Dictionary of ligands

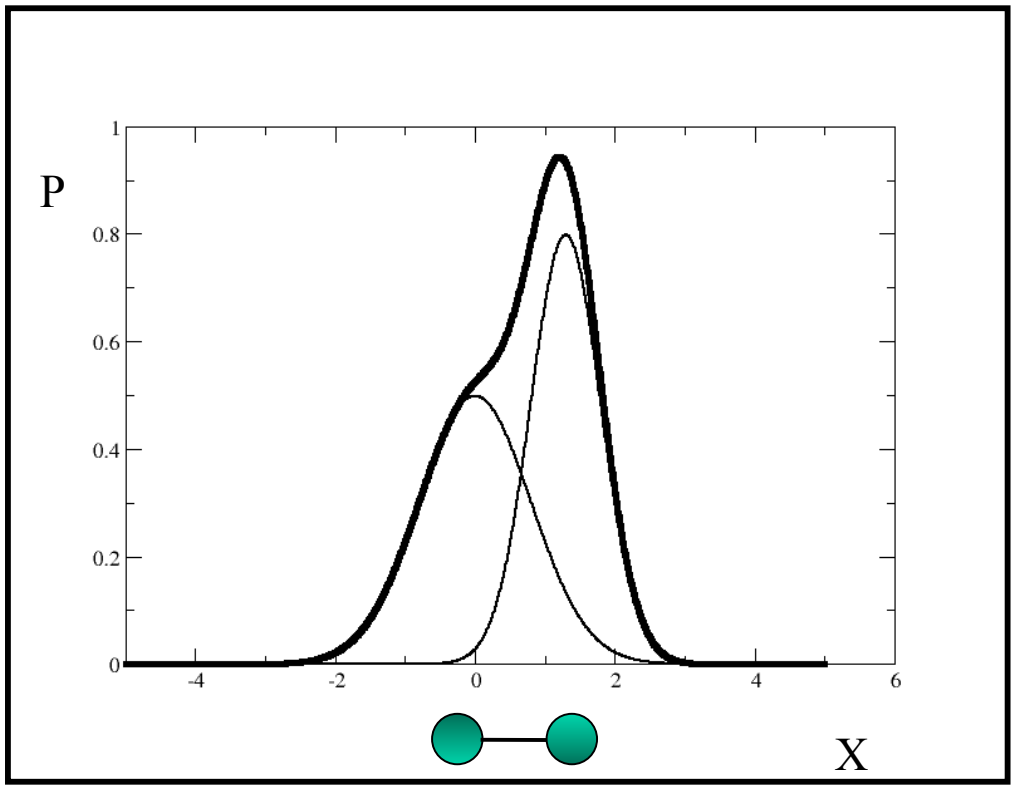
Some of the web and other resources

- 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/pdbe/
  - 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

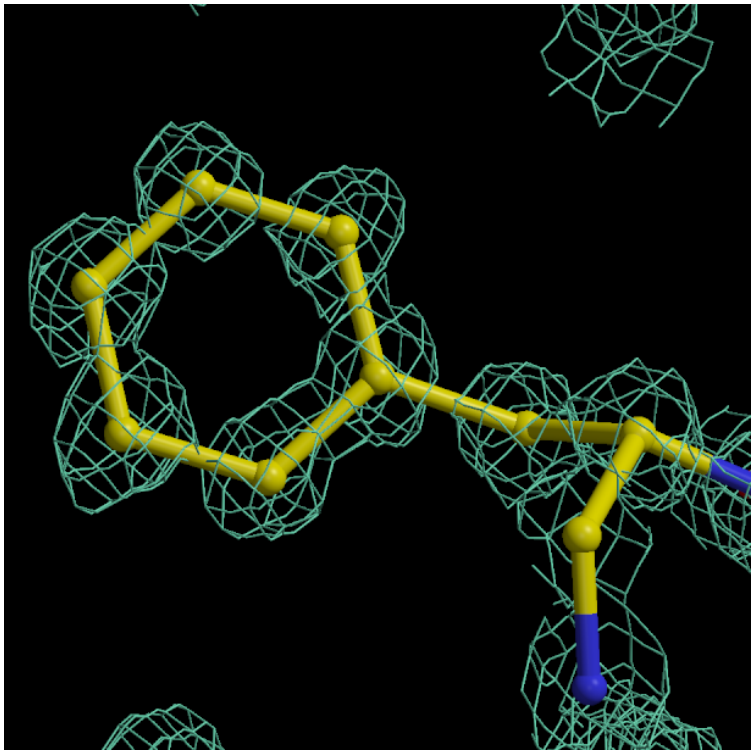
- Distance between atoms 1.3Å. B values 20 and 50



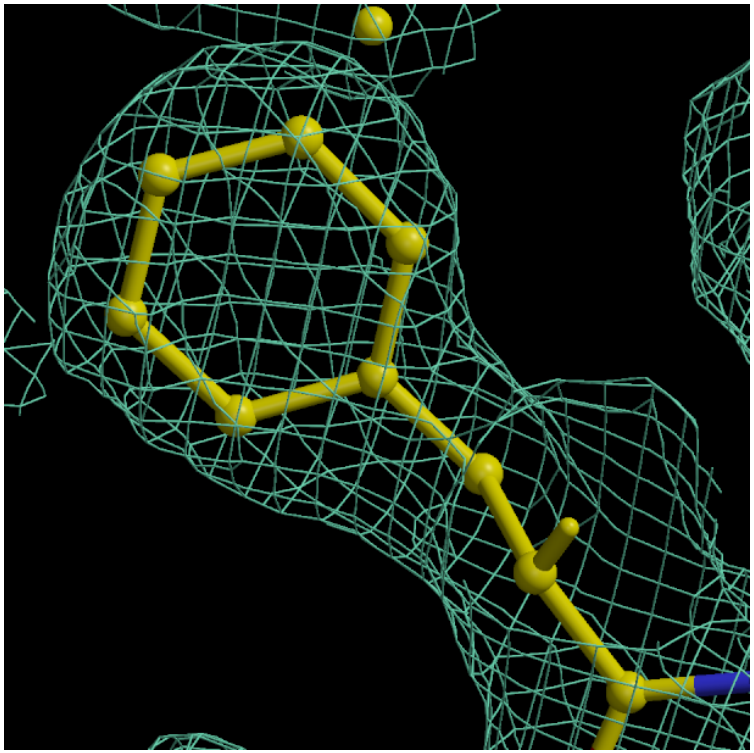
- Thin lines – single atoms
- Bold line - sum of the two atoms

