Dictionary of ligands

Some of the web and other resources

Small molecules

DrugBank: http://www.drugbank.ca/

ZINC: http://zinc.docking.org/index.shtml

PRODRUG: http://www.compbio.dundee.ac.uk/Web_Servers/prodrg_down.html

CACTVS: http://www2.chemie.uni-erlangen.de/software/cactvs/

Cambridge structural database - CSD: http://www.ccdc.cam.ac.uk/products/csd/

Macromolecules

PDB:

European EBI: http://www.ebi.ac.uk/msd/

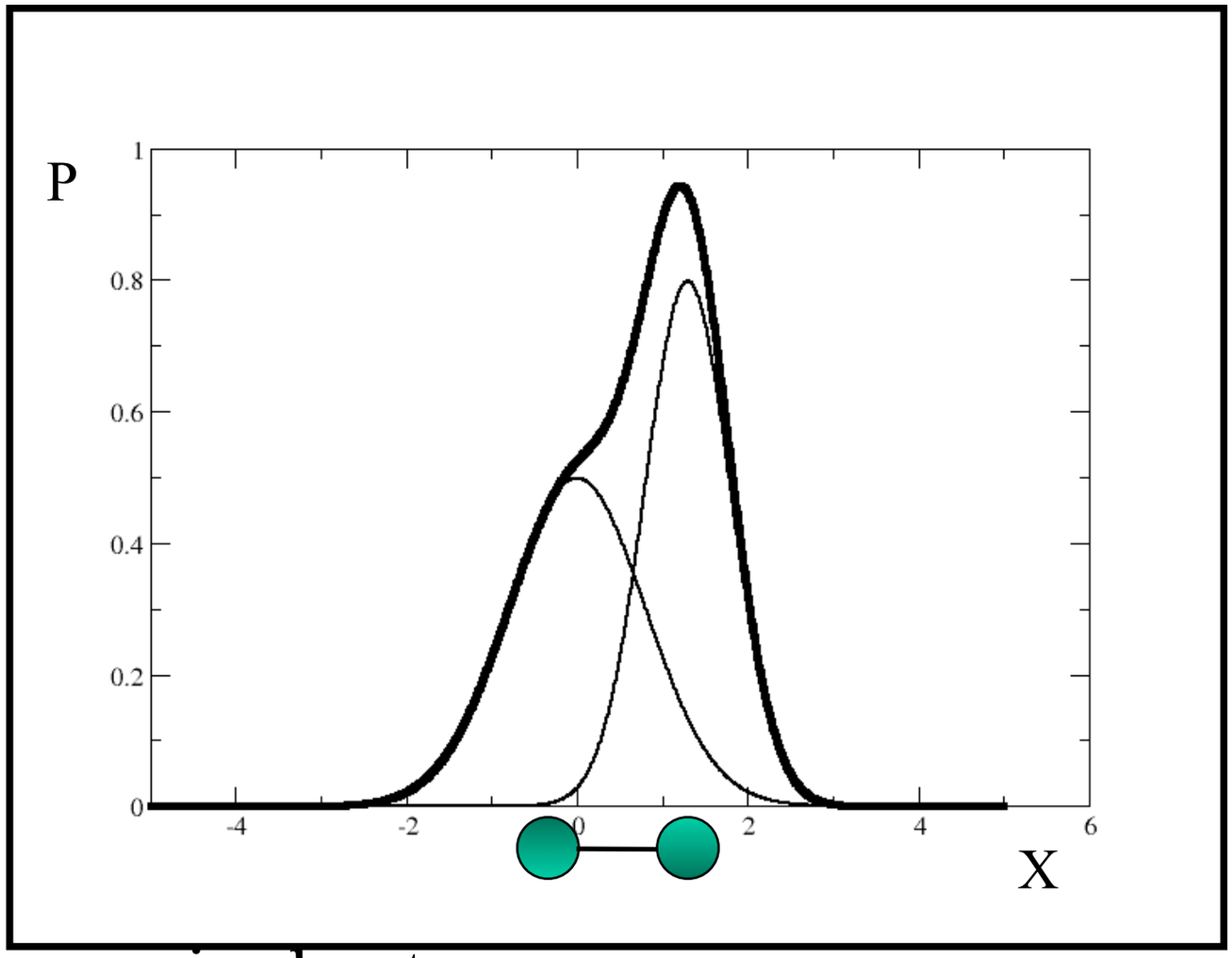
USA RSCB: http://www.rcsb.org/pdb/download/download.do

RASMOL (visualisation tool): http://rasmol.org/

JMOL (Java based visualisation tool): http://jmol.sourceforge.net/

Why restraints:
Two atoms ideal case

Distance between atoms 1.3Å. B values 20 and 50

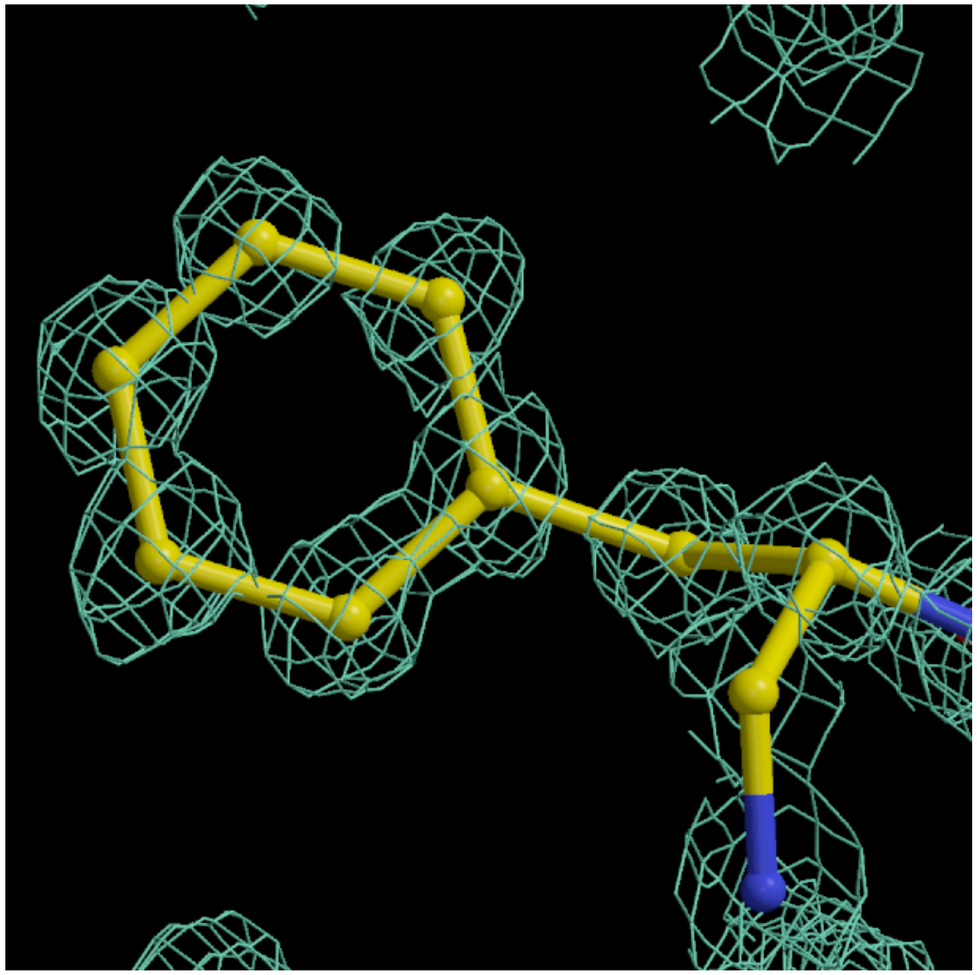


Thin lines – single atoms

Bold line - sum of the two atoms

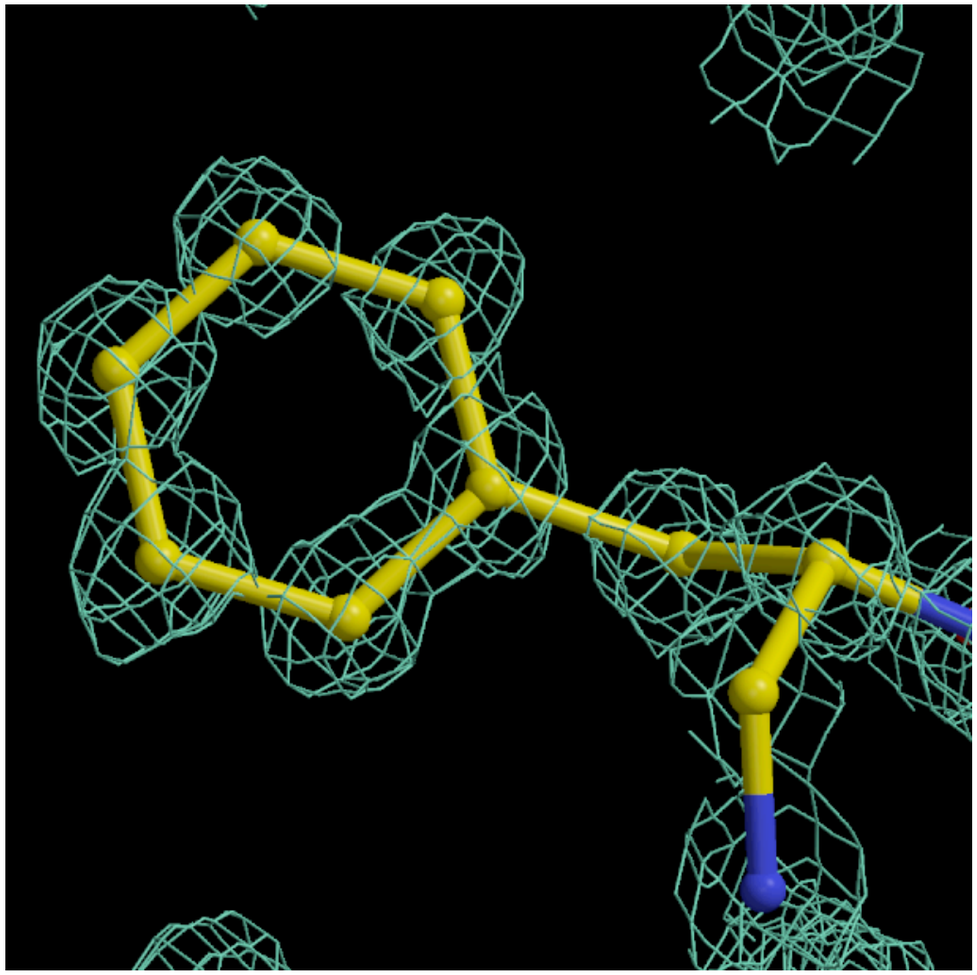
Chemical information: Phe at two different resolutions

