

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/>

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.

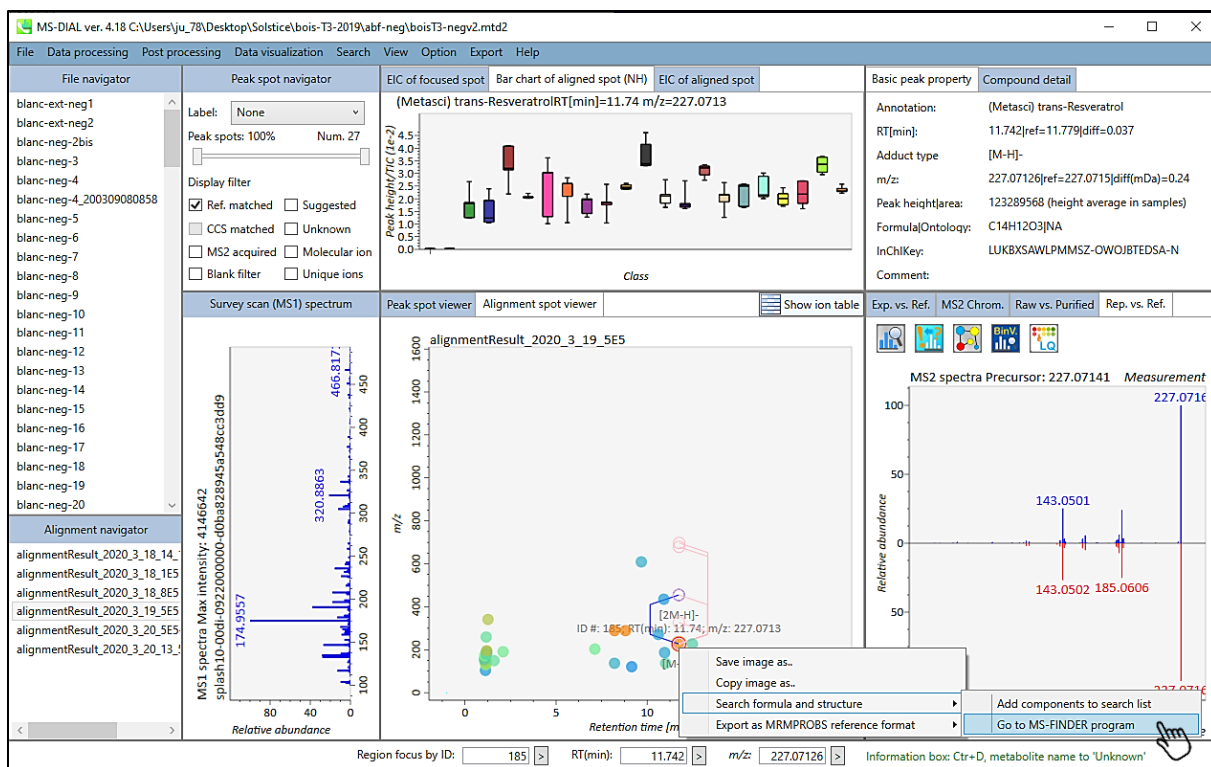
The screenshot displays the MS-DIAL software interface. The main window is divided into several panels. On the left, there is a 'File navigator' showing a list of files. The central area contains a 'Survey scan (MS1) spectrum' and a 'Peak spot viewer' showing a 2D plot of m/z vs. abundance. A right-click context menu is open over the peak spot viewer, with the option 'Add components to search list' highlighted. Below this, a dialog box titled 'Store MS annotation tag (MAT)' is shown, with the 'Path' field set to 'C:\Users\ju_78\Desktop\bois-T3-2019\script-boisT3\neg\peaks'. The 'Export option' is set to 'All peaks'. A red arrow points to the 'Export' button in the dialog box.