Chemical information: Phe at two different resolutions

- 0.88 Å



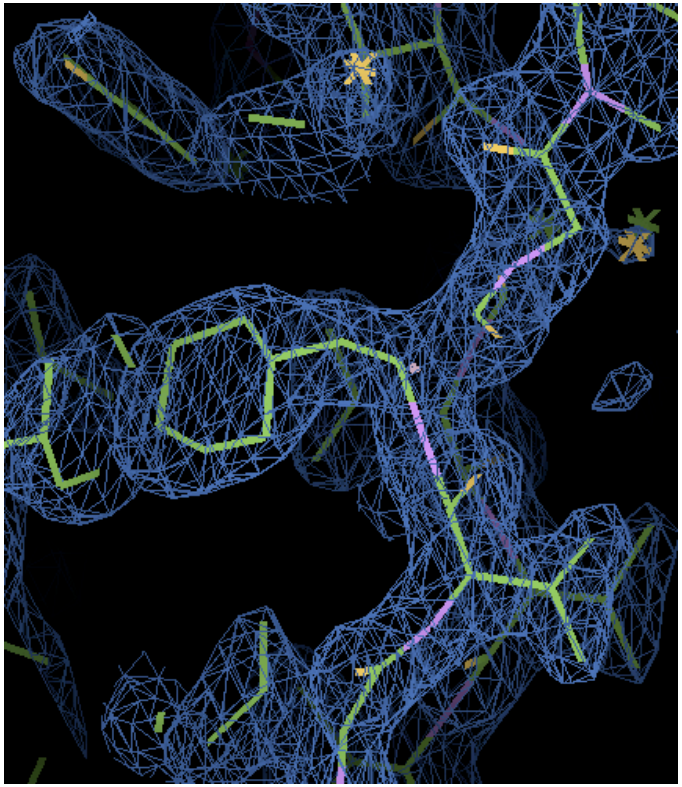
2 Å and High mobility



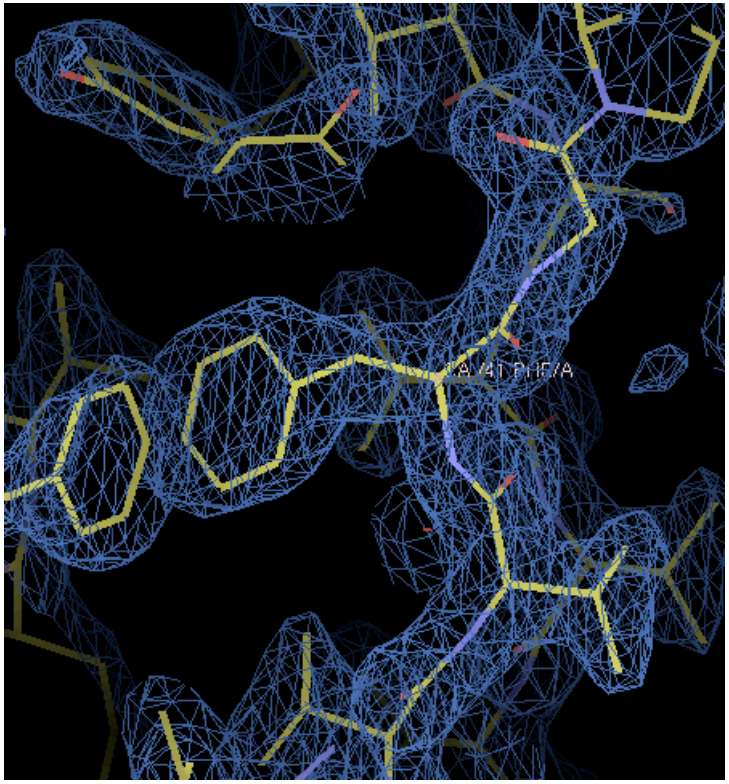
Example

- Data -
  1.9A

Unrestrained



Restrained

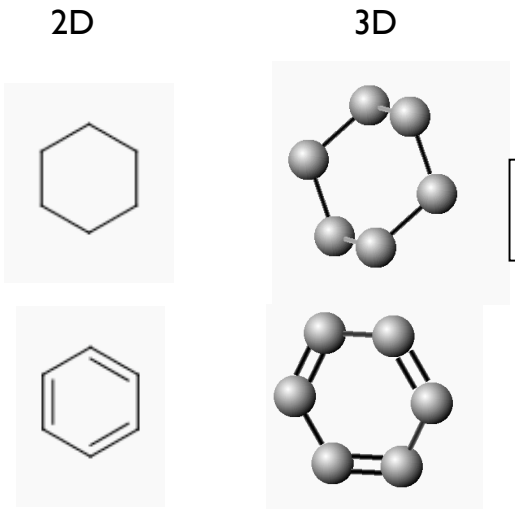


Role of restraints

- 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 to retain chemistry of molecule 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

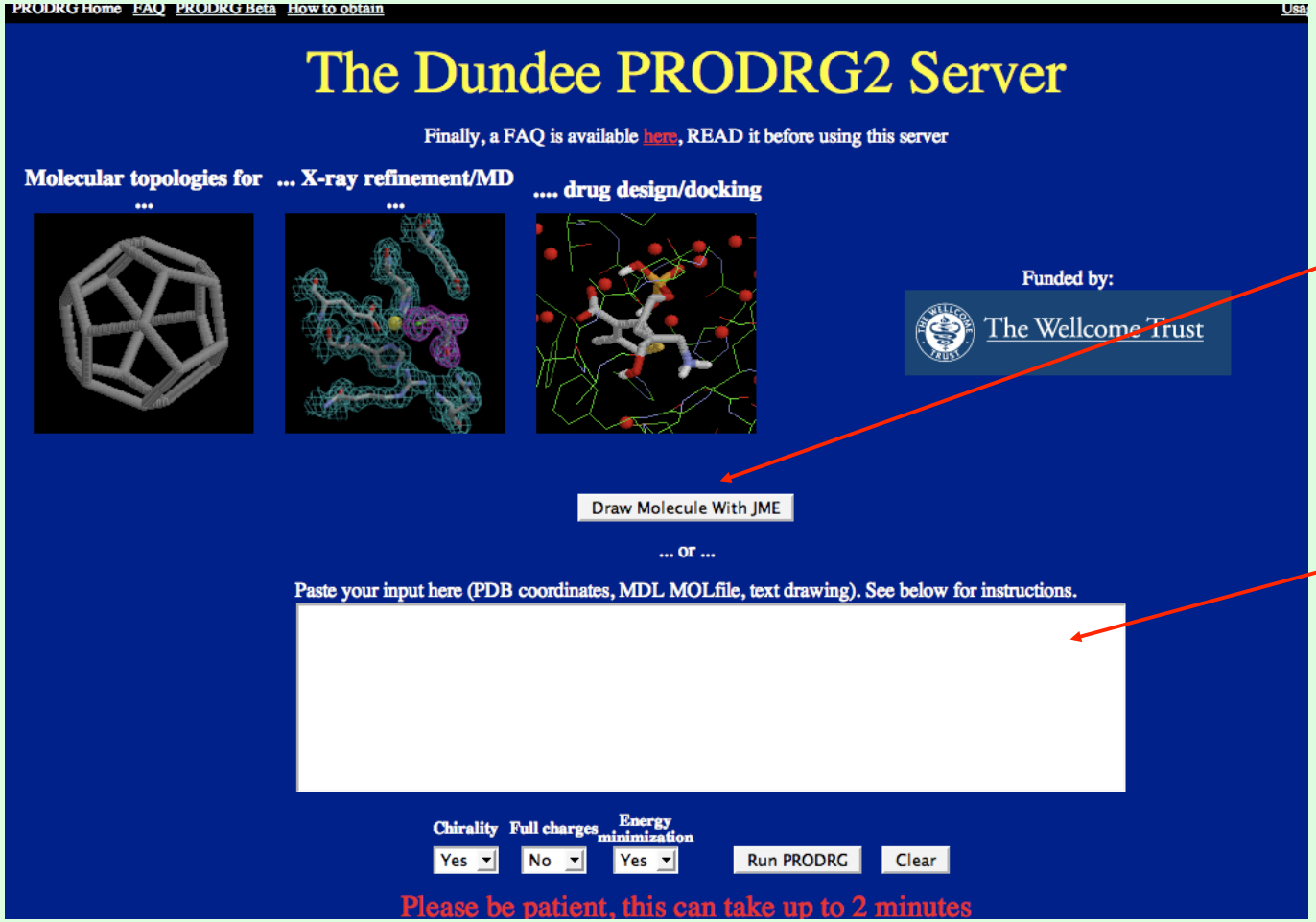
Using restraints

- Standard dictionary has description of around 10000 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.



These two molecules will refine very differently (oviously)

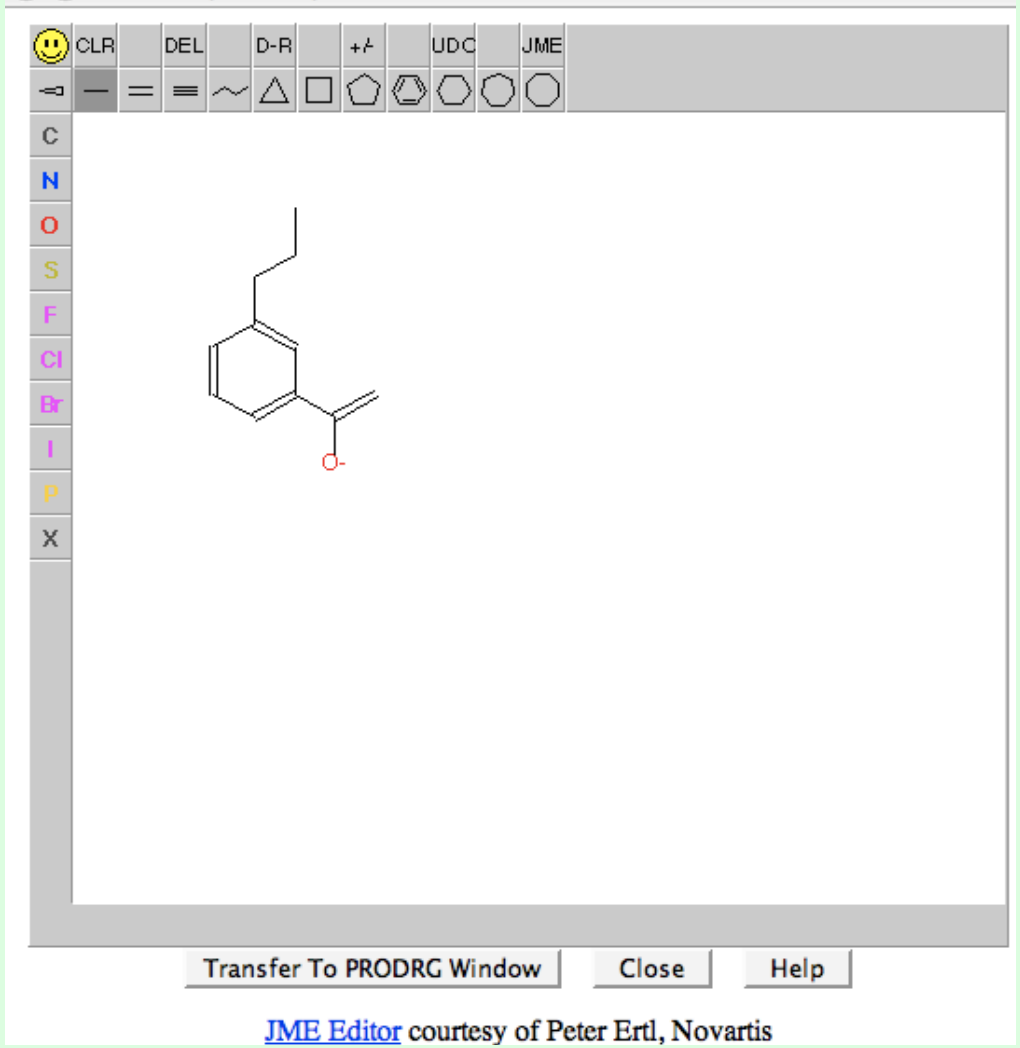
PRODRG server



JME

Load your file

PRODRG: JME



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

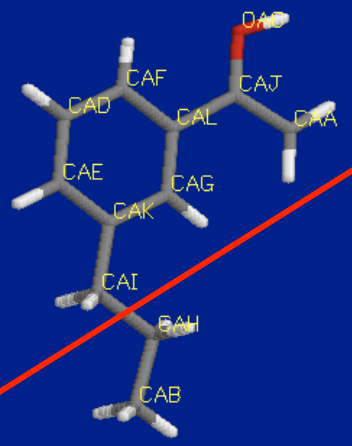
Draw your ligand, transfer to PRODRG window and run

