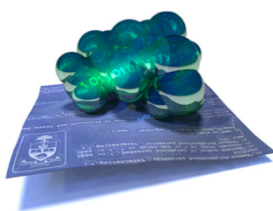


A
BIOINFORMATICS
COURSE

SMALL MOLECULE STRUCTURE



BORIS STEIPE

*DEPARTMENT OF BIOCHEMISTRY – DEPARTMENT OF MOLECULAR GENETICS
UNIVERSITY OF TORONTO*

SOURCES FOR SMALL MOLECULE STRUCTURE

Many structures of chemical compounds are freely available:

NCI database browser: search by CAS number, Formula

~250,000 structures

CSD (Cambridge Structural Database) commercial, ~400,000 structures

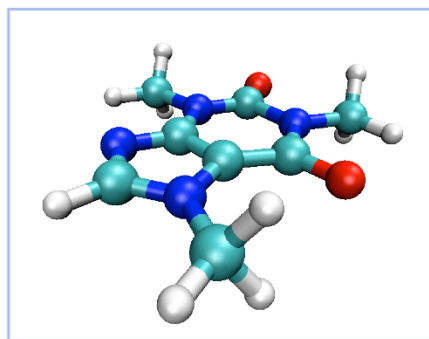
CHEBI (at the EBI) public

PDB

PubChem (at the NCBI) public

DrugBank

... or build your own model.



Rich and varied resources are available for “Cheminformatics”, and they are very well interfaced with support for exploring the structures of small molecules, and the context of their complexes with macromolecules.

"HETERO" COMPOUNDS IN PDB FILES

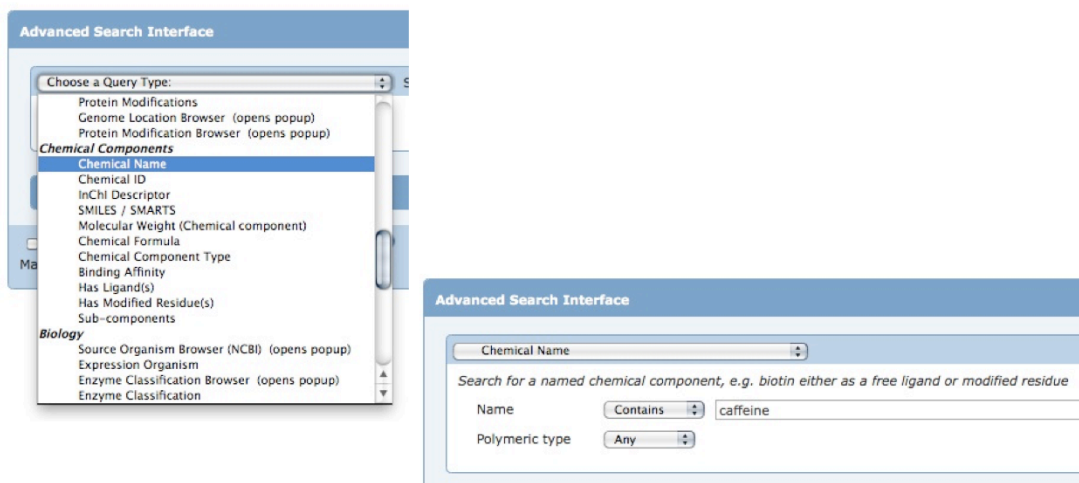
... solvent, substrates, inhibitors, products, ligands, cofactors, prosthetic groups ...

```
[...]  
HETATM 877 O HOH 1 -4.169 60.050 40.145 1.00 3.00 2IMM 937  
[...]
```

"Hetero" compounds are treated separately from protein and DNA chains and stored in the HETATM records of PDB files.

PDB maintains their own database of hetero-compounds cocrystallized in PDB structures – the [Chemical Component Dictionary](#).

Access via the Advanced Search Interface ...



OR ...

Access via the Ligand or Drug Search Window...

Search by Ligand
 Search ligands bound to molecules in the PDB by:
 • Identifier or Name
 • SMILES String, InChI
 • Chemical Formula or Weight
 • Chemical Substructure

Search by Drugs & Drug Targets
 Drugs & Drug Targets in the PDB have been mapped to DrugBank.