Chemical information: Phe at two different resolutions



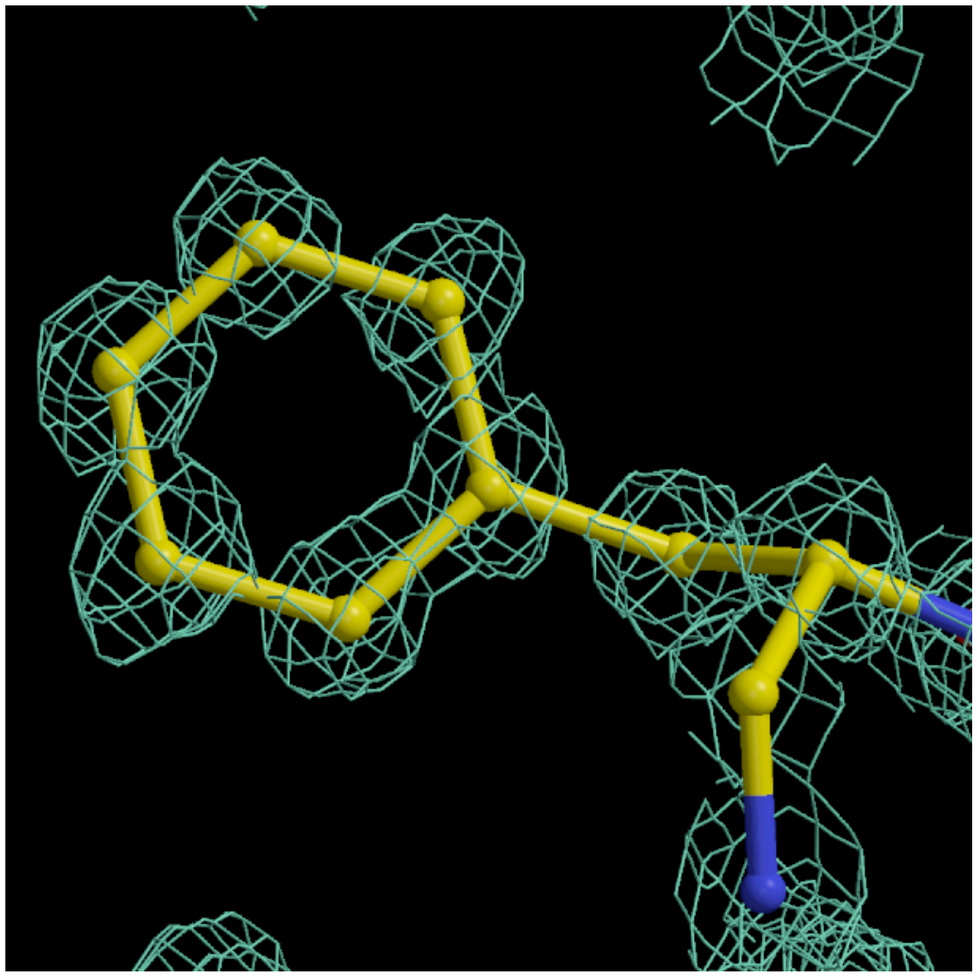
Chemical information: Phe at two different resolutions

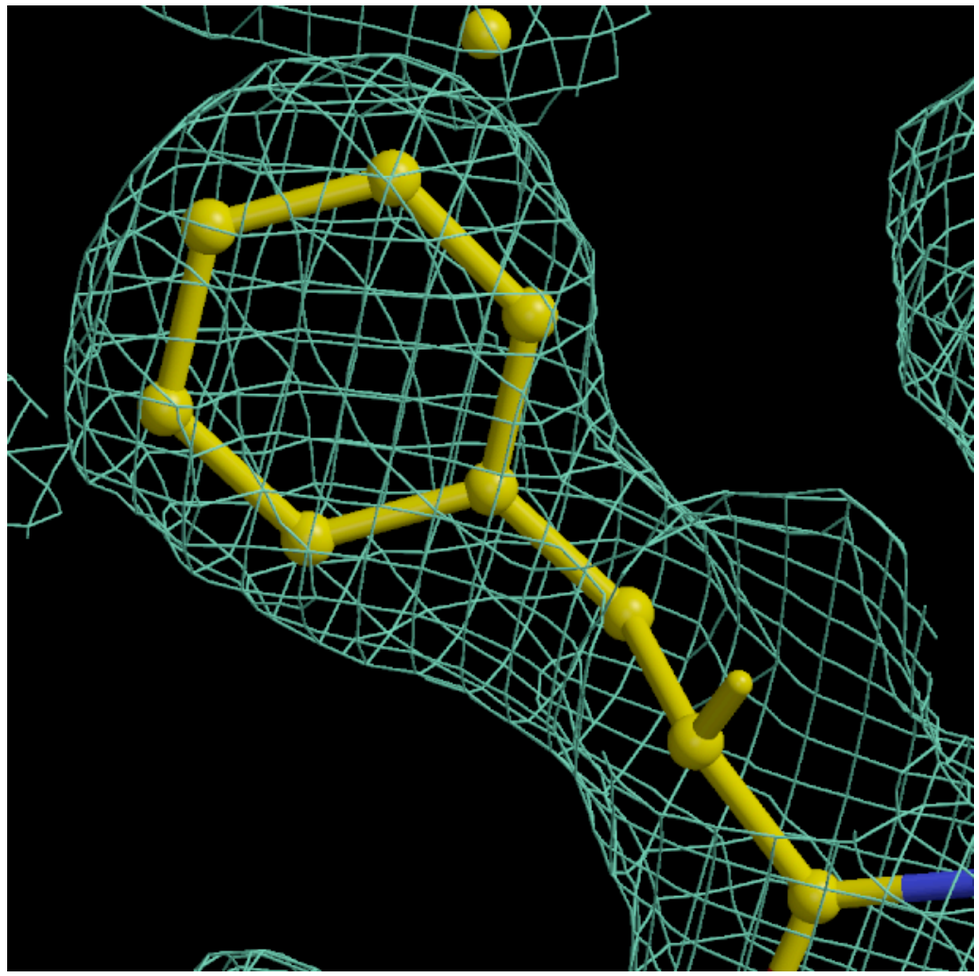
0.88 Å



Chemical information: Phe at two different resolutions

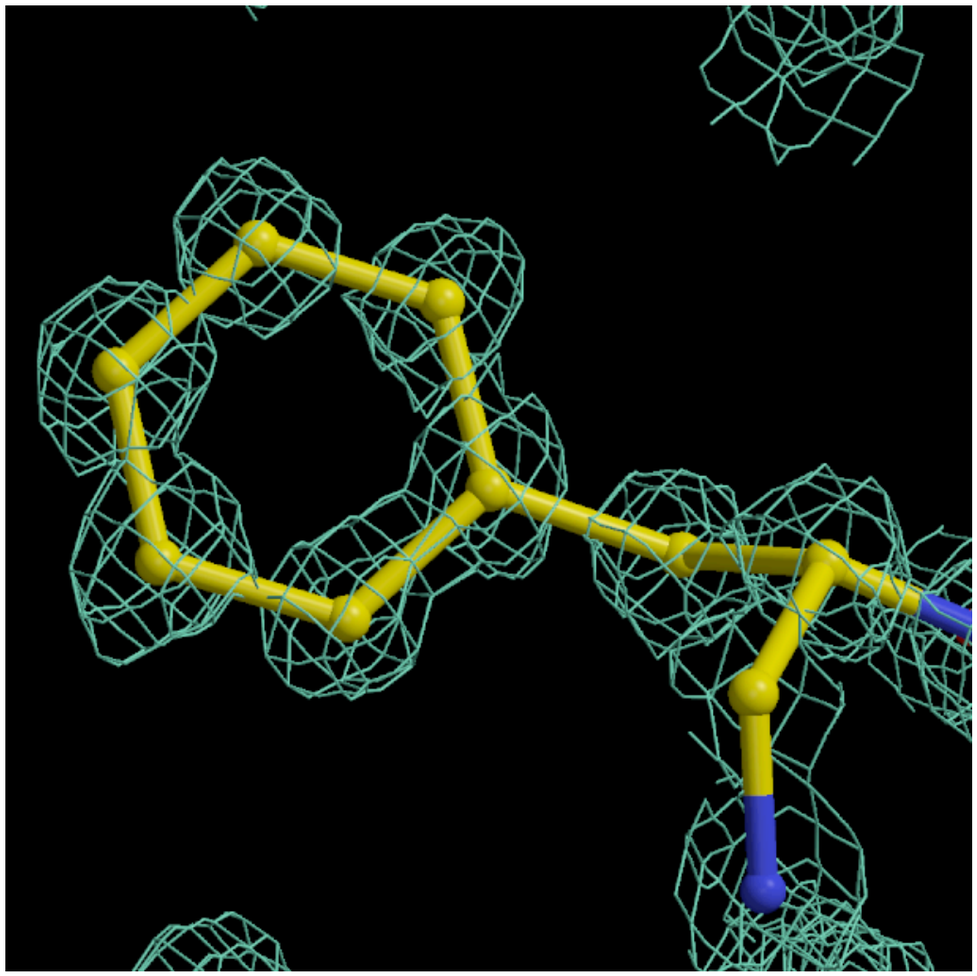
0.88 Å



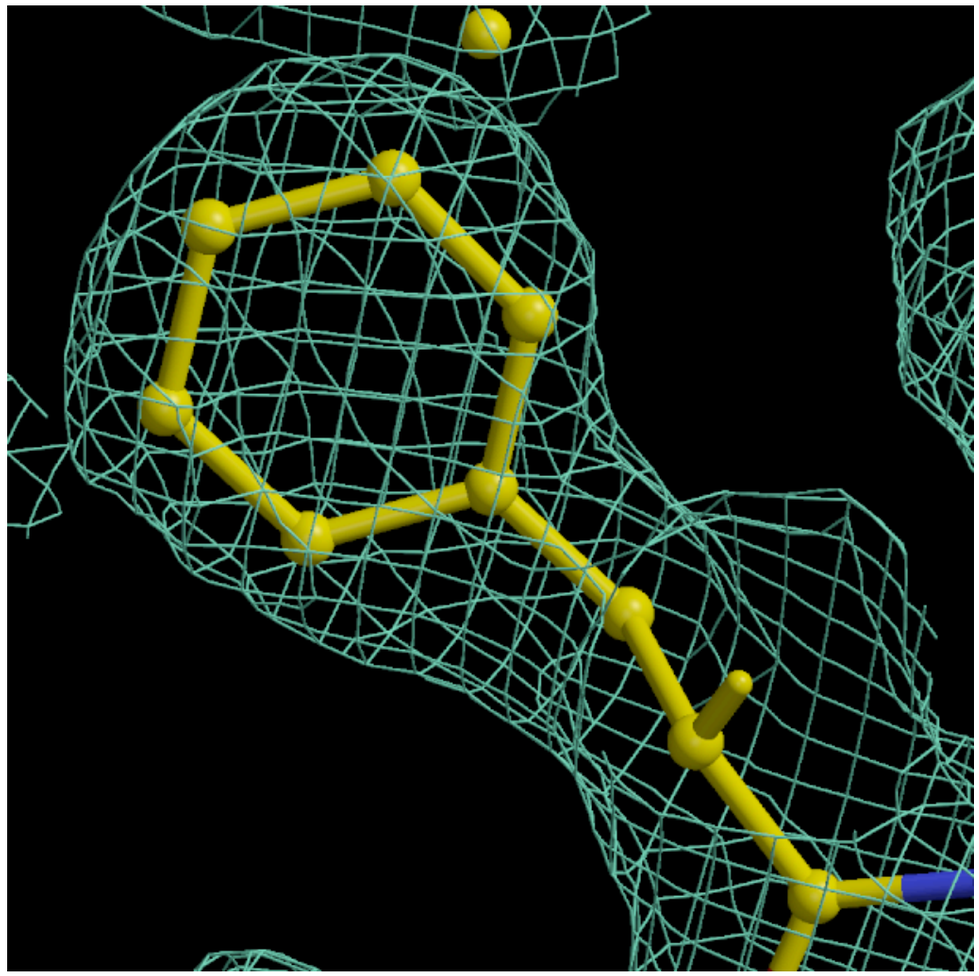


Chemical information: Phe at two different resolutions

0.88 Å



2 Å and High mobility



Role of restraints

When atoms have high B values and/or data are at low resolution then electron density may not show separate peaks

If restraints would not be used then chemistry of molecule would be unreasonable.