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



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

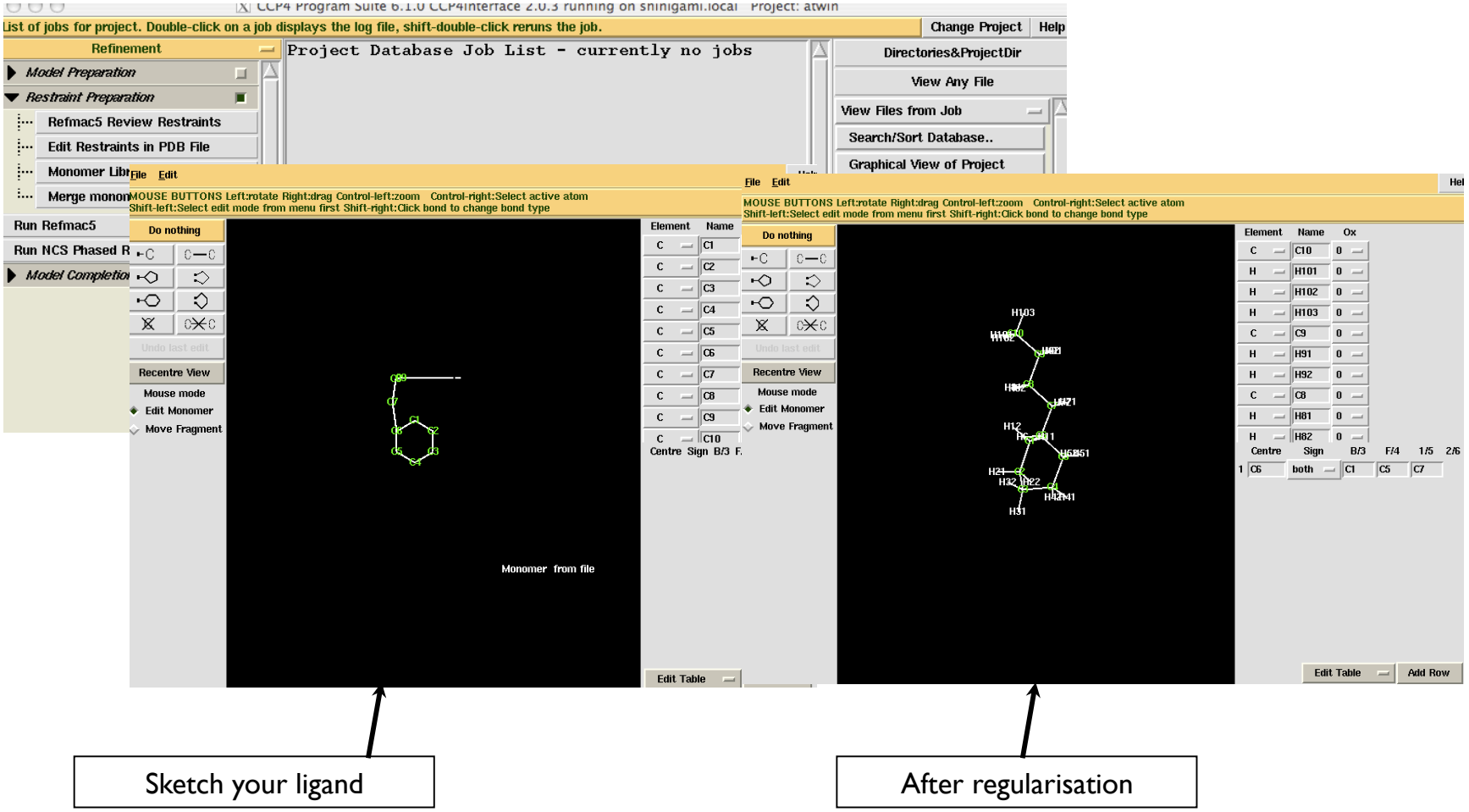
- CNS (parameters and topology)
- REFMAC5
- SHELX
- O (pre-9.x torsion entry, pre-9.x refi dictionary and 9.x dictionary)

Done

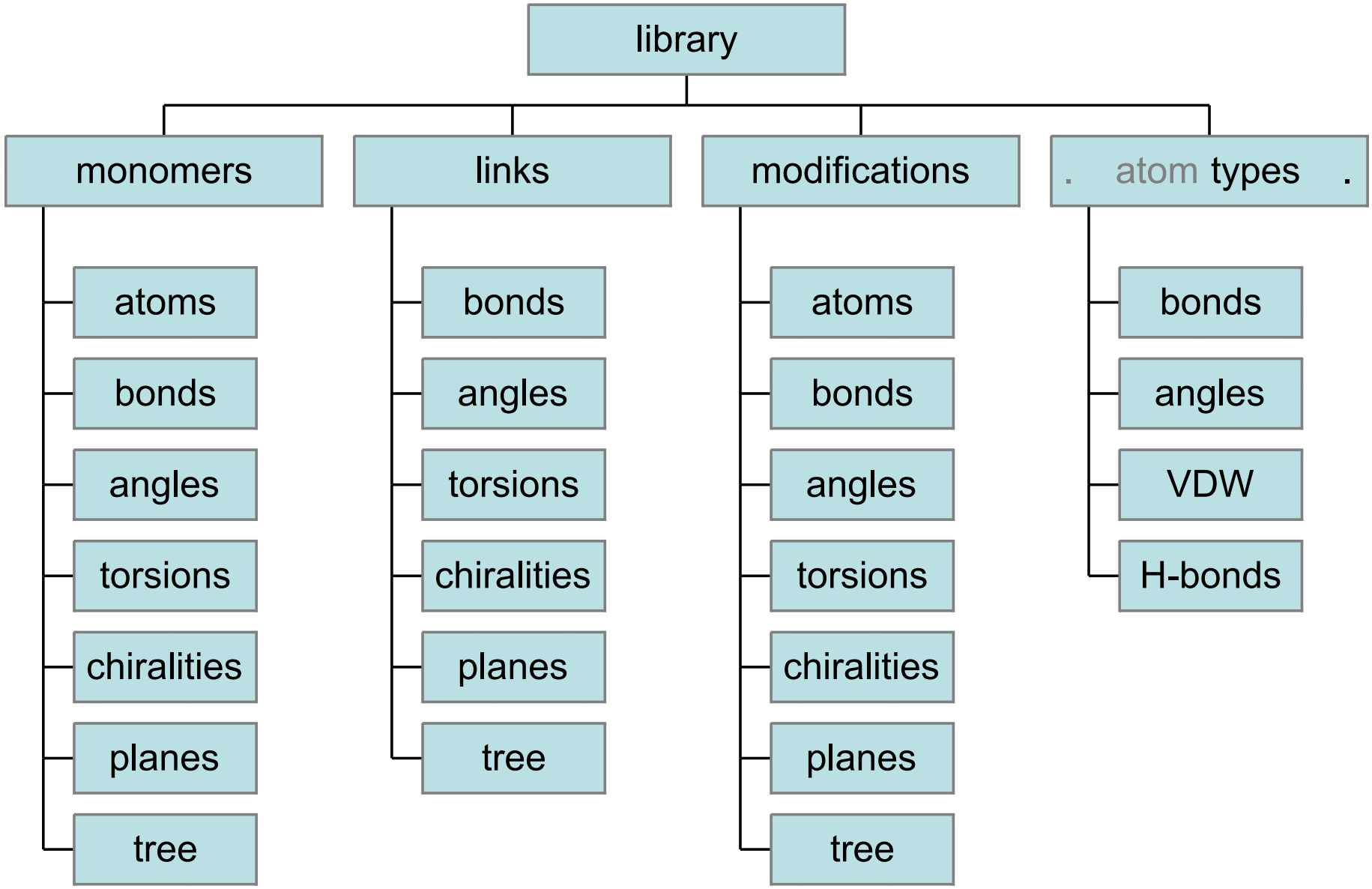
It can write out representation in various formats suitable for various popular software

Using resources from ccp4

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



CCP4 library of restraints



Monomers

A *monomer* entry describes an individual compound

CCP4 library contains:

- All amino acids

- All nucleic acids

- Common sugars

- Other organic and inorganic compounds:

in ccp4-6.13	2,500
in the next ccp4 release (atom names as in PDB-v3)	10,500

New monomer

CCP4 tools:

LIBCHECK - uses *atom types*
- creates *monomer* description from molecular graph
- creates coordinates from the monomer description

SKETCHER - GUI for LIBCHECK

MAKECIF - creates restraints for a particular macromolecule
- a separate program wrapping a subset of REFMAC subroutines

Modifications and links

The idea of this mechanism is that
- while *monomer* records describe individual compounds
- *modifications* and *links* describe changes resulted from chemical reactions

Modification formalism allows to change a monomer
Link formalism allows to join modified monomers together

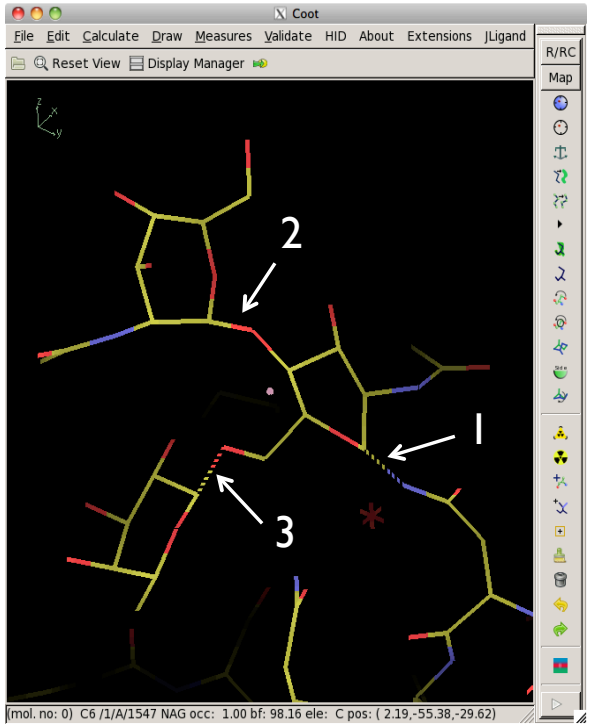
(details later)

CCP4 library contains modifications
- terminal peptides and nucleotides
- methylated nucleotides
- deprotonated states of some peptide side chains

CCP4 library contains links and corresponding modifications for:
- polypeptide chains (CIS, TRANS), S-S bridges
- polynucleotide chains
- glycosylated proteins

Standard links

2xmb



FAQ: How to create links for glycosylated proteins?

