Manual updated on November 21, 2020

Compatible with version 1.0.2

1. About

MS2Compound is a user-friendly Graphical User Interface (GUI) for the identification of the compounds from MS/MS metabolomics data. The tool features the use of a custom database for the identification of compounds. The current version of MS2Compound (v1.0.2) is compatible with the customized database prepared using the fragment prediction using CFM-ID tool (https://cfmid.wishartlab.com/).

The current version of the tool is compatible with Mascot Generic Format (.mgf) files as input.

The tool has been developed and maintained by Center for Systems Biology and Molecular Medicine (http://csbmm.yenepoya.res.in/), Yenepoya Research Centre, Yenepoya (Deemed to be University).

2. Installing MS2Compound

2.1 System requirements

2.1.1 Operating system

Windows 7, 8 and 10

2.1.2 Minimum system configuration

2GB or more RAM

Intel i3 64-bit CPU

2.2 Installation

2.2.1 Dependencies

The tool is independent of all the prerequisite dependencies.

2.2.2 MS2Compound installation

- Download the application "MS2Compound_v1.0.2.exe" from https://sourceforge.net/projects/ms2compound/

- Click on “MS2Compound_v1.0.2.exe” to install the program. The default installation directory will be “C:\MS2Compound”, user can select the installation directory of choice, however, avoid spaces (“ ”) in the directory pathname.
- Please select to create a shortcut of the MS2Compound in Desktop for easy access.
- A successful installation will create a directory “C:\MS2Compound”. This directory will contain the executable “MS2Compound.exe”.

3. Start MS2Compound

Click on the shortcut “MS2Compound_v1.0.2”, created in Desktop after successful completion of installation. If the user has not opted for desktop shortcut during installation, the executable can be started by clicking the “MS2Compound.exe” in “C:\MS2Compound” directory.

4. Availability of test dataset

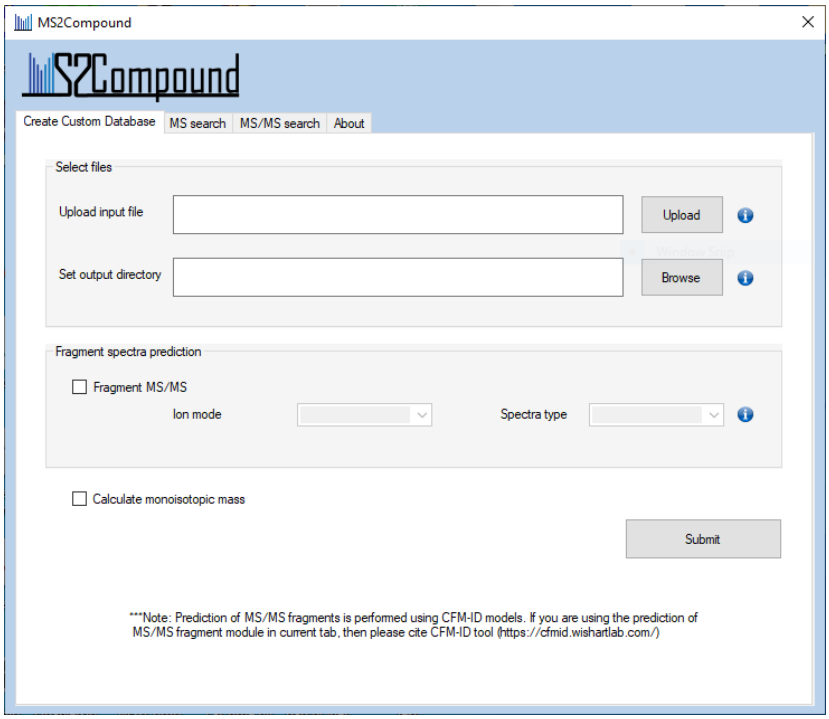
The test dataset for MS2Compound are available at https://github.com/beherasan/MS2Compound/tree/master/TestData

The current version of MS2Compound performs the following five tasks (Link for the test data for each function are provided in parentheses).

a. Create a custom MS/MS database using CFM-id prediction models. Please cite CFM-id (https://cfmid.wishartlab.com/) if the user has used this module in MS2Compound. (https://github.com/beherasan/MS2Compound/tree/master/TestData/FragmentPrediction)
b. Identification of compounds from LC-MS data using the in-built database. (https://github.com/beherasan/MS2Compound/tree/master/TestData/MS_InBuiltDB)
c. Identification of compounds from LC-MS data using custom (user-defined) database. (https://github.com/beherasan/MS2Compound/tree/master/TestData/MS_CustomDB)
d. Identification of compounds from LC-MS/MS data using the in-built database. (https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_InBuiltDB)
e. Identification of compounds from LC-MS/MS data using custom (user-defined) database. (https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_CustomDB)

4. Run MS2Compound

Clicking the shortcut “MS2Compound_v1.0.2” in Desktop or “MS2Compound.exe” in C:\MS2Compound will start the GUI as shown below.