The screenshot displays the MS-FINDER software interface. The top menu bar includes File, Analysis, Setting, Export, Tool, and Help. The main window is divided into several sections:

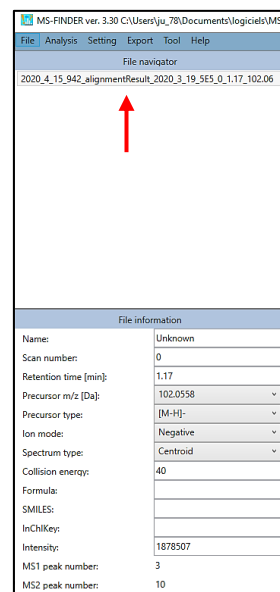
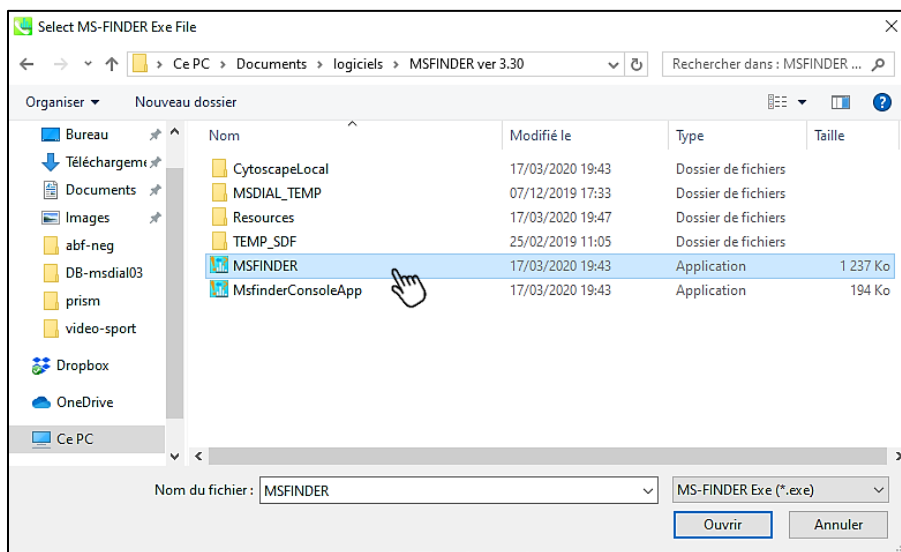
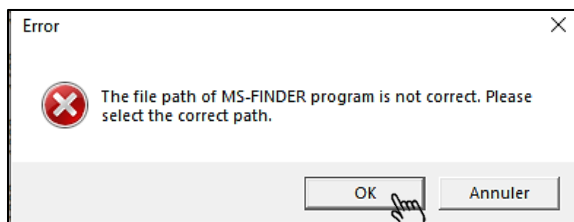
- Import...** (with a magnifying glass icon) and **Selector** tabs.
- Create a query** section with a text input field and a button labeled "Select a folder including MAT or MSP files."
- AlignmentID** list: A scrollable list of alignment IDs, with the first few visible as:
 - AlignmentID 100_1,23_207.0508_2020_3_24_920_alignmentRes
 - AlignmentID 102_1,2_209.0301_2020_3_24_920_alignmentRes
 - AlignmentID 103_1,35_209.0665_2020_3_24_920_alignmentRes
 - AlignmentID 104_1,21_209.0665_2020_3_24_920_alignmentRes
 - AlignmentID 105_12,17_211.0763_2020_3_24_920_alignmentRes
 - AlignmentID 106_14,65_211.0765_2020_3_24_920_alignmentRes
 - AlignmentID 107_1,4_212.0234_2020_3_24_920_alignmentRes
 - AlignmentID 108_1,06_212.9206_2020_3_24_920_alignmentRes
 - AlignmentID 109_1,19_215.0327_2020_3_24_920_alignmentRes
 - AlignmentID 10_1,46_115.0033_2020_3_24_920_alignmentRes
 - AlignmentID 110_0,56_216.8535_2020_3_24_920_alignmentRes
 - AlignmentID 111_1,16_216.9091_2020_3_24_920_alignmentRes
 - AlignmentID 112_1,18_217.048_2020_3_24_920_alignmentRes
- File information** section (indicated by a red arrow):
 - 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: (empty)
 - SMILES: (empty)
 - InChIKey: (empty)
 - Intensity: 1878507
 - MS1 peak number: 3
 - MS2 peak number: 10
- Molecular formula finder** section:
 - Formula: (empty)
 - Error [mDa]: (empty)
 - Error [ppm]: (empty)
 - Score: (empty)
 - Resource: (empty)
 - Select: (empty)
- Structure finder** section:
 - Name: (empty)
 - Score (max=10): (empty)
 - Ontology: (empty)
 - InChIKey: (empty)
- MS1 spectrum** plot: Shows relative abundance (0 to 100) versus m/z (95 to 110). A single prominent peak is visible at m/z 102.0558.
- MS/MS spectrum** plot: Shows relative abundance (0 to 100) versus m/z (50 to 100). A single prominent peak is visible at m/z 102.0558.
- Spectrum** section:
 - Structure: (empty)
 - Meta data: (empty)
- Measurement vs. Reference** plot: Shows relative abundance (0 to 100) versus m/z (0 to 100). The plot is titled "Actual MS/MS" and "In silico MS/MS".



