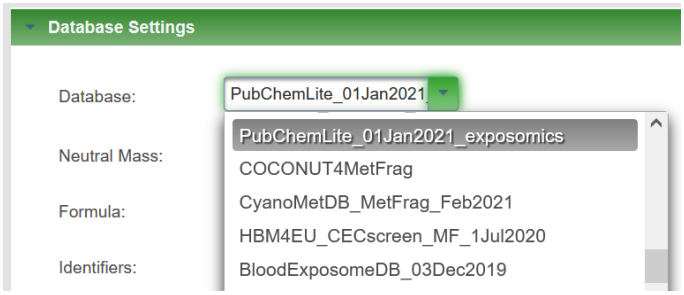
Example 1: Nicotine

Practicing MetFrag (https://msbi.ipb-halle.de/MetFrag/) with a formula search using ***PubChemLite for Exposomics*** Local CSV from the dropdown menu:



Using the Mass Spectrum of Nicotine from MassBank:

https://massbank.eu/MassBank/RecordDisplay?id=EQ300801

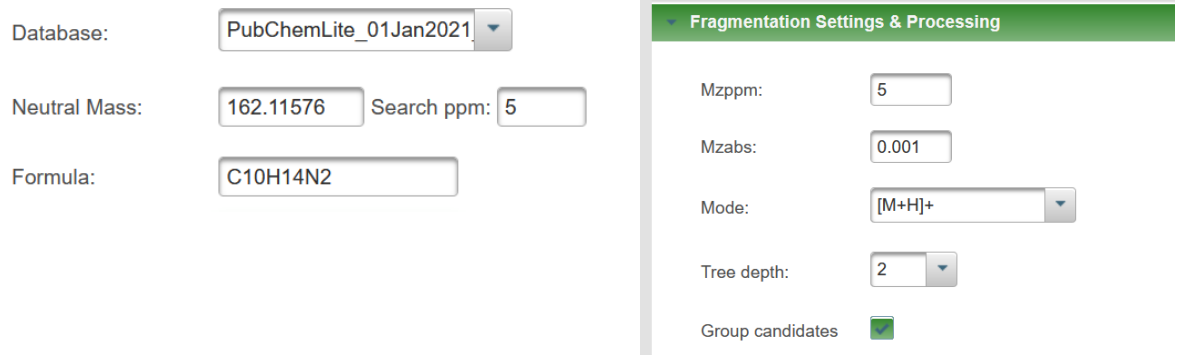
Peaks:

```
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]+

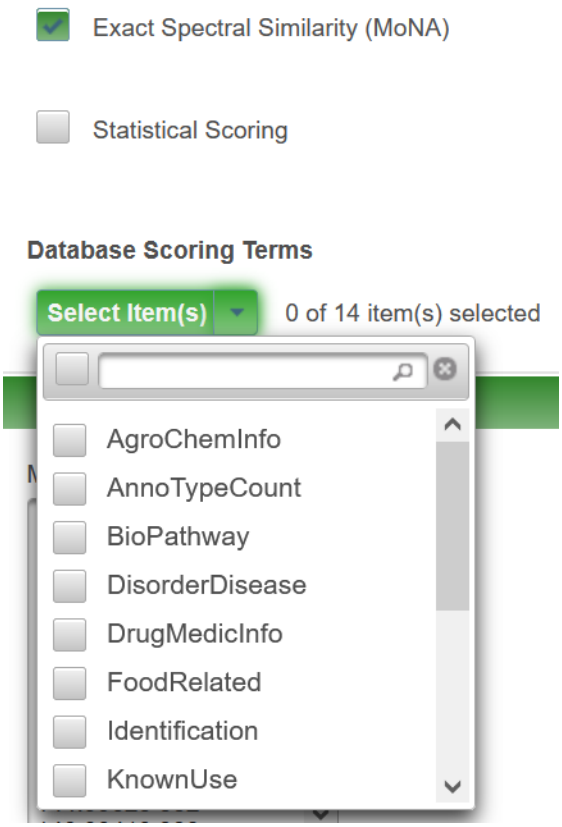
Accuracy settings: 5 ppm + 0.001 Da



Try with various scoring terms (see next page) to see the influence of different metadata terms.

