

# Ligands in Phenix

Generating & modifying for all scenarios

Nigel W. Moriarty

User's Meeting  
ACA, July 2024

# What are restraints?

- Restraints are harmonic functions that provide (mostly) chemical information via residuals & gradients
- Needed because experimental data is rarely sufficient for structure determination
- Weighted with the experimental information to find the optimal result

# Resolution dependence

- Ultra-hi res – Not needed
- Hi res – Can have large deviations because the experimental data dominates
- Lo res – Generally approaches the ideal values
  - If not, large scale problems

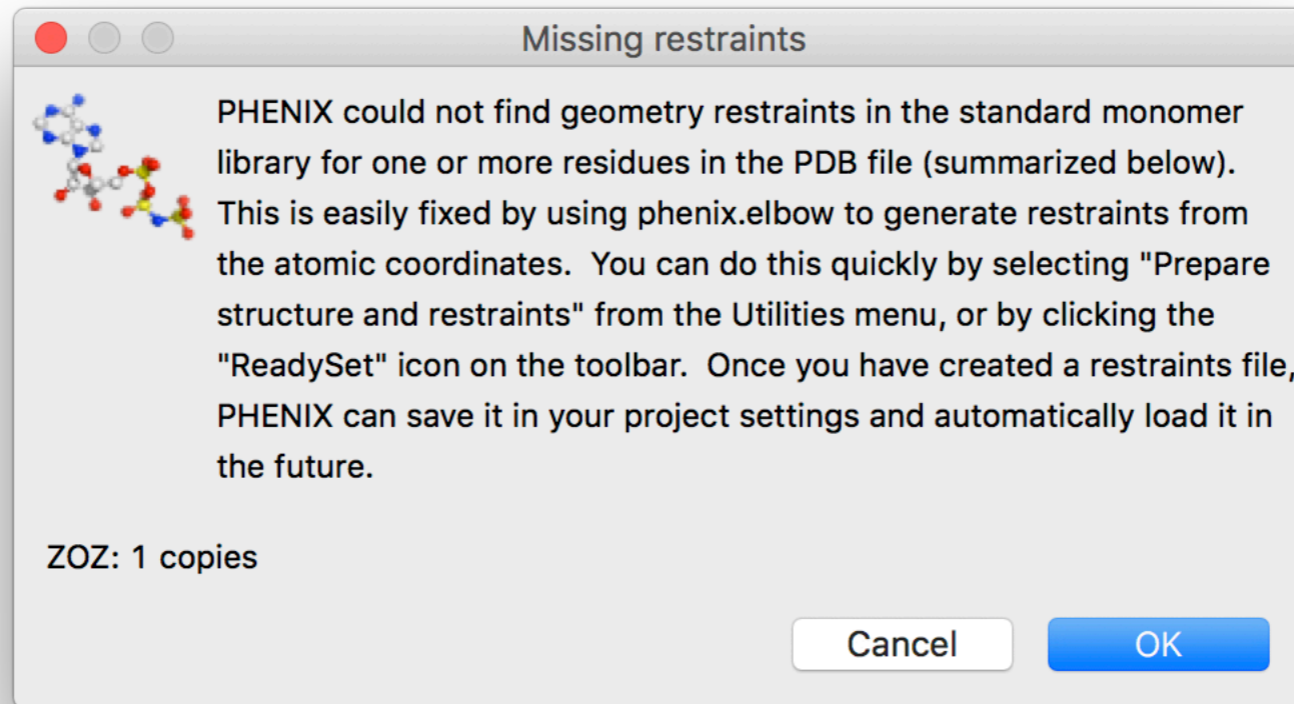
# Restraints in Action

- Libraries
  - Monomer Library
  - GeoStd
- Algorithms
  - Polymer
  - Links

# GeoStd

- All standard amino acids
- Current list of non-standard amino acids
- All standard RNA/DNA
- Current list of non-standard RNA/DNA
- Others – 37k Mogul validated restraints using PBEh-3c/CPCM and higher QM

# What you will see



# Or

Sorry: Fatal problems interpreting model file:

Number of atoms with unknown nonbonded energy type symbols: 21

Please edit the model file to resolve the problems and/or supply a CIF file with matching restraint definitions, along with `apply_cif_modification` and `apply_cif_link` parameter definitions if necessary.

# CIF

- Crystallographic Information File
- mmCIF – macro-molecular CIF
- Used for
  - Model
  - Data
  - Maps
  - Ligands
    - Information
    - Restraints



# Confusion

- All depositions of X-ray model use mmCIF from 1 July 2019
- “I need a CIF file.”
  - But what do you really need?

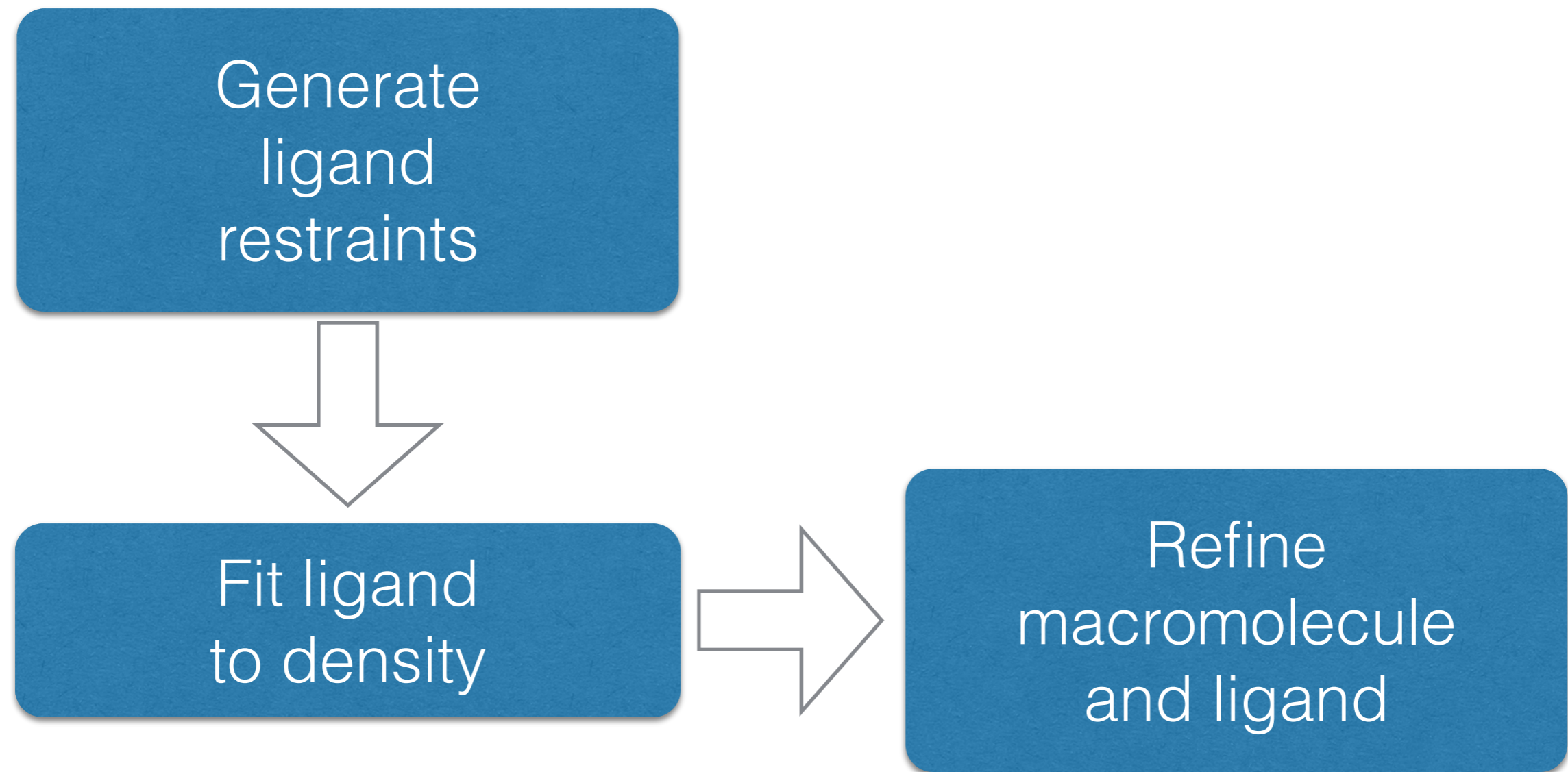
# Restraints?

- Provide a reasonable geometry during refinement particularly at low resolution
  - Bonds, angles, dihedrals, chirals, planes, ...
- Must be weighted against the experimental information

# Overview

- eLBOW - electronic Ligand Builder & Optimisation Workbench
- ReadySet! - One-stop preparation for your refinement needs
- REEL - Restraints Editor Essentially Ligands

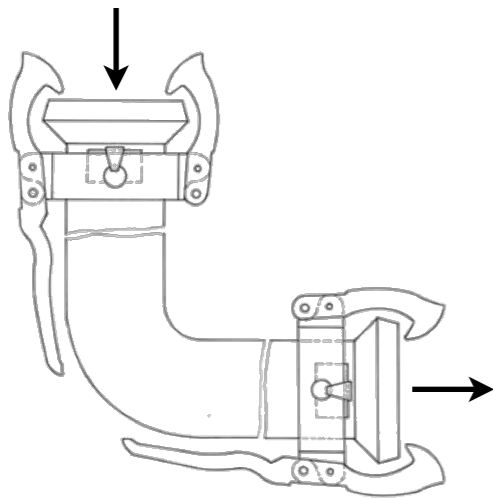
# Ligands in crystallography



# eLBOW goals

- Fast, simple and flexible procedure to include ligands
- Reduce the tedium of building 3D ligand models
- Automate generation of restraints for ligands
- Comparison of ligand structures

Chemical input



Chemical restraints (CIF)  
Cartesian coordinates (PDB)

Reflection data  
Protein information

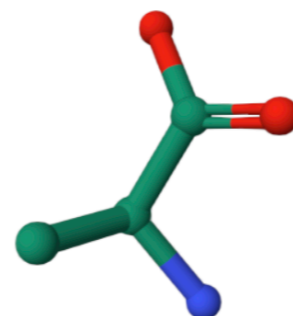
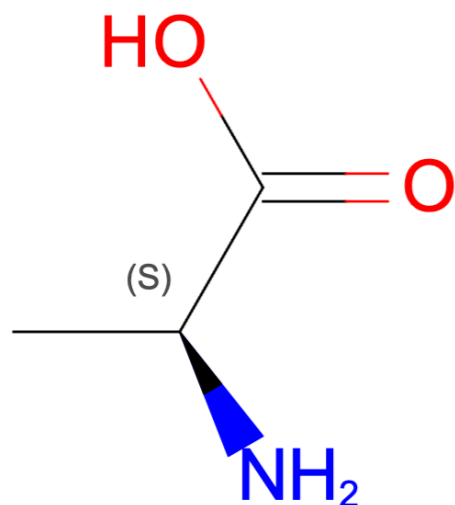
refinement

N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, (2009). Acta Cryst. D 65, 1074-1080.

