Manual for MFSearcher tool

version 1.6.0

August 28, 2019

License

MFSearcher GUI tool is available free for charge for academic purpose. The compounds registered in the following databases can be searched by mass values using this tool. Please preserve the license information of each databases when using the search results of this tool.

KEGG	http://www.genome.jp/kegg/
KNApSAcK	http://kanaya.naist.jp/KNApSAcK/
FlavonoidViewer	http://metabolomics.jp/wiki/Category:FL
HMDB	http://www.hmdb.ca/
LipidMAPS	http://www.lipidmaps.org/
UNPD	http://pkuxxj.pku.edu.cn/UNPD/
PubChem	http://pubchem.ncbi.nlm.nih.gov/

Requirement s

PC (Windows is recommended) in which Java runtime environment (1.7 or later) is installed and connected to the Internet is required.

The software accesses to the Internet in the following operations:
- When executing the search, the tool accesses to the MFSearcher web service (http://webs2.kazusa.or.jp/mfsearcher/) in the background process.
- When browsing the searched compound information, the tool accesses to the original databases by the Internet browsers.

If proxy server is used to connect to the Internet, please set the proxy server information according to the section “Proxy Server Settings”.

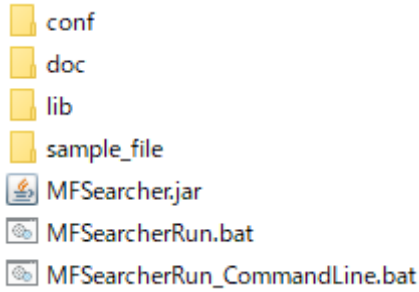
Please refer to the URL below for installation of Java.

https://www.java.com/ja/download/help/download_options.xml

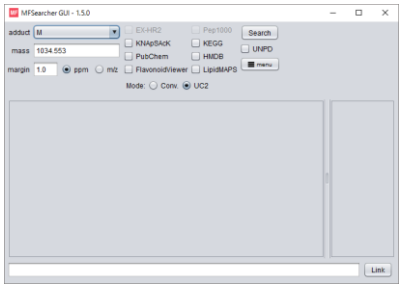
Run/Stop the tool

- Run as a GUI tool

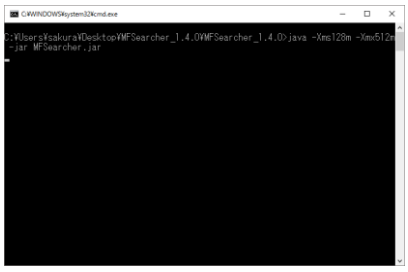
1) Decompress the downloaded .zip file. The following files and folders are appeared.



2) Double click the "MFSearcherRun.bat" file to run the tool. The main window of MFSearcher GUI is displayed.



* A console window shown below is appeared too. Don't close this window when MFSearcher program is running. If the window is closed by clicking "X" button on top-right of the window, the MFSearcher program stops too.



3) To stop the tool, click the "X" button on top-right of the main window.

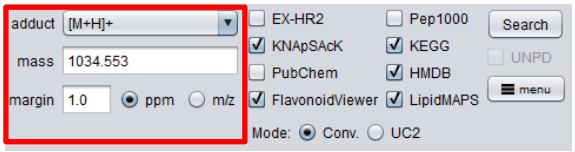
- Run as a command line tool

You can see a help of the command line tool by double clicking the "MFSearcher_CommandLine.bat" file in the extracted file set (see above). The command line tool is useful for batch processing for a large number of mass values (and their estimated adducts). The results are obtained as text files.

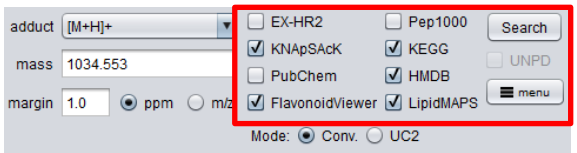
Details of the command line tool are described in the "Command line mode" section.

