An Introduction to PubChem for Life Scientists

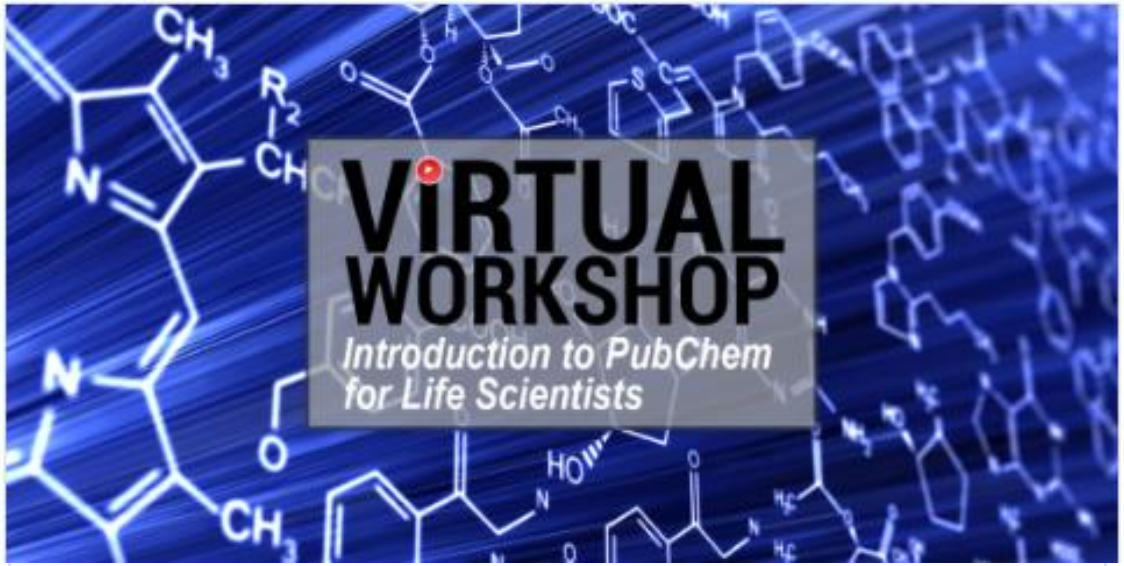
Alexa M. Salsbury, Ph.D.

National Library of Medicine
National Center for Biotechnology Information

Workshop Details

Overview

- Background
- Searching in PubChem
- Finding molecular & chemical safety information
- Using PubChem Sketcher
- Finding a potential inhibitor



An Introduction to PubChem for Life Scientists

Scientists and educators from many different fields need to find and understand chemical data to perform their work. As such, understanding PubChem, the world's largest collection of freely accessible chemical information, is a powerful skill for researchers, educators, clinicians, and more. In this workshop, participants will have access to NCBI experts and be introduced to the functionality of the PubChem Database.

POLL (1/4)

Which best describe your work, research, or educational background?

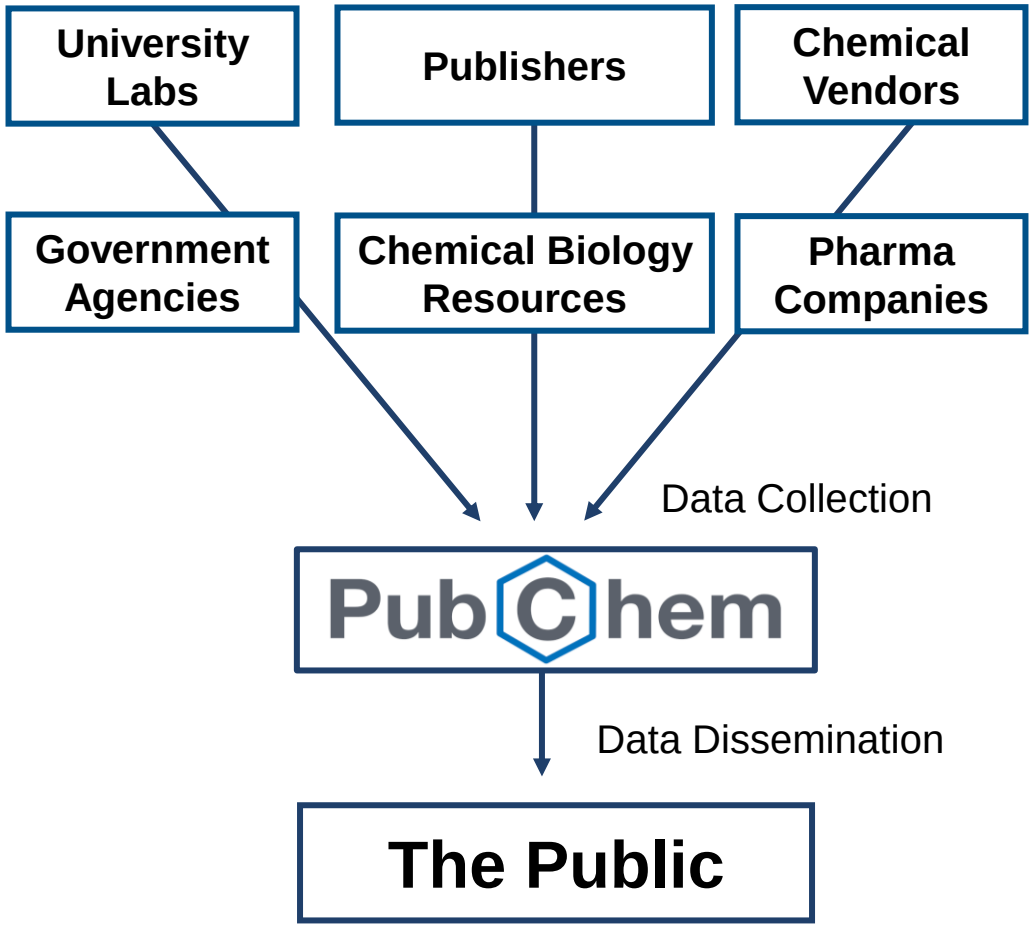
POLL (2/4)

Have you attended an NCBI workshop before?

What is PubChem?

- Created in 2004
- World's largest collection of freely accessible chemical information
  - Key chemical information resource for scientists, students, and the public
- Learn more:

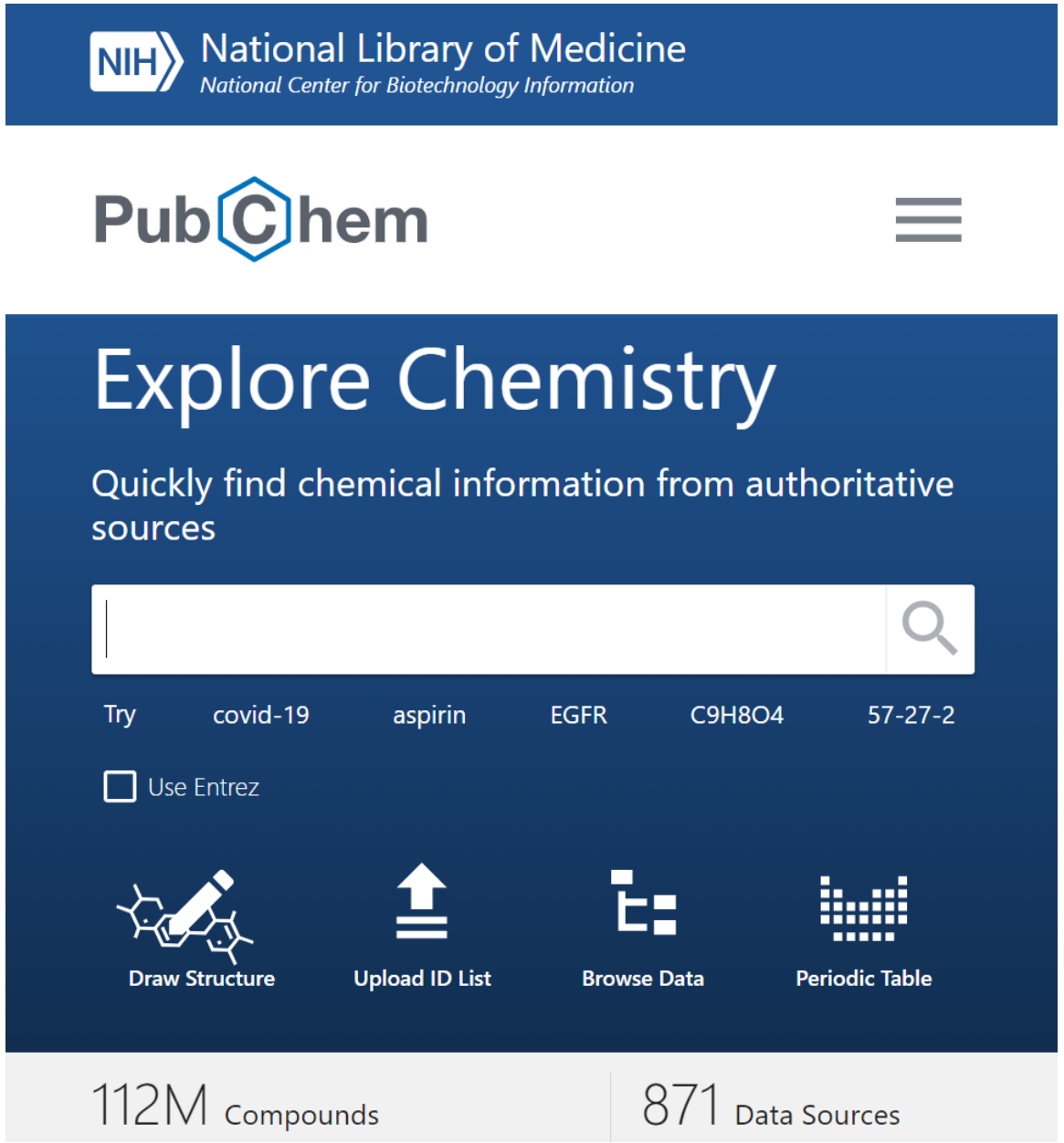
  https://pubchem.ncbi.nlm.nih.gov



What's in PubChem?

- Small molecules
- Larger molecules
  - Nucleotides
  - Carbohydrates
  - Lipids
  - Peptides
  - Chemically-modified macromolecules
- And more

PubChem



Common uses for PubChem

- Learn about a chemical, drug, or medication
- Get biological function, molecular structure, molecular weight, chemical safety, etc. data for fundamental research
- Mine data for biochemical research
  - Drug discovery, drug-gene targeting, toxicology

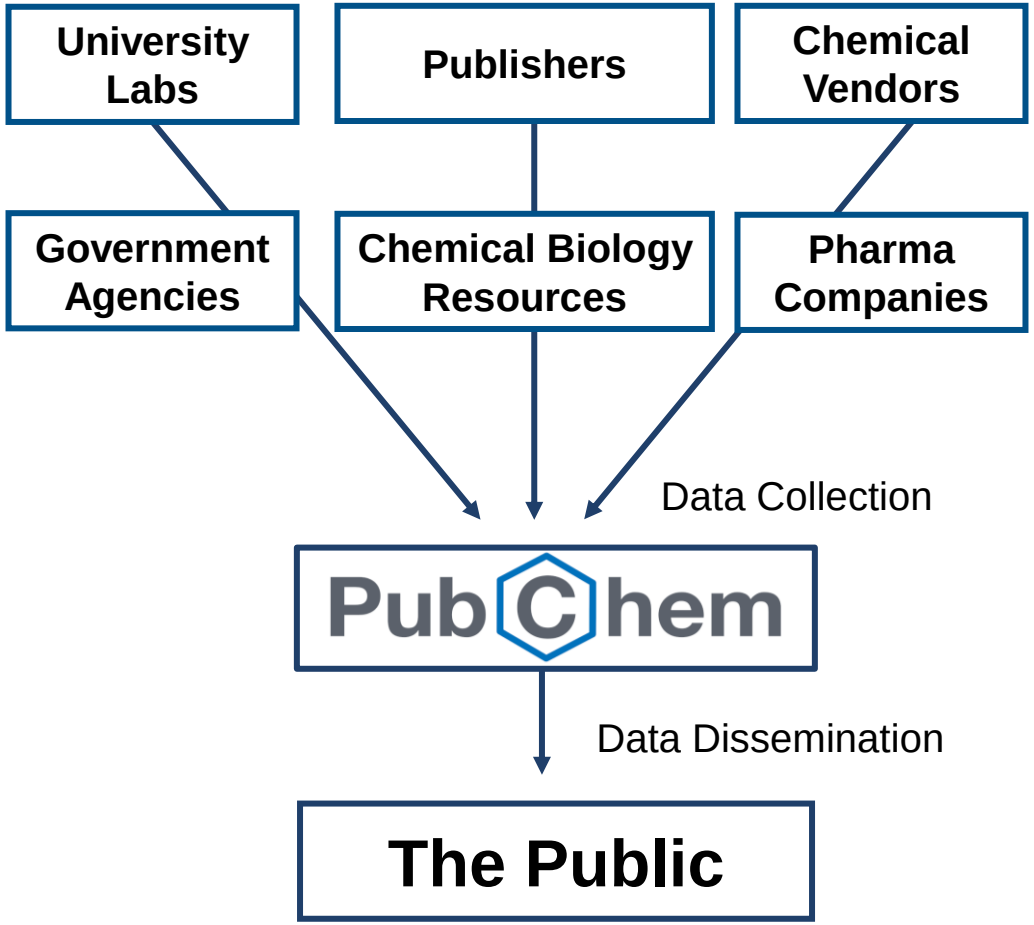
PubChem

POLL (3/4)

What do you currently use PubChem for?

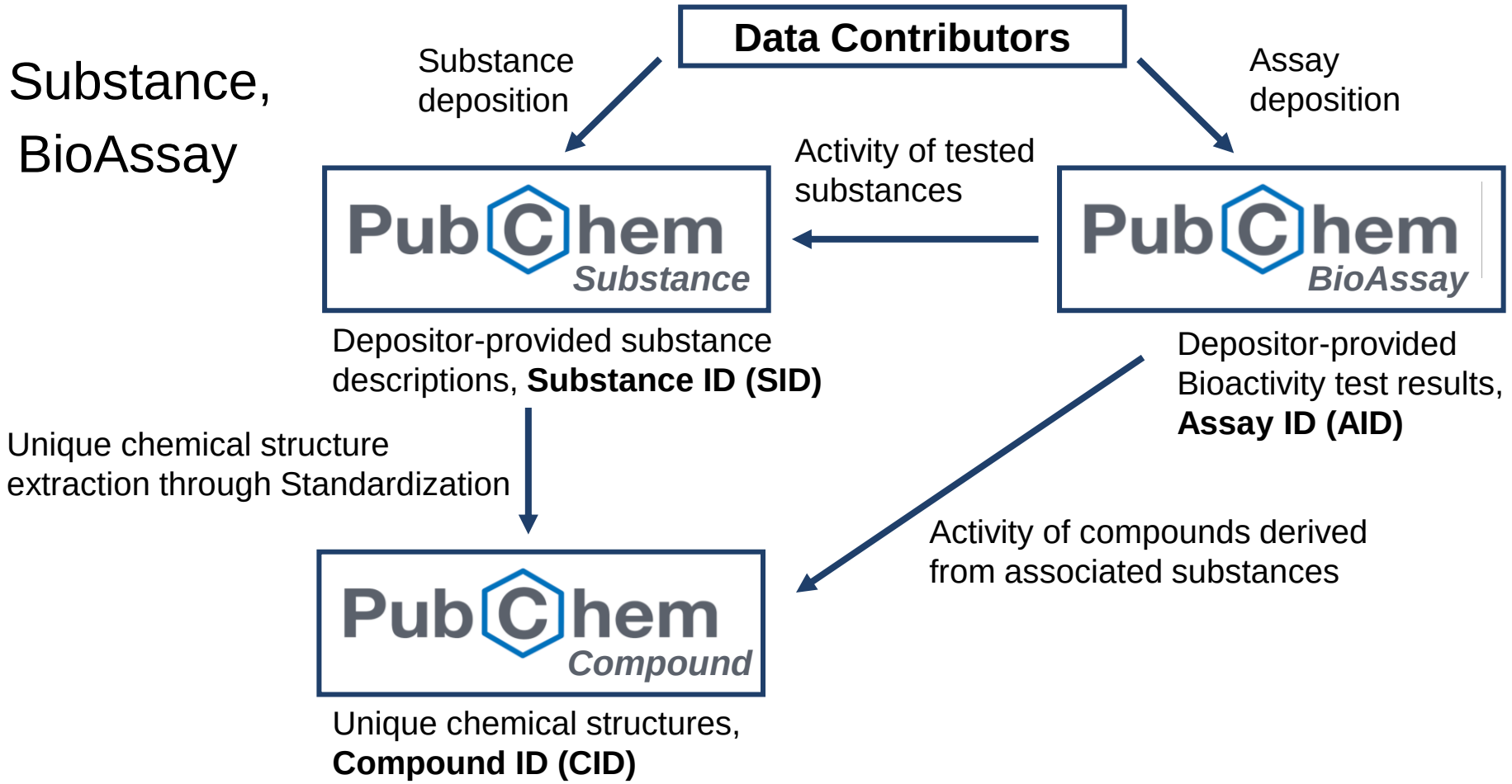
Who provides PubChem Data?