# Ligand Expo / Chemical Component Library

- List of all the entities in the Protein Data Bank
  - Amino acids, Nucleic acids
  - Ligands, Small molecule
  - Metal clusters
- In CIF format
  - Contents chemical information
    - SMILES, atom names, bonds
  - Not restraints

# Amino Acid



Toggle Hydrogen

Toggle Labels

 Display Files ▼

 Download Files ▼

 Data API

## ALA

### ALANINE

Find entries where: ALA

- is present as a standalone ligand in [172 entries](#)
- as a non-polymer is covalently linked to polymer or other heterogen groups [58 entries](#)
- is present in a polymer sequence [210,908 entries](#)

Find related ligands:

- [Similar Ligands \(Stereospecific\)](#)
- [Similar Ligands \(including Stereoisomers\)](#)
- [Similar Ligands \(Quick Screen\)](#)
- [Similar Ligands \(Substructure Stereospecific\)](#)
- [Similar Ligands \(Substructure including Stereoisomers\)](#)

# Water

## PDB Chemical Component HOH



JSmol

Ideal Model

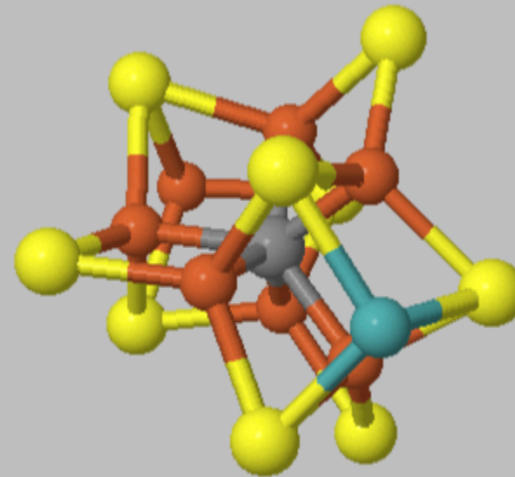
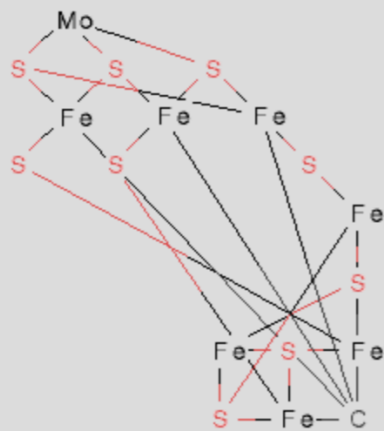
## Chemical Description

**Name** WATER  
**Formula** H<sub>2</sub> O  
**Formal charge** 0  
**Molecular weight** 18.015 g/mol  
**Component type** NON-POLYMER



# Metal clusters

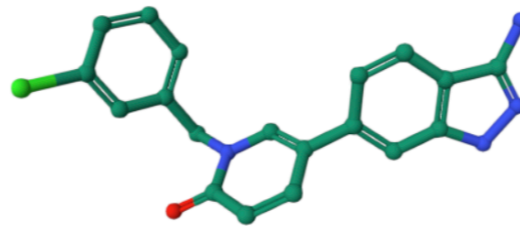
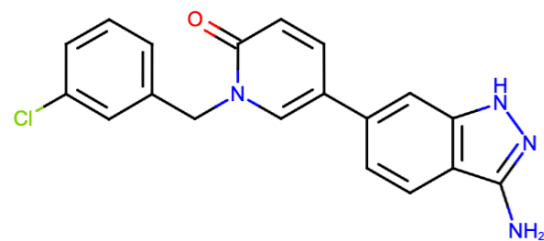
## PDB Chemical Component ICS



JSmol

Ideal Model

# 5-letter codes



Toggle Hydrogen

Toggle Labels

Display Files

Download Files

Data API

## A1LU6

5-(3-azanyl-1~{H}-indazol-6-yl)-1-[(3-chlorophenyl)methyl]pyridin-2-one

Find entries where: A1LU6

is present as a standalone ligand in 1 entries

search

Find related ligands:

[Similar Ligands \(Stereospecific\)](#)

[Similar Ligands \(including Stereoisomers\)](#)

[Similar Ligands \(Quick Screen\)](#)

[Similar Ligands \(Substructure Stereospecific\)](#)

[Similar Ligands \(Substructure including Stereoisomers\)](#)

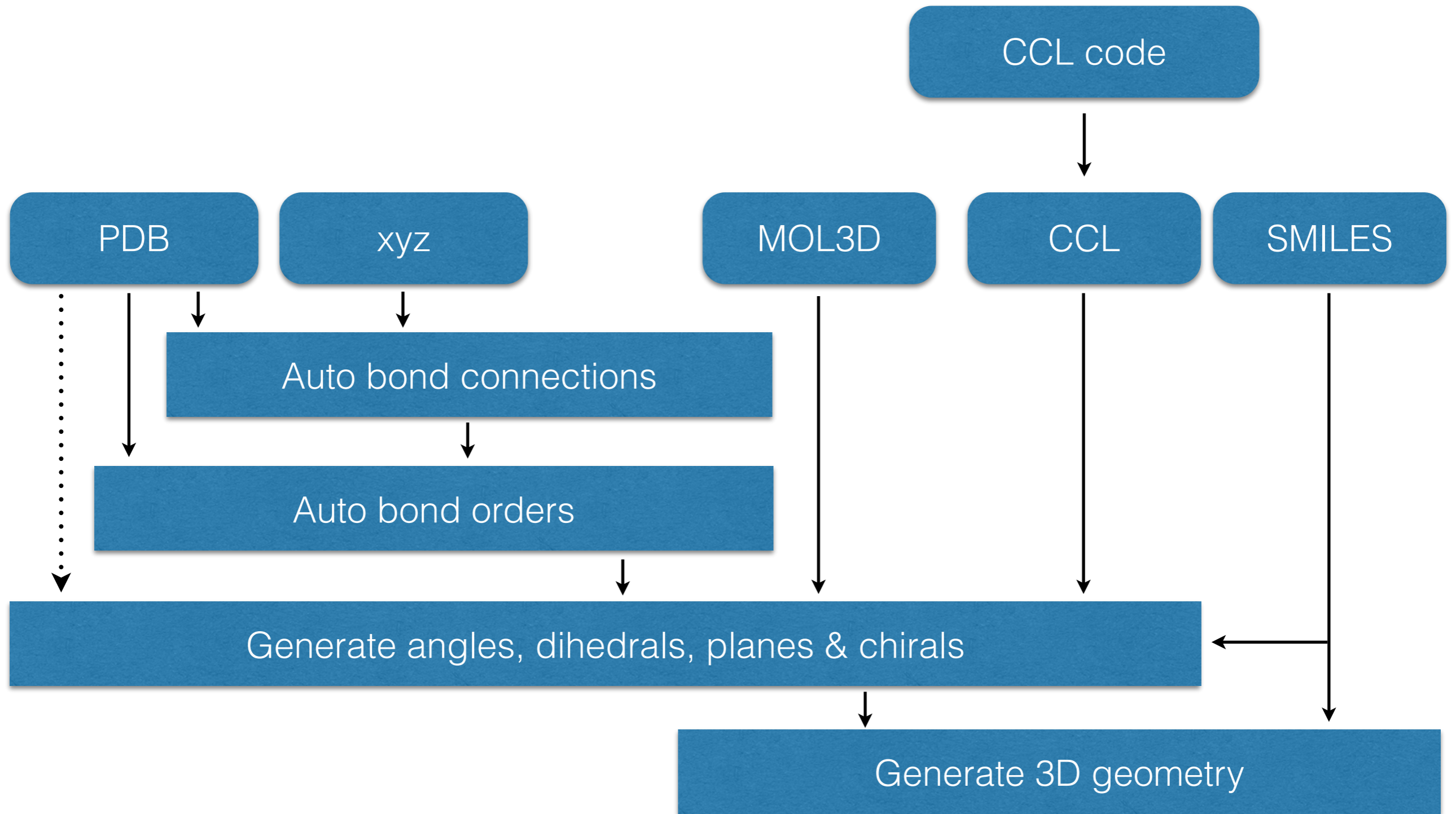
- 49k combinations for 3-letter codes

# Human readable

Standard		Human readable	
Uppercase	Lowercase	Uppercase	Lowercase
10I0	1oi0	1oi0	1oi0
1IJJ	1ijj	1iJJ	1ijj
40CL	4ocl	4oCL	4ocL
5SS2	5ss2	5ss2	5ss2

- Confusable letters are case-forced
  - L is always uppercase
  - i, o are always lowercase
- Somewhat confusable – “5” and “S”

# Topology



# Optimisation

Topology information - Atoms, bonds, angles, ...

Simple force field geometry optimisation

Add hydrogens

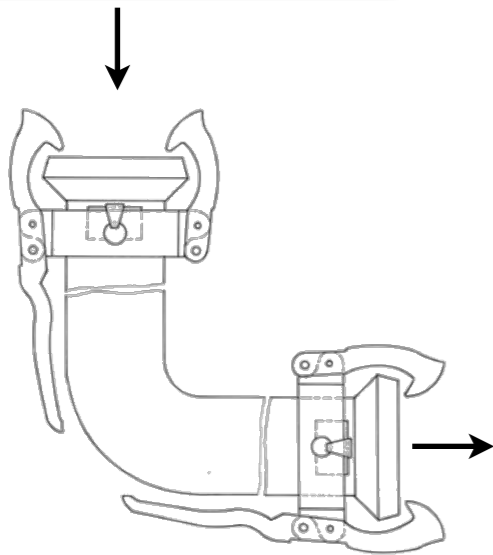
Advanced geometry optimisation

Output geometry (PDB) and restraints (CIF)

# Getting ready to refine

- Many details needed to prepare for structure refinement

Chemical input



Chemical restraints (CIF)  
Cartesian coordinates (PDB)

Protein Information

ReadySet!