A: For typical glycosylation cases this is not needed.

- necessary modifications and links are there in the standard ccp4 library

- by default REFMAC uses these library descriptions

NAG – NAG – ASN
|
FUL

FUL = Beta-L-Fucose
NAG = N-Acetyl-D-Glucosamine

Standard links used here:

(1) "NAG-ASN"
(2) "BETA1-4"
(3) "BETA1-6"

New links

When new link descriptions are needed:

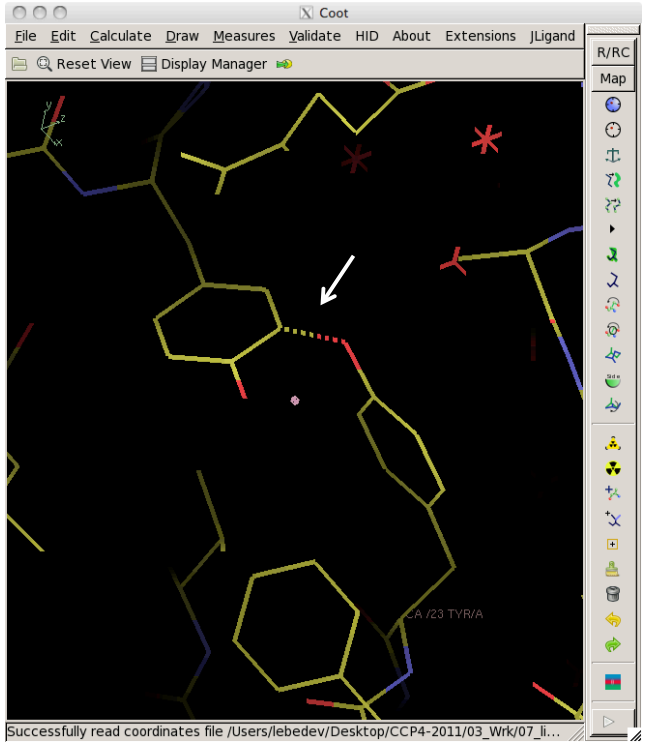
side chain – side chain (e.g. TYR – TYR on the figure)

side chain – main chain (e.g. LYS – Ubiquitin)

side chain – ligand (e.g. LYS – PLP)

JLigand:
- new GUI for LIBCHECK
- descriptions of monomers (functionality of SKETCHER)
- descriptions of links and corresponding modifications

TYR–TYR covalent link in
M. tuberculosis Hemoglobin O
PDB id 1ngk



JLigand: a graphical editor to create ligand and link descriptions

Andrey Lebedev, Paul Young, Alexei Vagin, Garib Murshudov

It works with java 1.5 or later versions

New link

Example:

– covalent linkage between LYS and Pyridoxal phosphate (PLP).

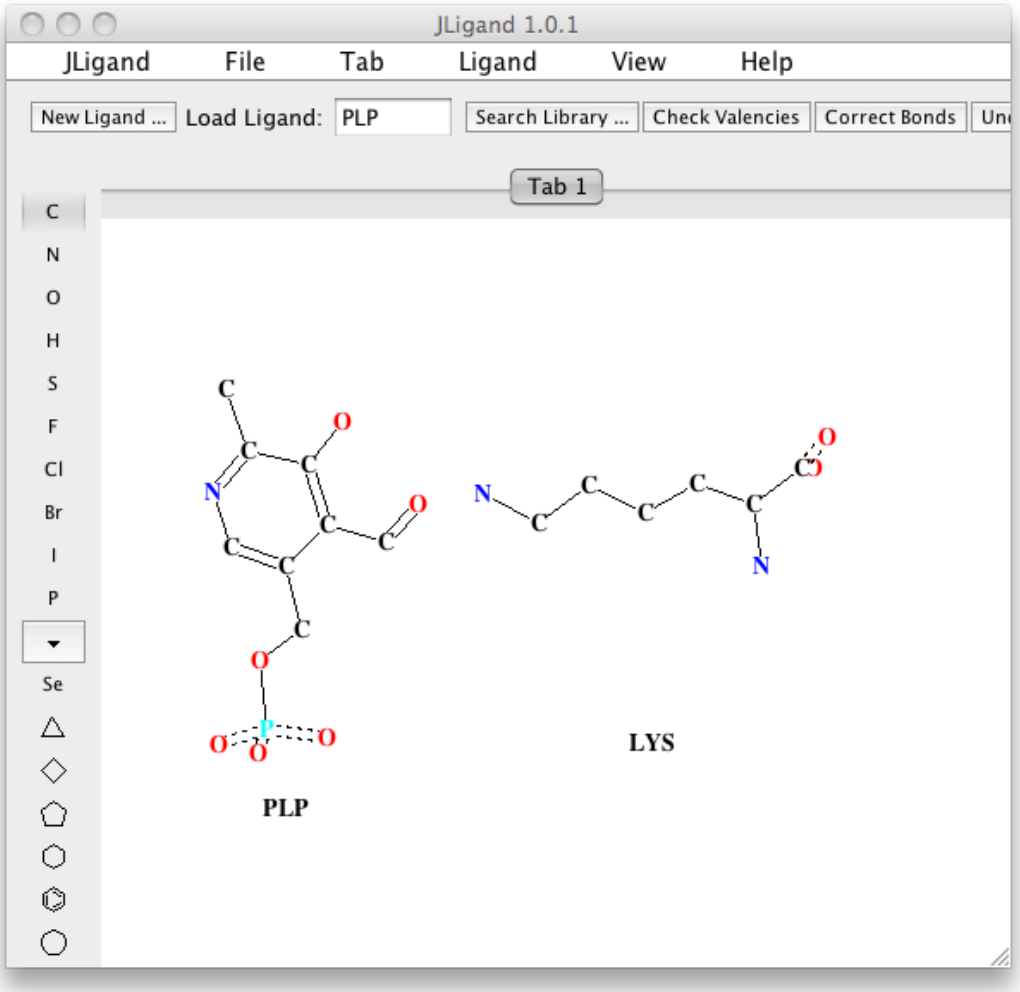
– describes PLP forming internal aldimine in aminotransferases.

Given:

- descriptions of LYS and PLP
  from the standard library

Needed:

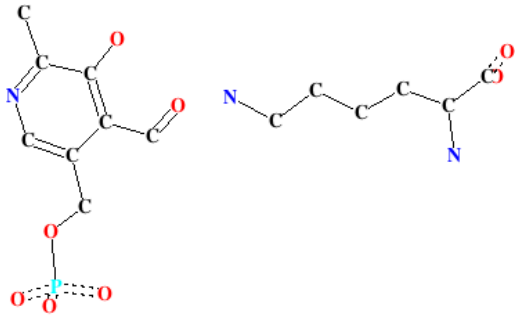
- additional library file with
  the description of link LYS–PLP



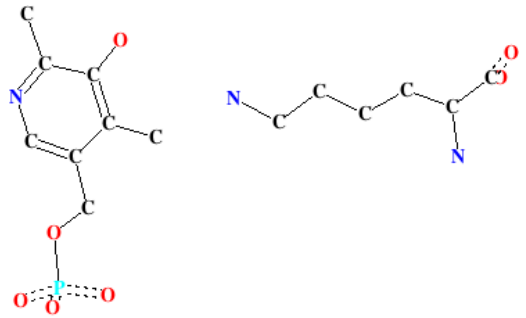
Creating a new link, as seen in JLigand GUI

The two monomers are in effect reacted in silico
Hydrogen atoms are dealt with automatically*)

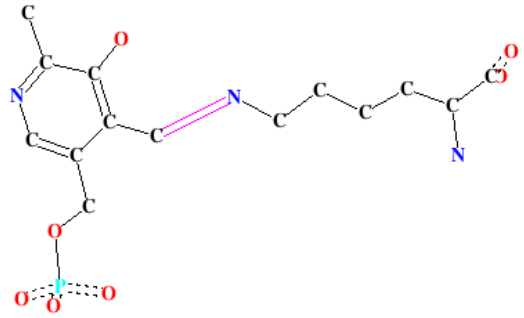
* it is also possible to visualise H-atoms and dealt with them explicitly



