

Manual for VirMolAnalyte

VirMolAnalyte AI-Driven Metabolite Identification Tool

Introduction

Task

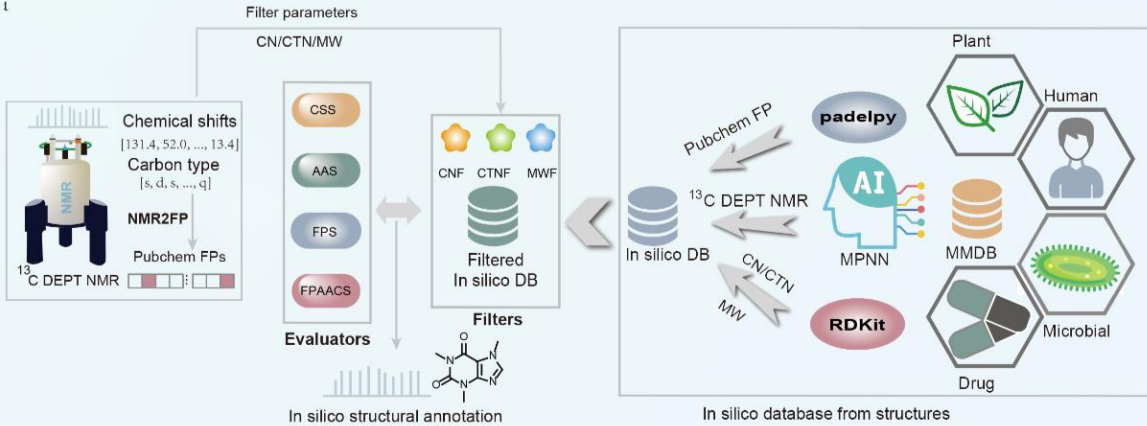
Download

Contact Us

VirMolAnalyte: Metabolite Identification Tool

Introduction

VirMolAnalyte is an advanced AI-driven tool designed for identifying metabolites from plant, human, microbial, and drug sources. Based on ^{13}C NMR (Nuclear Magnetic Resonance) data, this tool aims to provide researchers with an efficient and accurate solution for metabolite analysis.



The workflow diagram illustrates the process of metabolite identification. It begins with input data including NMR spectra, chemical shifts, carbon types, NMR2FPs, and Pubchem FPs. These are processed through a series of filters (CNF, CTNF, MWF) and evaluators (CSS, AAS, FPS, FPAACS) to generate an in silico structural annotation. The tool also utilizes an in silico database from structures, which includes Pubchem FP, padelpy, AI (MPNN), RDKit, and MMDB. The database is categorized by source: Plant, Human, Microbial, and Drug. The final output is an in silico structural annotation, represented by a chemical structure and a spectrum.

