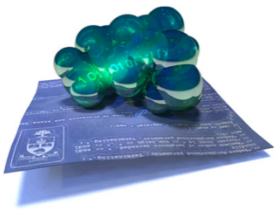
A
Bioinformatics
Course

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

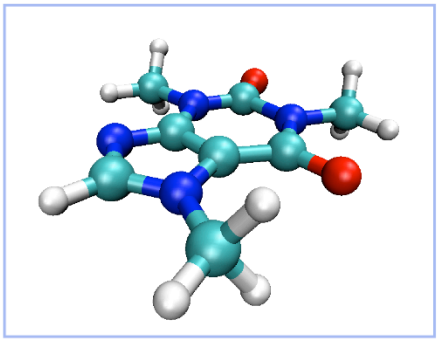
CHEBI (at the EBI) public

PDB

PubChem (at the NCBI) public

DrugBank

... or build your own model.



Richa 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

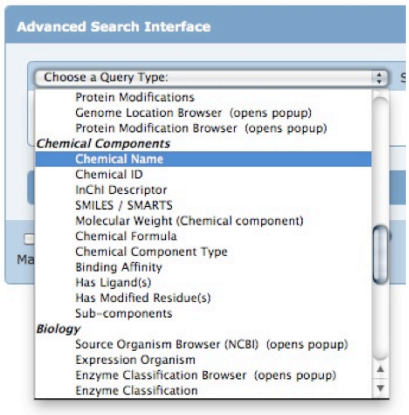
... solvent, substrates, inhibitors, products, ligands, cofactors, prosthetic groups ...

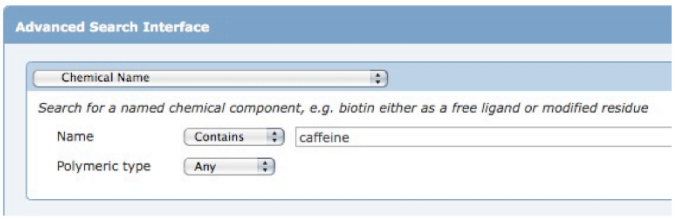
```
[...]
HETATM  877  O   HOH     1      -4.169  60.050  40.145  1.00  3.00      2IMM 937
[...]
```

"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 ...

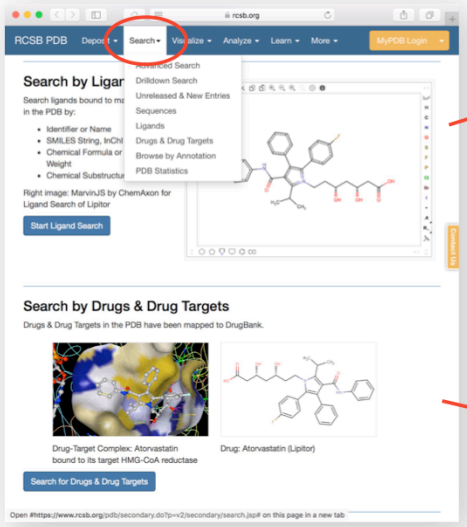


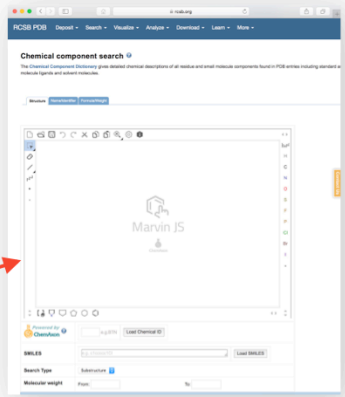


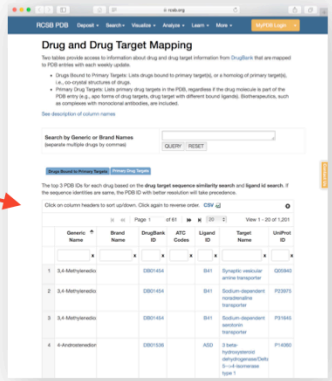
or ...

PDB Chemical Components Dictionary

Access via the Ligand or Drug Search Window...



