

Dictionary of ligands

Some of the web and other resources

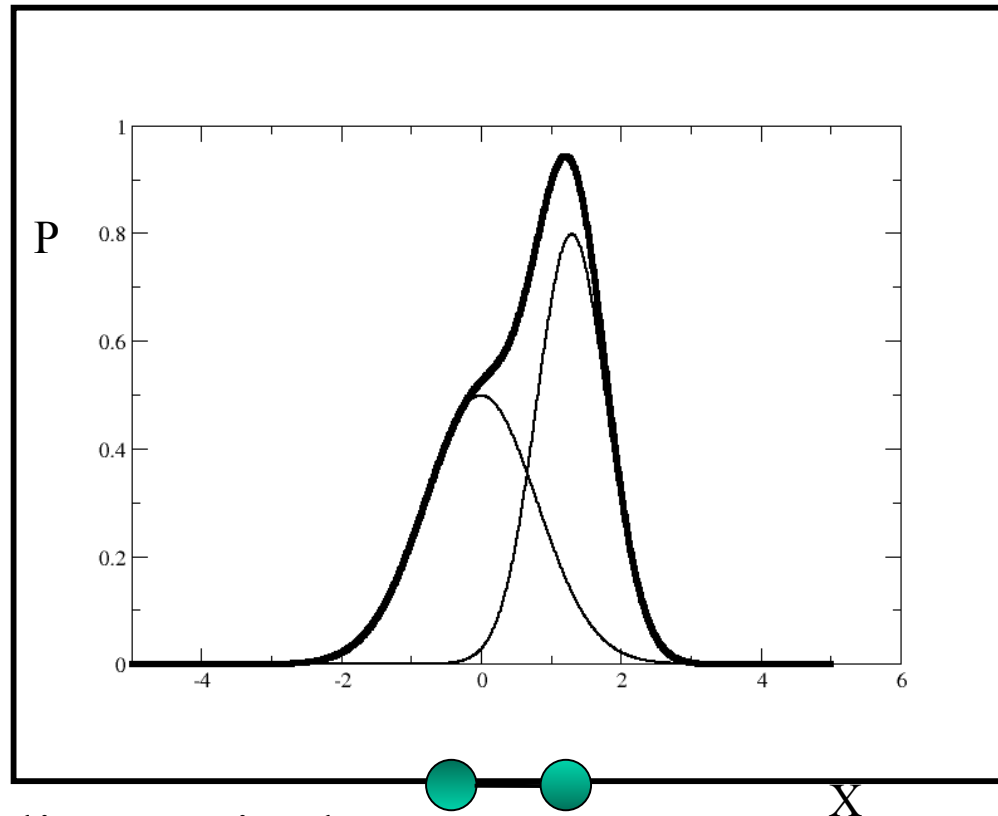
- Small molecules
- DrugBank: <http://www.drugbank.ca/>
- ZINC: <http://zinc.docking.org/index.shtml>
- PRODRUG: http://www.compbio.dundee.ac.uk/Web_Servers/prodrg_down.html
- CACTVS: <http://www2.chemie.uni-erlangen.de/software/cactvs/>
- Cambridge structural database - CSD: <http://www.ccdc.cam.ac.uk/products/csd/>

- 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

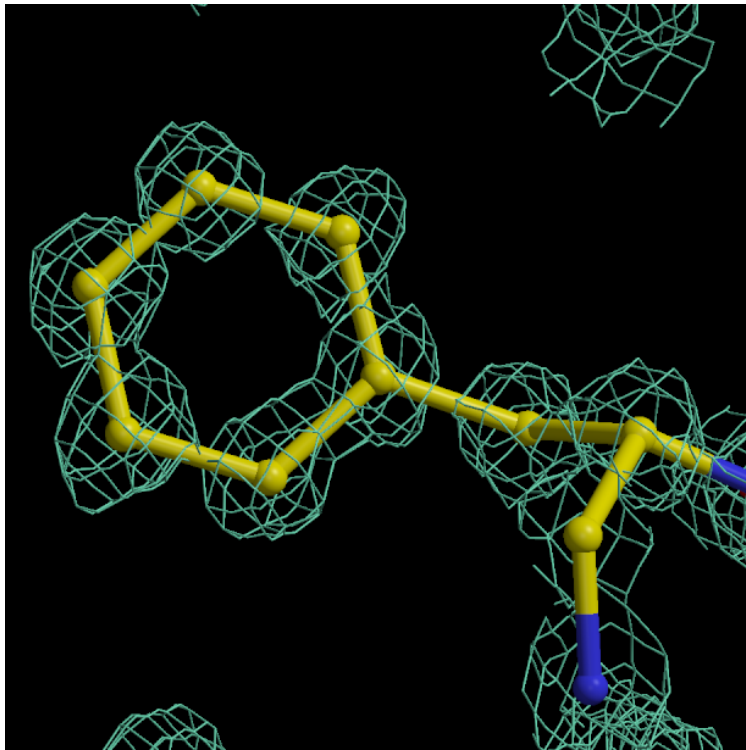
- Distance between atoms 1.3Å. B values 20 and 50



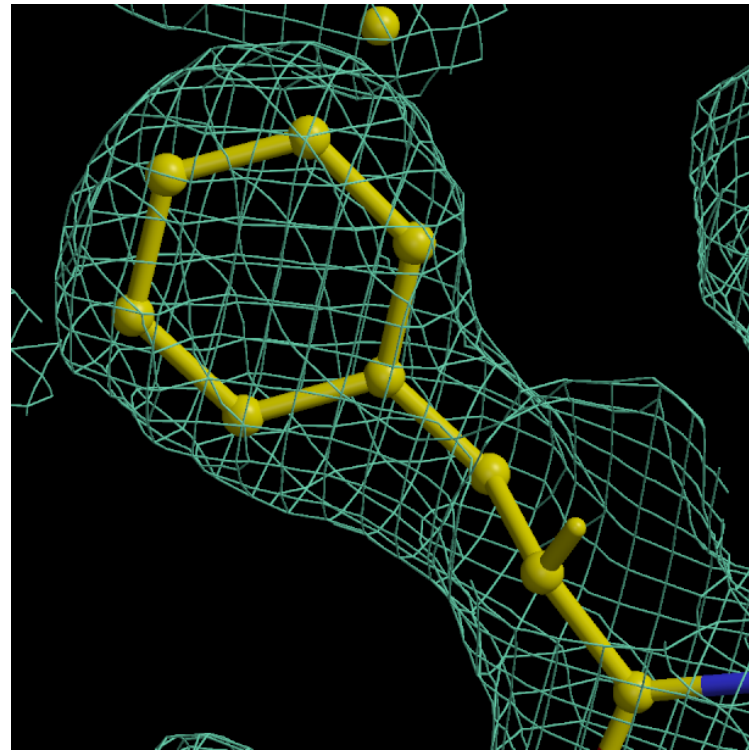
- Thin lines – single atoms
- Bold line - sum of the two atoms