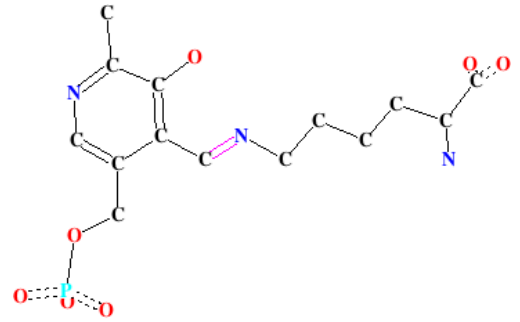
leaving O is removed



double bond C=N is added

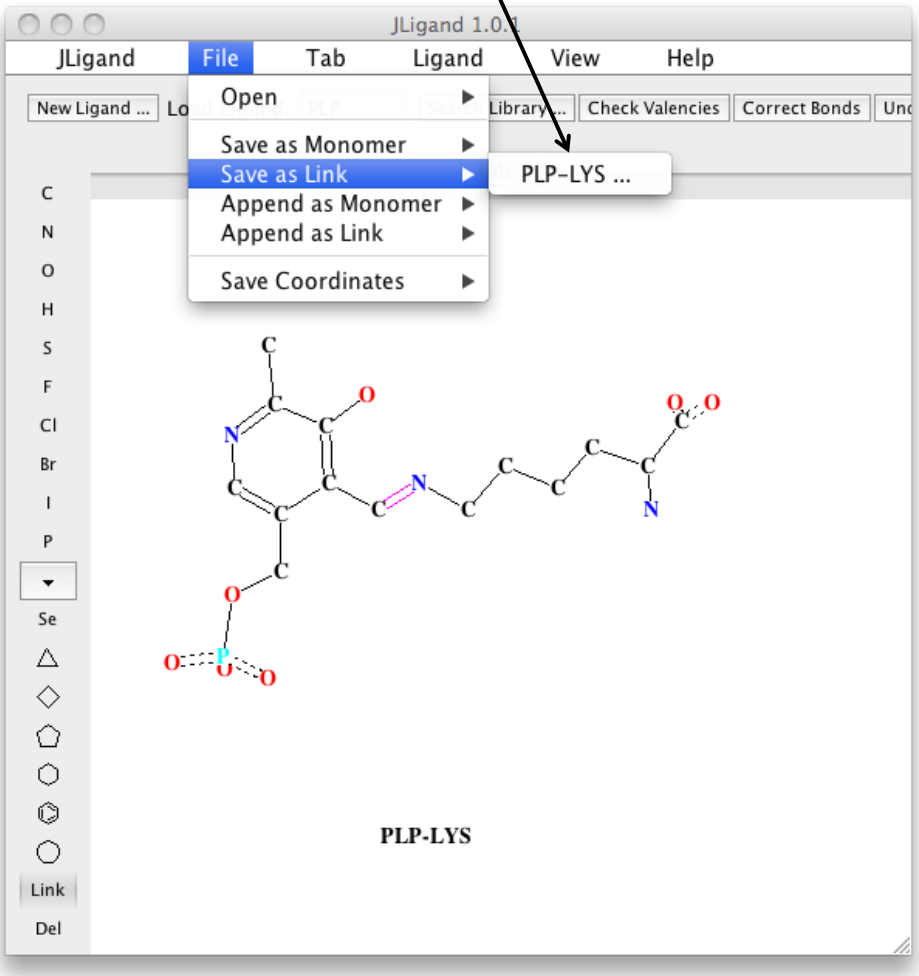


the new compound
is regularised



The new link, "file view"

To save into CIF-file (additional library)



Contents:
(1) modification "PLPmod1"
(2) modification "LYSmod1"
(3) link "PLP-LYS"
No monomers

1
2
3

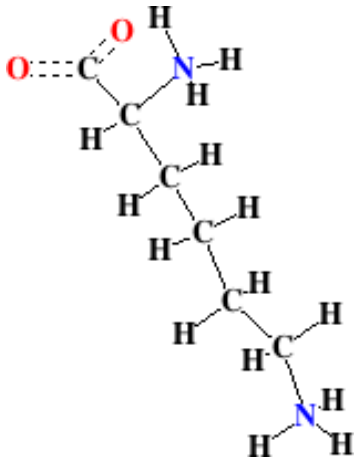
PLP-LYS

```
data_mod_list

loop_
_chem_mod.id
_chem_mod.name
_chem_mod.comp_id
_chem_mod.group_id
PLPmod1  "PYRIDOXAL-5'-PHOSPHATE          "       PLP     .
LYSmod1  'LYSINE                          '       LYS     .

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_2
_chem_link.name
PLP-LYS  PLP     PLPmod1 .       LYS     LYSmod1 .
 PLP-LYS
```

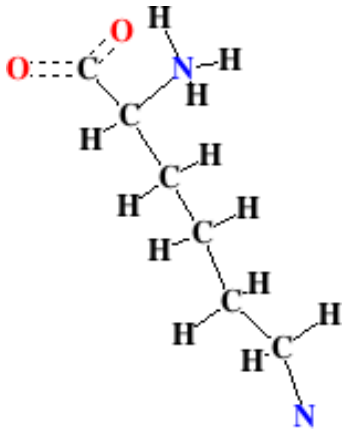
Close

The new link, "file view"

LYS



LYSmod1



Modification "LYSmod1": changes to LYS

```
data_mod_LYSmod1

loop_
_chem_mod_atom.mod_id
_chem_mod_atom.function
_chem_mod_atom.atom_id
_chem_mod_atom.new_atom_id
_chem_mod_atom.new_type_symbol
_chem_mod_atom.new_type_energy
_chem_mod_atom.new_partial_charge
 LYSmod1  change NZ    .    .  N      0.000
 LYSmod1  delete HZ1   .    .  .      .
 LYSmod1  delete HZ2   .    .  .      .
 LYSmod1  delete HZ3   .    .  .      .
```

Atoms

```
loop_
_chem_mod_bond.mod_id
_chem_mod_bond.function
_chem_mod_bond.atom_id_1
_chem_mod_bond.atom_id_2
_chem_mod_bond.new_type
_chem_mod_bond.new_value_dist
_chem_mod_bond.new_value_dist_esd
 LYSmod1  change CE   NZ     .      1.455   0.020
 LYSmod1  delete NZ   HZ3    .      .       .
 LYSmod1  delete NZ   HZ2    .      .       .
 LYSmod1  delete NZ   HZ1    .      .       .
```

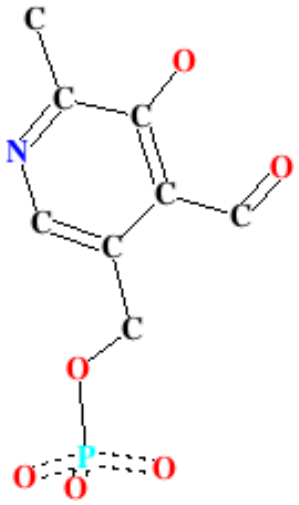
Bonds

Angles

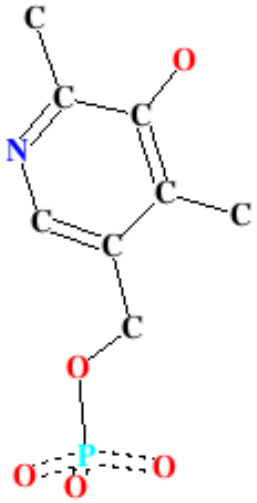
.....

The new link, "file view"

PLP



PLPmod1



```
data_mod_PLPmod1

loop_
_chem_mod_atom.mod_id
_chem_mod_atom.function
_chem_mod_atom.atom_id
_chem_mod_atom.new_atom_id
_chem_mod_atom.new_type_symbol
_chem_mod_atom.new_type_energy
_chem_mod_atom.new_partial_charge
 PLPmod1  delete O4A    .     .     .          .

loop_
_chem_mod_bond.mod_id
_chem_mod_bond.function
_chem_mod_bond.atom_id_1
_chem_mod_bond.atom_id_2
_chem_mod_bond.new_type
_chem_mod_bond.new_value_dist
_chem_mod_bond.new_value_dist_esd
 PLPmod1  delete C4A    O4A      .          .      .

loop_
_chem_mod_angle.mod_id
_chem_mod_angle.function
_chem_mod_angle.atom_id_1
_chem_mod_angle.atom_id_2
_chem_mod_angle.atom_id_3
_chem_mod_angle.new_value_angle
_chem_mod_angle.new_value_angle_esd
 PLPmod1  delete H4A    C4A    O4A        .      .
 PLPmod1  delete C4     C4A    O4A        .      .

loop_
_chem_mod_plane_atom.mod_id
_chem_mod_plane_atom.function
_chem_mod_plane_atom.plane_id
_chem_mod_plane_atom.atom_id
_chem_mod_plane_atom.new_dist_esd
 PLPmod1  delete plan-2    C4          .
 PLPmod1  delete plan-2    C4A         .
 PLPmod1  delete plan-2    H4A         .
 PLPmod1  delete plan-2    O4A         .
```