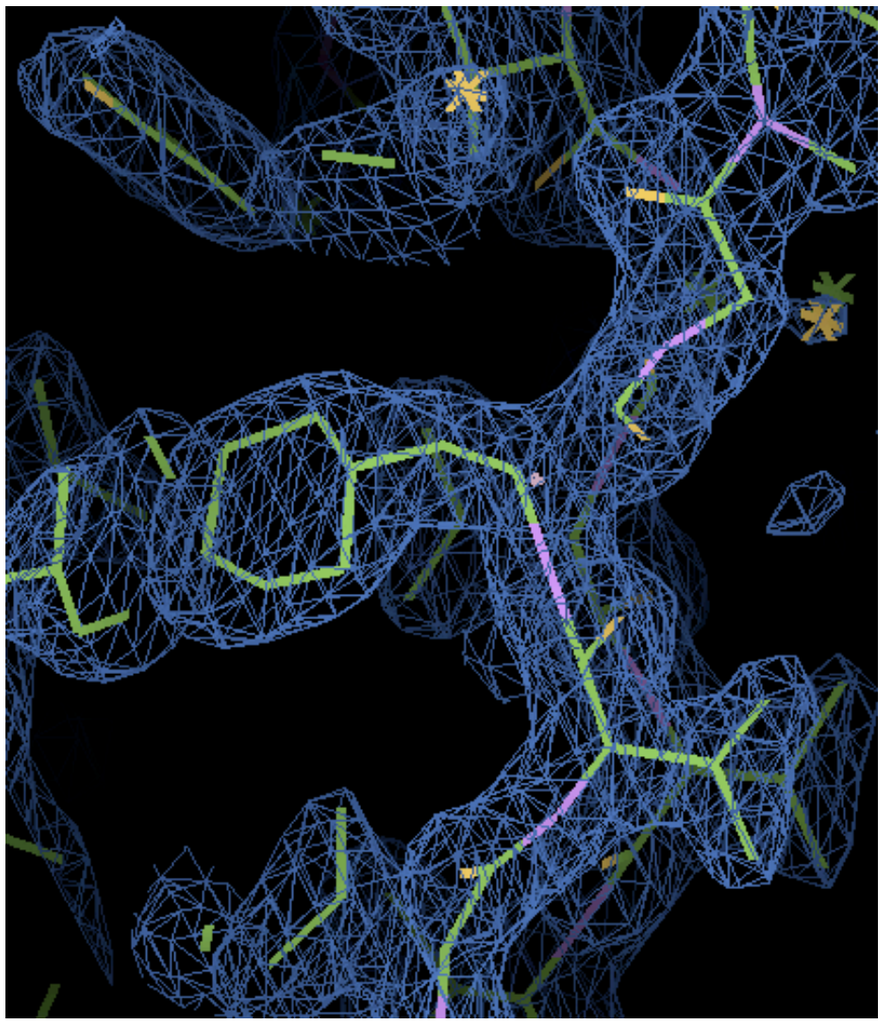
Role of restraints is that to retain chemistry of atoms and at the same time describe electron density optimally.

If atoms are close to each other it is unlikely that they will have hugely different B values

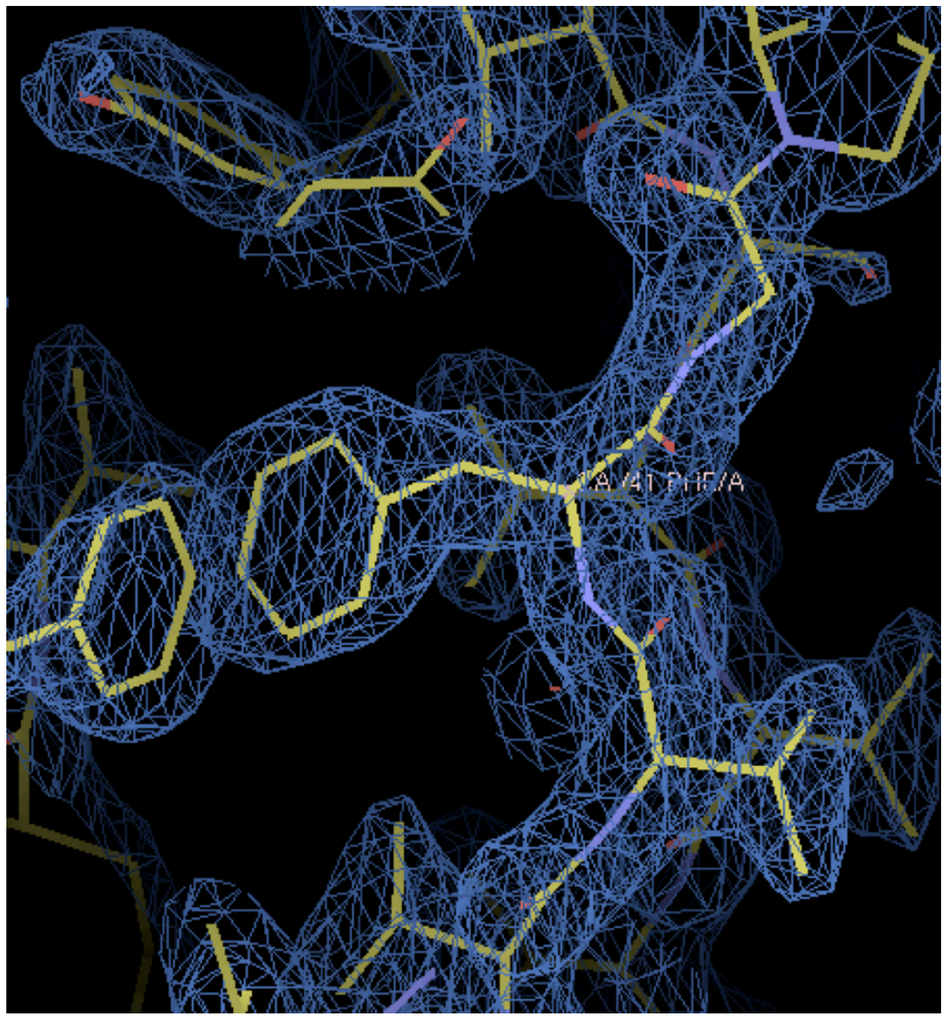
Example

Data - 1.9A

Unrestrained



Restrained



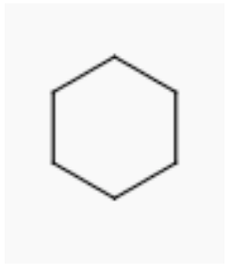
Using restraints

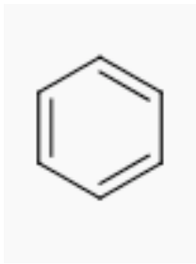
Standard dictionary has description of around 1 500 small molecules. If one of them is in your crystal then the will be used automatically. In the new version there will be more than 8 000.

What happens if you have a ligand that is not in the dictionary. Then it is your responsibility to create chemically sensible description.

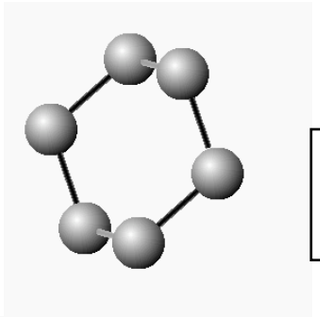
Before starting to create a description you need to study bonding structure of your ligand.

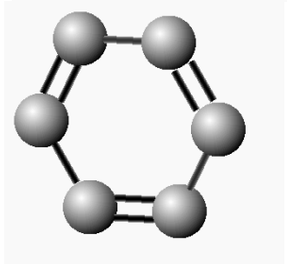
2D





3D

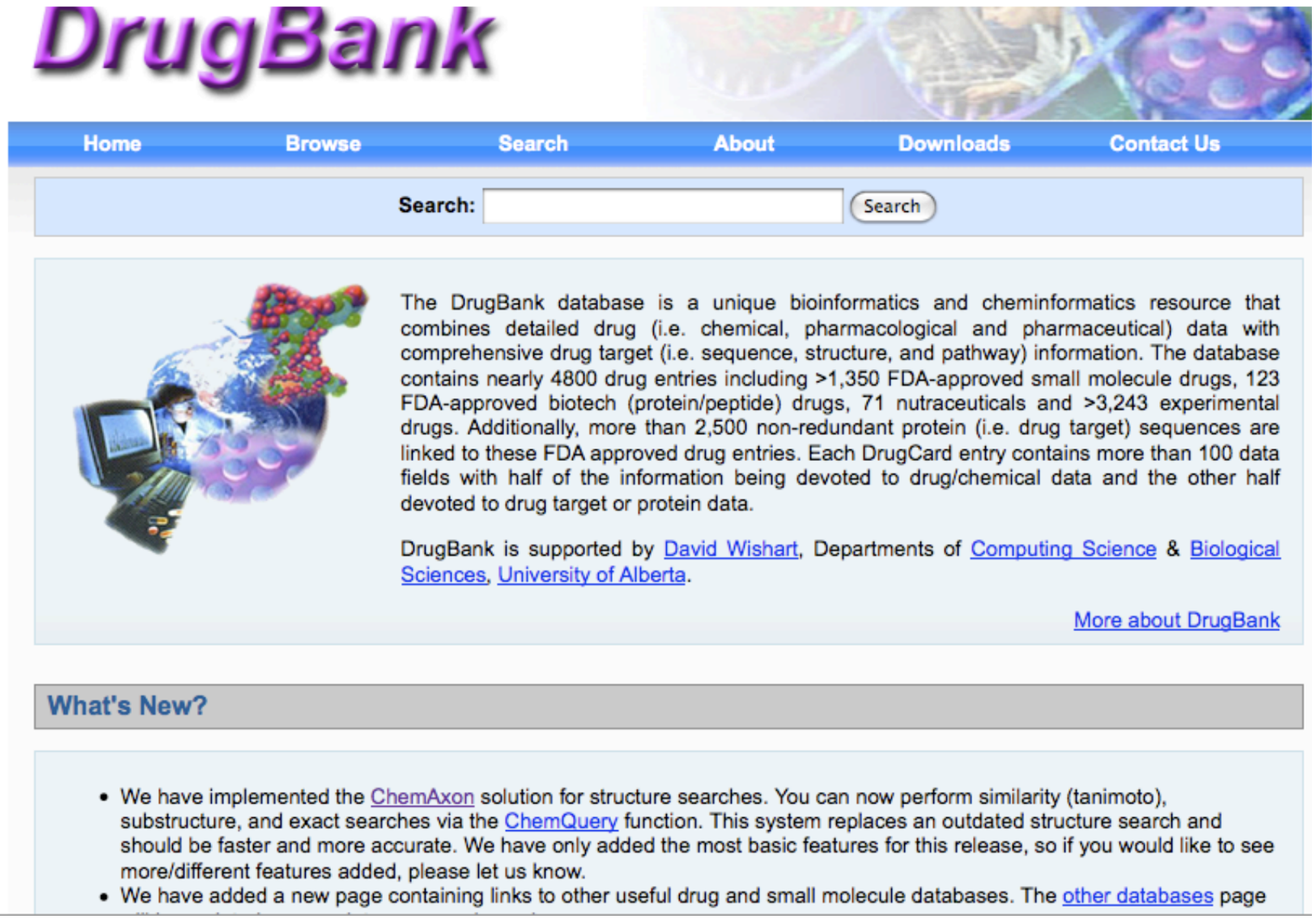




These two molecules will refine very differently (oviously)