Create Custom Database

This module will help the user to create their custom database from a list of compounds (given the SMILES id of the listed compounds)

The page for creating a custom database is the default start page for MS2Compound.

The input file as given the test dataset "inputSMILES.txt" is a tab-separated file containing two columns; the first column is the compound id or any unique identifier, and the second column containing the SMILES id as shown below.

```
challenge-001    c1cc(ccc1C1COc2cc(cc(c2C1=O)O)O)O
challenge-002    C=C(C)C(C)c1cc(/C=C/C(=O)c2ccc(cc2)O)c(cc1O)OC
challenge-003    COc1ccc(CCN=C(/C=C/c2ccccc2)O)cc1OC
challenge-004    c1cc(ccc1c1c(c(=O)c2ccc(cc2o1)O)O)O
challenge-005    c1cc(c(cc1/C=C\1/C(=O)c2ccc(cc2O1)O)O)O
```

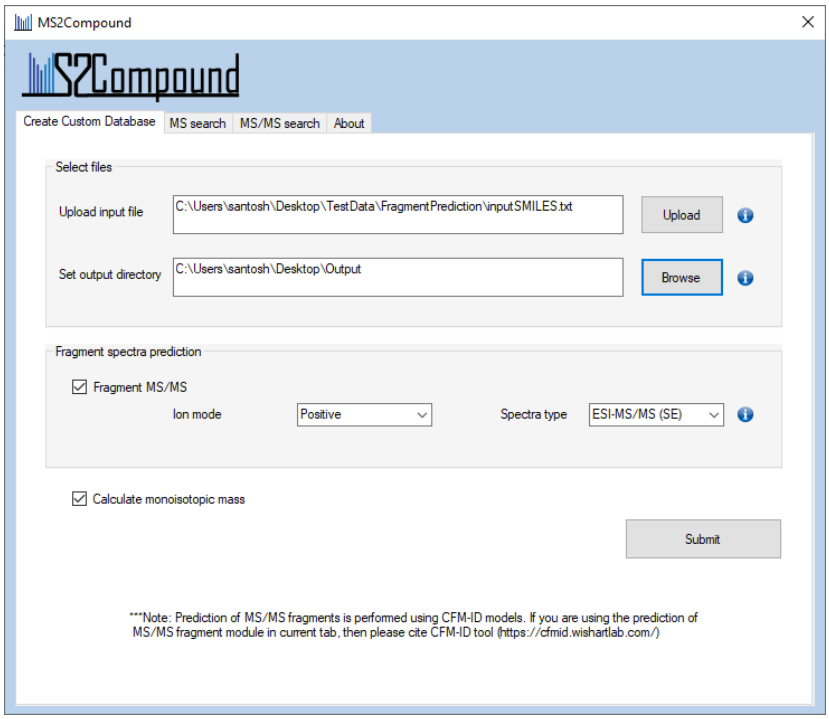
Select the Output directory: Please avoid spaces in the name of the directory or the directory path

User can select both the option "Fragment MS/MS" and "Calculate monoisotopic mass" or anyone according to the need.

After selecting the "Fragment MS/MS", please select the desired "Ion mode" and "Spectra type" for the prediction.

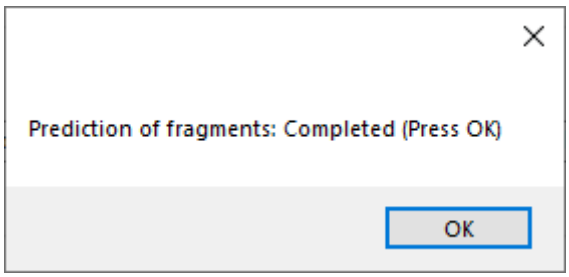
Click “Submit after selecting the desired options”. After submitting, it will start a command prompt, and the main program cannot be assessed until the current work completed.