Drug-Target Complex: Atonvastatin bound to its target HMG-CoA reductase
 Drug: Atonvastatin (Lipitor)

Chemical component search

Drug and Drug Target Mapping

Search by Generic or Brand Names

Generic Name	Brand Name	DrugBank ID	ATC Code	Ligand ID	Target Name	UniProt ID
3,4-Methylenedioxyamphetamine		DB01454	B01	B01	Synaptic vesicle amine transporter	Q22945
3,4-Methylenedioxyamphetamine		DB01454	B01	B01	Sodium-dependent neurotransmitter transporter	P28975
3,4-Methylenedioxyamphetamine		DB01454	B01	B01	Sodium-dependent neurotransmitter transporter	P23365
4-Aminobiphenyl		DB01528	AD03	AD03	2 beta-Hydroxyprostanoid synthase (cyclooxygenase-2)	P14600

OR ...

PDB CHEMICAL COMPONENTS DICTIONARY

Access via PDBeChem (at the EBI) or Ligand Expo (at the PDB).

The image displays two side-by-side screenshots of web interfaces for searching chemical components in the Protein Data Bank (PDB).

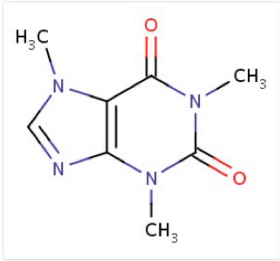

Left Screenshot: PDBeChem (EBI)
The interface is titled "Chemical Components in the PDB" and is part of the "Protein Data Bank in Europe" (PDBe). It features a navigation menu on the left with options like "General Information", "How to Use", "Overview", "Search By", "Useful Links", and "Latest Releases". The main content area is titled "PDBeChem: Search for a chemical component" and includes a search form with fields for "Code", "Molecule Name", "Formula", "Non-Stereo SMILES", and "Fragments". There are also "Search" and "Reset" buttons and a "Combine Criteria With" option set to "AND".

Right Screenshot: Ligand Expo (PDB)
The interface is titled "Ligand Expo" and is part of the "Protein Data Bank" (PDB). It features a navigation menu on the left with options like "Home", "Search", "Browse", "Download", and "Ligand Expo Help". The main content area is titled "Chemical Component Search Tools" and includes a search form with fields for "Search term" and "Search type". There are also "Launch" buttons and a "Component identifier (3-letter code)" field.

PDB
CHEMICAL
COMPONENTS
DICTIONARY

RCSB PDB - Ligand Summary Page
DrugBank: Caffeine

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

View / Download Files

CFF

CAFFEINE

CFF is found in [22 entries](#).

CFF as free ligands, exist in [22 entries](#).
Examples include [1C8L](#), [1GFZ](#), [1LSQ](#).

Find related ligands: [Stereoisomers](#) [Similar ligands](#) [Chemical Structure Search](#)

View summary at Ligand Expo

Rotate Hydrogens Labels

Chemical Component Summary	
Name	CAFFEINE
Identifiers	1,3,7-trimethylpurine-2,6-dione
Formula	C ₈ H ₁₀ N ₄ O ₂
Molecular Weight	194.19 g/mol
Type	NON-POLYMER
Isomeric SMILES	Cn1cnc2n(C)c(=O)n(C)c(=O)c12
InChI	InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
InChIKey	RYYVLZVUVIJVGH-UHFFFAOYSA-N

Chemical Details	
Formal Charge	0
Atom Count	24
Chiral Atom Count	0
Chiral Atoms	n/a
Bond Count	25
Aromatic Bond Count	10
Leaving Atoms	n/a