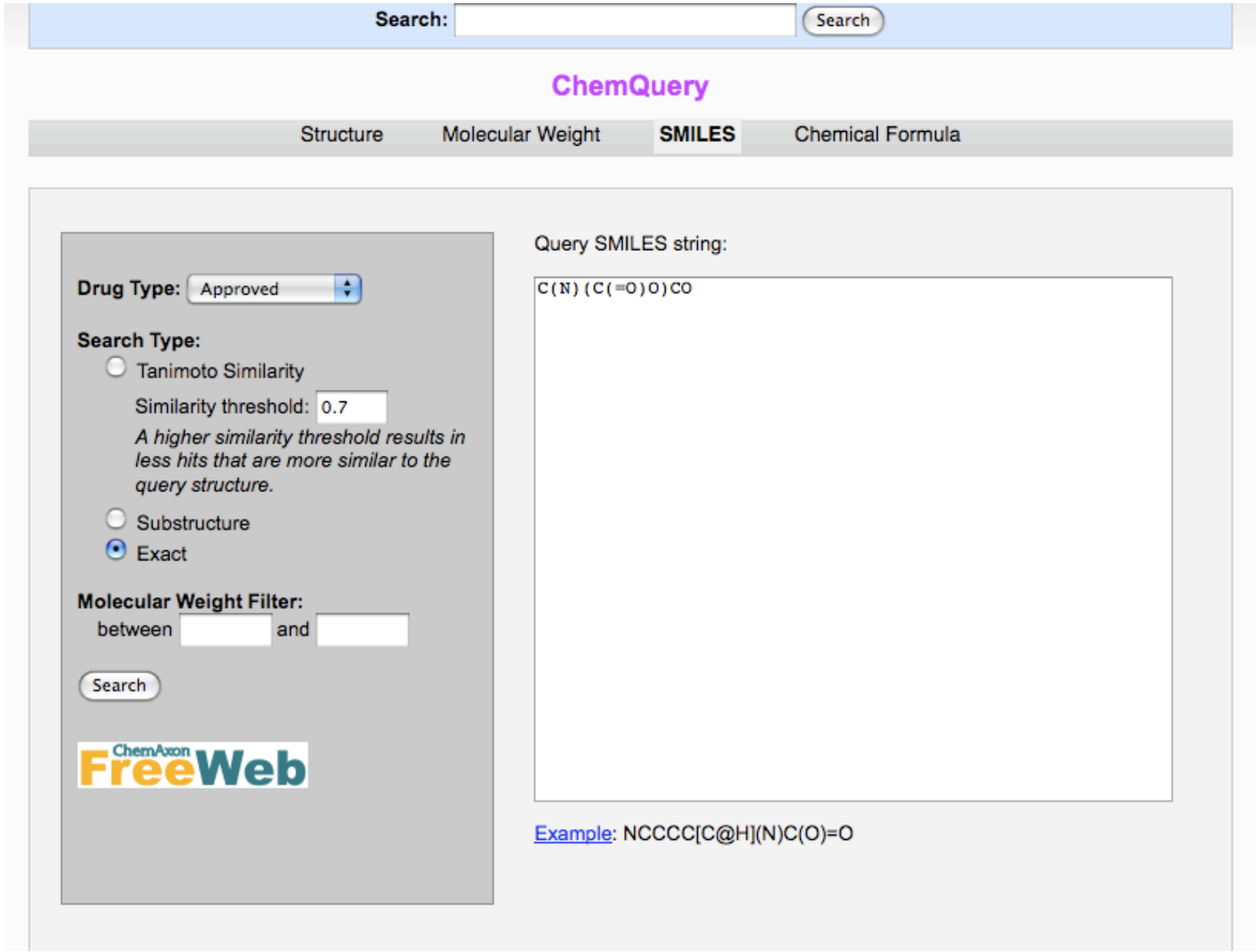
DrugBank

There are various options like “Search”, “Download”



DrugBank

Search can be performed using different tools. One of them is smile string Search can be exact or substructure



SMILES

SMILES notation is the most popular notation and almost all computational chemical websites, programs use this notation. They can read and write SMILES.

It is based on several simple rules. Full description of SMILES can be find from daylight websites.
http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html

SMILES stands for Simplified Molecular Input Line Entry System.

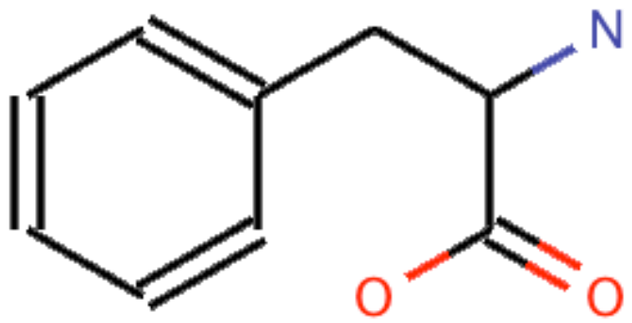
It is concise and widely spread. It is very easy to learn. It was originally designed for manual input using text only editors. SMILES has become as a standard and it is a useful thing to know about.

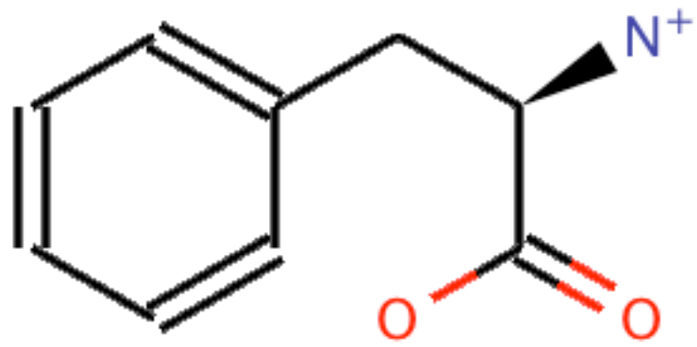
SMILES

SMILES uses several very simple rules (these rules are sufficient to generate SMILES from structure and structure from SMILES).
Rules:
1. Atomic symbols used for atoms
2. Hydrogen atoms as a rule are implicit. They are deduced using valence information about atoms
3. Neighbouring atoms stand one after another
4. Single, double, triple and aromatic bonds are denoted using "-", "=", "#" and ":" respectively. Single and aromatic bonds are usually not shown.
5. Branches represented by parentheses
6. Cycles are added by using matching digits on connecting atoms
7. Aromatic atoms are denoted using lower cases.

These rules are sufficient to describe most of the cases. Let us consider some examples

SMILES





C(O)(=O)C(N)Cc1ccccc1

Hydrogens are implicit, aromatic and single bonds are not shown. Stereochemistry is not clear.

This representation is not unique. Here is the SMILES produced by MarvinSketch:
NC(CC1=CC=CC=C1)C(O)=O

[NH3+][C@H](CC1=CC=CC=C1)C(O)=O

Explicit charges are shown as an attributer of atom. Stereochemistry is shown using @ or @@. Chirality and other stereochemical information in SMILES are local and can be understood using immediate neighbourhood of the atom for which @ or @@ symbols are defined.

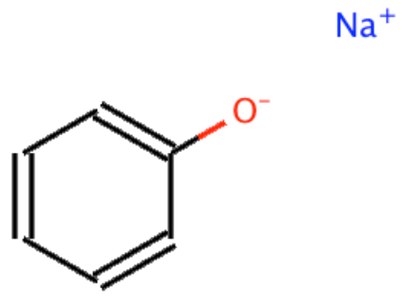
SMILES: atom specification

Atoms are specified using their atomic symbol. Atoms that are not in the organic subset (C, N, O, S, P, F, Cl) or valency is different from "normal" or are isotopes or charged then they are shown inside square brackets. General notation for atoms is:
[weight atom chirality charge]
For example
[2H+] - shows deuterium with atomic charge +1.
[N@H+] - shows positively charged nitrogen atom that is a chiral centre
[Fe++2] - iron atom with atomic charge +2

Hydrogen atoms are not usually specified. In this case the number of hydrogens that would satisfy "normal" valency of atoms is assumed. However hydrogens can be specified explicitly. The number of hydrogens is shown using a digit immediately after hydrogen. It should not cause problem with ring closure since hydrogens can make only one bond (except in reaction intermediates).

SMILES: disconnected atoms

If consecutive atoms are not connected then between them "." is added. For example: [Na+].c1ccccc1[O-] id for.



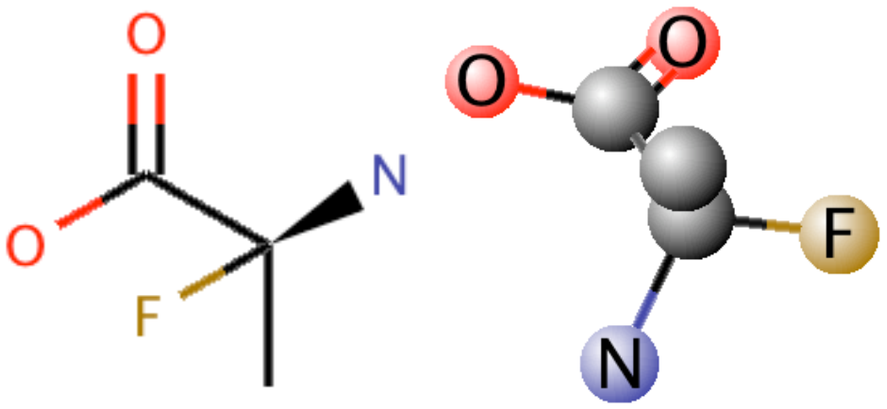
Note that if a SMILES string has "." it does not mean that it is a disconnected structure. For example: C1.C1 is same as CC (note that matching digits show that these atoms must be connected). This notation is used in extension of SMILES to represent reaction (e.g. reagents are separated by ".")

SMILES: stereochemistry

If an atom makes bond with four atoms (e.g. sp^3 carbon) and all these atoms are different then we can arrange these atoms in general in two different ways. One structure cannot be generated from another using rotations and translations only. To distinguish these two structures SMILES uses chirality notations - @ or @@. If @ sign after an atom is specified then it means that if we take the first atom attached to chiral atom and the three remaining atoms appear anti-clockwise. For example in case of N[C@](C)(F)C(O)=O we look down to the chiral carbon from the first atom N then methyl group, F and carboxyl group appear anti-clockwise. If we write @@ then they appear clockwise.
If chiral atom is not the first atom and it has an implicit hydrogen then it is taken as the first atom. For example in case N[C@](C)C(O)=O neighbours H (implicit hydrogen), methyl and carboxyl groups appear anti-clockwise. To avoid confusion in these cases hydrogen can be written explicitly.
If chiral atom is the first atom then implicit hydrogen is an atom "from" where we look at the chiral atom.

I could not find specification what happens if a chiral atom is the first atom and it has no implicit hydrogens.

SMILES: examples of chiral centres written by MarvinSketch



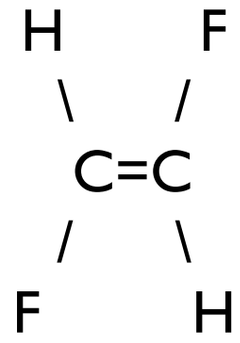
C[C@](N)(F)C(O)=O

C[C@@H](O)F

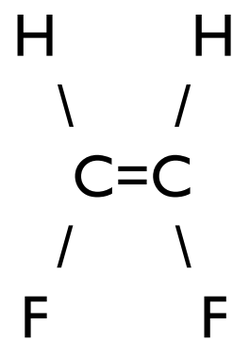
SMILES: stereochemistry

Configuration around double bonds are denoted using matching "\" and/or "/". For example

F\C=C\F (or F/C=C/F) denotes



F\C=C/F (or F/C=C\F) denotes



Note that SMILES chirality and are local chirality.

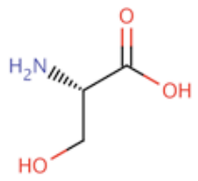
DrugBank

The list of compounds

It gives various info about each structure. They can be visualised using 2D or 3D visualisation tools

ChemQuery Search Results

ChemQuery search returned 1 results.