Chemical information: Phe at two different resolutions

- 0.88 Å



2 Å and High mobility



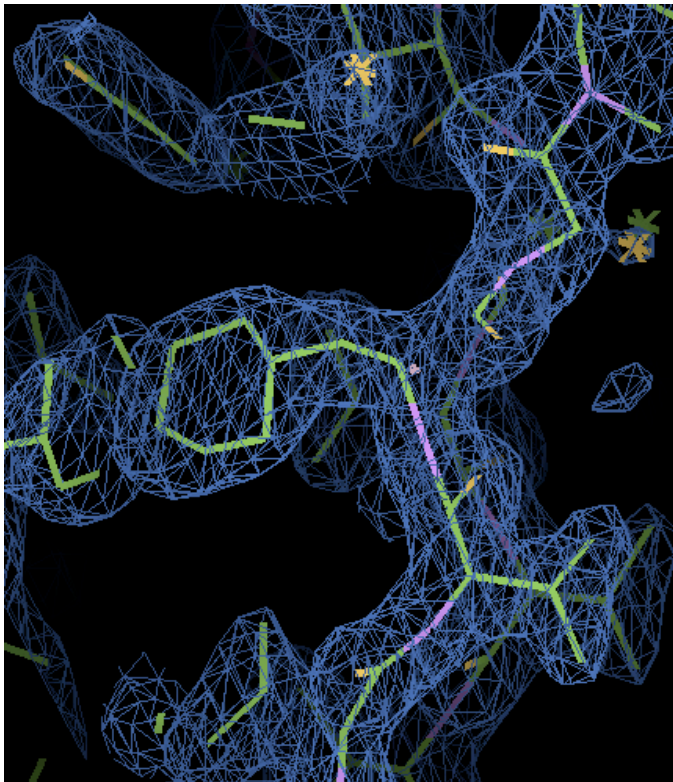
Role of restraints

- 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

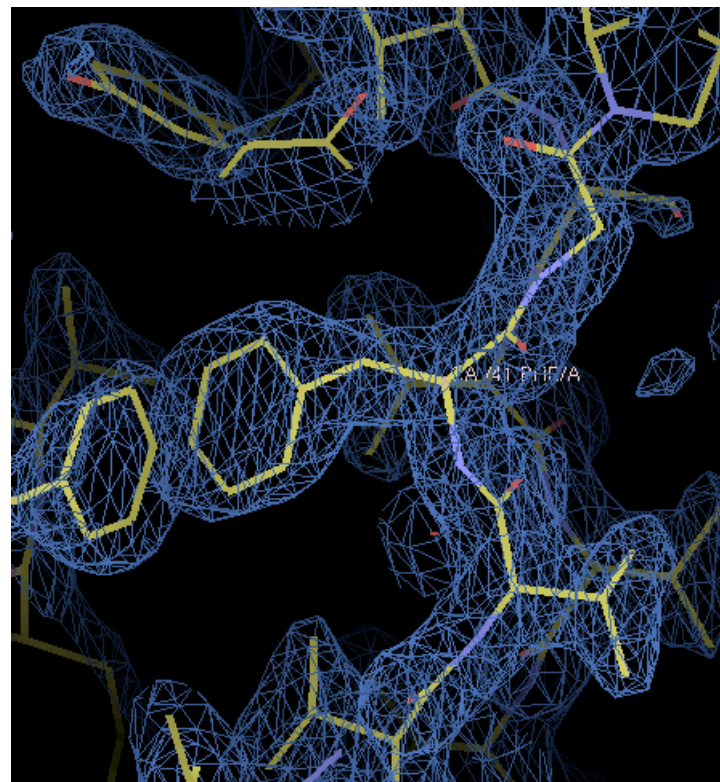
Example

- Data -
1.9Å

Unrestrained

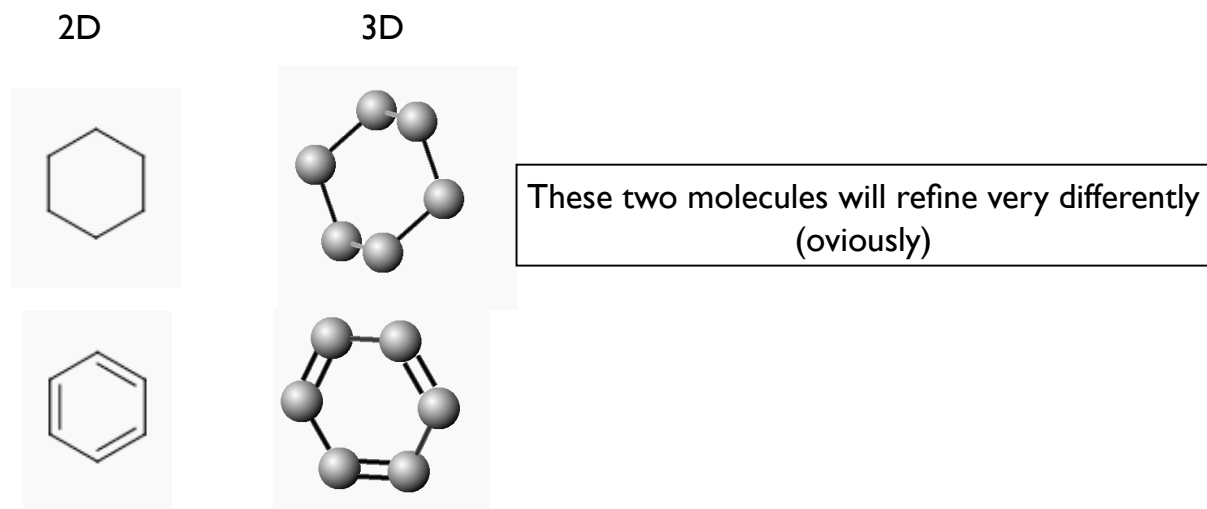


Restrained



Using restraints

- Standard dictionary has description of around 1 500 small molecules. If one of them is in your crystal then it 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.



DrugBank

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



[Home](#) [Browse](#) [Search](#) [About](#) [Downloads](#) [Contact Us](#)

Search:



The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains nearly 4800 drug entries including >1,350 FDA-approved small molecule drugs, 123 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals and >3,243 experimental drugs. Additionally, more than 2,500 non-redundant protein (i.e. drug target) sequences are linked to these FDA approved drug entries. Each DrugCard entry contains more than 100 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

DrugBank is supported by [David Wishart](#), Departments of [Computing Science](#) & [Biological Sciences](#), [University of Alberta](#).

[More about DrugBank](#)

What's New?

- We have implemented the [ChemAxon](#) solution for structure searches. You can now perform similarity (tanimoto), substructure, and exact searches via the [ChemQuery](#) function. This system replaces an outdated structure search and should be faster and more accurate. We have only added the most basic features for this release, so if you would like to see more/different features added, please let us know.
- We have added a new page containing links to other useful drug and small molecule databases. The [other databases](#) page

DrugBank

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

Search:

Search

ChemQuery


StructureMolecular WeightSMILESChemical Formula

Drug Type:

Search Type:
☐ Tanimoto Similarity
Similarity threshold:
A higher similarity threshold results in less hits that are more similar to the query structure.
☐ Substructure
☒ Exact

Molecular Weight Filter:
between and

Search



Query SMILES string:

[Example:](#) NCCCC[C@H](N)C(O)=O

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

PRODRG server

PRODRG Home [FAQ](#) [PRODRG Beta](#) [How to obtain](#) [Usa](#)

The Dundee PRODRG2 Server

Finally, a FAQ is available [here](#), READ it before using this server

Molecular topologies for ... X-ray refinement/MD ... drug design/docking

...

...

...

