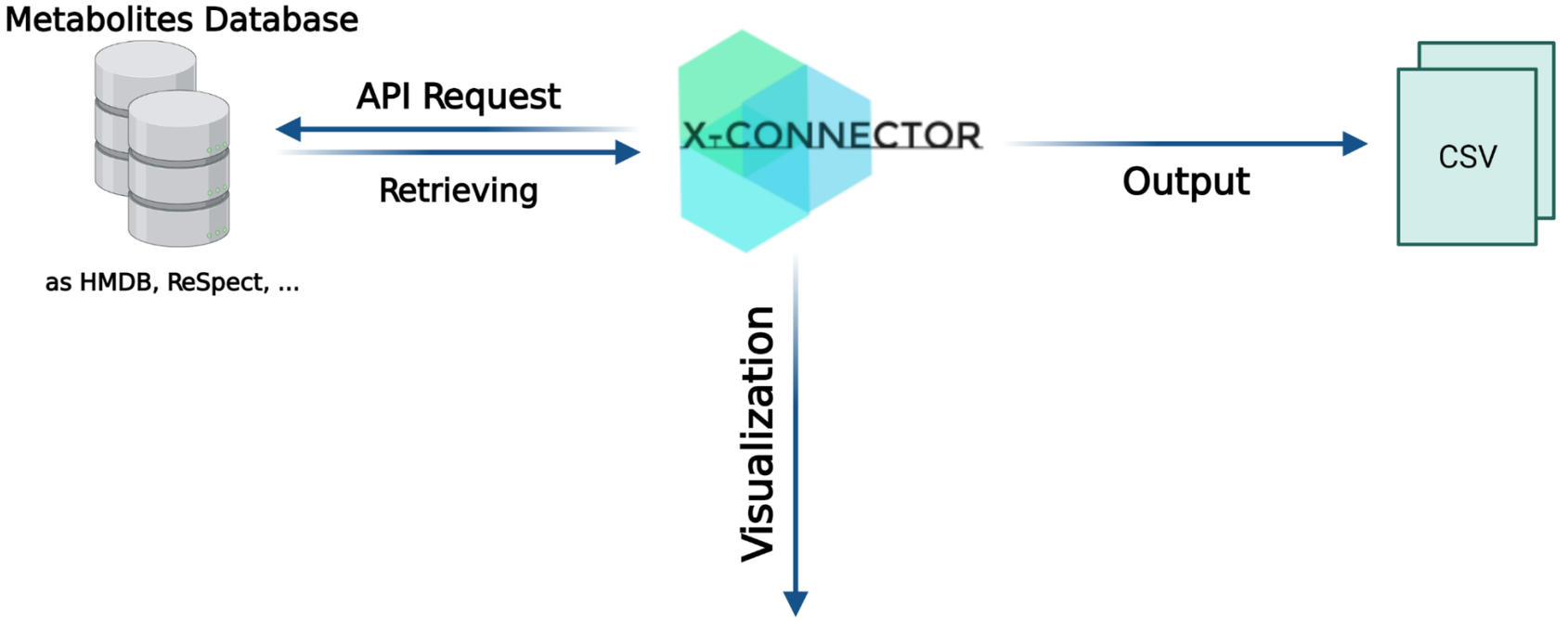
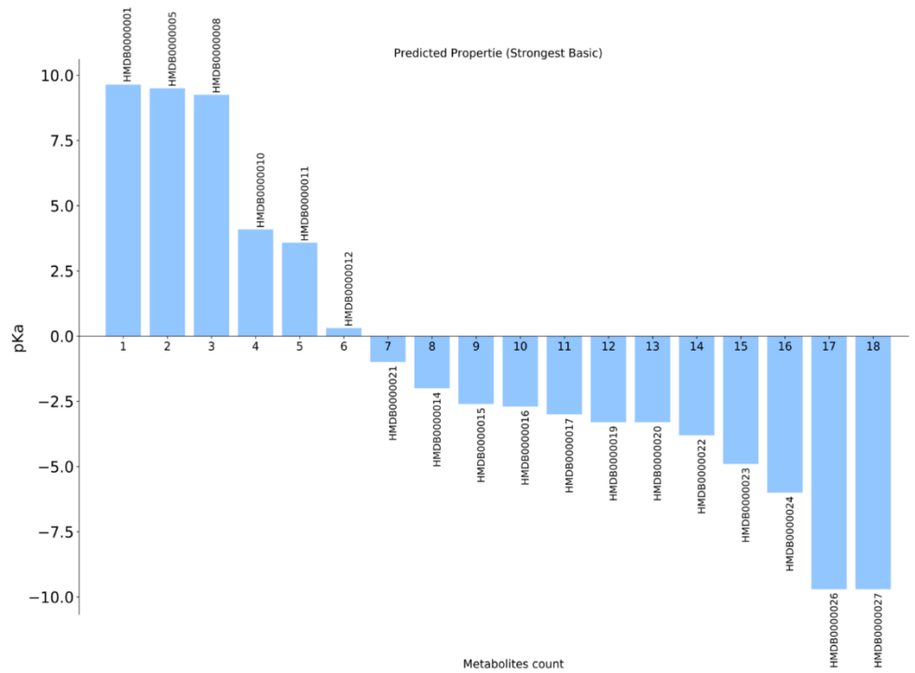
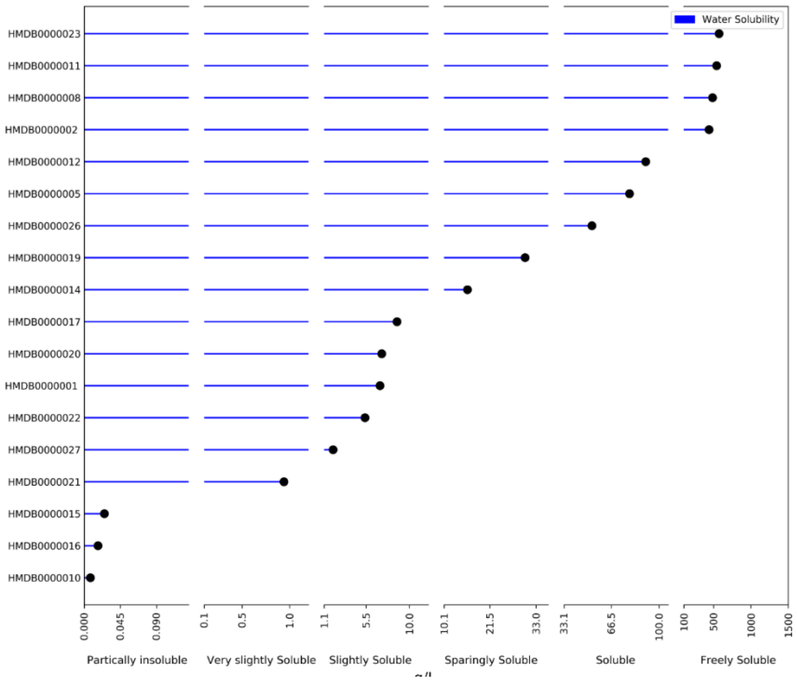






Documentation

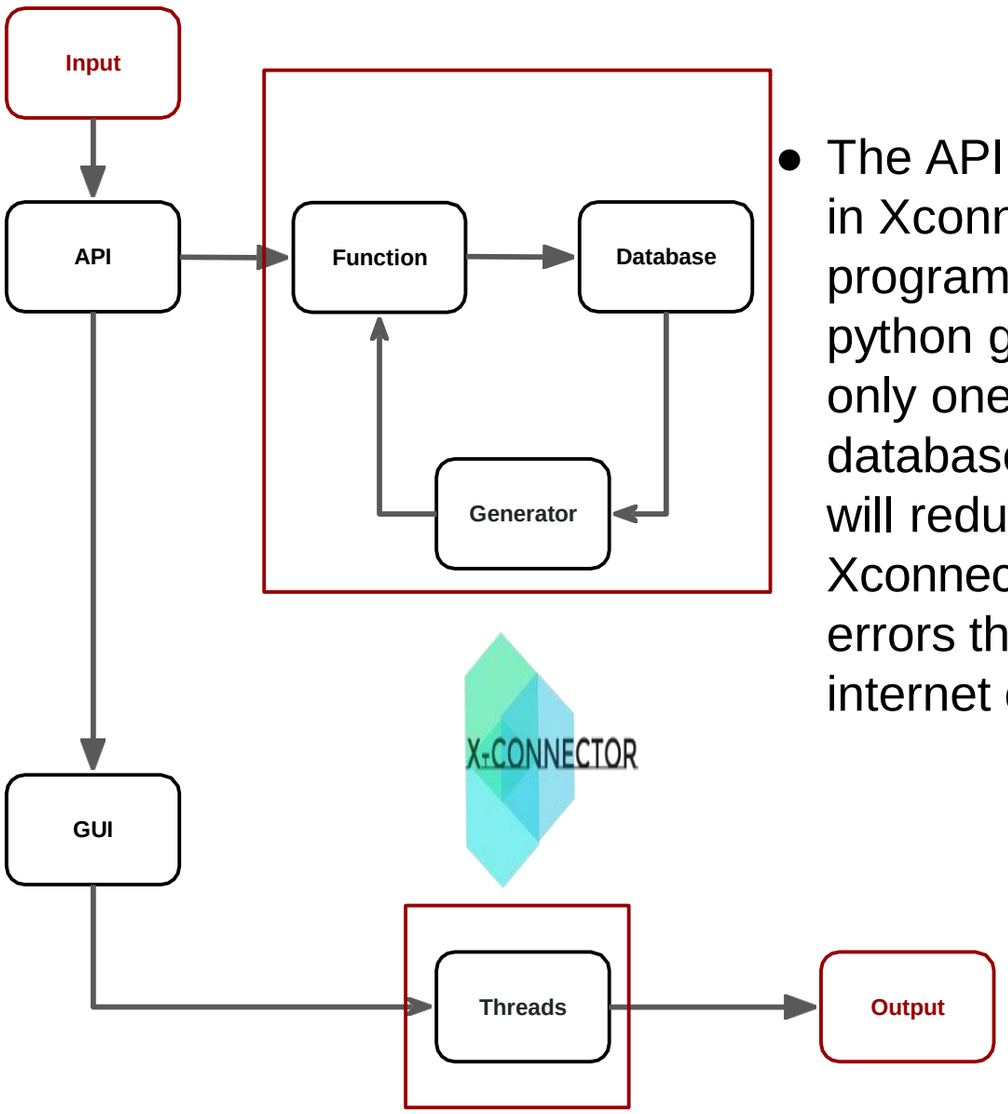
Introduction

Xconnector is a software package designed to easily retrieve, and visualize metabolomics data from different database sources. The goal of **Xconnector** is to connect different metabolomics databases in one place. The nine databases implemented in **Xconnector** are:

1. The Human Metabolome Database (HMDB).
2. The Livestock Metabolome Database (LMDB).
3. The Yeast Metabolome Database (YMDB).
4. The *Toxin and Toxin Target Database (T3DB).*
5. ReSpect for Phytochemicals DataBase.
6. KEGG: Kyoto Encyclopedia of Genes and Genomes.
7. The Small Molecule Pathway Database (SMPDB).
8. The Blood Exposome Database (BEDB).
9. The Phenol-Explorer Database (PEDB).

In future, we aim to include the most used databases for metabolites data.

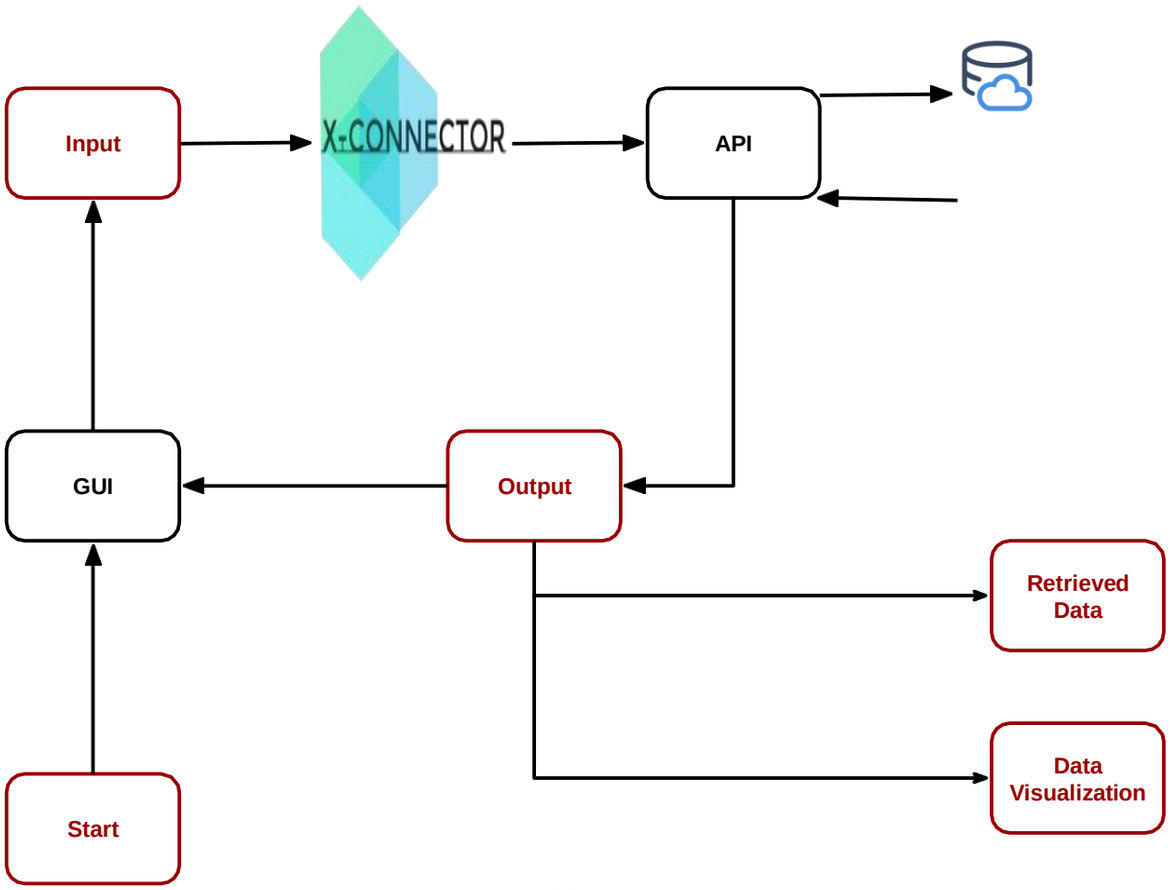
API & GUI Implementation



- The API function connects databases in Xconnector is made to be programmatically efficient. Using python generators implementation, only one query is called from the database each time by the API. This will reduce the memory used by Xconnector, as well as overcome the errors that could occur during the slow internet connection.

- After the API sends the output to the GUI. Xconnector utilises multithreading to allow efficient execution for the GUI, which allow multitasking and converting data between the GUI and the API.

Implementation workflow



- This workflow summarizes how Xconnector works. Using a graphical user interface (GUI), the user could import different type of input as database IDs or keyword from a CSV file or text file. Then send it to the application programming interface (API), the API connect one or all databases to retrieve and parse the information that made hits with the user input. Finally, the API send it back to the GUI to be displayed for the user.

Software requirements

- **A main requirement is a speed internet connection.**

- **For Windows:**
  1- Windows 10 64-bit system.

For Windows 10 users, two options to run Xconnector:
a. Download it as normal Windows 10 software (executable program), with no need for any dependencies.
b. Install it as a python package using pip. All the dependencies needed will be installed automatically (Note: Python >=3.7 is required).

- **For Linux:**
  Install Xconnector as a python package using pip. All the dependencies needed will be installed automatically (Note: Python >=3.7 is required).

- **For Mac:**
  Install Xconnector as a python package using pip. All the dependencies needed will be installed automatically (Note: Python >=3.7 is required).

Software availability

- The source code for Xconnector is available at **GitHub**:

  https://github.com/Proteomicslab57357/Xconnector

- The Package can be downloaded using **pip** as:

  pip install Xconnector

- For Windows 10 users the **Executable** program can be downloaded from:

  https://github.com/Proteomicslab57357/Xconnector

  or

  https://www.57357.org/en/department/proteomics-unit-dept/in-house-bioinformatics-tools/

  or directly from

  https://beta.57357.org/wp-content/themes/57357/programs/SASA%20new%20Link.zip

How to download and install

- **For Windows 10 users only.**

1 Download the zip file from.

2 Extract it and double click on the icon named Install to run the installation.

3 Choose where to install Xconnector. And press install.

4 After, a shortcut called Xconnector will be created on you desktop.

- **For any operating system.**

Download and install Python >=3.7. Using pip download the Xconnector package as (pip install Xconnector). Then, Open python and run the package to run the GUI as:

```
>>> From Xconnector import GUI
>>> GUI( )
```

Databases available and different functions

In this version, 1.0.0 of Xconnector there are nine different databases:

1. The Human Metabolome Database (HMDB).
2. The Livestock Metabolome Database (LMDB).
3. The Yeast Metabolome Database (YMDB).
4. The Toxin and Toxin Target Database (T3DB).
5. ReSpect for Phytochemicals DataBase.
6. KEGG: Kyoto Encyclopedia of Genes and Genomes.
7. The Small Molecule Pathway Database (SMPDB).
8. The Blood Exposome Database (BEDB).
9. The Phenol-Explorer Database (PEDB).

Each with different functions. Also, Xconnector contains a function which connect the first 4 databases together. Hence, user, for example, can search on thus databases at the same time with different IDs or keywords. The functions for each database to retrieve, parse and visualize where be described in the next section.

The Start window

The start window contains 8 buttons:

1- X-connector button: to connect (HMDB, LMDB, YMDB, T3DB) together at the same time.