Experimental data

refinement

# ReadySet!

- Add hydrogens
  - Default: adds hydrogens to protein, ligands
    - Protein - Reduce
    - Ligands - eLBOW
  - Add hydrogens to water
  - Add deuteriums instead of hydrogens
  - Add hydrogen & deuteriums appropriately
- Generate restraints

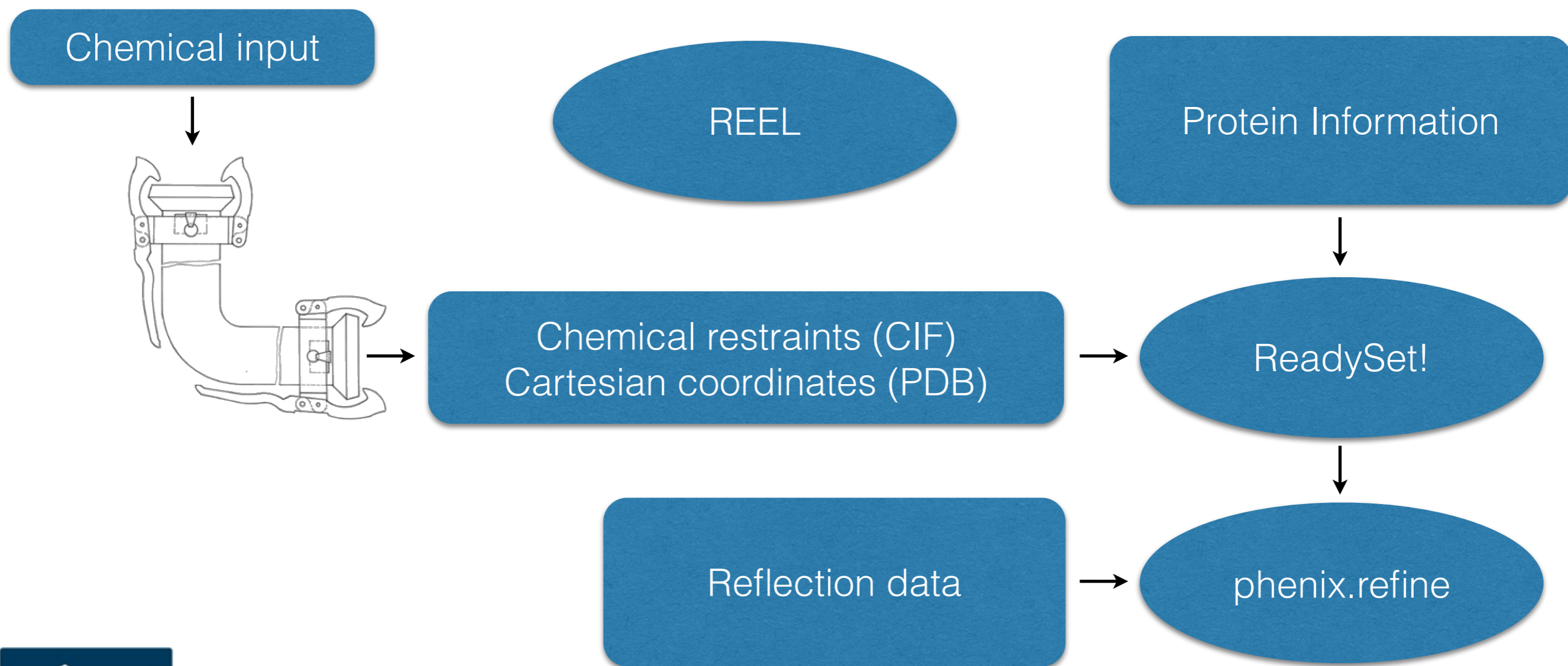
# ReadySet!

- Restraints CIF filename
- Restraints CIF directory
- LINKS to “edits”
- --dry-run to show ligand process pathway
- Metal coordination



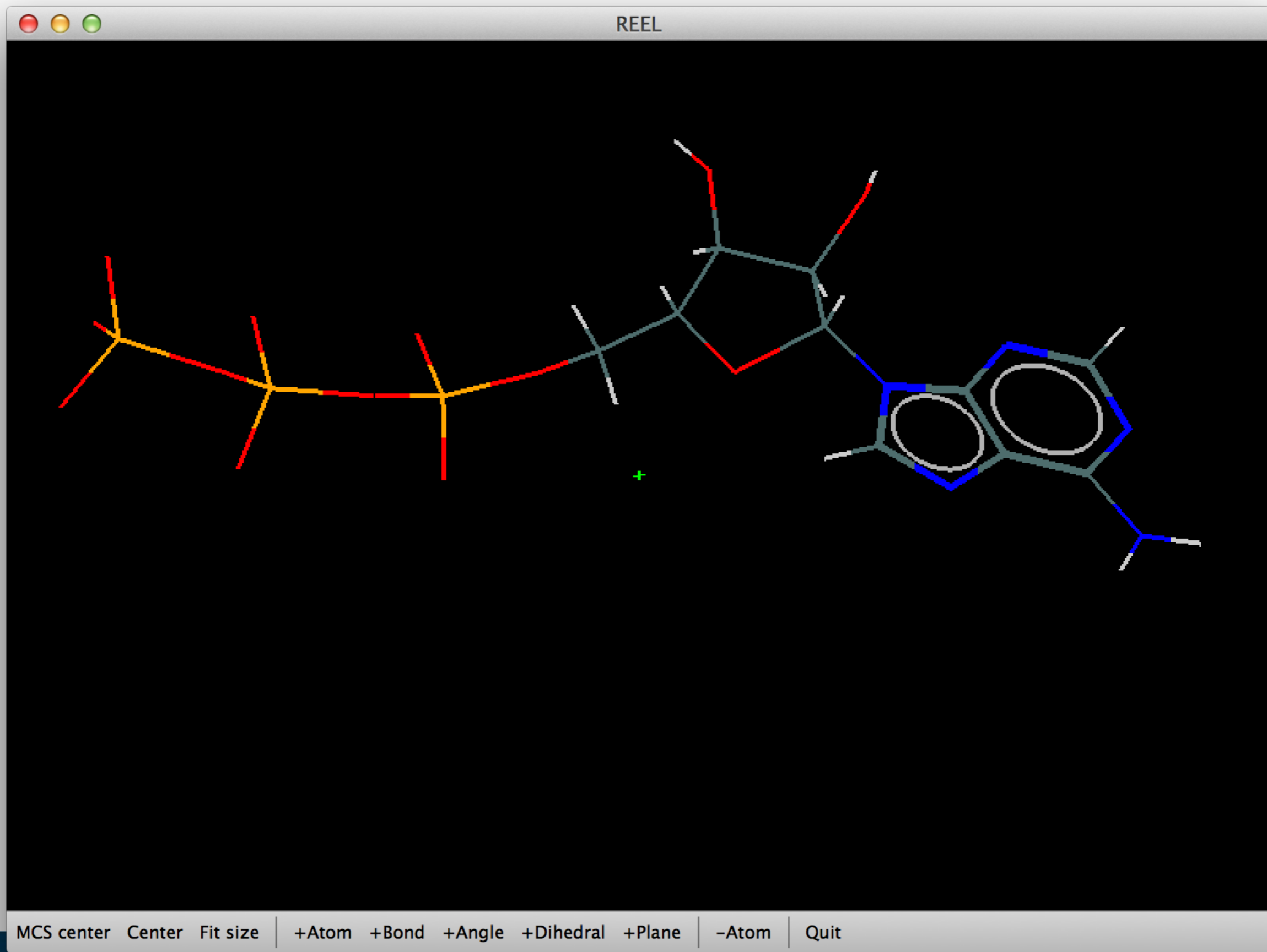
# Restraints editing

- Visualisation of primary and secondary restraints
- Adjusting restraints to user preference



# Restraints Editor, Essentially Ligands

- Generate a geometry for a set of restraints
- Modify restraints and generate new geometry
- Fast editing in menu items
- Highlight atom and restraints are highlighted
- Multiple ligands simultaneously
- Highlight restraint and atoms are highlighted
- Save restraints (CIF)
- Save geometry (PDB)
- Run eLBOW



Restraints Editor Especially Ligands (REEL)

Simple Optimisation eLBOW Optimisation AM1 Optimisation Search Components Find unique code

ATP

Atoms(43) Bonds(45) Angles(78) Dihedrals(30) Planes(17) Chirals(4) CisTrans Chirals Implicit(7) BoatChair

	?	comp_id	atom_id	type_symbol	type_energy	charge	partial_charge	x	y	z
1		ATP	PG	P	P	0	.	-2.009900	-8.939900	0.927100
2		ATP	O1G	O	O	0	.	-0.919700	-9.907100	1.322000
3		ATP	O2G	O	OP	-1	.	-3.257700	-9.235200	1.724400
4		ATP	O3G	O	OP	-1	.	-2.306500	-9.089900	-0.545900
5		ATP	PB	P	P	0	.	-0.243200	-6.679500	0.469800
6		ATP	O1B	O	O	0	.	1.050300	-7.257000	0.992700
7		ATP	O2B	O	OP	-1	.	-0.341900	-6.929700	-1.016000
8		ATP	O3B	O	O2	0	.	-1.514200	-7.398500	1.233500
9		ATP	PA	P	P	0	.	0.670000	-3.968400	-0.041400
10		ATP	O1A	O	O	0	.	0.729800	-4.332500	-1.505600
11		ATP	O2A	O	OP	-1	.	2.063300	-3.990800	0.540200
12		ATP	O3A	O	O2	0	.	-0.276300	-5.056600	0.757100
13		ATP	O5'	O	O2	0	.	0.029800	-2.458300	0.123600
14		ATP	C5'	C	CH2	0	.	0.490600	-1.443000	-0.721300
15		ATP	C4'	C	CH1	0	.	-0.438000	-0.204600	-0.596500
16		ATP	O4'	O	O2	0	.	0.055400	0.644000	0.263400
17		ATP	C3'	C	CH1	0	.	-0.496500	0.556500	-1.940500
18		ATP	O3'	O	OH1	0	.	-1.718900	0.408900	-2.511600
19		ATP	C2'	C	CH1	0	.	-0.245900	2.038900	-1.588600
20		ATP	O2'	O	OH1	0	.	-1.282600	2.903300	-2.232200
21		ATP	C1'	C	CH1	0	.	-0.345300	2.109700	-0.288400
22		ATP	N9	N	NR5	0	.	0.564300	3.101200	0.226900
23		ATP	C8	C	CR15	0	.	1.894800	3.031600	0.318500

View preferences loaded

Restraints Editor Especially Ligands (REEL)

Simple Optimisation eLBOW Optimisation AM1 Optimisation Search Components Find unique code

ATP

Atoms(43) **Bonds(45)** Angles(78) Dihedrals(30) Planes(17) Chirals(4) CisTrans Chirals Implicit(7) BoatChair