How confident are you in the top ranked candidate?

 DOI: 10.5281/zenodo.6856116



Can you find out information about the relevance of the candidates in PubChem?

https://pubchem.ncbi.nlm.nih.gov/compound/89594#section=Associated-Disorders-and-Diseases

PubChem Nicotine (Compound)

14 Associated Disorders and Diseases

Page 41 of 246 items View More Rows & Details Download

Disease	Evidence Type	Evidence PMID
Respiratory System Abnormalities	marker/mechanism	20363831
Respiratory Tract Diseases	marker/mechanism	27806211
Retinal Diseases	marker/mechanism	20448088

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

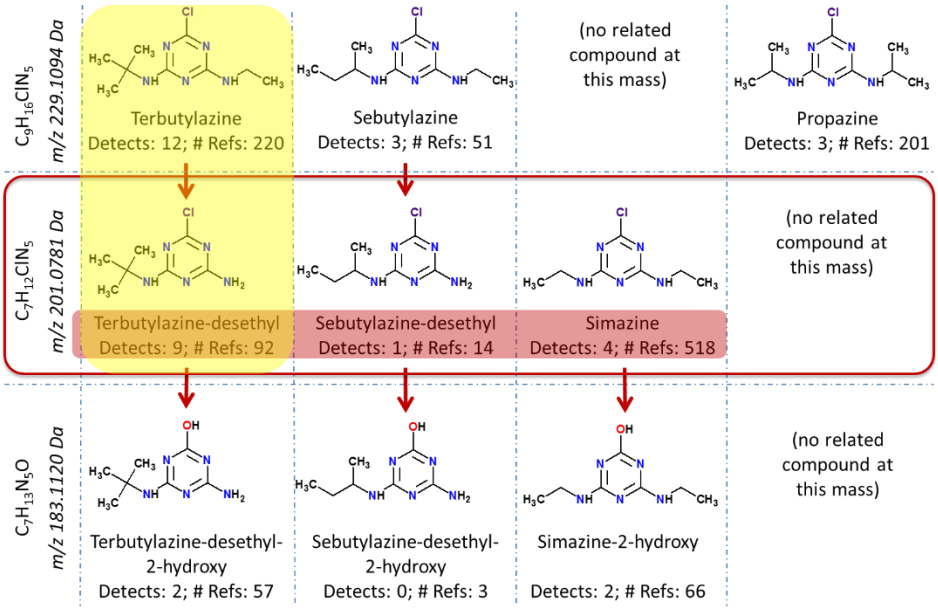
https://massbank.eu/MassBank/RecordDisplay?id=EQ300804

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

 DOI: 10.5281/zenodo.6856116

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 ***PubChemLite for Exposomics***, 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?id=EA067112

```
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?id=EA026205

```
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 in the top ranked candidate?
Is the best candidate always ranked first?