DrugBank ID	Name / CAS Number	Structure	Formula	Average Weight / Monoisotopic Weight
DB00133 DRUGCARD	**L-Serine** / 56-45-1		$C_3H_7NO_3$	105.0926 / 105.0426
	Canonical SMILES: NC(CO)C(O)=O Isomeric SMILES: N[C@@H](CO)C(O)=O			

Properties

Showing drug card for L-Serine (DB00133)

Legend: drug field | target field | enzyme field — Show Similar Structures for Approved drugs

Version	2.5
Creation Date	2005-06-13 13:24:05
Update Date	2008-08-26 14:00:56
Primary Accession Number	DB00133
Secondary Accession Number	• NUTR00053
Name	**L-Serine**
Drug Type	• Approved • Nutraceutical • Small Molecule
Description	A non-essential amino acid occurring in natural form as the L-isomer. It is synthesized from glycine or threonine. It is involved in the biosynthesis of purines; pyrimidines; and other amino acids. [PubChem]
	1. (-)-Serine 2. (S)-2-Amino-3-hydroxypropanoic acid 3. (S)-Serine 4. (S)-a-Amino-b-hydroxypropionic acid 5. 2-Amino-3-hydroxypropionic acid

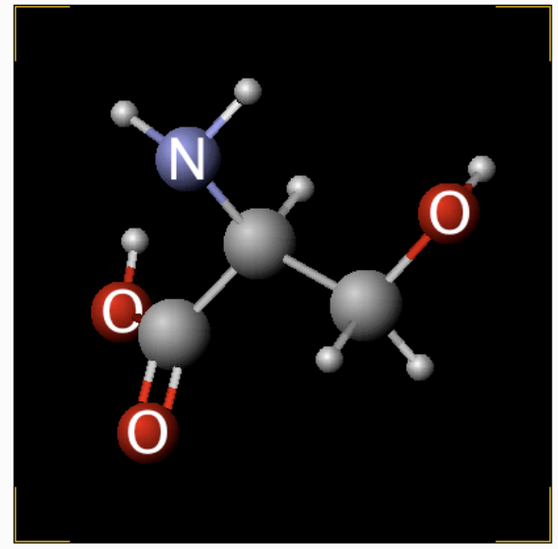
3D view

Structure Viewer

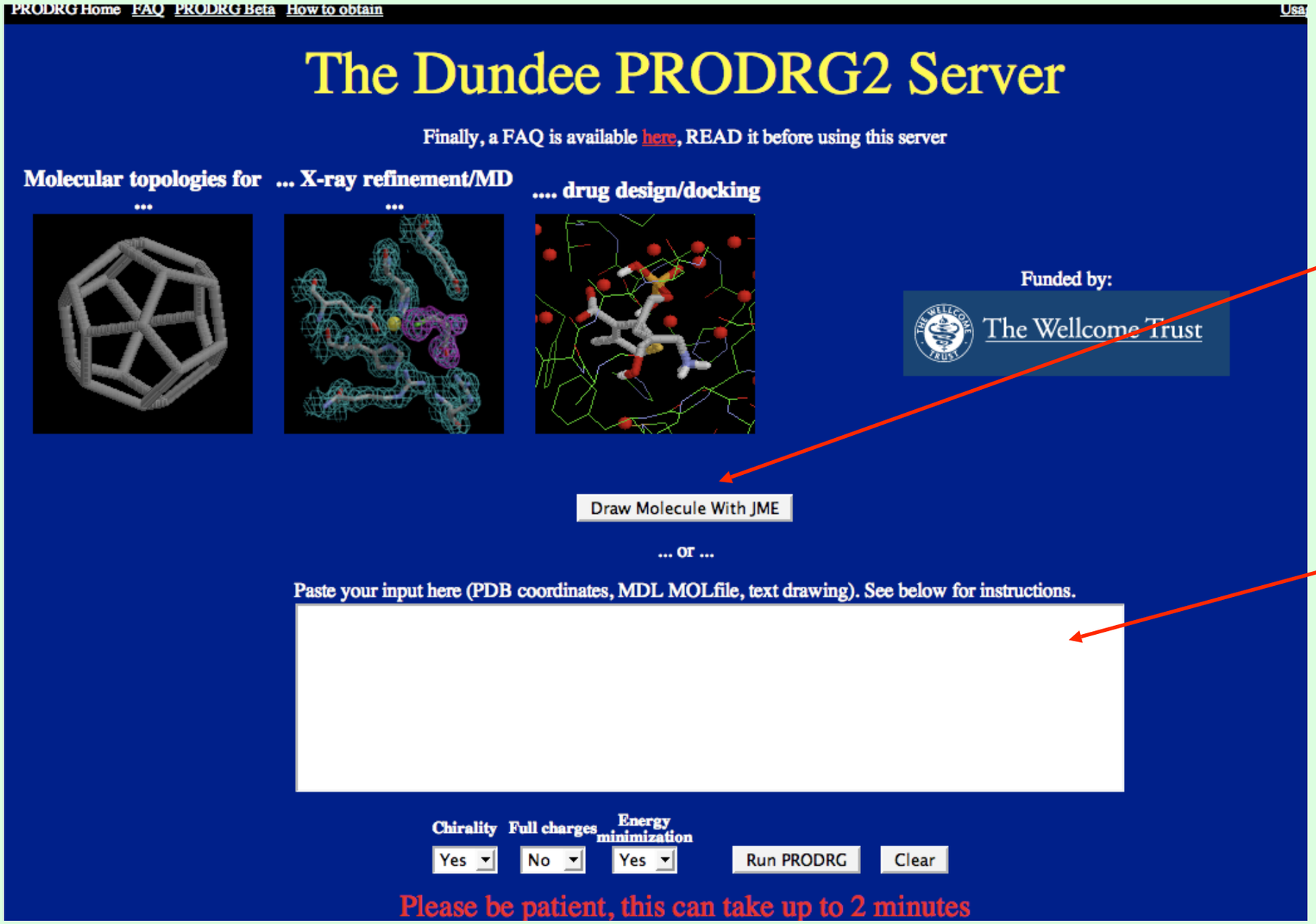
Displaying structure for L-Serine

Double click on the structure to save and manipulate the structure and view

Large molecules may take a while to load



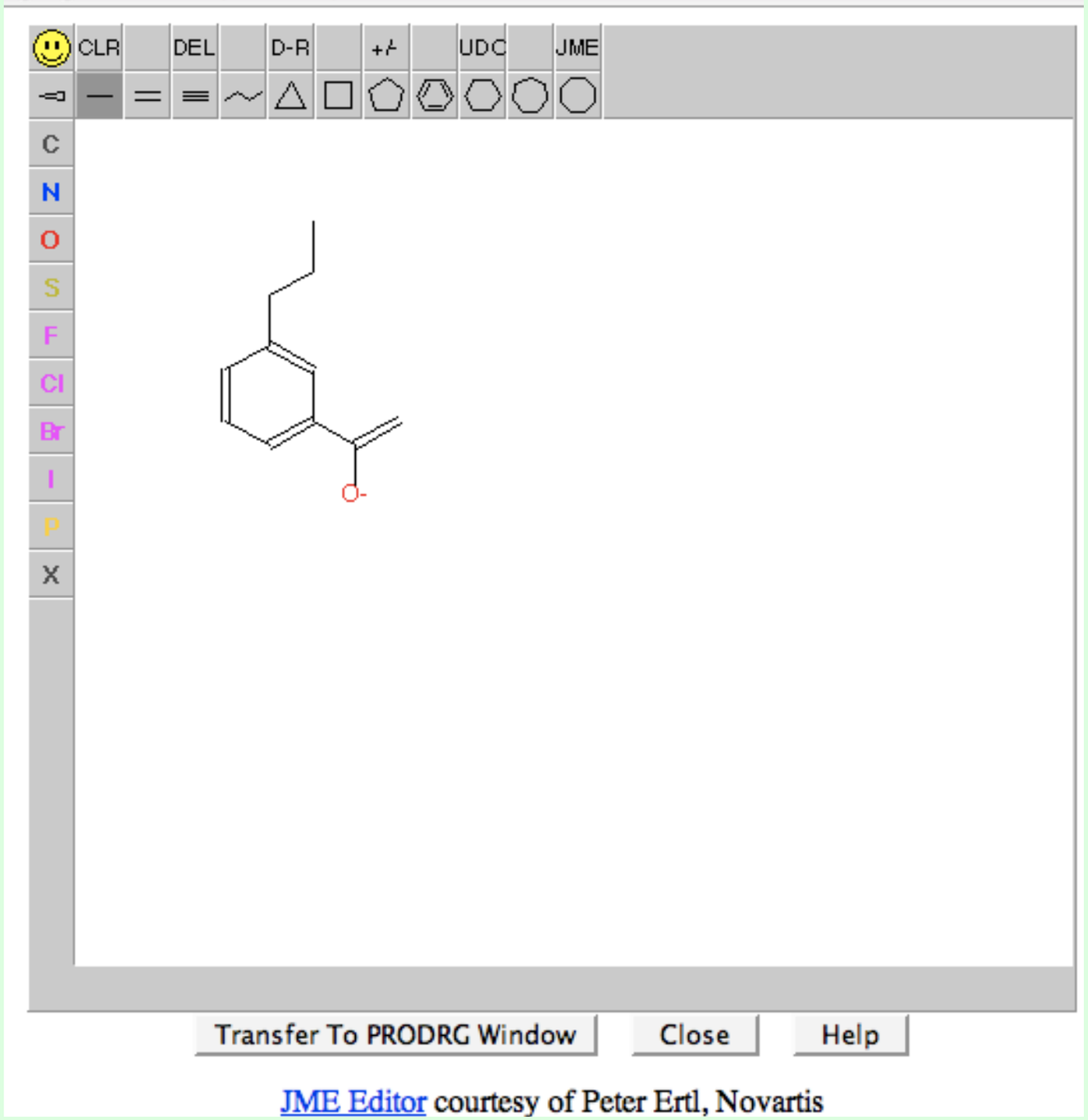
PRODRG server



JME

Load your file

PRODRG: JME



JME is java based program for 2D drawing of small compounds. It is used in PRODRG2, MSDchem etc

Draw your ligand,

transfer to PRODRG

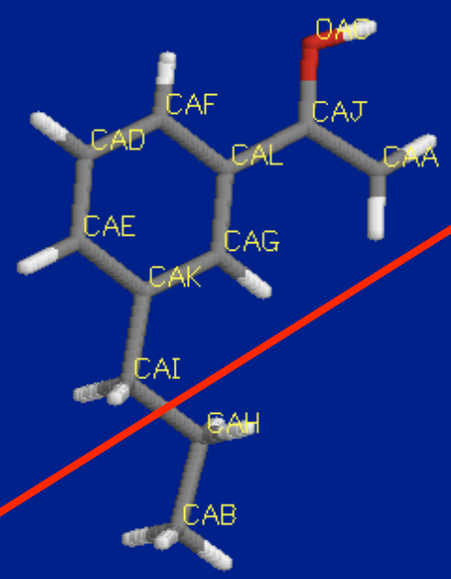
window and run

PRODRG output

PRODRG Home FAQ PRODRG Beta How to obtain

```
PRODRG> Starting up PRODRG version 061128.0522
PRODRG> PRODRG written/copyrighted by Daan van Aalten
PRODRG> and Alexander Schuettelkopf
PRODRG>
PRODRG> Questions/comments to dava@davapc1.bioch.dundee.ac.uk
PRODRG>
PRODRG> When using this software in a publication, cite:
PRODRG> A. W. Schuettelkopf and D. M. F. van Aalten (2004).
PRODRG> PRODRG - a tool for high-throughput crystallography
PRODRG> of protein-ligand complexes.
PRODRG> Acta Crystallogr. D60, 1355--1363.
PRODRG>
PRODRG>
PRODRG> MOL mode detected.
PRODRG> No stereo information found in input file.
PRODRG> Molecule complexity index: 2.00.
PRODRG>    1 hydrogen(s) added.
PRODRG>   13 bonds                 1 ambiguous
PRODRG>   16 bond angles            3 ambiguous
PRODRG>    9 improper dihedrals     1 ambiguous
PRODRG>    4 dihedrals              0 ambiguous
PRODRG>    2 partial charges        0 ambiguous
PRODRG> Net charge on molecule:    0.000
PRODRG> Using charge groups.
PRODRG> Writing GROMACS topology.
PRODRG> GROMACS topology quality on 0-10 scale:  7.7
GENDRG> Best structure was iteration  841 with   0.70210928
PRODRG> Spawning GROMACS version 3.2.1...
PRODRG> RMSD from GROMOS bond ideality (Angstrom) :    0.017
PRODRG> RMSD from GROMOS angle ideality (degrees) :    2.257
PRODRG> RMSD from GROMOS plane ideality (degrees) :    0.432
PRODRG> Number of improper improper dihedrals     :        0
PRODRG> Writing: SCRHWMMPG
PRODRG> Normal program end
```

Your molecule + added hydrogens



It can write out representation in various formats suitable for various popular software

Click to go to the following output:

Coordinates

- PDB (all H's, polar H's only or no H's)
- MDL Molfile (all H's, polar H's only or no H's)
- GROMOS87/GROMACS (polar H's only)

X-ray refinement

- CNS (parameters and topology)
- REFMAC5
- SHELX
- O (pre-9.x torsion entry, pre-9.x refi dictionary and 9.x dictionary)

Done

PDB

PDB is Protein Data Bank. It has all macromolecular structures determined experimentally as well as theoretically. There are more than 56000 macromolecular structures available in the PDB.