Funded by:

 The Wellcome Trust

Draw Molecule With JME

... Or ...

Paste your input here (PDB coordinates, MDL MOLfile, text drawing). See below for instructions.

Chirality Full charges Energy minimization

Yes No Yes

Run PRODRG Clear

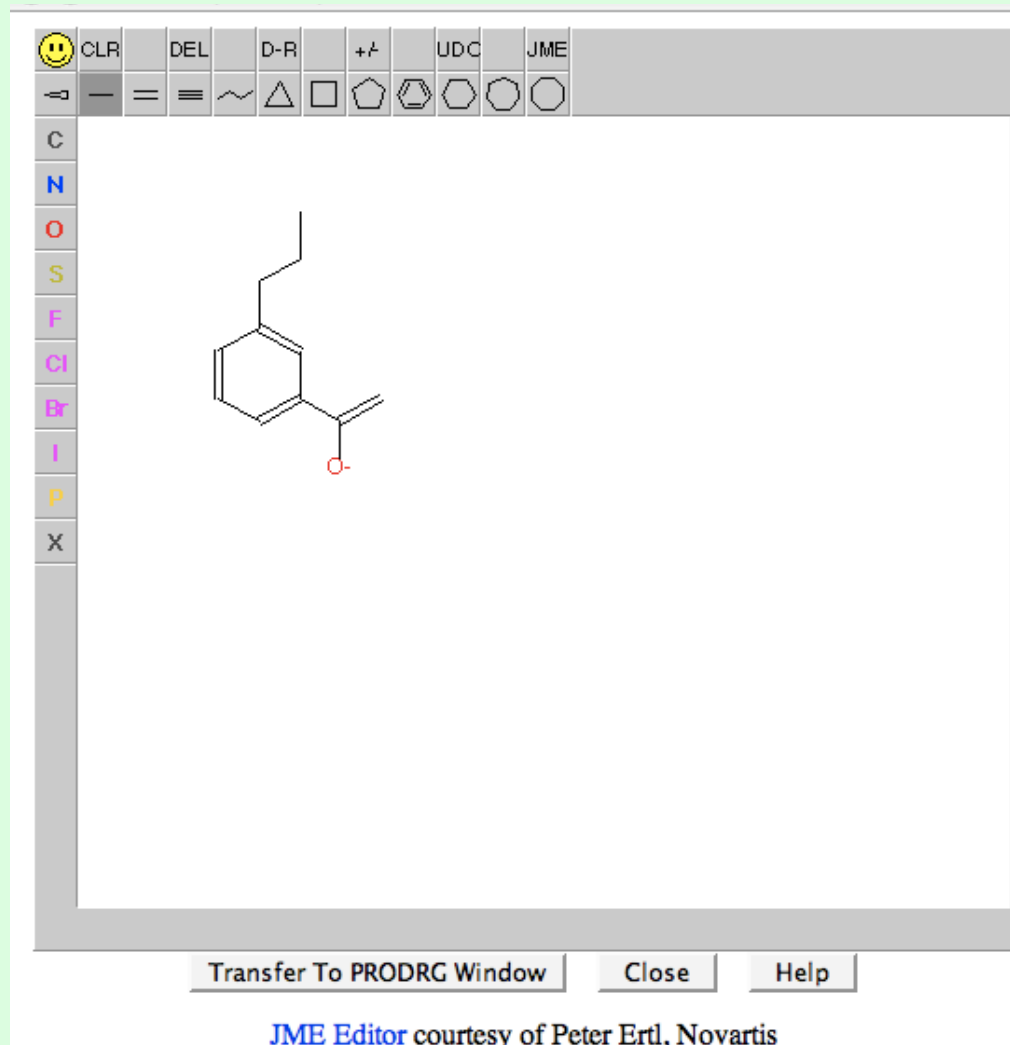
Please be patient, this can take up to 2 minutes

JME

Load your file

PRODRG: JME

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



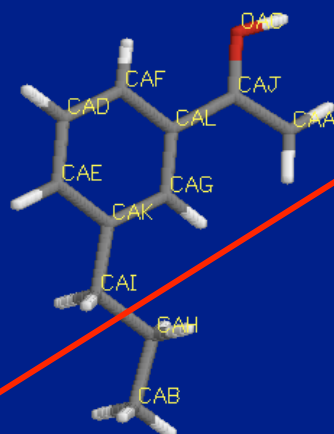
Draw your ligand,
transfer to PRODRG
window and run

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
PRODRG> 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)

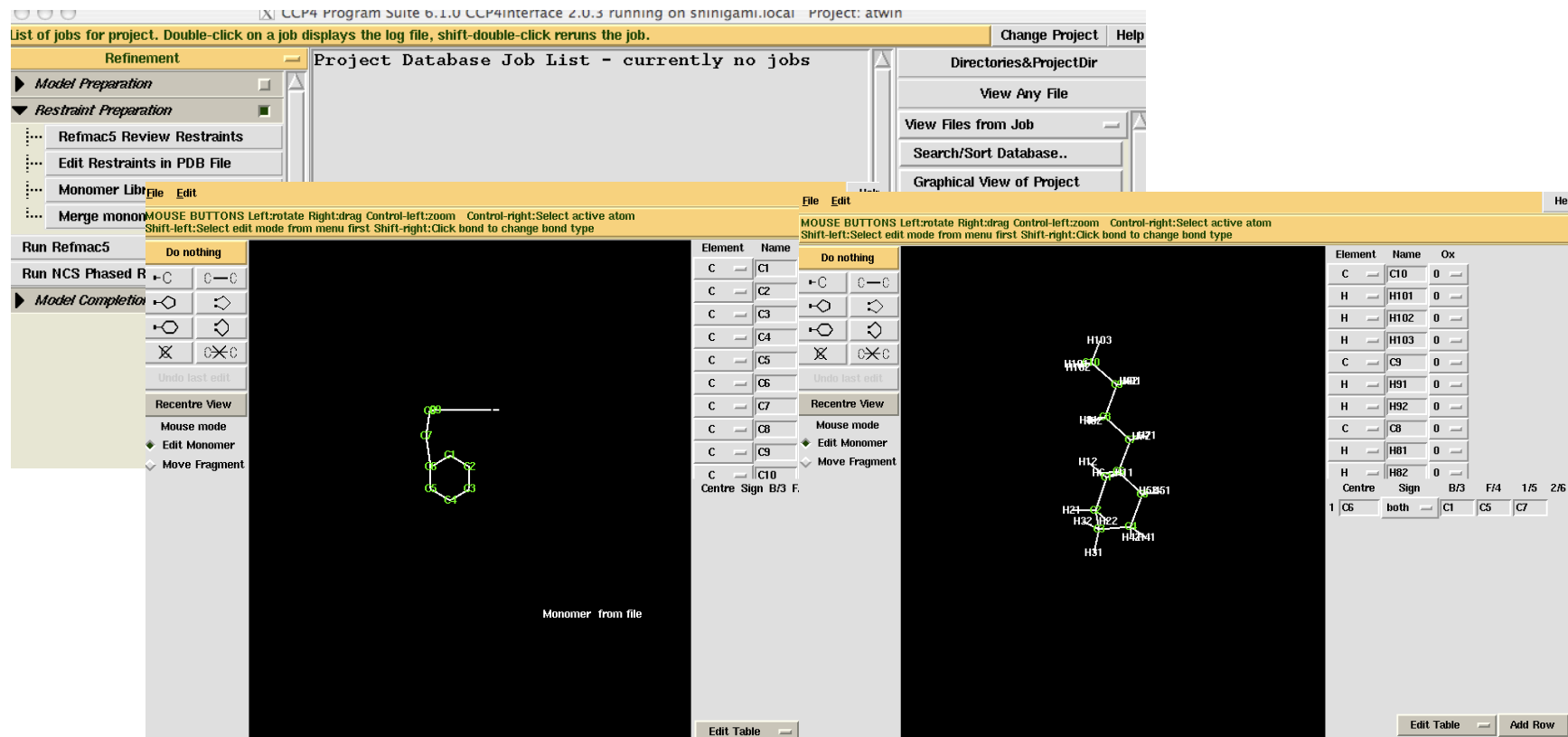
Done

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.