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.



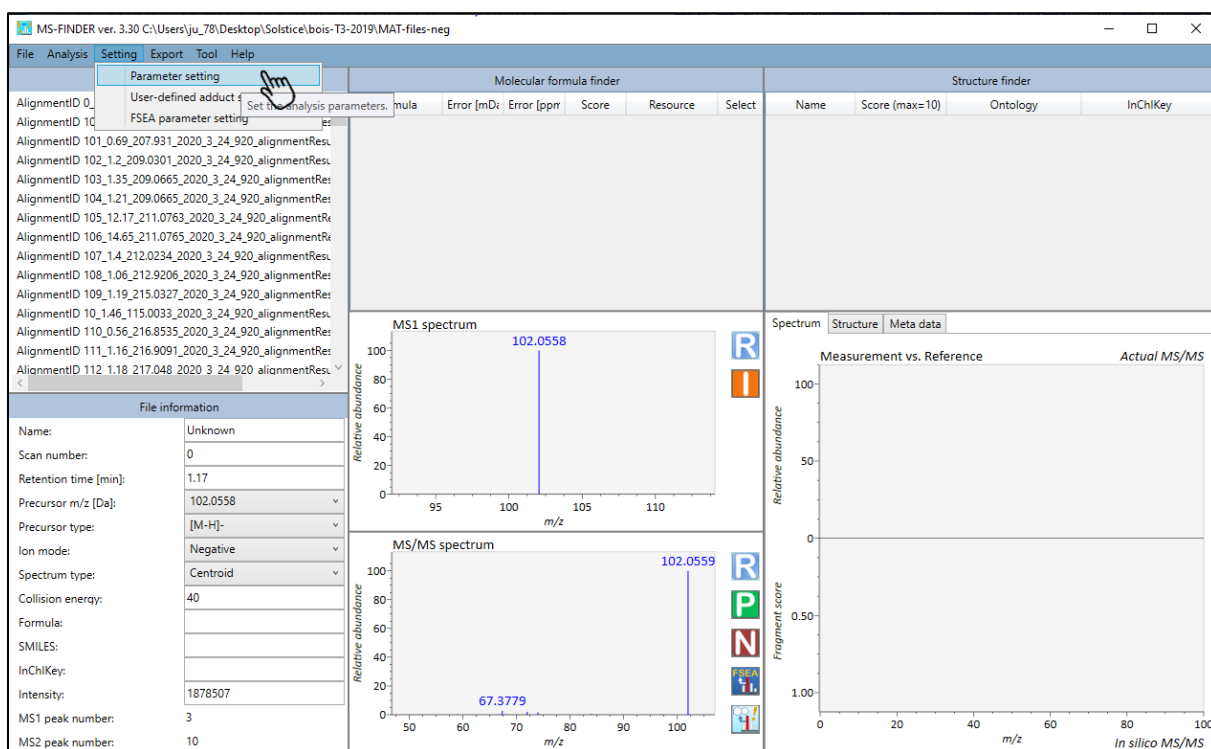
The following error window appears and you have to click on “OK” and go find the application MSFINDER.exe in your computer and open it.



