

## 1. About

MS2Compound v1.0.0 is a user friendly Graphical User Interface (GUI) for the identification of the compounds from MS/MS metabolomics data. The tool features use of custom database for the identification of compounds. The current version is compatible with the customized database prepared using the fragment prediction using CFM-id tool.

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 from all the pre-requisite dependencies.

#### 2.2.2 MS2Compound installation

- Download the application “MS2Compound\_Setup.exe” from <https://github.com/beherasan/MS2Compound> or <https://sourceforge.net/projects/ms2compound/>
- Click on “MS2Compound\_Setup.exe” to install. The default installation directory will be “C:\MS2Compound”, user can select the installation directory of choice, however avoid spaces (“ ”) in the directory path name.
- 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”, created in desktop after successful completion of installation. If the user has not opted for desktop shortcut during installation, the executable can be start 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>

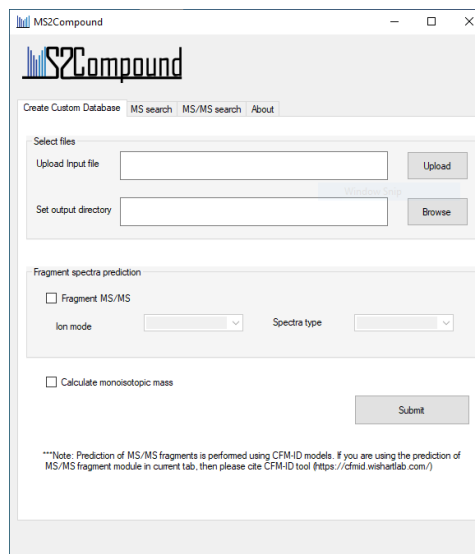
The current version of MS2Compound majorly performs the following five tasks (Link for the test data for each function are provided in parentheses).

- Create 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>)
- Identification of compounds from LC-MS data using in-built database. ([https://github.com/beherasan/MS2Compound/tree/master/TestData/MS\\_InBuiltDB](https://github.com/beherasan/MS2Compound/tree/master/TestData/MS_InBuiltDB))
- Identification of compounds from LC-MS data using custom (user-defined) database. ([https://github.com/beherasan/MS2Compound/tree/master/TestData/MS\\_CustomDB](https://github.com/beherasan/MS2Compound/tree/master/TestData/MS_CustomDB))
- Identification of compounds from LC-MS/MS data using in-built database. ([https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS\\_InBuiltDB](https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_InBuiltDB))
- Identification of compounds from LC-MS/MS data using custom (user-defined) database. ([https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS\\_CustomDB](https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_CustomDB))

### 4. Run MS2Compound

Clicking the “MS2Compound.exe” will start the GUI as shown in figure 1.

Figure 1. Start page of MS2Compound



## 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 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; first column is compound id or any unique identifier, and second column containing the SMILES id for the same (Figure 2).

challenge-001	<chem>c1cc(ccc1C1C0c2cc(cc(c2C1=O)O)O)O</chem>
challenge-002	<chem>C=C(C)C(C)c1cc(/C=C/C(=O)c2ccc(cc2)O)c(cc1O)OC</chem>
challenge-003	<chem>COc1ccc(CCN=C(/C=C/c2ccccc2)O)cc1OC</chem>
challenge-004	<chem>c1cc(ccc1c1c(c(=O)c2ccc(cc2O1)O)O)O</chem>
challenge-005	<chem>c1cc(c(cc1/C=C\1/C(=O)c2ccc(cc2O1)O)O)O</chem>

Figure 2. Sample input file for creating custom database for MS/MS search.

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

User can select both the option “Fragment MS/MS” and “Calculate monoisotopic mass” or any one 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 submit it will start a cmd.exe and the main program cannot be assessed until the current work completed.

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 selected output directory, 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 compounds 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.