	?	comp_id	atom_id_1	atom_id_2	type	value_dist	value_dist_esd	value_dist_neutron
1		ATP	PG	O1G	deloc	1.510000	0.020000	1.51
2		ATP	PG	O2G	deloc	1.510000	0.020000	1.51
3		ATP	PG	O3G	deloc	1.510000	0.020000	1.51
4		ATP	PG	O3B	single	1.648000	0.020000	1.648
5		ATP	PB	O1B	deloc	1.510000	0.020000	1.51
6		ATP	PB	O2B	deloc	1.510000	0.020000	1.51
7		ATP	PB	O3B	single	1.648000	0.020000	1.648
8		ATP	PB	O3A	single	1.648000	0.020000	1.648
9		ATP	PA	O1A	deloc	1.510000	0.020000	1.51
10		ATP	PA	O2A	deloc	1.510000	0.020000	1.51
11		ATP	PA	O3A	single	1.648000	0.020000	1.648
12		ATP	PA	O5'	single	1.648000	0.020000	1.648
13		ATP	O5'	C5'	single	1.399000	0.020000	1.399
14		ATP	C5'	C4'	single	1.553000	0.020000	1.553
15		ATP	C5'	H5'1	single	0.970000	0.020000	1.09
16		ATP	C5'	H5'2	single	0.970000	0.020000	1.09
17		ATP	C4'	O4'	single	1.305000	0.020000	1.305
18		ATP	C4'	C3'	single	1.546000	0.020000	1.546
19		ATP	C4'	H4'	single	0.970000	0.020000	1.09
20		ATP	O4'	C1'	single	1.617000	0.020000	1.617
21		ATP	C3'	O3'	single	1.357000	0.020000	1.357
22		ATP	C3'	C2'	single	1.544000	0.020000	1.544
23		ATP	C3'	H3'	single	0.970000	0.020000	1.09

View preferences loaded



# Restraints Editor Especially Ligands (REEL)

Simple Optimisation

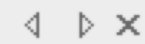
eLBOW Optimisation

AM1 Optimisation

Search Components

Find unique code

ATP



Angles(78)

Dihedrals(30)

Planes(17)

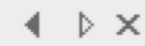
Chirals(4)

CisTrans

Chirals Implicit(7)

BoatChair

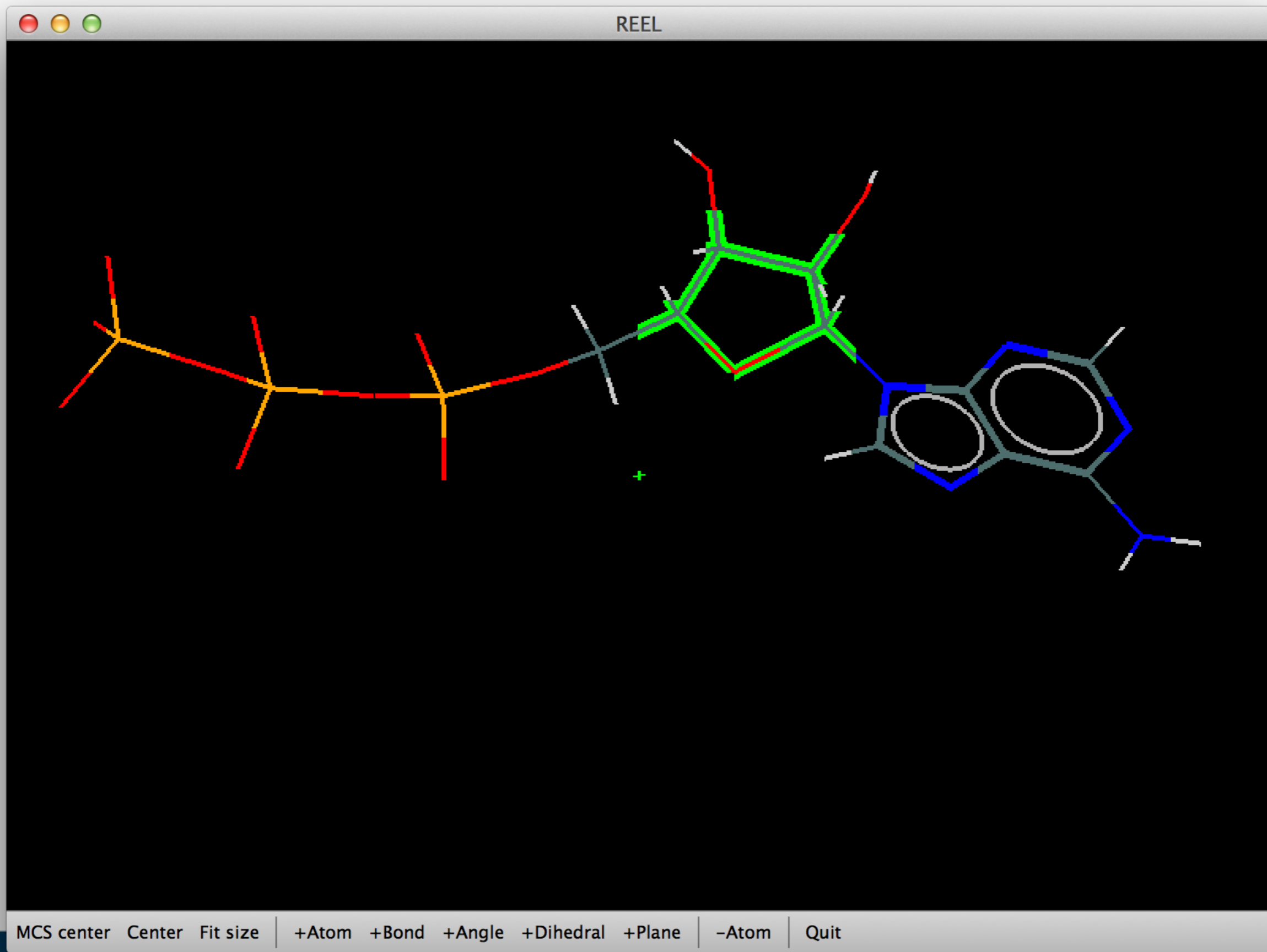
**Ribose(1)**



	?	atom_id_1	atom_id_2	atom_id_3	atom_id_4	atom_id_5	restrained_to	Pucker
1		C4'	C3'	C2'	C1'	O4'		--++--

View preferences loaded





Restraints Editor Especially Ligands (REEL)

Simple Optimisation    eLBOW Optimisation    AM1 Optimisation    Search Components    Find unique code





LIG

Atoms(12)    Bonds(11)    Angles(18)    Dihedrals(3)    Planes(6)    Chirals    **CisTrans(1)**    Chirals Implicit    BoatChair    Ribose

	?	atom_id_1	atom_id_2	restrained_to	E/Z
1		C03	C02	E	E



Chemical Components Search

 View in REEL    
  View info    
  View similar    
  View help

	Code	Name
1	HOH	WATER
2	1CU	COPPER ION, 1 WATER COORDINATED
3	2OF	FERROUS ION, 2 WATERS COORDINATED
4	3OF	HYDRATED FE (III) ION, 2 WATERS COORDINATED
5	543	CALCIUM ION, 6 WATERS PLUS ETHANOL COORDINATED
6	CD1	CADMIUM ION, 1 WATER COORDINATED
7	CD3	CADMIUM ION, 3 WATERS COORDINATED
8	CD5	CADMIUM ION, 5 WATERS COORDINATED
9	CO5	COBALT ION, 5 WATERS COORDINATED
10	DOD	DEUTERATED WATER
11	HC0	2 IRON/2 SULFUR/6 CARBONYL/1 WATER INORGANIC CLUST...
12	HC1	2 IRON/2 SULFUR/5 CARBONYL/2 WATER INORGANIC CLUST...
13	HCN	2 IRON/2 SULFUR/3 CARBONYL/2 CYANIDE/WATER/METHYLE...
14	KO4	POTASSIUM ION, 4 WATERS COORDINATED
15	MN5	MANGANESE ION, 5 WATERS COORDINATED
16	MN6	MANGANESE ION, 6 WATERS COORDINATED
17	MO1	MAGNESIUM ION, 1 WATER COORDINATED
18	MO2	MAGNESIUM ION, 2 WATERS COORDINATED
19	MO3	MAGNESIUM ION, 3 WATERS COORDINATED
20	MO4	MAGNESIUM ION, 4 WATERS COORDINATED
21	MO5	MAGNESIUM ION, 5 WATERS COORDINATED
22	MO6	MAGNESIUM ION, 6 WATERS COORDINATED
23	MTO	BOUND WATER
24	MW1	MANGANESE ION, 1 WATER COORDINATED

Input search string

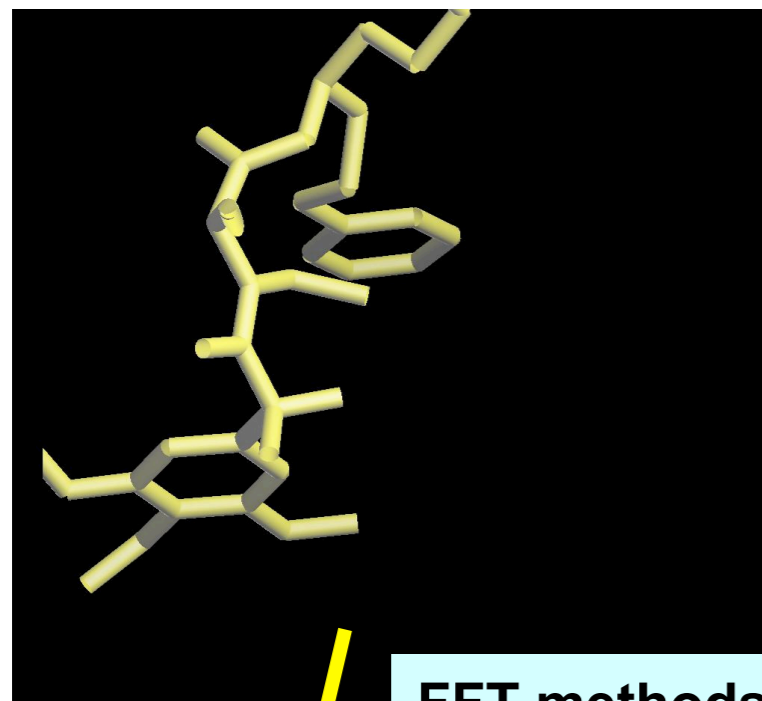
# Ligand Fitting



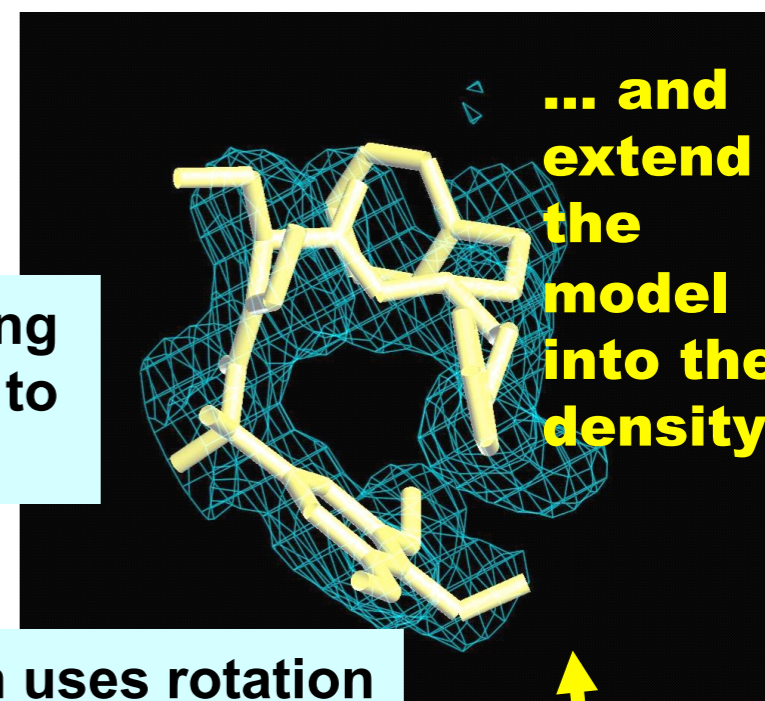
# Approach

- Where is the ligand?
  - Choose the largest region of contiguous density
- What are rotatable bonds?
  - Analyze ligand for allowed rotations
- What is the orientation of the ligand?
  - Fit core of ligand
- What is the conformation of the ligand?
  - Trace the ligand out from the core