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

3. Define search parameters

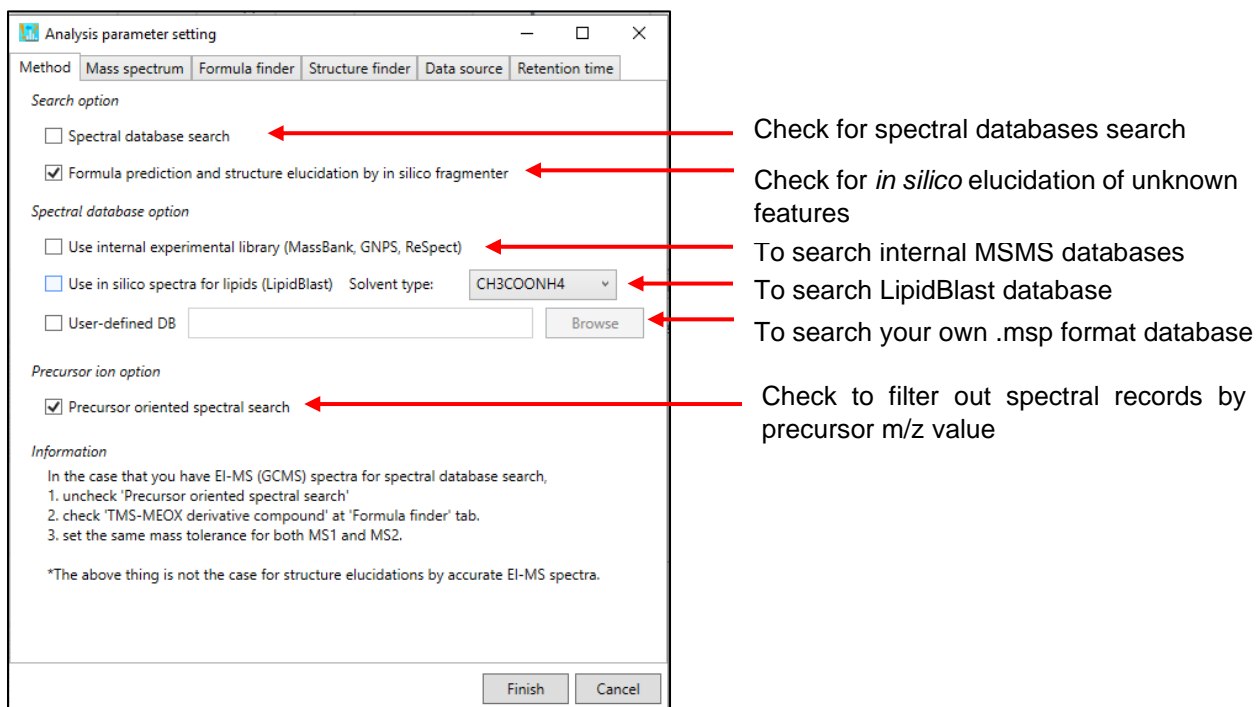
On the menu "Setting", select "Parameter setting" to define all the parameters of formula and structure search.



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



b. Mass spectrum parameter

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

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

The 'Analysis parameter setting' dialog box has several tabs: Method, Mass spectrum, Formula finder, Structure finder (selected), Data source, and Retention time. Under the 'Structure finder' tab, the 'In silico MS/MS or EI-MS fragmenter setting' section includes a 'Tree depth' input field with the value '2' and a range '[1-3]'. Below this are two unchecked checkboxes: 'Use the fragmentation library for electron ionization (EI)' and 'Use the fragmentation library for low energy CID'. The 'Options' section contains 'Maximum report number' set to '50' (with a note 'up to 100') and 'Time out (-1 means infinite):' set to '1' (with a note 'min'). At the bottom are 'Finish' and 'Cancel' buttons.