Features

Database selection<Filter selection<Evaluator selection<NMR data input<Analyze

VirMolAnalyte AI-Driven Metabolite Identification Tool

Introduction

Task

Download

Contact Us

Database Selection

AIIDB

Filter Selection

☒ CNF 5

☒ CTNF 2

☐ MWF 3

Evaluator Selection

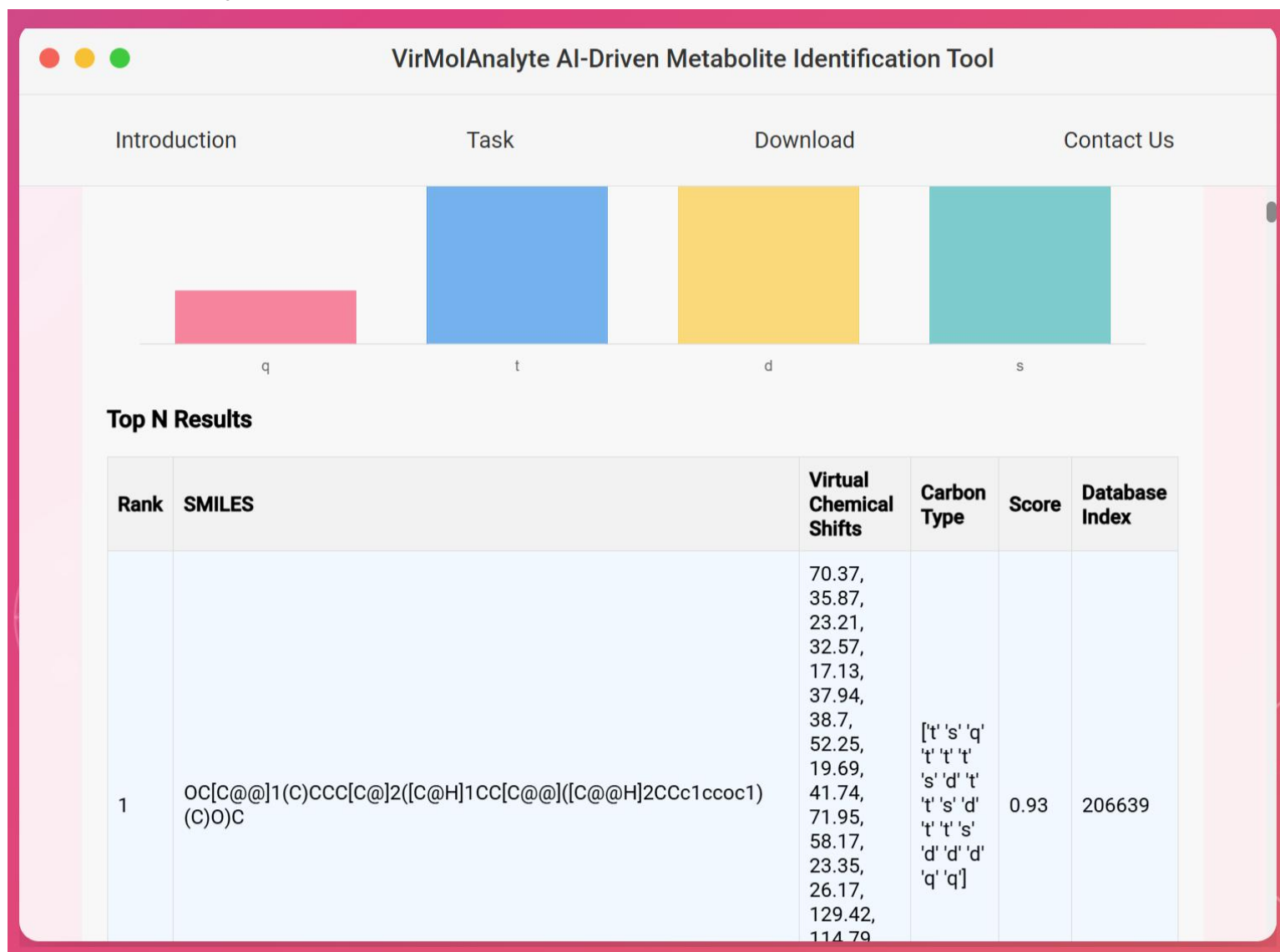
FPS

13. 31, q
66. 29, t
53. 37, t
40. 84, t
38. 16, t
35. 71, t
26. 03, t
23. 07, t
20. 61, t
140. 56, d
108. 29, d

Use Example Data

Analyze

Result analysis



VirMolAnalyte AI-Driven Metabolite Identification Tool

Introduction

Task

Download

Contact Us

Download

Software Download

We recommend downloading and installing the software (Contain all databases) through the following link: [Download VirMolAnalyte software](#)

For detailed instructions on how to use VirMolAnalyte, please download the user manual: [Download User Manual](#)

The full version of the software offers additional functions, including:

- Automatic peak recognition and impurity removal based on raw data
- Rapid construction of in silico databases tailored to user requirements

These advanced capabilities will significantly enhance your NMR data analysis workflow, saving time and improving accuracy. Download now to experience the full potential of our NMR analysis tools!

Source Code

We also provide the source code for the software, which is open-source and free to use. The source code is available via link: [Download Source Code](#)

In Silico Metabolite Databases

We offer a range of specialized in silico metabolite databases to support your research. Each database is curated to

Installation of VirMolAnalyte

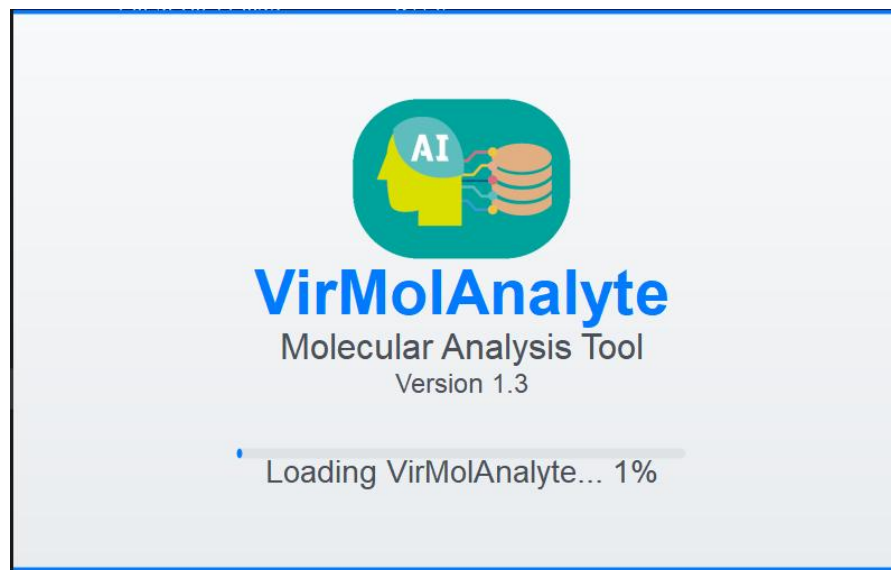


1. Extract the installation package.



名称	修改日期	类型	大小
_internal	2025/10/4 18:51	文件夹	
Database	2025/10/4 20:33	文件夹	
GUI_result_files	2025/10/4 20:50	文件夹	
NMR2FP	2025/10/4 20:51	文件夹	
test	2025/10/4 20:33	文件夹	
VirMolAnalyte	2025/10/4 18:49	文件夹	
1d_pipe.fid	2025/10/4 20:54	FID 文件	1,026 KB
1d_pipe90.fid	2025/10/4 20:54	FID 文件	1,026 KB
1d_pipe135.fid	2025/10/4 20:54	FID 文件	1,026 KB
Logo	2024/10/23 3:46	PNG 文件	46 KB
NMR-1D	2025/10/4 20:54	XLS 工作表	1 KB
Result	2025/10/4 20:51	XLS 工作表	43 KB
VirMolAnalyte	2025/10/4 18:29	应用程序	33,260 KB

2. Click **VirMolAnalyte.exe** to run the software



STEP1: Select the path of ^{13}C NMR, DEPT 90, DEPT 135 spectra.