A filled form is shown below:

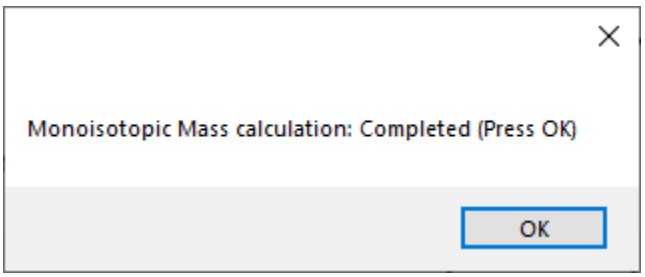


Clicking “Submit” will start a command prompt as follows:

After completion of the prediction program, the above command prompt will disappear, and it will create a message box as follows:

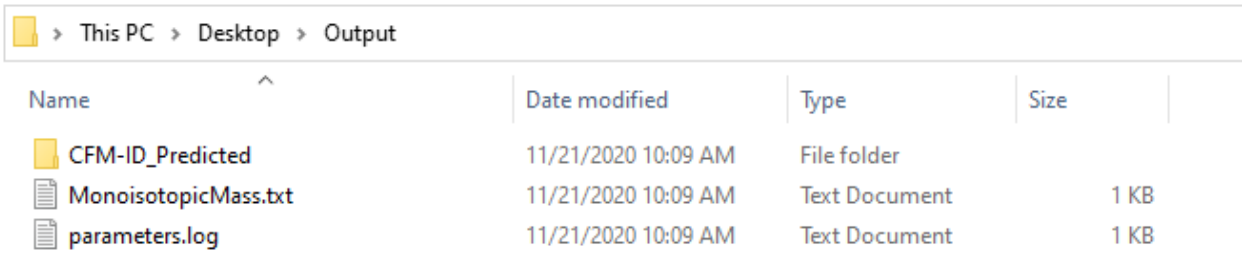


Click “OK” for the monoisotopic mass calculation program to execute. After the completion of the mass calculation, it will generate a message box as follows:



Press “OK” to finish the task.

This program will generate a directory “CFM-ID_Predicted” inside the selected output directory for the MS/MS fragments. If the directory “CFM-ID_Predicted” already exists in the selected output directory, the program will be halted with an error. Please rename the existing file or delete it before running the command. After successful completion of the prediction, it will create separate log files for each compound in “CFM-ID_Predicted” directory; this directory can be used as the custom database in the MS/MS search using custom Database module in the tool. The output directory will be generated as follows:



The parameters used for prediction of MS/MS fragments were saved in “parameters.log” file

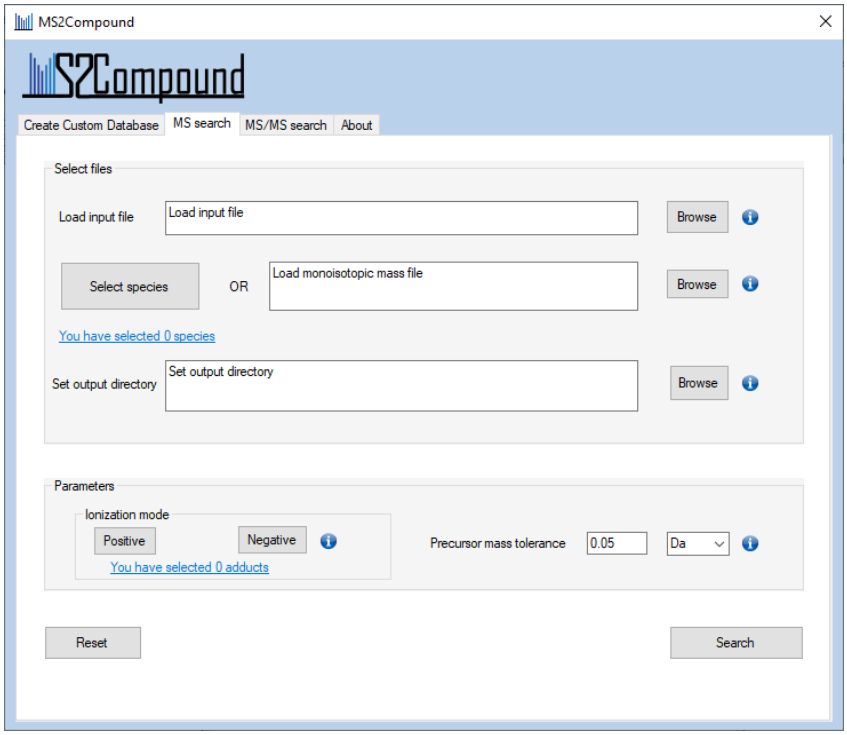
Alongside the program will create an output file named “MonoisotopicMass.txt” for the calculated monoisotopic masses.