# Automated Ligand Fitting



The best fit is measured using the correlation of the model to the electron density



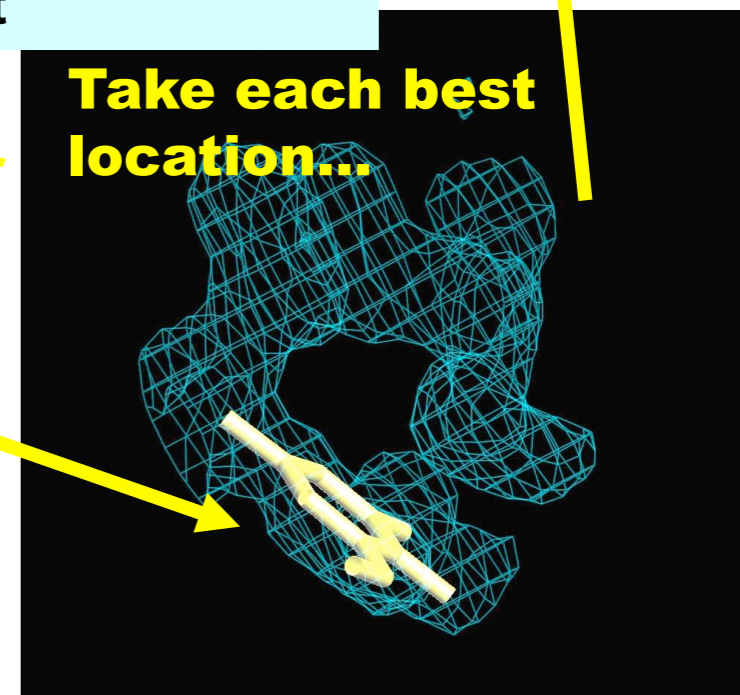
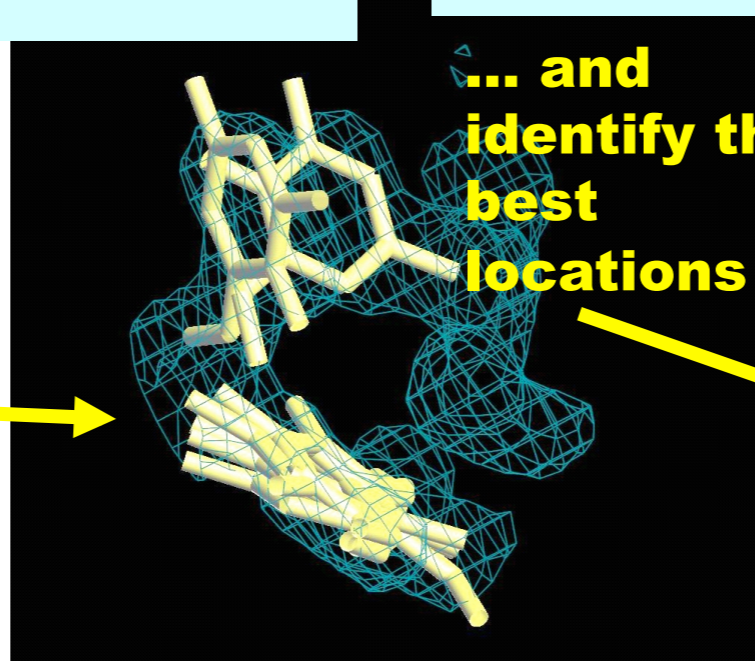
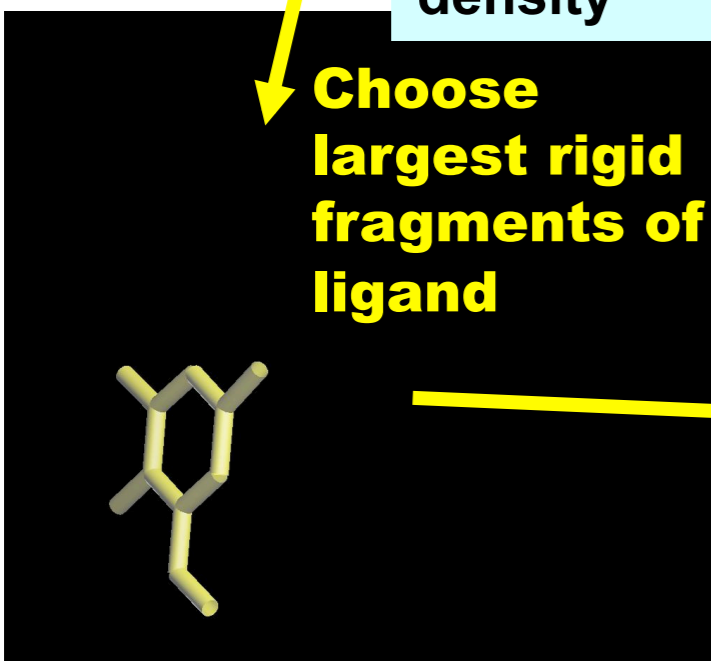
FFT methods are used to rapidly fit fragments to density

The extension uses rotation around torsion angles to find the best fit

**Choose largest rigid fragments of ligand**

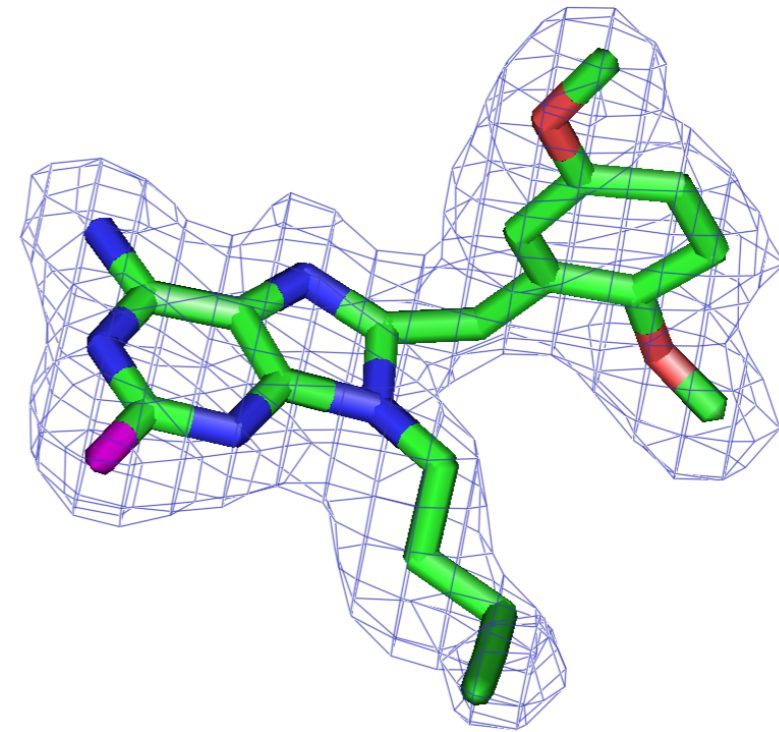
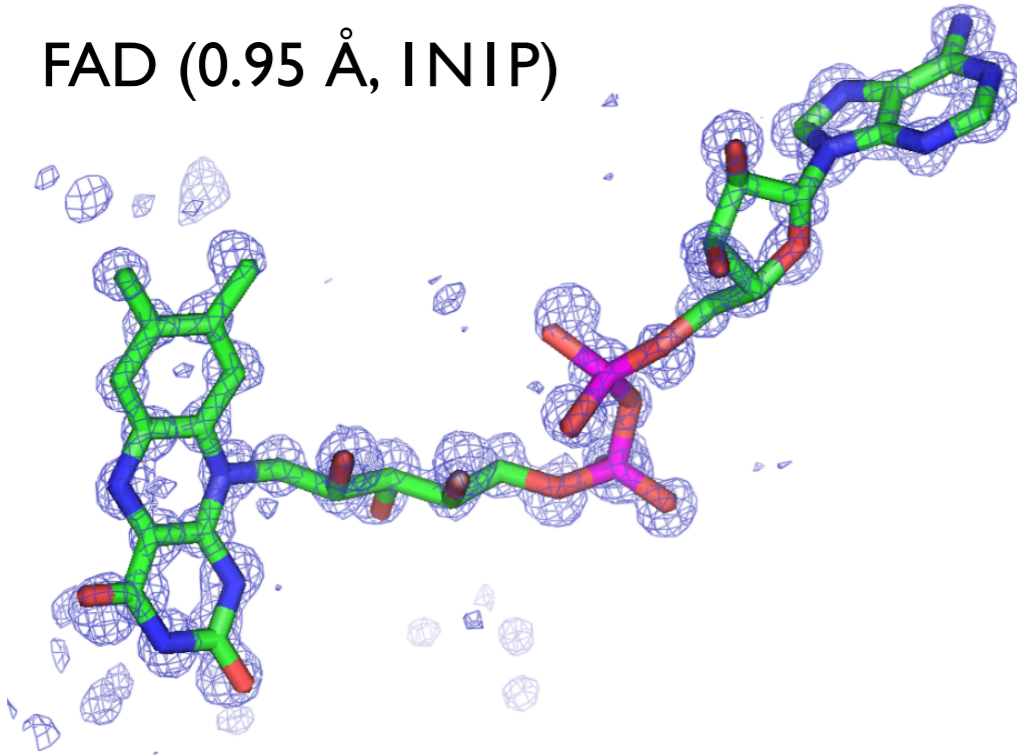
**... and identify their best locations**