2- HMDB-connector button: to connect HMDB.

3- LMDB-connector button: to connect LMDB.

4- YMDB-connector button: to connect YMDB.

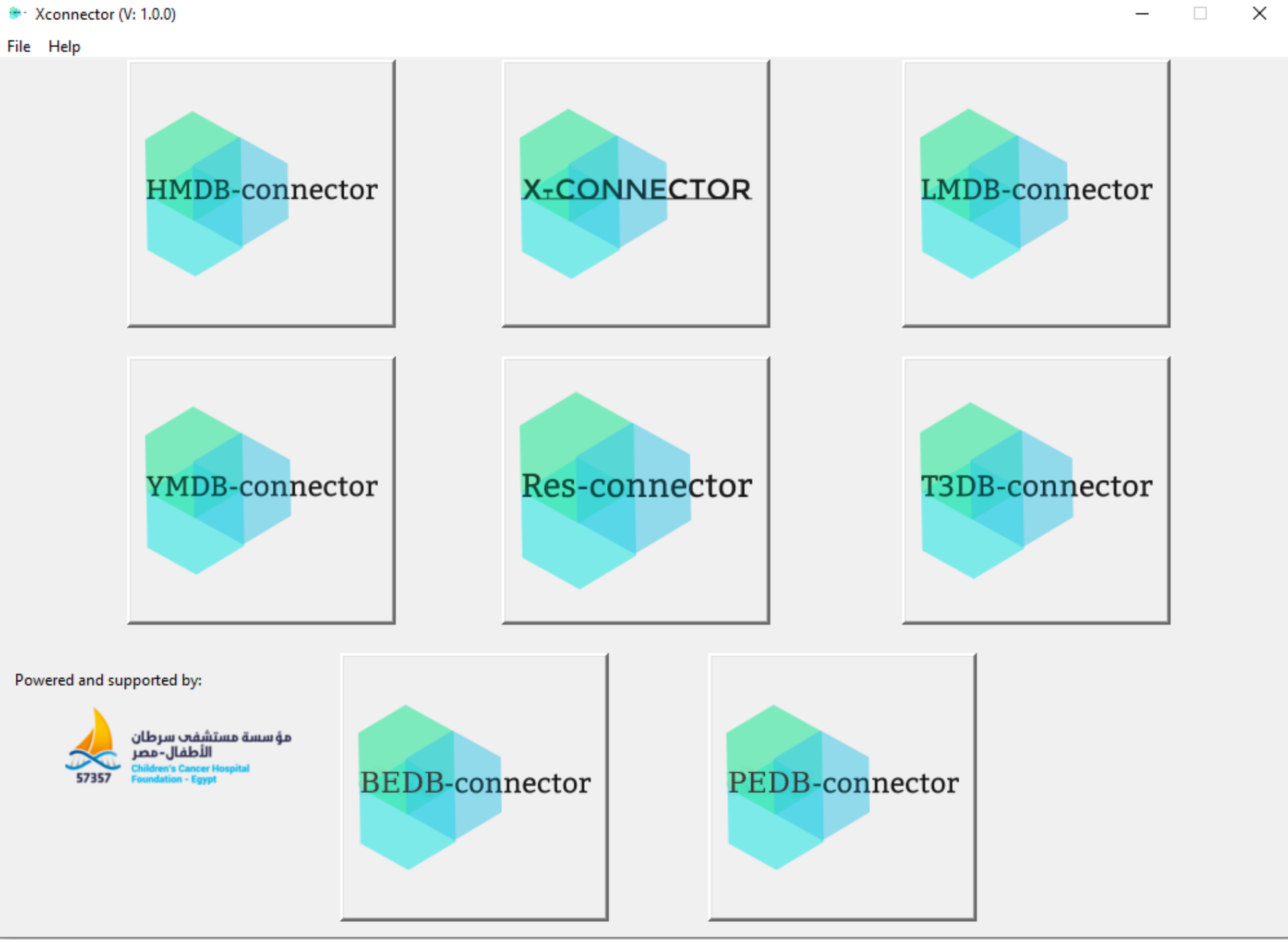
5- T3DB-connector button: to connect T3DB.

6- Res-connector button: to connect ReSpect for Phytochemicals.

7- BEDB-connector button: to connect Blood Exposome Database.

8- PEDB-connector button: to connect Phenol-Explorer Database.

The Start window



1- The Human Metabolome Database (HMDB)

1- Searching Using IDs.

Description:

This function is used to retrieve and parse different information from HMDB using IDs.

Input:

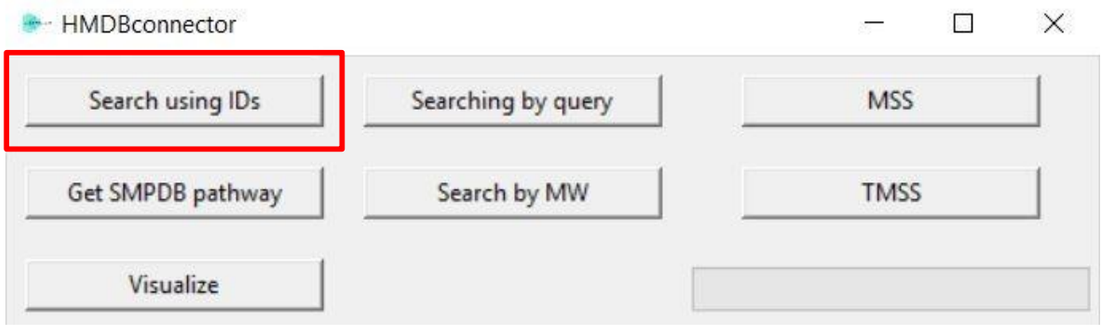
The input is CSV file with column name (IDs_HMDB) and HMDB IDs.

Output:

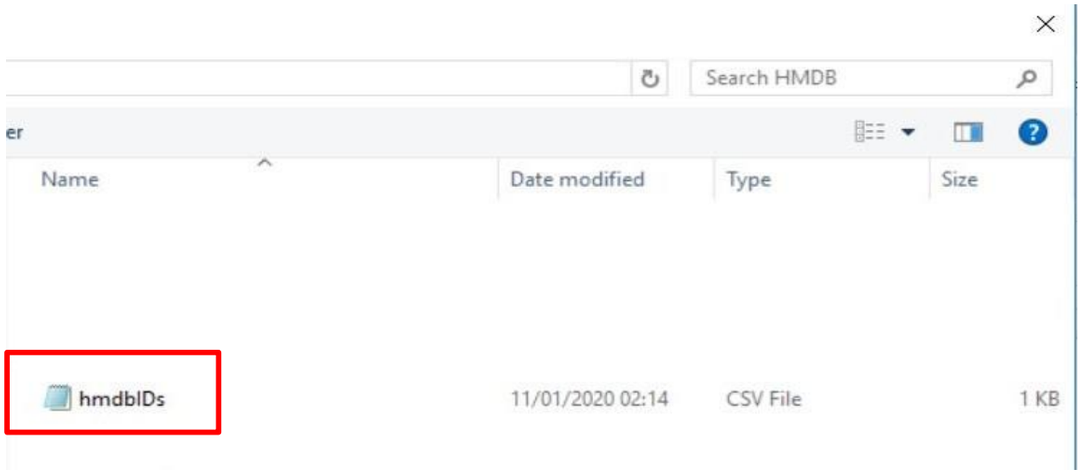
The outputs are 7 CSV files with different information:

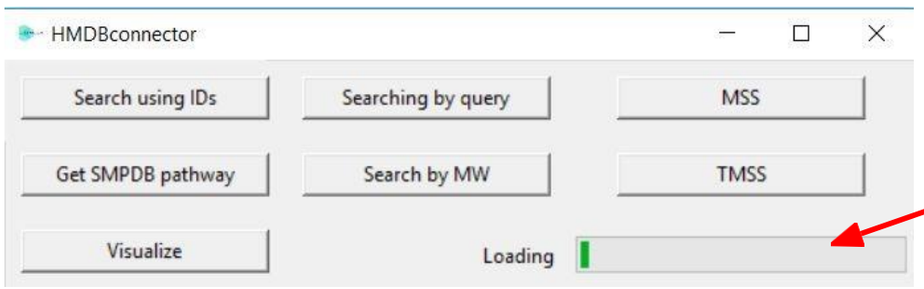
1- General information, 2- Synonyms information, 3- Experimental Properties, 4- Predicted Properties, 5- Normal Concentrations information, 6- Abnormal Concentrations information, and 7- Pathways summary information.

1- Select "Searching Using IDs"



2- Choose CSV contains HMDB ids with a column named "IDs_HMDB"





3- Wait until the loading bar finish

An example for the CSV input file

A
IDs_HMDB
HMDB0000001
HMDB0000002
HMDB0000005
HMDB0000008
HMDB0000010
HMDB0000011
HMDB0000012
HMDB0000021
HMDB0000014
HMDB0000015
HMDB0000016
HMDB0000017
HMDB0000019
HMDB0000020
HMDB0000021
HMDB0000022
HMDB0000023
HMDB0000024
HMDB0000026
HMDB0000027

Note: the header must be named IDs_HMDB

Sample file:
https://github.com/Proteomicslab57357/Xconnector/blob/master/examples/hmdbIDs.csv

2- Searching by query.

Description:

This function is used to retrieve and parse metabolites general information from HMDB using keyword (query).

Input:

The input is Keyword or query to search HMDB with.

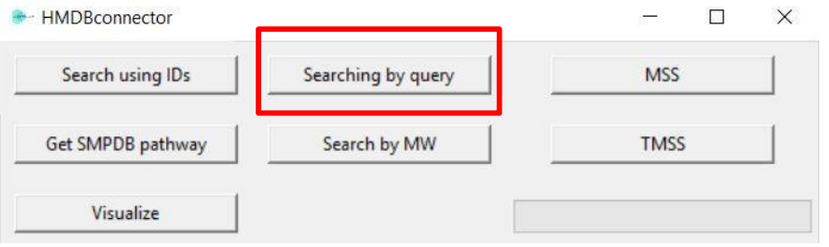
Output:

The output is one CSV file name "Search", contains metabolites general information, retrieved and parsed from the search using the user query.

Note:

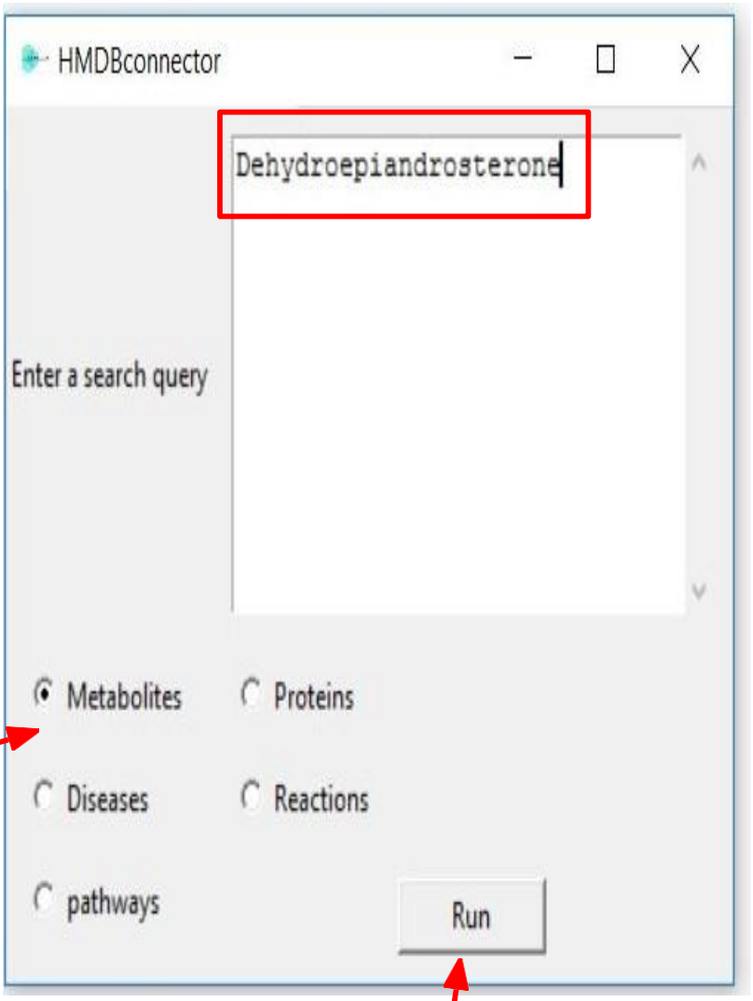
There are 5 different filters user can use (metabolites, proteins, disease, reaction, and pathways)

1- Select "Searching by query"

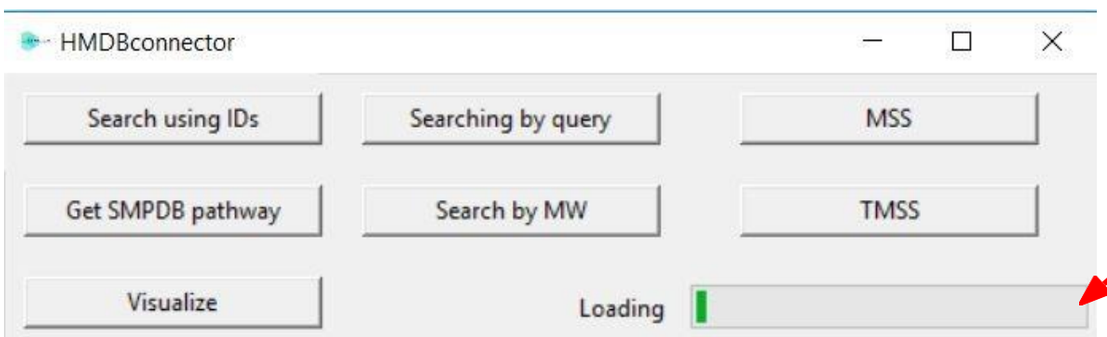


2- Enter a keyword to search with.

3-You have to choose a filter.



4- Press "Run" and wait until the loading bar finish



3- Searching by Molecular weight (MW).

Description:

This function is used to retrieve and parse metabolites general information from HMDB using molecular weight.

Input:

The inputs are a start MW and an end MW to search within.

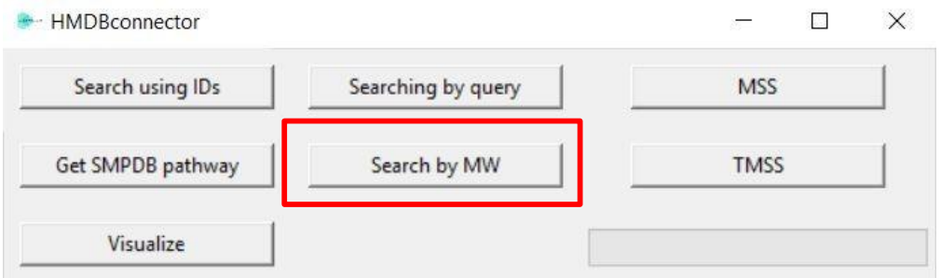
Output:

The output is one CSV file name "ChemQuery", contains metabolites general information for the result from the search by molecular weight.

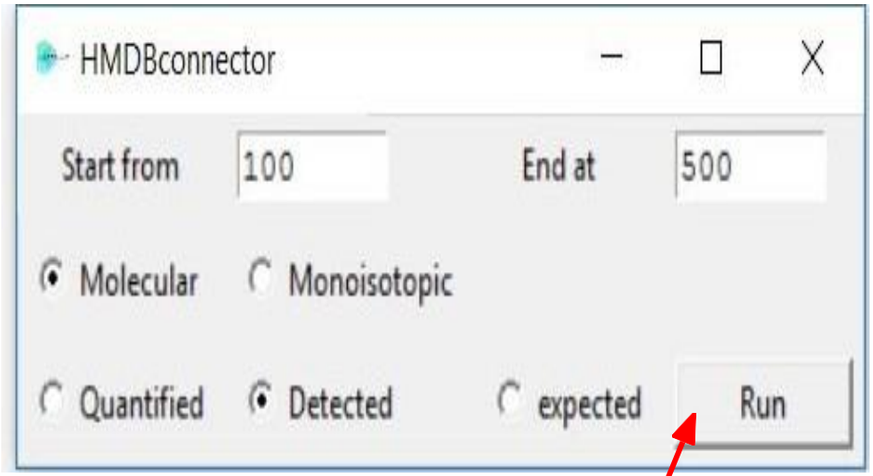
Note:

There are 2 groups of different filters the user has to select one from each. The first group (Molecular weight / Average mass or Monoisotopic mass). The second group (quantified, detected, or expected).

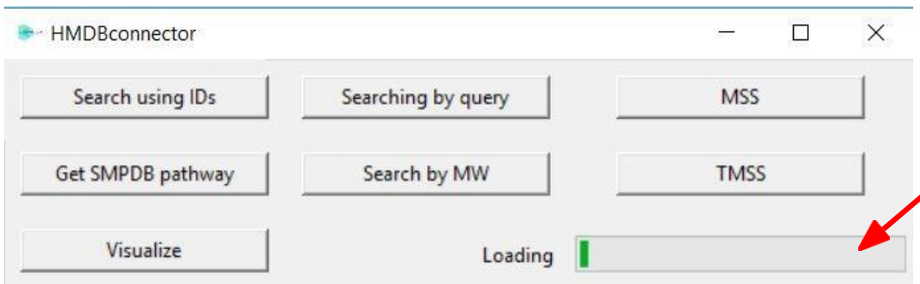
1- Select "Searching by MW"



2- Enter a start mass and the end mass you want to search with, then select the mass to unite you want to use, after, select a filter you want to use.