VirMolAnalyte - Molecular Analysis Tool

File Edit View Tools Help

Molecular Analysis Database Management

Spectral Data Input

^{13}C _NMR: test\NMRsample\4\data\1

DEPT90: test\NMRsample\8\data\1

DEPT135: test\NMRsample\6\data\1

Peak Detection Parameters

Impurity Signal Removal

Database Analysis

NMR Spectrum Display

Peak Detection Plot Merged Spectrum

Peak Information (NMR-ID.csv)

Chemical Shift (pp)	Carbon Type	Intensity
---------------------	-------------	-----------

Analysis Results

Compound ID	SMILES	Score	Chemical Shifts	Carbon Types
-------------	--------	-------	-----------------	--------------

Selected Compound Details

SMILES:

Copy SMILES

Memory: 353.9 MB

STEP2: Adjust the threshold to extract peaks

Set threshold<Submit

VirMolAnalyte - Molecular Analysis Tool

File Edit View Tools Help

Molecular Analysis Database Management

Spectral Data Input

Peak Detection Parameters

C: 40000000 C90: 90000000

C135pos: 50000000 C135neg: -50000000

Submit Merge

Impurity Signal Removal

Database Analysis

NMR Spectrum Display

Peak Detection Plot Merged Spectrum

Peak Information (NMR-ID.csv)

Chemical Shift (pp)	Carbon Type	Intensity
---------------------	-------------	-----------

Analysis Results

Compound ID	SMILES	Score	Chemical Shifts	Carbon Types
-------------	--------	-------	-----------------	--------------

Selected Compound Details

SMILES:

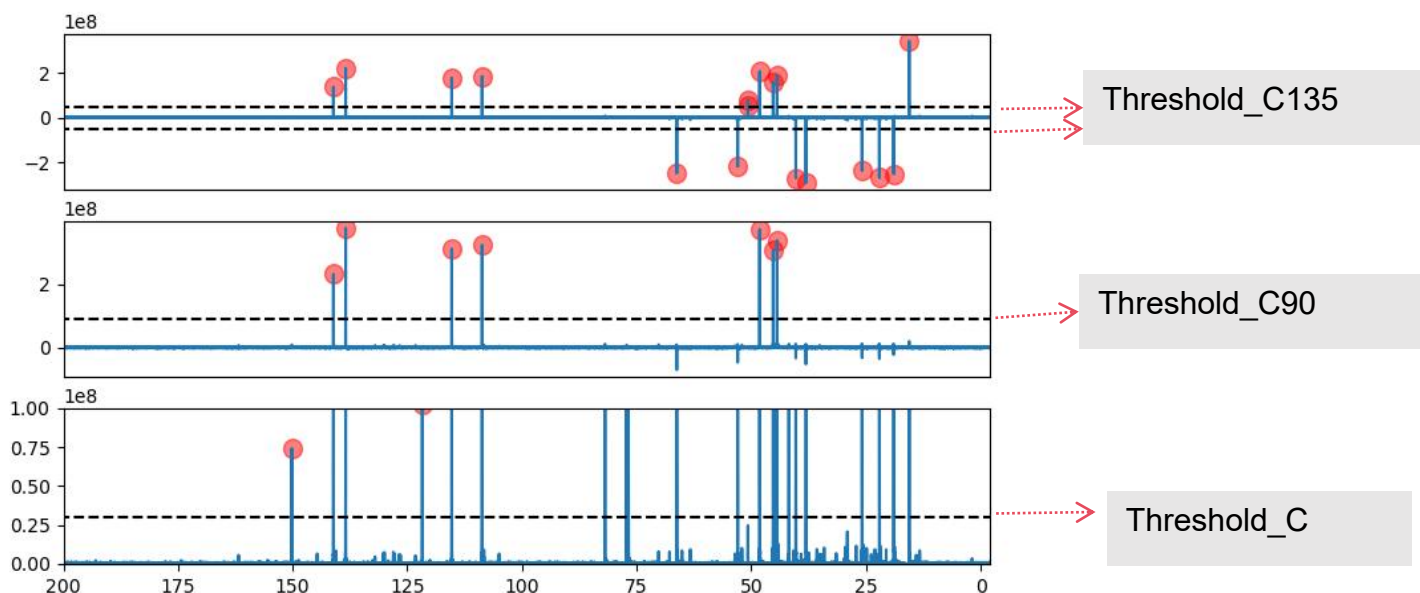
Copy SMILES

Peak detection completed

Memory: 458.7 MB

STEP2: Adjust the threshold to extract peaks

After clicking **submit**, the tool will extract peaks from the raw data and adjust the threshold to ensure reasonable peak extraction



STEP3: Merge spectra to establish ^{13}C DEPT NMR spectra

VirMolAnalyte - Molecular Analysis Tool

File Edit View Tools Help

Molecular Analysis

Database Management

Spectral Data Input

Peak Detection Parameters

C: 40000000

C90: 90000000

C135pos: 50000000

C135neg: -50000000

Submit

Merge

Impurity Signal Removal

Database Analysis

NMR Spectrum Display

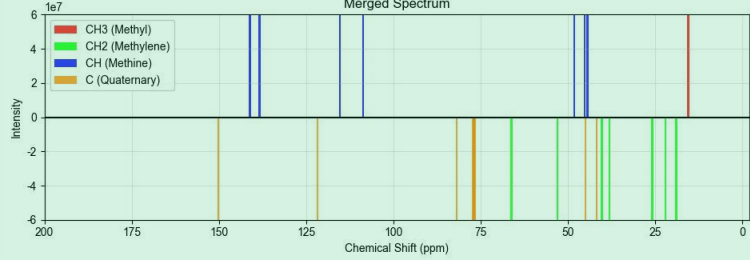
Peak Detection Plot

Merged Spectrum

Peak Information (NMR-ID.csv)

	ppm	Type	Intensity
1	81.89	C	127,850,362
2	77.21	C	179,628,976
3	77.00	C	185,302,593
4	76.79	C	187,153,731
5	66.24	CH2	280,649,585
6	52.97	CH2	310,244,657
7	48.19	CH	382,401,015