- Submitters and contributors are vetted
  - Algorithmically
  - Manually
  - Post-production



Data Organization

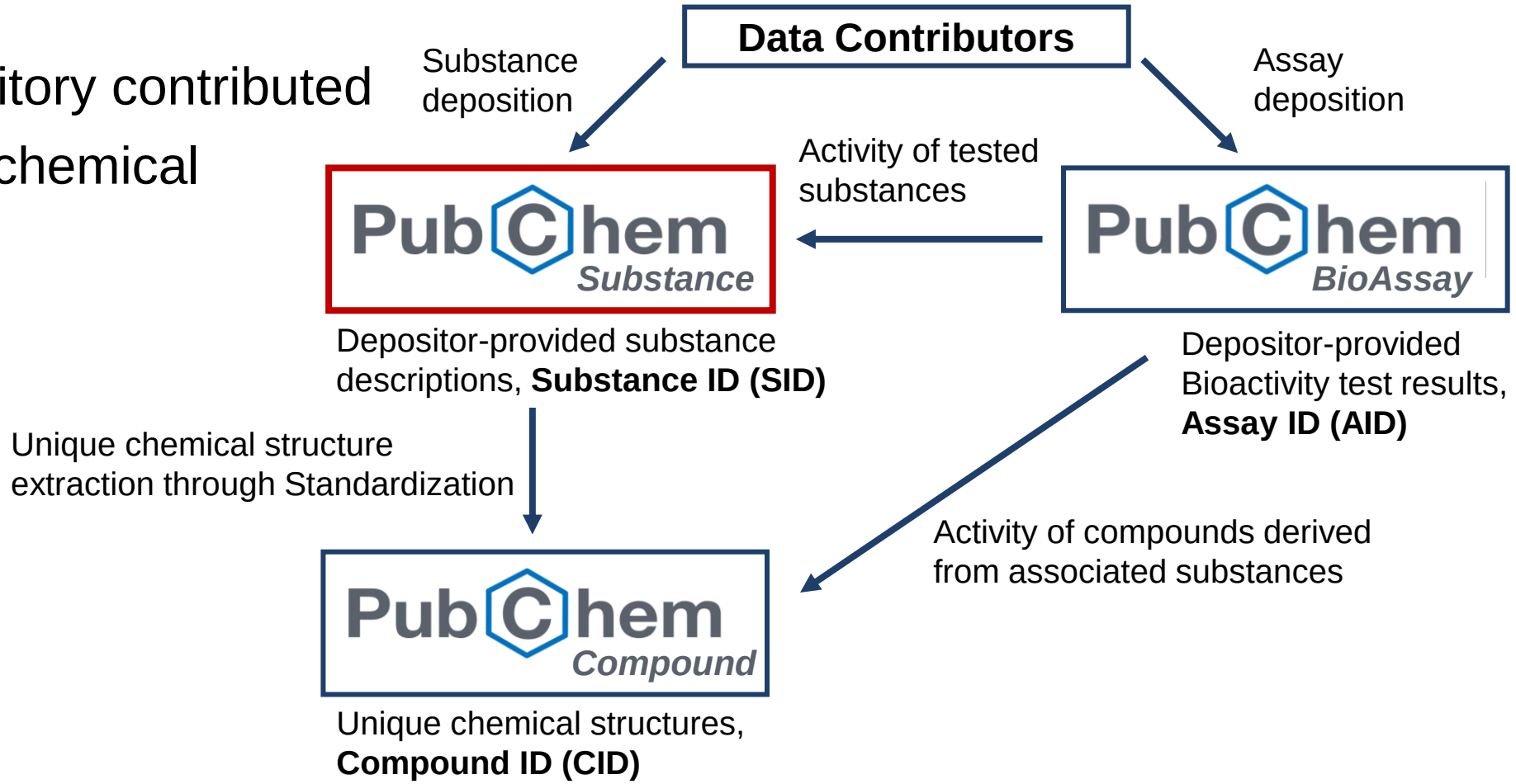
- Three Linked to Substance, Compound, and BioAssay Databases



More information here

Substance Database

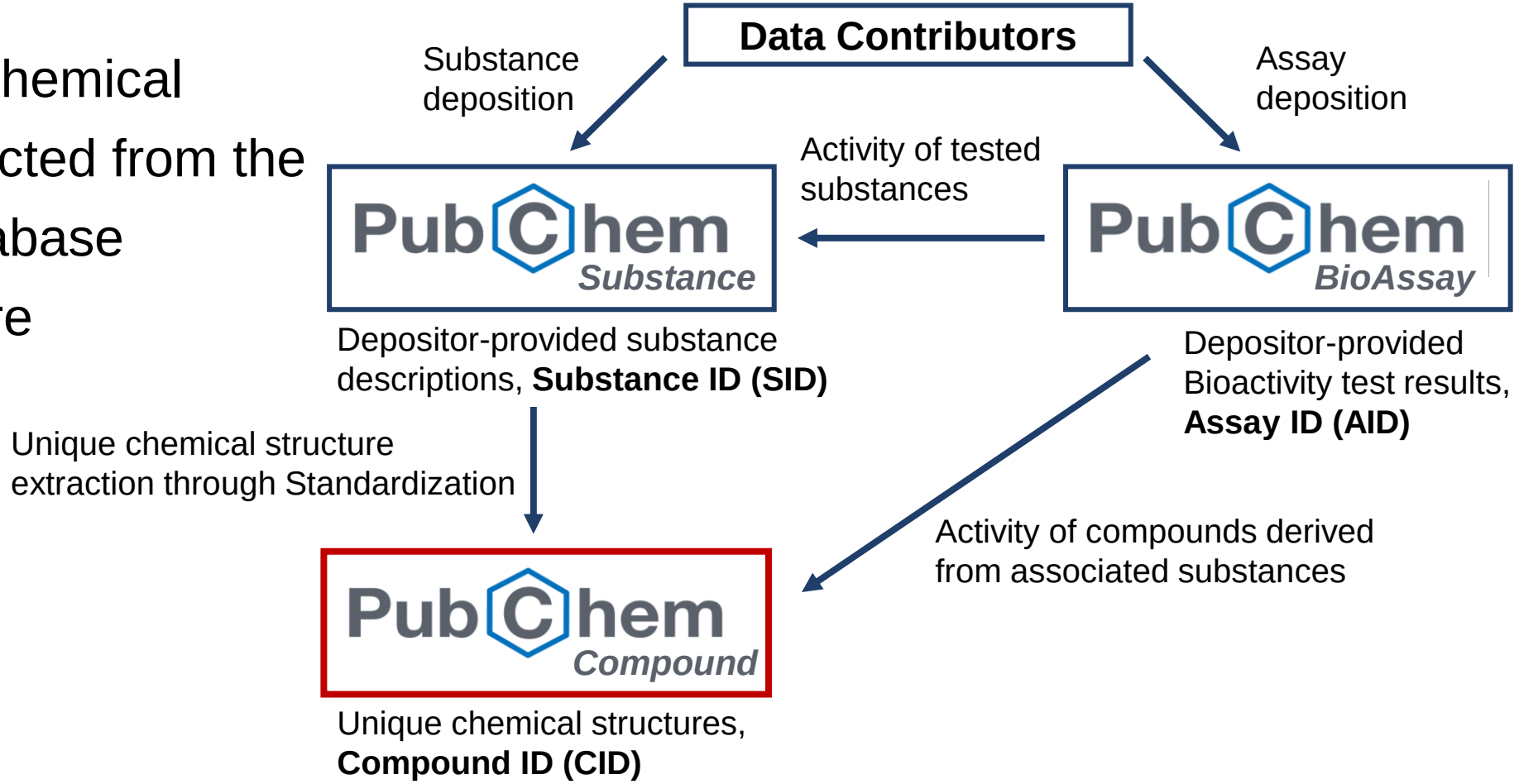
- Archives depository contributed descriptions of chemical substances



More about tools and services in Substance & Compound

Compound Database

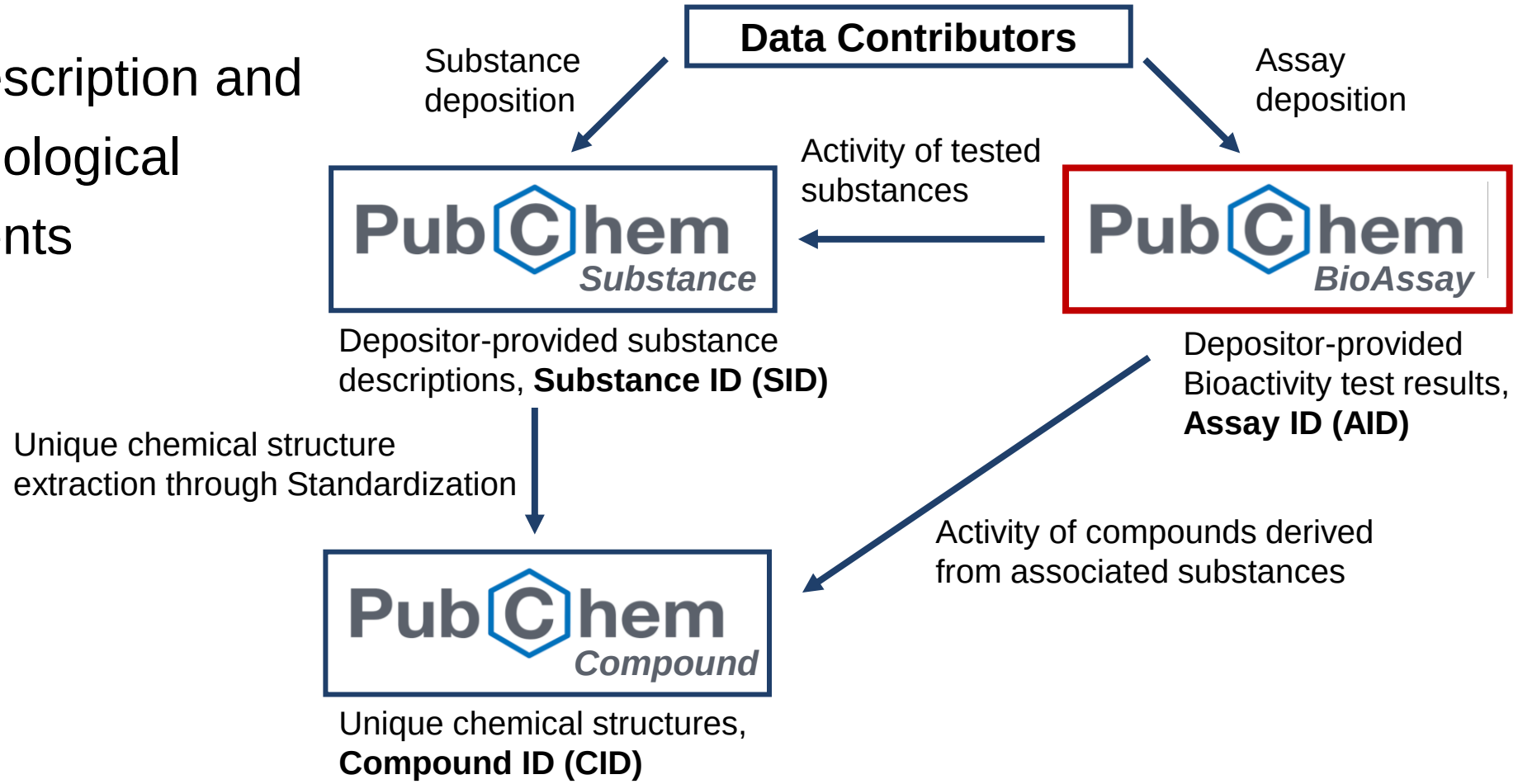
- Stores unique chemical structures extracted from the Substance Database through structure standardization



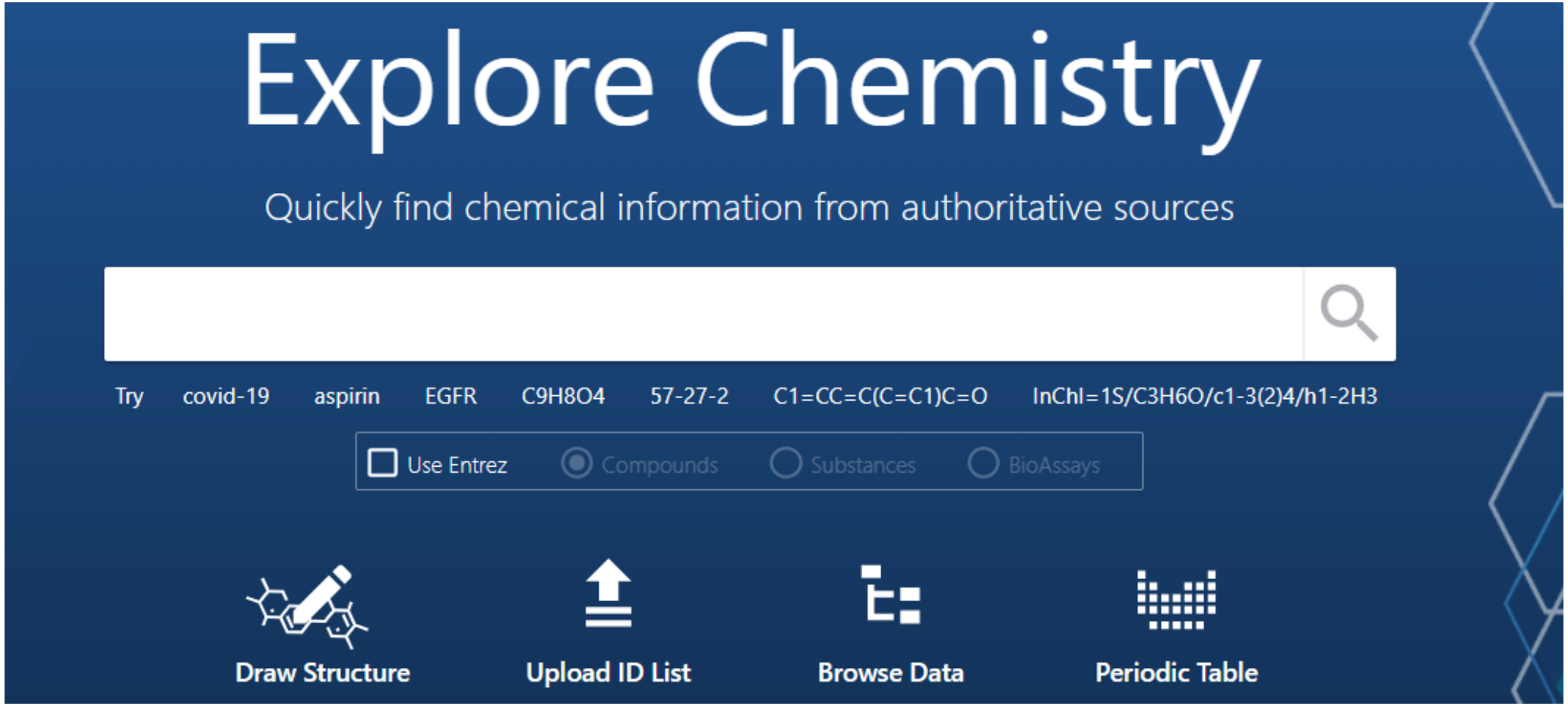
More about tools and services in Substance & Compound

