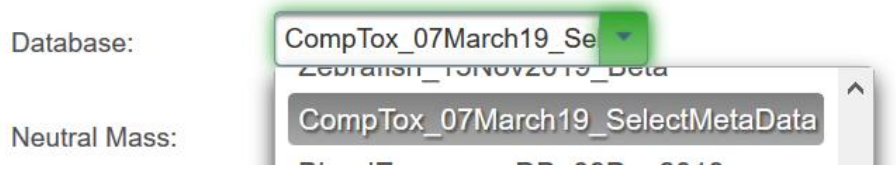
Example 1: Nicotine

Practicing MetFrag (https://msbi.ipb-halle.de/MetFrag/) with a formula search using the CompTox Chemistry Dashboard Local CSV from the dropdown menu:



Using the Mass Spectrum of Nicotine from MassBank:

https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300801&dsn=Eawag

Peaks:

```
80.0494 6261028.7 23
84.0807 13197924.1 50
94.065 967625.9 3
106.065 24640249.3 93
117.0572 3192413.5 12
120.0807 8648923.7 32
130.0651 24669353.9 93
132.0807 80112590.7 304
163.1229 263120223.6 999
```

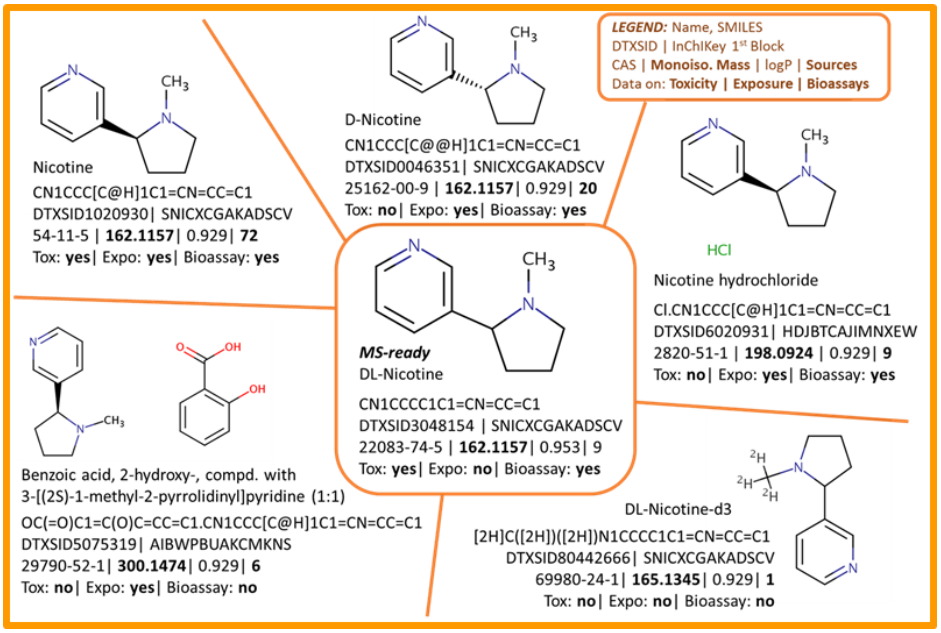
Formula: $C_{10}H_{14}N_2$

Mode: [M+H]+

Accuracy settings: 5 ppm + 0.001 Da

MS-ready form: try grouped and ungrouped (fragmentation settings)