3- Press "Run" and wait until the loading bar finish



4- Mass Spectrum Search (MSS)

Description:

This function is used to retrieve and parse metabolites general information from HMDB using mass spectrum search (MSS).

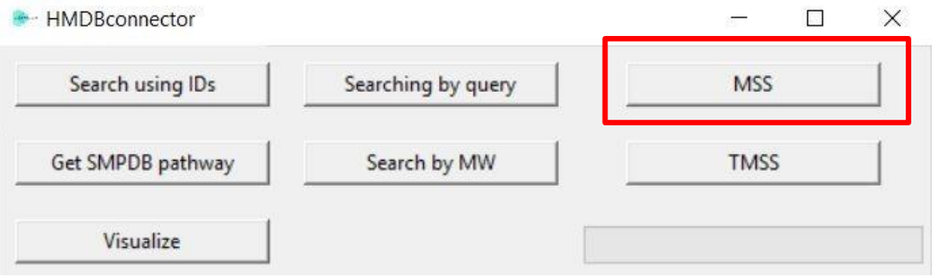
Input:

The inputs are masses to search with, ion mode, adducts, and molecular weight tolerance also the unit used for it.

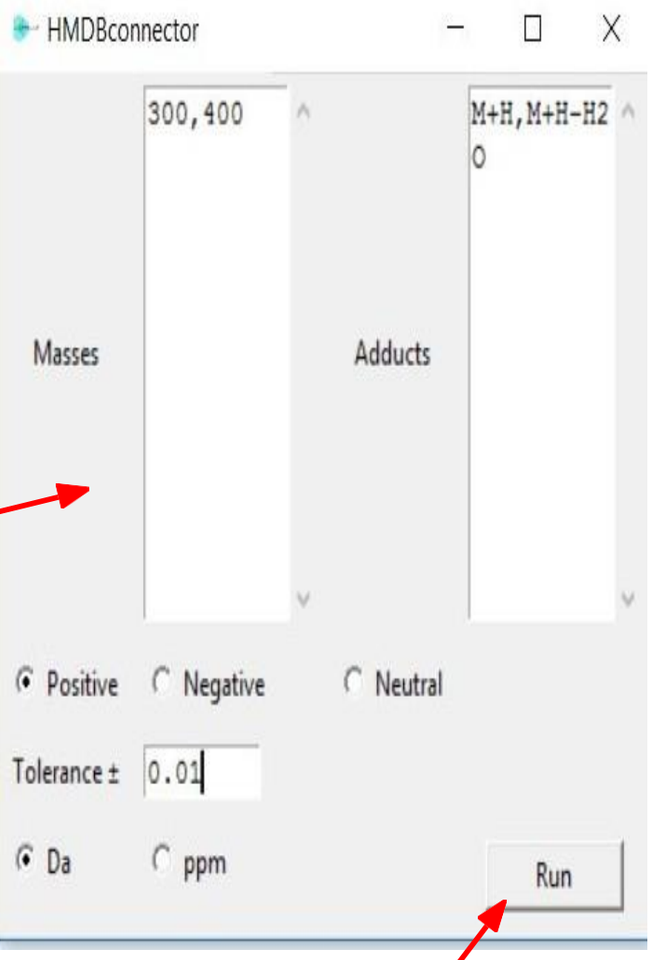
Output:

The output is one CSV file name "LC-MS", contains metabolites general information for the result from the search.

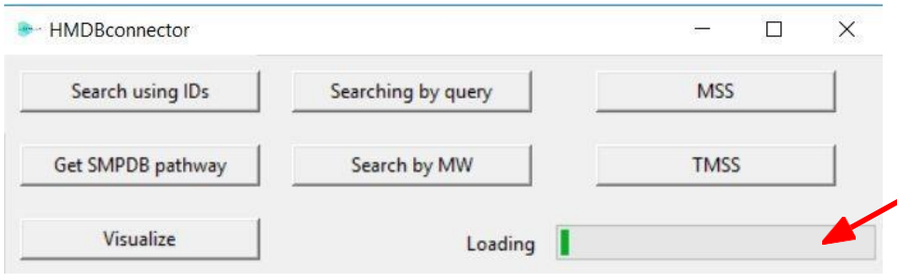
1- Select "MSS"



2- Enter masse(s) separated by a comma, enter adduct(s) separated by a comma. Then select an Ion mode. Finally, Enter a tolerance and select the unit you want to use.



3- Press "Run" and wait until the loading bar finish



5- Tandom Mass Spectrum Search (TMSS)

Description:

This function is used to retrieve and parse metabolites general information from HMDB using tandom mass spectrum search (MSS).

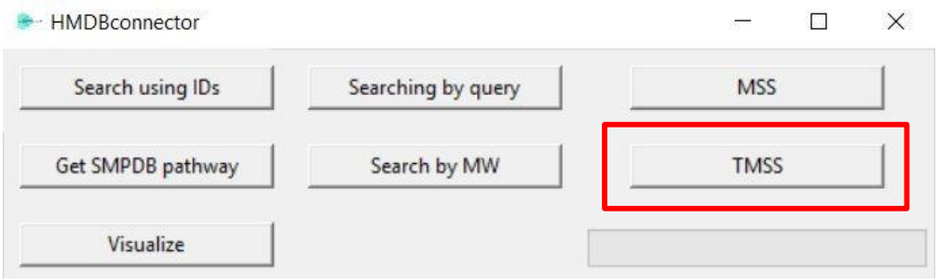
Input:

The inputs are MS/MS peak list, parent ion mass, parent ion mass tolerance ±, mass/charge (m/z) tolerance ±, ionization mode, and CID energy.

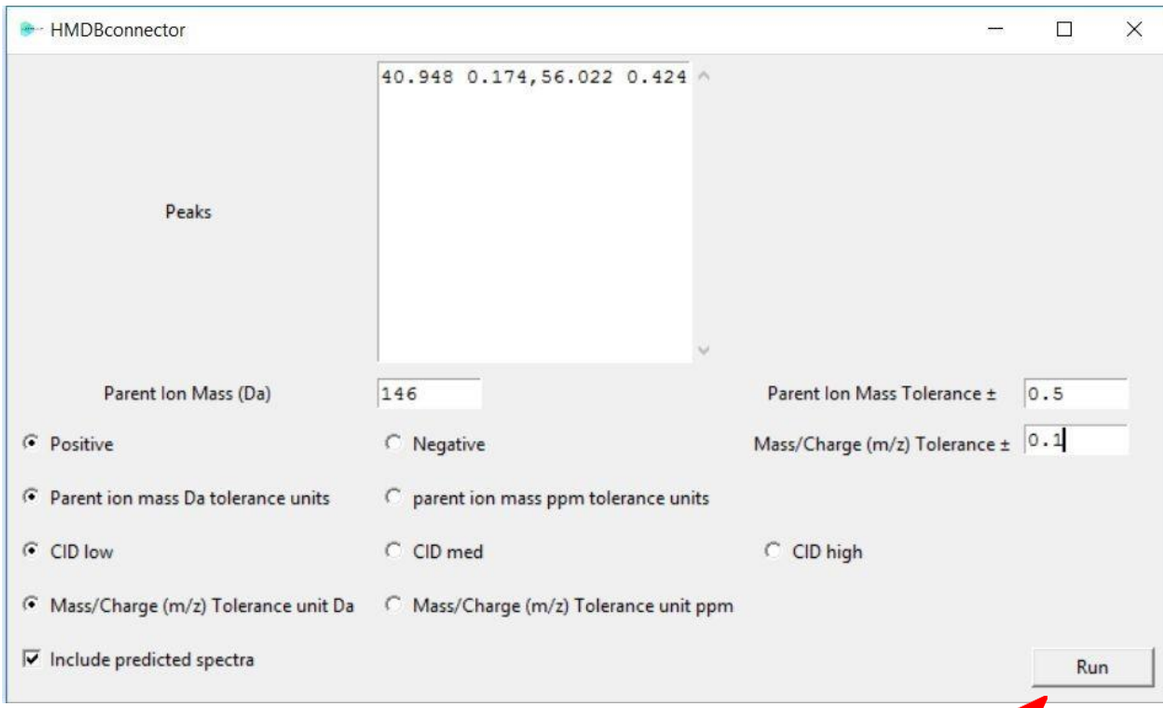
Output:

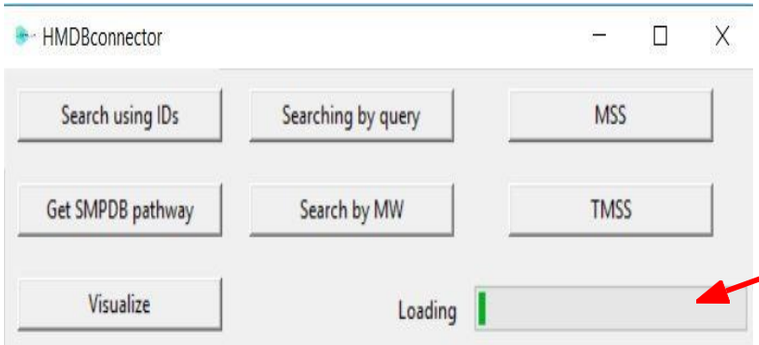
The output is one CSV file name "LC-MS/MS", contains metabolites general information for the result from the search.

1- Select "TMSS"



2- First, Enter masses and m/z for each with same format here separated by a comma. Enter the parent mass and its ion mass tolerance, also enter the m/z tolerance. Then, select the ion mode, the parent tolerance mass unit (Da or ppm), also select the m/z tolerance unit (Da or ppm). Finally, select the level for CID and if to include predicted spectra or not.





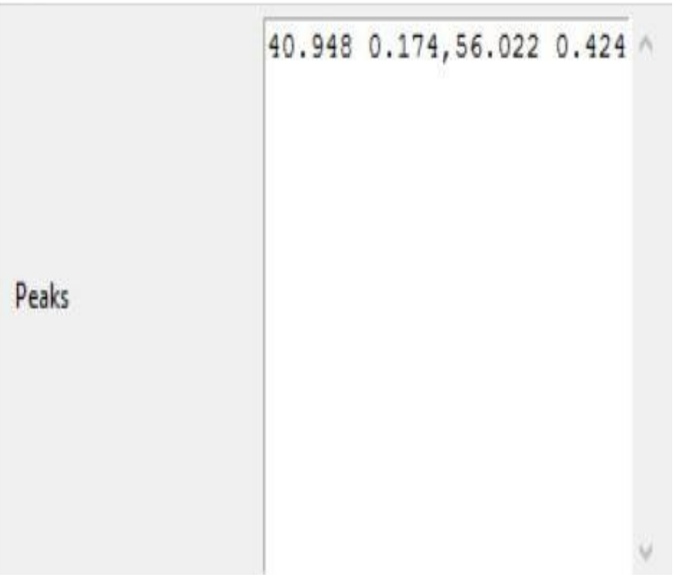
3- Press "Run" and wait until the loading bar finish

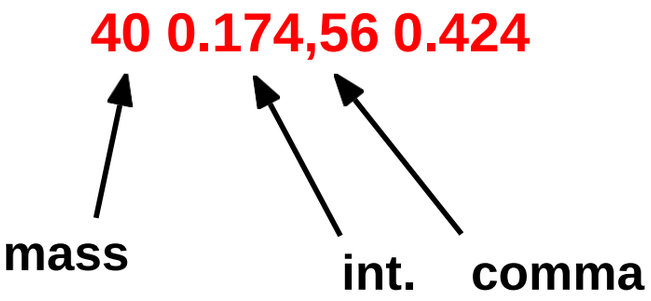
Note: You must enter the masses and the m/z in the correct format.

The correct format is:

mass[space]int.,mass[space]int., mass[space]int.,……

example:





6- Get SMPDB pathway

Description:

This function is used to retrieve and parse pathways information from SMPDB. Using metabolites IDs from HMDB.

Input:

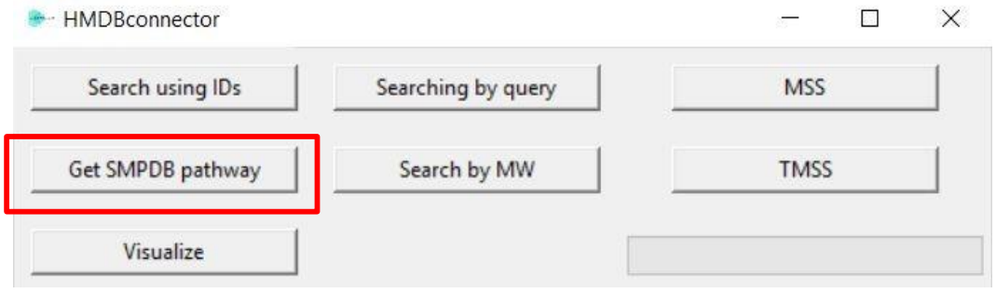
The input is CSV file with column name (IDs_HMDB) and HMDB IDs.

Output:

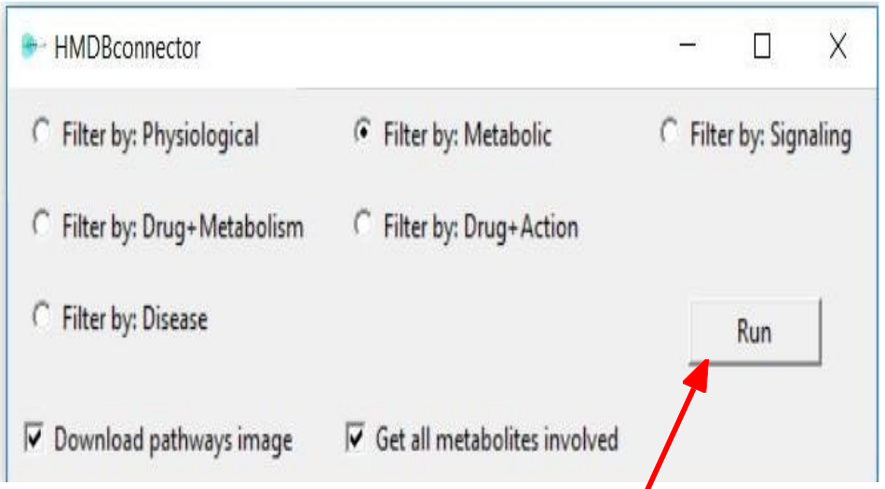
The main output is a CSV file named "SMPDB_HMDB_pathway" containing each metabolite from the input ids and the pathways this metabolite involve in. Also, the user can use different filters to search with as (Metabolic, Drug+Action, etc.).

As well as, there are two options (Download pathways image and Get all metabolites involved). For the first option, this could be selected if the user wants to download all pathways images that are appeared in the result. The second option is "Get all metabolites involved" to get all the metabolites that are involved in all pathways that appeared in the result.

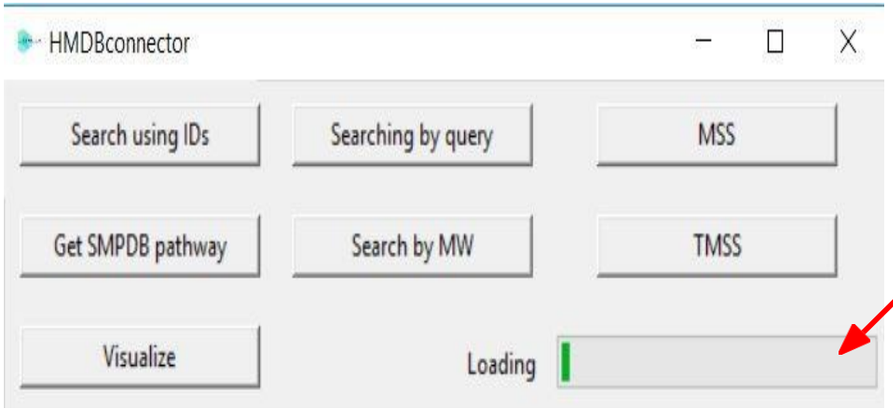
1- Select "Get SMPDB pathway"



2- First, Select a filter to search with or not. Two options (Download pathways image and Get all metabolites involved) you can select one from them, both or none. The first option allows Xconnector to download all the pathways (returned from the search) image in the form of SCV files. The second option allows Xconnector to retrieve all metabolites involved in each pathway returned from the search.



3- Press "Run" and wait until the loading bar finish



Visualization

For HMDB there are three categories of plots each display different information.

The first category named "Biological Properties".

It contains three different bar plots showing the **biospecimen locations**, **cellular locations** and **tissue locations**.

- Biospecimen locations
- Cellular locations
- Tissue locations

Examples : https://github.com/Proteomicslab57357/Xconnector/tree/master/examples/HMDB

The second category named "Disposition".

It contains six different bar plots showing the **Route of exposure**, **Source**, **Organ and components, Subcellular, Cell and elements, and Biofluid and excreta.**

Route of exposure example plot:

Source example plot:

Organ and components example plot:

Subcellular example plot:

Cell and elements example plot:

Biofluid and excreta example plot:

Examples :
https://github.com/Proteomicslab57357/Xconnector/tree/master/examples/HMDB

The third category named "Predicted Physical Properties".

It contains four different plots showing the **Water Solubility**, **pKa (Strongest Acidic)**, **pKa (Strongest Basic),** and both **pKa (Strongest Acidic)**, **pKa (Strongest Basic)**

- Water Solubility
- pKa (Strongest Acidic)
- pKa (Strongest Basic)
- pKa (Strongest Acidic) Vs. pKa (Strongest Basic)

Examples :

https://github.com/Proteomicslab57357/Xconnector/tree/master/examples/HMDB

2- The Livestock Metabolome Database (LMDB)

1- Searching Using IDs.

Description:

This function is used to retrieve and parse different information from LMDB using IDs.