The sample outputs from this module using the given test data, and “Positive” ionization mode and “ESI-MS/MS (CE)” spectra type is provided in https://github.com/beherasan/MS2Compound/tree/master/TestData/FragmentPrediction/OUTPUT.

********* *The time taken for execution of CFM-id prediction will be based on the number of the compounds in the input file. The current monoisotopic mass calculation cannot be used for the isotopes of compounds.*

MS-Search using in-built Database

Click on the “MS search” tab. This will open a window as follows:

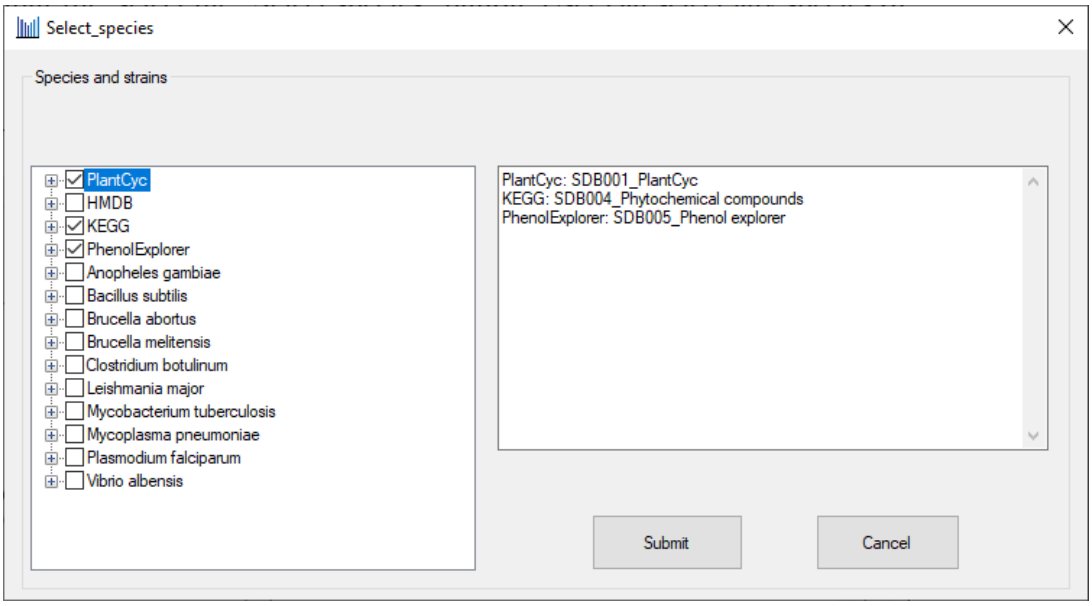


The input file should be a tab-separated file with three columns; the first column containing unique identifiers; second containing the precursor m/z value, and the third column with charge information. The format of the charge should be “a numeric number followed by + or - sign”, for example, “1+”. A sample file has been shown below:

1	145.54	1+
2	530.3015	1+
3	758.2237	1+
4	336.2166	2+
5	747.3701	1+
6	345.2206	2+
7	445.2429	2+
8	536.1644	
9	610.1874	
10	182.9848	1+

If the charge information is provided with the match will consider particular adducts for the given charge, else it will search for all adducts selected.

After uploading the input file, select the “Select species” button. User can select any species of interest, or the search can be strain-specific by clicking the “+” sign of species, and then select the desired strain. The species selection is as follows:

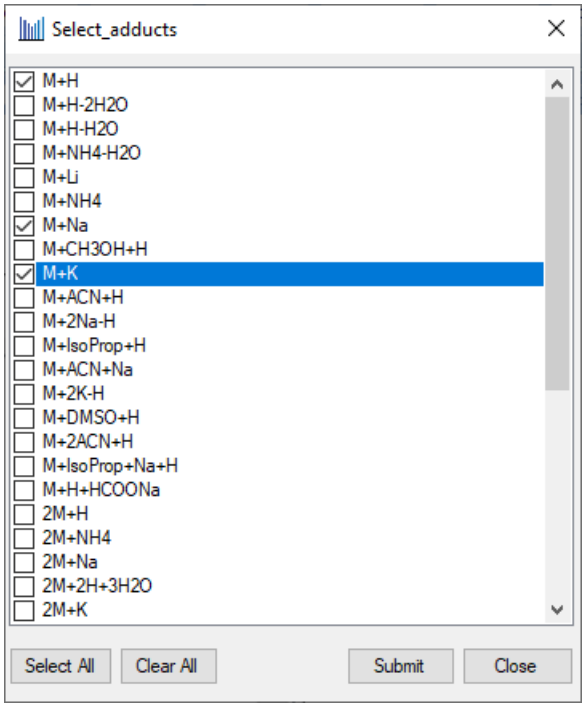


Please ignore the “SID…” as these are the in-built ids provided by MS2Compound backend database.