In many cases protein structures are determined with some ligands (small molecular compounds). These small molecular structures are available from PDB. There are 8000-9000 such small molecules in the PDB.

There are websites that allows people to view macromolecular structures as well as small molecular compounds. These sites are located in USA, Europe and Japan.

PDB in Europe: MSD at EBI, Cambridge

- General information
- How to use it
- Overview

Search by:
- References in macromolecules ?
- Molecule Classification ?
- Atom energy types ?

- **Binding sites**
- **Chemical search**
- **Ligand Export**
- **Ligand index - Download**

Export a single file with the complete ligand dictionary in XML or view alphabetical listings of ligands and download the zipped collection

Software contributions:
- Cactvs
- JME
- Molecular Networks
- VEGA
- Jmol

EBI > Databases > Structure Databases > MSD > Services contact msd

Ligand Chemistry ? Energy types ?

Consistent and enriched library of ligands, small molecules and monomers that are referred as residues and hetgroups in any PDB entry (8702 currently in the database - Release:03-2008_02_22)

MSDchem : Molecule Search Reset

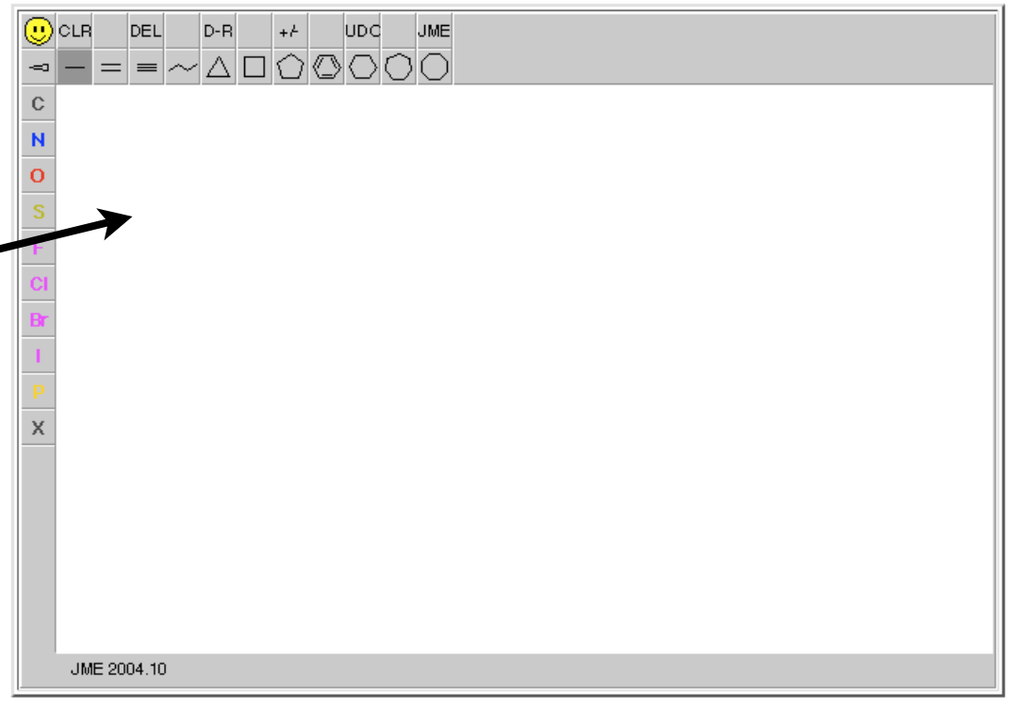
3 letter code	=		
Code	=		
Molecule name	like		
Formula	formula range		edit
Non stereo smile	has substructure		edit
Stereo smile	exact stereo structure		edit
Fragments	fragment expression		edit
Fingerprint	common segments		edit

And / Or

Retrieve: Html

Substructure, common segment, exact stereo

Draw Structure (how to use editor):



Load OK Cancel

or Load File Browse... **(SDF,MOL,mmCif,PDB e.t.c.)**

or give Smile String **(i.e. c1ccccc1)**

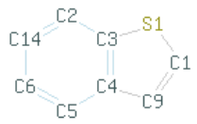
or give Code of Existing Molecule **(i.e. ATP)**

(Press the Load Button to Load the Molecule with that smile or 3 letter code, or file into the editor)

Fragment expression

Select chemical fragment pattern

acetylurea	acridine	acridone			
actinophenoxazine	adenine	alkaloid			
barbit	barbiturates	barbiturgroup			
benzimidazole	benzodiazepine	benzofuran			
benzoisoquinoline	benzothiadiazide	benzothiazole			
benzothiophen	benzoxazole	bilirubin			
biotin	carbazole	cephalosporin			
chromen	cinnoline	coumarine			
cyclobutane	cyclohexane	cyclopentane			
cyclopropane	cytosine	deoxyribose			
dibenzofuran	dibenzothiophen	dithiolane			
flavin	furan	furanose			
glycerophos	guanine	imidazole			
indole	inosine	isoquinoline			
isoxadiazole	isoxazole	naphtyridine			
napthalene	oxadiazole	oxazole			
oxazolidinedione	oxepin	peptide			
penicillin	phenanthrene	phenanthridine			
phenanthroline	phenazine	phenothiazine	phenyl	phthalazine	piperazine
porphin	prosto	pteridine	pteroyl	purine	pyran
pyranose	pyrazine	pyrazole	pyridazine	pyridine	pyrimidine
pyrole	quinazoline	quinoline	quinoxaline	rauwolfia	ribose
steroid	succinimide	thiadiazole	thiazole	thiepin	thiophen
tocol	vitaminAcore	xanthen			