Modification "PLPmod1": changes to PLP

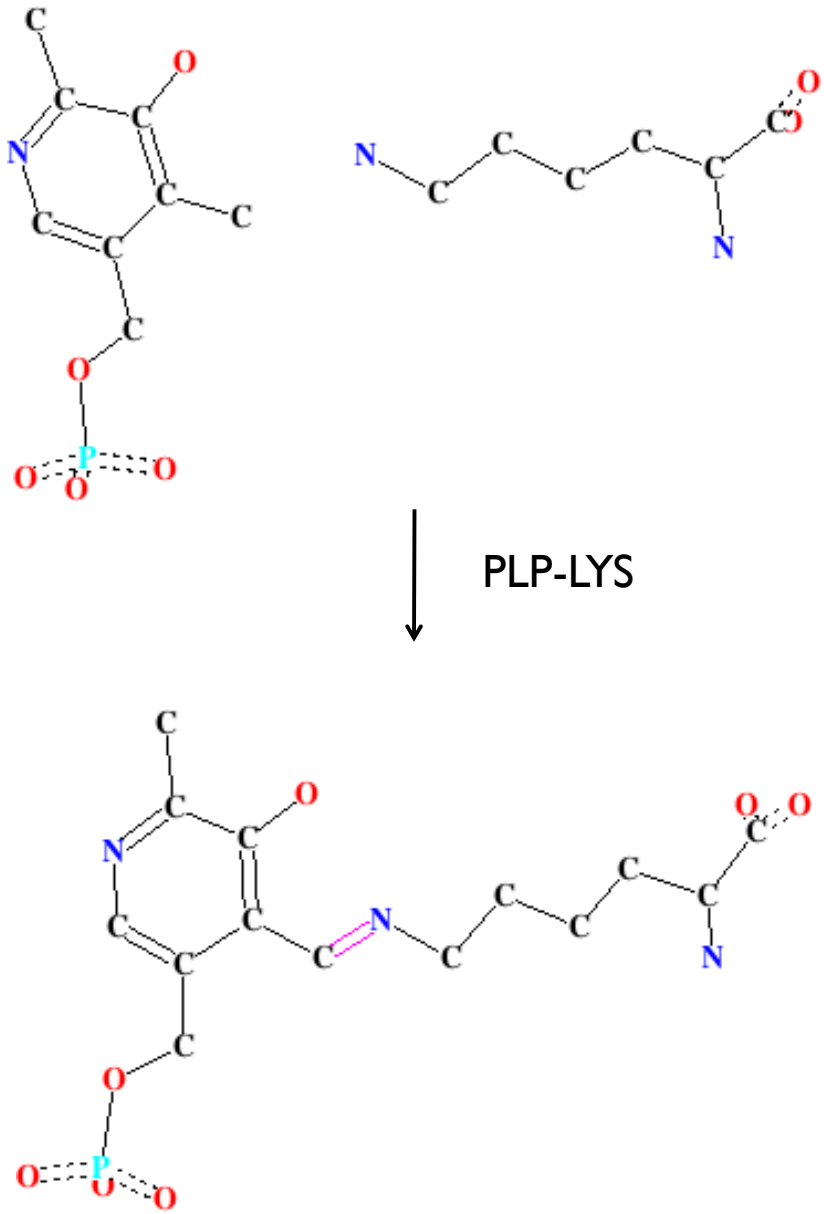
Atom

Bond

Angles

Plane

The new link, "file view"



Link "PLP-LYS": changes associated with covalent linkage between modified PLP and LYS

```
data_link_PLP-LYS

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
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
 PLP-LYS  1 C4A    2 NZ       double      1.260    0.020
```

Bond

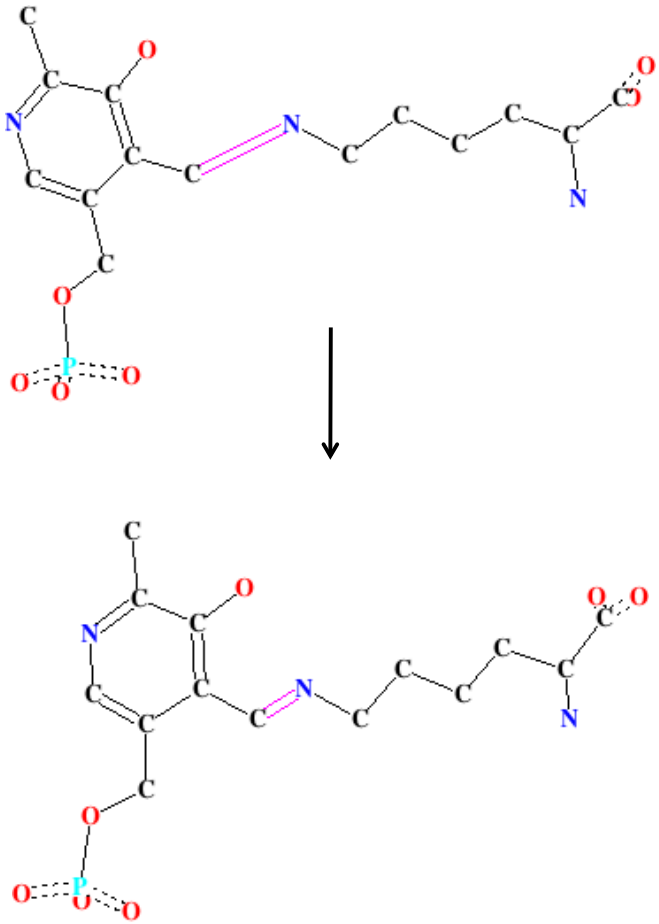
```
loop_
_chem_link_angle.link_id
_chem_link_angle.atom_1_comp_id
_chem_link_angle.atom_id_1
_chem_link_angle.atom_2_comp_id
_chem_link_angle.atom_id_2
_chem_link_angle.atom_3_comp_id
_chem_link_angle.atom_id_3
_chem_link_angle.value_angle
_chem_link_angle.value_angle_esd
 PLP-LYS  1 C4A    2 NZ    2 CE     120.000    3.000
 PLP-LYS  1 H4A    1 C4A   2 NZ     120.000    3.000
 PLP-LYS  1 C4     1 C4A   2 NZ     120.000    3.000
```

Angles

```
loop_
_chem_link_plane.link_id
_chem_link_plane.plane_id
_chem_link_plane.atom_comp_id
_chem_link_plane.atom_id
_chem_link_plane.dist_esd
 PLP-LYS  plan-2    1 C4       0.020
 PLP-LYS  plan-2    1 C4A      0.020
 PLP-LYS  plan-2    1 H4A      0.020
 PLP-LYS  plan-2    2 CE       0.020
 PLP-LYS  plan-2    2 NZ       0.020
```

Plane

Regularisation



– The molecular graph of the total compound (not coordinates!) is loaded into LIBCHECK

– LIBCHECK generates target values and sigmas for restraints

– LIBCHECK generates initial coordinates

– REFMAC ("mode newentry", i.e. no X-ray data) refines these initial coordinates using restraints from LIBCHECK

– The linked compound is displayed in JLigand GUI

– On request ("View" or "Save") but not before regularisation, the description of the total compound is split into two modifications and one link

– Descriptions of original non-modified monomers are discarded when they are available from the standard library

Utilising new link description

Good news: there is no need to manually edit the additional CIF-library, even to see its contents. JLigand does the job.

Three remaining steps:

– docking monomer(s) into electron density

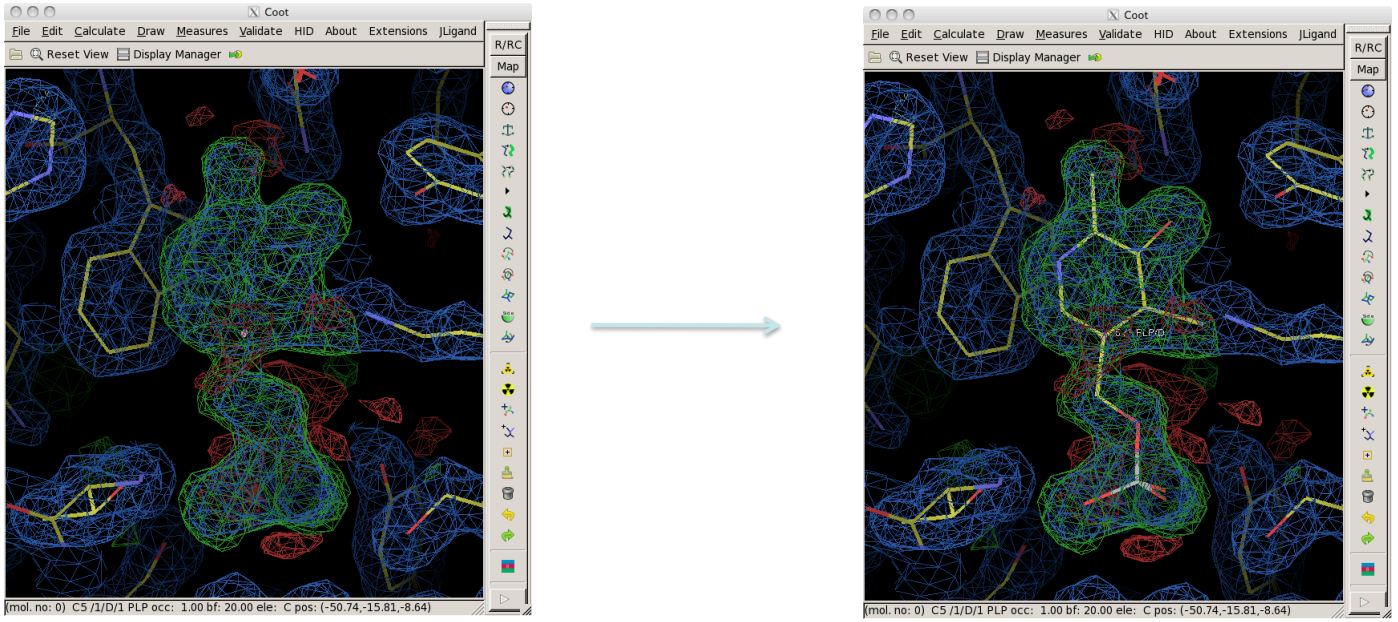
– defining link in the pdb-file

– refinement of the structure with linked ligand using additional library

Docking into the electron density

In our example, this is completely independent step: the additional library is not used.