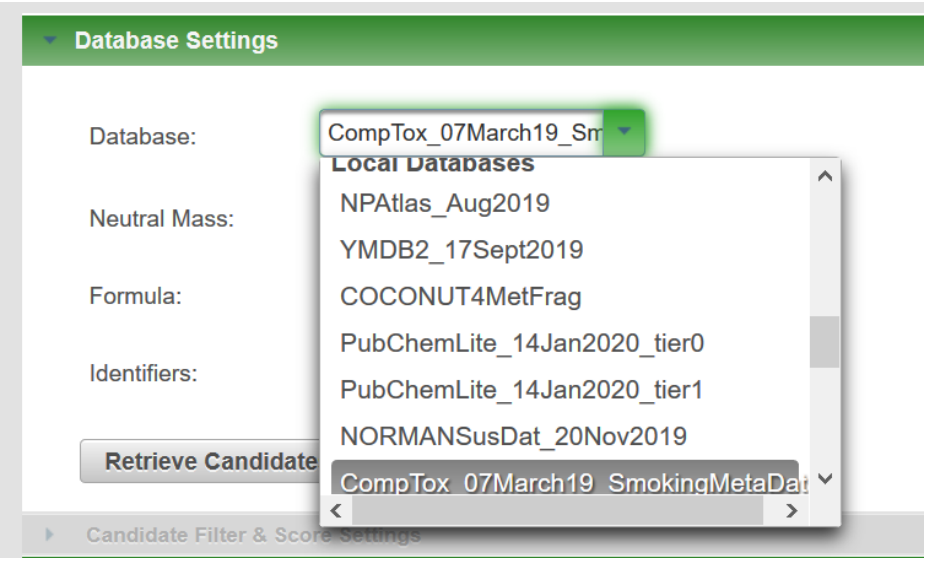
Try with various CompTox Metadata (MetFrag scoring terms) to see the influence of different metadata terms. How confident are you in the top ranked candidate?



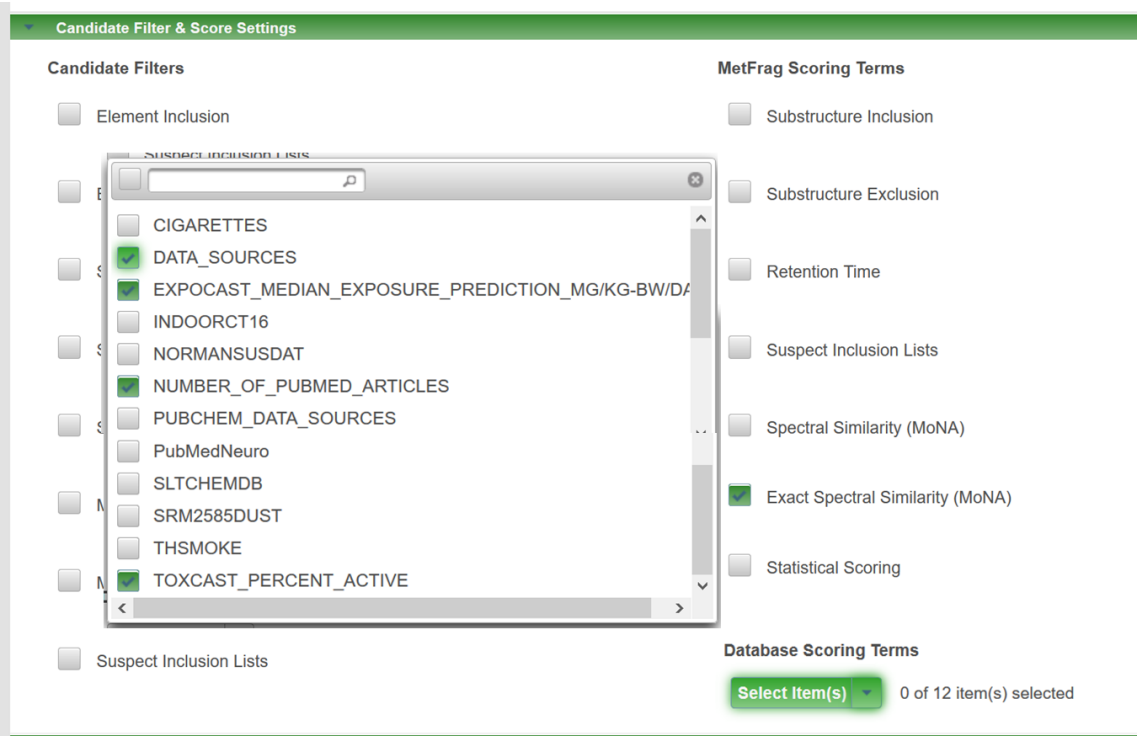
 DOI: 10.5281/zenodo.3953683

Example 1a:

Investigate the same example using the CompTox "smoking" file to add some exposomics context.



The additional metadata terms indicate presence in cigarettes (CIGARETTES), thirdhand smoke (THSMOKE), indoor environments (INDOORCT16, SRM2585DUST) and potential neurotoxic endpoints (PubMedNeuro, literature mined data, see next page). More information and cross-links to the lists are available at DOI: 10.5281/zenodo.3364463



The PubMedNeuro (LITMINEDNEURO) list comes from an excel macro developed by Nancy Baker, the whole process is documented in Schymanski *et al.* 2019 DOI: 10.1039/C9EM00068B. The PubMedNeuro list is an extraction of the PMID Count column shown below; this excel file can also be used to create specific lists relevant to only certain endpoints (see Example 3).