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

The 'Analysis parameter setting' dialog box has the 'Data source' tab selected. The 'Local Databases' section lists 22 databases with checkboxes: 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) [checked], ChEBI (Biomolecules) [checked], T3DB (Toxin), STOFF (Environment), BLEX (blood exposome), KNApSack (Natural product) [checked], NNPDB (Natural product) [checked], PubChem (Biomolecules), UNPD (Natural product) [checked], and User-defined DB [checked] with a file path 'C:\Users\ju_78\Desktop\bois-T3-2019\BDD-vigne\Pr' and a 'Browse' button. The 'MINEs (Metabolic In silico Network Expansions) setting' section has three radio buttons: 'Never use it.' (selected), 'Only use when there is no query in local DBs.', and 'Always use it.'. The 'PubChem Online setting' section also has three radio buttons: 'Never use it.' (selected), 'Only use when there is no query in local DBs.', and 'Always use it.'. At the bottom are 'Finish' and 'Cancel' buttons.

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.

Analysis parameter setting

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

☒ Use predicted RT/RI values for structure elucidation

RT tolerance: min

RT resource: ☐ From user-library ☒ Calculated by XLogP based equation

User-defined RT-InChIKey library

File path: Browse

XLogP based RT prediction and cut off setting for structure elucidation

File path: Browse **Load**

Summary: 1.586* XLogP + 6.813 (min); R-squared: 0.71057

Retention time setting for spectral searching

☐ Use experimental RT/RI values for searching

☐ As retention index (RI)

RT(min) / RI tolerance:

☐ Use retention time to 'exclude' candidates

Finish Cancel

RT-MSFINDER-IRGA - Bloc notes

Name	RT (min)	SMILES
ALLOTHREONINE	1.16655	C[C@H](O)[C@H](N)C(O)=O
AMINOADIPATE	1.316817	NC(CCCC(O)=O)C(O)=O
ANISERINE	1.061133	CN1C=NC=C1C[C@H](NC(=O)CCN)C(C
ARABITOL	1.23445	OC[C@H](O)C(O)[C@H](O)CO
ARGININE	1.069483	N[C@H](CCCN=C(N)N)C(O)=O
ASPARAGINE	1.139567	N[C@H](CC(N)=O)C(O)=O
ASPARTATE	1.157167	N[C@H](CC(O)=O)C(O)=O
CAFFEATE	8.76485	OC(=O)C=C/C=C(C(O)C)=C1
CARNOSINE	1.055267	NC(=O)N[C@H](CC1=CN=CN1)C(C
CIS-4-HYDROXY-D-PROLINE	1.17475	O[C@H]1CN[C@H](C1)C(O)=O
CITRAMALATE	3.323433	CC(O)(CC(O)=O)C(O)=O
CITRULLINE	1.22105	N[C@H](CCCN(N)=O)C(O)=O
CORTISONE	12.776	[H][C@]12CC[C@](O)(C(=O)CO)[C@]1(C)C
CREATINE	1.306933	CN(CC(O)=O)C(N)=N
CREATININE	1.21425	CN1CC(=O)NC1=N
CYCLICAMP	5.269484	NC1=NC=NC2=C1N=CN2[C@H]10[C@
CYSTATHIONINE	1.108667	N[C@H](CCSC[C@H](N)C(O)=O)C(C
CYSTEATE	1.1794	NC(CS(O)=O)C(O)=O
CYSTINE	1.11125	N[C@H](CSC[C@H](N)C(O)=O)C(O)=O
D-ALANINE	1.158167	C[C@H](N)C(O)=O
DETHIOBIOTIN	9.591784	C[C@H]1NC(=O)N[C@H]1CCCCC(C
DIAMINOPIMELATE	1.0866	N[C@H](CCC[C@H](N)C(O)=O)C(O)=O
D-MANNOSAMINE	1.074033	C(C1C(C(C(C(O1)N)O)O)O)O
D-ORNITHINE	1.002467	NC(=O)N[C@H](N)C(O)=O
FERULATE	10.27797	CCOC1=C(O)C=CC(\C=C\C(O)=O)=C1
FOLATE	7.969867	NC1=NC(=O)C2=NC(CNC3=CC=C(C(=C3)C(=O)N[
GALACTITOL	1.206083	OC[C@H](O)[C@H](O)[C@H](O)[C
GLUCOSAMINATE	1.124483	C(C(C(C(C(C(=O)O)N)O)O)O)O