Input:

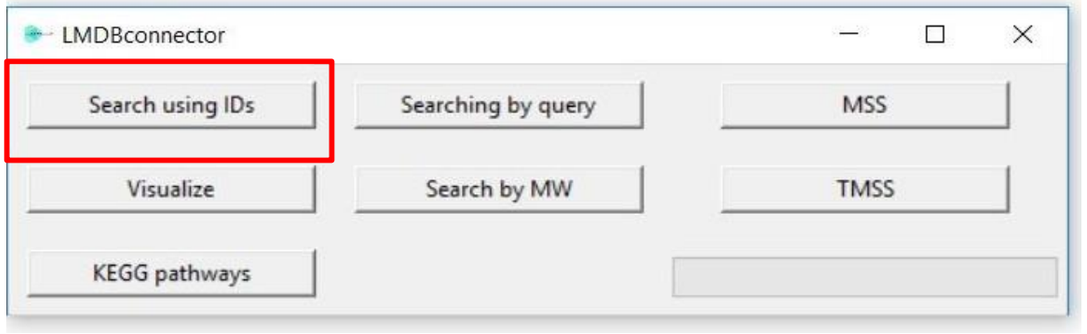
The input is CSV file with column name (IDs_LMDB) and LMDB IDs.

Output:

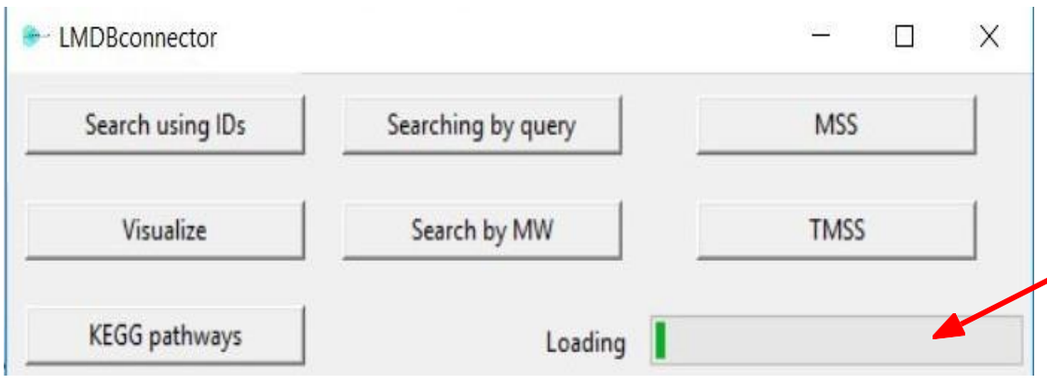
The outputs are 5 CSV files with different information:

1- General information, 2- Synonyms information, 3- Experimental Properties, 4- Predicted Properties, 5- Concentrations information.

1- Select "Searching Using IDs"



2- Choose CSV contains LMDB ids with a column named "IDs_LMDB"



3- Wait until the loading bar finish

An example for the CSV input file

A
IDs_LMDB
LMDB00001
LMDB00002
LMDB00048
LMDB00003
LMDB00004
LMDB00005
LMDB00006
LMDB00007
LMDB00008
LMDB00009
LMDB00010
LMDB00011
LMDB00012
LMDB00013
LMDB00014
LMDB00016
LMDB00017
LMDB00018
LMDB00019
LMDB00020
LMDB00021

Note: the header must be named IDs_LMDB

Sample file:
https://github.com/Proteomicslab57357/Xconnector/blob/master/examples/LMDB_ids.csv

2- Searching by query.

Description:

This function is used to retrieve and parse metabolites general information from LMDB using keyword (query).

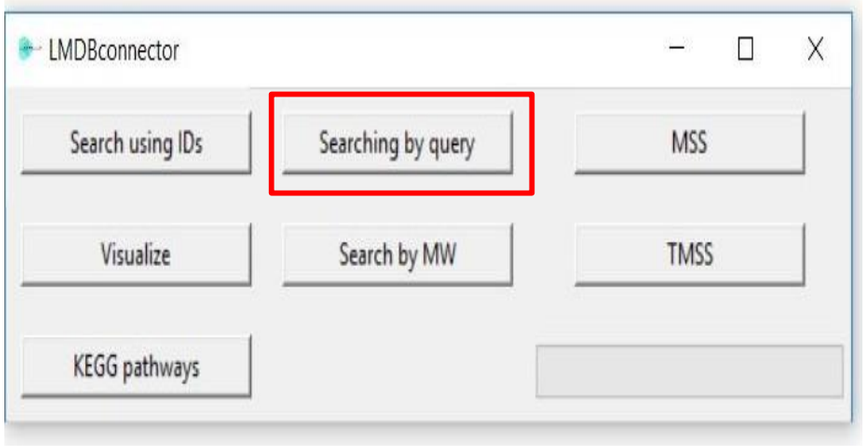
Input:

The input is Keyword or query to search LMDB with.

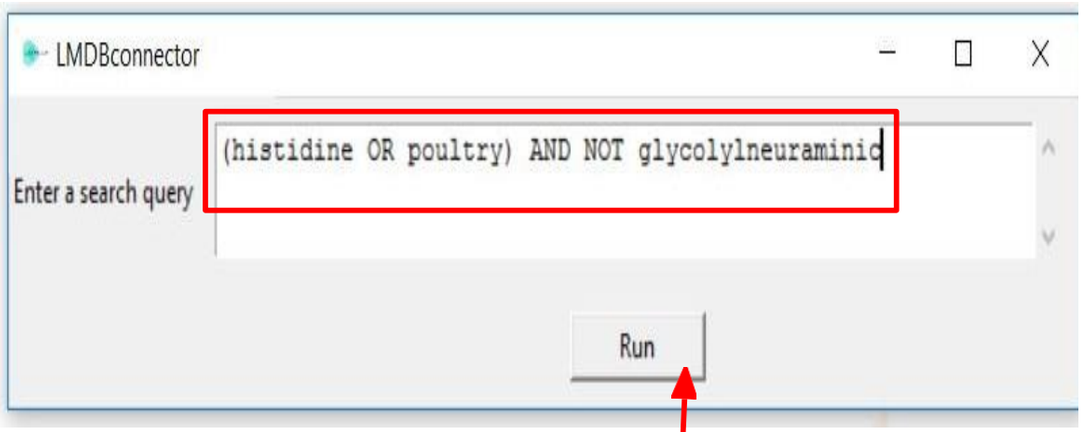
Output:

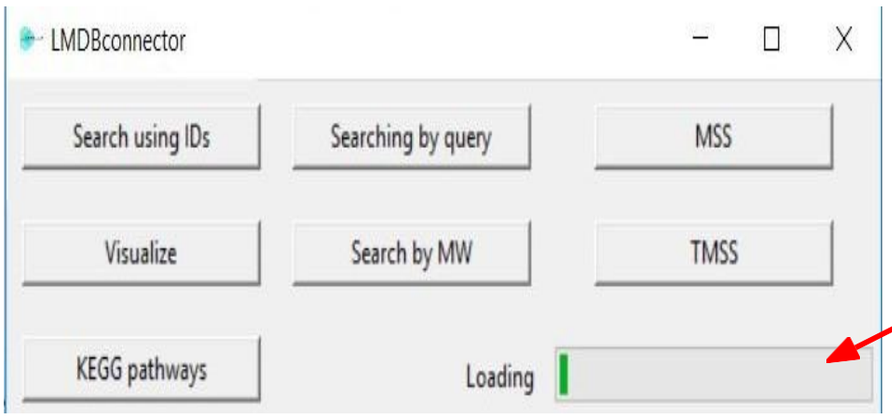
The output is one CSV file name "txtsearch", contains metabolites general information, retrieved and parsed from the search using the user query.

1- Select "Searching by query"



2- Enter a keyword to search with.





3- Press "Run" and wait until the loading bar finish

3- Searching by Molecular weight (MW).

Description:

This function is used to retrieve and parse metabolites general information from LMDB using molecular weight.

Input:

The inputs are a start MW and an end MW to search within.

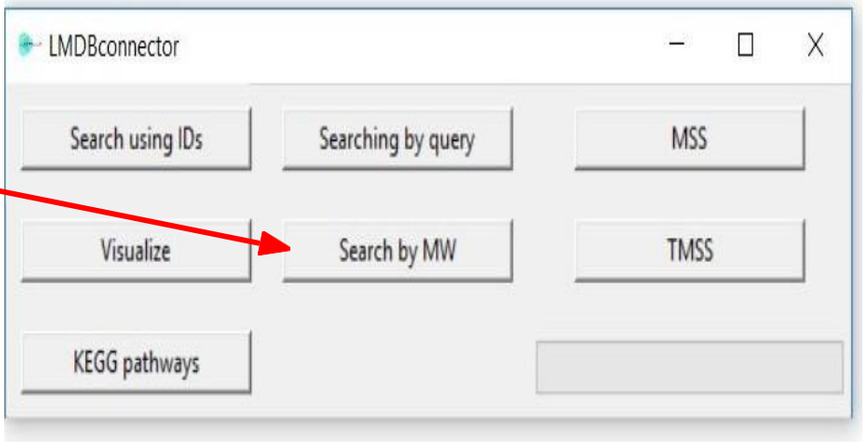
Output:

The output is one CSV file name "ChemQuery", contains metabolites general information for the result from the search by molecular weight.

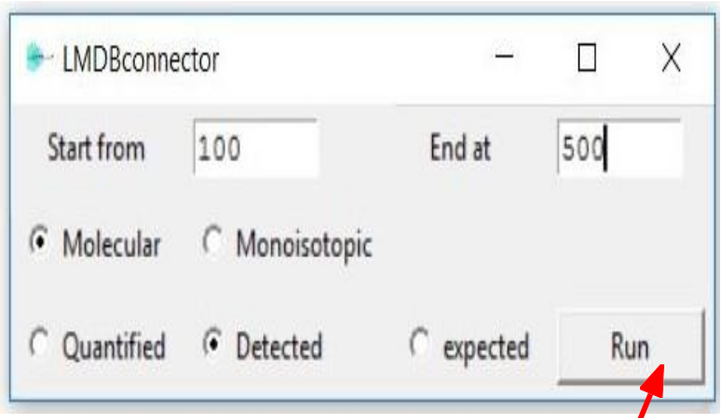
Note:

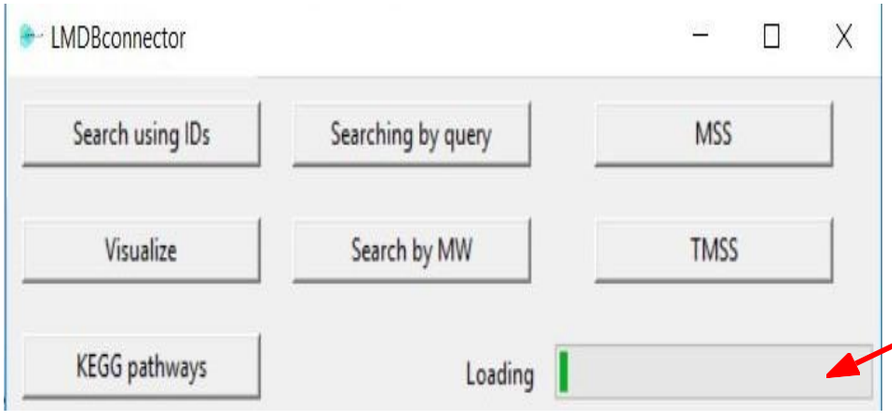
There are 2 groups of different filters the user has to select one from each. The first group (Molecular weight / Average mass or Monoisotopic mass). The second group (quantified, detected, or expected).

1- Select "Searching by MW"



2- Enter a start mass and the end mass you want to search with, then select the mass to unite you want to use, after, select a filter you want to use.





3- Press "Run" and wait until the loading bar finish

4- Mass Spectrum Search (MSS)

Description:

This function is used to retrieve and parse metabolites general information from LMDB using mass spectrum search (MSS).

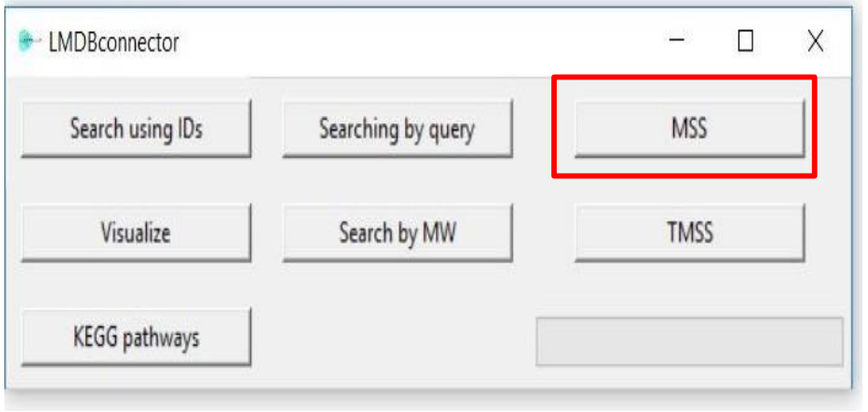
Input:

The inputs are masses to search with, ion mode, adducts, and Molecular Weight Tolerance also the unit used for it.

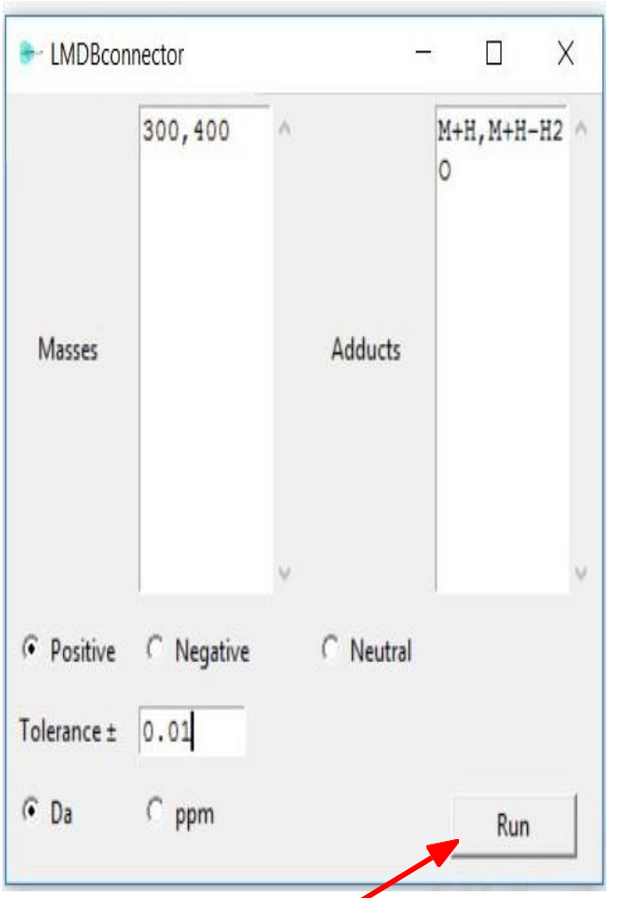
Output:

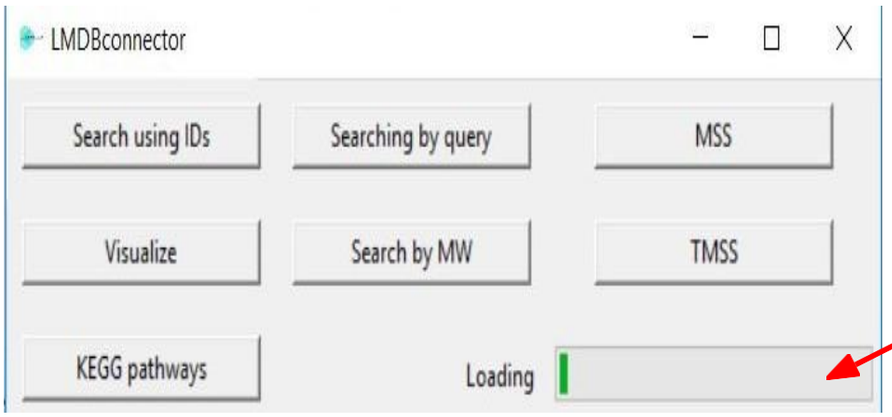
The output is one CSV file name "LC-MS", contains metabolites general information for the result from the search.

1- Select "MSS"



2- Enter masse(s) separated by a comma, enter adduct(s) separated by a comma. Then select an Ion mode. Finally, Enter a tolerance and select the unit you want to use.





3- Press "Run" and wait until the loading bar finish

5- Tandom Mass Spectrum Search (TMSS)

Description:

This function is used to retrieve and parse metabolites general information from LMDB using tandom mass spectrum search (MSS).

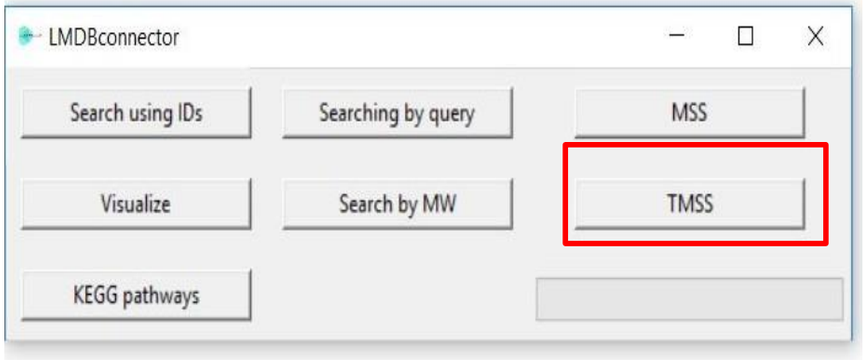
Input:

The inputs are MS/MS peak list, parent ion mass, parent ion mass tolerance ±, mass/charge (m/z) tolerance ±, ionization mode, and CID energy.