**Take each best location...**



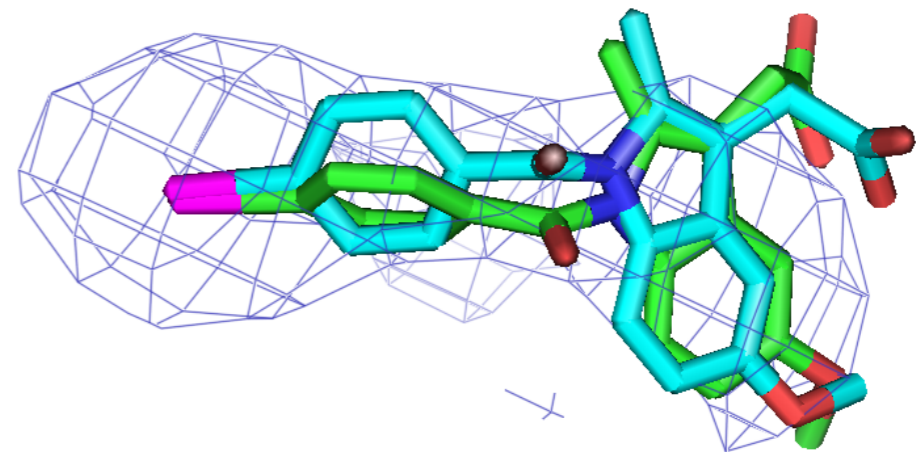
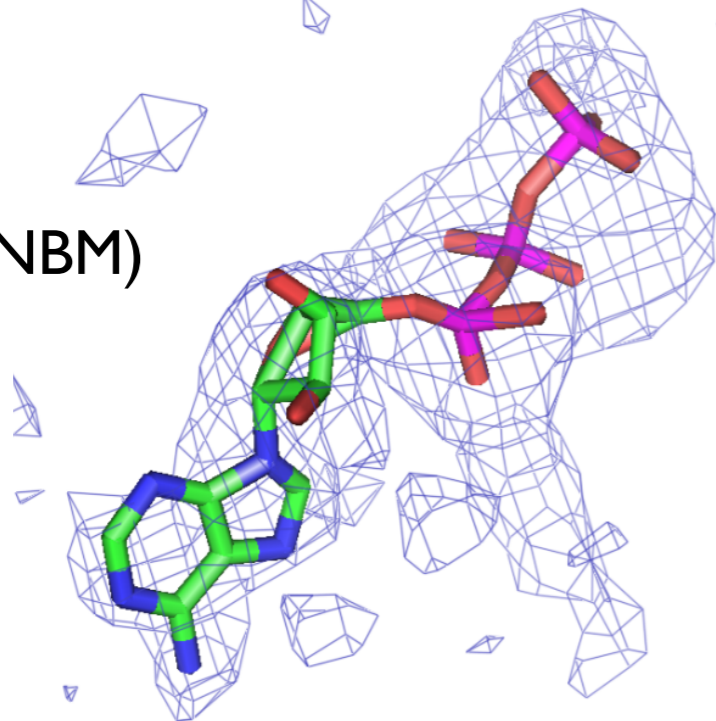
# Fitting Over a Range of Resolutions

FAD (0.95 Å, 1NIP)



8-(2,5-DIMETHOXY-BENZYL)-2-FLUORO-9-PENT-9H-PURIN-6-YLAMINE (2.2 Å 1UYI)

ATP (3 Å, 1NBM)



(1-(4-IODOBENZOYL)-5-METHOXY-2-METHYL-INDOLE-3-ACETIC ACID (4.5 Å, 1PGF)

# Restraints in phenix.(real\_space\_)refine

- LINK records have no impact
- Automatically accesses the “standard” residues restraints
- Automatically links the “standard” residues
- Parameter “link\_all=True” links
  - Covalent ligands
  - Carbohydrates
  - Metal ions

# phenix.(real\_space\_)refine (continued)

- RNA/DNA restraints
  - Base pair hydrogen bonding
  - Base pair planarity
  - Base stacking (parallelity)
- Secondary Structure restraints
- NCS restraints
- Custom bonds & angles using edits
- Restraints are written to .geo file including non bonded interactions

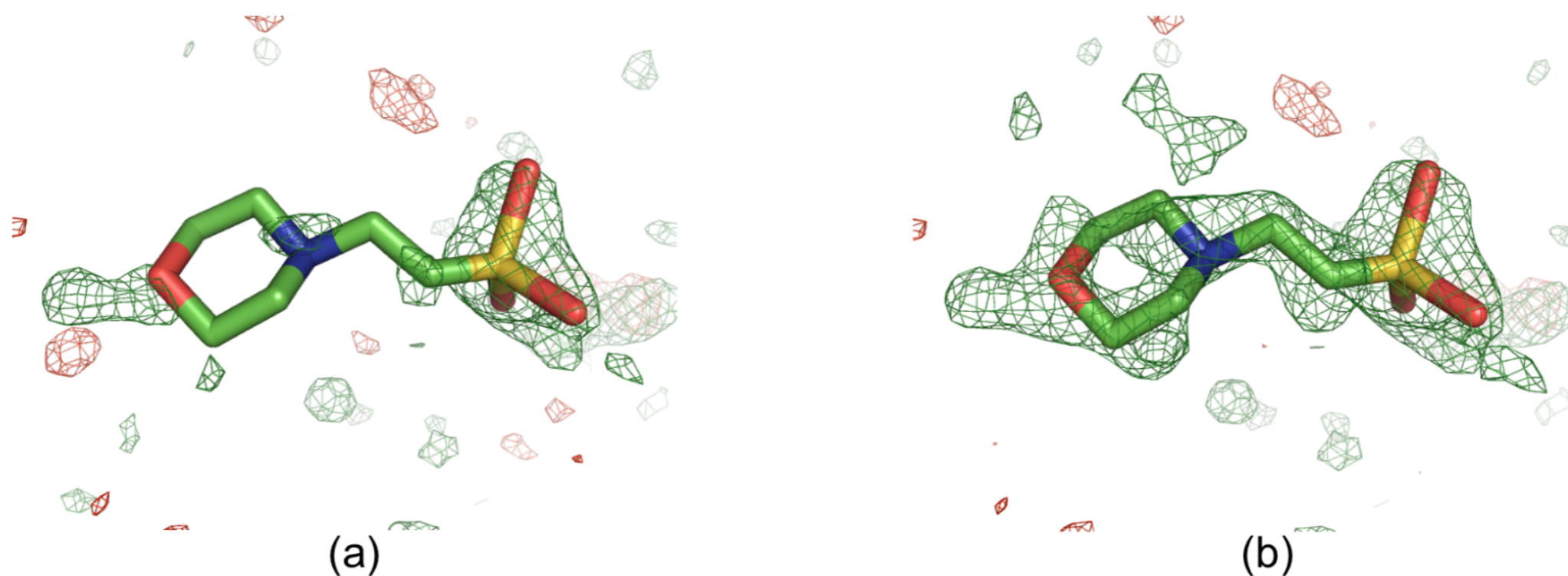
# Ligands? Water?

- Be careful when inheriting a model
- Can contain ligands and water molecules
- Must be able to “see” them in the map



# Ligands

# Polder OMIT Maps



Solvent molecule MES88 in structure 1ABA. The positive and negative mFobs-DFmodel OMIT difference density is displayed in green and red, respectively. (a) OMIT map contoured at  $\pm 3\sigma$ . (b) Polder map contoured at  $\pm 3\sigma$ . In the OMIT map, there is only some density for the O, N and S atoms. The polder map shows difference electron density for the entire molecule

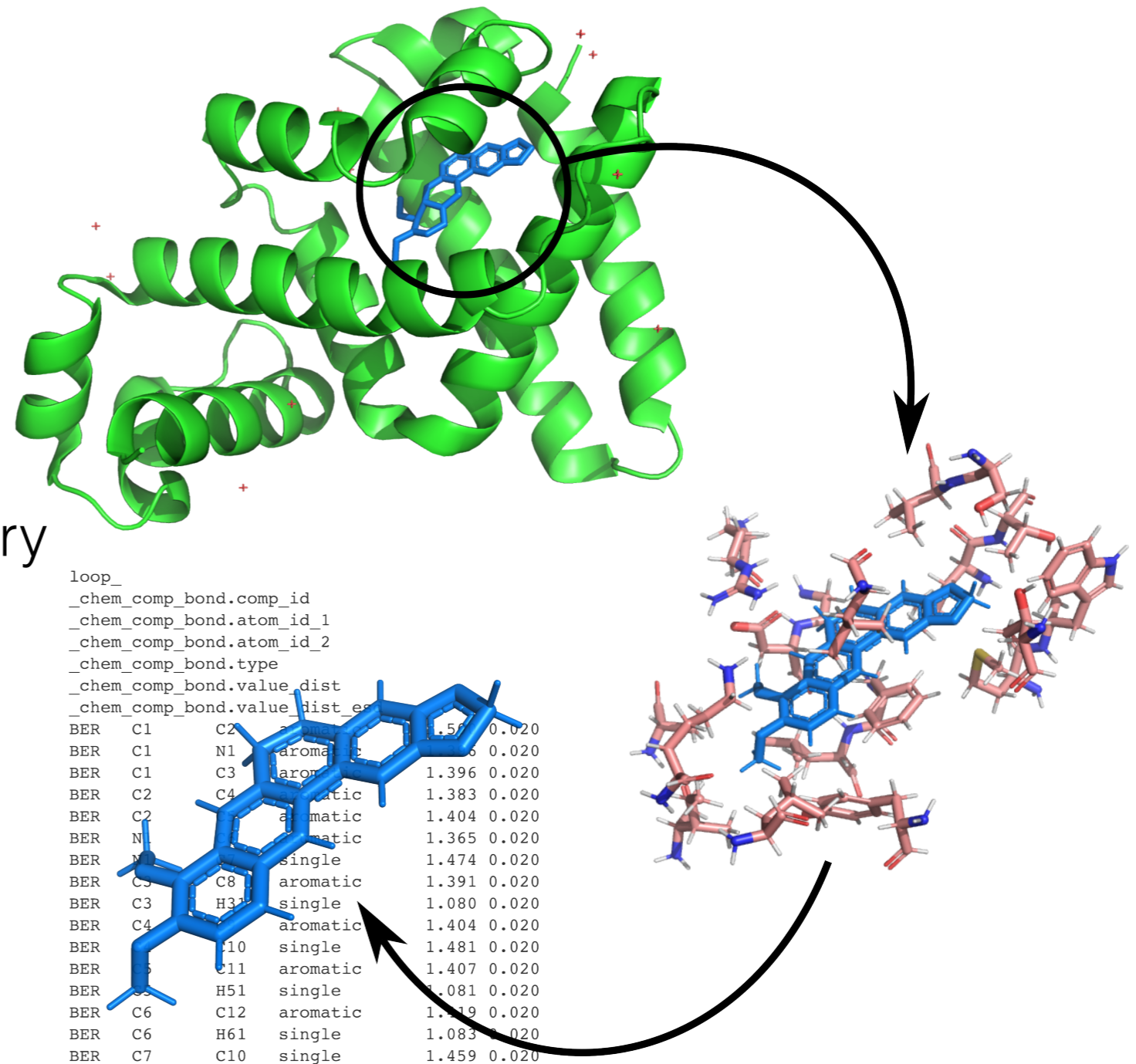
# QM Restraints

- Generates restraints of ligands using Quantum Mechanics *in situ*
- There are two ways of using QMR
  - In *phenix.refine*
  - In a standalone program *mmtbx.quantum\_interface*
- Python3 installers from the bottom of the download page
- For ORCA, set \$PHENIX\_ORCA

