

Databases: Compound Databases and Spectral Databases

Answers for Suggested Exercises

ChemSpider and PubChem

1. Find out more about Ibuprofen

- a) Molecular formula – **C₁₃H₁₈O₂**
- b) Exact mass – **206.1307**
- c) ChemSpider (**3544**), PubChem (**3672**) Record Number, CAS Number (**15687-27-1**)?
(CAS number is tricky!)
- d) How many data sources and references?
119 Sources, 6132 References, 7925 PubMed entries via ChemSpider
- e) What is the experimental log P? How does this compare to calculated values?
Experimental: 3.722 Vitas-M, 3.97 EPISuite.
Predicted: ACD/LogP: 3.502, EPISuite 3.79, Chemicalize: 3.84.
These all agree pretty well!
- f) How many structures contain the structure of Ibuprofen in them?
(hint: substructure search – but the answer is not 100, it's **1297 via ChemSpider**)
- g) How many active bioassays? **Active in 98 of 1654 BioAssays via PubChem: in the search summary already**
- h) How many records have the same CAS Number associated with them? **5 (ChemSp)**
Are they all the same? **No, three are similar but stereochemistry and ion different.**
Other two different structures (although related).
- i) How many records have the same formula? **1748 via ChemSpider**

2. What is this? Clc1nc(nc(n1)NC(C)C)NCC. **Atrazine**

- a) Molecular formula – **C₈H₁₄ClN₅**
- b) Exact mass – **215.0937**
- c) ChemSpider (**2169**), PubChem (**2256**) Record Number, CAS Number (**1912-24-9**)?
- d) How many data sources and references?
87 Sources, 1847 References, 2586 PubMed entries via ChemSpider
- e) What is the experimental log P? How does this compare to calculated values?
Experimental: 2.632 Vitas-M, 2.61 EPISuite.
Predicted: ACD/LogP: 2.636, EPISuite 2.82, Chemicalize: 2.20.
These all agree pretty well, chemicalize a little low.
- f) How many structures contain the structure of this compound in them?
(hint: substructure search – but the answer is not 100, it's **827 via ChemSpider**)
- g) How many active bioassays? **Active in 8 of 716 BioAssays via PubChem**
- h) How many records have the same CAS Number associated with them? **1 (ChemSp)**
Are they all the same? **Only 1.**
- i) How many records have the same formula? **16 via ChemSpider**

3. What is this? JPMIIZHYWMDHT-UHFFFAOYSA-N. **2-octylisothiazol-3(2H)-one**

- a) Molecular formula - **C₁₁H₁₉NOS**
- b) Exact mass – **213.1187**
- c) ChemSpider (**30932**), PubChem (**33528**) Record Number, CAS Number (**122667-23-6, 26530-20-1, 12673-72-2**)? (CAS number is tricky!)
- d) How many data sources and references?
37 Sources, 93 References, 0 PubMed entries via ChemSpider
- e) What is the experimental log P? How does this compare to calculated values?
Experimental: 2.45 EPISuite.
Predicted: ACD/LogP: 3.685, EPISuite 2.61, Chemicalize: 3.33.
These are quite different!
- f) How many structures contain the structure in them? **80 via ChemSpider**
- g) How many active bioassays? **Active in 97 of 460 BioAssays via PubChem.**
- h) How many records have the same CAS Number associated with them? **1 associated with 122667-23-6, 12673-72-2, but two associated with 26530-20-1 (ChemSp)**
Are they all the same? **The two are different structures – don't trust the database all the time!**
- i) How many records have the same formula? **72 via ChemSpider**

Extra: Can you buy these compounds? From where? **Yes, look at the data sources**

MassBank and METLIN

1. Try the METLIN example – what neat features does it have?

Look at the substructures for the peaks!

2. Are these first three compounds in MassBank?

<http://www.massbank.jp/> ? Ibuprofen: **7 spectra**. Atrazine: **15 spectra**. Othilinone: **14 spectra**.

<http://massbank.normandata.eu/MassBank/> ? Ibuprofen: **10 spectra**. Atrazine: **44 spectra**. Othilinone: **14 spectra**

Is there a difference? **Yes**

3. Are they in METLIN?

Ibuprofen: **yes, plus 4 others related compounds – only MSMS for ibuprofen**

Atrazine: **yes, plus 5 related compounds – but only MSMS for Atrazine**

Othilinone: **yes, but no MSMS**