Chemical	CAS RN	DSSToxID	PMID Ct	Seizures	Nervous System Diseases	Peripheral Nervous System Diseases	Brain Diseases	Muscular Diseases	Basal Ganglia Diseases	Parkinson Disease, Secondary	Coma	Hallucinations	Tremor	Memory Disorders	Central Nervous
Cisplatin	15663-27-1	DTXSID4024983	1032	20	47	140	13	0	4	1	1	0	1	2	4
Ethanol	64-17-5	DTXSID9020584	768	100	23	11	18	26	1	3	20	6	17	54	2
Lead	7439-92-1	DTXSID2024161	740	28	107	68	102	4	2	2	1	3	4	19	30
Lithium	7439-93-2	DTXSID5036761	689	30	50	9	22	5	36	13	25	6	93	12	15
Valproic Acid	76584-70-8	DTXSID70227388	666	32	10	3	65	6	10	18	45	5	18	4	2
1-Methyl-4-phe	28289-54-5	DTXSID8040933	638	1	24	0	11	0	6	289	0	0	5	0	1
Vincristine	2068-78-2	DTXSID8044331	567	17	59	125	15	5	1	1	5	3	2	1	8
Phenytoin	57-41-0	DTXSID8020541	560	37	24	25	16	9	3	1	9	3	8	4	6
Haloperidol	52-86-8	DTXSID4034150	555	6	6	1	10	6	153	51	4	4	11	1	0
Cocaine	50-36-2	DTXSID2038443	530	151	16	0	8	0	2	3	3	8	6	12	11
Aspirin	50-78-2	DTXSID5020108	489	8	3	0	3	2	2	0	9	4	1	0	5
Paclitaxel	33069-62-4	DTXSID9023413	485	4	43	217	9	14	0	0	0	0	0	1	2
Aluminum	7429-90-5	DTXSID3040273	477	13	41	1	105	4	0	0	1	0	1	13	12
Lidocaine	6108-05-0	DTXSID80209953	464	150	26	15	3	2	0	0	8	4	6	2	10
Methotrexate	59-05-2	DTXSID4020822	451	17	25	1	79	4	0	1	5	0	1	9	18
Mercury	7439-97-6	DTXSID1024172	450	6	79	22	23	2	3	5	2	2	38	7	25

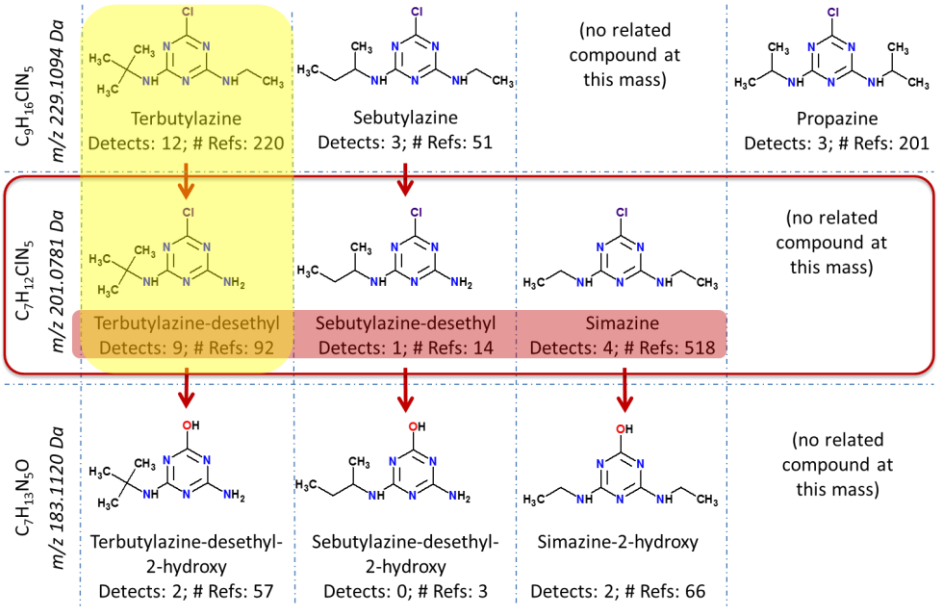
While familiarizing yourself with MetFrag, you can also try other Nicotine spectra in MassBank, e.g.

https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300804&dsn=Eawag

Before you move on … what level of confidence do you have in these identifications?