– non-modified monomer is taken from the standard library
– docking is performed, e.g. using coot:

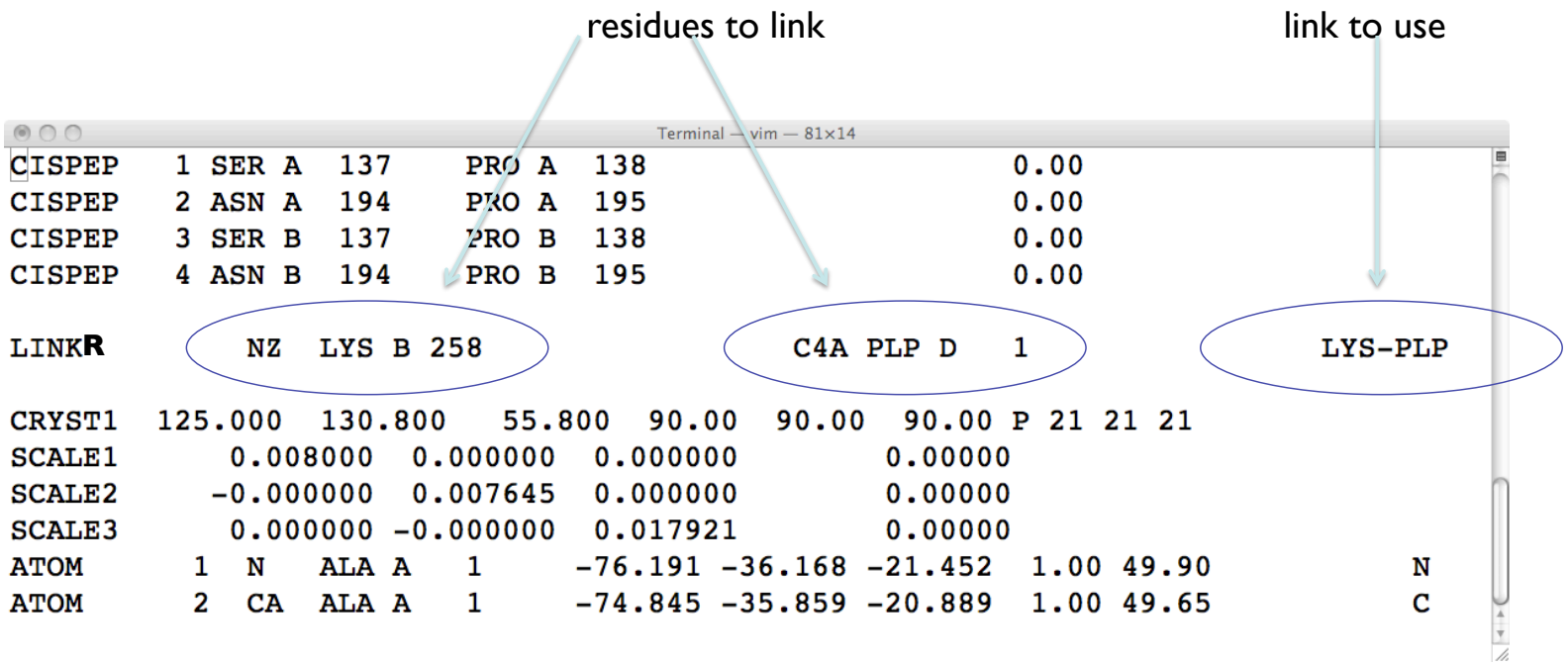


– leaving atoms (O4A of PLP in this example) are removed
– in our example, one of the monomers (LYS) is already in the model

Defining link in the pdb-file

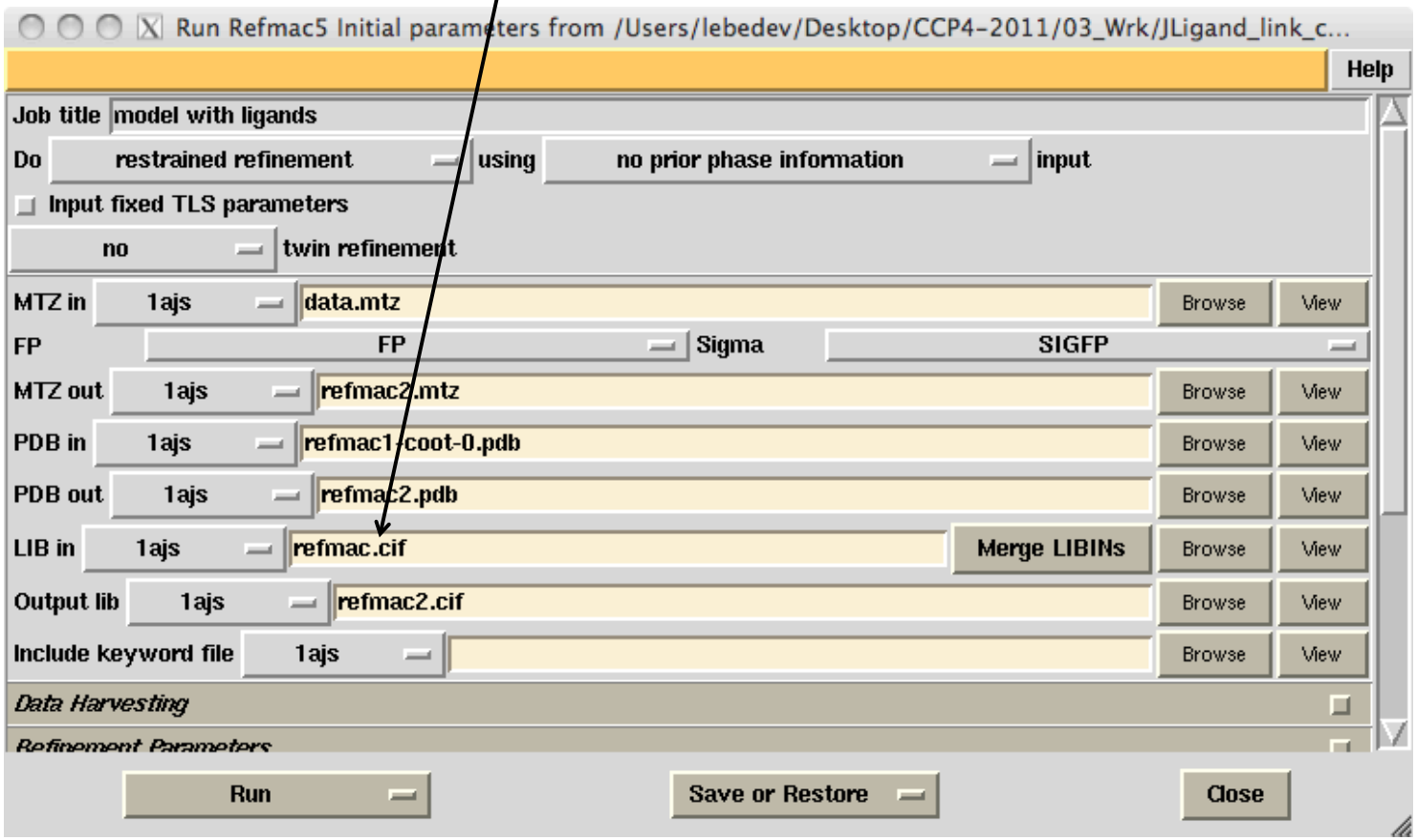
In general case, link cannot be applied automatically.
For example:
– e.g. the same two atoms of the same two compounds can form single or double bond
– H-atom are not defined in the PDB-file

Therefore REFMAC needs additional instructions:



Refinement using additional library

Additional library is defined here



CCP4 library of restraints

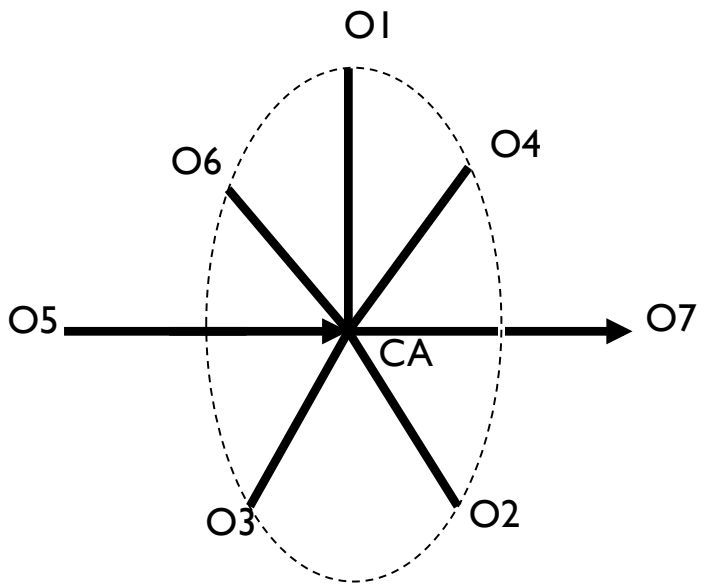
JLigand: new link

JLignad: metal coordination

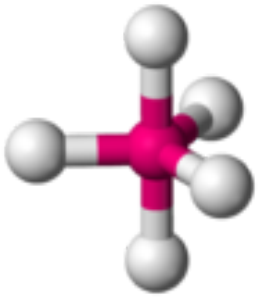
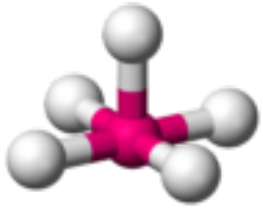
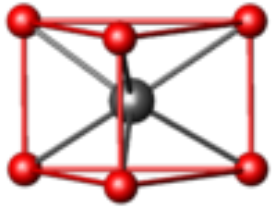
Metal chiralities in LIBCHECK

A fragment of CIF-library:

```
loop_
_chem_comp_chir.comp_id
_chem_comp_chir.id
_chem_comp_chir.atom_id_centre
_chem_comp_chir.atom_id_1
...
_chem_comp_chir.atom_id_8
_chem_comp_chir.volume_sign
OC7 chir_01 CA O5 O7 O1 O4 O2 O3 O6   . cross5
```