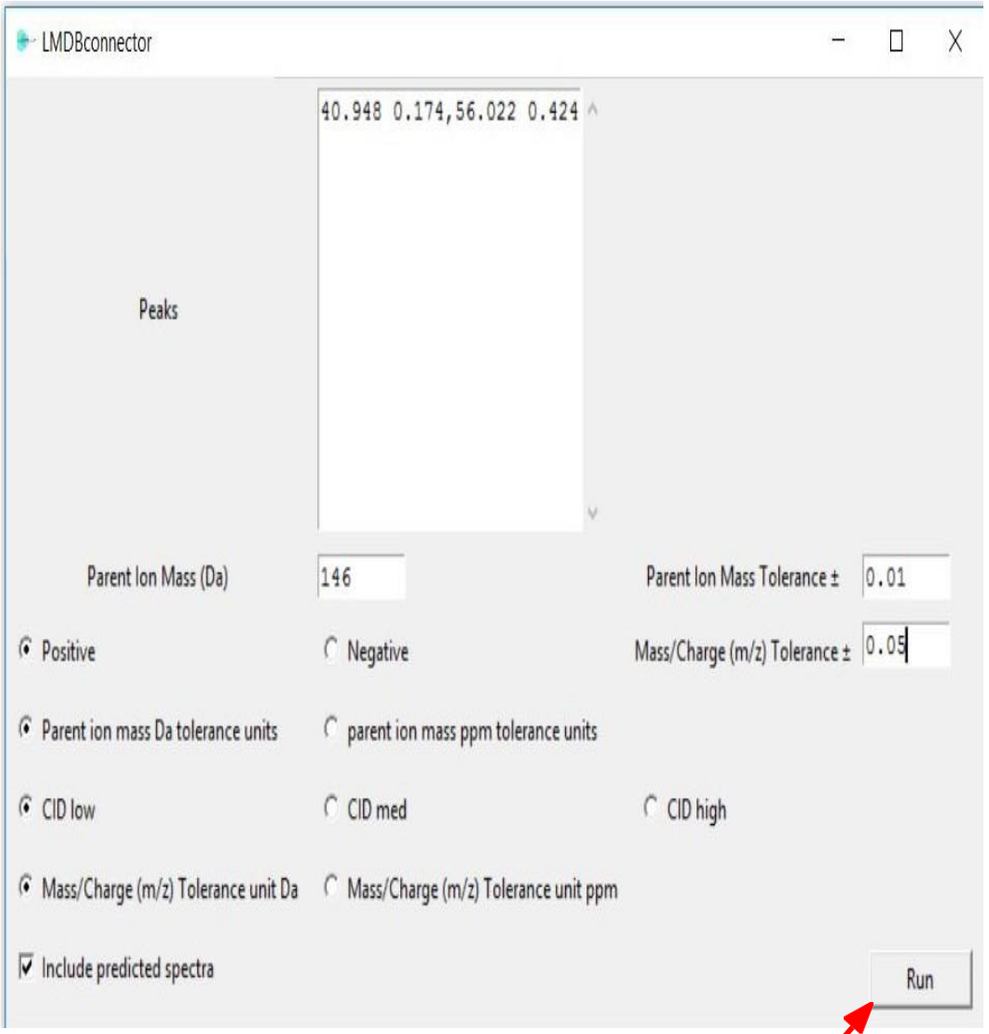
Output:

The output is one CSV file name "LC-MS/MS", contains metabolites general information for the result from the search.

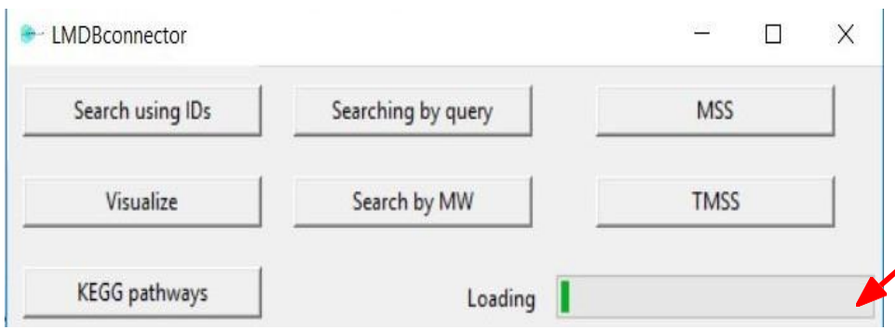
1- Select "TMSS"



2- First, Enter masses and m/z for each with same format here separated by a comma. Enter the parent mass and its ion mass tolerance, also enter the m/z tolerance. Then, select the ion mode, the parent tolerance mass unit (Da or ppm), also select the m/z tolerance unit (Da or ppm). Finally, select the level for CID and if to include predicted spectra or not.



3- Press "Run" and wait until the loading bar finish

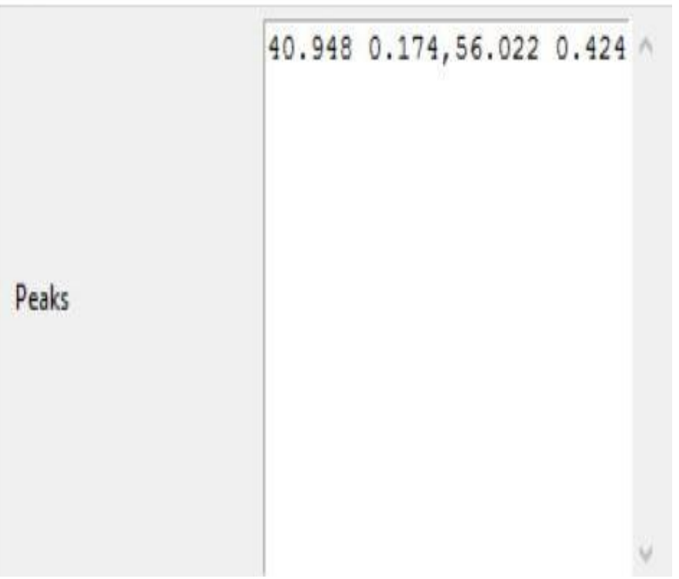


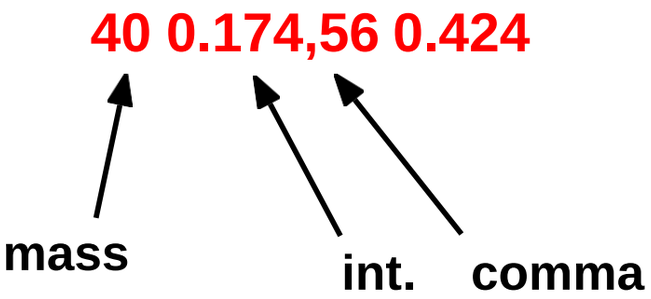
Note: You must enter the masses and the m/z in the correct format.

The correct format is:

mass[space]int.,mass[space]int., mass[space]int.,.....

example:





6- Get KEGG pathway

Description:

This function is used to retrieve and parse pathways information from KEGG. Using metabolites IDs from LMDB.

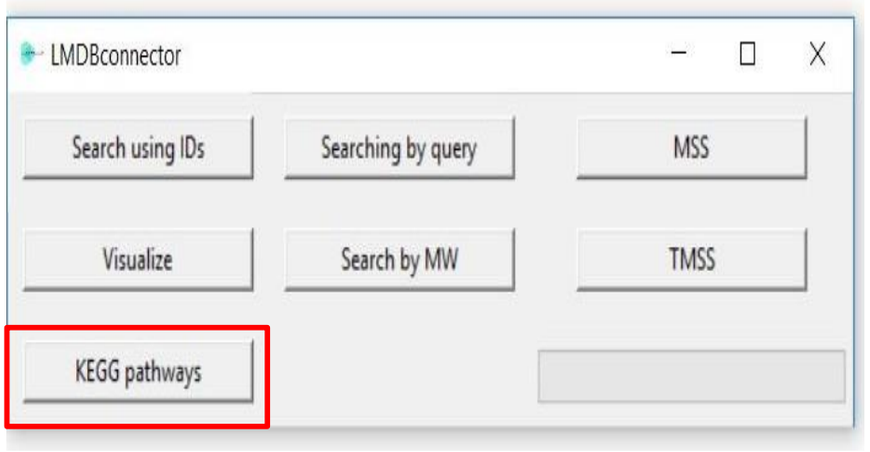
Input:

The input is CSV file with column name (IDs_LMDB) and LMDB IDs.

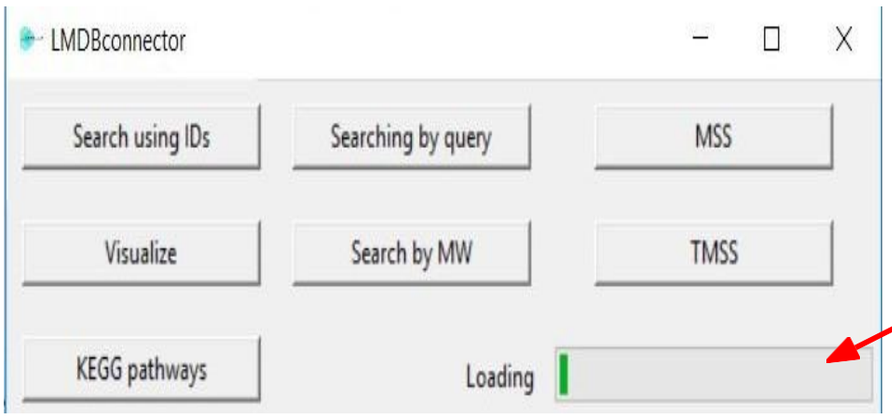
Output:

The main output is a CSV file named "Kegg_pathway" containing each metabolite from the input ids and the pathways this metabolite involve in.

1- Select "KEGG pathway"



2- Choose CSV contains LMDB ids with a column named "IDs_LMDB"



3- Wait until the loading bar finish

Visualization

For LMDB only one plot named “Biofluid Locations and Tissue Locations”, used to show the biofluid and tissue locations of the metabolites.

Biofluid Locations and Tissue Locations example plot: https://github.com/Proteomicslab57357/Xconnector/tree/master/examples/LMDB

3- The Yeast Metabolome Database (YMDB)

1- Searching Using IDs.

Description:

This function is used to retrieve and parse different information from YMDB using IDs.

Input:

The input is CSV file with column name (IDs_YMDB) and YMDB IDs.

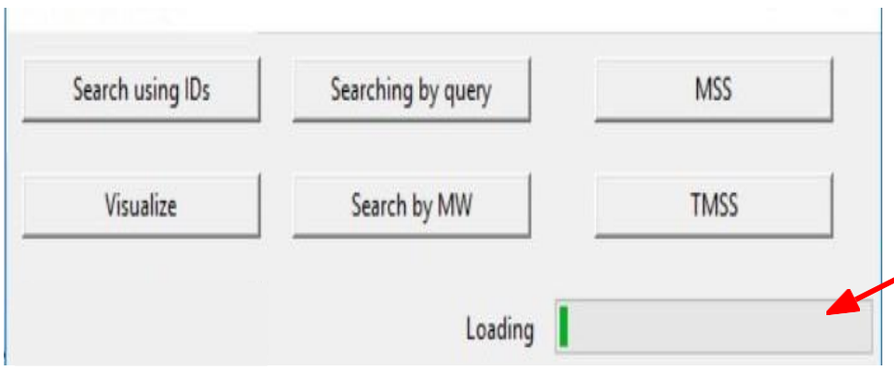
Output:

The outputs are 3 CSV files with different information:

1- General information, 2- Experimental Properties, 3- Predicted Properties.

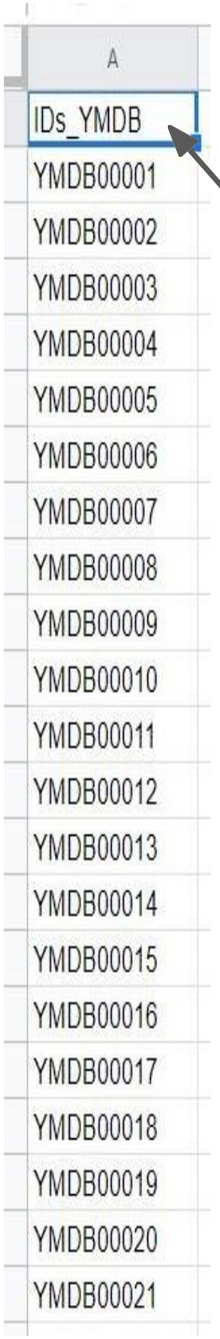
1- Select "Searching Using IDs"

2- Choose CSV contains YMDB ids with a column named "IDs_YMDB"



3- Wait until the loading bar finish

An example for the CSV input file



A
IDs_YMDB
YMDB00001
YMDB00002
YMDB00003
YMDB00004
YMDB00005
YMDB00006
YMDB00007
YMDB00008
YMDB00009
YMDB00010
YMDB00011
YMDB00012
YMDB00013
YMDB00014
YMDB00015
YMDB00016
YMDB00017
YMDB00018
YMDB00019
YMDB00020
YMDB00021

Note: the header must be named IDs_YMDB

Sample file:
https://github.com/Proteomicslab57357/Xconnector/blob/master/examples/YMDB_ids.csv

2- Searching by query.

Description:

This function is used to retrieve and parse metabolites general information from YMDB using keyword (query).

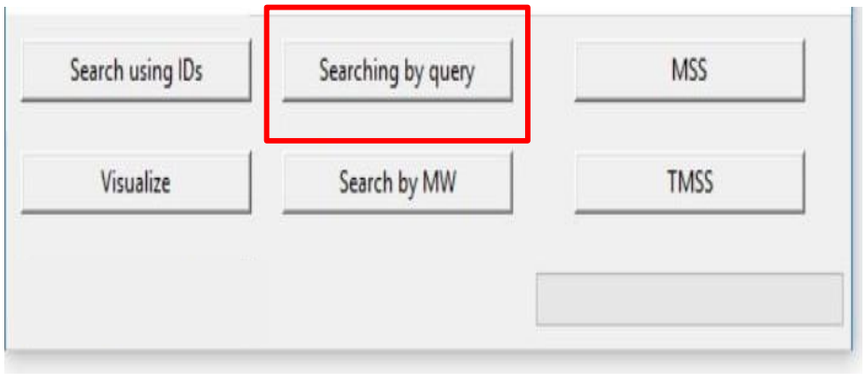
Input:

The input is Keyword or query to search YMDB with.

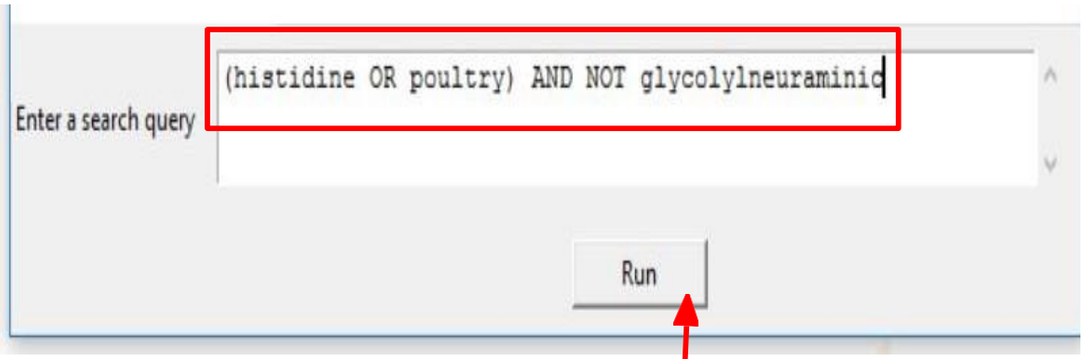
Output:

The output is one CSV file name "txtsearch", contains metabolites general information, retrieved and parsed from the search using the user query.

1- Select "Searching by query"



2- Enter a keyword to search with.



3- Press "Run" and wait until the loading bar finish

3- Searching by Molecular weight (MW).

Description:

This function is used to retrieve and parse metabolites general information from YMDB using molecular weight.

Input:

The inputs are a start MW and an end MW to search within.

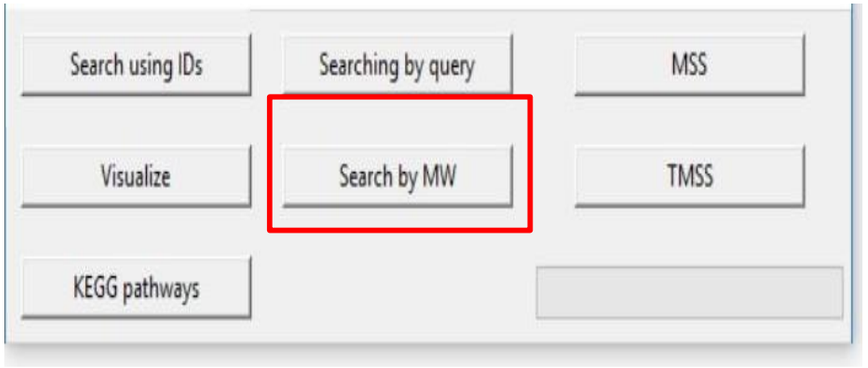
Output:

The output is one CSV file name "ChemQuery", contains metabolites general information for the result from the search by molecular weight.