BioAssay Database

- Contains the description and test results of biological assay experiments



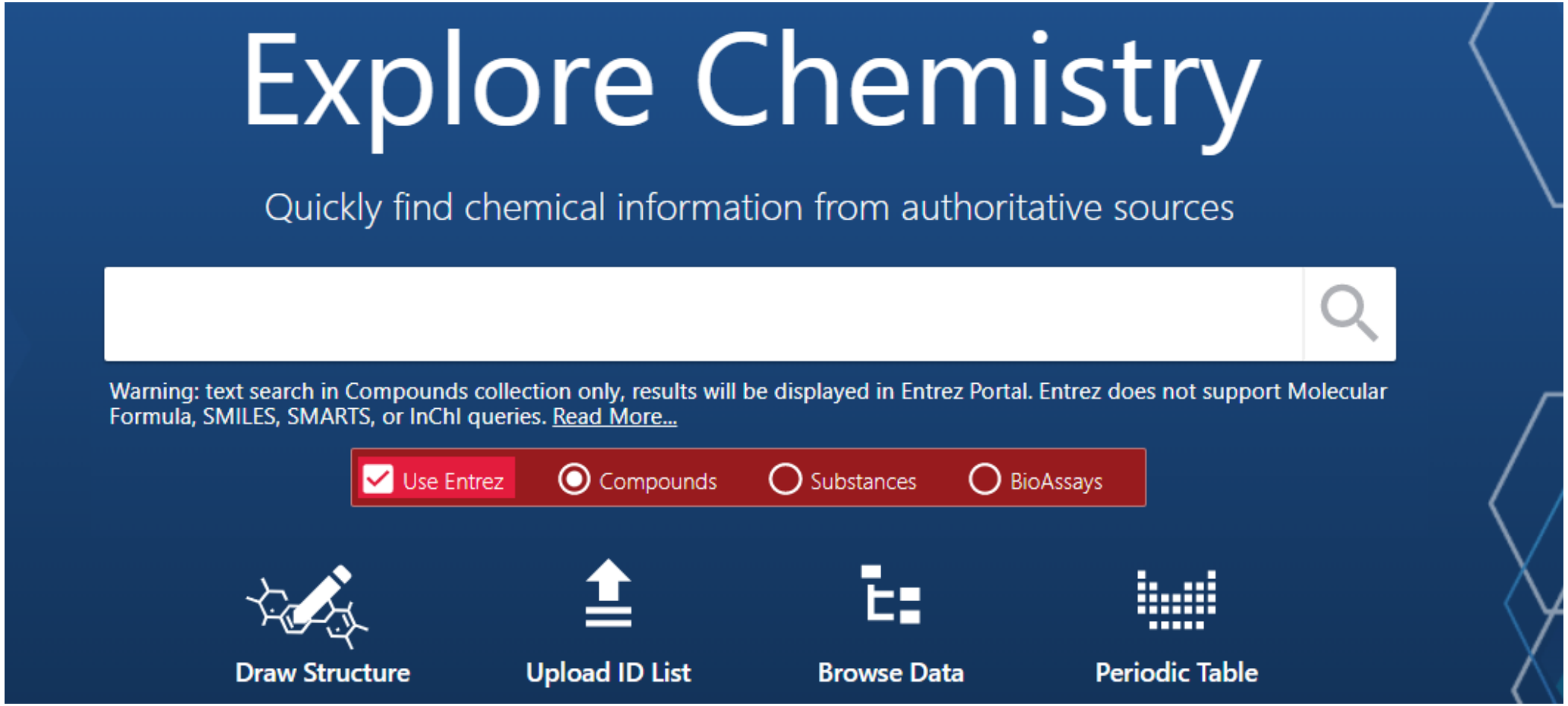
Searching in PubChem



Common Search Strategies

Text	covid-19 OR aspirin
Molecular Formula	C9H8O4
Simplified Molecular Input Line Entry System (SMILES)	C1=CC=C(C=C1)C=O
InChI Identifier	InChI=1S/C3H6O/c1-3(2)4/h1-2H3

Searching with Entrez



Search Tips

Entrez is a molecular biology database system that provides access to a wealth of NCBI data

- More Entrez Help is available on the NCBI website

Finding structures with Entrez

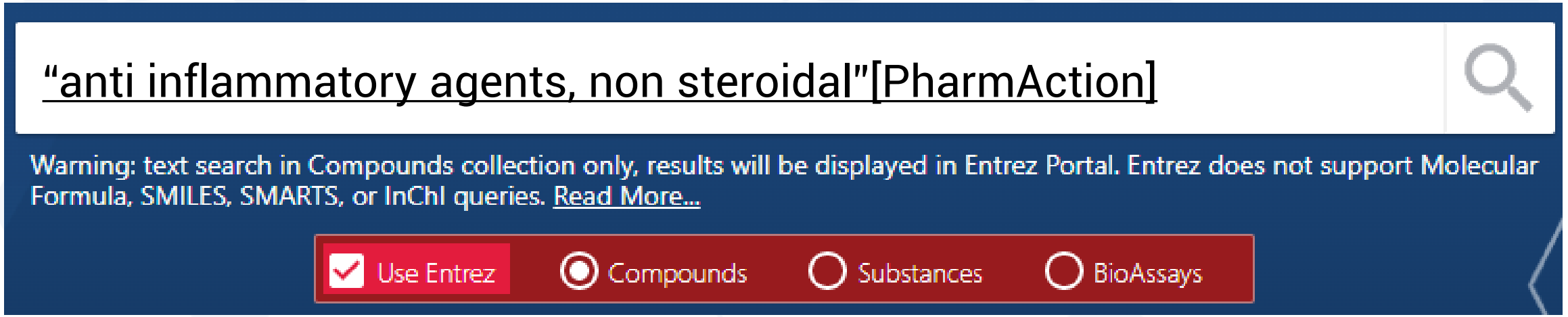
```
"anti inflammatory agents, non steroidal"[pharmaction]
```

- Use field limits and Boolean operators
- Put phrases in quotes

More on Entrez Indices and Filters

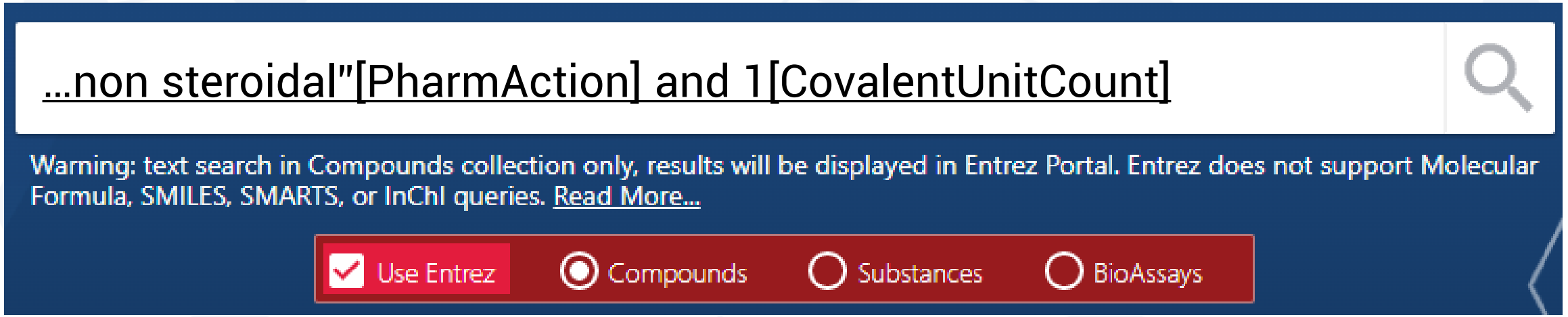
Text searching in PubChem (1/8)

Click Use Entrez and Compounds, then use the field PharmAction to limit by Pharmacological action:



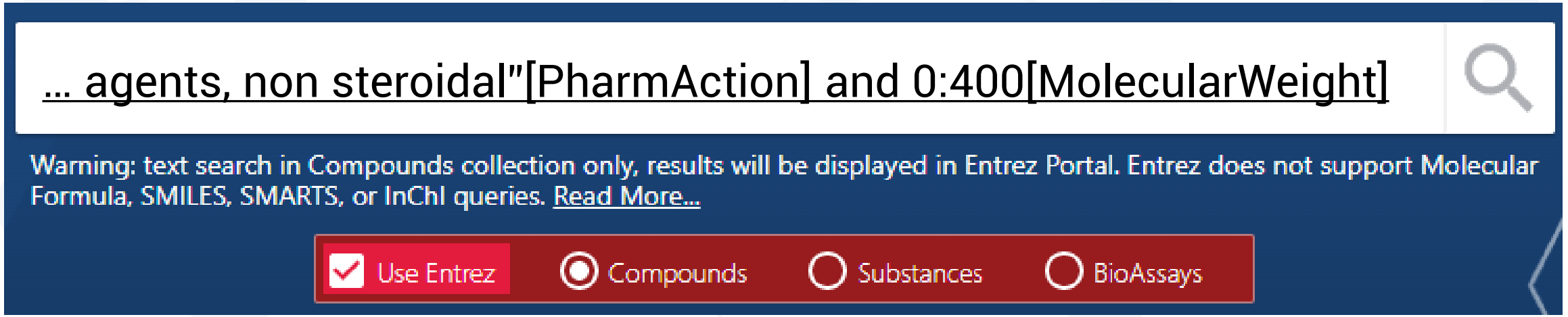
Text searching in PubChem (2/8)

Narrow your search to single chemical NSAIDS by adding another term, **and 1[CovalentUnitCount]:**



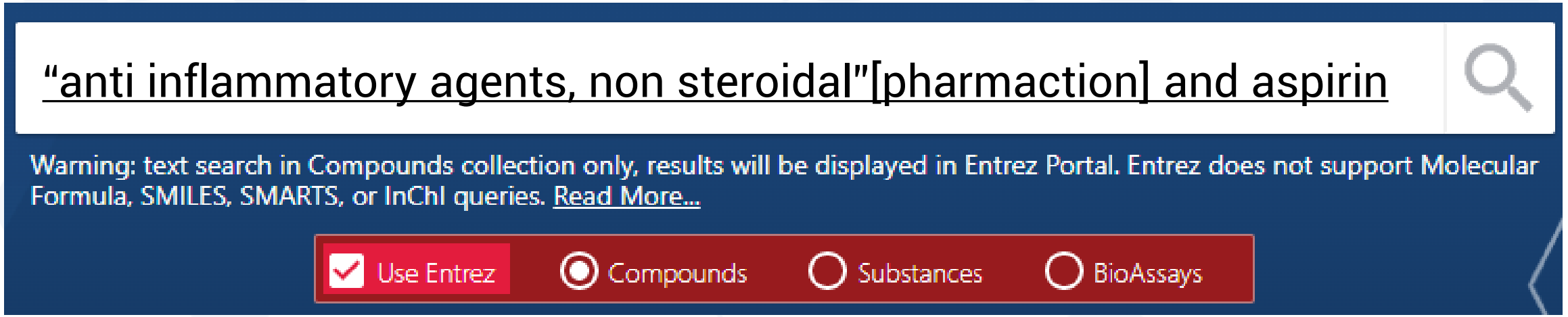
Text searching in PubChem (3/8)

Or narrow your search to smaller chemical molecules (<400 g/mol) with, **and 0:400[MolecularWeight]**



Text searching in PubChem (4/8)

Or narrow your search by including a simple term, and aspirin:



Text searching in PubChem (5/8)

Answer the following:

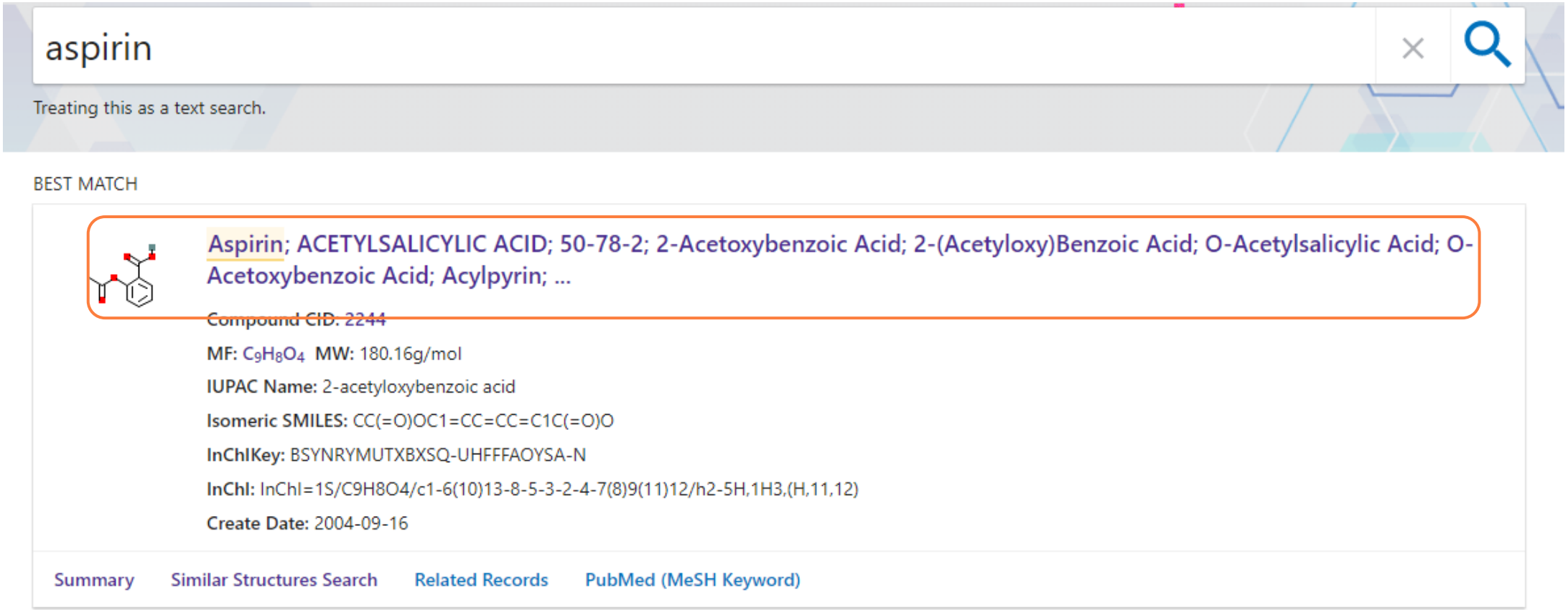
5-minute exercise!