RT prediction result

Reflect Cancel

Name	SMILES	XLogP	Exp. RT	Pred. RT
ALLOTHREON	O=C(O)C(N)C(O)C	-3.50	1.17	1.27
AMINOADIP	O=C(O)CCC(N)C(=O)O	-2.99	1.32	2.07
ANISERINE	O=C(O)C(NC(=O)CCN)CC1=CN=CN1C	-3.80	1.06	0.79
ARABITOL	OC[C@H](O)C(O)CO	-3.22	1.23	1.70
ARGININE	O=C(O)C(N)CCN=C(N)N	-3.59	1.07	1.13
ASPARAGINE	O=C(O)C(N)CC(=O)N	-4.44	1.14	-0.22
ASPARTATE	O=C(O)C(N)C(=O)O	-3.71	1.16	0.93
CAFFEATE	O=C(O)C=C1=CC=C(C(O)C)=C1	1.26	8.76	8.81
CARNOSINE	O=C(O)C(NC(=O)CCN)CC1=CN=CN1	-4.10	1.06	0.31
CIS-4-HYDRO	O=C(O)C1NCC(O)C1	-0.79	1.17	5.57
CITRAMALAT	O=C(O)C(C(O)C(=O)O)C	-1.36	3.32	4.65
CITRULLINE	O=C(N)NCCC(N)C(=O)O	-3.91	1.22	0.61
CORTISONE	O=C4C=C3CCC2C(C(=O)CC1(C)C2(CCC1(O)C)=C	0.14	12.78	7.04
CREATINE	O=C(O)C(N)C(N)C	-2.40	1.31	3.00
CREATININE	O=C1NC(=N)N(C)C1	0.03	1.21	6.87
CYCLICAMP	O=P3(O)OCG4OC(N2C=NC=1C(=NC=NC=12)N	-2.49	5.27	2.87
CYSTATHIONI	O=C(O)C(N)CCSC(N)C(=O)O	-3.21	1.11	1.72
CYSTEATE	O=C(O)C(N)C(S(=O)(=O)O)O	-4.36	1.18	-0.10
CYSTINE	O=C(O)C(N)C(SCC(N)C(=O)O)O	-3.31	1.11	1.56
D-ALANINE	O=C(O)C(N)C	-2.82	1.16	2.33
DETHIOBIOTI	O=C1NC(C)C(N1)CCCCC(=O)O	0.82	9.59	8.11
DIAMINOPIM	O=C(O)C(N)CCCC(N)C(=O)O	-3.46	1.09	1.32
D-MANNOSA	OC[C@H](O)C(N)C(O)C1(O)	-1.76	1.07	4.01
D-ORNITHINE	O=C(O)C(N)CCN	-3.31	1.00	1.57
FERULATE	O=C(O)C=CC(=C)C(=O)C(=O)C=C1	1.15	10.28	8.64
FOLATE	O=C(O)CCC(NC(=O)C1=CC=C(C(=C1)NCC=2N=C	-2.25	7.97	3.24
GALACTITOL	OC[C@H](O)C(O)C(O)CO	-3.90	1.21	0.63
GLUCOSAMIN	O=C(O)C(N)C(O)C(O)CO	-5.97	1.12	-2.66
GLUCOCHOLA	O=C(O)C(NC(=O)CCCC(C)C2CCC3C4C(O)CC1CC(O)K	3.27	14.03	11.99
GUANIDINOA	O=C(O)C(NC(=N)N	-2.97	1.20	2.11
HIPPURATE	O=C(O)C(NC(=O)C1=CC=CC=C1	0.81	8.44	8.10
HISTIDINE	O=C(N)C(NC(=O)C1=CC=CC=C1	-3.12	1.05	1.87

Equation: 1.586* XLogP + 6.813

R-squared: 0.71057

RT error Max (min): 6.254

RT error Min (min): 0.015

Prediction result

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. The 'File navigator' on the left lists alignment files. The 'Molecular formula finder' table in the center shows candidates for formula C₁₅H₁₄O₆, with the top result having a score of 7.11. The 'Structure finder' table on the right lists structure candidates, with the top result being (+)-epicatechin. Below these tables, the 'MS1 spectrum' and 'MS2 spectrum' are shown, along with a comparison of the 'Experimental spectrum vs. In silico spectrum'. The 'File information' panel on the left provides metadata for the selected file, including the name, scan number, retention time, precursor m/z, precursor type, ion mode, spectrum type, collision energy, formula, SMILES, InChIKey, intensity, MS1 peak number, and MS2 peak number.