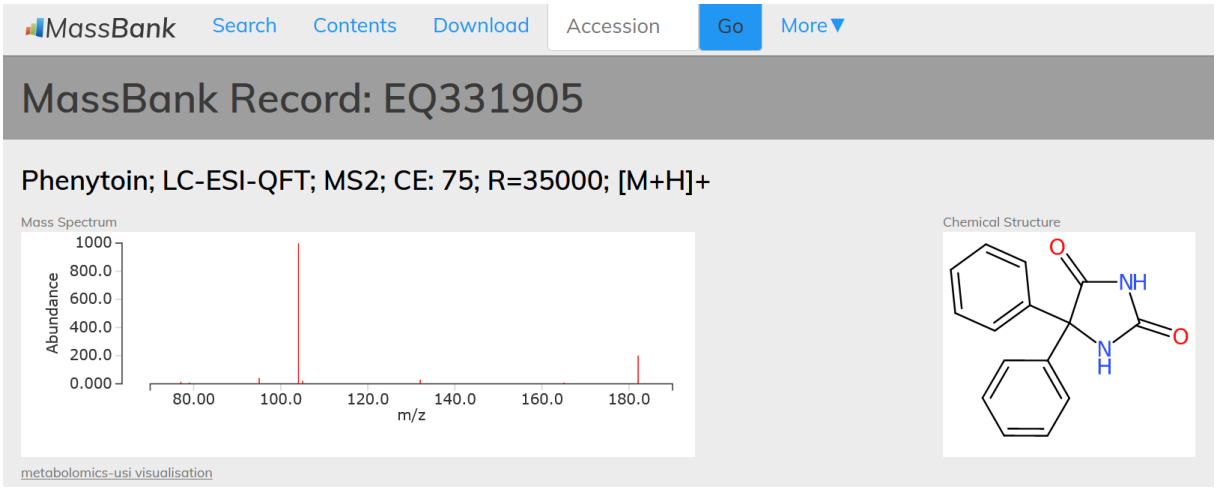
Extra Examples

These are examples from Schymanski *et al* 2019 DOI: 10.1039/C9EM00068B, with selected formulas that have multiple candidates with neurological disease endpoints that also have spectra in MassBank as a "proof of concept". These serve to show how metadata may (or may not) help you select chemicals relevant for specific questions. These examples were chosen using keywords from the participant descriptions during the 2019 Exposome Boot Camp.

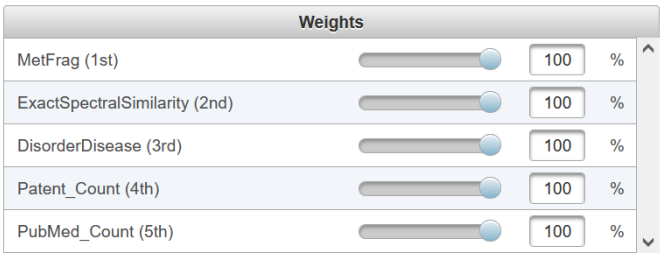
Example for $C_{15}H_{12}N_2O_2$ (Spectrum of Phenytoin):

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

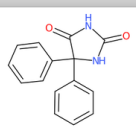
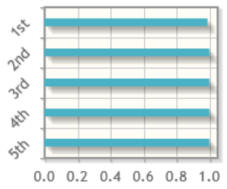
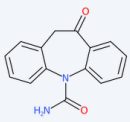
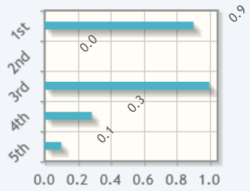
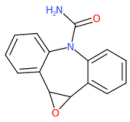
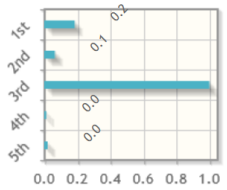
```
77.0385 5324013 15
79.0542 3557309.5 10
81.0334 388320.8 1
95.0491 14019273 41
103.0412 409314.6 1
104.0494 339629280 999
105.0333 6008052 17
105.0445 6911438 20
122.0599 695733.5 2
132.0443 10507750 30
165.0699 3174707.5 9
167.073 690844.3 2
180.0808 721253.2 2
182.0964 68547840 201
```



Select different metadata terms … below is one example:



For instance 3 candidates have associated disease and disorder information:

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 5,5-diphenylimidazolidine-2,4-dione	1775 InChIKeyBlock1 = CXOFVDLJLONNDW	252.08988	$C_{15}H_{12}N_2O_2$		4.9876	Peaks: 9 / 14 Fragments Scores Download
2	 5-oxo-6H-benzo[b][1]benzazepine-11-carboxamide	34312 InChIKeyBlock1 = CTRLABGOLIVAIY	252.08988	$C_{15}H_{12}N_2O_2$		2.2948	Peaks: 11 / 14 Fragments Scores Download
3	 3-oxa-11-azatetracyclo[10.4.0.02,4.05,10]hexadeca-1(16),5,7,9,12,14-hexaene-11-	2555 InChIKeyBlock1 = ZRWWEEVEIOGMMT	252.08988	$C_{15}H_{12}N_2O_2$		1.2811	Peaks: 7 / 14 Fragments Scores Download

Try browsing the information for these candidates in PubChem to find relevant diseases:

https://pubchem.ncbi.nlm.nih.gov/compound/1775#section=Associated-Disorders-and-Diseases

https://pubchem.ncbi.nlm.nih.gov/compound/34312#section=Associated-Disorders-and-Diseases

https://pubchem.ncbi.nlm.nih.gov/compound/2555#section=Associated-Disorders-and-Diseases

Do candidates 2 and 3 look similar to another common medication?