Example 2: Simazine and Desethylterbutylazine

This is the "tricky" example shown in the second part of the demo, where sample context and metadata play a role in candidate ranking. Using the **CompTox Local CSV** or **PubChem** or **PubChemLite**, start with desethylterbutylazine and take the details from the records and the slide set. Choose different metadata categories and explore the influence of weighting.



Desethylterbutylazine: https://massbank.eu/MassBank/RecordDisplay.jsp?id=EA067112&dsn=Eawag

```
57.0699 34824.5 11
61.9792 27260 8
68.0244 499632.7 162
79.0059 755457.7 245
104.0011 904831.9 293
110.0462 250173.3 81
128.0568 46548.7 15
146.023 3077763.2 999
```

Simazine: https://massbank.eu/MassBank/RecordDisplay.jsp?id=EA026205&dsn=Eawag

```
61.9791 64844.4 25
68.0243 942037.9 364
71.0604 574767 222
79.0058 90581.2 35
90.0106 62889.1 24
96.0557 1273147.5 493
104.0011 1577666.6 610
107.0373 28681.6 11
124.087 1936398.2 749
132.0324 2579841.5 999
138.0774 89990.1 34
146.0229 101986.5 39
166.1088 466684.3 180
174.0542 710100.7 274
200.0706 22069.8 8
202.0853 2368133.5 917
```

What is your level of confidence with the top ranked candidate? Is the best candidate always ranked first?

Example 3: Own CSV and Disease-specific Endpoints

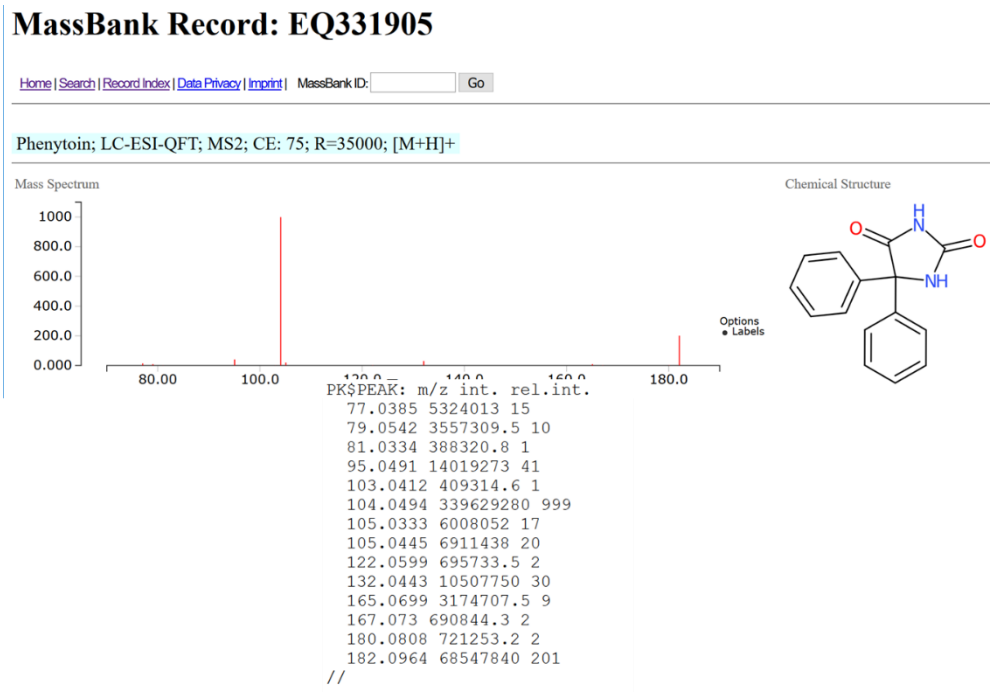
These are examples from Schymanski *et al* 2019 DOI: 10.1039/C9EM00068B, with selected formulas that have multiple candidates and multiple endpoints, with spectra in MassBank as "proof of concept". These serve to show how metadata may (or may not) help you select chemicals relevant for specific questions. The file for this example is a subset of the full excel macro (all disease endpoints would result in a huge dropdown menu) for direct upload into MetFrag. The endpoints were chosen based on those with the most data, plus keywords from participant description (from the 2019 Exposome Boot Camp). Note these are all neurological outcomes.