Merged Spectrum



Analysis Results

Compound ID	SMILES	Score	Chemical Shifts	Carbon Types

Selected Compound Details

SMILES:

Copy SMILES

Data merge completed

Memory: 460.7 MB

STEP4: Remove the signal peak of the solvent.

Select the solvent to be used in the experiment from the drop-down menu of solvents and choose the carbon signal category to be removed, then **submit**

Molecular Analysis

Database Management

Peak Detection Parameters

Impurity Signal Removal

Solvent Removal

Impurity Removal

Database Analysis

Solvent: Chloroform

Type: CH3

Submit

Type: CH3

Threshold: ... e7

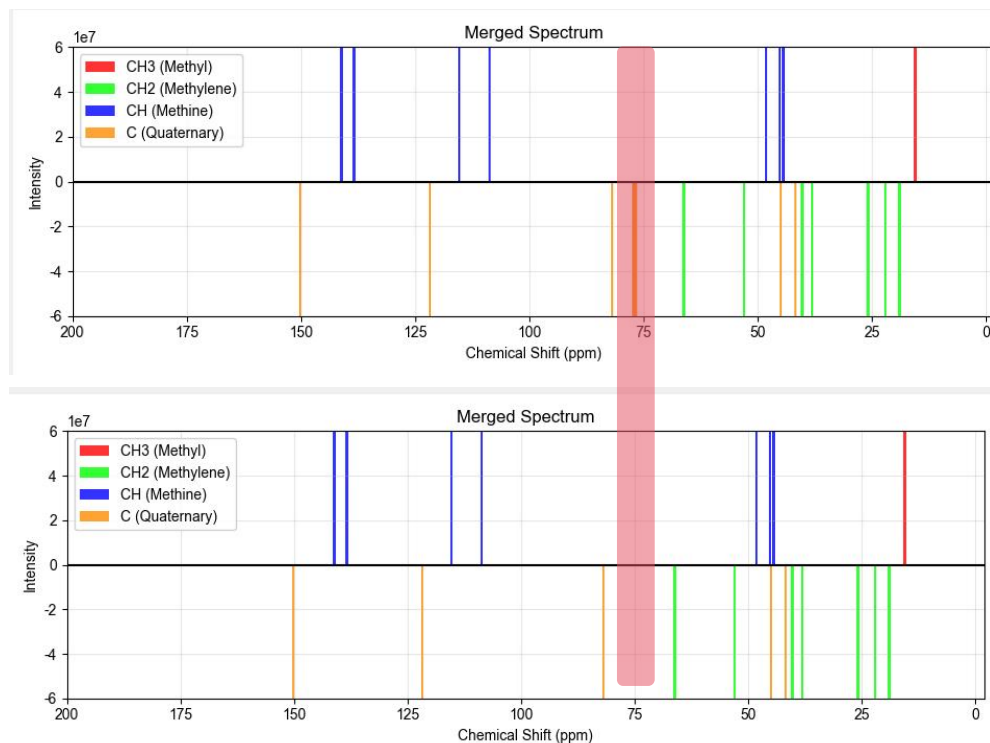
Submit

CH3 (Methyl)

CH2 (Methylene)

CH (Methine)

C (Quaternary)



Note: Pay attention to distinguishing the carbon categories of solvent peaks

STEP4: Remove impurities

For a mixture where the impurity signal differs from the target compound, impurities can be removed by establishing a carbon-atom-based threshold.

VirMolAnalyte - Molecular Analysis Tool

File Edit View Tools Help

Molecular Analysis Database Management

Peak Detection Parameters

Impurity Signal Removal

Solvent Removal

Solvent: Chloroform Type: C Submit

Impurity Removal

Type: CH3 Threshold: ... e7 Submit

Database Analysis

Merged Spectrum

Intensity

Chemical Shift (ppm)

CH3 (Methyl)
CH2 (Methylene)
CH (Methine)
C (Quaternary)

Peak Information (NMR-ID.csv)

	ppm	Type	Intensity
1	81.89	C	127,850,362
2	66.24	CH2	280,649,585
3	52.97	CH2	310,244,657
4	48.19	CH	382,401,015
5	45.24	CH	304,772,004
6	45.00	C	144,174,257
7	44.39	CH	366,120,717

Analysis Results

Compound ID	SMILES	Score	Chemical Shifts	Carbon Types
-------------	--------	-------	-----------------	--------------