- What are synonyms for aspirin?
- What is the Molecular Formula?
- Molecular weight?
- Are there patents associated with this entry?
- Do you recognize any of the information sources?

Share an information source you recognize in the chat!

Text searching in PubChem (6/8)

Click on the Best Match result:



Text searching in PubChem (7/8)

Review the compound summary page:

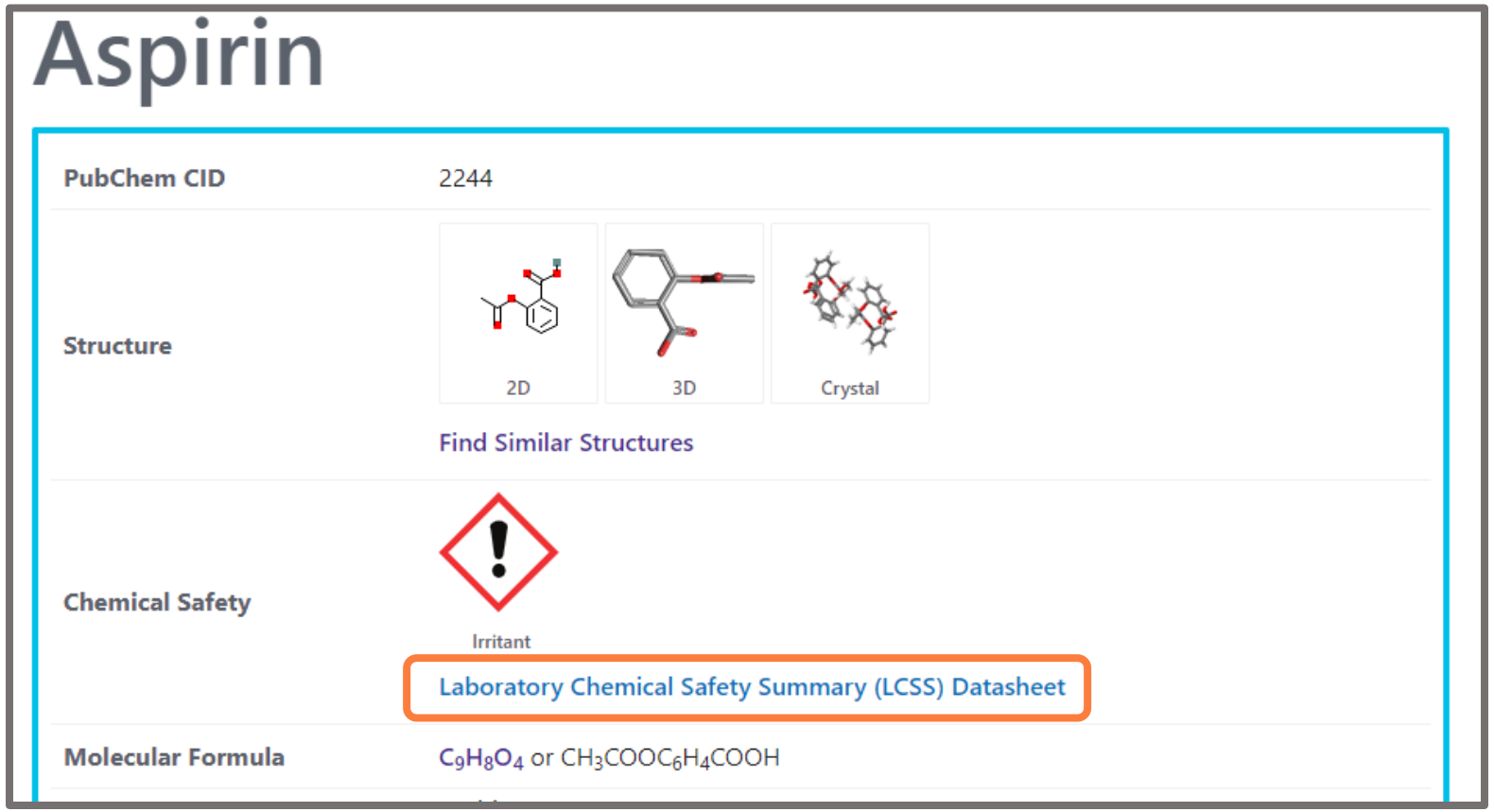
Aspirin

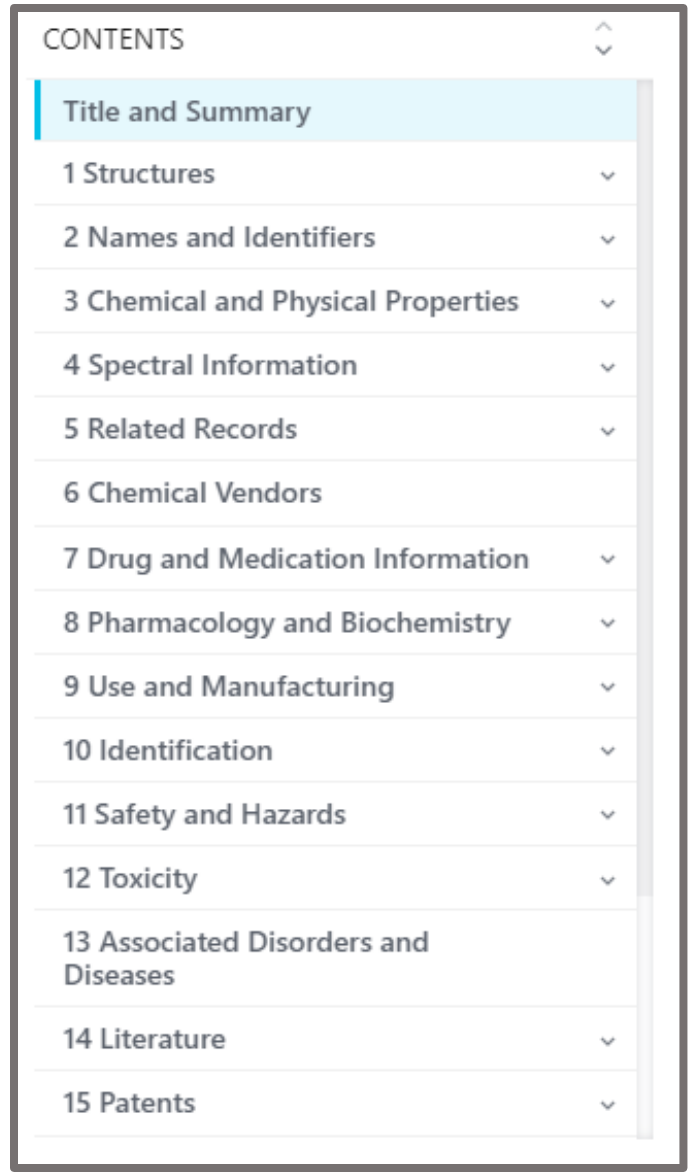
PubChem CID	2244
Structure	 2D 3D Crystal Find Similar Structures
Chemical Safety	 Irritant Laboratory Chemical Safety Summary (LCSS) Datasheet
Molecular Formula	$C_9H_8O_4$ or $CH_3COOC_6H_4COOH$

CONTENTS

Text searching in PubChem (8/8)

Click on the LCSS Datasheet:





Finding safety information

COMPOUND SUMMARY > LABORATORY CHEMICAL SAFETY SUMMARY (LCSS)

Aspirin

PubChem CID	2244
Structure	 2D Find Similar Structures
Synonyms	aspirin ACETYLSALICYLIC ACID 50-78-2 2-Acetoxybenzoic acid 2-(Acetyloxy)benzoic acid More...
Molecular Formula	$C_9H_8O_4$ or $CH_3COOC_6H_4COOH$
Molecular Weight	180.16

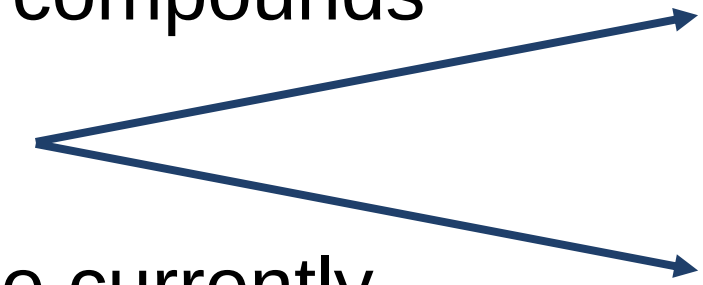
Learn More About LCSS Project >

Chemical safety information

- Lab safety is an essential topic in all labs
- PubChem's Laboratory Chemical Safety Summary (LCSS) is a thorough online resource for chemical safety information

PubChem LCSS

- You can find LCSS from the summary page of applicable compounds
- 12 main headings
- >170,000 LCSS are currently available in PubChem



Contents

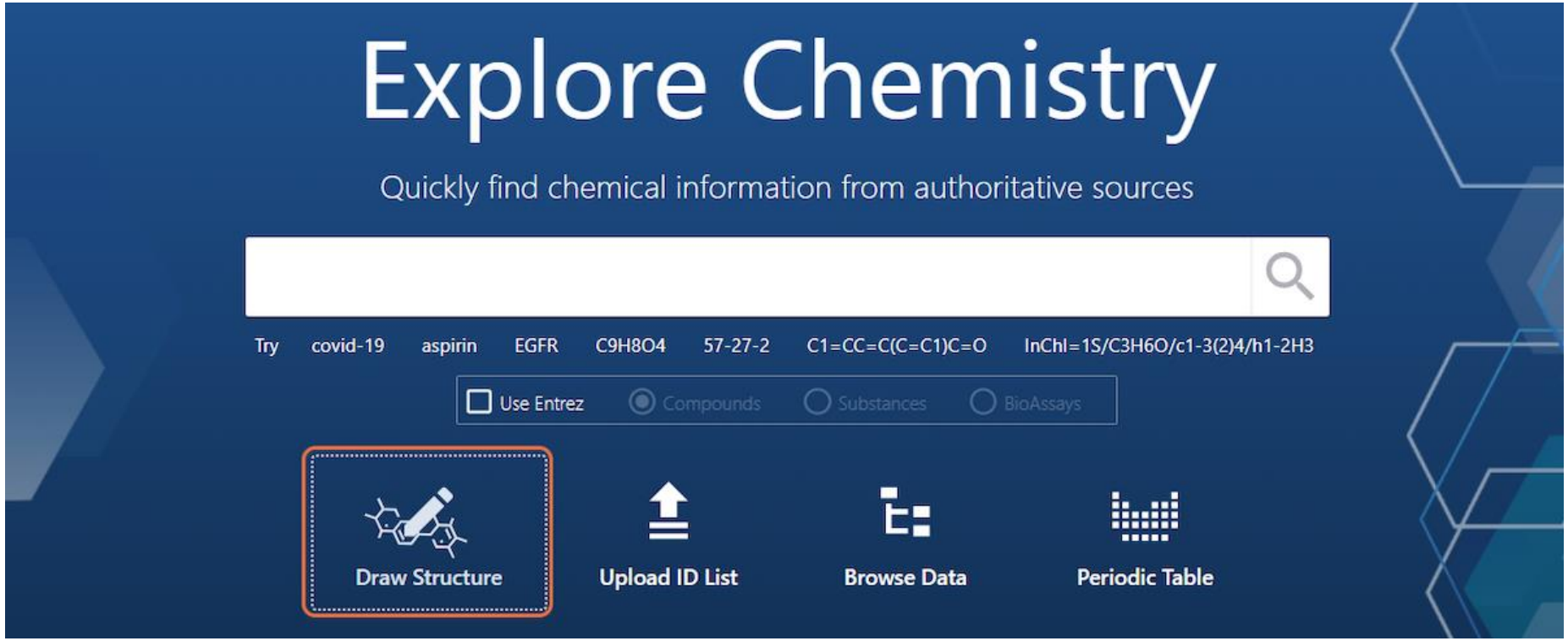
1 GHS Classification Learn more
2 Identifiers
3 Physical Properties
4 Toxicity Information
5 Exposure Limits
6 Health and Symptoms
7 First Aid
8 Flammability and Explosivity
9 Stability and Reactivity
10 Storage and Handling
11 Cleanup and Disposal
12 Information Sources

Searching with PubChem Sketcher

- PubChem Sketcher is a web-based chemical structure sketching tool
- Allows users to search a manually drawn or modified structure
- Supports CID, SMILES, SMARTS, InChI, Molecular Formula, and select uploaded file formats
- More information at PubChem Sketcher Help

Searching with PubChem Sketcher (1/10)

Go to the PubChem homepage . You can search by structure from the start by clicking on Draw Structure:



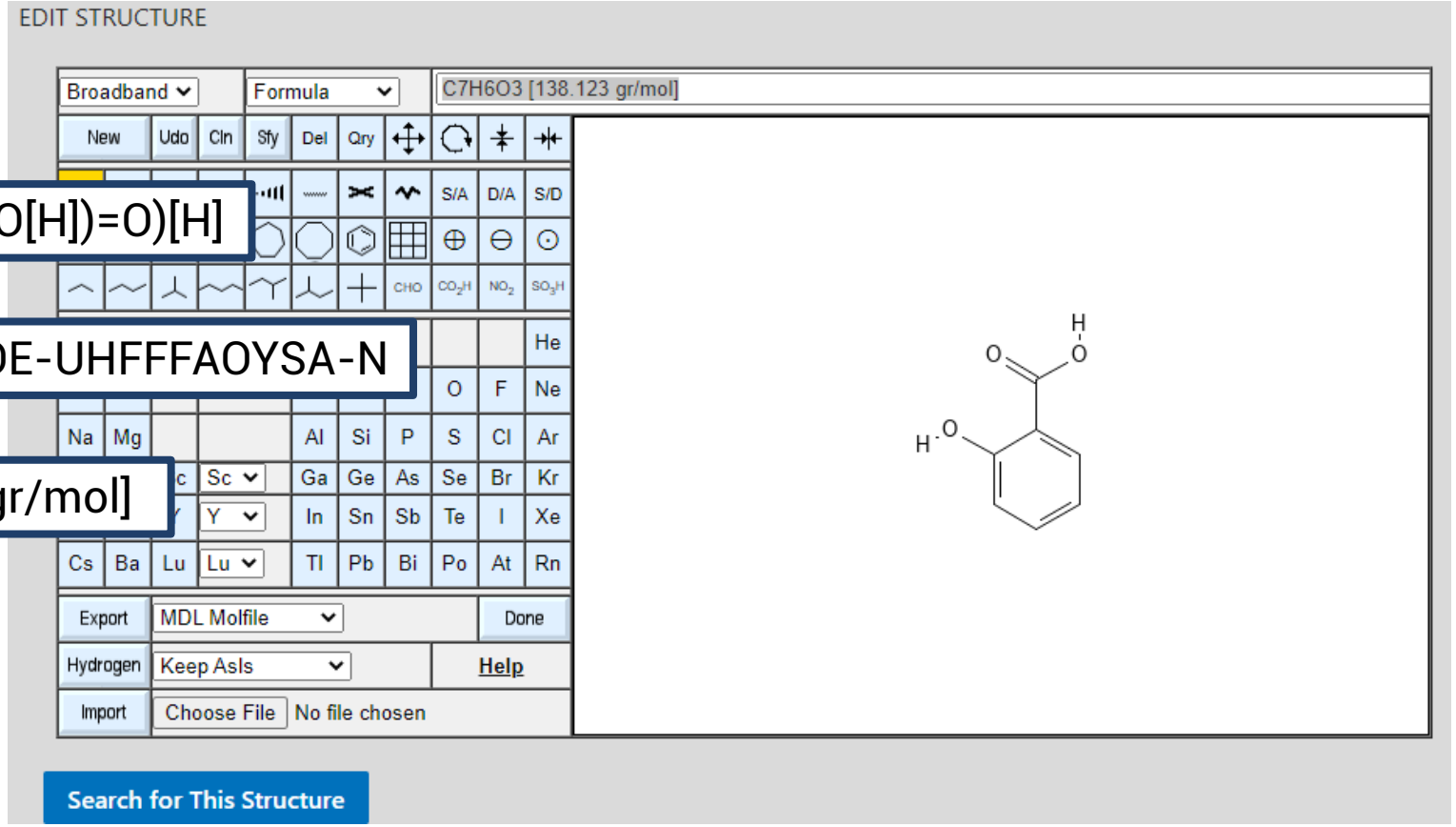
Searching with PubChem Sketcher (2/10)