Basic Use

1) Enter search conditions.
adduct: Select a type of adduct.
mass: Enter the mass value
margin: Enter the mass tolerance, and select the units, ppm or m/z (Da).



2) Check the databases for search.

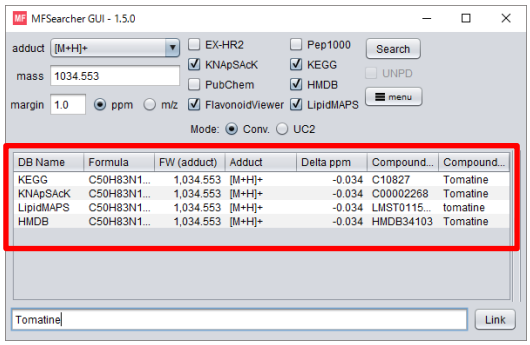


Following databases can be used. See also MFSearcher website (http://webs2.kazusa.or.jp/mfsearcher) for the details.

EX-HR2	A database of computationally calculated elemental compositions

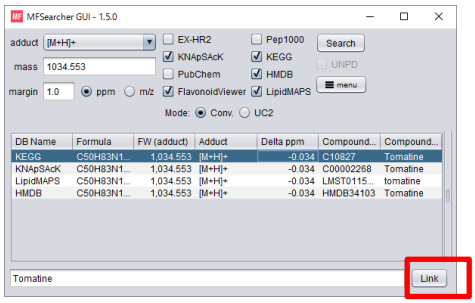
	of possible molecules with correct valences.
Pep1000	A database of computationally calculated linear polypeptides whose molecular weight are up to 1000.
KNApSAcK	A database of natural products. http://kanaya.naist.jp/KNApSAcK/
KEGG	Compound database in the KEGG pathway. http://www.genome.jp/kegg/
PubChem	A huge compound database maintained by NCBI. https://pubchem.ncbi.nlm.nih.gov/
HMDB	A database of compounds detected and predicted in human body. http://www.hmdb.ca/
FlavonoidViewer	A database of flavonoids. Flavonoid is a group of secondary metabolites primarily biosynthesized in plants. http://metabolomics.jp/wiki/Category:FL
LipidMAPS	A database of lipids, polyphenols and so on. http://www.lipidmaps.org/

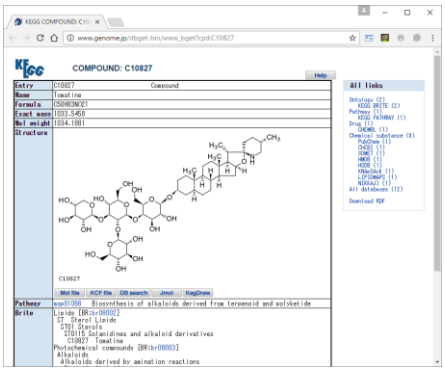
1) Execute searching by clicking the "Search" button. The results are listed in the table.



Click one of the results to see the name of the compound. The name is displayed in the text field at the bottom of the main window.

Click the "Link" button to see the information of the compound. Web browser will open and the original page of the selected compound in the database can be seen.





UC2 Search

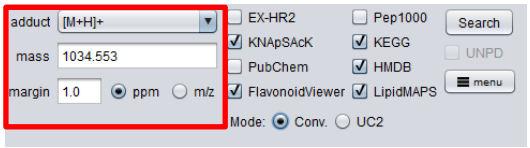
UC2 database compiles the information of the same compounds in multiple databases into one result. UC2 has following features too:

- Charged compounds in the original databases, such as cyanidines ([M]+), are searched with correct charge states and mass values.
- Compounds registered as multiple components in the original databases, such as salts, are searched as a single component (the component of the largest molecular weight is used).

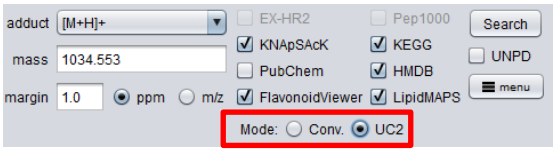
Therefore, UC2 search excludes the miss hits to the charged compound and salts, and reduce the candidates.