Using resources from ccp4

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



Sketch your ligand

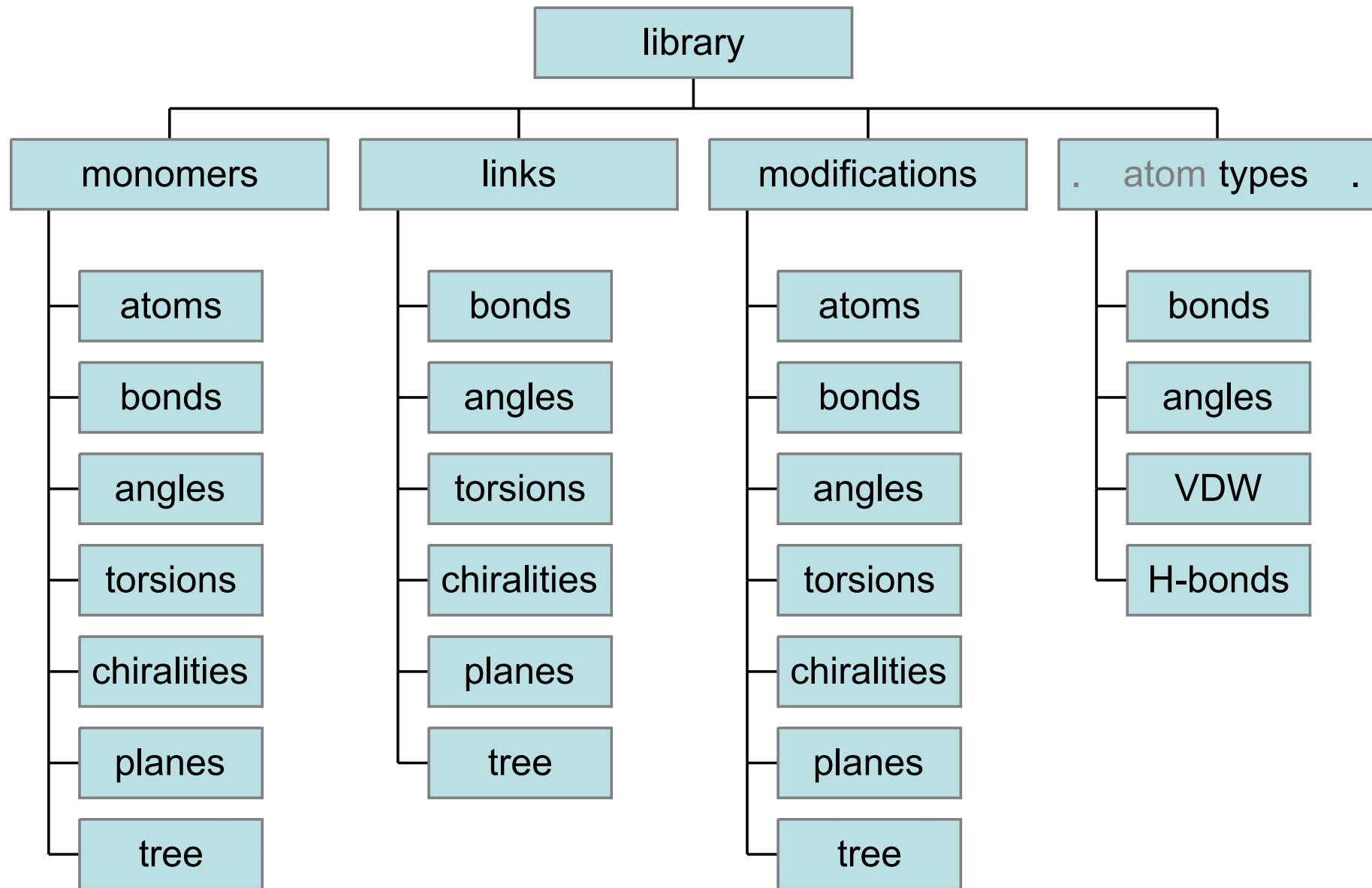
After regularisation

Jligand and Links

CCP4 monomer library: modifications and links

New link description

CCP4 monomer library (library of restraints)



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

Generic links for peptides

Generic peptide modification "DEL-OXT":



Generic peptide modification "NH1":

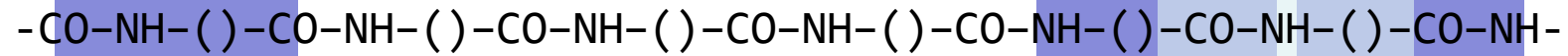


Generic peptide link "TRANS":

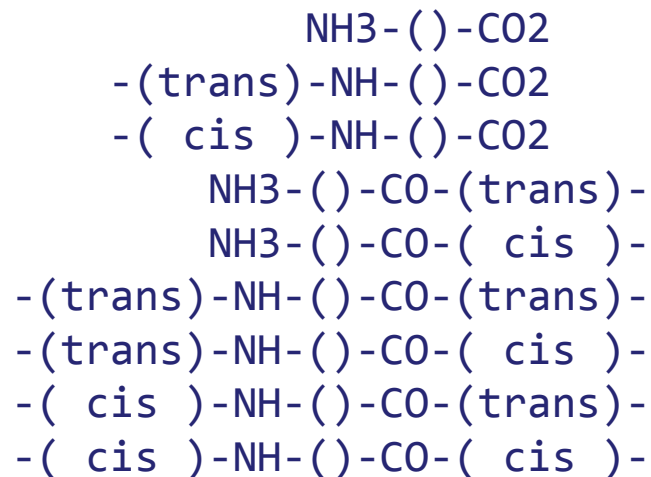


These define: bond length, angles and a plane associated with the bond C-N

Specialised monomers vs. generic links



Specialised monomers:



9 versions

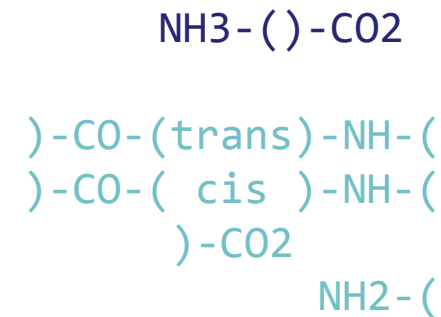
×

20 aminoacids

=

180 library entries

Generic links:



20 aminoacids

+

2 links

+

7 modifications

=

29 library entries

Links for peptides

generic

– peptide-peptide: TRANS, CIS

generic from one side

– peptide-PRO: PTRANS, NMTRANS, PCIS, NMCIS
– C-terminal modification: FOR_C-C, DFO_N-C, STA_N-C, ...
– N-terminal modification: FOR_C-N, ACE_C-N, DFO_C-N, ...
– pyranose-(ASP,THR,SER): NAG-SER, NAG-THR, NAG-ASN

specialised

– S-S bridge: CYS-CYS
– pyranose-peptide: XYS-SER, XYS-THR, XYS-ASN, ...
– metal-peptide: ZN-CYS, FE-CYS

Standard modifications and links (generic and specialised)

CCP4 library contains modifications for:

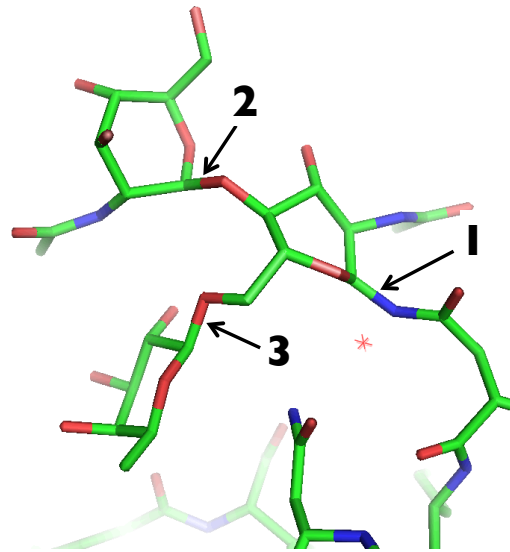
- terminal peptides and nucleotides
- methylated nucleotides
- deprotonated states

CCP4 library contains links and corresponding modifications for:

- polypeptide chains (CIS,TRANS), S-S bridges
- polynucleotide chains
- glycosylated proteins

Generic links for sugars

2xmb



For typical glycosylation cases

- necessary modifications and links are there in the standard ccp4 library
- by default REFMAC uses these library descriptions and does not need any additional instructions

NAG – NAG – ASN

|
FUL

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

Standard links used here:

- (1) "NAG-ASN"
- (2) "BETA1-4"
- (3) "ALPHA1-6"

Sugar links: refmac checkpoints

✓ refmac terminated normally

✓ output pdb-file contains expected LINKR records, e.g.

LINKR...
...NAG A1547 FUL
A1549...
...ALPHA1-6

```
Terminal — vim — 87x16
REMARK      3
MODRES      NAG A 1547  NAG-b-D
MODRES      NAG A 1548  NAG-b-D
SSBOND      1 CYS A   65   CYS A   92
SSBOND      2 CYS A  252   CYS A  263
SSBOND      3 CYS A  400   CYS A  519
LINKR        C1  NAG A1547
LINKR        NAG A1547
LINKR        NAG A1547
CISPEP      1 ALA A  101   PRO A  102
CISPEP      2 VAL A  377   ASP A  378
CISPEP      3 GLN A  380   ARG A  381
CRYST1      154.800 154.800 134.650 90.00 90.00 90.00 I 4 2 2
SCALE1       0.006460 0.000000 0.000000 0.000000
SCALE2      -0.000000 0.006460 0.000000 0.000000
ND2 ASN A 241
NAG A1548
FUL A1549
NAG-ASN
BETA1-4
ALPHA1-6
```

✓ log-file contains warnings saying e.g. that

... link:ALPHA1-6 is
found
... res:1547 NAG ...
... res:1549 FUL ...

(WARNING = OK)

```
Terminal — vim — 87x21
<!--SUMMARY_END--></FONT></B>
Number of atoms : 4627
Number of residues : 910
Number of chains : 5
I am reading library. Please wait.
mon_lib.cif
WARNING : link:SS is found dist = 2.070 ideal_dist= 2.031
ch:AA res: 65 CYS at:SG .->AA res: 92 CYS at:SG .
WARNING : link:NAG-ASN is found dist = 1.468 ideal_dist= 1.439
ch:AA res: 241 ASN at:ND2 .->Aa res:1547 NAG at:C1 .
WARNING : link:SS is found dist = 2.092 ideal_dist= 2.031
ch:AA res: 252 CYS at:SG .->AA res: 263 CYS at:SG .
WARNING : link:SS is found dist = 2.108 ideal_dist= 2.031
ch:AA res: 400 CYS at:SG .->AA res: 519 CYS at:SG .
WARNING : link:BETA1-4 is found dist = 1.491 ideal_dist= 1.439
ch:Aa res:1547 NAG at:O4 .->Ab res:1548 NAG at:C1 .
WARNING : link:ALPHA1-6 is found dist = 1.432 ideal_dist= 1.439
ch:Aa res:1547 NAG at:O6 .->Ac res:1549 FUL at:C1 .
-----
--- title of input coord file ---
```

User-defined links

When new link descriptions are needed:

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

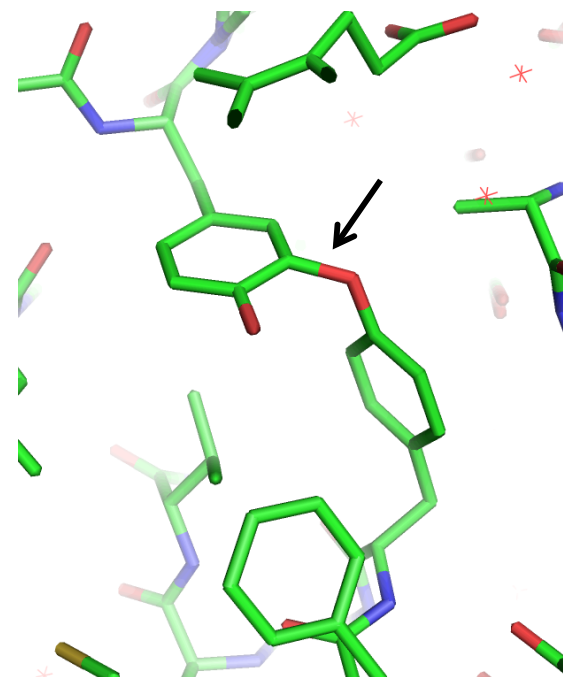
side chain – main chain (e.g. LYS – Ubiquitin)

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



CCP4 monomer library: modifications and links

New link description

New link

Example:

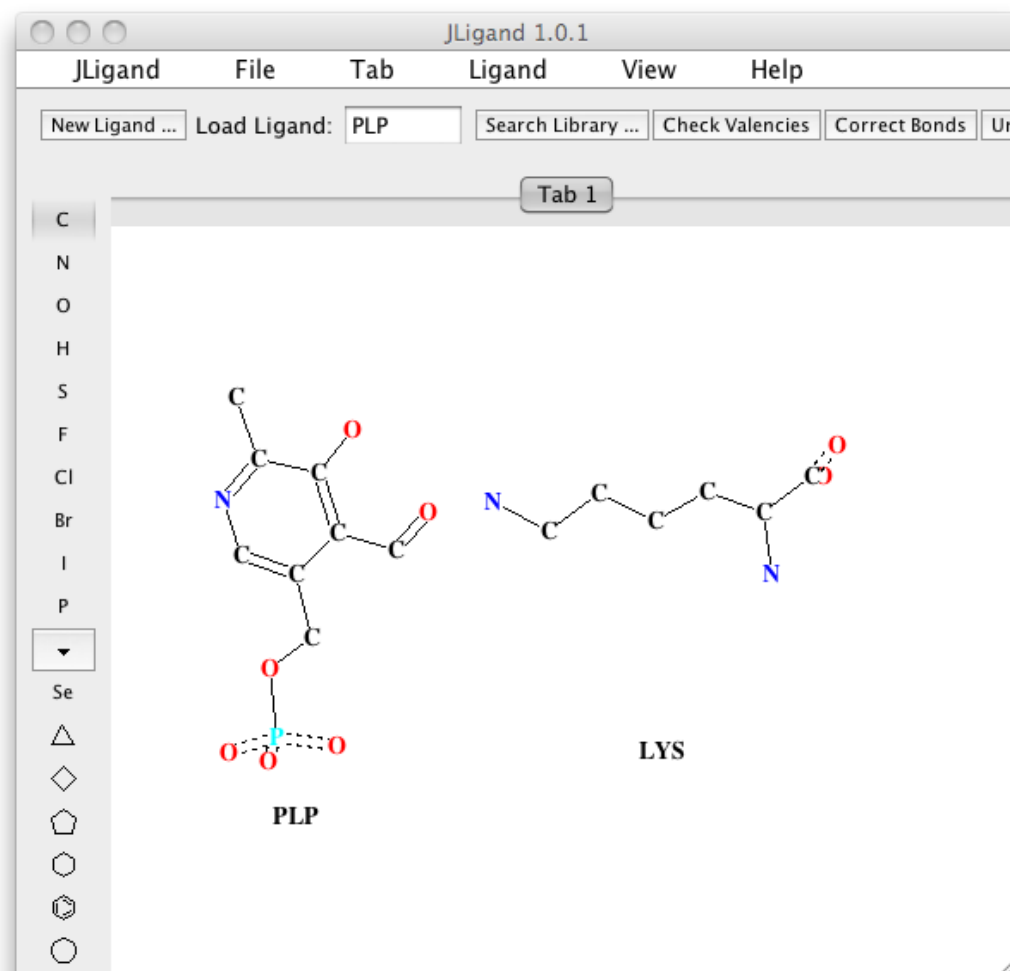
- covalent linkage between LYS and Pyridoxal phosphate (PLP).
- describes PLP forming internal aldimine in aminotransferases.

Given:

- descriptions of LYS and PLP from the standard library

Needed:

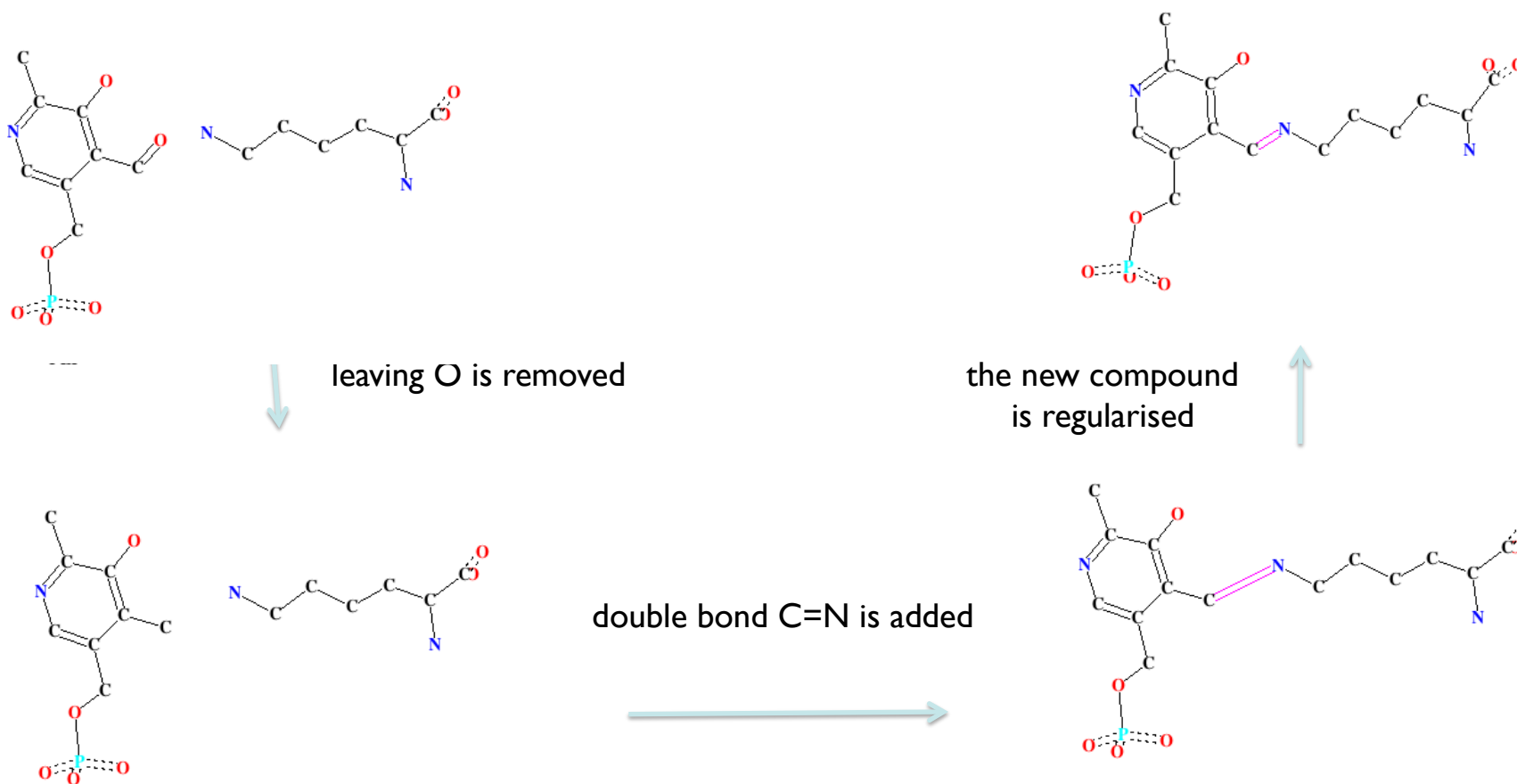
- additional library file with the description of link LYS–PLP