Selected Compound Details

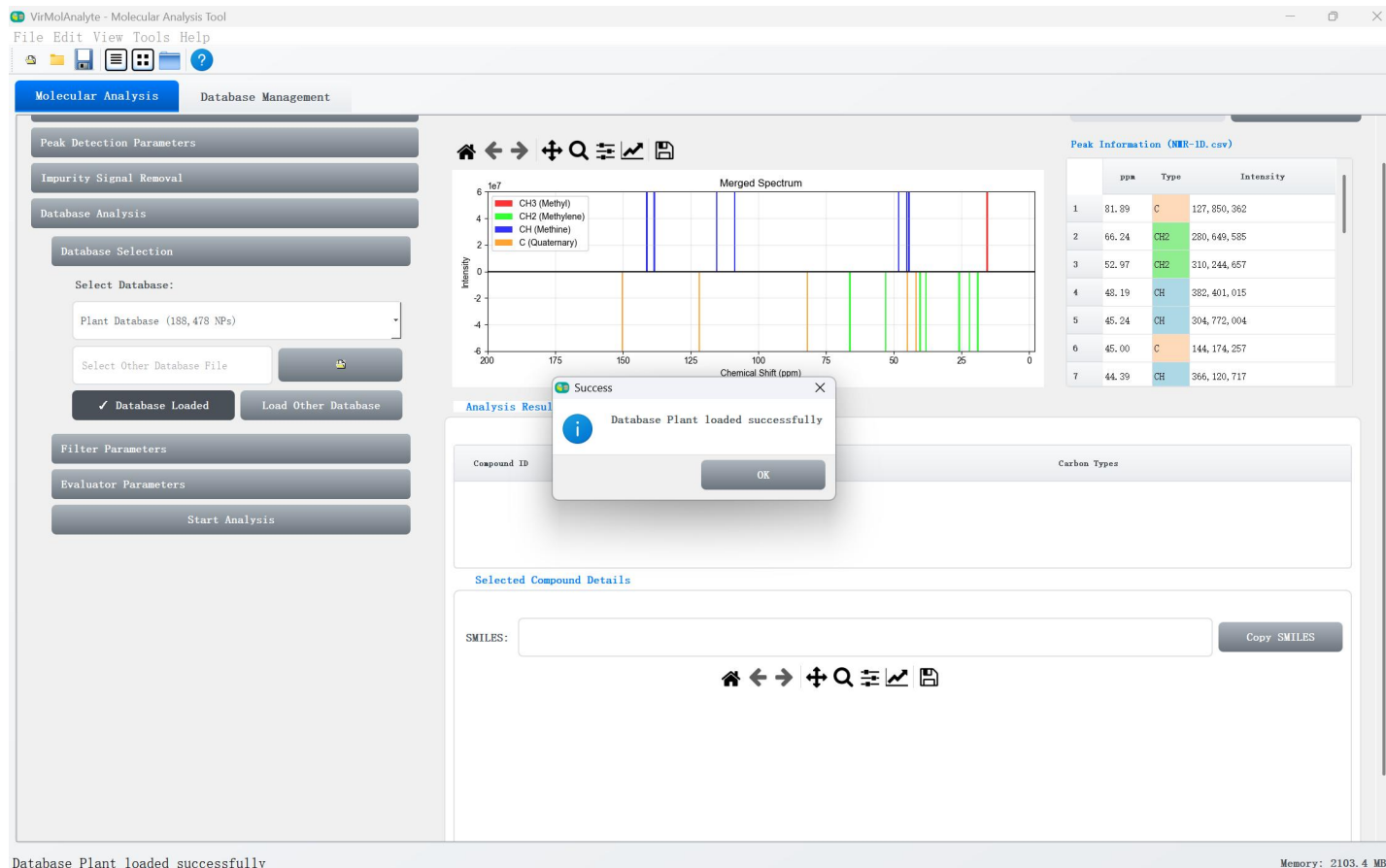
SMILES: Copy SMILES

Solvent removal completed

Memory: 463.0 MB

STEP5: Load database

Select the appropriate database and click 'load database' to load the database;



The screenshot displays the VirMolAnalyte - Molecular Analysis Tool interface. On the left, the 'Database Selection' panel shows the 'Plant Database (188,478 NPs)' selected. A 'Load Other Database' button is visible. The main window shows a 'Merged Spectrum' plot with peaks categorized by carbon type: CH3 (Methyl), CH2 (Methylene), CH (Methine), and C (Quaternary). A 'Peak Information (NMR-ID.csv)' table is displayed on the right. A 'Success' dialog box indicates 'Database Plant loaded successfully'. The 'Analysis Result' section shows 'Compound ID' and 'Carbon Types'. The 'Selected Compound Details' section includes a 'SMILES' input field and a 'Copy SMILES' button. The status bar at the bottom indicates 'Database Plant loaded successfully' and 'Memory: 2103.4 MB'.

	ppm	Type	Intensity
1	81.89	C	127,850,362
2	66.24	CH2	280,649,585
3	52.97	CH2	310,244,657
4	48.19	CH	382,401,015
5	45.24	CH	304,772,004
6	45.00	C	144,174,257
7	44.39	CH	366,120,717