* The identity of the compound structure is judged by the connectivity of the atoms using the first block (14 letters) of InChIKey, and the stereochemistry is not distinguished. For example, D-type and L-type of an amino acid is compiled in a single result.

1) Set the adduct, mass and margin, same as those in the normal search.



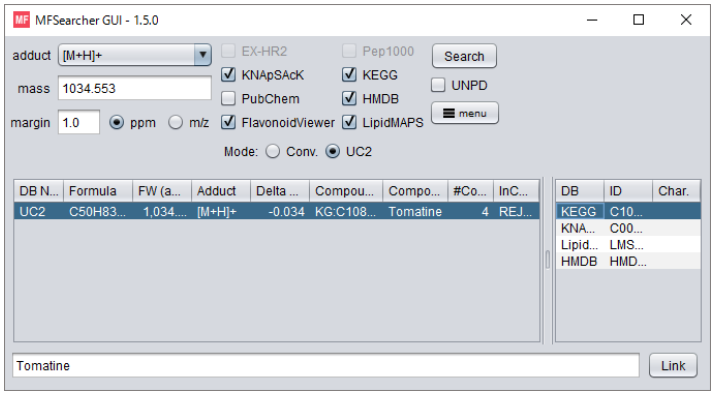
2) Check the "UC2" radio button to enter the UC2 mode.



In UC2 mode, the databases, EX-HR2 and Pep100, cannot be selected and searched, because these databases does not contain chemical structural information. The UNPD database can be selected in UC2 search.

UNPD	A large compound database of natural product http://pkuxxj.pku.edu.cn/UNPD/

3) Click "Search" button to execute the search. The results are listed in the table.



In UC2 mode, the result table is separated into two tables. The left table showed the list of the same connectivity where several compounds are compiled. Click one of the list, then the database names and the IDs of the compiled compounds are displayed in the right table.

Select a compound in the right table, and click "Link" button. The details of the compound information can be read in the Internet Browser.

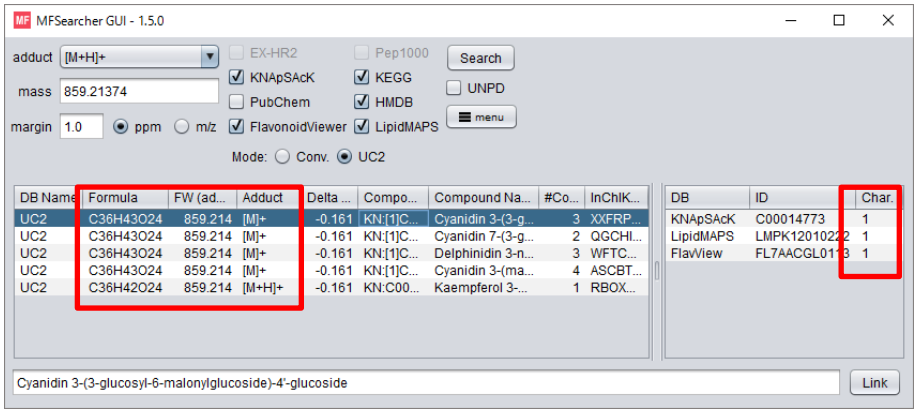
* Records in the UNPD are not linked because of the specification of the original website.

The compound name displayed in the left table is a representative of those that has the shortest text length among the compiled compounds.