Click “Submit” after selecting the species.

Select the “Output directory”

Select the “Ionization mode” by clicking either “Positive” or “Negative” button. This will open a new window for adducts.

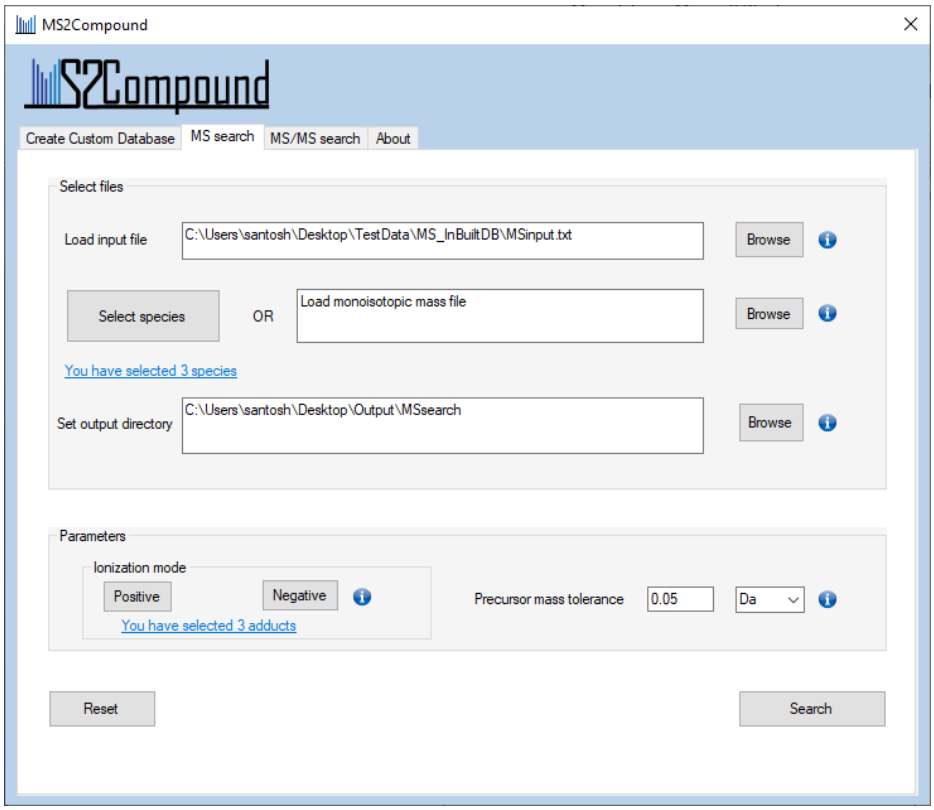


If adducts are not known, please select all adducts.

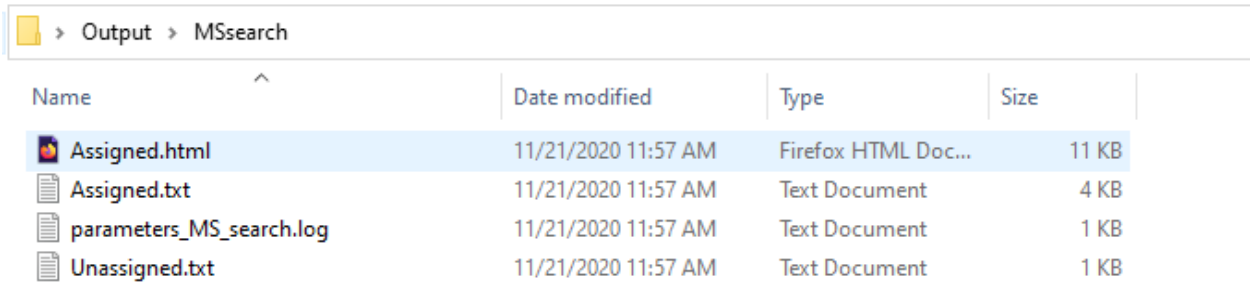
Select the “Precursor mass tolerance”, and the unit.