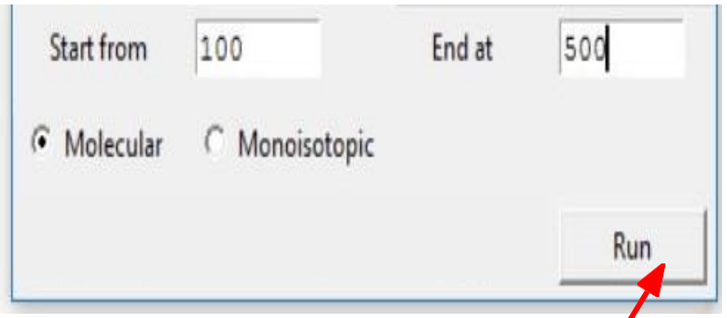
Note:

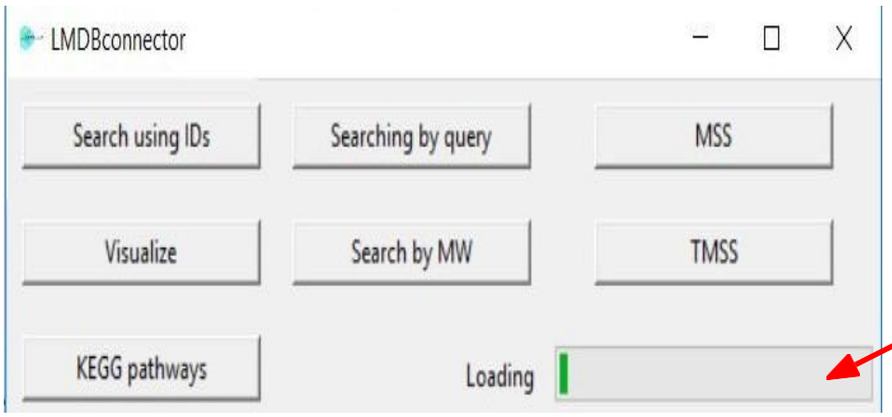
The user can select Molecular weight / Average mass or Monoisotopic mass to search with.

1- Select "Searching by MW"



2- Enter a start mass and the end mass you want to search with, then select the mass to unite you want to use.





3- Press "Run" and wait until the loading bar finish

4- Mass Spectrum Search (MSS)

Description:

This function is used to retrieve and parse metabolites general information from YMDB using mass spectrum search (MSS).

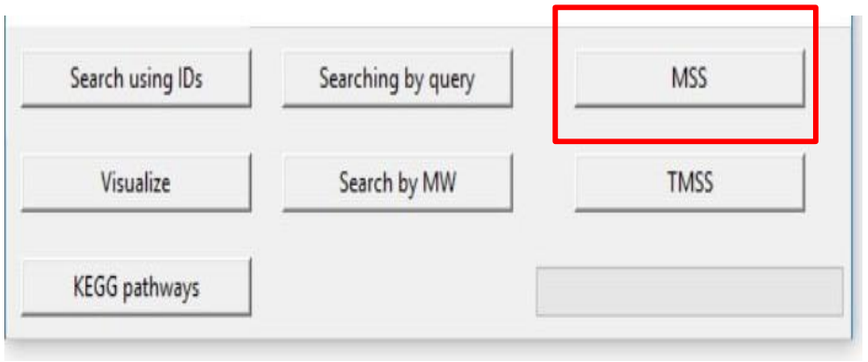
Input:

The inputs are masses to search with, ion mode, adducts, and Molecular Weight Tolerance also the unit used for it.

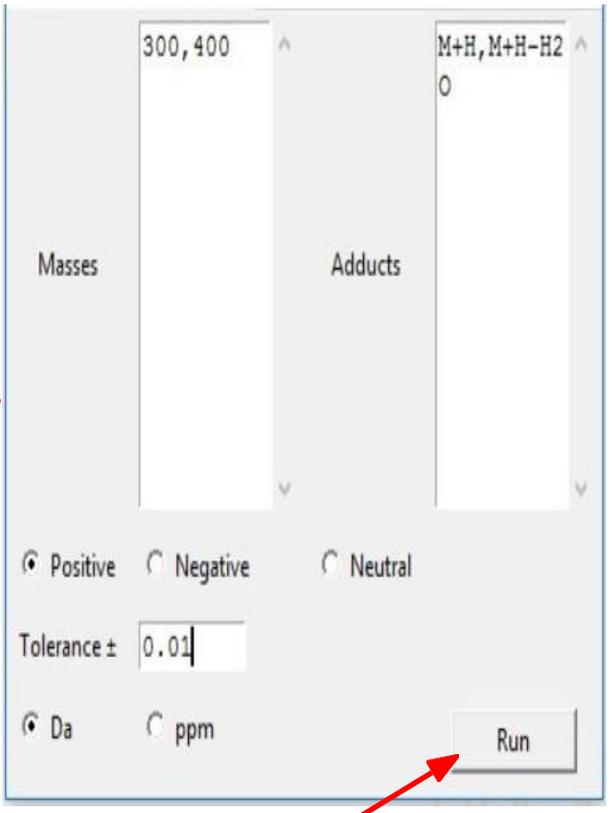
Output:

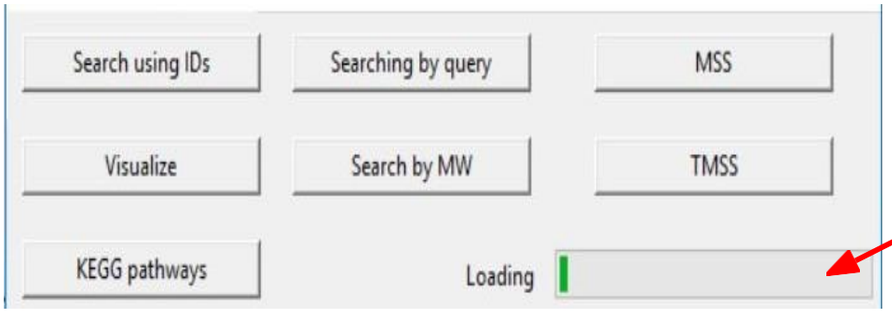
The output is one CSV file name "LC-MS", contains metabolites general information for the result from the search.

1- Select "MSS"



2- Enter masse(s) separated by a comma, enter adduct(s) separated by a comma. Then select an Ion mode. Finally, Enter a tolerance and select the unit you want to use.





3- Press "Run" and wait until the loading bar finish

5- Tandom Mass Spectrum Search (TMSS)

Description:

This function is used to retrieve and parse metabolites general information from YMDB using tandom mass spectrum search (MSS).

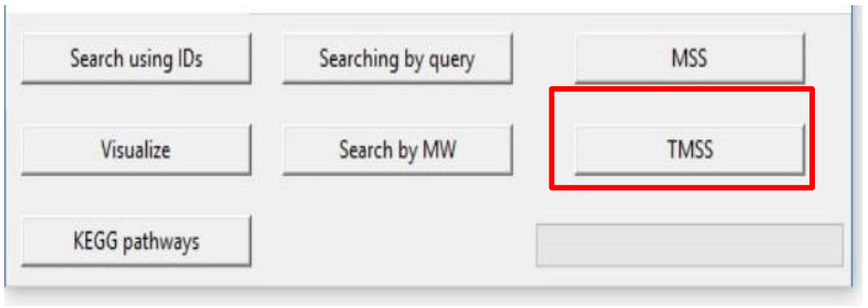
Input:

The inputs are MS/MS peak list, parent ion mass, parent ion mass tolerance ±, mass/charge (m/z) tolerance ±, ionization mode, and CID energy.

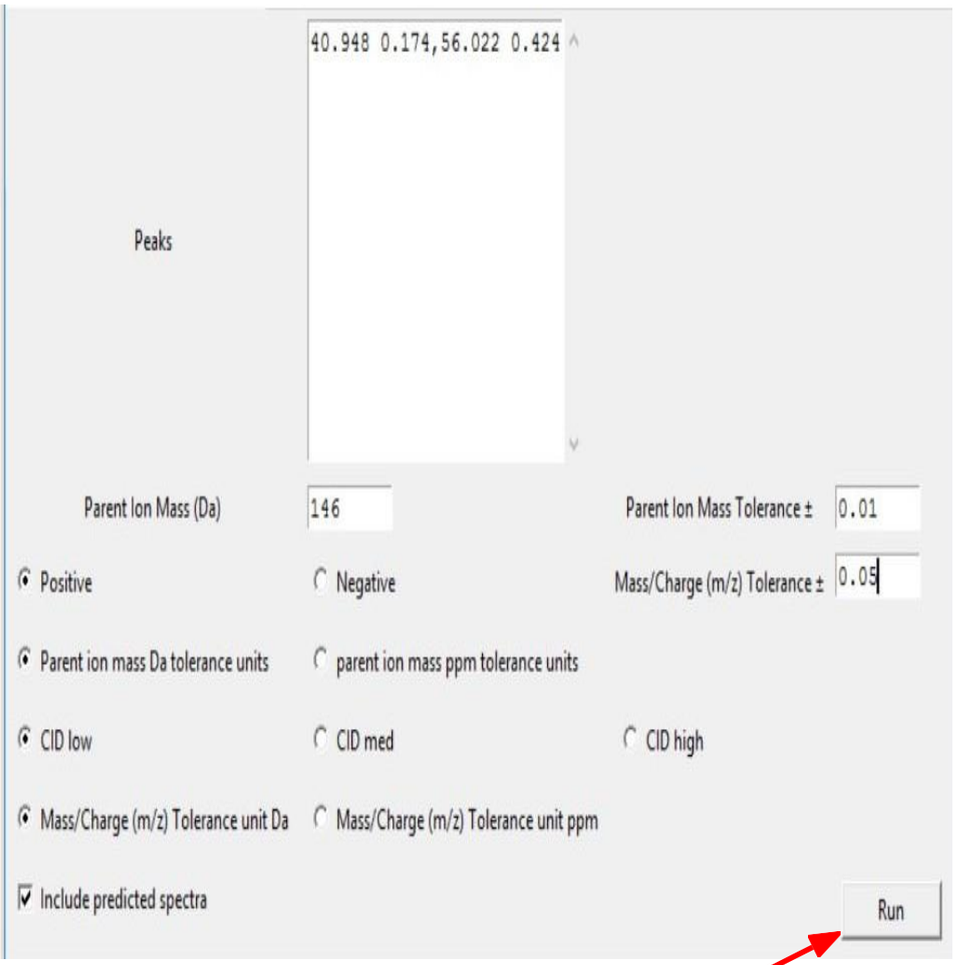
Output:

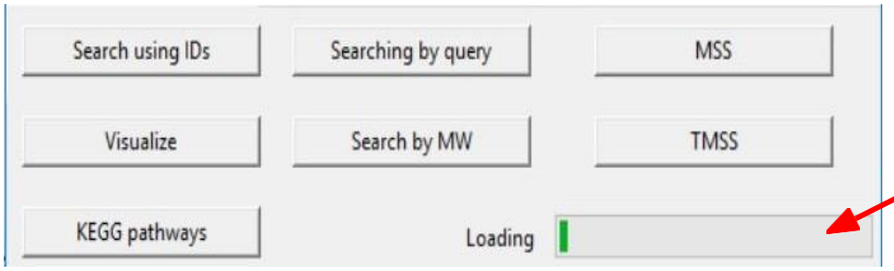
The output is one CSV file name "LC-MS/MS", contains metabolites general information for the result from the search.

1- Select "TMSS"



2- First, Enter masses and m/z for each with same format here separated by a comma. Enter the parent mass and its ion mass tolerance, also enter the m/z tolerance. Then, select the ion mode, the parent tolerance mass unit (Da or ppm), also select the m/z tolerance unit (Da or ppm). Finally, select the level for CID and if to include predicted spectra or not.





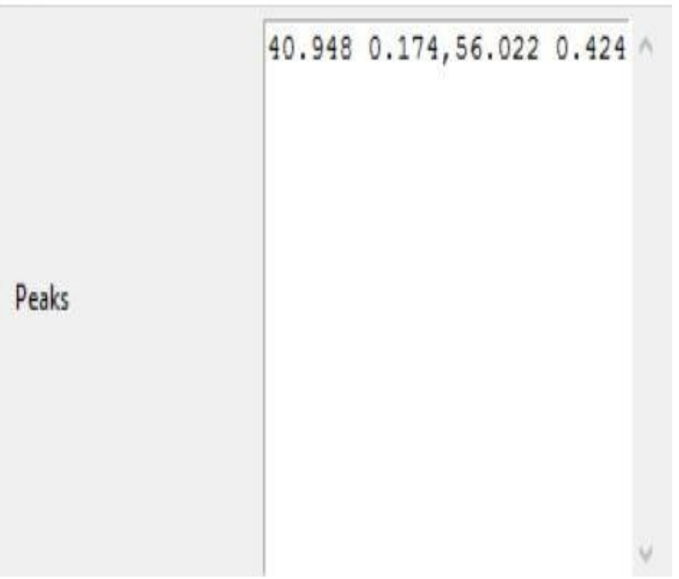
3- Press "Run" and wait until the loading bar finish

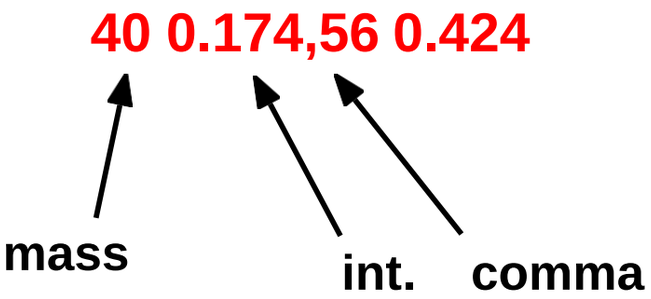
Note: You must enter the masses and the m/z in the correct format.

The correct format is:

mass[space]int.,mass[space]int., mass[space]int.,.....

example:





Visualization

For YMDB only one plot named “Cellular Locations”, used to show the cellular locations of the metabolites.

Cellular Locations example plot:
https://github.com/Proteomicslab57357/Xconnector/tree/master/examples/YMDB

4- The *Toxin and Toxin Target Database (T3DB)*

1- Searching Using IDs.

Description:

This function is used to retrieve and parse different information from T3DB using IDs.

Input:

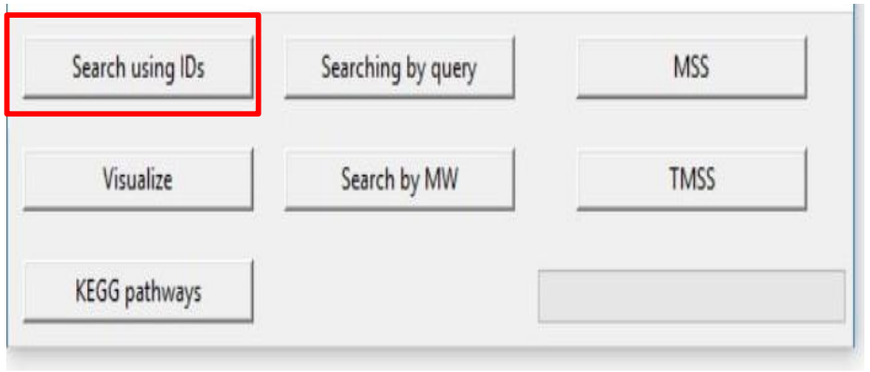
The input is CSV file with column name (IDs_YMDB) and T3DB IDs.

Output:

The outputs are 3 CSV files with different information:

1- General information, 2- Experimental Properties, 3- Predicted Properties.

1- Select "Searching Using IDs"



2- Choose CSV contains T3DB ids with a column named "IDs_T3DB"

3- Wait until the loading bar finish

An example for the CSV input file

A
IDs_T3DB
T3D0001
T3D0002
T3D0003
T3D0004
T3D0006
T3D0007
T3D0009
T3D0010
T3D0011
T3D0012
T3D0013
T3D0014
T3D0015
T3D0016
T3D0017
T3D0018
T3D0019
T3D0020
T3D0021
T3D0022
T3D0024

Note: the header must be named IDs_T3DB

Sample file:
https://github.com/Proteomicslab57357/Xconnector/blob/master/examples/T3DB_id.csv

2- Searching by query.

Description:

This function is used to retrieve and parse metabolites general information from T3DB using keyword (query).

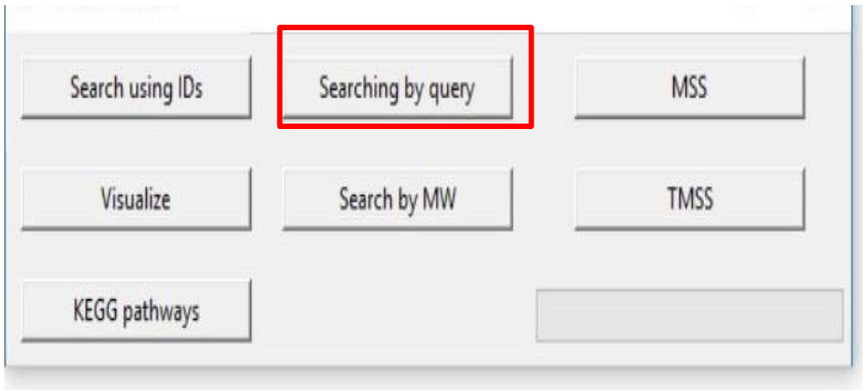
Input:

The input is Keyword or query to search T3DB with.