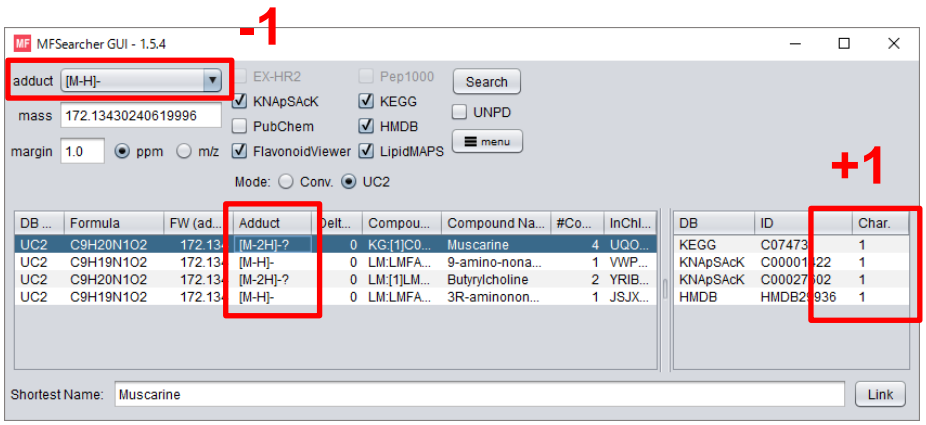
The column “Char.” in the right table shows the features of the compounds in the original database. Each letter in the column shows followings:

Number	The charge of compound.
f	The compound was registered as multiple components, such as a salt.
r	The compound was a radical.

When the query is recognized as apparent molecular ion, the information of the molecular ion is displayed in the Formula, Adduct and FW (molecule) column. For example, when a mass is searched as $[M+H]^+$ and all the results are records with +1 charge, the information of $[M]^+$ will be displayed.

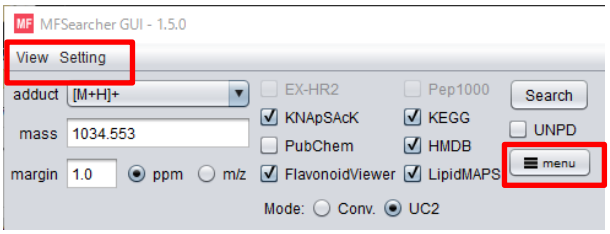


When all the results have opposite charge to the queried adduct, a candidate adduct ion with “?” will be displayed in the Adduct column.



Advanced Settings

Click the "menu" button to show the menu.

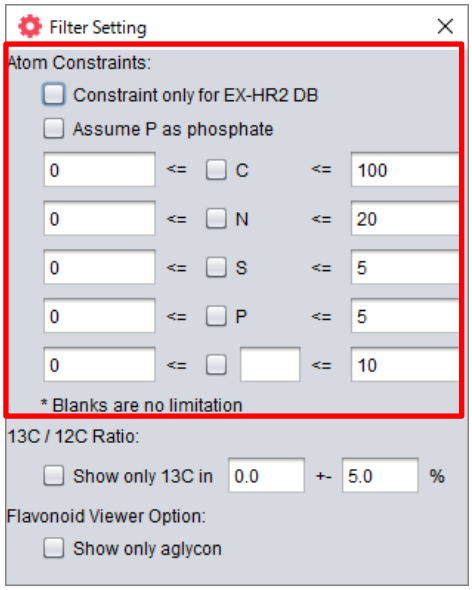


Following items are included in the menu.

View	Show/Hide Status Bar	This menu item switches On/Off of the status bar at the bottom of the main window.
	Show DB Info.	A window that informs the record numbers and updated dates of the databases will open. * The databases listed here is only for the normal search. The data used in UC2 search are not the same one of them and are not displayed here.
	Hide Menu	This menu item hide the menu.
Setting	Column Settings	Users can change the columns in the result table.
	Filter Settings	A setting window for filtering the results by atom numbers etc., will open.

Filter Settings

Open a setting window by selecting "Filter Settings" in the "Setting" menu.



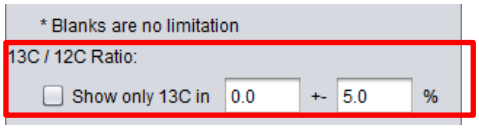
1) Filtering by atom numbers

Users can filter the results by number of atoms C, N, S, and P in the chemical formula. Check the check boxes adjacent to the atom labels to enable the filtering. The minimum and maximum number of the atoms are set by the text fields at left and right of the atom label. An optional atom label can be set in the text field under "P".