Panel (A) shows the 'Experimental spectrum vs. In silico spectrum' plot, comparing the experimental spectrum (blue) with the in silico spectrum (red). The x-axis is m/z (50 to 300) and the y-axis is Relative abundance (0 to 100). The plot shows several peaks, with the base peak at m/z 289.072. The in silico spectrum is labeled 'Actual MS/MS'.

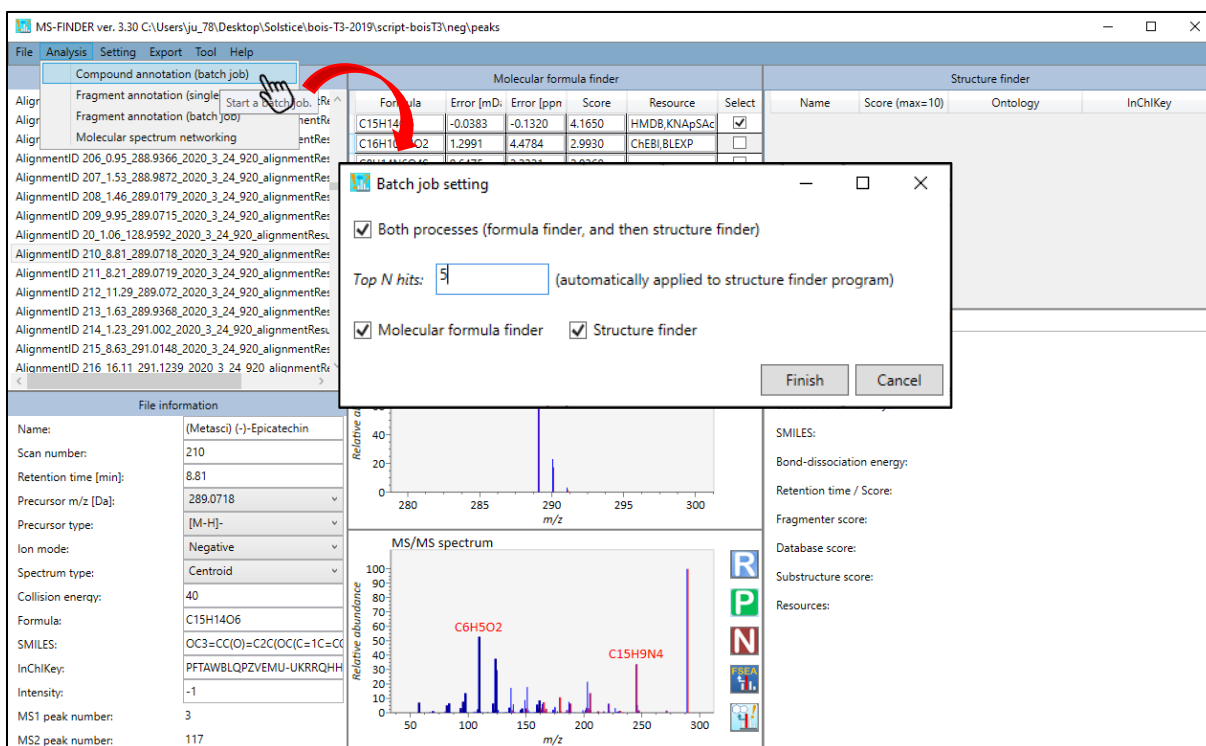
Panel (B) shows the chemical structure of (+)-epicatechin, a flavan-3-ol. The structure is a catechin derivative with a phenyl ring and a pyrogallol B-ring.