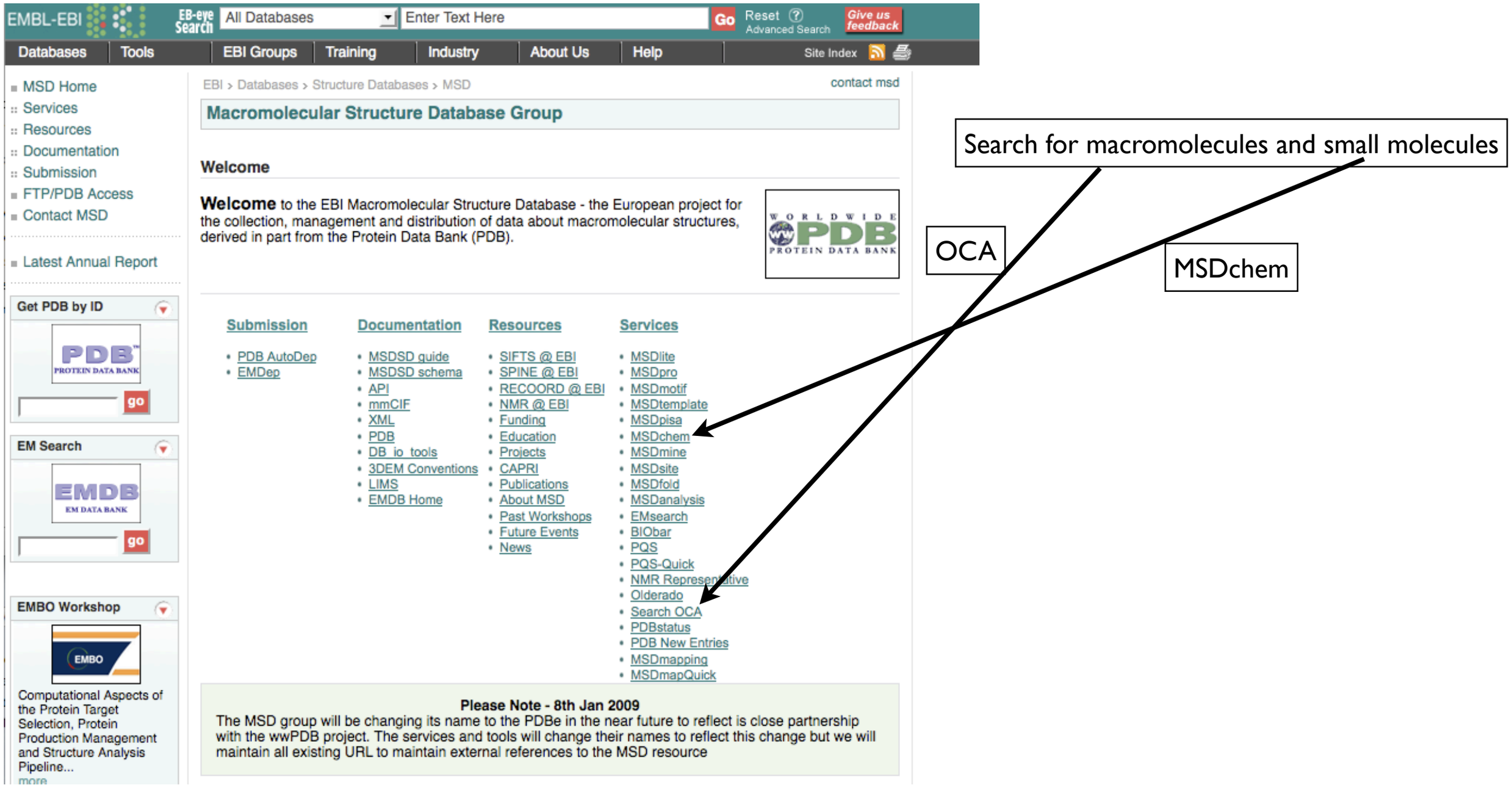


Fragment: min: max: none: any: Add

Fragment expression

Reset OK

PDB in Europe: MSD at EBI, Cambridge



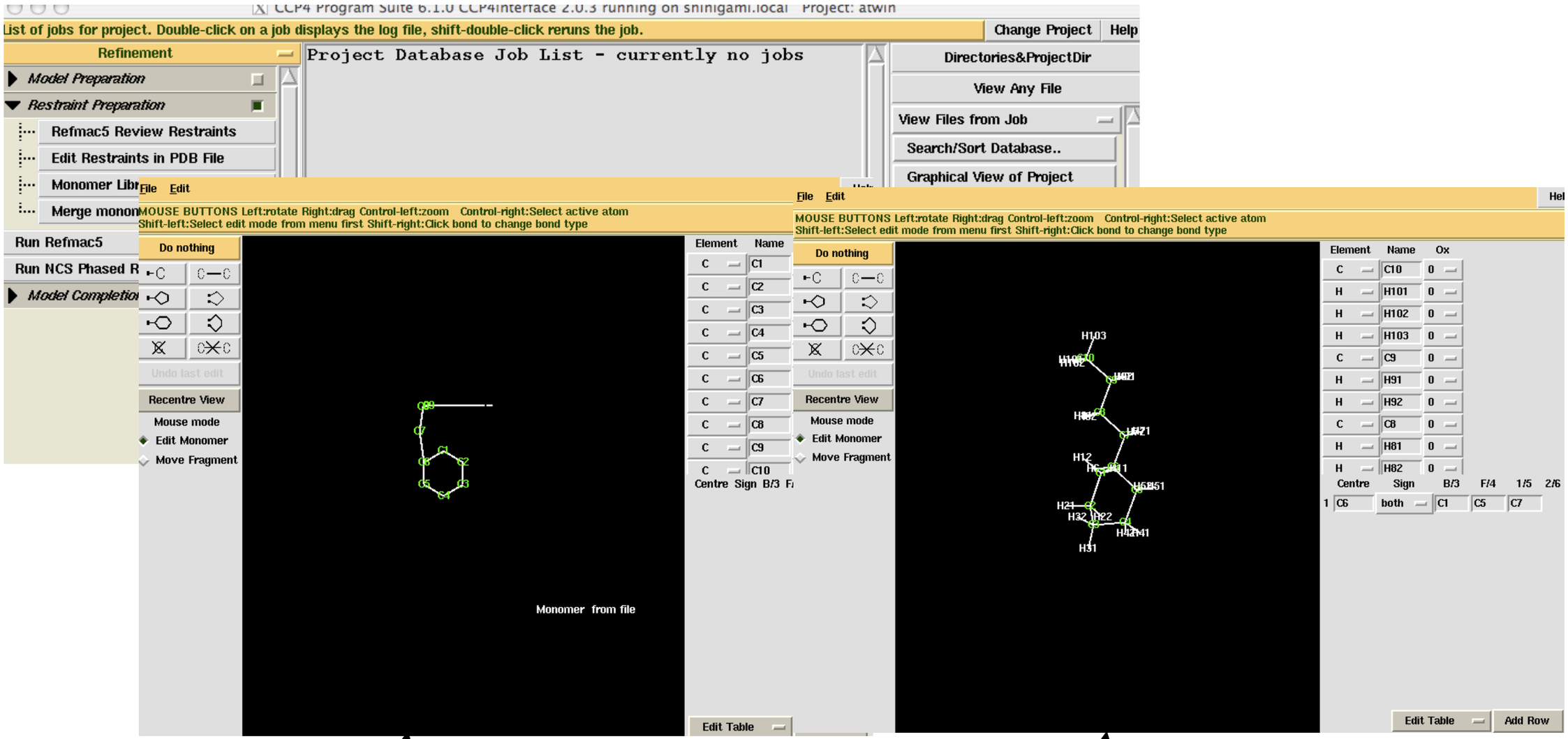
Search for macromolecules and small molecules

OCA

MSDchem

Using resources from ccp4

Sketcher is under Refinement/Restraint Preparation/Monomer library sketcher.



Sketch your ligand

After regularisation

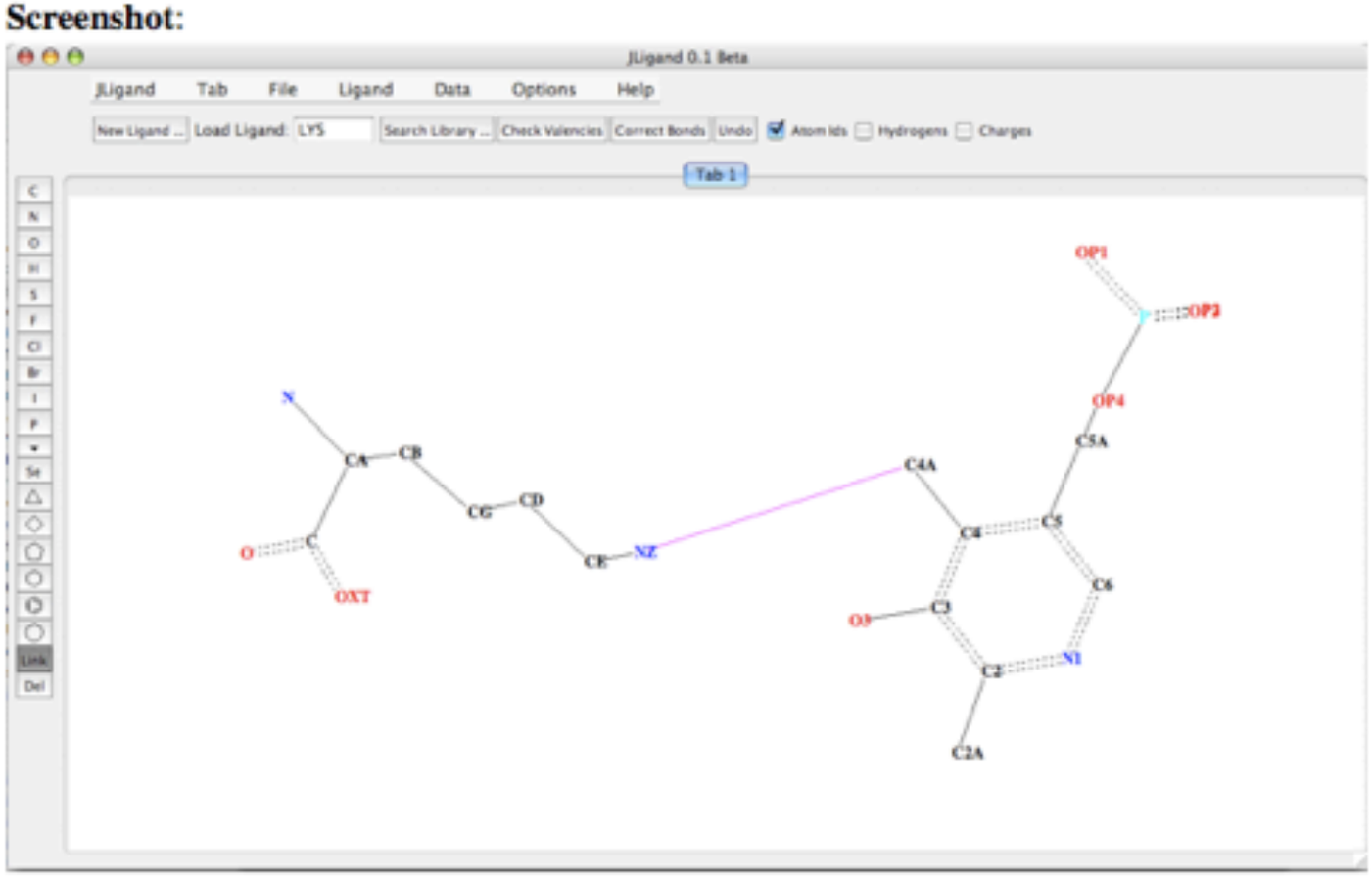
http://www.ysbl.york.ac.uk/~pyoung/JLigand/JLigand.html

JLigand

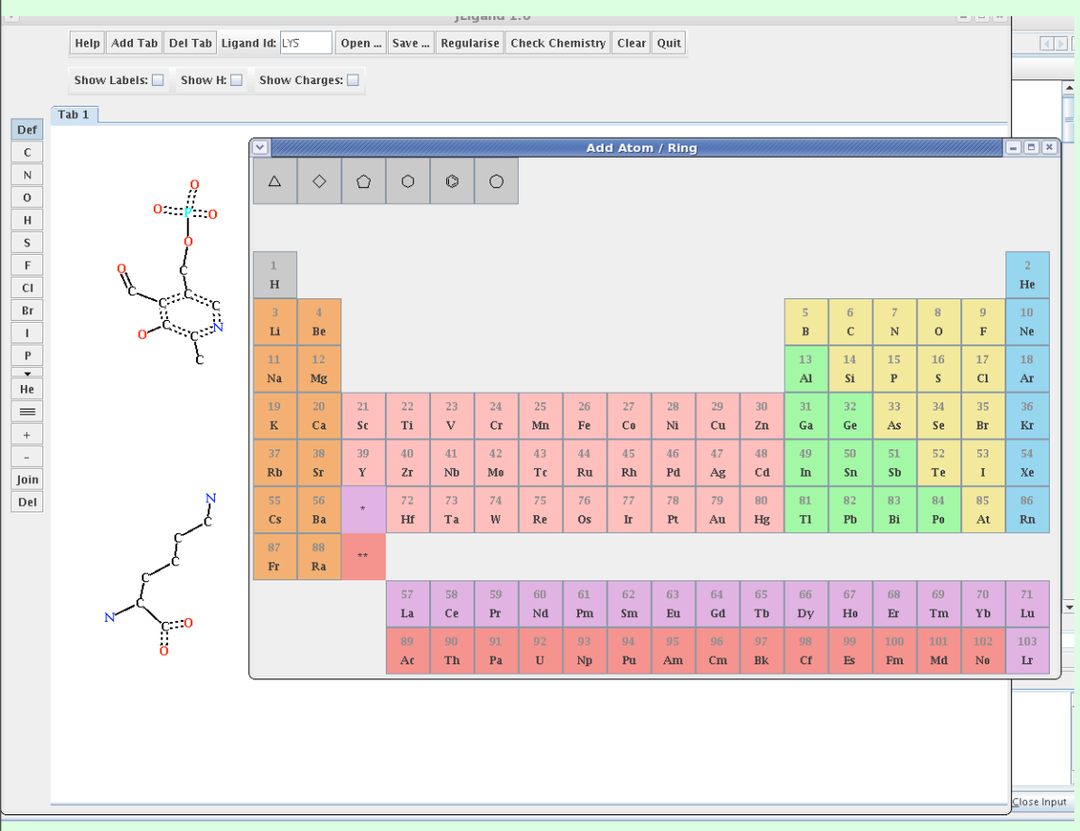
JLigand 0.1 Beta	Download
Java Source (Netbeans Project - Java 1.5)	java source
Java Executable	java executable
Readme file	README
JLigand Info	JLigand Info

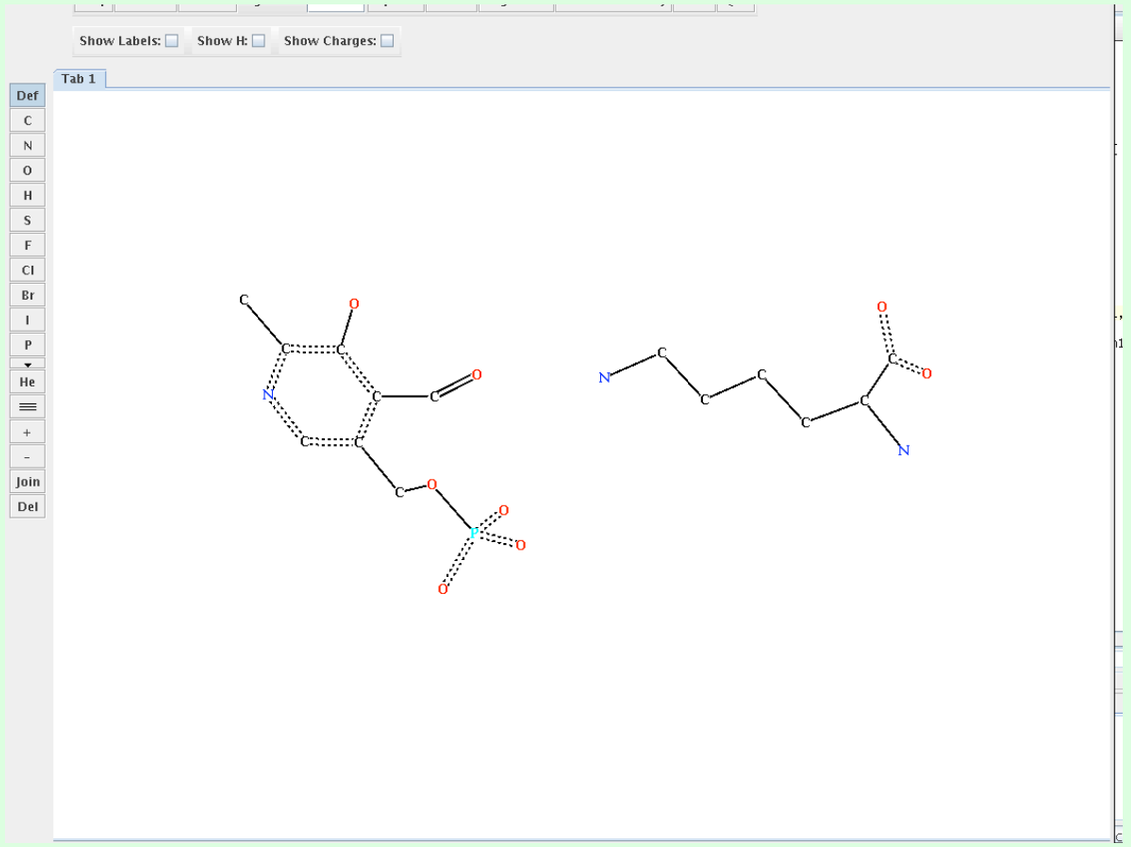
You can download the latest refmac, libcheck and monomers ligand dictionary from Garib's Refmac pages (Note: libcheck is part of the r