File: PubMedNeurotox_14Jul2019_MSready_DiseaseSubset.csv (upload via the "CSV" menu option)

Example spectra:

Phenytoin ($C_{15}H_{12}N_2O_2$):
https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ331905&dsn=Eawag

```
77.0385 5324013 15
79.0542 3557309.5 10
81.0334 388320.8 1
95.0491 14019273 41
103.0412 409314.6 1
104.0494 339629280 999
105.0333 6008052 17
105.0445 6911438 20
122.0599 695733.5 2
132.0443 10507750 30
165.0699 3174707.5 9
167.073 690844.3 2
180.0808 721253.2 2
182.0964 68547840 201
```



MassBank Record: EQ331905

Home | Search | Record Index | Data Privacy | Imprint | MassBank ID: Go

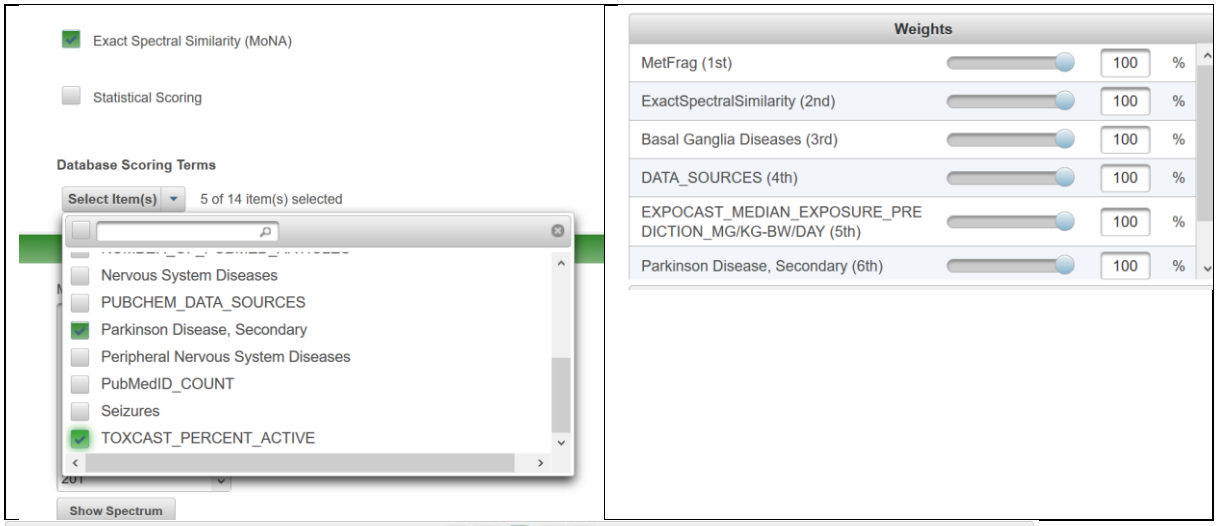
Phenytoin; LC-ESI-QFT; MS2; CE: 75; R=35000; [M+H]+