And manually drawing a structure of interest or inputting SMILES, SMARTS, InChI, and InChiKey information:

SMILES O(C1=CC=CC=C1C(O[H])=O)[H]

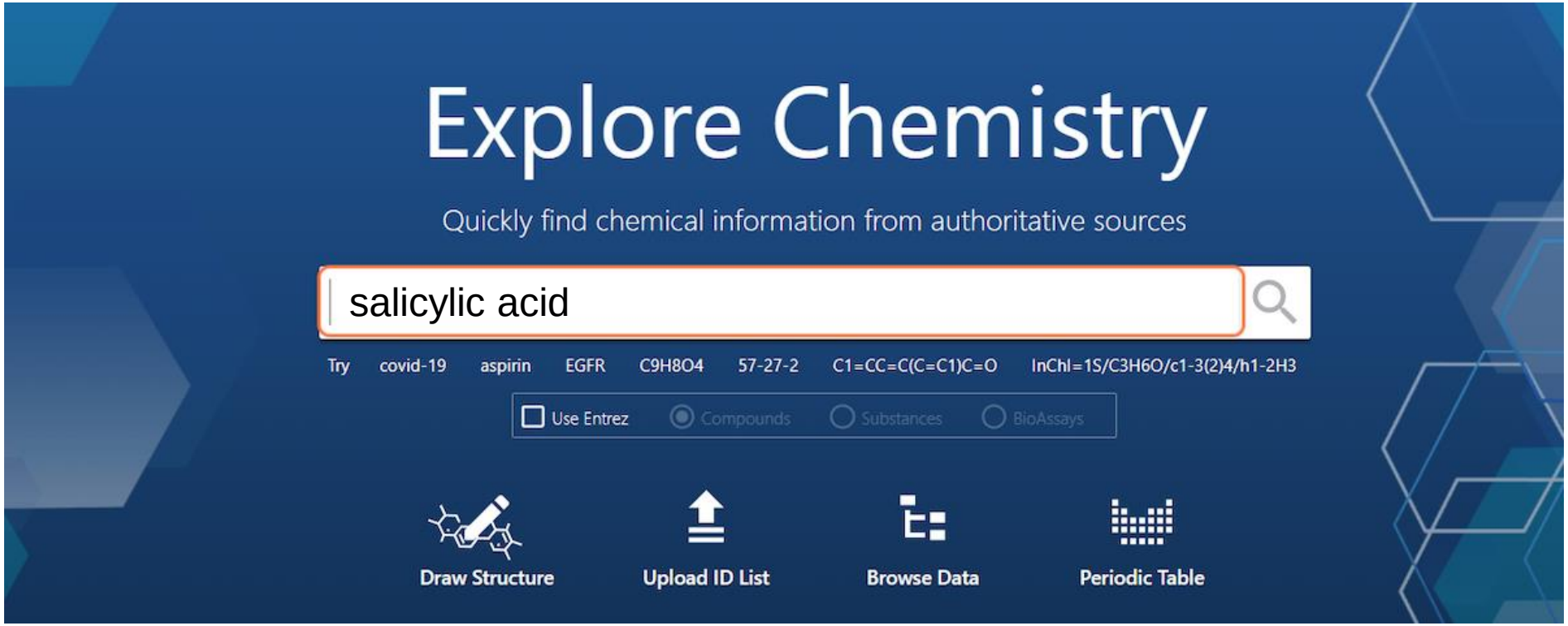
InChIKey=YGSDEFSMJLZEOE-UHFFFAOYSA-N

Formula C7H6O3 [138.123 gr/mol]



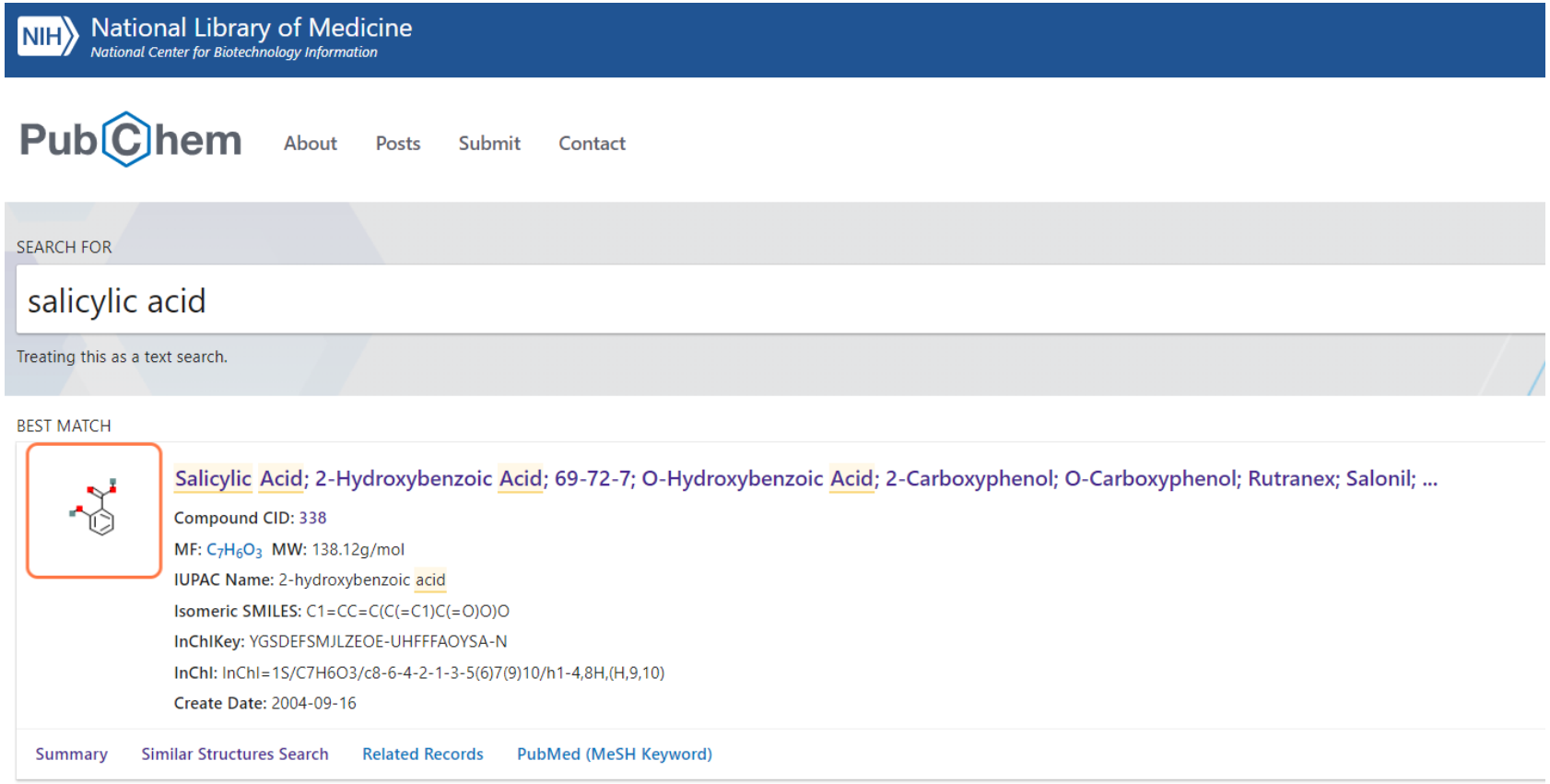
Searching with PubChem Sketcher (3/10)

Or you can find a similar structure during a text query. Here we will type "salicylic acid" into input:



Searching with PubChem Sketcher (4/10)

And click on the Structure image:

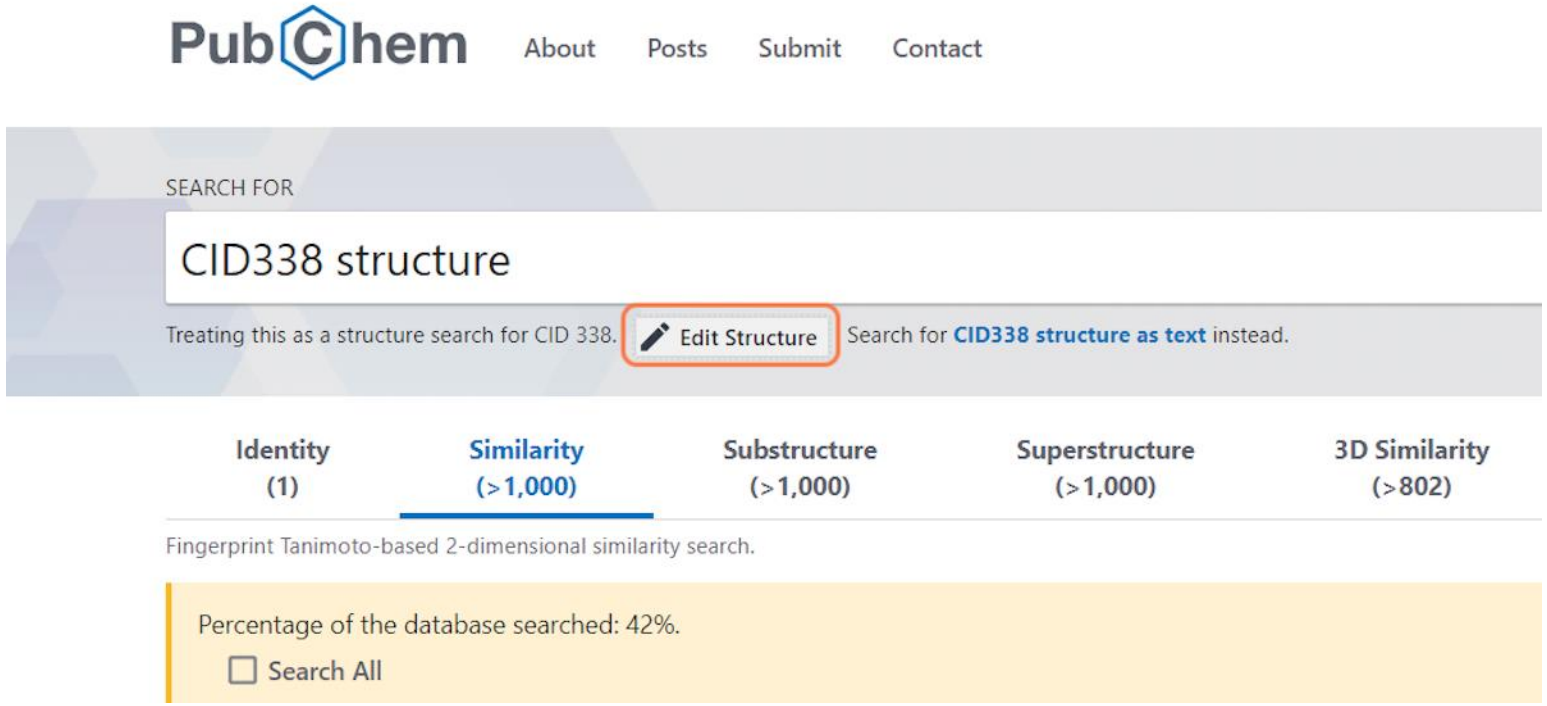


Searching with PubChem Sketcher (5/10)

Select Find Similar Structures:

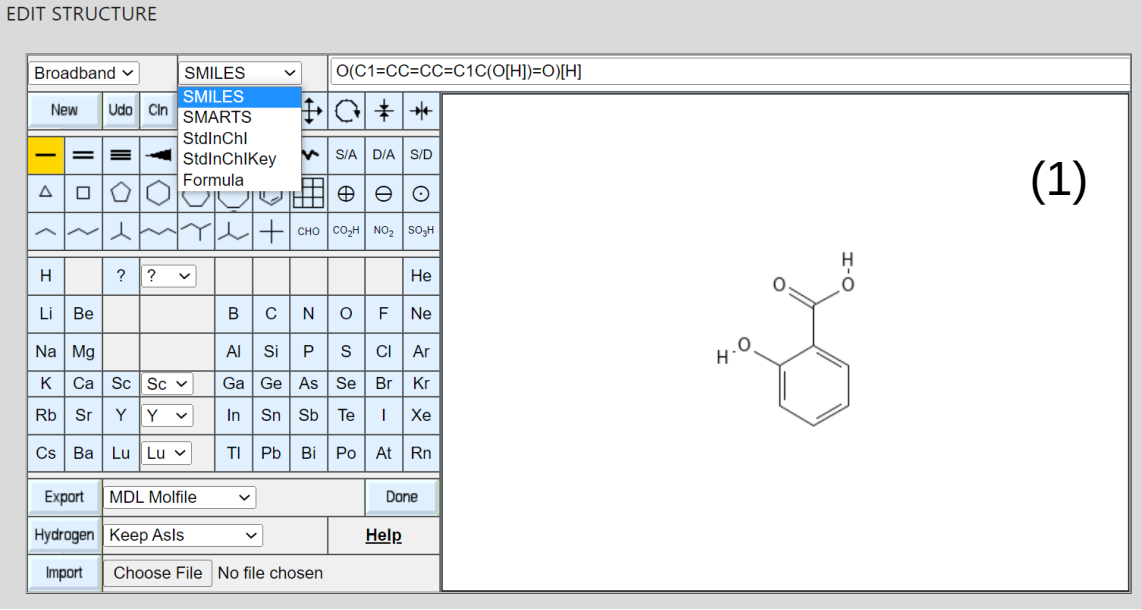
Searching with PubChem Sketcher (6/10)