Check the "Constraint only for EX-HR2 DB" to limit the filter settings only to the EX-HR2 database.

2) Filtering by $^{13}C/^{12}C$ ratio

Users can filter the results by formula that match to the ratio of the intensities of $^{13}C_1$ stable isotope peak and to ^{12}C monoisotopic peak.



Enter the $^{13}C/^{12}C$ ratio (obtained by experimental measurement etc.) in the text field at the left.

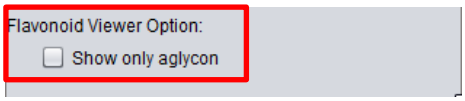
Enter the tolerance of the error in % in the text field at the right.

Check the check box "Show only 13C in" to enable the filtering.

3) Limit the FlavonoidViewer results to aglycones

Check the " Show only aglycon" at the "FlavonoidViewer Option" to limit the results

from FlavonoidViewer to aglycones.



Command Line Mode

Command line mode is useful for batch processing for a large number of mass values (and their estimated adduct ions) for formula prediction and compound database searching. Results are obtained as text files.

1) Preparation of an input file

A list of mass values is prepared as a text file. Following three formats are acceptable. You can see examples in the sample_file folder.

a) A list of mass values (input_text_plain.txt)

Mass values are written in each line. The mass values are directly used for calculation.

b) A list of pairs of mass value and adduct (input_text.txt)

Pairs of mass value and estimated adduct ion are separated by tab and written in each line. The labels of the adduct ions should be defined in “adduct.ini” file. See “Adduct Settings” section for details.

c) MassChroViewer CheckFile (input_mcv_checkFile.txt)

A file used in the MassChroViewer tool. The file is generated by MassChroViewer and PowerGetBatch.

MassChroViewer: http://www.kazusa.or.jp/komics/software/MassChroViewer
PowerGetBatch: http://www.kazusa.or.jp/komics/software/PowerGetBatch

2) Run command prompt tool (terminal tool) of the OS.

3) Using “cd” command, move to the directory where you have extracted the fileset of the tool. You can see “MFSearcher.jar” file there.

4) Command line help is available by executing following command.

java -jar MFSearcher.jar -h

5) Similarly, you can execute MFSearcher tool in command line mode by defining the input file, output directory, compound databases for search, mass tolerance for formula prediction and database search as options.

Ex) A case using a list of pairs of mass value and adduct ion (input_text.txt)

java -jar MFSearcher.jar -i sample_file/input_text.txt -od outdir_plain -db KG,KN,FL,HM,LM -fppms 1,3,5 -dppms 3,5,15

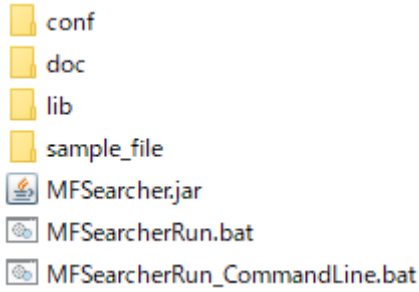
Multiple values for mass tolerances can be set by separating the values by comma. In this case, formula prediction and/or database search are attempted using smaller tolerances and continued till any result obtained. The UC2 database is used for database search.

6) Result files are obtained at the directory defined by “-od” option.

ResFormula_[Input File Name]: Full results of formula prediction.
ResDatabase_[Input File Name]: Full results of database search.
ResSummary_[Input File Name]: Summary of the two files above.

Adduct Settings

See the "conf" folder in the file sets of MFSearcher program.



A file named "adduct.ini" is included in the conf folder. By editing the adduct.ini file, the information of adducts used in the MFSearcher program can be changed.
The adduct.ini file is written in a tab delimited text format. Use a proper software such as a text editor (note pad) to edit the file.

An example of the file structure is as below.

