechin, InChIKey: PFTAWBLQPZVEMU-DDOUBSINA-N, Substructure inChIKeys: XLVQNOQVBJNP-OKKJLVBELLTKV-ATUQVWHBWRKTHZ, SMILES: OC3=CC(O)=C2C(OC(C=1C=C...), Bond-dissociation energy: 15572, Retention time / Score: 9.131 (min) / Score: 99.18, Fragmenter score: 45.70, Database score: 50.00, Substructure score: 70.97, Resources: Higher_biosoc=Animalia Arthropoda Insecta[Animalia Cni], Family=Arecales Areaceae[Asparagales Asparagaceae]Aur, Biosoc=Acacia adunca[Acacia baileyana]Acacia calamifolia, Classifyre_class=Flavonoids, Classifyre_subclass=Flavans, Compound_level=1a, Internal_id=172494, Links=knapsackC00000956(cas:35323-91-2).

-  Show raw data spectrum
-  Show the isotopic ions
-  Show product ion
-  Show neutral loss ion
-  Show the result of fragment set enrichment analysis
-  Show assigned substructures

b. Batch analysis

You can also perform a batch analysis for both formula and structure searches. On the **“Analysis”** menu, select **“Compound annotation (Batch job)”**. You can process both formula and structure finders by selecting “Both processes” or choose to do them separately. Fix the number of Top formula hits to be processed by structure finder. *More than 2 hits greatly increase processing time.*



The screenshot shows the MS-FINDER ver. 3.30 interface. The 'Analysis' menu is open, and 'Compound annotation (batch job)' is selected. A 'Batch job setting' dialog box is displayed in the foreground. The dialog box has the following settings:

- ☒ Both processes (formula finder, and then structure finder)
- Top N hits: (automatically applied to structure finder program)
- ☒ Molecular formula finder
- ☒ Structure finder

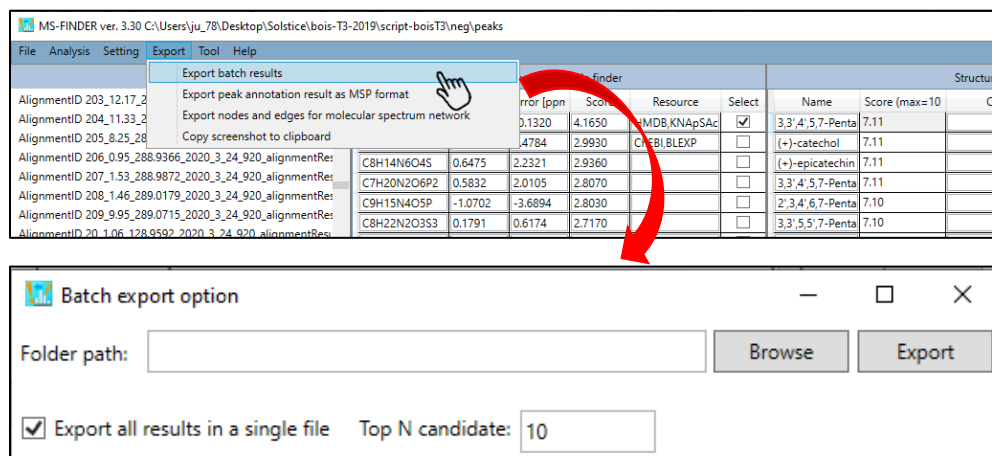
The background interface shows the 'Molecular formula finder' table with columns: Formula, Error [mD], Error [ppm], Score, Resource, and Select. The table contains two rows:

Formula	Error [mD]	Error [ppm]	Score	Resource	Select
C15H14O6	-0.0383	-0.1320	4.1650	HMDB, KnapSAC	<input checked="" type="checkbox"/>
C16H16O6	1.2991	4.4784	2.9930	ChEBI, BLEX	<input type="checkbox"/>

The 'Structure finder' table is also visible, with columns: Name, Score (max=10), Ontology, and InChIKey. The 'File information' section on the left shows details for the sample: (Metasci) (-)-Epicatechin, Scan number: 210, Retention time [min]: 8.81, Precursor m/z [Da]: 289.0718, Precursor type: [M-H]⁺, Ion mode: Negative, Spectrum type: Centroid, Collision energy: 40, Formula: C15H14O6, SMILES: OC3=CC(O)=C2C(OC(C=C1C=CC=C1C2)O)C3, InChIKey: PFTAWBLQPVEMU-UKRRQHH, Intensity: -1, MS1 peak number: 3, MS2 peak number: 117. The 'MS/MS spectrum' plot shows relative abundance versus m/z, with peaks labeled C6H5O2 and C15H9N4. The 'Resources' section on the right lists SMILES, Bond-dissociation energy, Retention time / Score, Fragmenter score, Database score, Substructure score, and Resources.

5. Export results

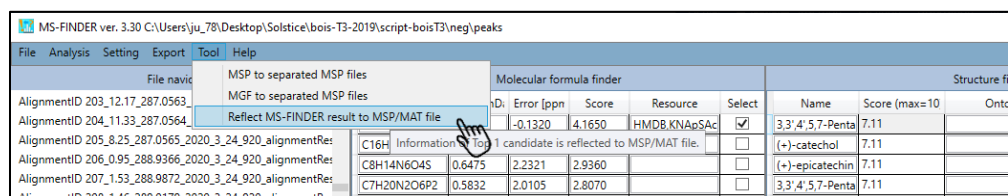
You can export Formula and Structure results on the menu “**Export**” by selecting “**Export batch results**”. Define the folder of arrival and fix the number of Top candidates you want to export.



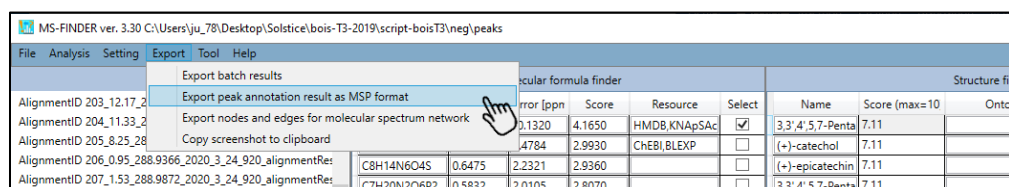
Two text format files will then be created:

Nom	Modifié le	Type	Taille
Formula result-2097	24/03/2020 13:38	Document texte	515 Ko
Structure result-2098	24/03/2020 14:43	Document texte	1 296 Ko

You can also choose to export results as .msp file. On the “**Tool**” menu select “**Reflect MS-FINDER results to MSP/MAT file**”.



Then on the “**Export**” menu, select “**Export peak annotation result as MSP format**”. Choose the folder of arrival and name your file.



Follow the MS-CleanR tutorial to merge results from MS-DIAL and MS-FINDER.