Panel (C) shows the 'Meta data' for (+)-epicatechin. The data includes the Name, InChIKey, Substructure inChIKeys, SMILES, Bond-dissociation energy, Retention time / Score, Fragmenter score, Database score, Substructure score, and Resources. The SMILES is OC3=CC(O)=C2C(OC(C=1C=CC=C1)C2)=C3. The Retention time is 9.131 (min) and the Score is 99.18. The Fragmenter score is 45.70, the Database score is 50.00, and the Substructure score is 70.97. The Resources include Higher_biosoc=Animalia Arthropoda Insecta|Animalia Cni|Family=Arecales Arecaceae|Asparagales Asparagaceae|Asteraceae|Acacia adunca|Acacia baileyana|Acacia calamifolia|Classifyre_class=Flavonoids|Classifyre_subclass=Flavans|Compound_level=1a|Internal_id=172494|Links=knapsackC00000956|cas35323-91-2.

-  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.*



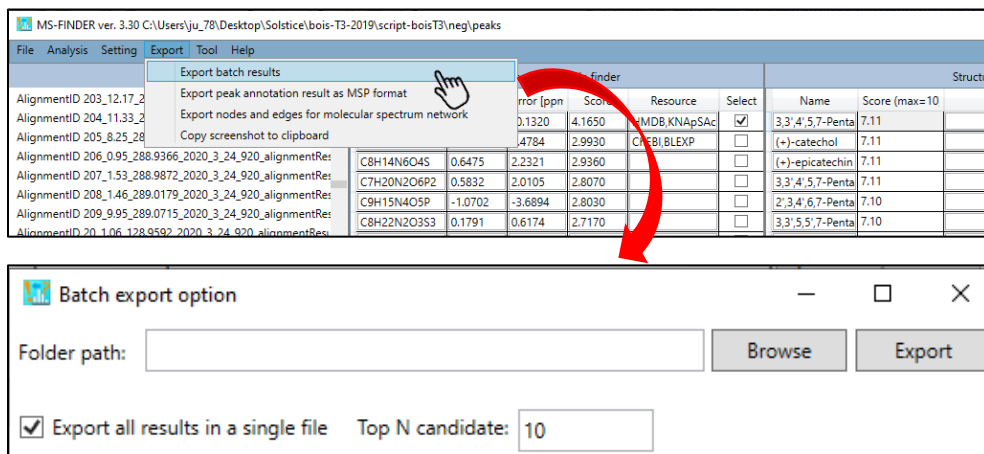
The screenshot shows the MS-FINDER ver. 3.30 interface. The 'Analysis' menu is open, and 'Compound annotation (batch job)' is selected. A 'Batch job setting' dialog box is displayed in the foreground. The dialog box has the following settings:

- ☒ 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 'Structure finder' table has columns: Name, Score (max=10), Ontology, and InChIKey. The 'File information' section on the left displays 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: PFTAWBLQPZVEMU-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

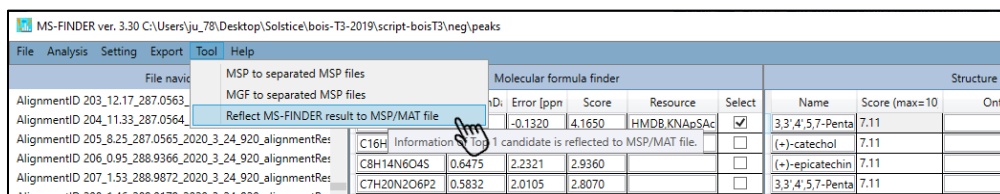
You can export Formula and Structure results on the menu “**Export**” by selecting “**Export batch results**”. Define the folder of arrival and fix the number of Top candidates you want to export.



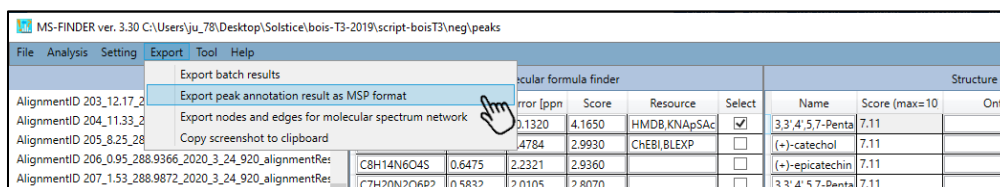
Two text format files will then be created:

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.