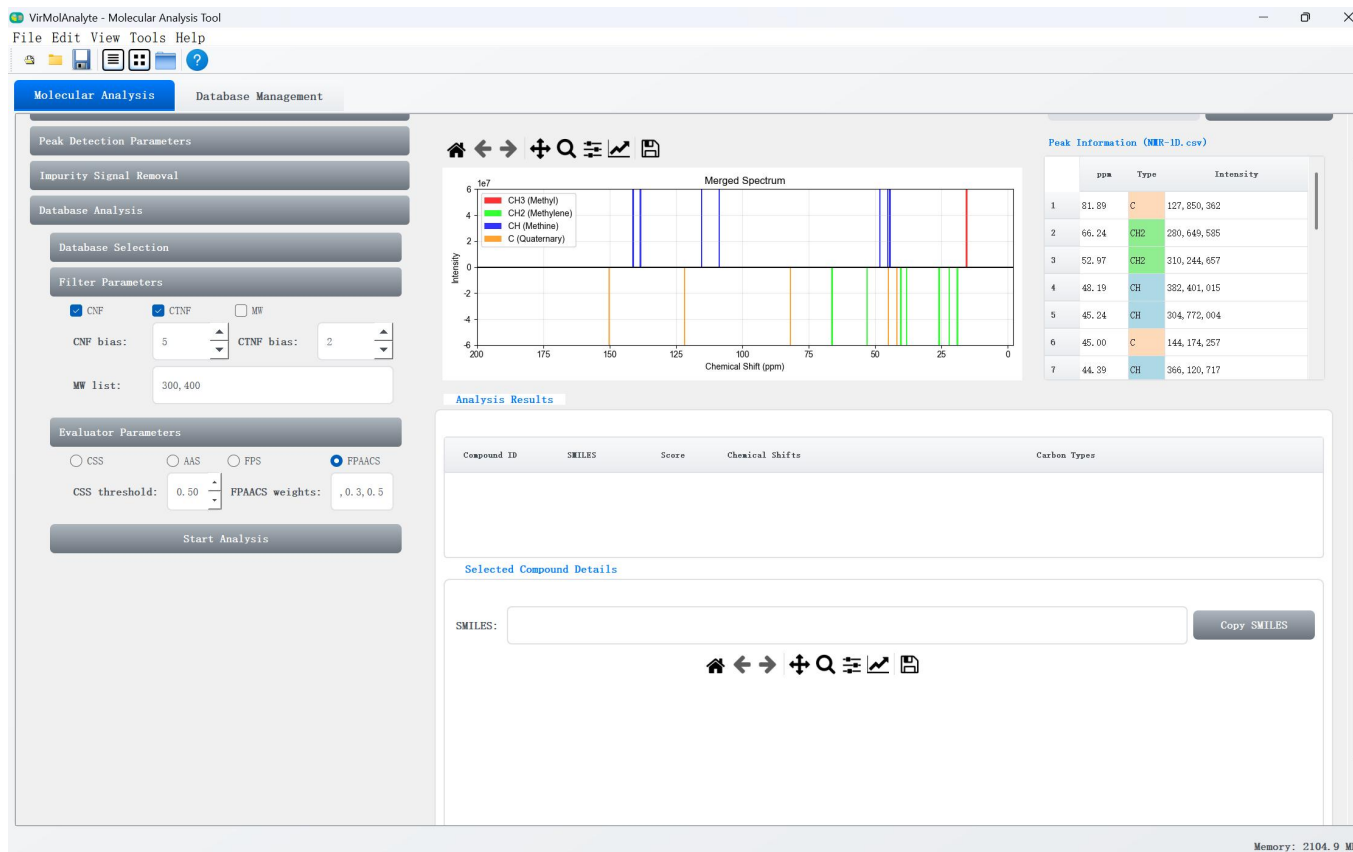
All database:
650,375 metabolites
Plant database:
188,478 metabolites
Human database:
217,347 metabolites
Microorganism:
36,427 metabolites

Note: The database only needs to be loaded during the first use, and subsequent tasks do not require repeated loading

STEP6: Set database search parameters

Select filter (Multiple Choice): By default, CNF (Number of Carbon Filters) and CTNF (Number of Carbon Filter Types) are selected. When the compound has an MS spectrum, MW can be checked and possible molecular weights such as "312, 345, 413" can be entered

Select evaluator (Single Choice): It is recommended to prioritize the FPAACS evaluator.
recommended FPAACS_weight: 0.36, 0.16, 0.48



The screenshot displays the VirMolAnalyte - Molecular Analysis Tool interface. The 'Molecular Analysis' tab is active, showing various parameters for peak detection and evaluation. The 'Filter Parameters' section includes checkboxes for CNF, CTNF, and MW, with CNF and CTNF selected. The 'Evaluator Parameters' section shows the FPAACS evaluator selected. The 'Merged Spectrum' plot shows intensity versus chemical shift (ppm) with peaks color-coded by carbon type: CH3 (Methyl), CH2 (Methylene), CH (Methine), and C (Quaternary). The 'Peak Information (NMR-ID.csv)' table lists peak data.

	ppm	Type	Intensity
1	81.89	C	127,850,362
2	66.24	CH2	280,649,585
3	52.97	CH2	310,244,657
4	48.19	CH	382,401,015
5	45.24	CH	304,772,004
6	45.00	C	144,174,257
7	44.39	CH	366,120,717

The 'Analysis Results' section shows a table with columns: Compound ID, SMILES, Score, Chemical Shifts, and Carbon Types. The 'Selected Compound Details' section shows a text input for SMILES and a 'Copy SMILES' button.