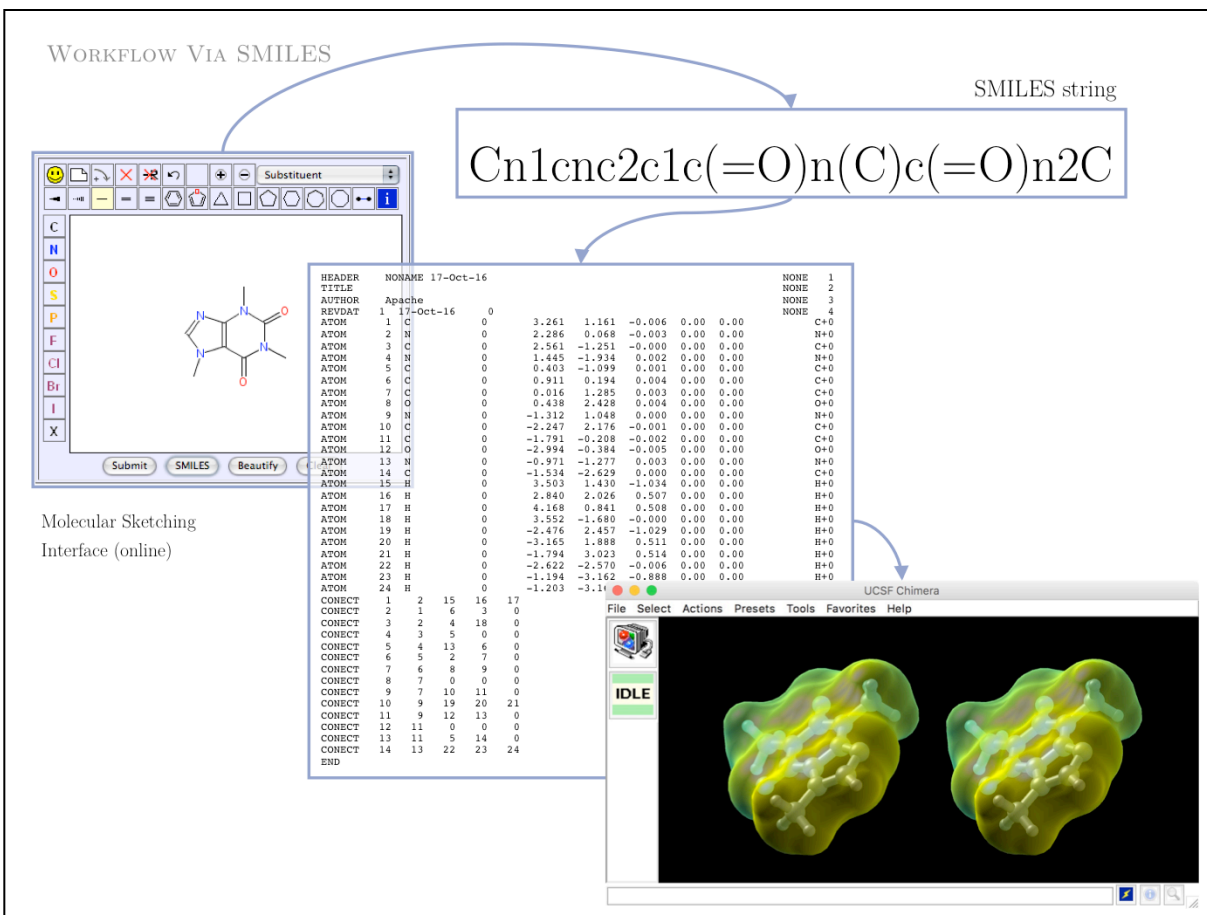
Drug Info: DrugBank

DrugBank ID	DB00201 (Stereoisomeric match)
Name	Caffeine
Grading	approved

Display a menu Contact Us

Caffeine is found in 22 different PDB entries.

The link to <http://www.drugbank.ca> contains pages upon pages of relevant information: such as physical and chemical data including spectra, pharmacology, drug interactions, cross-references to other databases like KEGG, and Wikipedia, and literature references.



Creating small-molecule structures from scratch can easily be done by sketching a molecule with a molecular editor such as the one provided in the PDB search-by-ligand window, saving its SMILES string, and converting that into 3D coordinates directly in Chimera.

<http://steipe.biochemistry.utoronto.ca/abc>

B O R I S . S T E I P E @ U T O R O N T O . C A

DEPARTMENT OF BIOCHEMISTRY & DEPARTMENT OF MOLECULAR GENETICS
UNIVERSITY OF TORONTO, CANADA