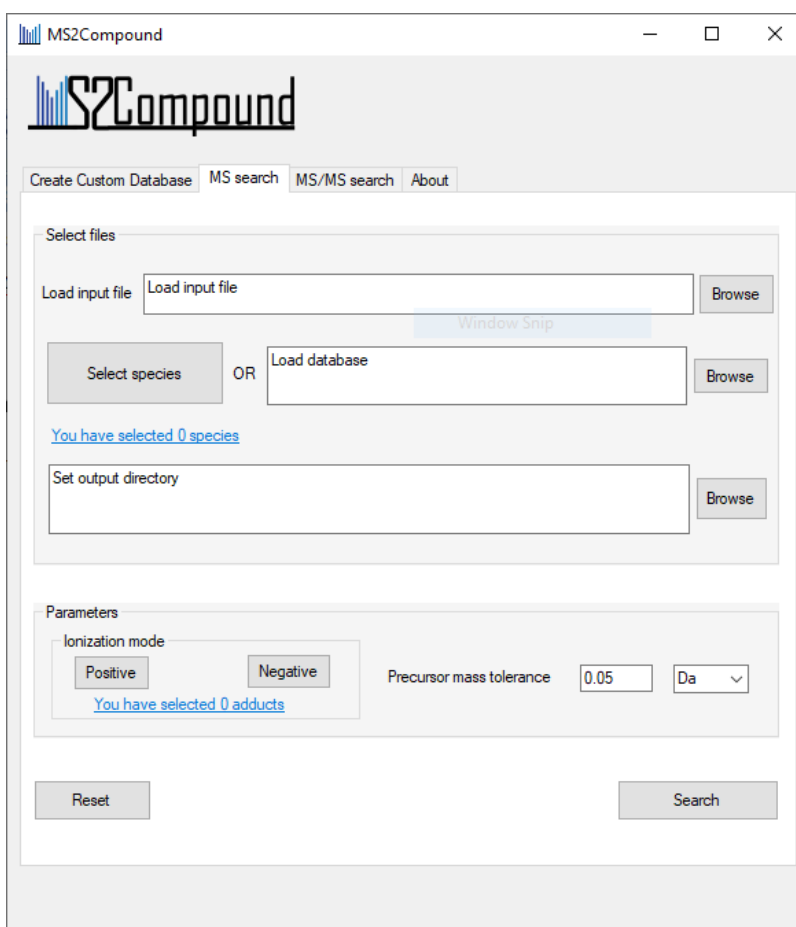
Alongside the program will create a output file named “MonoisotopicMass.txt” for the calculated monoisotopic masses.

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

## MS-Search using in-built Database

Click on MS search tab. This will open a window as following:



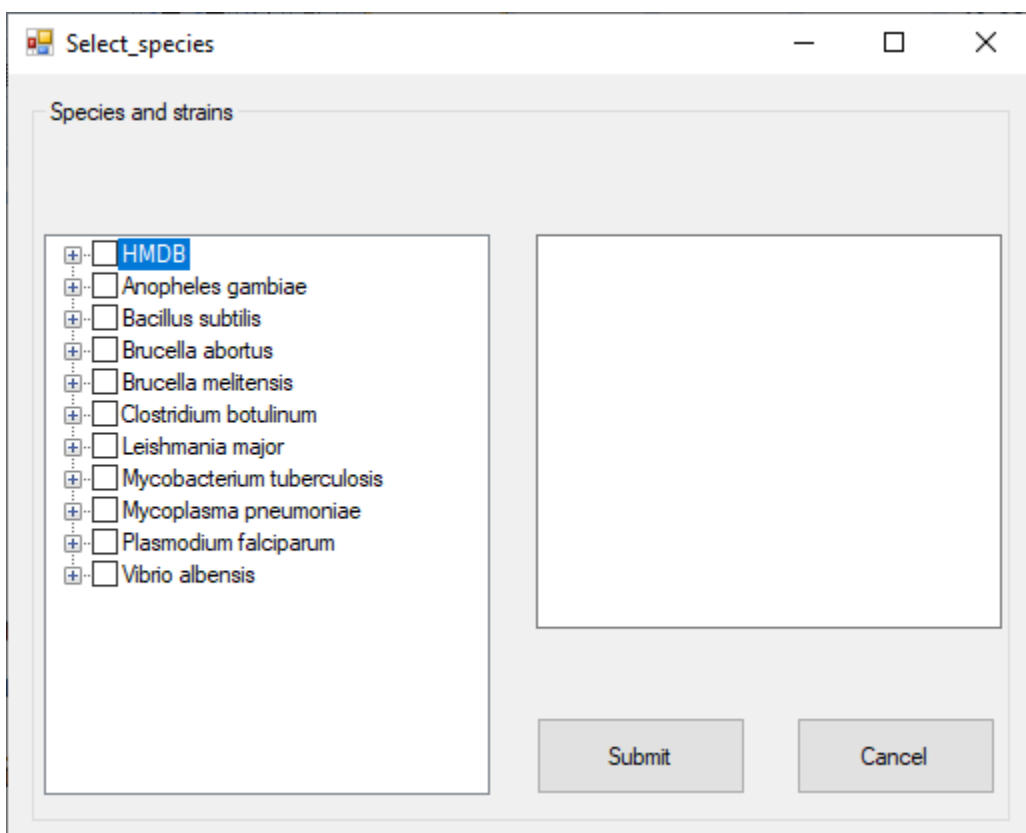
The screenshot shows the MS2Compound application window with the 'MS search' tab selected. The window has a title bar with standard OS controls. Below the title bar is a logo and a tabbed interface with 'Create Custom Database', 'MS search', 'MS/MS search', and 'About'. The 'MS search' tab is active, showing a 'Select files' section with a 'Load input file' text box and a 'Browse' button. Below this is a 'Select species' button, an 'OR' label, and a 'Load database' text box with a 'Browse' button. A status message 'You have selected 0 species' is displayed. The 'Set output directory' text box also has a 'Browse' button. The 'Parameters' section includes an 'Ionization mode' dropdown with 'Positive' and 'Negative' options, a 'Precursor mass tolerance' text box with '0.05' and a unit dropdown with 'Da' selected. A status message 'You have selected 0 adducts' is shown. At the bottom are 'Reset' and 'Search' buttons.

The input file should be a tab separated file with three columns; first column containing unique identifier, 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:

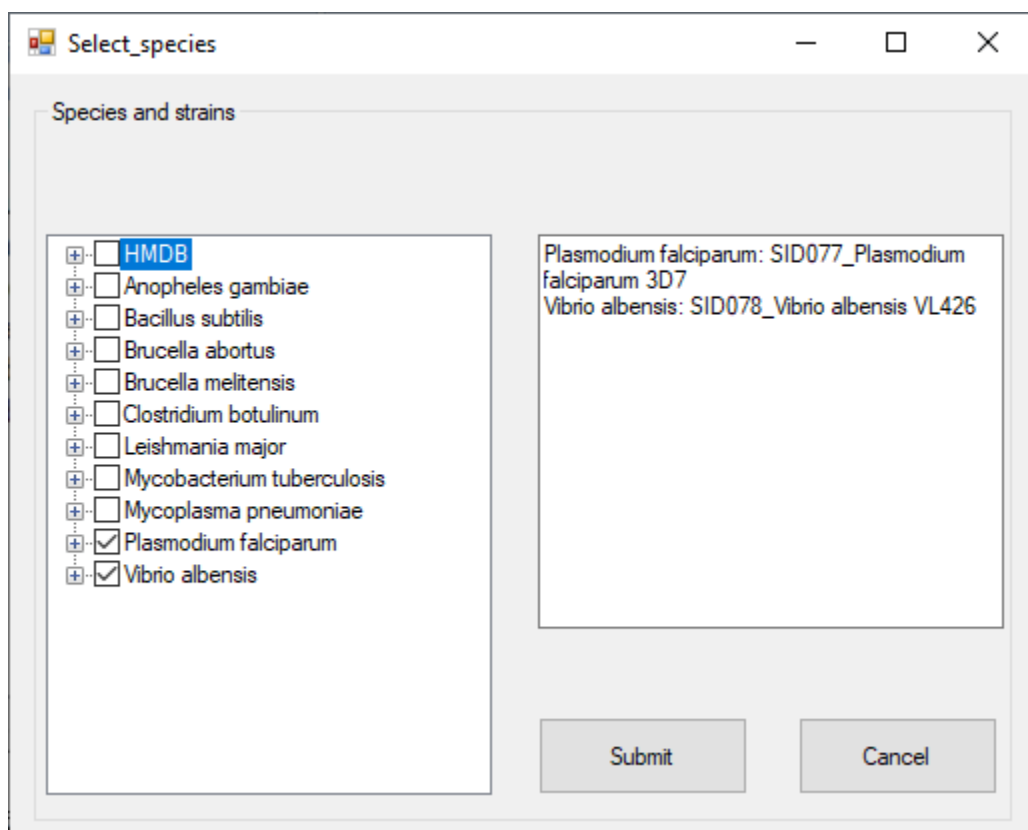
1	145.54	1+
2	530.3015	1+
3	758.2237	1+
4	336.2166	2+
5	747.3701	1+
6	345.2206	2+
7	445.2429	2+
8	536.1644	
9	610.1874	
10	182.9848	1+