Click “Search”

A filled form is shown below:



This generates three output files in the selected output directory, and a parameters log file as follows:



The output files are:

1. Assigned.txt

2. Assigned.html

3. Unassigned.txt

4. parameters_MS_search.log

Assigned.html can be opened with any web-browser. A screenshot for the same is shown below:

Please Cite MS2Compound if you are using this as a part of your analysis

ID	mz	Rank	Adduct	Charge	MS2Compound ID	Da error	Source ID	Source	Name	Monoisotopic Mass	HMDB	PubChem
2	530.3015	1	M+Na	1+	KG_C08679	0.029064700000049	C08679	KEGG	Deltaline	507.2832173	-	441728
3	758.2237	1	M+H	1+	KG_C12646	0.002694700000000638	C12646	KEGG	Cyanidin 3-O-rutinoside 5-O-beta-D-glucoside	757.2191187	-	443921
3	758.2237	1	M+H	1+	PE_21	0.002694700000000638	21	PhenolExplorer	Cyanidin 3-O-glucosyl-rutinoside	757.2191187	-	
3	758.2237	1	M+H	1+	PE_741	0.002694700000000638	741	PhenolExplorer	Peonidin 3-O-sambubioside-5-O-glucoside	757.2191187	-	
3	758.2237	2	M+Na	1+	KG_C08327	0.002966900000004673	C08327	KEGG	Anthemis glycoside B	735.2374489	-	11953899

The sample output files can be found at https://github.com/beherasan/MS2Compound/tree/master/TestData/MS_InBuiltDB/OUTPUT.

MS-Search using Custom Database

This module allows the user to search MS data against a user-defined monoisotopic mass file for a selected list of compounds.

The input file will be similar to that of the input for "**MS-Search using in-built Database**".

The monoisotopic mass file should be a tab-separated file with two columns; one for a unique identifier, and the second one for corresponding monoisotopic mass.

Remaining parameters and the output will be similar to that of "**MS-Search using in-built Database**".

MS/MS-Search using in-built Database

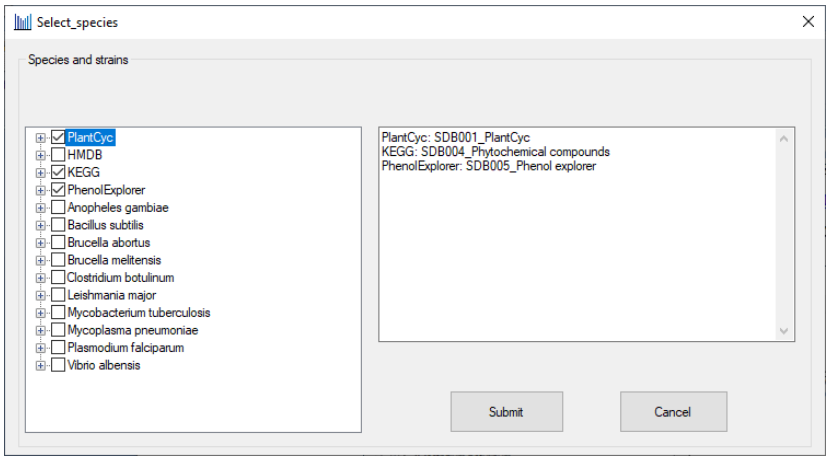
This module allows the user to search MS/MS data against the in-built MS/MS database. The current version (1.0.2) of MS2Compound supports Mascot Generic Format (MGF) file as input.

The screenshot for a sample input is shown below:

```
BEGIN IONS
FEATURE_ID=2
PEPMASS=245.1289
RTINSECONDS=145.54
CHARGE=1+
MSLEVEL=2
75.72264099121094 0.0921952008202265
76.5239486694336 0.105720467448089
77.03893280029297 0.129845590210377
82.06517791748047 0.157307460571153
85.88349914550781 0.0928718868766921
87.49651336669922 0.100132314717048
91.05423736572266 0.650206850612891
93.07019805908203 0.282088145881232
96.9111328125 0.109116483030762
98.06049346923828 20.5369436388339
102.48599243164062 0.0955491634067138
103.05403900146484 0.858005857680907
105.0699234008789 0.199647936413473
108.80220031738281 0.0939097949997296
113.07038879394531 0.0998121146098996
120.08045196533203 100
121.7412338256836 0.0913062209685591
125.07101440429688 0.462762398350365
126.07872772216797 0.274709301805809
131.04966735839844 1.98450060608678
137.071044921875 0.153365507697125
148.0762939453125 0.511827029879847
```

The MS/MS search can be performed with the following steps:

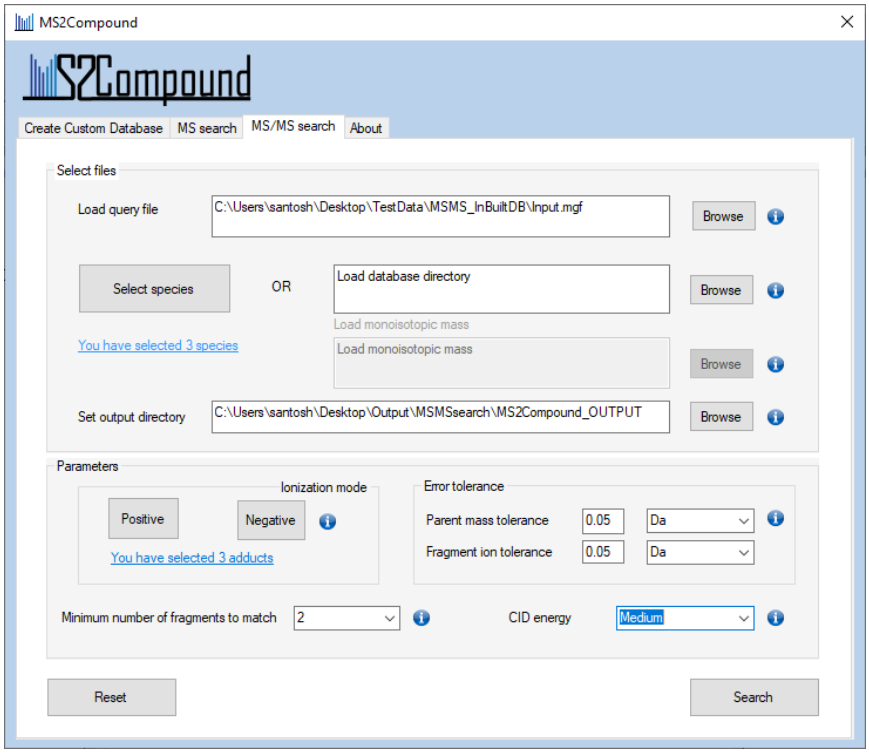
- Load MGF file as Query file. The sample input is provided at https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_InBuiltDB
- Select the species of interest by clicking the “Select species” button.



- Set the output directory
- Click on ionization mode (either Positive or Negative), and select adducts. If the adducts are not known, please select all the adducts

- Select the error tolerance for precursor and fragment match
- Select the “Minimum number of fragments to match” for the number of fragment matches.
- Select the CID energy for MS/MS fragmentation
- Select the “Search” button

A filled form is shown below:



This will create a directory “MS2Compound_OUTPUT” in the selected output directory. If the “MS2Compound_OUTPUT” directory already exists in the output directory path, it will create “MS2Compound_OUTPUT_1” for a new search. In the same search, it will replace the results in “MS2Compound_OUTPUT”.

On clicking “Search”, it will start a command prompt as follows:

After completion of search, it generates four output files in "MS2Compound_OUTPUT" directory.

1. Assigned.txt

2. Assigned.html

3. Unassigned.mgf

4. allMapped.txt

5. parameters_MSMS_search.log

The sample output files can be found at https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_InBuiltDB/MS2Compound_OUTPUT

"Assigned.html" can be open with any browser (such as firefox). A screenshot of "Assigned.html" is shown below:

Please Cite MS2Compound if you are using this as a part of your analysis

Feature ID	Rank	Adduct	Charge	MS2Compound ID	mS-score	Source Id	Source	Name	Monoisotopic Mass	HMDB	PubChem
2	1	M+H	1+	KG_C10782	2.24816	C10782	KEGG	Sophoramine	244.1575633	-	169014
2	2	M+H	1+	KG_C10789	2.24118	C10789	KEGG	Thermospine	244.1575633	-	442967
2	2	M+H	1+	KG_C10750	2.24118	C10750	KEGG	Anagyrine	244.1575633	-	442938
2	3	M+H	1+	KG_C09778	0.84760	C09778	KEGG	Fulvoplumierin	244.0735589	-	5281541
2	4	M+H	1+	KG_C10273	0.83450	C10273	KEGG	Oxyresveratrol	244.0735589	-	5281717
2	5	M+H	1+	KG_C10790	0.69634	C10790	KEGG	Tinctorine	244.1575633	-	442968
2	6	M+H	1+	KG_C09404	0.50742	C09404	KEGG	Encelin	244.1099444	-	72540

The assigned compounds to a particular Feature ID are ranked based on the *mS-score* and sorted in order (Rank 1 is the best match).