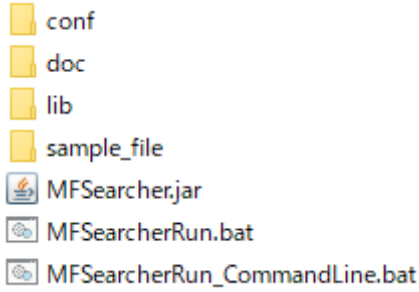
1) Caption	2) Number of molecule	3) Formula added	4) Formula subtracted	5) Optional mass value	6) Charge	7) Not used
M	1				0	TRUE
[M]+	1				1	TRUE
[M+H]+	1	H			1	TRUE
[M+NH4]+	1	NH4			1	TRUE
[M+Na]+	1	Na			1	TRUE
[M+K]+	1	K			1	TRUE
[M−H2O+H]+	1	H	H2O		1	TRUE
[M−2(H2O)+H]+	1	H	H2OH2O		1	TRUE
[2M+H]+	2	H			1	TRUE
[2M+NH4]+	2	NH4			1	TRUE
[2M+Na]+	2	Na			1	TRUE
[2M+K]+	2	K			1	TRUE
[2M+ACN+H]+	2	C2H3NH			1	TRUE
[M+2H]2+	1	H2			2	TRUE
[M+2Na]2+	1	Na2			2	TRUE
[M+Na+H]2+	1	NaH			2	TRUE
[M+3H]3+	1	H3			3	TRUE
[M+3Na]3+	1	Na3			3	TRUE
[M−H]−	1		H		−1	TRUE
[M+HCOO]−	1	HCOO			−1	TRUE
[M+K−2H]−	1	K	H2		−1	TRUE
[M+Na−2H]−	1	Na	H2		−1	TRUE
[M+2Na−3H]−	1	Na2	H3		−1	TRUE
[M+HCOO+Na−H]−	1	HCOONa	H		−1	TRUE
[M+HCOO+K−H]−	1	HCOOK	H		−1	TRUE
[M−2H]2−	1		H2		−2	TRUE
[M−3H]3−	1		H3		−3	TRUE

1) An arbitral name can be set. Avoid to enter the same name in multiple rows.
2) In the case of [2M+H]+, 2 should be entered.
5) If the mass difference is difficult to described by formula, enter a numerical value here.
7) This column is not used.

Save the file and restart the MFSearcher tool to reflect the editing.

Proxy Server Settings

See the "conf" folder in the file sets of MFSearcher program.



Find a file named “system.ini” in the “conf” folder. Open the “system.ini” file by a text editor (such as Note Pad). A default setting is shown below.

```
# network proxy setting
network.proxy.enabled=false
network.proxy.server=your.proxy.server
network.proxy.port=8080
```

Edit the text in the right side of “=” as below.

network.proxy.enabled	true
network.proxy.server	Name or IP address of the proxy server
network.proxy.port	Port number of the proxy server

Example)

```
# network proxy setting
network.proxy.enabled=true
network.proxy.server=123.45.67.89
network.proxy.port=1234
```

Overwrite the “system.ini” file, and then restart the MFSearcher.

Publications

Sakurai N, Narise T, Sim JS, Lee CM, Ikeda C, Akimoto N and Kanaya S (2018) UC2 search: Using unique connectivity of uncharged compounds for metabolite annotation by database searching in mass spectrometry-based metabolomics. Bioinformatics 34: 698-700
[PMID: 29040459]

Sakurai N, Ara T, Kanaya S, Nakamura Y, Iijima Y, Enomoto M, Motegi T, Aoki K, Suzuki H and Shibata D (2012) Application of a relational database system for high-throughput prediction of elemental compositions from accurate mass values. Bioinformatics 29 (2): 290-291
[PMID: 23162084]

Contact Information

National Institute of Genetics, Mishima, Japan
Nozomu SAKURAI Ph.D.
E-mail: sakurai AT nig.ac.jp (replace AT with @)

The earlier version of this tool was developed by NS at Kazusa DNA Research Institute, Kisarazu, Japan.