If the charge information is provided 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. This will open a window as following:



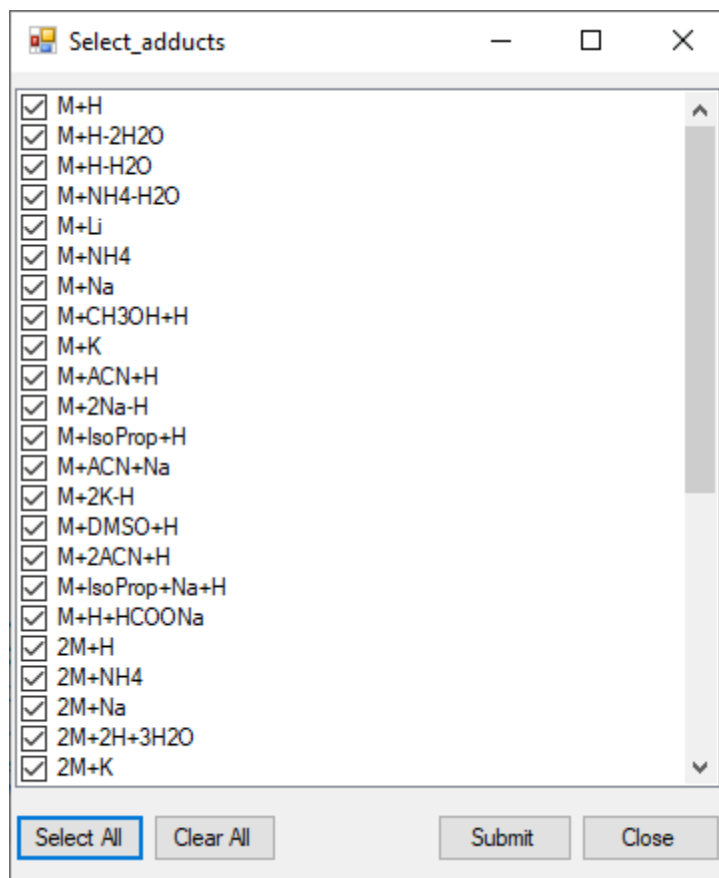
User can select any species of interest, or the search can be strain specific by clicking the “+” sign of species, and then select the desired strain. The species selection is shown as following:



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”

This generates three output files in selected output directory.

1. Assigned.txt
2. Assigned.html
3. Unassigned.txt

The sample output files can be found at  
“[https://github.com/beherasan/MS2Compound/tree/master/TestData/MS\\_InBuiltDB/OUTPUT](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 as 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 unique identifier, and the second one for corresponding monoisotopic mass.

Remaining parameters and the output will be similar as 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 in-built MS/MS database. The current version (1.0.0) of MS2Compound supports Mascot Generic Format (MGF) file as input.