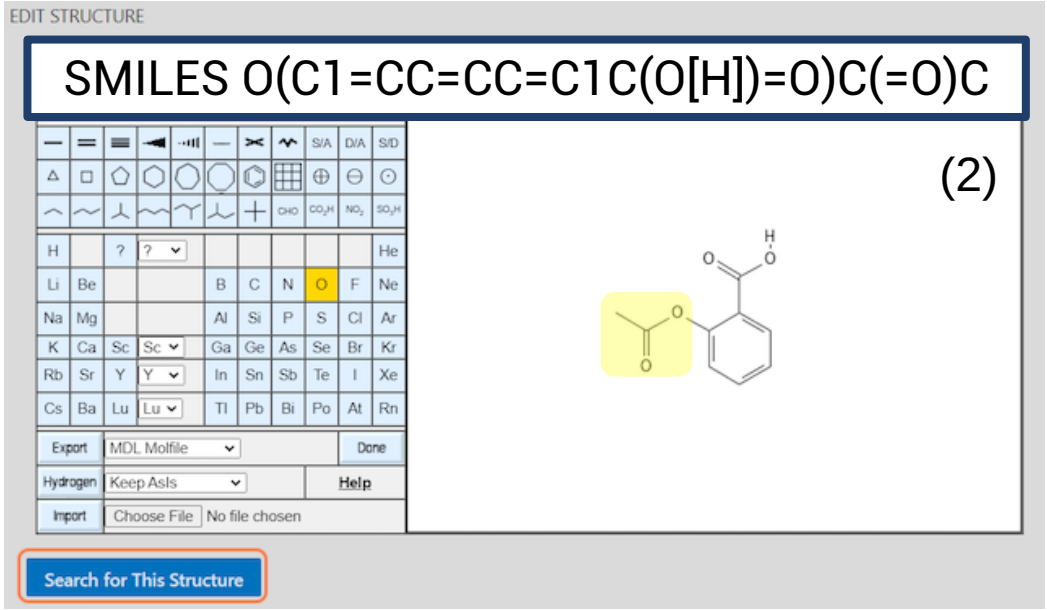
And select Edit Structure to make any changes:



Searching with PubChem Sketcher (7/10)

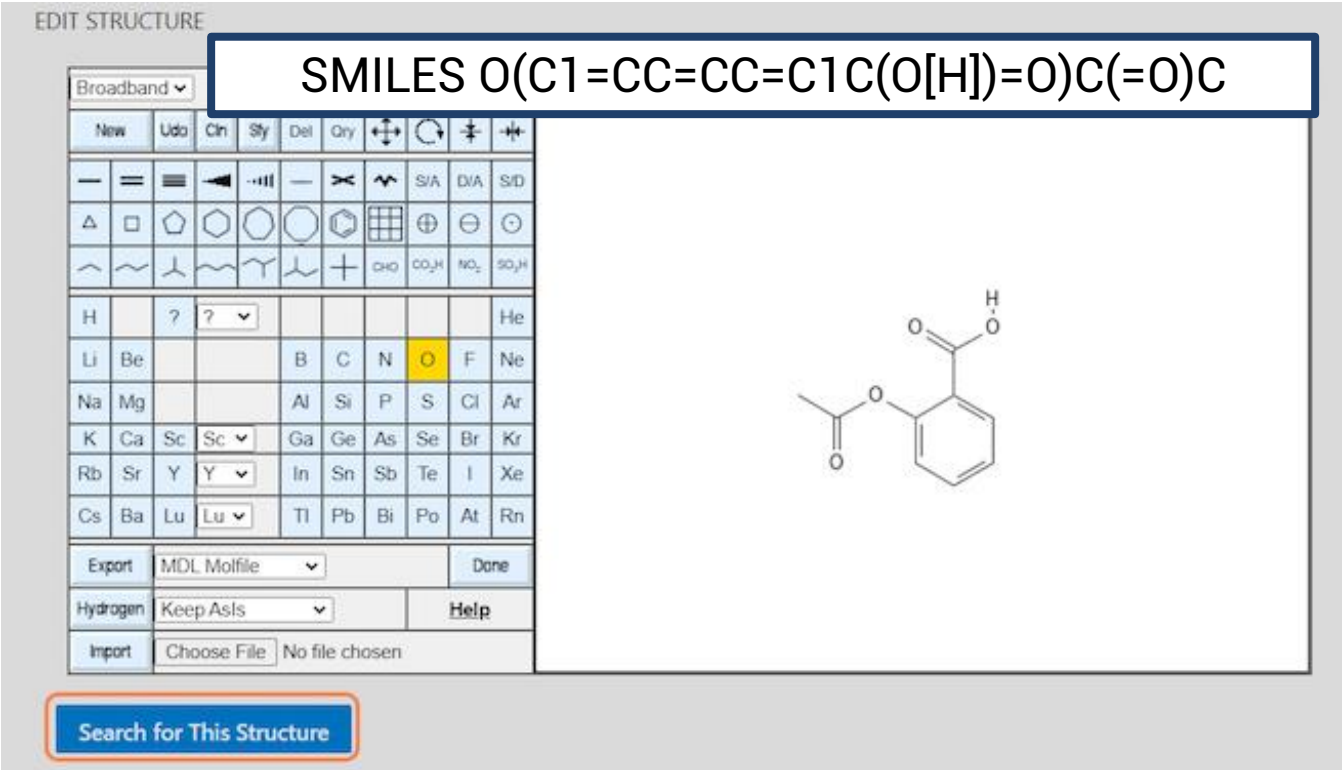
Here, you can replace the phenol (1) with a carboxyl manually (2) or change the SMILES String to O(C1=CC=CC=C1C(O[H])=O)C(=O)C:





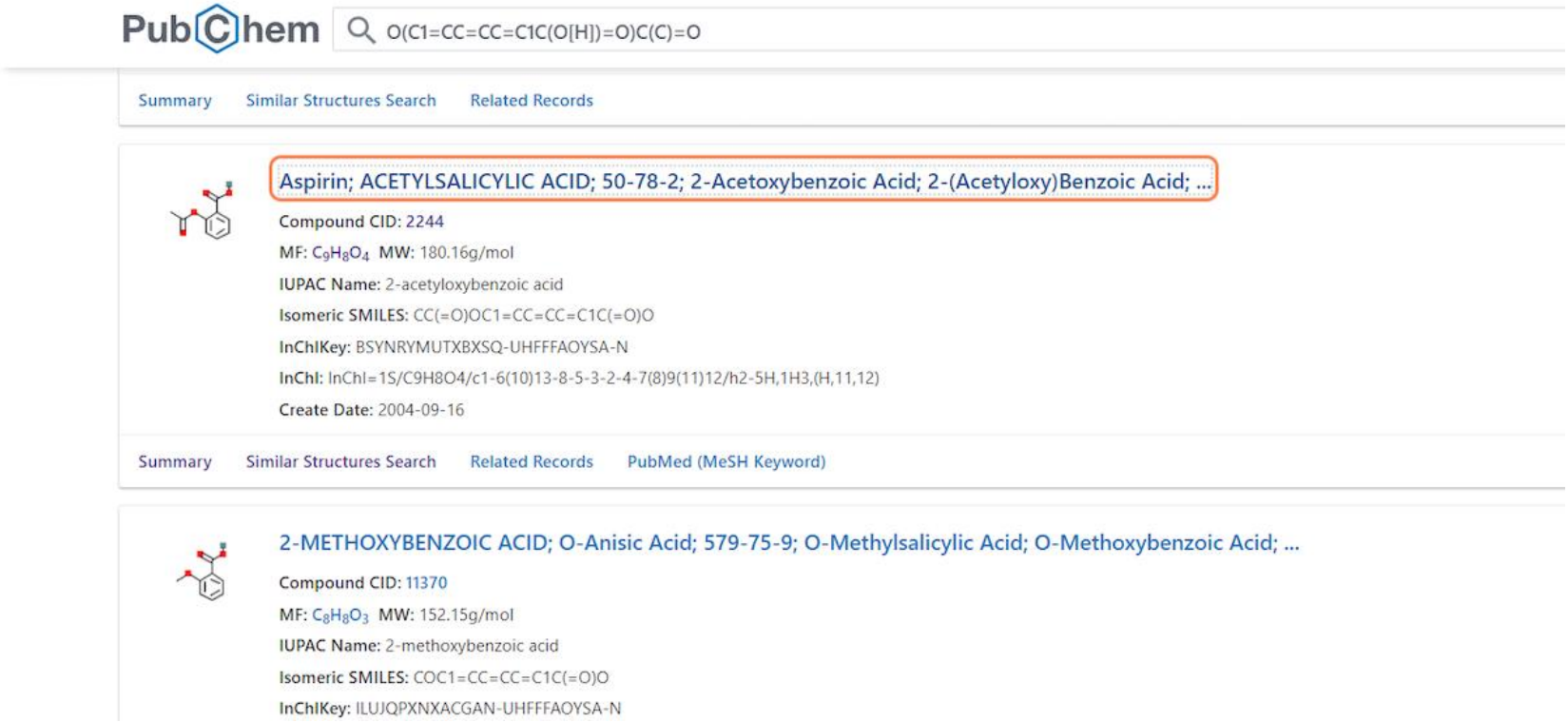
Searching with PubChem Sketcher (8/10)

And click on Search for This Structure to see how this effects the PubChem results:



Searching with PubChem Sketcher (9/10)

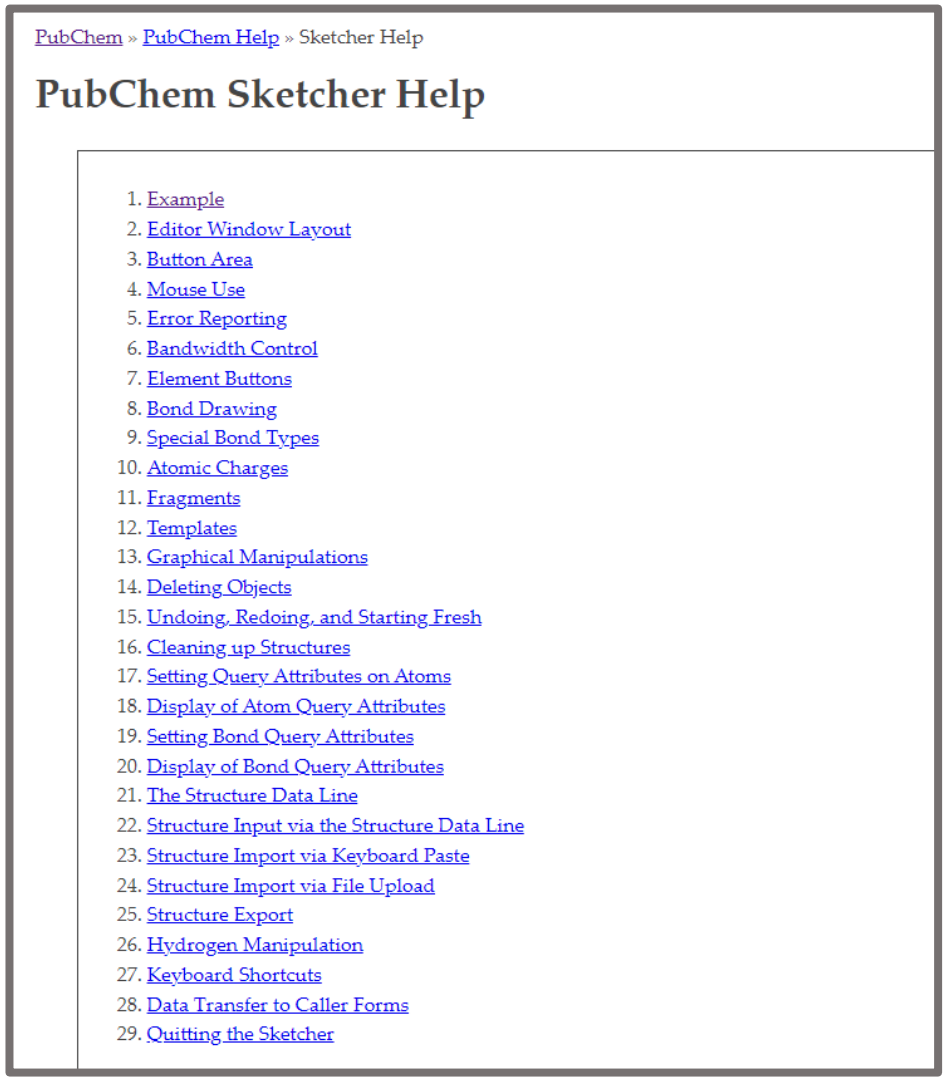
You will find a handful of results, one that may be of interest is Aspirin; ACETYLSALICYLIC ACID. You can open this entry to learn more:



Searching with PubChem Sketcher (10/10)

More information about the PubChem Sketcher can be found here:

5-minute break!


PubChem » PubChem Help » Sketcher Help

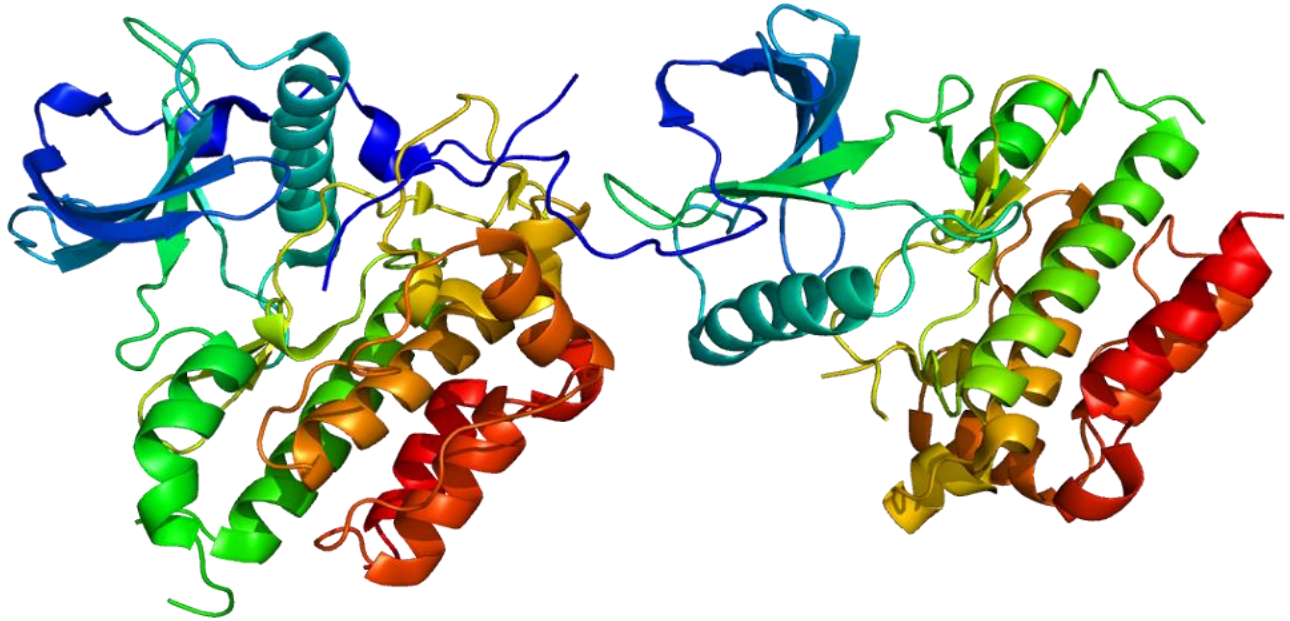
PubChem Sketcher Help

1. Example
2. Editor Window Layout
3. Button Area
4. Mouse Use
5. Error Reporting
6. Bandwidth Control
7. Element Buttons
8. Bond Drawing
9. Special Bond Types
10. Atomic Charges
11. Fragments
12. Templates
13. Graphical Manipulations
14. Deleting Objects
15. Undoing, Redoing, and Starting Fresh
16. Cleaning up Structures
17. Setting Query Attributes on Atoms
18. Display of Atom Query Attributes
19. Setting Bond Query Attributes
20. Display of Bond Query Attributes
21. The Structure Data Line
22. Structure Input via the Structure Data Line
23. Structure Import via Keyboard Paste
24. Structure Import via File Upload
25. Structure Export
26. Hydrogen Manipulation
27. Keyboard Shortcuts
28. Data Transfer to Caller Forms
29. Quitting the Sketcher