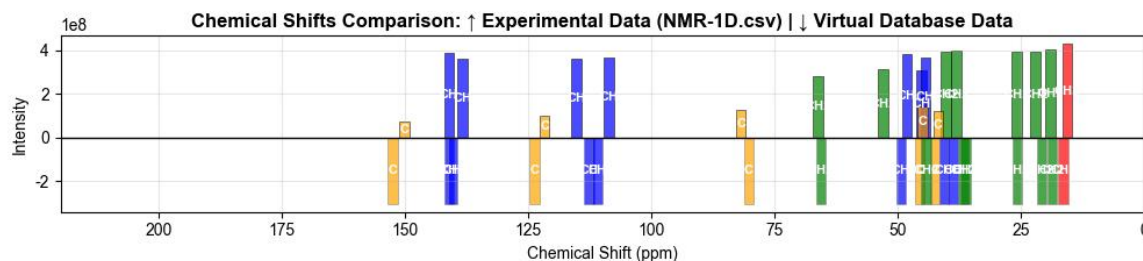
Click the **Start Analysis** button to perform database analysis

Memory: 2111.3 MB

After clicking on the result entry, you can see the structure and comparison between the experimental ^{13}C DEPT NMR and the virtual ^{13}C DEPT NMR

	Compound ID	SMILES	Score	Chemical Shifts (ppm)	Carbon Types
1	4004	<chem>CC12C=Cc3occc3C1CCCC13CC(CCC12)C(O)(CO)C3</chem>	65.002	[16.38527 42.486748 141.12566 112.7092 152.52473 140.4448 110.979485 123.7509 40.602833 20.820316 36.719612 45.52128...	['q' 's' 'd' 'd' 's' 'd' 'd' 's'...
2	165069	<chem>CC1=CCCC2C1(C)CC(O)C(C)C2(CO)CCc1ccoc1</chem>	62.302	[20.624367 141.7491 121.40474 26.313866 18.379913 43.30023 39.38704 24.387255 44.966553 67.609764 38.62079 12.806296...	['q' 's' 'd' 't' 't' 'd' 's' 'q'...
3	3078	<chem>CC1CCC2C=CC3CCC1CCC2(CCC12)C(O)(CO)C3</chem>	60.046	[16.387362 36.65864 34.748543 18.326387 150.30975 140.08382	['q' 's' 't' 't'