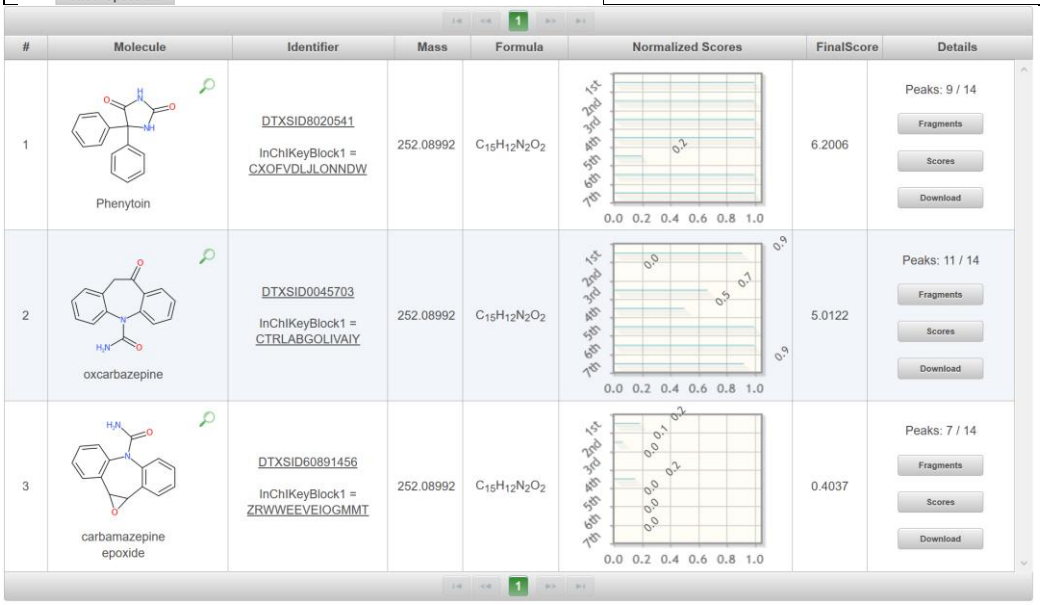
Mass Spectrum

Chemical Structure

```
PK$PEAK: m/z int. rel.int.
  77.0385 5324013 15
  79.0542 3557309.5 10
  81.0334 388320.8 1
  95.0491 14019273 41
  103.0412 409314.6 1
  104.0494 339629280 999
  105.0333 6008052 17
  105.0445 6911438 20
  122.0599 695733.5 2
  132.0443 10507750 30
  165.0699 3174707.5 9
  167.073 690844.3 2
  180.0808 721253.2 2
  182.0964 68547840 201
//
```

Select the metadata terms … some examples …





Scores View

Candidate Name: Phenytoin
Candidate Identifier: DTXSID8020541|5

	Name	Normalized Value	Raw Value
	MetFrag	1.0	92.9258
	ExactSpectralSimilarity	1.0	1.0
	Basal Ganglia Diseases	1.0	3.0
	DATA_SOURCES	1.0	105.0
	EXPOCAST_MEDIAN_EXPOSURE_PREDICTION_MG/KG-BW/DAY	0.2006	0.0
	Parkinson Disease, Secondary	1.0	1.0
	TOXCAST_PERCENT_ACTIVE	1.0	2.5605

Further examples: Another example is below, other formulas to try (formula search on MassBank for spectra) are C14H22N2O3, C10H13N5O4, C10H15NO, C11H9I3N2O4, C12H12N2O3. As shown on the next page, you can pick different spectra for different compounds with the same formula to see the effect of metadata vs experimental evidence …

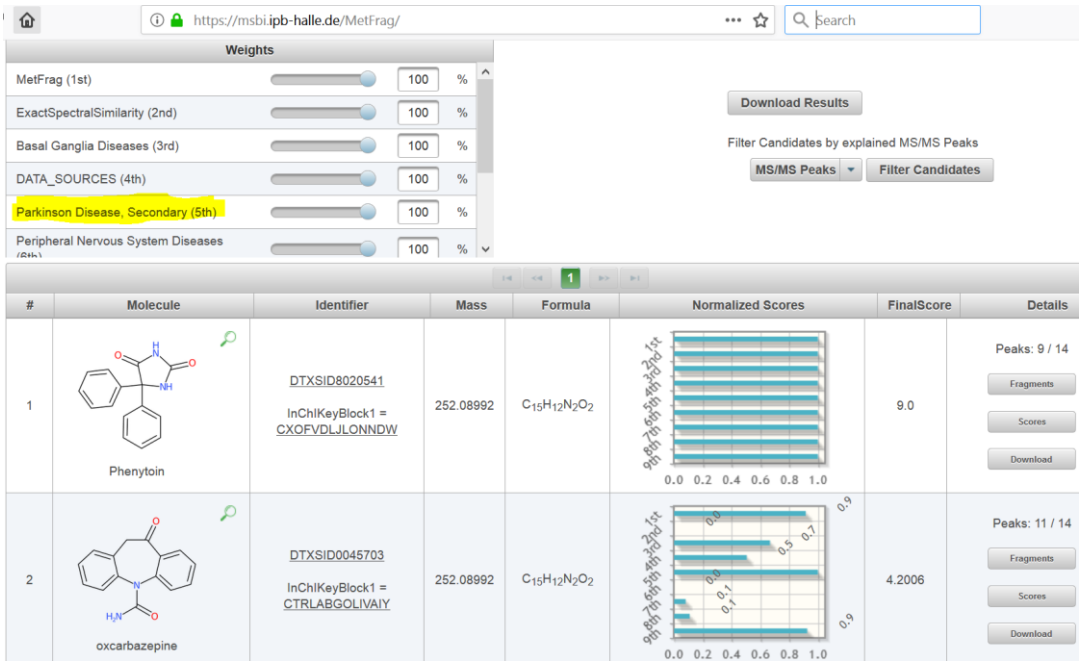
MDA ($C_{10}H_{13}NO_2$): https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ371504&dsn=Eawag

```
51.0228 956728.1 3
53.0384 727981.6 2
55.0178 6833328 25
65.0383 821205.1 3
77.0383 3254964 11
79.0541 27205822 100
91.0541 1550512 5
93.0334 2123720.2 7
95.049 2343883.5 8
103.0541 16062469 59
105.0697 271695296 999
107.0489 1063558.9 3
111.0438 806308.1 2
115.0541 2447483.5 8
121.0282 1584766.6 5
122.0361 5458170 20
123.0436 344165.6 1
131.0491 469800.7 1
133.0646 177603088 653
135.0438 193504560 711
145.0647 504020.8 1
148.0515 1192720.1 4
151.0752 2766502 10
163.0751 63095424 231
```