Finding a potential inhibitor (Background)

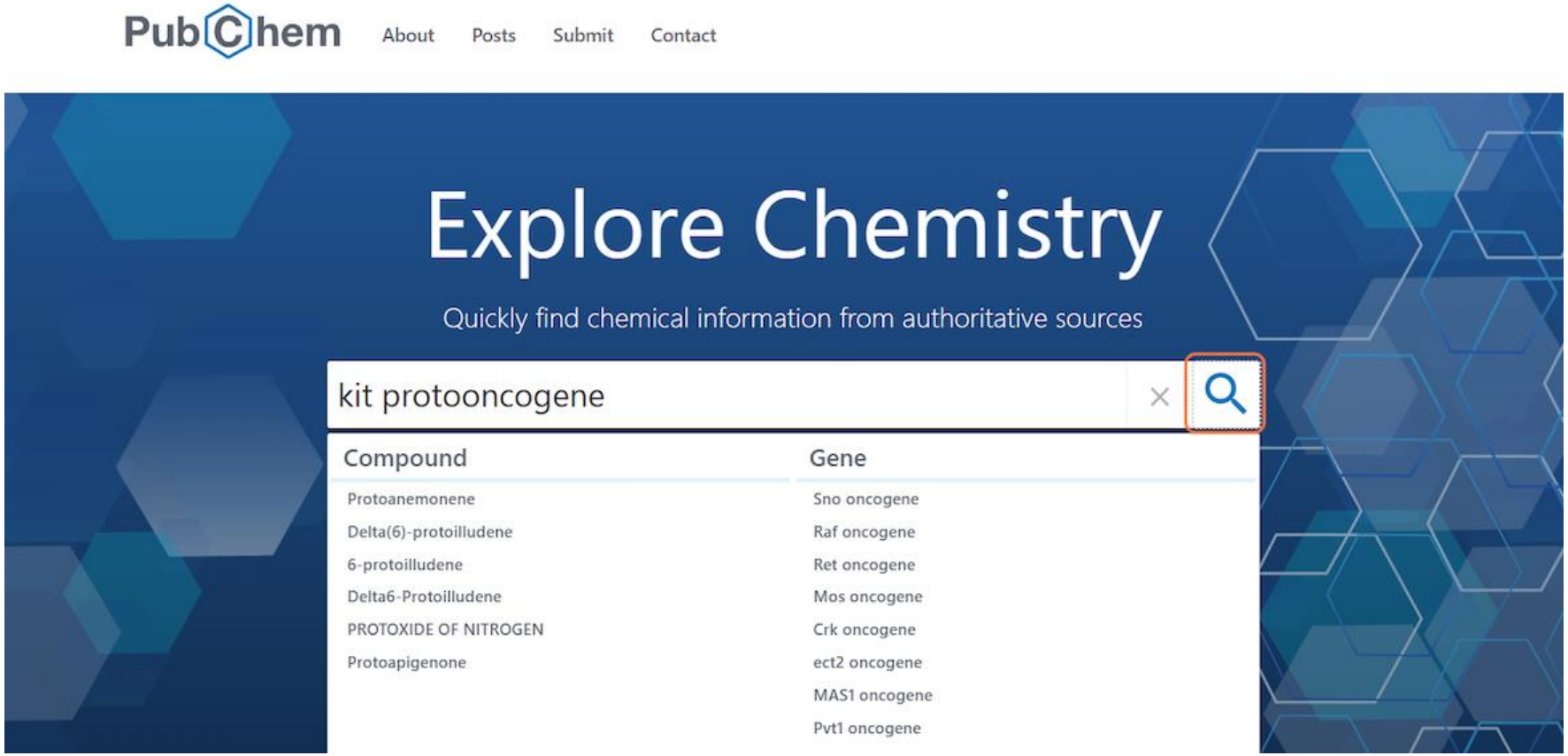
- Increased gene expression leads to disease
- Research on KIT inhibitors can help drug discovery

The KIT protooncogene encodes for a receptor tyrosine kinase protein



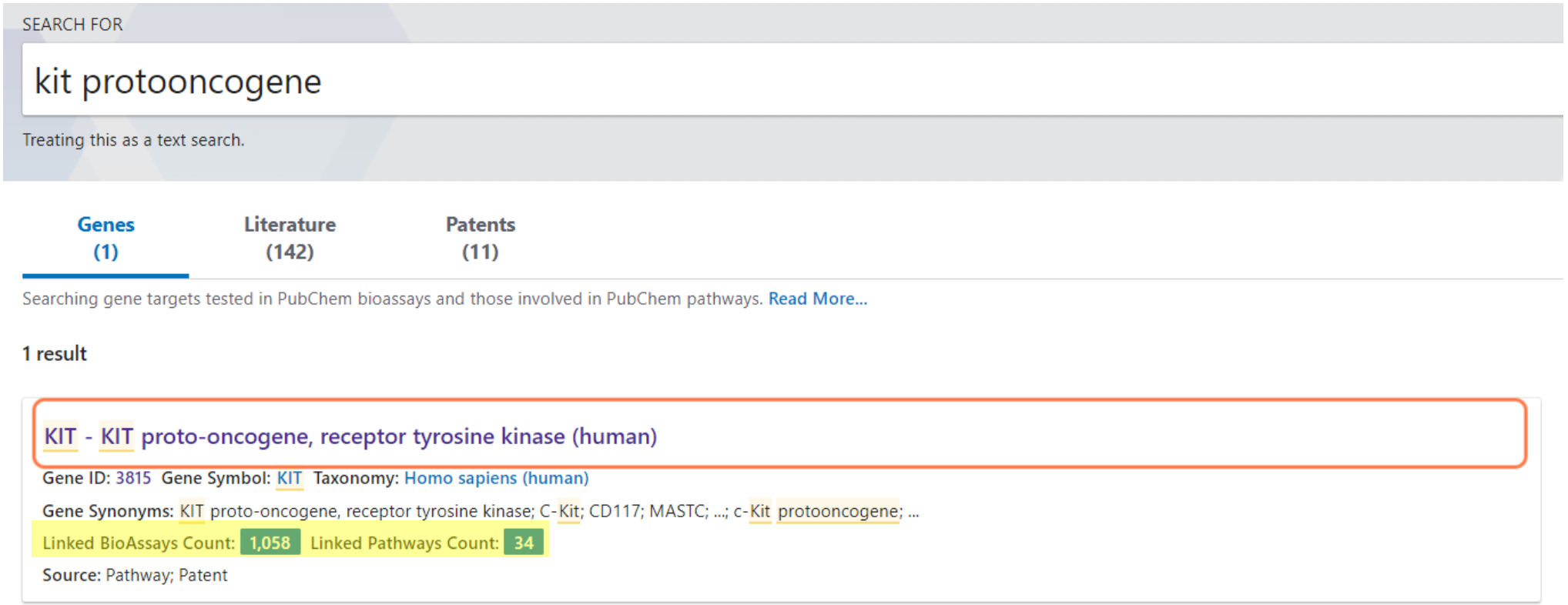
Finding a potential inhibitor (1/11)

Go to the PubChem homepage and search for "kit protooncogene":



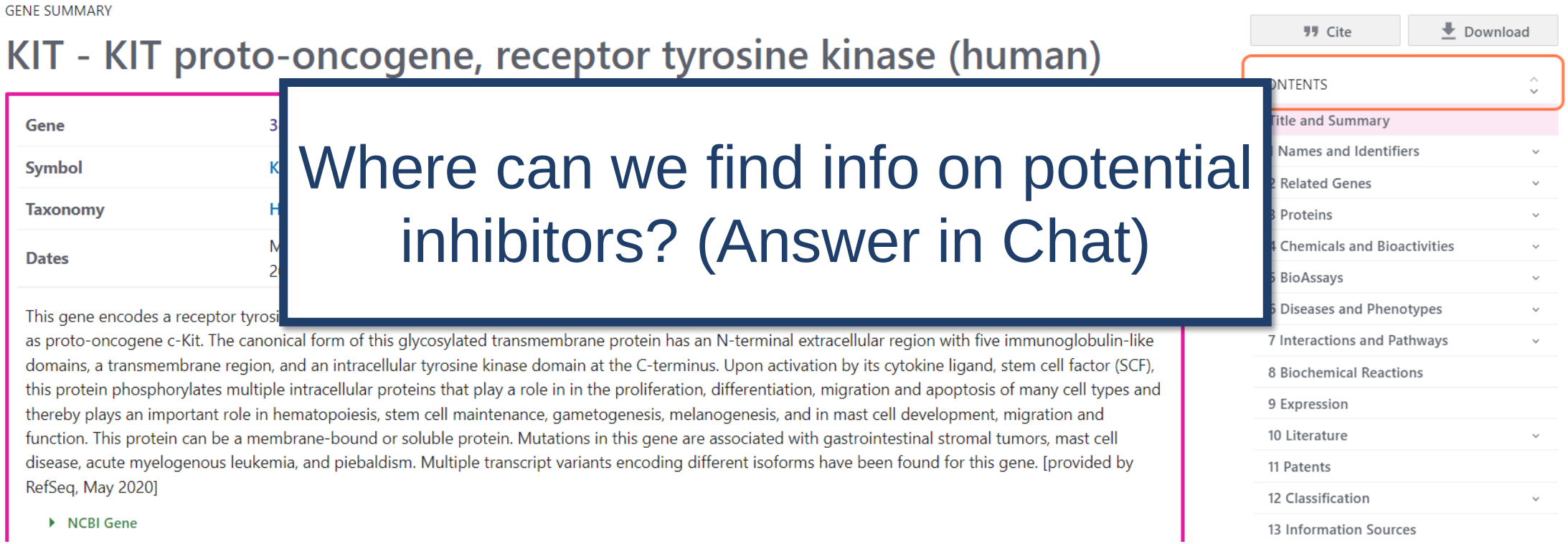
Finding a potential inhibitor (2/11)

Click on the result, which will direct you to the gene summary page:



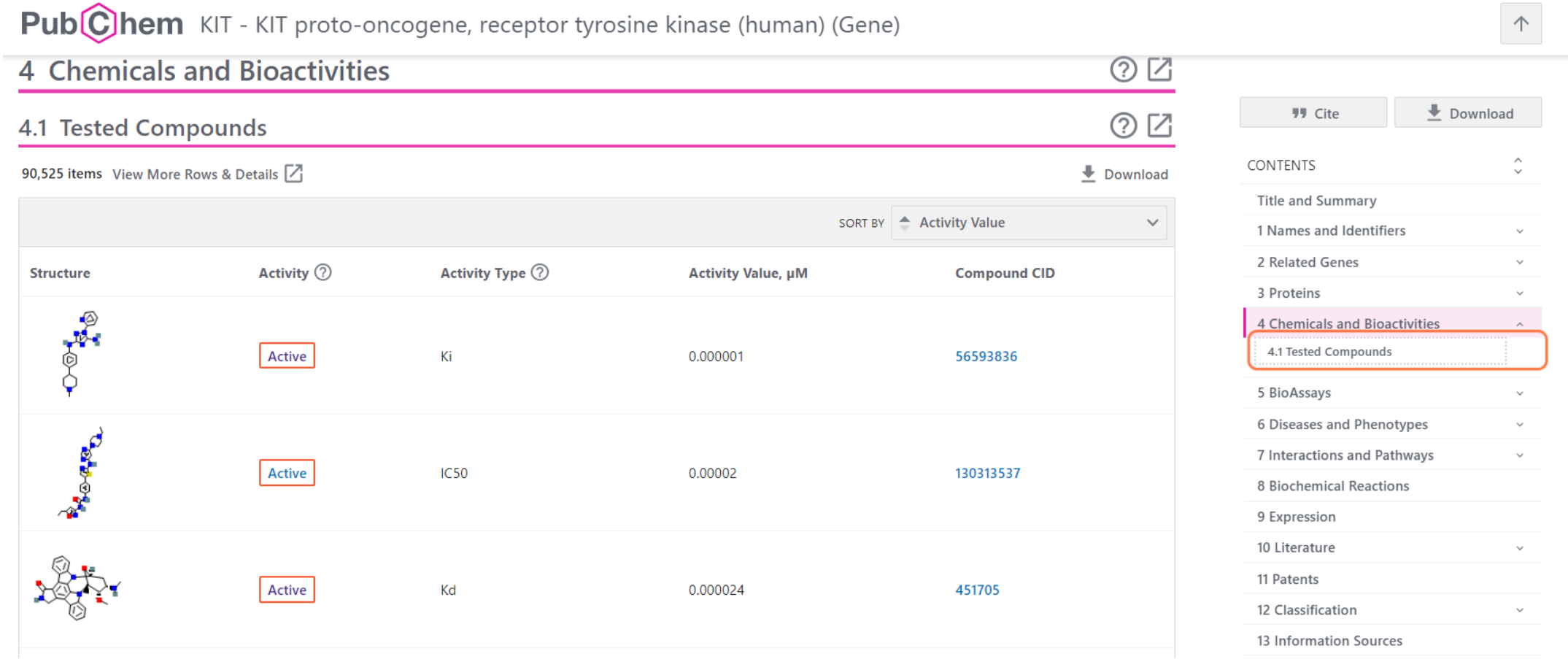
Finding a potential inhibitor (3/11)

Review fundamental information about the KIT protooncogene and navigate to the Contents table:



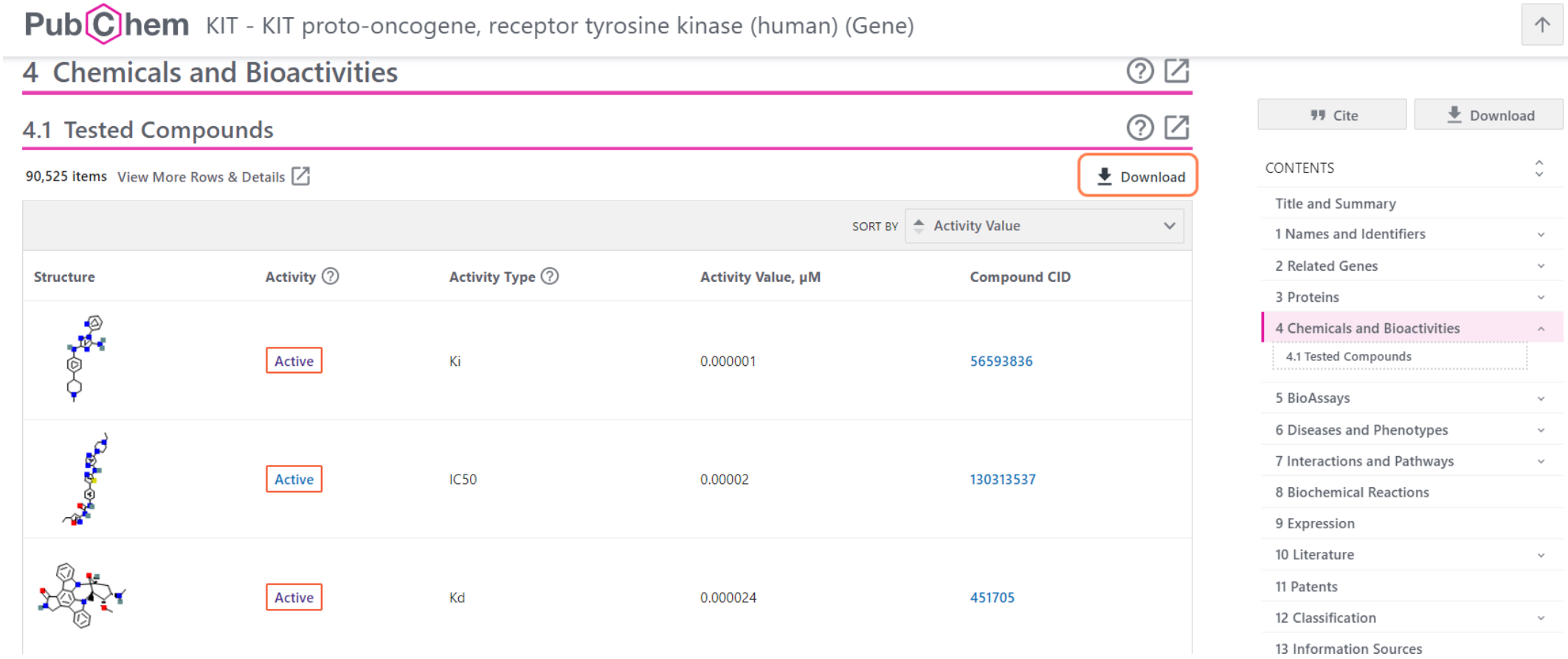
Finding a potential inhibitor (4/11)