Creating a new link, as seen in JLigand GUI

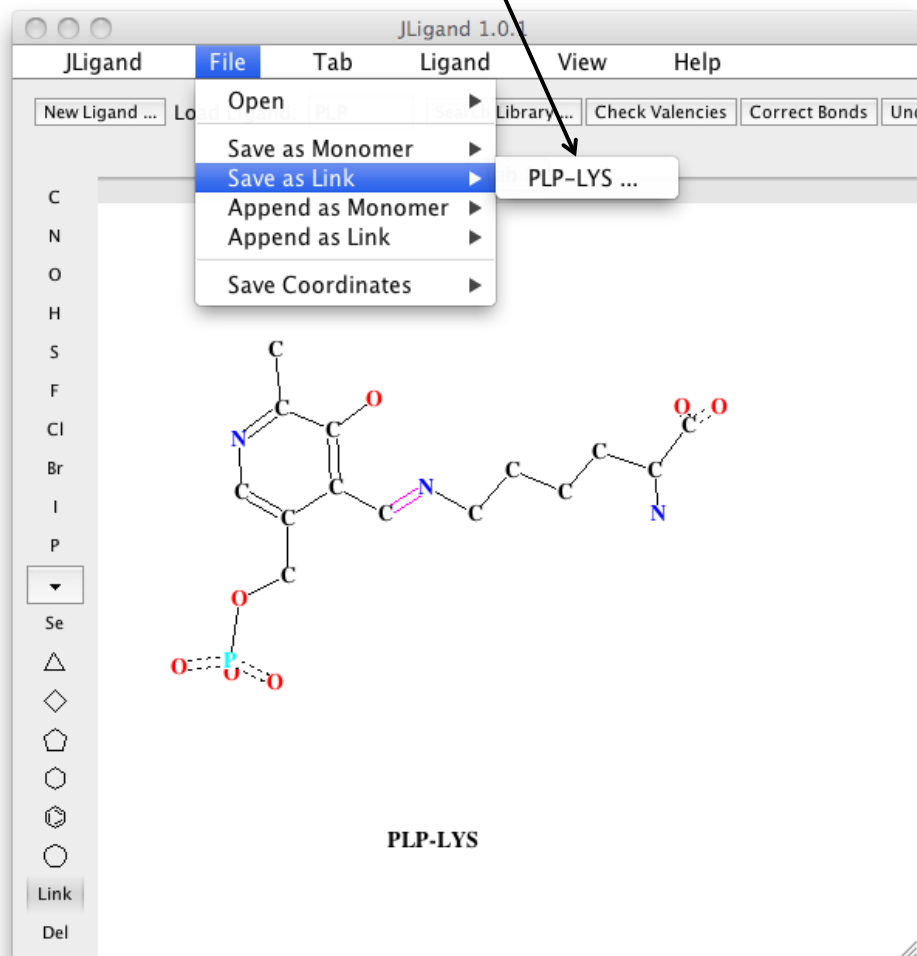
The two monomers are in effect reacted in silico
Hydrogen atoms are dealt with automatically*)

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



The new link, "file view"

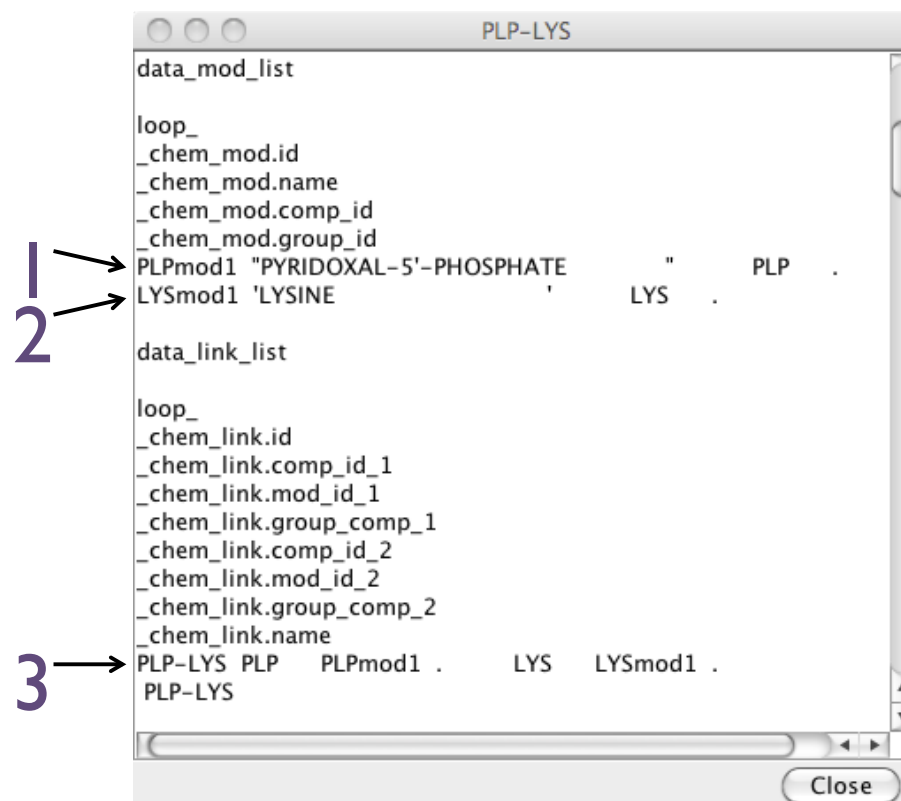
To save into CIF-file (additional library)



Contents:

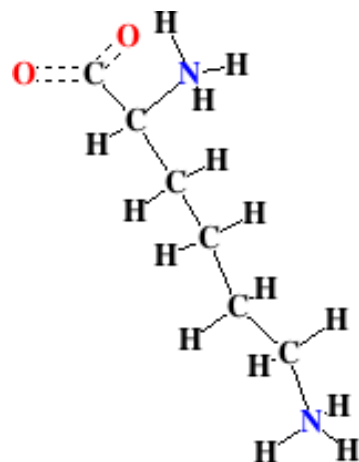
- (1) modification "PLPmod1"
- (2) modification "LYSmod1"
- (3) link "PLP-LYS"

No monomers

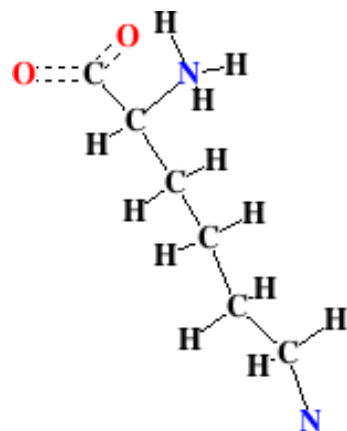


The new link, "file view"

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     .      .      .
```

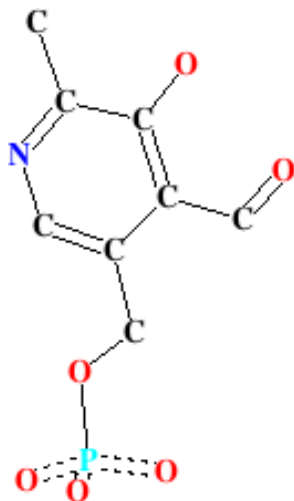
Bonds

Angles

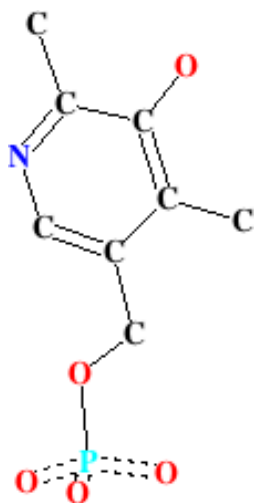
.....

The new link, "file view"

PLP



PLPmod1



Modification "PLPmod1":
changes to PLP

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 . . . .
```