Out of Examples?

Search for your favourite chemicals on MassBank and try them! Try the disease-specific file or CompTox or PubChemLite first, not PubChem, to avoid large candidate numbers!

 DOI: 10.5281/zenodo.3953683



Search Parameters :
Formula: **C15H12N2O2**

Instrument Type: **CE-ESI-QTOF** , **ESI-ITFT** , **ESI-ITTOF**
ESI-QTOF , **ESI-TOF** , **LC-ESI-IT**
LC-ESI-ITFT , **LC-ESI-ITTOF** , **LC-ESI-Q**
LC-ESI-QFT , **LC-ESI-QIT** , **LC-ESI-QQ**
LC-ESI-QTOF , **LC-ESI-TOF** , **APCI-ITFT**
APCI-ITTOF , **LC-APCI-ITFT** , **LC-APCI-Q**
LC-APCI-QTOF , **LC-APPI-QQ**
MS Type: **MS2**
Ion Mode: **Both**

Edit / Resubmit Query

Results : 59 Hit. (1 - 59 Displayed)

Open All Tree

First Prev **1** Next Last (Total **1** Page)

Results End

Name	Formula / Structure	ExactMass	ID
2-Hydroxycarbamazepine 19 spectra	**C15H12N2O2**	**252.08990**	
Carbamazepine-10,11-epoxide 24 spectra	**C15H12N2O2**	**252.08990**	
Carbamazepine10,11-epoxide 4 spectra	**C15H12N2O2**	**252.08990**	
Carbamazepine10,11-epoxide 4 spectra	**C15H12N2O2**	**252.08990**	
Phenytoin 12 spectra	**C15H12N2O2**	**252.08987**	
LC-ESI-QFT; MS2; CE: 15; R=35000; [M+H]+			EQ331901
LC-ESI-QFT; MS2; CE: 15; R=35000; [M-H]-			EQ331951
LC-ESI-QFT; MS2; CE: 30; R=35000; [M+H]+			EQ331902
LC-ESI-QFT; MS2; CE: 30; R=35000; [M-H]-			EQ331952
LC-ESI-QFT; MS2; CE: 45; R=35000; [M+H]+			EQ331903
LC-ESI-QFT; MS2; CE: 45; R=35000; [M-H]-			EQ331953
LC-ESI-QFT; MS2; CE: 60; R=35000; [M+H]+			EQ331904
LC-ESI-QFT; MS2; CE: 60; R=35000; [M-H]-			EQ331954
LC-ESI-QFT; MS2; CE: 75; R=35000; [M+H]+			EQ331905
LC-ESI-QFT; MS2; CE: 75; R=35000; [M-H]-			EQ331955
LC-ESI-QFT; MS2; CE: 90; R=35000; [M+H]+			EQ331906
LC-ESI-QFT; MS2; CE: 90; R=35000; [M-H]-			EQ331956

 DOI: 10.5281/zenodo.3953683