https://pubchem.ncbi.nlm.nih.gov/compound/2555#section=Transformations

https://pubchem.ncbi.nlm.nih.gov/compound/34312#section=Human-Metabolite-Information

https://smpdb.ca/view/SMP0000634

Do the spectral similarity values here help you have confidence in the correct structure?

Would you get the same result if you did not consider exact spectral match?

On the basis of information here … would you screen for related compounds in your samples?

https://pubchem.ncbi.nlm.nih.gov/compound/1775#section=Transformations&fullscreen=true

 DOI: 10.5281/zenodo.6856116

Other Examples

Other formulas to try (formula search on MassBank for spectra) are $C_{14}H_{22}N_2O_3$, $C_{10}H_{13}N_5O_4$, $C_{10}H_{15}NO$, $C_{11}H_9I_3N_2O_4$, $C_{12}H_{12}N_2O_3$. Try to pick different spectra for different compounds with the same formula to see the effect of metadata vs experimental evidence. On the next page is one suggested example.

Example for $C_{10}H_{13}NO_2$ (Spectrum of MDA):

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

```
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
```

 DOI: 10.5281/zenodo.6856116

Further Reading:

If you would like to find out more about PubChem [1], PubChemLite [2], the disease-related content in PubChem [3], earlier work with CompTox [4] and our latest collective efforts [5], or provide data to expand our knowledgebase of exposomics content [6, 7], please consider the following references:

1. Kim S, Chen J, Cheng T, et al (2021) PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Research 49:D1388–D1395. https://doi.org/10.1093/nar/gkaa971

2. Schymanski EL, Kondić T, Neumann S, et al (2021) Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. J Cheminform 13:19. https://doi.org/10.1186/s13321-021-00489-0

3. Zaslavsky L, Cheng T, Gindulyte A, et al (2021) Discovering and Summarizing Relationships Between Chemicals, Genes, Proteins, and Diseases in PubChem. Front Res Metr Anal 6:689059. https://doi.org/10.3389/frma.2021.689059

4. Schymanski EL, Baker NC, Williams AJ, et al (2019) Connecting environmental exposure and neurodegeneration using cheminformatics and high resolution mass spectrometry: potential and challenges. Environ Sci: Processes Impacts 21:1426–1445. https://doi.org/10.1039/C9EM00068B

5. Talavera Andújar B, Aurich D, Aho VTE, et al (2022) Studying the Parkinson's disease metabolome and exposome in biological samples through different analytical and cheminformatics approaches: a pilot study. Anal Bioanal Chem. https://doi.org/10.1007/s00216-022-04207-z

6. Schymanski EL, Bolton EE (2021) FAIR chemical structures in the Journal of Cheminformatics. J Cheminform 13:50. https://doi.org/10.1186/s13321-021-00520-4

7. Schymanski EL, Bolton EE (2021) FAIR-ifying the Exposome Journal: Templates for Chemical Structures and Transformations. Exposome osab006. https://doi.org/10.1093/exposome/osab006