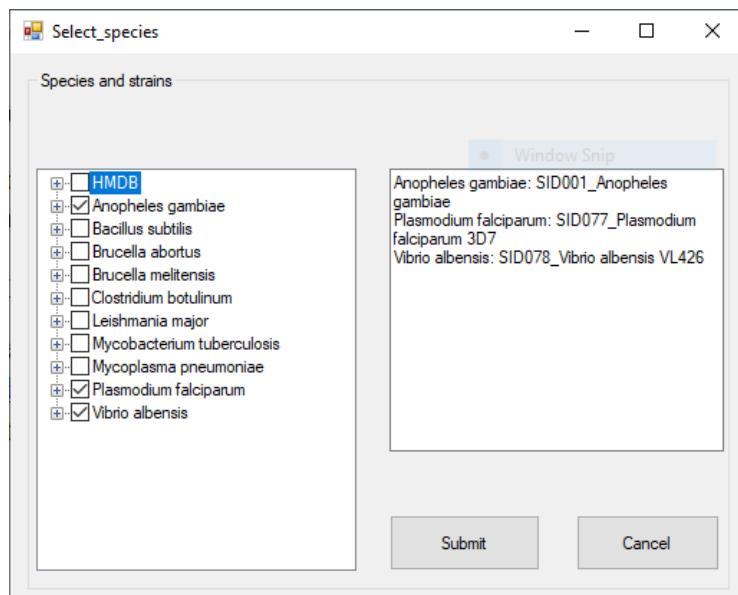
The screenshot for a sample input is shown below:

```
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 following steps:



- Load MGF file as Query file. The sample input is provided at [https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS\\_InBuiltDB](https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_InBuiltDB)
- Select the species of interest by clicking “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 number of fragment matches.
- Select the CID energy for MS/MS fragmentation
- Select “Search” button

This will create a directory “MS2Compound\_OUTPUT” in 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 same search it will replace the results in “MS2Compound\_OUTPUT”. This directory contains the

This generates three output files in “MS2Compound\_OUTPUT” directory.

1. Assigned.txt
2. Assigned.html
3. Unassigned.mgf
4. allMapped.txt

The sample output files can be found at “[https://github.com/beherasan/MS2Compound/tree/master/TestData/MS\\_InBuiltDB/OUTPUT](https://github.com/beherasan/MS2Compound/tree/master/TestData/MS_InBuiltDB/OUTPUT)”

“Assigned.html” can be open with any browser. 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	Compound ID	mS-score	BioCyc UniqId	Common Name	Molecular Weight	Monoisotopic Mass	HMDB	PubChem
2	1	2M+H	1+	CID002969	1.99212	CPD-7035	2-phenylethanol	122.166	122.0731649431	HMDB33944	<a href="#">6054</a>
2	2	2M+H	1+	CID000058	1.96664	PHENYLETHYLAMINE	2-phenylethylamine	122.189	122.0969743904	HMDB12275	<a href="#">448751</a>
2	3	M+DMSO+H	1+	CID002075	1.91389	PHE	L-phenylalanine	165.191	166.0868036346	HMDB00159	<a href="#">6925665</a>
2	4	M+IsoProP+H	1+	CID002755	0.98871	L-EPINEPHRINE	(R)-adrenaline	184.214	184.0973683209	HMDB00068	<a href="#">4102253</a>
2	5	M+IsoProP+H	1+	CID000245	0.77687	CPD-11875	normetanephrine	184.214	184.0973683209	HMDB00819	<a href="#">1237</a>
2	6	M+IsoProP+Na+H	1+	CID002572	0.46007	N-ACETYL-L-24-DIAMINOBUTANOATE	N-acetyl-L-2,4-diaminobutanoate	160.172	161.092617294		<a href="#">25245066</a>
2	7	M+IsoProP+H	1+	CID000272	0.36721	CPD-11497	3-methoxy-4-hydroxyphenylglycol	184.191	184.0735588736	HMDB01490	<a href="#">10805</a>
2	8	M+NH4	1+	CID001879	0.08235	SMA-0000133	SMA-0000133		227.0694765487		

The assigned compounds to a particular Feature ID are ranked based on the *mS-score* and sorted in order.

“Assigned.txt” has the 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 (1.0.0) 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](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](https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_CustomDB)) instead of selecting species of interest.

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

The MS/MS fragment file names should be UID.log in “ReferenceDB”. Please maintain the naming for 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.

The result file will be similar as 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](https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_CustomDB/MS2Compound_OUTPUT)