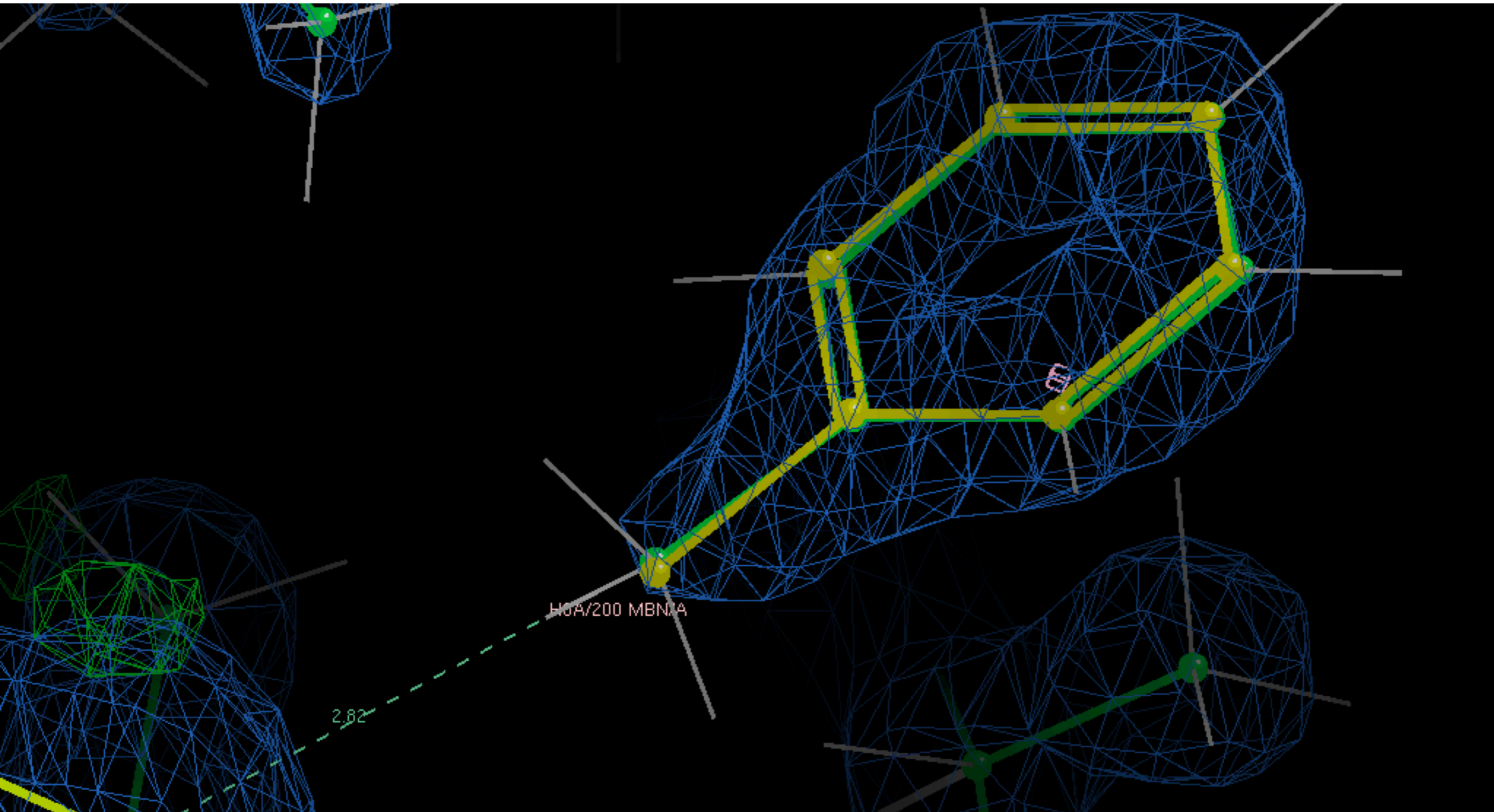
# QMR

*In situ* restraints generation

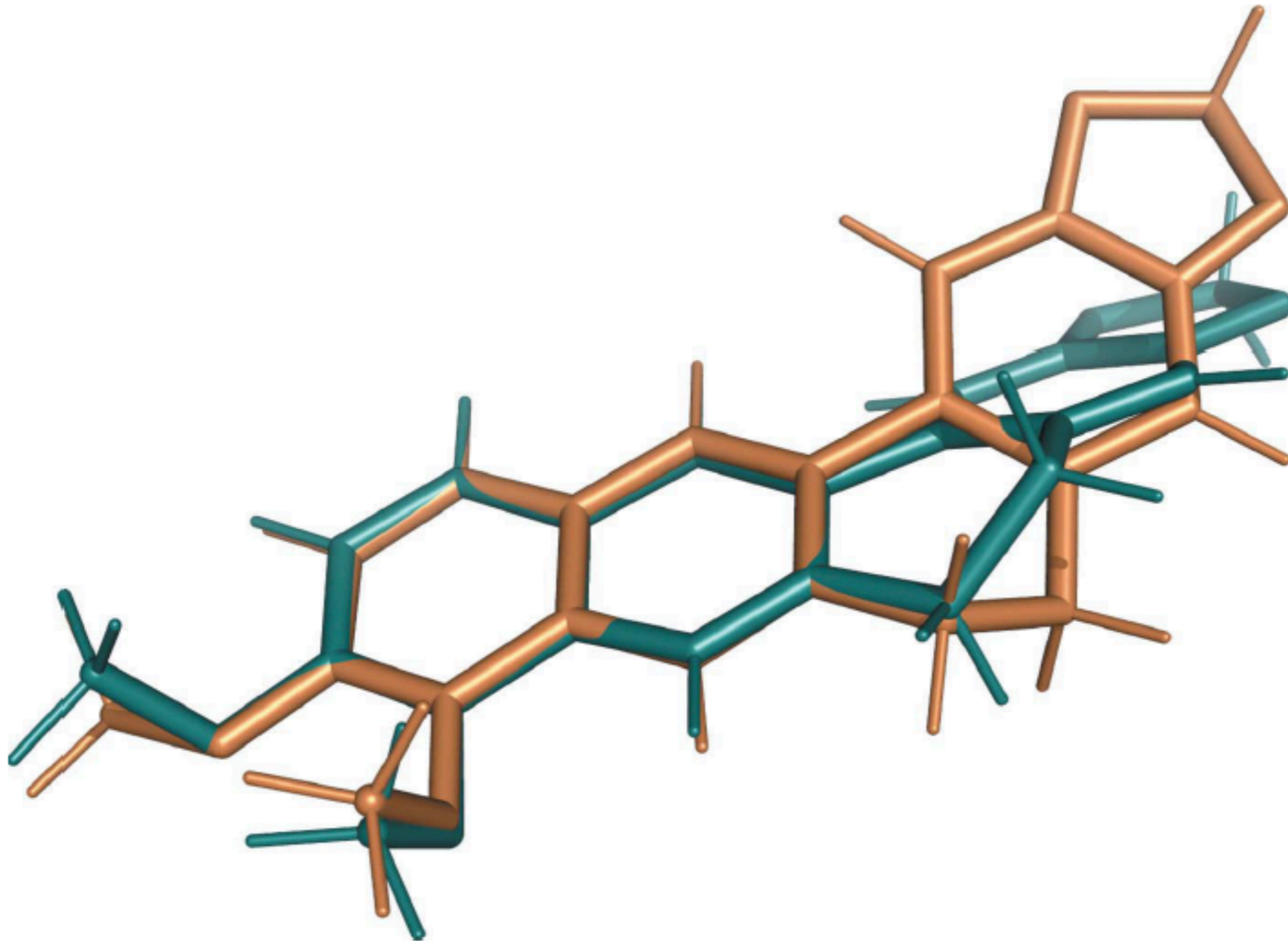
- Carve out the ligand environment
- Minimise the ligand geometry *in situ*
- Transfer geometry values to restraints (and write to disk)
- Refinement with modified restraints