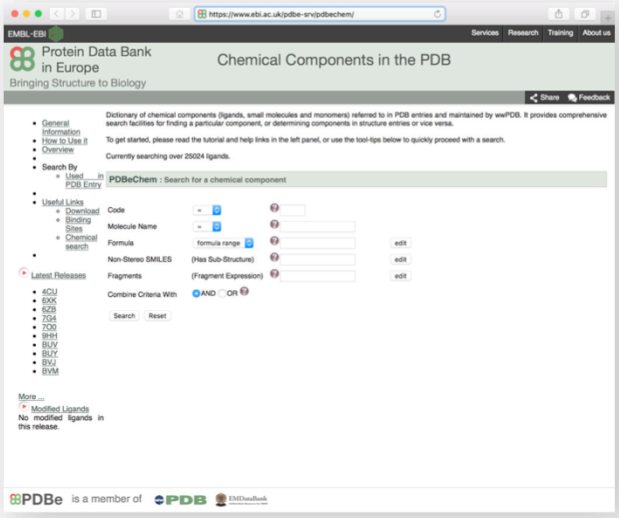


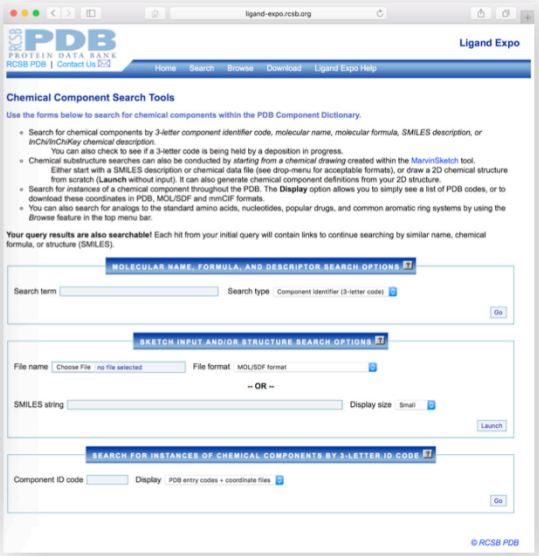


or ...