"Assigned.txt" has similar information as that of "Assigned.html", and can be opened in any text editor or excel sheet to view the results.

MS/MS-Search using Custom Database

This module allows the user to search MS/MS data against custom/user-generated MS/MS database. The current version (1.0.2) of MS2Compound supports Mascot Generic Format (MGF) file as input as described in "**MS/MS-Search using in-built Database**" section.

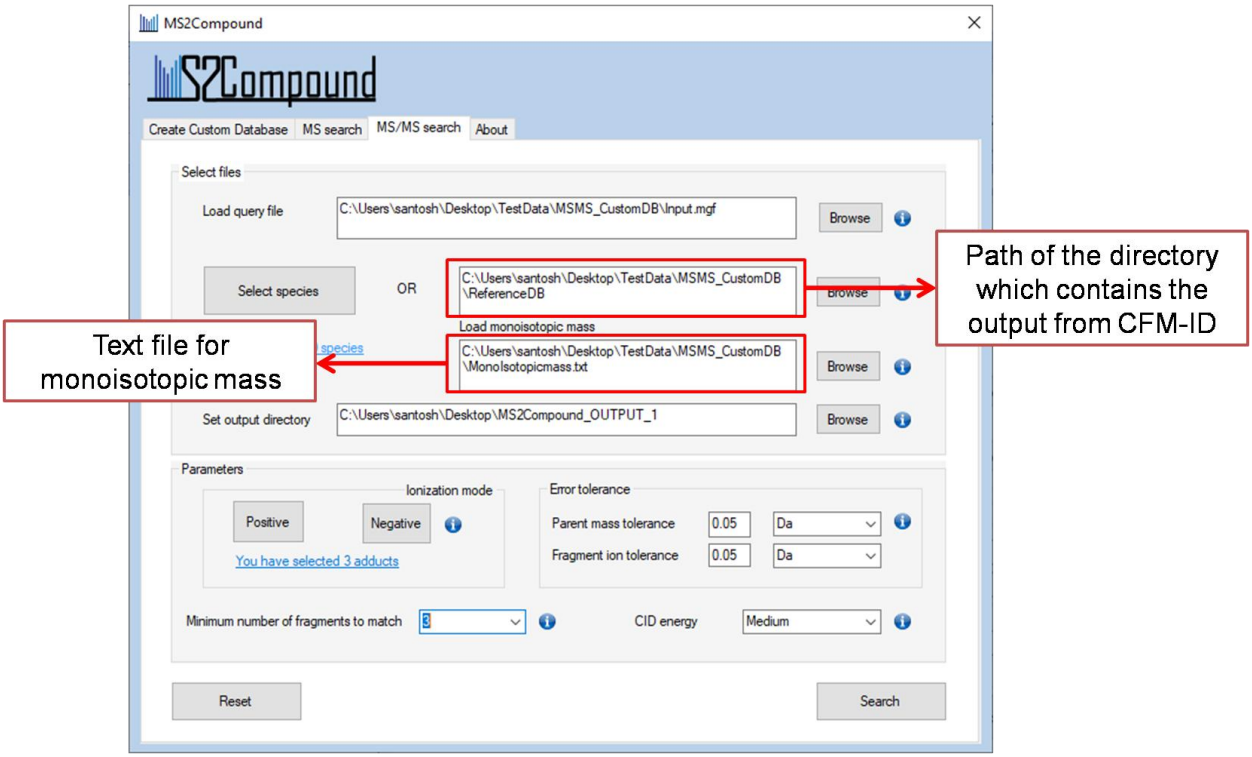
Unlike that of "**MS/MS-Search using in-built Database**", in this module user have to upload a Database directory (as "ReferenceDB" in https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_CustomDB), and the monoisotopic mass file (as "MonoIsotopicmass.txt" in https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_CustomDB) instead of selecting species of interest.

"MonoIsotopicmass.txt" is a tab-separated file containing a unique compound identifier (UID) and the corresponding monoisotopic mass.

The MS/MS fragment file names should be UID.log in "ReferenceDB". Please maintain the naming for the proper functioning of the tool.

If you are using our module "**Create Custom Database**" for creating your database, then you can directly load the Database directory and the monoisotopic mass file for the search.

A filled form is shown below:



The resulting file will be similar to that of the "**MS/MS-Search using in-built Database**"; however, it will not provide the compound details.

The sample output files are provided at https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_CustomDB/MS2Compound_OUTPUT.