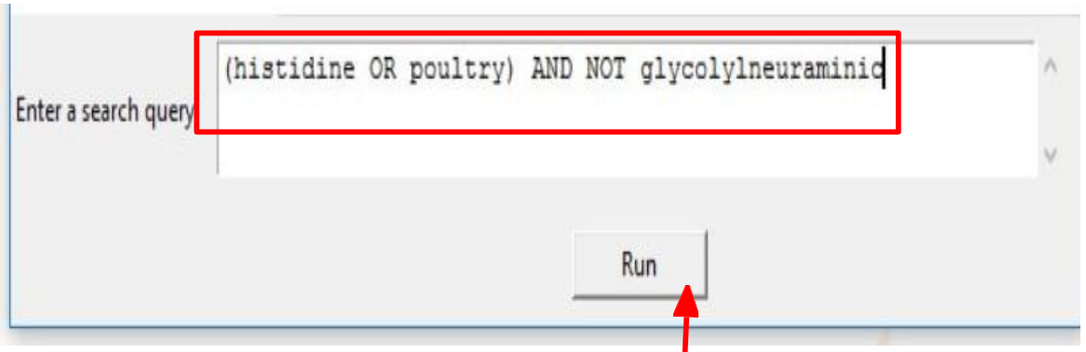
Output:

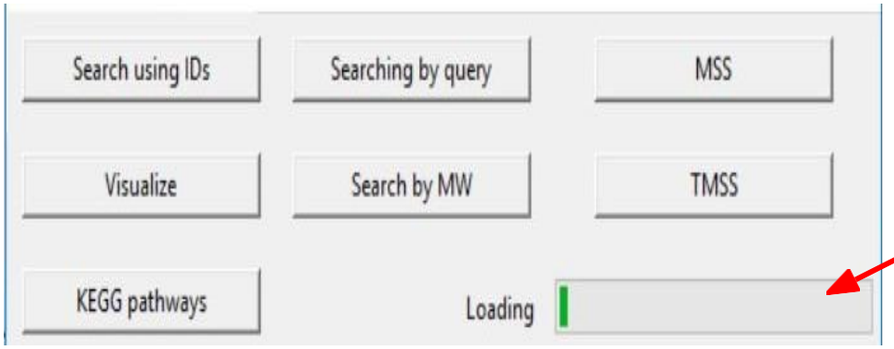
The output is one CSV file name "txtsearch", contains metabolites general information, retrieved and parsed from the search using the user query.

1- Select "Searching by query"



2- Enter a keyword to search with.





3- Press "Run" and wait until the loading bar finish

3- Searching by Molecular weight (MW).

Description:

This function is used to retrieve and parse metabolites general information from T3DB using molecular weight.

Input:

The inputs are a start MW and an end MW to search within.

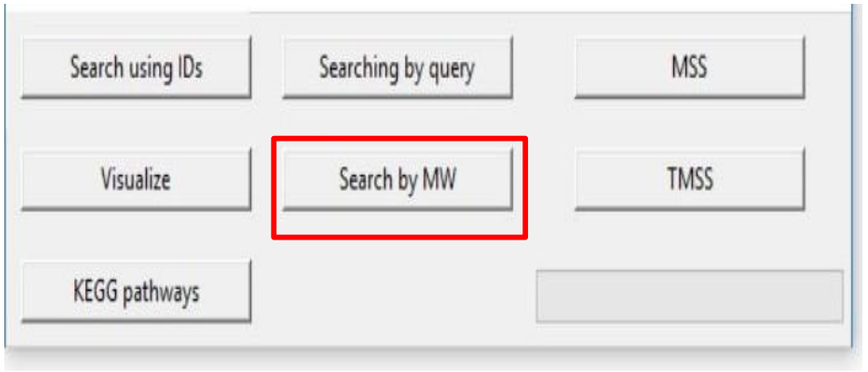
Output:

The output is one CSV file name "ChemQuery", contains metabolites general information for the result from the search by molecular weight.

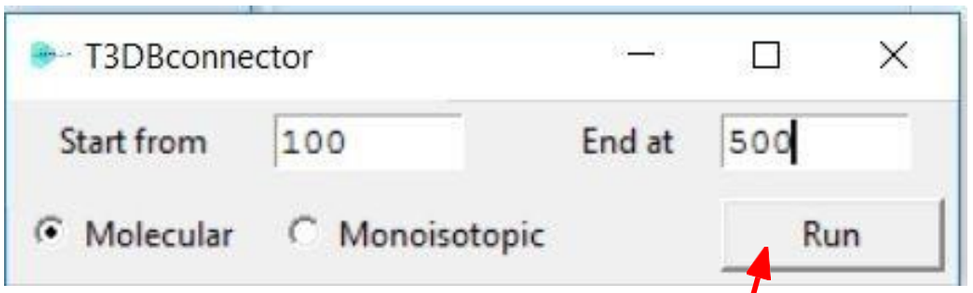
Note:

The user can select Molecular weight / Average mass or Monoisotopic mass to search with.

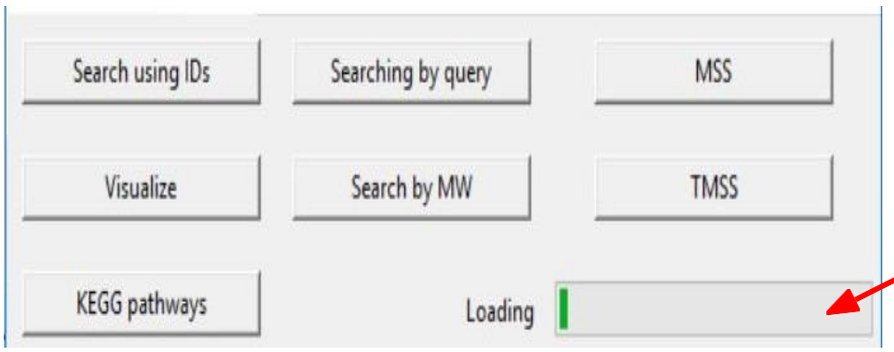
1- Select "Searching by MW"



2- Enter a start mass and the end mass you want to search with, then select the mass to unite you want to use.



3- Press "Run" and wait until the loading bar finish



4- Mass Spectrum Search (MSS)

Description:

This function is used to retrieve and parse metabolites general information from T3DB using mass spectrum search (MSS).

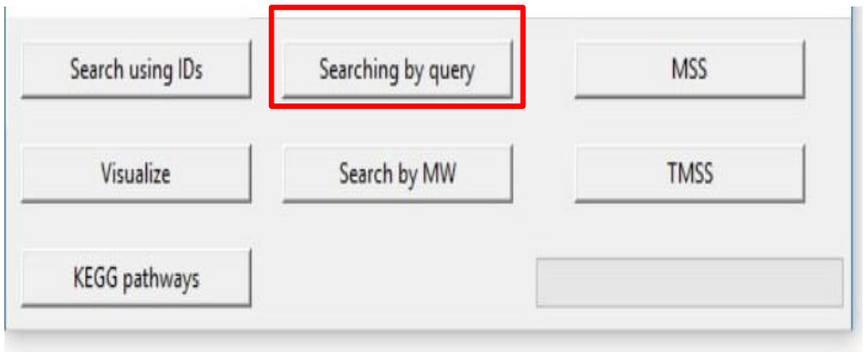
Input:

The inputs are masses to search with, ion mode, adducts, and Molecular Weight Tolerance also the unit used for it.

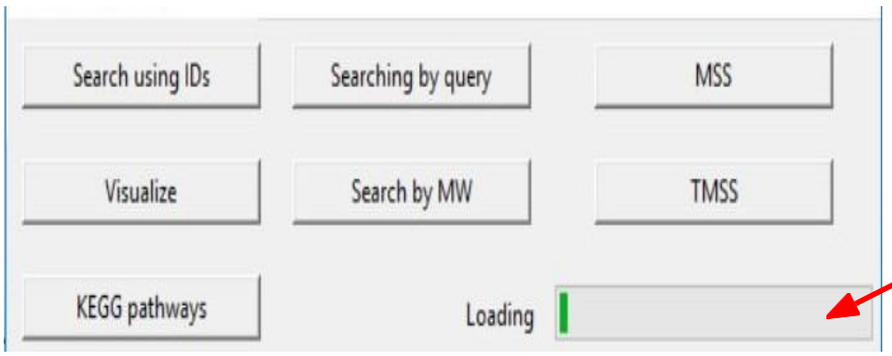
Output:

The output is one CSV file name "LC-MS", contains metabolites general information for the result from the search.

1- Select "MSS"



2- Enter masse(s) separated by a comma, enter adduct(s) separated by a comma. Then select an Ion mode. Finally, Enter a tolerance and select the unit you want to use.



3- Press "Run" and wait until the loading bar finish

5- Tandom Mass Spectrum Search (TMSS)

Description:

This function is used to retrieve and parse metabolites general information from T3DB using tandom mass spectrum search (MSS).

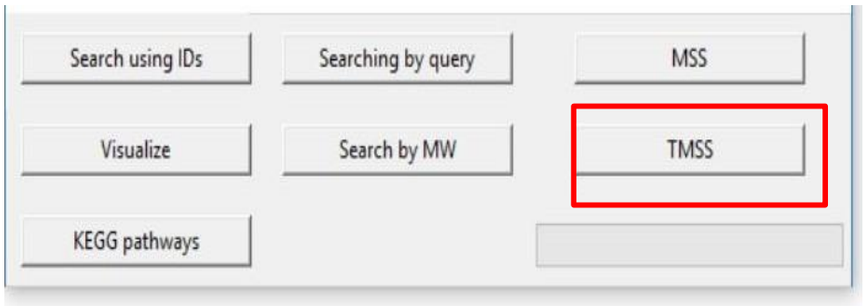
Input:

The inputs are MS/MS peak list, parent ion mass, parent ion mass tolerance ±, mass/charge (m/z) tolerance ±, ionization mode, and CID energy.

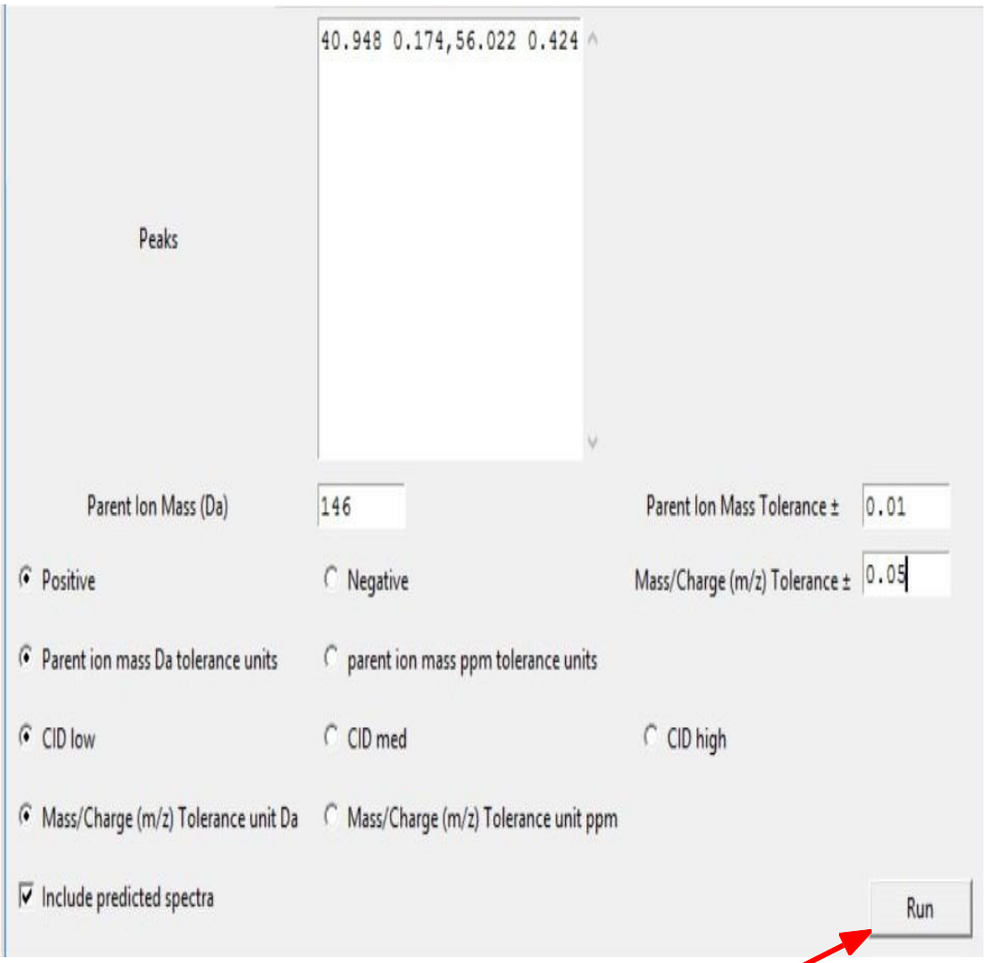
Output:

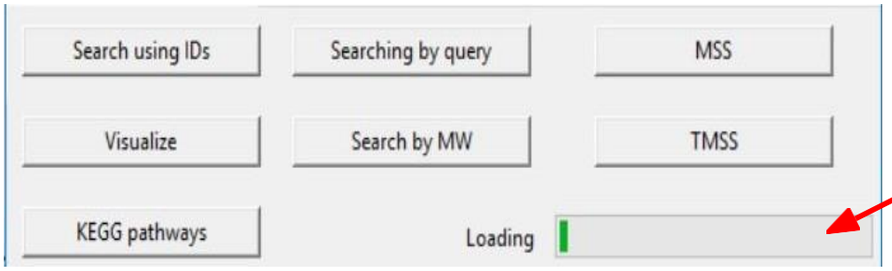
The output is one CSV file name "LC-MS/MS", contains metabolites general information for the result from the search.

1- Select "TMSS"



2- First, Enter masses and m/z for each with same format here separated by a comma. Enter the parent mass and its ion mass tolerance, also enter the m/z tolerance. Then, select the ion mode, the parent tolerance mass unit (Da or ppm), also select the m/z tolerance unit (Da or ppm). Finally, select the level for CID and if to include predicted spectra or not.





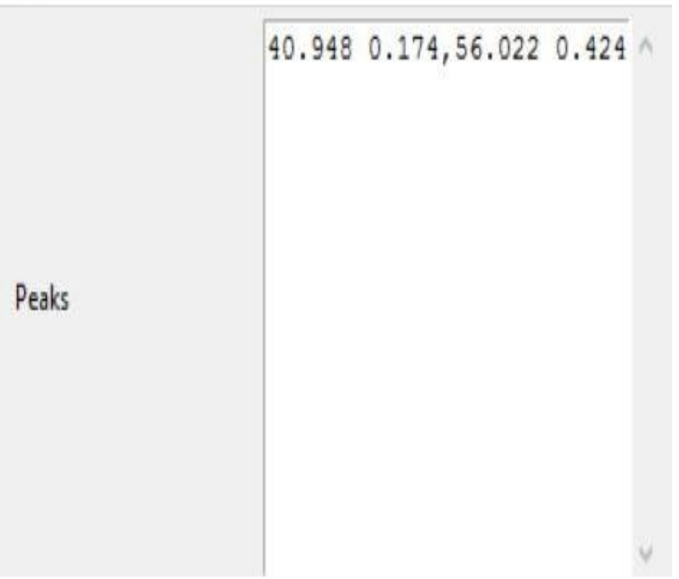
3- Press "Run" and wait until the loading bar finish

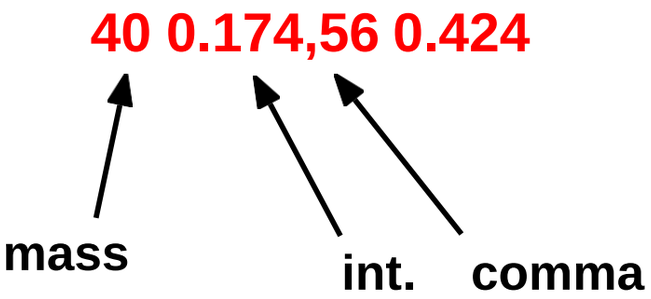
Note: You must enter the masses and the m/z in the correct format.

The correct format is:

mass[space]int.,mass[space]int., mass[space]int.,.....

example:





6- Get KEGG pathway

Description:

This function is used to retrieve and parse pathways information from KEGG. Using metabolites IDs from YMDB.

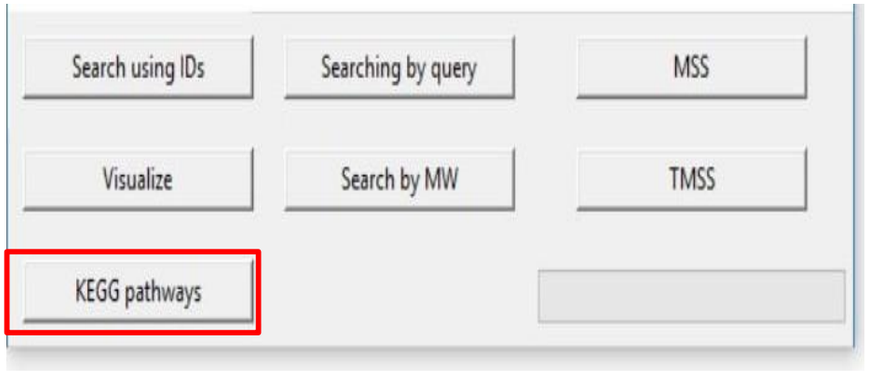
Input:

The input is CSV file with column name (IDs_T3DB) and T3DB IDs.

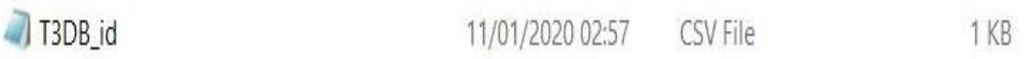
Output:

The main output is a CSV file named "Kegg_pathway" containing each metabolite from the input ids and the pathways this metabolite involve in.

1- Select "KEGG pathway"



2- Choose CSV contains T3DB ids with a column named "IDs_T3DB"



3- Wait until the loading bar finish

Visualization

For T3DB two plots named “Cellular Locations” and "Tissues Locations", used to show the cellular locations and tissues locations of the metabolites, respectively.

Cellular Locations

Tissues Locations

Examples:
https://github.com/Proteomicslab57357/Xconnector/tree/master/examples/T3DB

5- ReSpect for Phytochemicals (ResDB)

1- Searching Using IDs.

Description:

This function is used to retrieve and parse different information from ResDB using IDs.