Atom

```
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 . . . .
```

Bond

```
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 . . . .
```

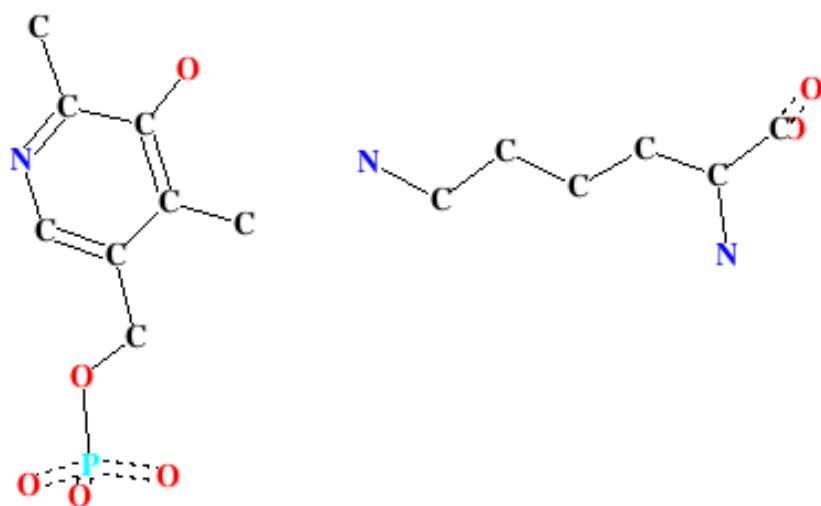
Angles

```
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 . . . .
```

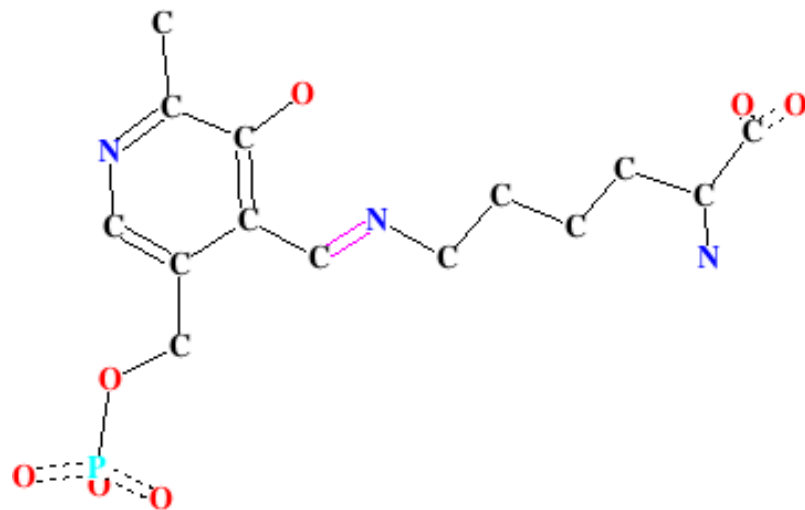
Plane

The new link, "file view"

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



PLP-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

Utilising new link description

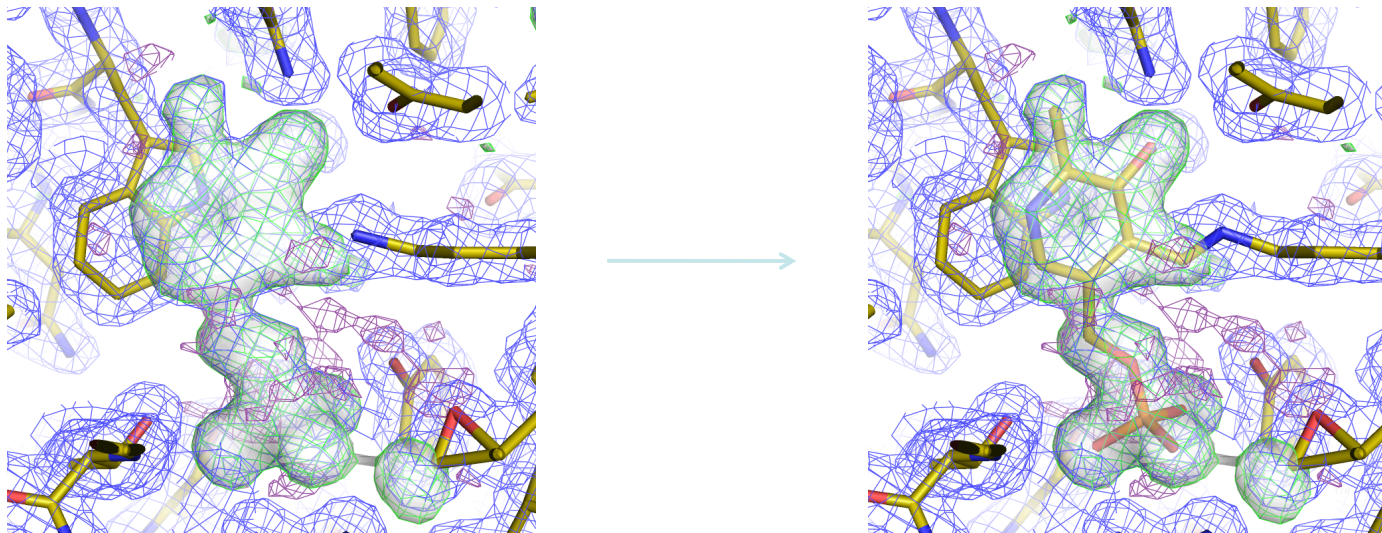
Three remaining steps:

- docking monomer(s) into electron density
- defining link in the pdb-file
- refinement of the structure with linked ligand using additional library

(I) 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

(2) 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:

residues to link

link to use

```
Terminal - vim - 81x14
CISPEP  1 SER A  137  PRO A  138      0.00
CISPEP  2 ASN A  194  PRO A  195      0.00
CISPEP  3 SER B  137  PRO B  138      0.00
CISPEP  4 ASN B  194  PRO B  195      0.00

LINKR    NZ  LYS B 258      C4A PLP D  1      LYS-PLP

CRYST1 125.000 130.800 55.800 90.00 90.00 90.00 P 21 21 21
SCALE1  0.008000 0.000000 0.000000      0.000000
SCALE2 -0.000000 0.007645 0.000000      0.000000
SCALE3  0.000000 -0.000000 0.017921      0.000000
ATOM    1  N  ALA A  1      -76.191 -36.168 -21.452 1.00 49.90      N
ATOM    2  CA ALA A  1      -74.845 -35.859 -20.889 1.00 49.65      C
```

(3) Refinement using additional library

Additional library is defined
here

Run Refmac5 Initial parameters from /Users/lebedev/Desktop/CCP4-2011/03_Wrk/JLigand_link_c...

Help

Job title **model with ligands**

Do **restrained refinement** using **no prior phase information** input

☐ Input fixed TLS parameters

no twin refinement

MTZ in **1ajs** **data.mtz** Browse View

FP **FP** Sigma **SIGFP**

MTZ out **1ajs** **refmac2.mtz** Browse View

PDB in **1ajs** **refmac1-coot-0.pdb** Browse View

PDB out **1ajs** **refmac2.pdb** Browse View

LIB in **1ajs** **refmac.cif** Merge LIBINs Browse View

Output lib **1ajs** **refmac2.cif** Browse View

Include keyword file **1ajs** Browse View

Data Harvesting ☐

Refinement Parameters ☐

Run Save or Restore Close

Acknowledgment

York

Andrey Lebedev

Alexei Vagin

Fei Long

CCP4, YSBL people

Jligand is available from CCP4 or from York's ftp site:

www.ysbl.york.ac.uk/mxstat/JLigand

or google jligand

This and other presentations can be found on:

www.ysbl.york.ac.uk/refmac/Presentations/