Click on 4 Chemicals and Bioactivites and 4.1 Tested Compounds:



Finding a potential inhibitor (5/11)

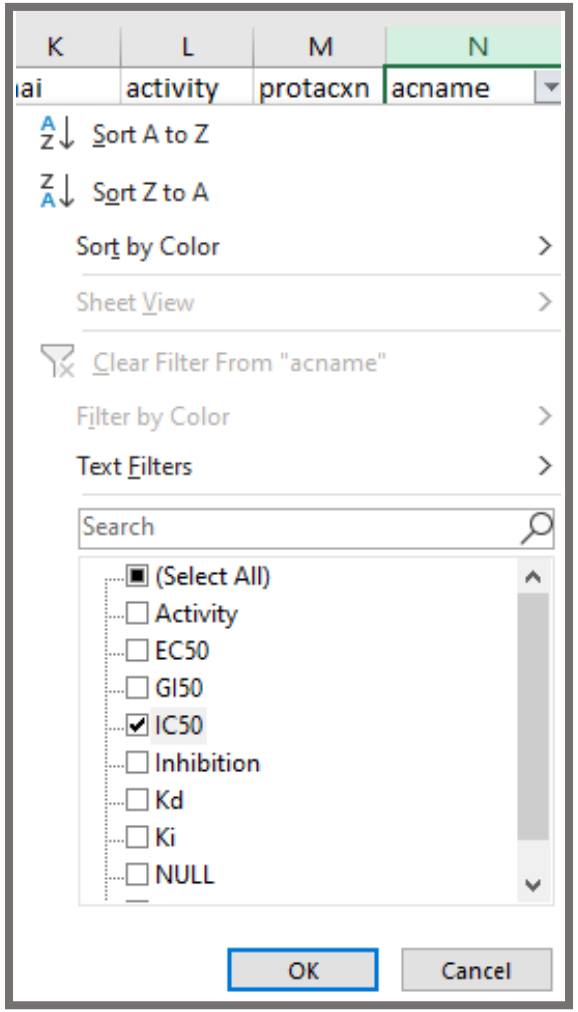
Click Download to download the bioactivity data in CSV format:



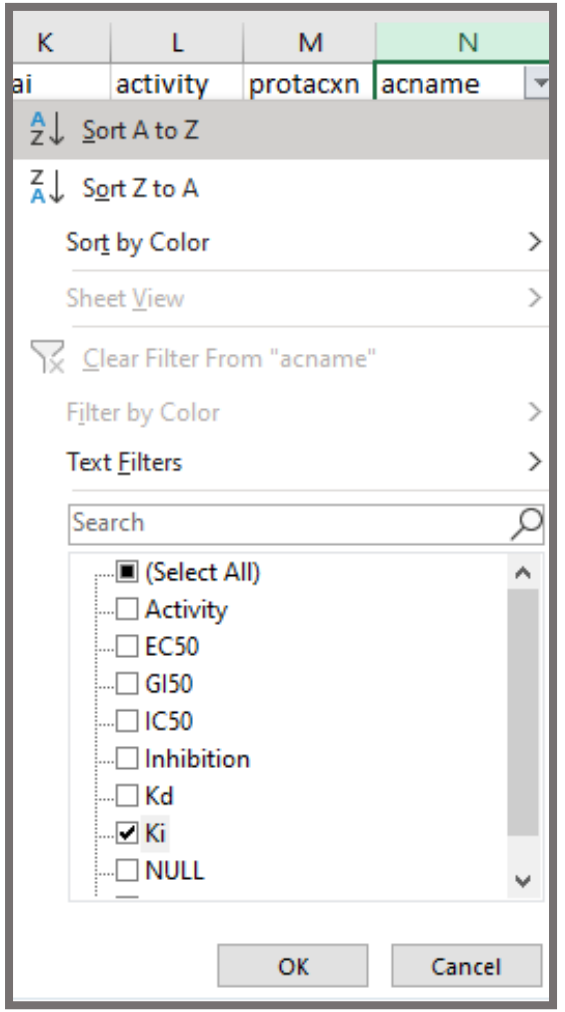
Finding a potential inhibitor (6/11)

Open the CSV file:

- The CSV file contains more detailed information
- You can filter or order the dataset based on measures of inhibition (K_i or IC50 values) or substrate binding (K_D)

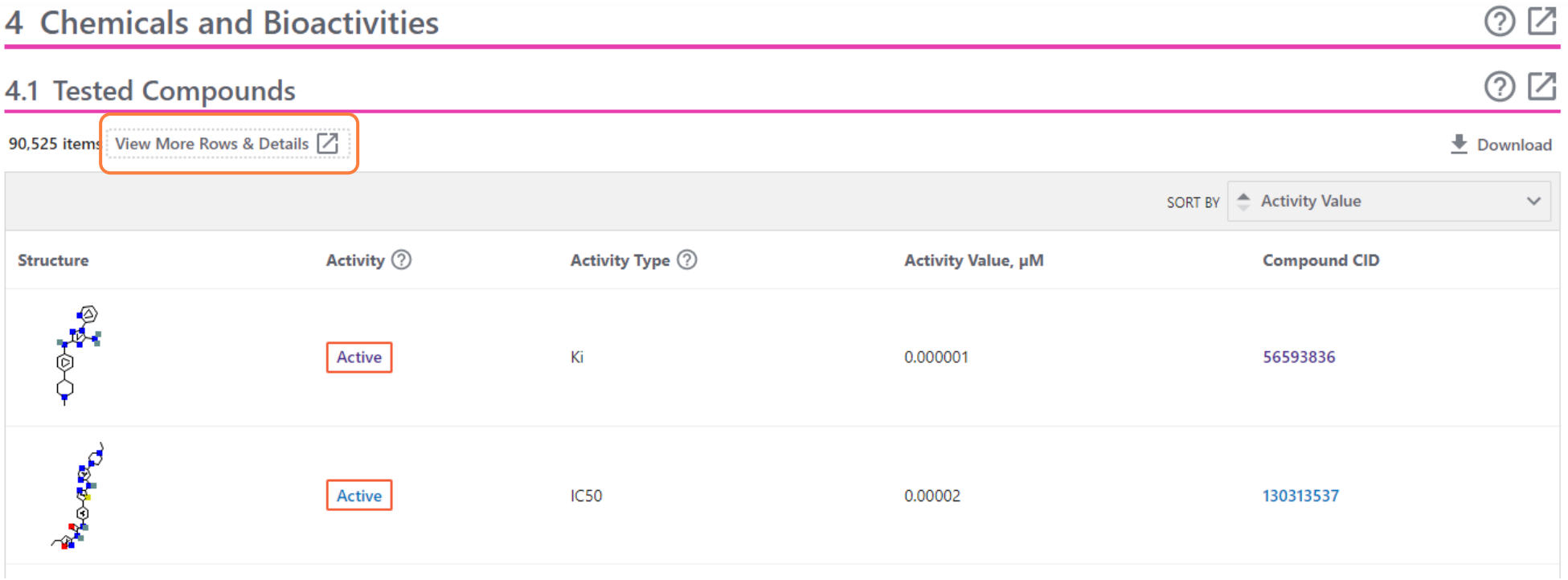


OR



Finding a potential inhibitor (7/11)

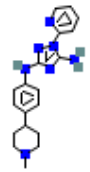
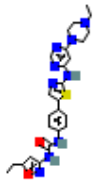
Click View More Rows & Details for more information without downloading a file:



Finding a potential inhibitor (8/11)

Sort Structures by Activity Value

SORT BY Activity Value

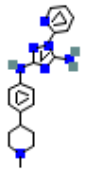
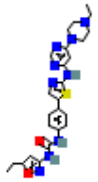
#	Structure	Activity	Activity Type	Activity Value, µM	Compound CID	Substance SID	BioAssay AID	BioAssay Name
1		Active	Ki				46	Displacement of [33P]ATP from human recombinant c-KIT domain after 20 mins by scintillation counting
2		Active	IC				63	Inhibition of human c-KIT A loop exon 17 D820Y single mutant using poly (Glu,Tyr) 4:1 as substrate in presence of 33P-gamma-ATP by hotspot kinase assay
3		Active	Kd	0.000024	451705	50100119	1433	Kinase Inhibitor Selectivity Profiling Assay

Find a compound to learn more about!

Finding a potential inhibitor (9/11)

Click on a relevant Compound CID for more information, including how to buy the compound:

SORT BY Activity Value

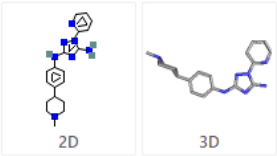
#	Structure	Activity	Activity Type	Activity Value, µM	Compound CID	Substance SID	BioAssay AID	BioAssay Name
1		Active	Ki	0.000001	56593836	134461981	623446	Displacement of [33P]ATP from human recombinant c-KIT domain after 20 mins by scintillation counting
2		Active	IC50	0.00002	130313537	440123730	1587963	Inhibition of human c-KIT A loop exon 17 D820Y single mutant using poly (Glu,Tyr) 4:1 as substrate in presence of 33P-gamma-ATP by hotspot kinase assay

Finding a potential inhibitor (10/11)

Click on 5 Chemical Vendors:

COMPOUND SUMMARY

3-N-[4-(1-methylpiperidin-4-yl)phenyl]-1-pyridin-2-yl-1,2,4-triazole-3,5-diamine

PubChem CID	56593836
Structure	 2D 3D Find Similar Structures
Molecular Formula	$C_{19}H_{23}N_7$
Synonyms	CHEMBL1835867 SCHEMBL4038099 BDBM50355489 ZINC72127621
Molecular Weight	349.4
Dates	Modify 2022-11-05 Create 2012-02-06

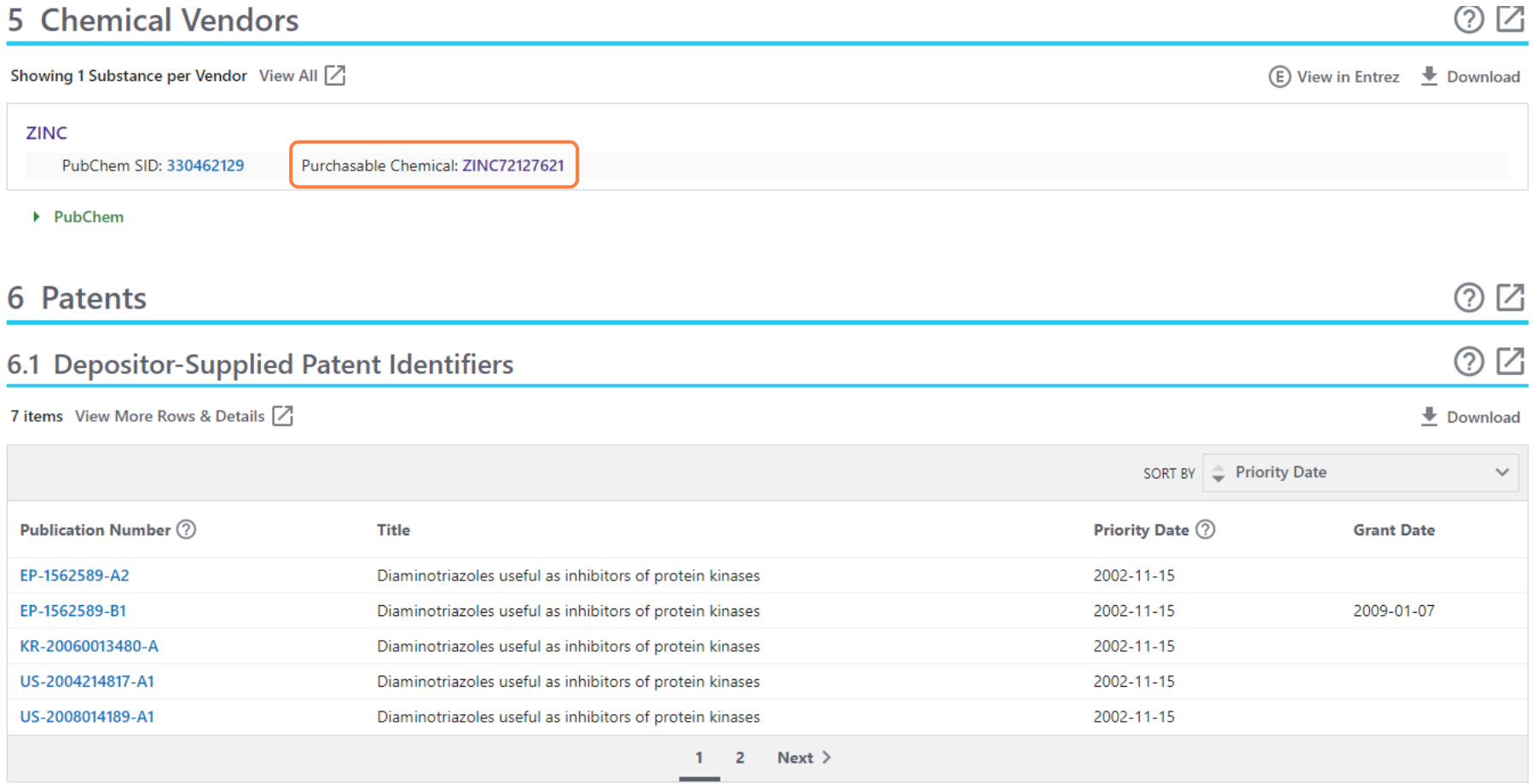
Cite Download

CONTENTS

- Title and Summary
- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Related Records
- 5 Chemical Vendors
- 6 Patents
- 7 Biological Test Results
- 8 Classification
- 9 Information Sources

Finding a potential inhibitor (11/11)

Click on Purchasable Chemical ID for more information:



Conclusion

- Background ✓
- Searching in PubChem ✓
- Finding molecular & chemical safety information ✓
- Using PubChem Sketcher ✓
- Finding a potential inhibitor ✓

POLL (4/4)

What aspects of PubChem would you like to learn more about in the future?

Continue learning about PubChem

- PubChem Tutorials and a training course
- Visit the PubChem Help Guide
- Use the Contact PubChem page
- Browse the PubChem Publications page for journal articles about PubChem

Continue learning about NCBI Resources

- Join us for workshops, webinars, or codeathons!

  NCBI Insights Blog

- Follow us on social media:
  - Twitter
  - LinkedIn
  - Facebook