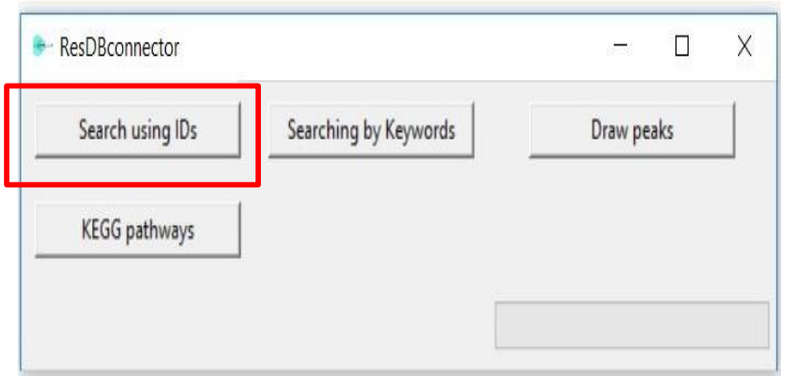
Input:

The input is CSV file with column name (IDs_respectDB) and ResDB IDs.

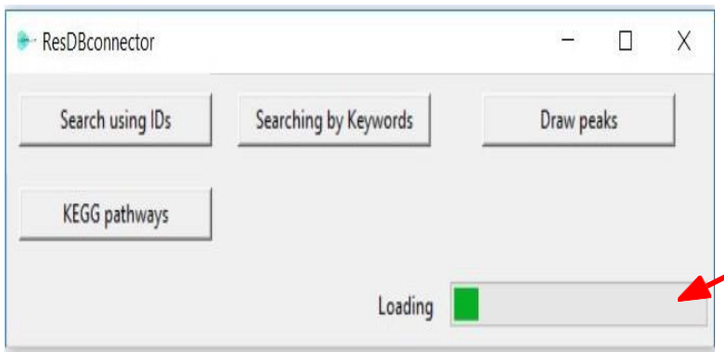
Output:

The outputs is CSV files contains general information about each metabolite.

1- Select "Searching Using IDs"



2- Choose CSV contains ReSpect ids with a column named "IDs_respectDB"



3- Wait until the loading bar finish

An example for the CSV input file

Note: the header must be named IDs_respectDB

Sample file:
https://github.com/Proteomicslab57357/Xconnector/blob/master/examples/res_id.csv

2- Searching by Keywords.

Description:

This function is used to retrieve and parse metabolites general information from ResDB using keyword (query).

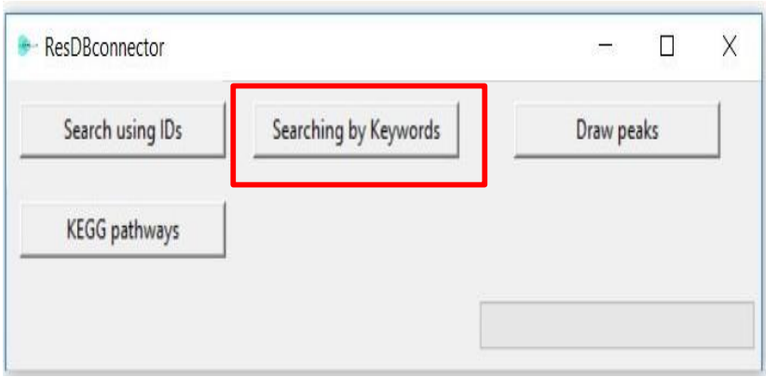
Input:

The input is Keyword or query to search ResDB with.

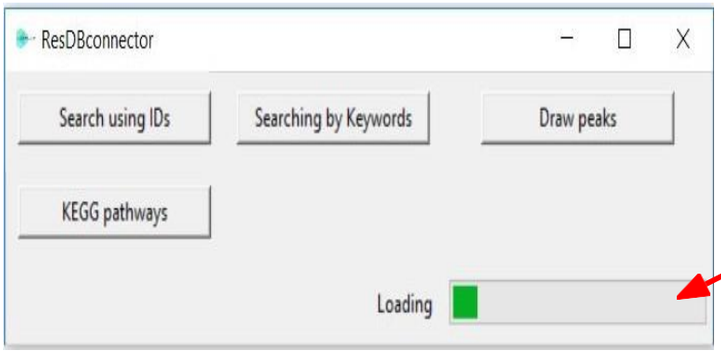
Output:

The output is one CSV file name "Keyword", contains metabolites general information, retrieved and parsed from the search using the user query.

1- Select "Searching by Keywords"



2- The search can be by all of this criteria or just one of them.



3- Press "Search" and wait until the loading bar finish

3- Get KEGG pathway

Description:

This function is used to retrieve and parse pathways information from KEGG. Using metabolites IDs from YMDB.

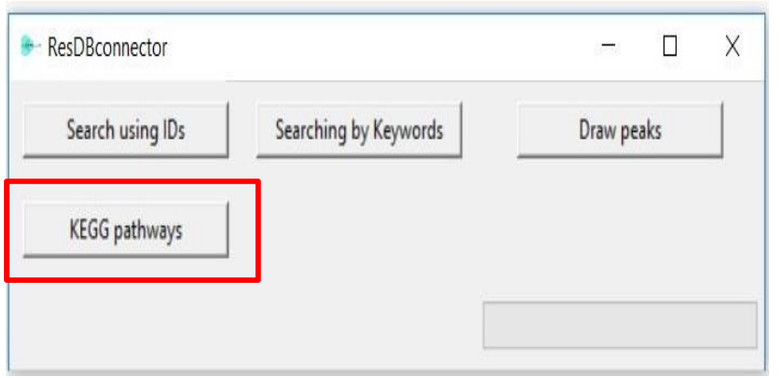
Input:

The input is CSV file with column name (IDs_YMDB) and YMDB IDs.

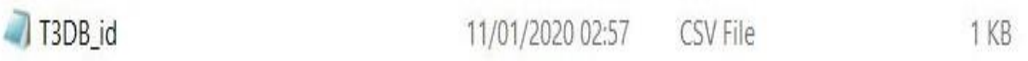
Output:

The main output is a CSV file named "Kegg_pathway" containing each metabolite from the input ids and the pathways this metabolite involve in.

1- Select "KEGG pathway"



2- Choose CSV contains T3DB ids with a column named "IDs_T3DB"



3- Wait until the loading bar finish

Visualization

In ReSpect for Phytochemicals database, the visualization function in Xconnector could be used to generates spectral peaks plot, using only the metabolites IDs.

Example plot:
https://github.com/Proteomicslab57357/Xconnector/tree/master/examples/ResDB

6- The Blood Exposome Database.

1- Searching.

Description:

Currently, this database has only one function. The user can search for different compounds available in the Blood Exposome DB (https://bloodexposome.org/#/dashboard). The user can use

Input:

The input is CSV file with any column name and HMDB IDs, CID (pubchem ID) , KEGG IDs, chemical formula, SMILES, or InChIKey, as row. Also a mixed row with all of these input can be used (example).

Output:

The outputs is CSV files contains general information about each metabolite.

	A
1	Anyname
2	HMDB0031106
3	HMDB0144295
4	HMDB0144295
5	Sphingosine 1-phosphate
6	N-heptadecanoyl-D-erythro-sphingosine
7	Triheptadecanoin
8	441
9	6287
10	6287
11	C4H6O5

7- The Phenol-Explorer Database.

1- Searching.

Description:

Currently, this database has only one function. The user can search for different compounds available in the Phenol-Explorer Database (https://bloodexposome.org/#/dashboard). The user can use

Input:

The input is CSV file with any column name andCID (pubchem ID), chemical formula, or name, as row

(example).

Output:

The outputs is CSV files contains general information about each metabolite.

441
6287
6287
C4H6O5

Xconnector

- The Idea behind the Xconnector function (not as software) is to connect the possible databases together as one database. Hence, (for now) there are 4 databases connected together in the Xconnector function (HMDB, LMDB, YMDB, and T3DB).

- There are three functions avilable: Searching using IDs, Searching using the metabolic name and Visualization.

1- Searching Using IDs.

Description:

This function is used to retrieve and parse different information from HMDB, LMDB, YMDB, and T3DB using IDs.

Input:

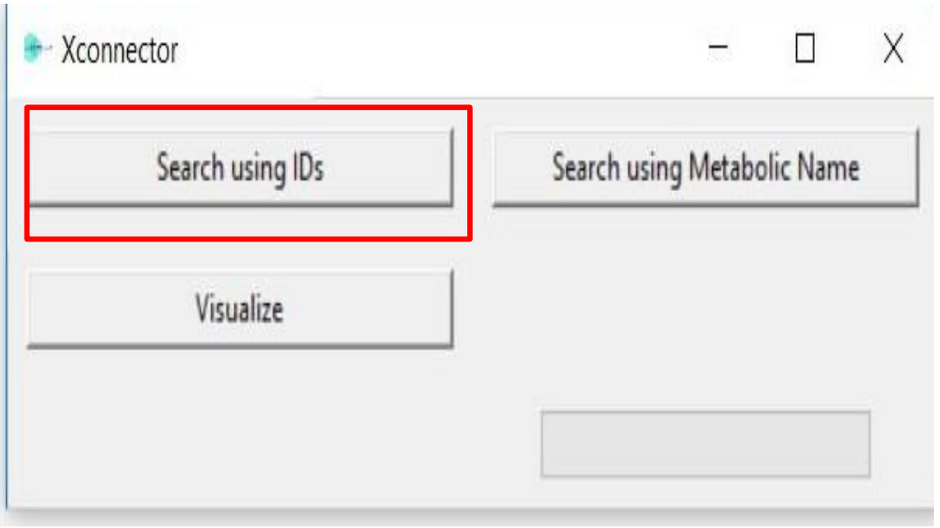
The input is CSV file with column name (IDs) and any ID from thur 4 databases IDs.

Output:

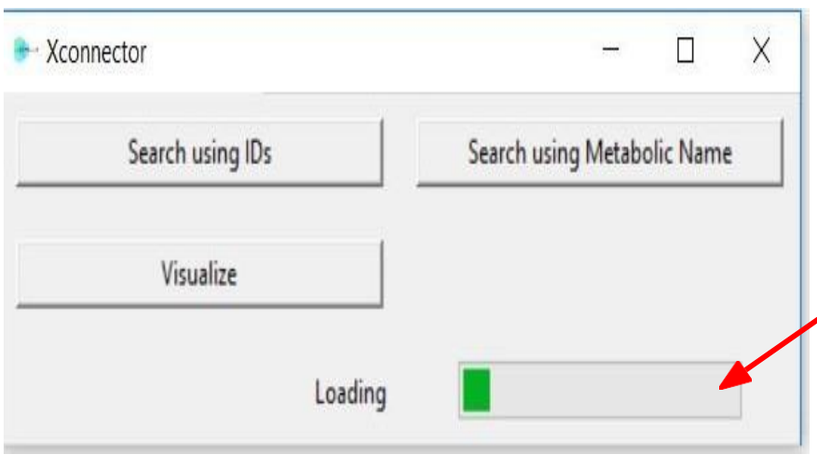
The outputs is CSV files contains general information about each metabolite.

Sample file:
https://github.com/Proteomicslab57357/Xconnector/blob/master/examples/all_ids.csv

1- Select "Searching Using IDs"



2- Choose CSV contains ReSpect ids with a column named "IDs_respectDB"



3- Wait until the loading bar finish

2- Searching Using the metabolic name.

Description:

This function is used to retrieve and parse different information from HMDB, LMDB, YMDB, and T3DB using metabolites name as query search.

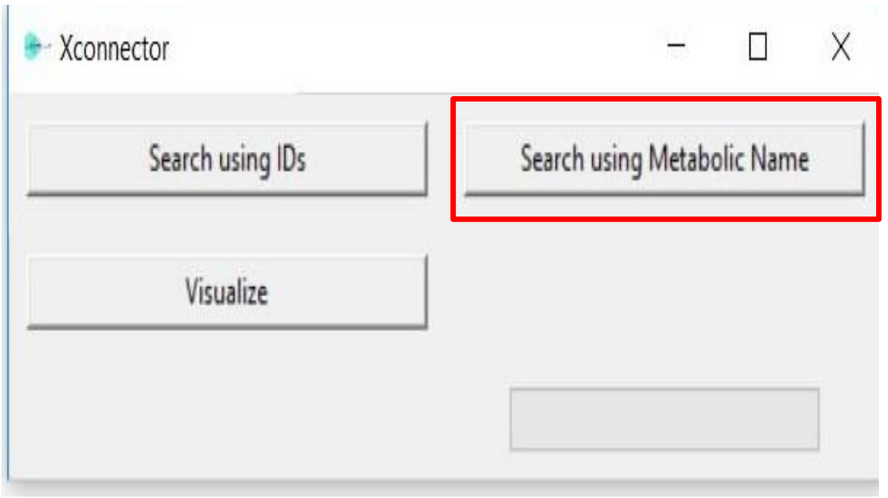
Input:

The input is Keyword of metabolite name .

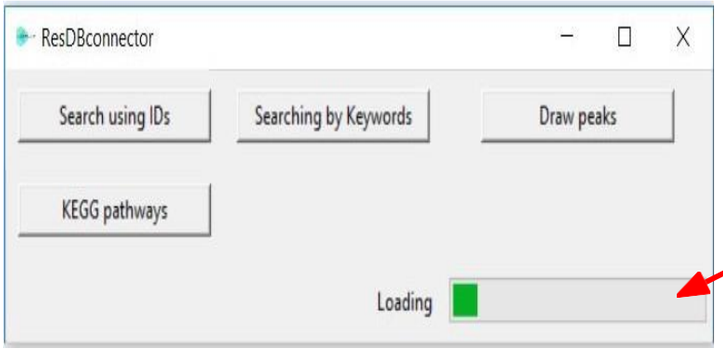
Output:

The outputs is CSV files contains general information about each metabolite.

1- Select "Searching by Keywords"



2- Enter a compound name



3- Press "Search" and wait until the loading bar finish

Visualization

For Xconnector function, it can visualize four plots, for the Predicted Physical Properties of the metabolites. Using HMDB, LMDB, YMDB and T3MD.

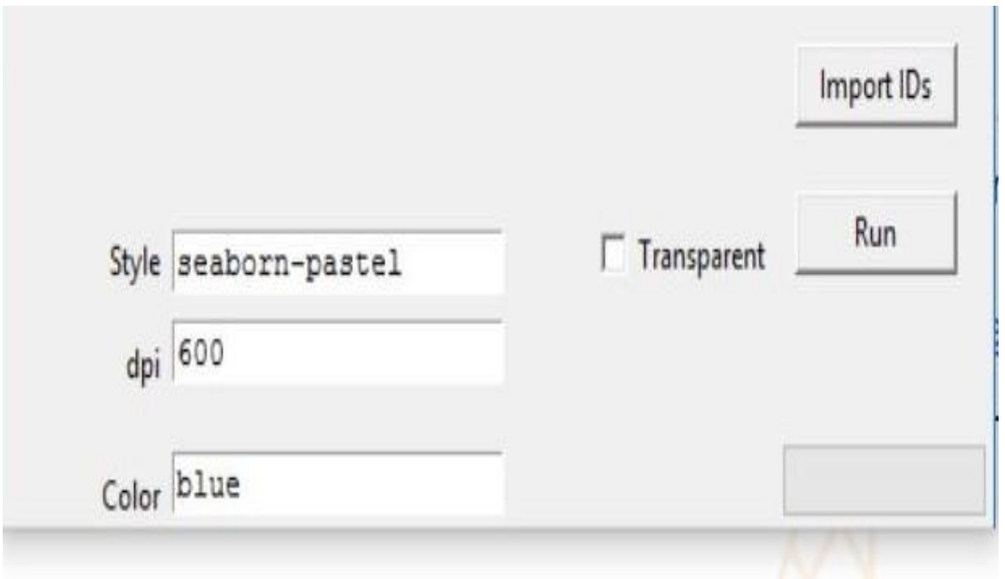
- Water Solubility
- pKa (Strongest Acidic)
- pKa (Strongest Basic)
- pKa (Strongest Acidic) Vs. pKa (Strongest Basic)

Examples:

https://github.com/Proteomicslab57357/Xconnector/tree/master/examples/Xconnector

Important Note

Xconnector is flexible regarding the generated plots. A user could choose a style, dpi and the colour of the plot. The style of the plots could be selected from style sheets reference available in the matplotlib python package "https://matplotlib.org/gallery/style_sheets/style_sheets_reference.html#style-sheets-reference". The default dpi is 600, and the colour could be any colour by its name or its HTML colour codes. As well as, the transparent option could be selected to generate plots with no background (transparent)



Links

GitHub: https://github.com/Proteomicslab57357/Xconnector

Our Website: https://www.57357.org/en/department/proteomics-unit-dept/about-department/

For more help: proteomics.lab@57357.org

HMDB: http://www.hmdb.ca/

LMDB: http://lmdb.ca/

YMDB: http://www.ymdb.ca/

T3DB: http://www.t3db.ca/

ReSpect DB: http://spectra.psc.riken.jp/

License

Proteomicslab57357/Xconnector is licensed under the

GNU General Public License v3.0

Permissions of this strong copyleft license are conditioned on making available complete source code of licensed works and modifications, which include larger works using a licensed work, under the same license. Copyright and license notices must be preserved. Contributors provide an express grant of patent rights.

Permissions	Limitations	Conditions
✓ Commercial use	✗ Liability	ⓘ License and copyright notice
✓ Modification	✗ Warranty	ⓘ State changes
✓ Distribution		ⓘ Disclose source
✓ Patent use		ⓘ Same license
✓ Private use		

Complete License Form

https://github.com/Proteomicslab57357/Xconnector/blob/master/LICENSE