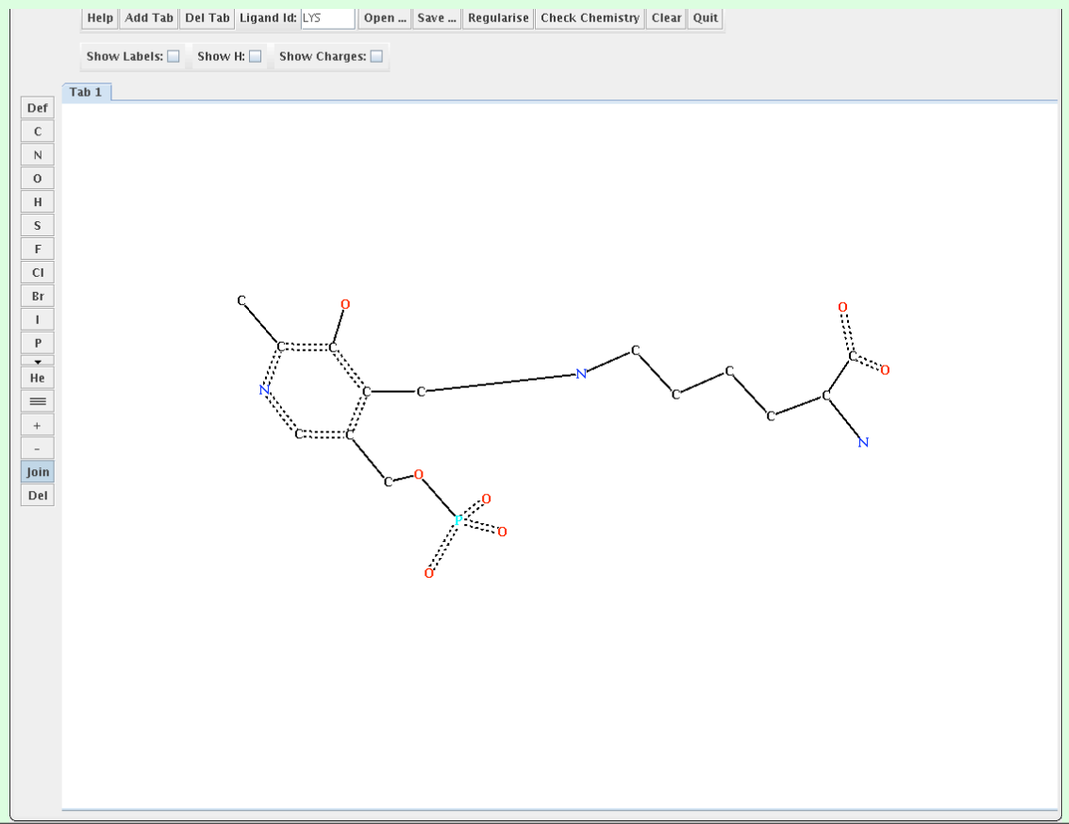
Screenshot:

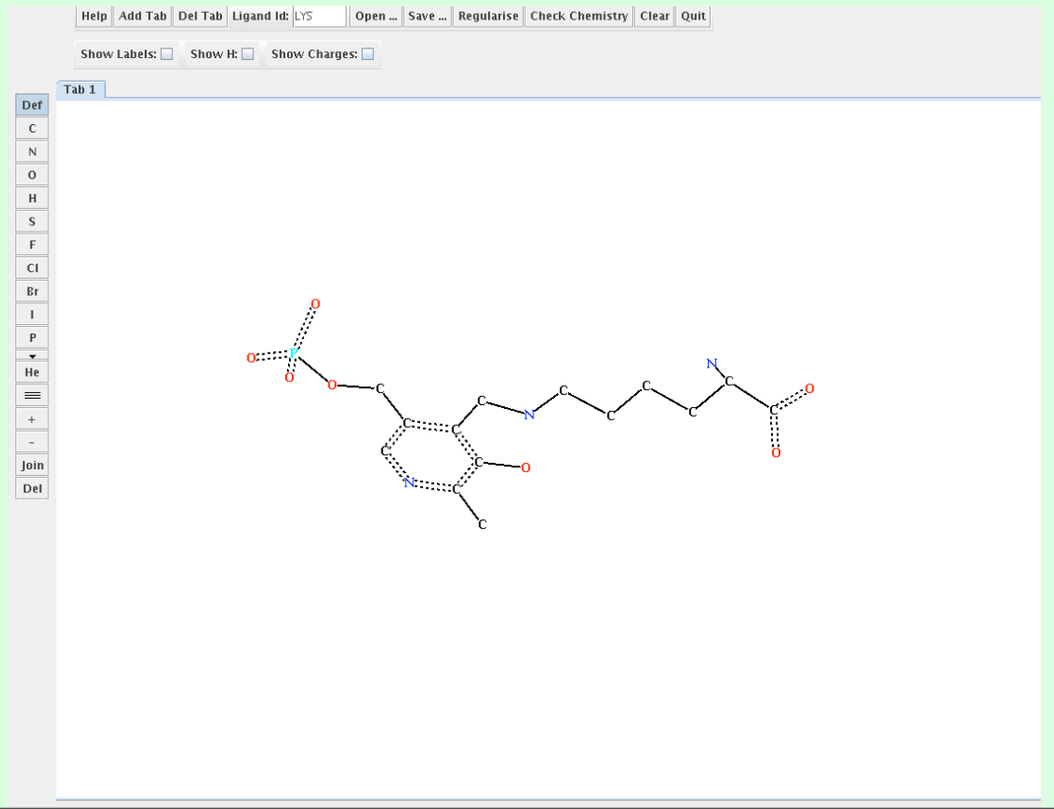


Jligand: New ligand design interface



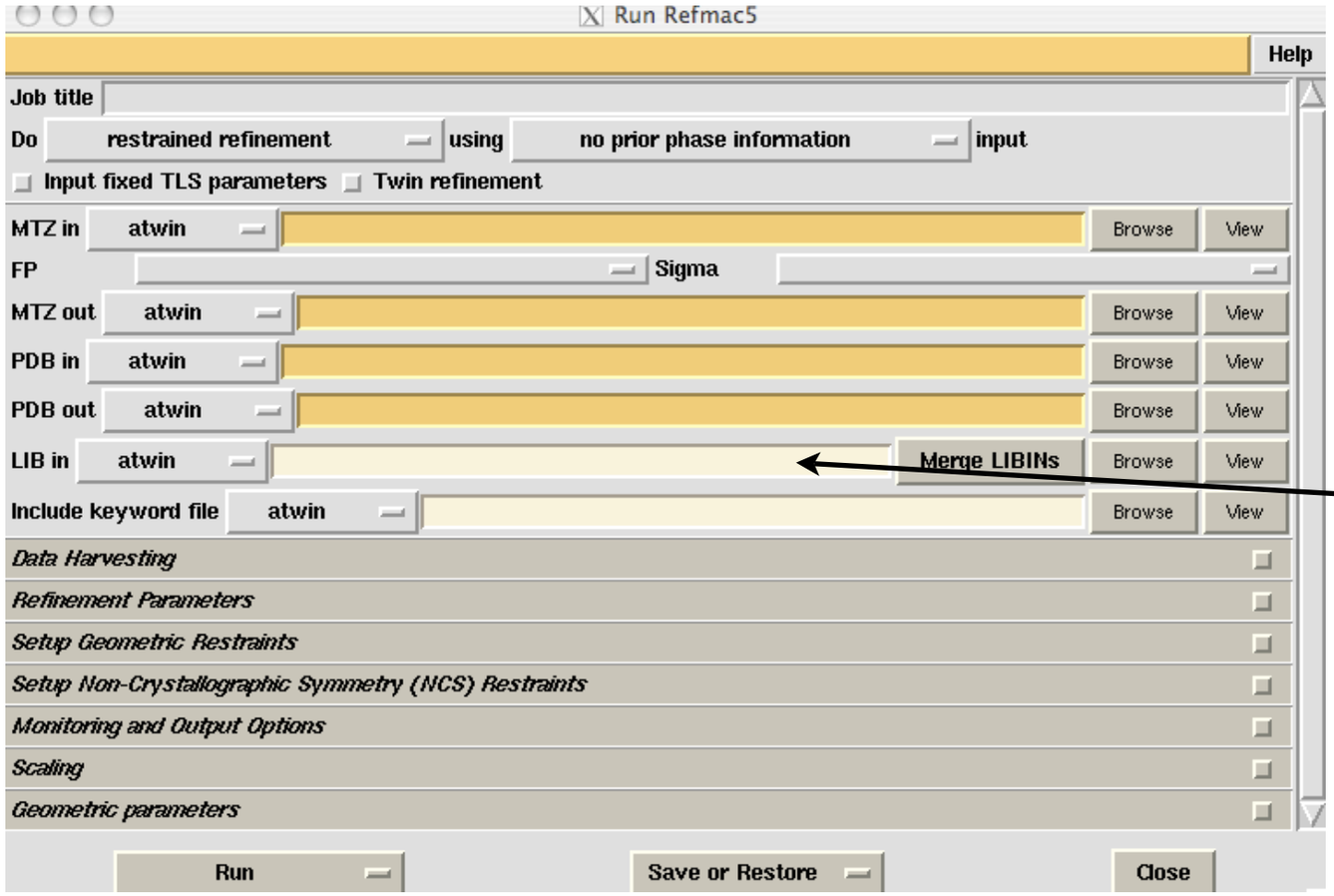






```
# created using: JLigand 1.0
# libcheck version: ()
# dictionary version: ()
# authors: Paul Young, Andrey Lebedev, Alexei Vagin, Garib Murshidov
#
#
#
data_link_list
loop_
_chem_link.id
_chem_link.comp_id_1
_chem_link.mod_id_1
_chem_link.group_comp_1
_chem_link.comp_id_2
_chem_link.mod_id_2
_chem_link.group_comp_id_2
_chem_link.name
PHEPHE1    PHE    PHEmod    .    PHE1    PHE1mod    .    'PHEPHE1'
#
data_mod_list
loop_
_chem_mod.id
_chem_mod.name
_chem_mod.comp_id
_chem_mod.group_id
PHEmod    'PHEmod'    PHE    .
PHE1mod    'PHE1mod'    PHE1    .
#
data_link_PHEPHE1
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
```

Once satisfactory dictionary has been created they can be used in refinement and model building. In refinement



Your dictionary should go here.