Metal chiralities in LIBCHECK

trigonal bipyramidal	square pyramidal	pentagonal bipyramidal	trigonal prismatic	
				...
"cross3"	"cross4" with one atom missing	"cross5"	cannot handle	

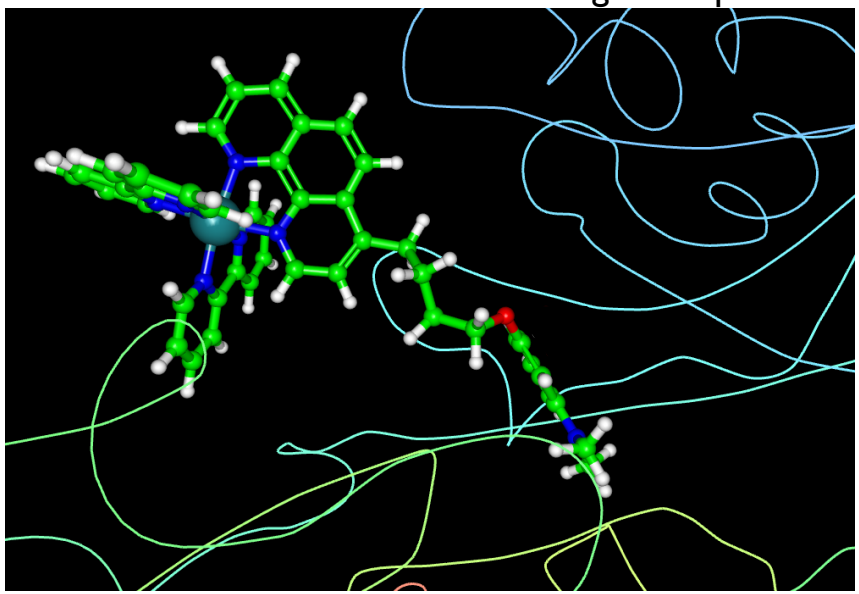
LIBCHECK can generate restraints for some of the coordination geometries:

– from crossX chirality defined in input CIF-file
– from 3D coordinates; this option is used in JLignad.

Example octahedral coordination

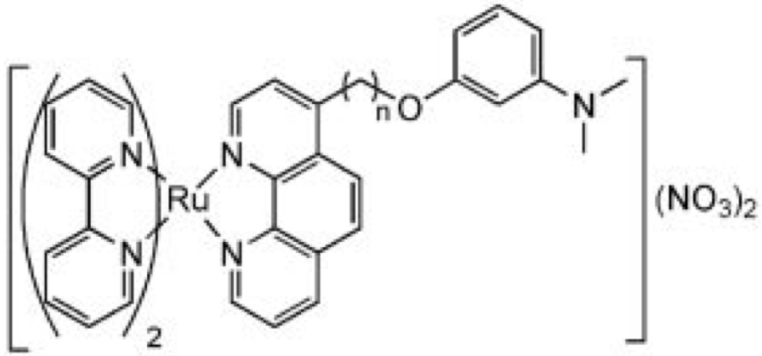
Ruthenium(II) "molecular wire",
an inhibitor of a copper amine oxidase,
PDB code 2BT3

RCSB - Ligand Explorer



3D-editing is essential here

2D representation:

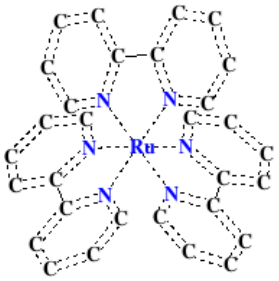


Coordination of Ru is not obvious:
– octahedral?
– trigonal prismatic?

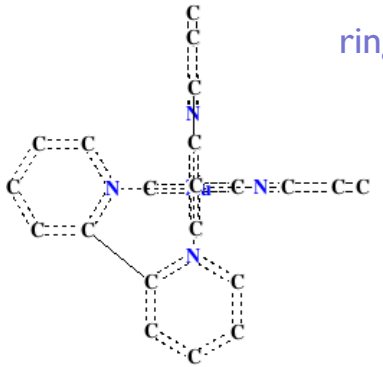
Example of 3-D editing

– Rings have the same Z-value as the atoms to which they are added

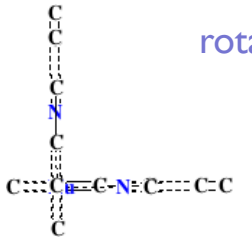
– In addition, positions of individual atoms can be adjusted



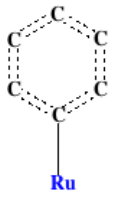
regularised



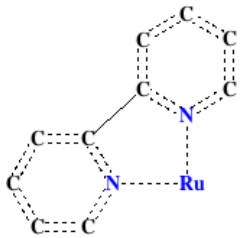
rings added



rotated 90o

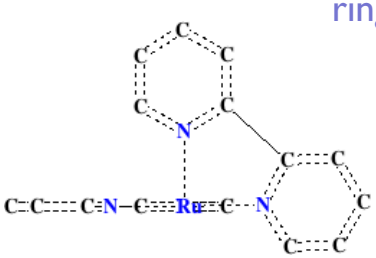


ring added



ring added

rotated 90o

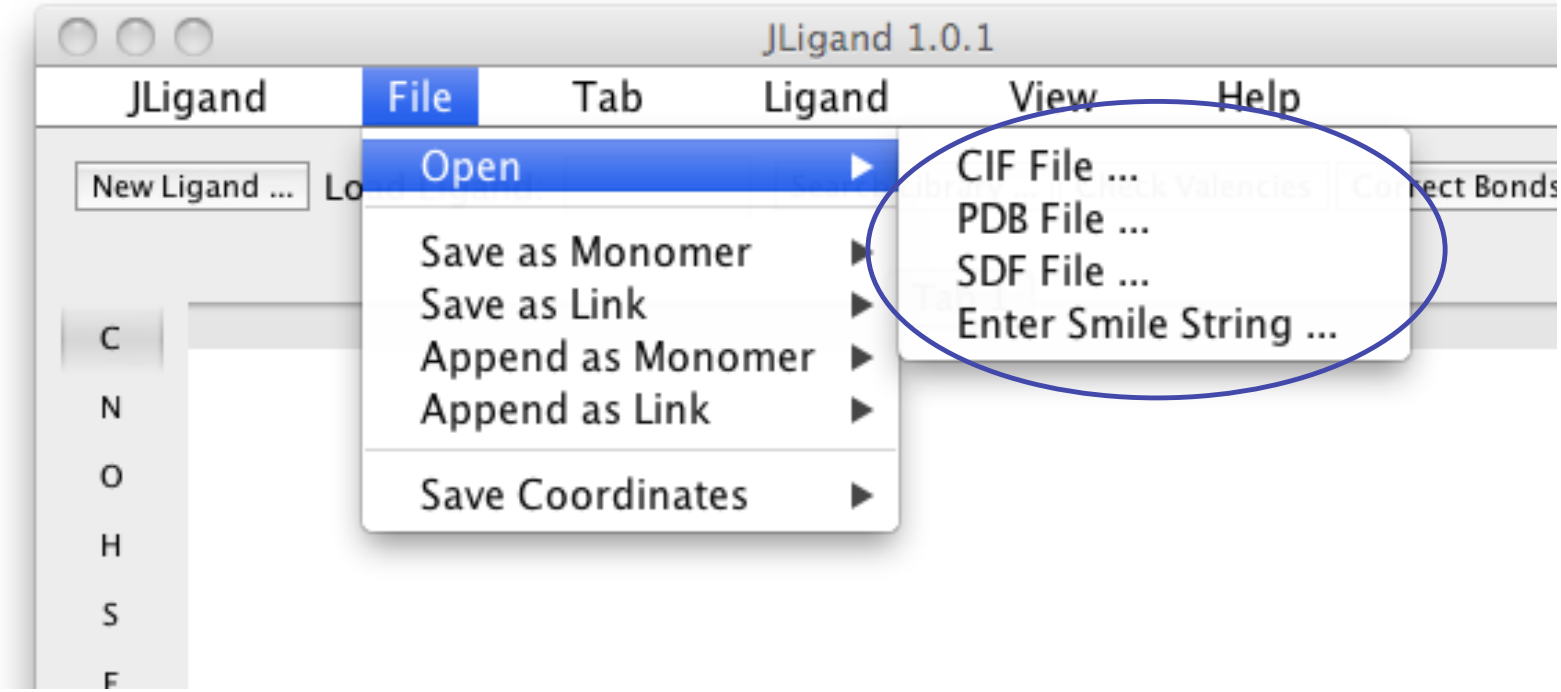


rings added

Input formats

The molecular graph for a new ligand can be drawn:
– from scratch as in previous example
– starting from imported molecular graph (mmCIF, SDF, SMILE)
– starting from molecular graph created from imported coordinates (PDB)

As for the most of other tasks, the actual job is done by LIBCHECK
while JLigand provides a GUI:



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:

- Atomic symbols used for atoms
- Hydrogen atoms as a rule are implicit. They are deduced using valence information about atoms
- Neighbouring atoms stand one after another
- Single, double, triple and aromatic bonds are denoted using “-”, “=”, “#” and “:” respectively. Single and aromatic bonds are usually not shown.

Branches represented by parentheses

Cycles are added by using matching digits on connecting atoms

- Aromatic atoms are denoted using lower cases.

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

Ligand tutorials and JLIgand:
www.ysbl.york.ac.uk/mxstat/Jligand

or search using google for jligand