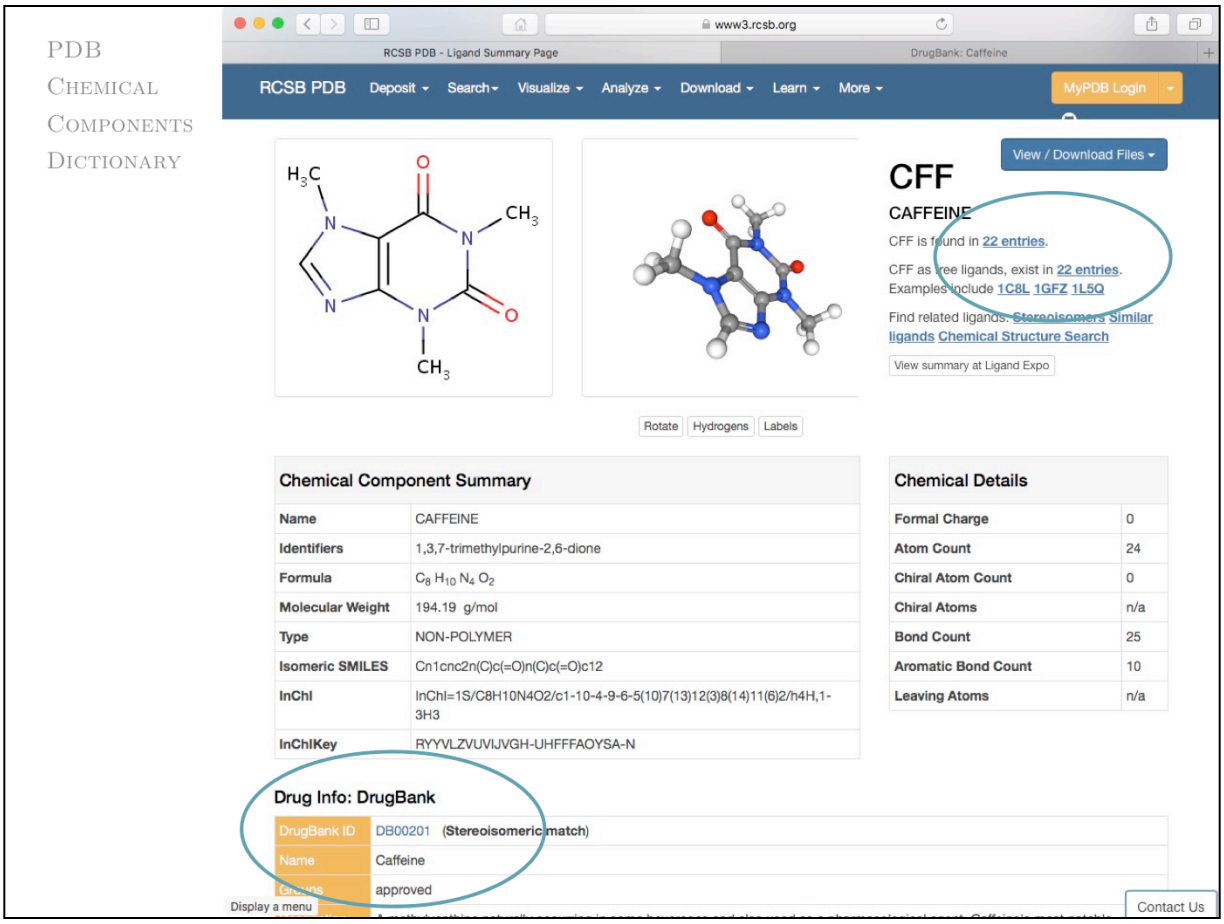
PDB Chemical Components Dictionary

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



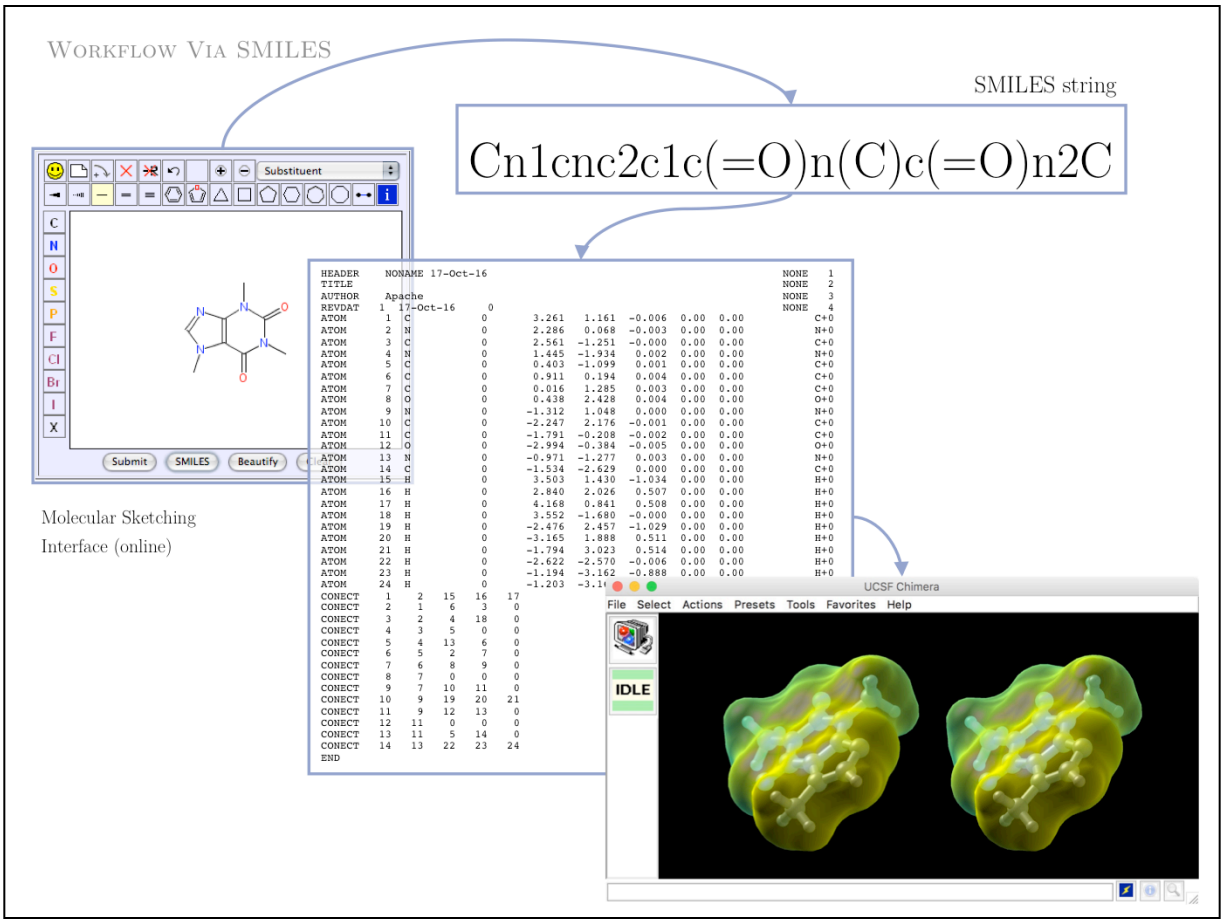


PDB
Chemical
Components
Dictionary



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