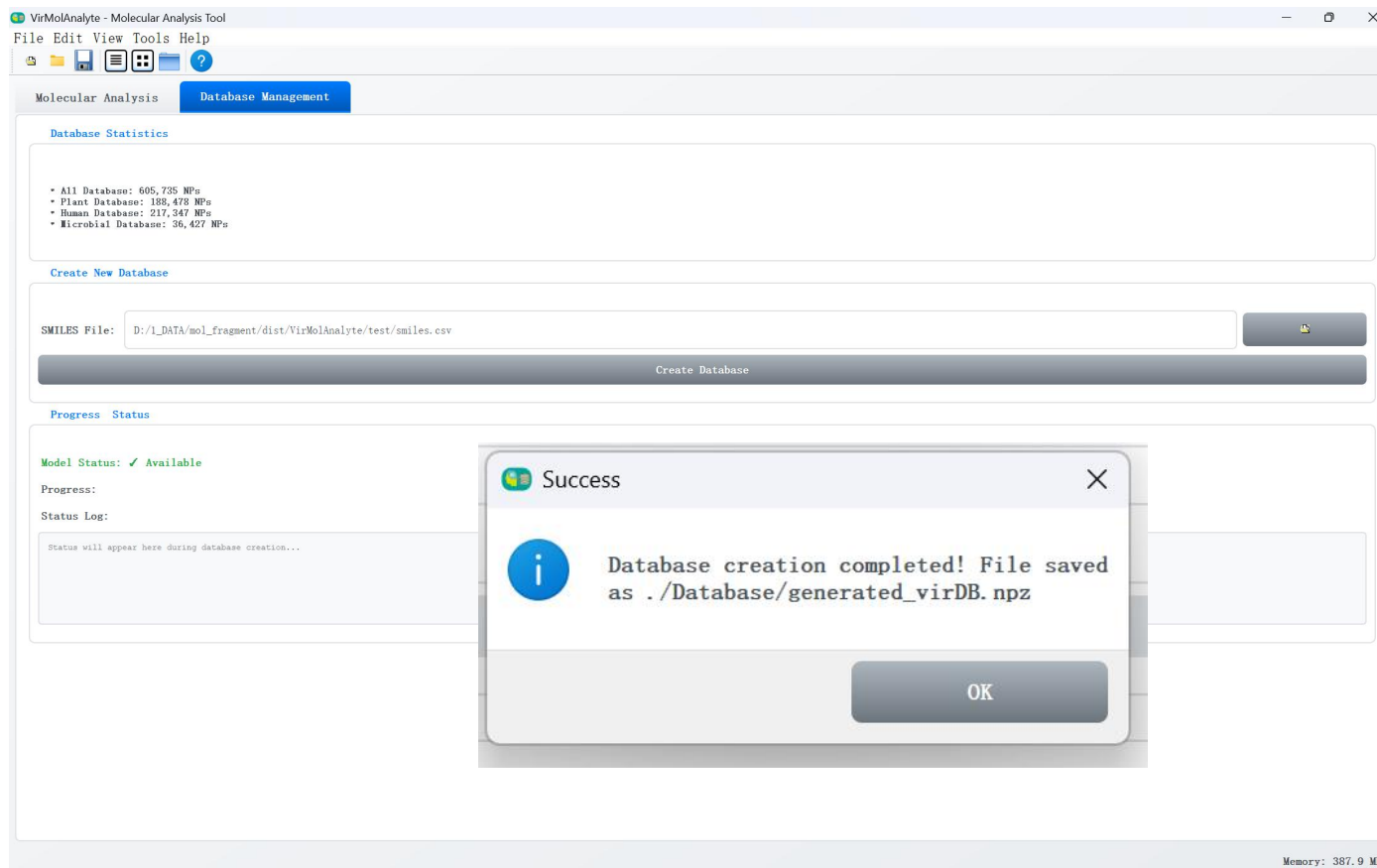
Copy SMILES



Usage of VirMolAnalyte

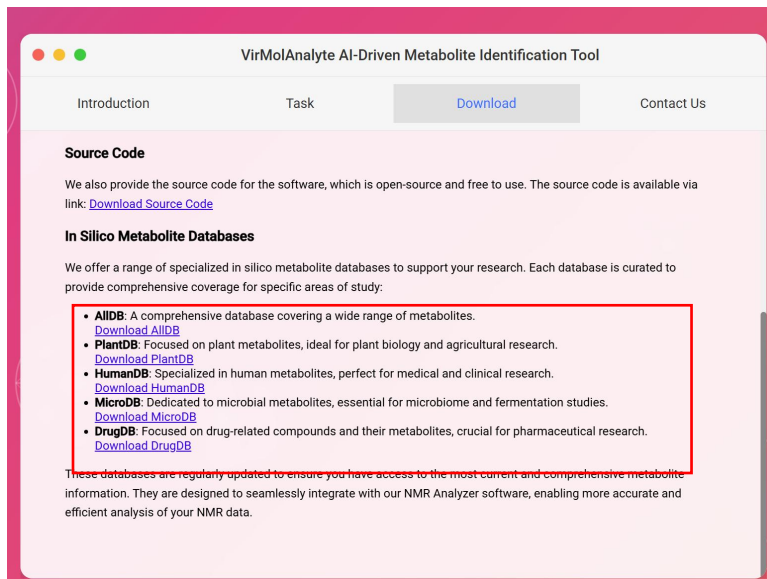
Generate your in house database

Place the molecular descriptors (smiles) used for database construction in a .csv file, and then run the construction method. *The test file is located at path “./test/smiles.csv”*

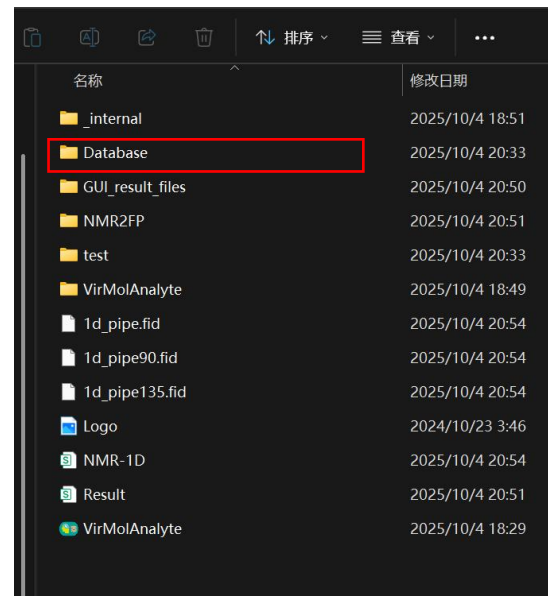


Note: In step 5 of the analysis, you can choose your database (Other database<load other database) for NMR analysis; Please ensure that Java is installed on your computer

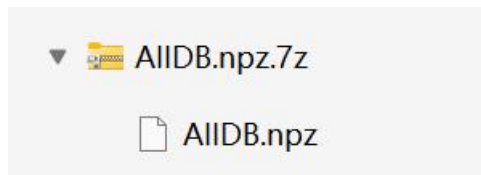
1. Download data files from the web page



3. Place the decompressed file (. npz) in the **Database** folder of the working directory



2. Retrieve database file from compressed file



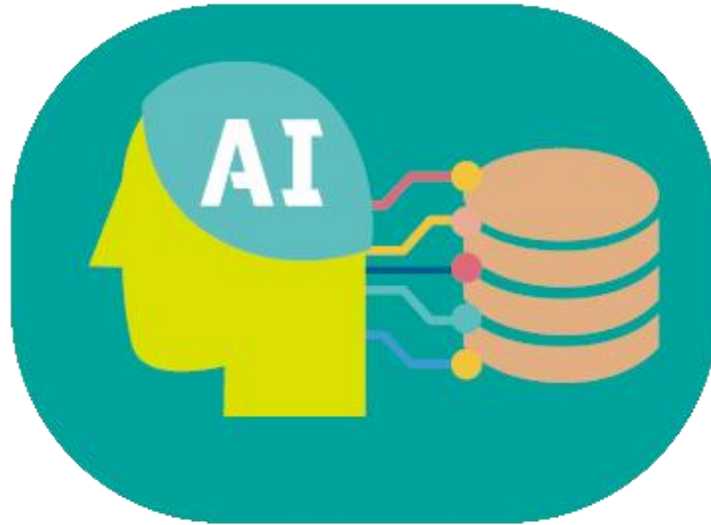
AIIDB: Also known as Metabolites Multidimensional In Silico Database (MMDB), comprising information on 605,735 metabolites.

PlantDB: An in silico database of plant-derived metabolites, containing 188,478 metabolites.

HMDB: An in silico database primarily focused on human-derived metabolites, including 217,347 metabolites.

MicroDB: An in silico database of microbial-derived metabolites, encompassing 36,427 metabolites.

DrugDB: An in silico database focused on drug-related compounds and their metabolites, including 4,280 metabolites.



Thanks