# *In situ*



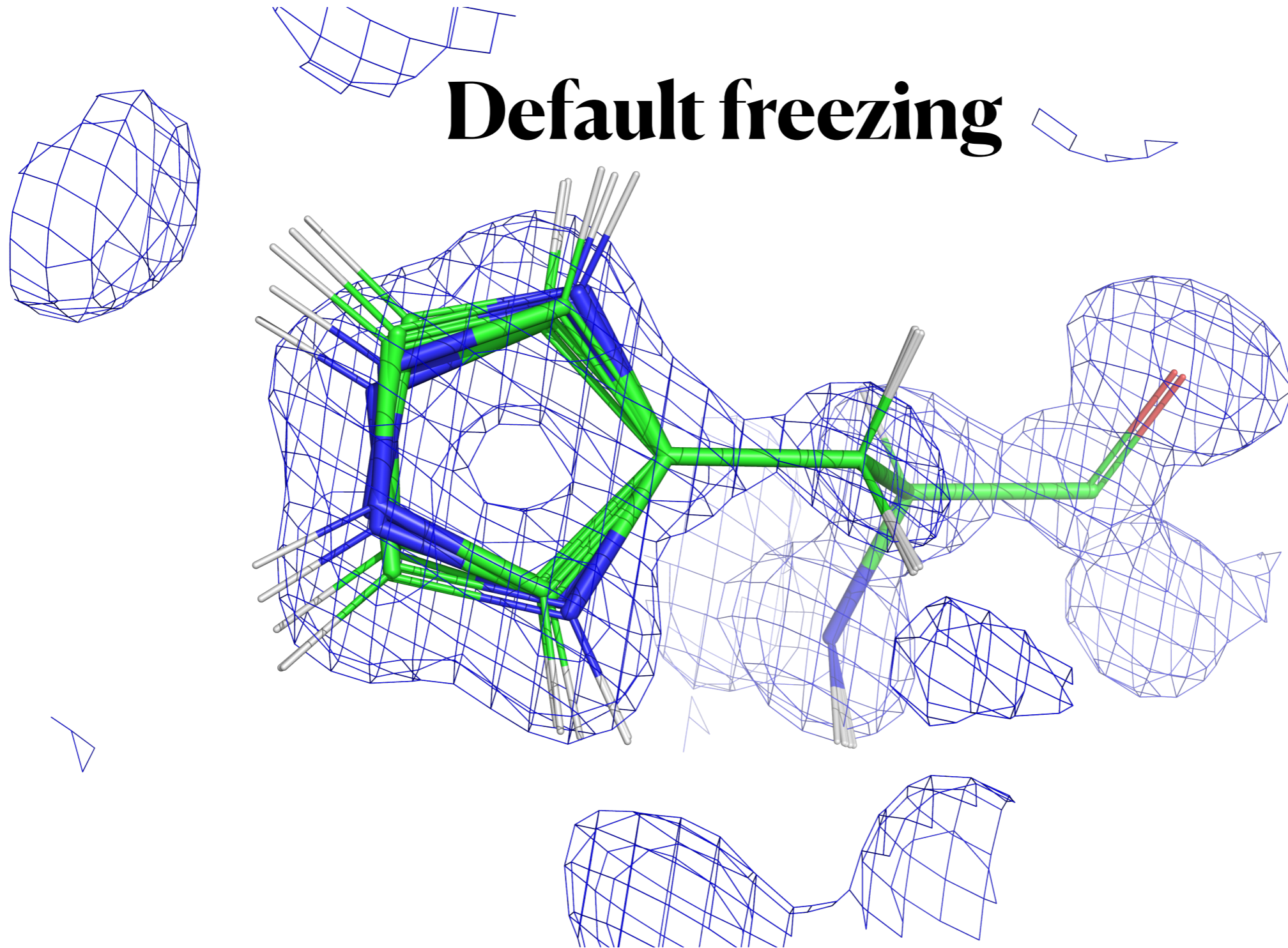
# BER in 3vw2



# QM Flipping

- Generate the three pronation states of HIS
- Flip  $\chi_2$  180 for total of six configurations
- Perform a QM geometry minimisation of the side-chain while freezing the heavy atoms (non-H) for rest of model
- Compare metrics
  - Energy
  - H-bonds
  - RMSD

# Default freezing



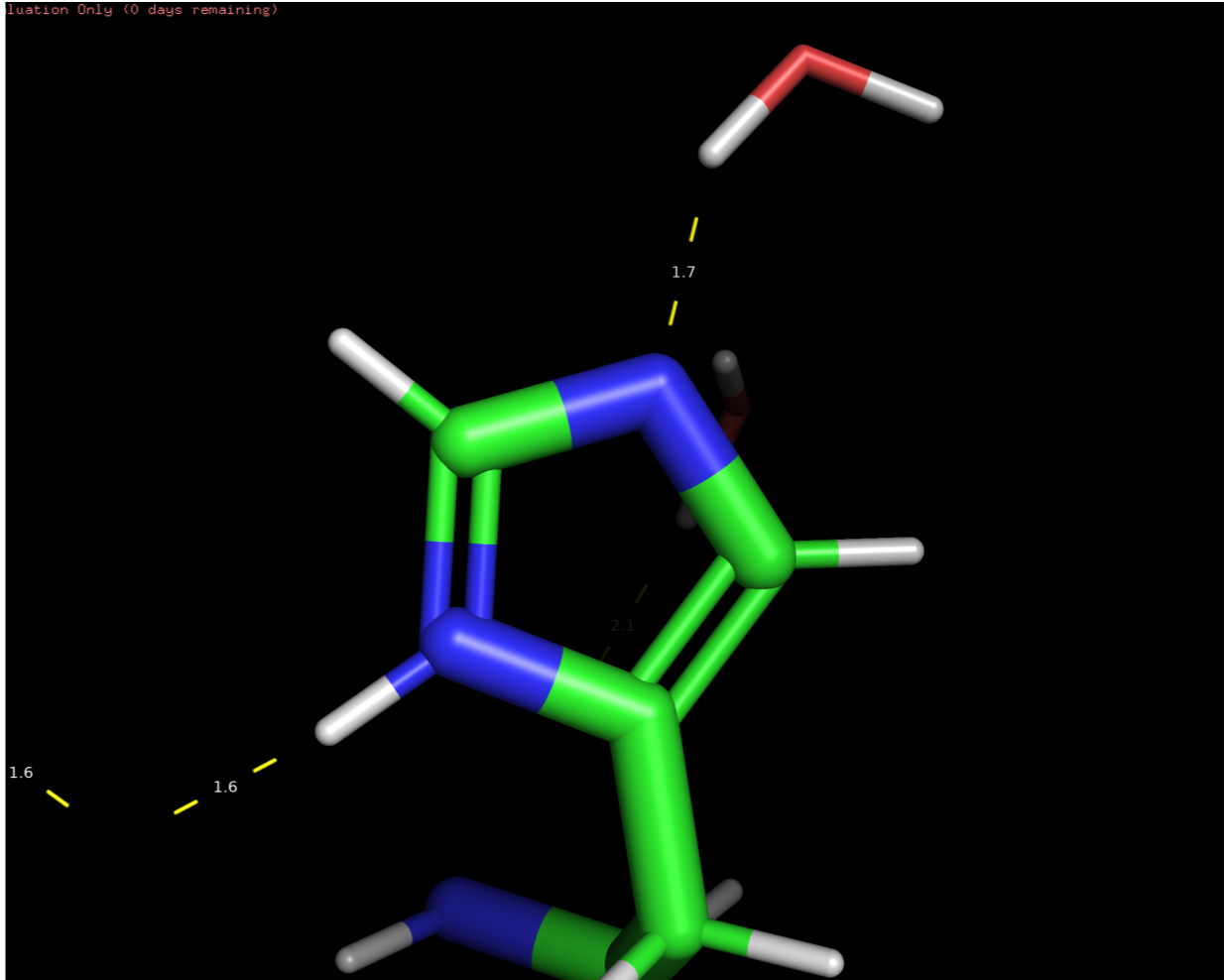
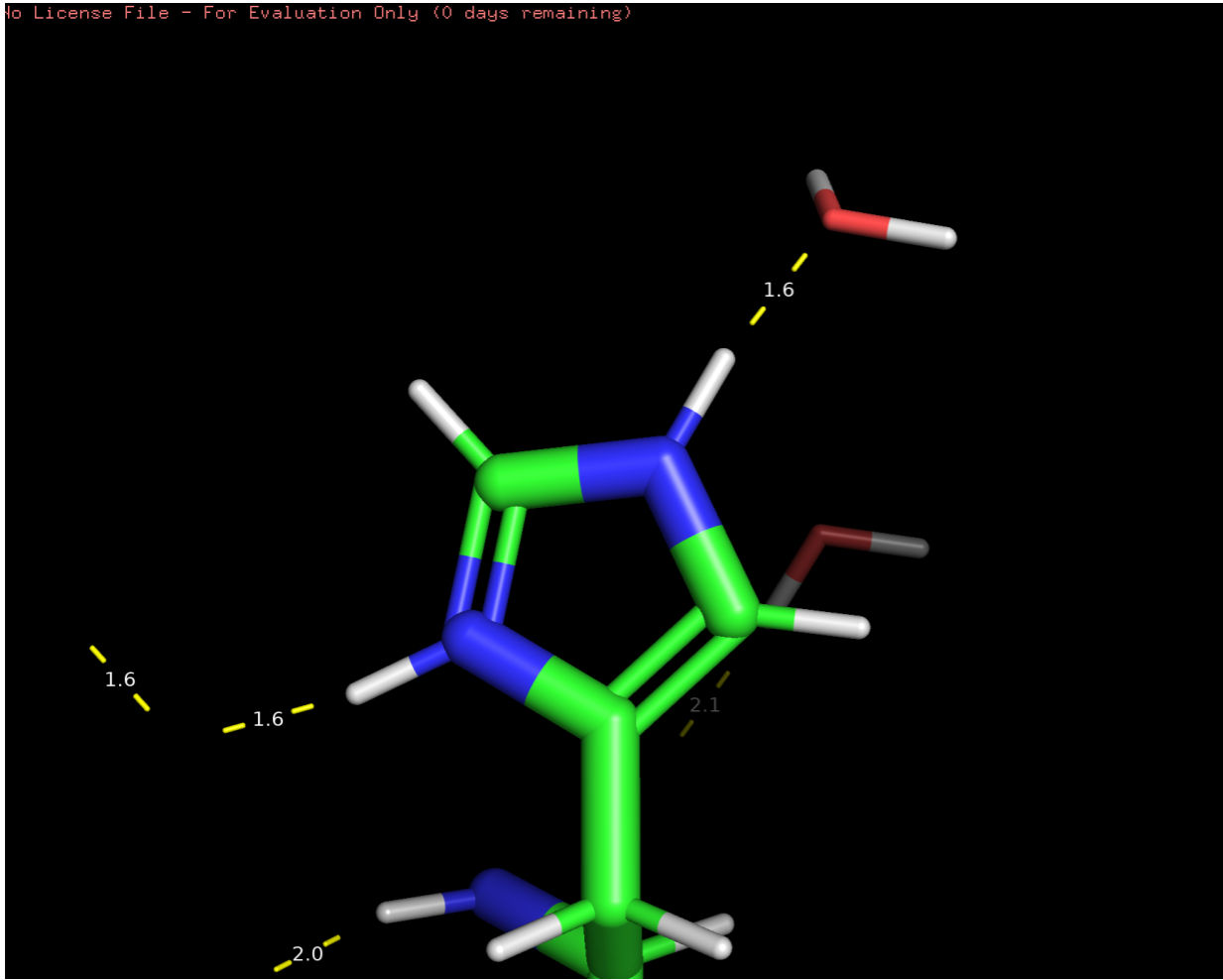


# Metrics

	Configuratio	Energy (kcal/	$\Delta E$	H bonds	r.m.s.d.	Rotamer
0	HD1, HE2					m90
1	HD1, HE2	-1019.6	2.6	14	0.04	m90
2	HD1 only	-1022.2	0.0	14	0.04	m90
3	HE2 only	-1003.0	19.3	13	0.05	m90
4	HD1, HE2	-1004.0	18.3	12	0.29	m-70
5	HD1 only	-1004.4	17.9	12	0.38	m-70
6	HE2 only	-1009.4	12.9	11	0.32	m-70

The QMF results for the histidine resseq 4 in chain A of PDB 4rj2

**Minimised geometry of two histidine configurations of histidine in chain A and resid 4 of PDB 4rj2.**



# Summary

- eLBOW & ReadySet! perform better when provided with better input. (GIGO)
- Need to know something about the ligand
  - Hierarchy of input file value
- Check your .geo file for confirmation of restraints