# Ligands in Phenix

Generating & modifying for all scenarios

Nigel W. Moriarty

User's Meeting
ACA, July 2025

# 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

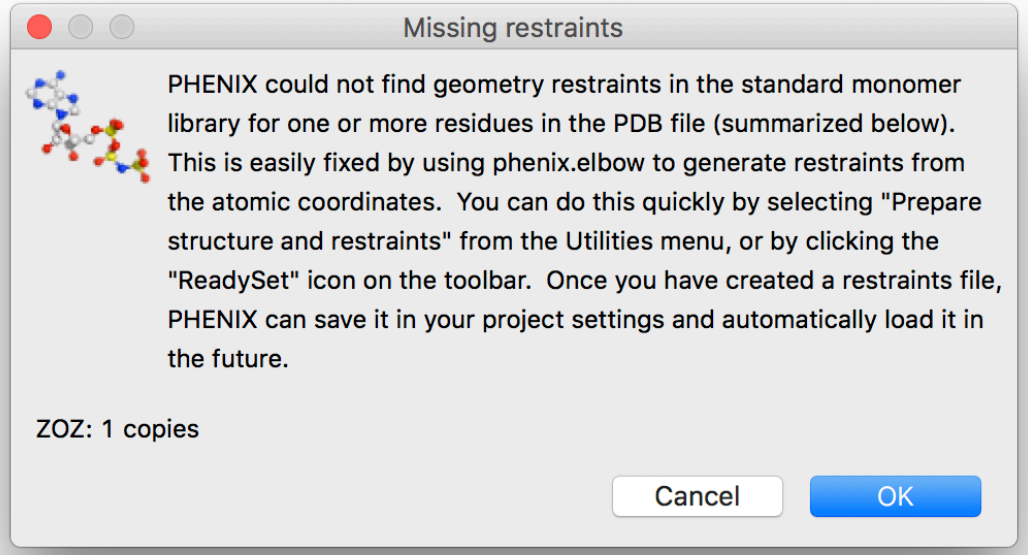
# Restraints in Action

- 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

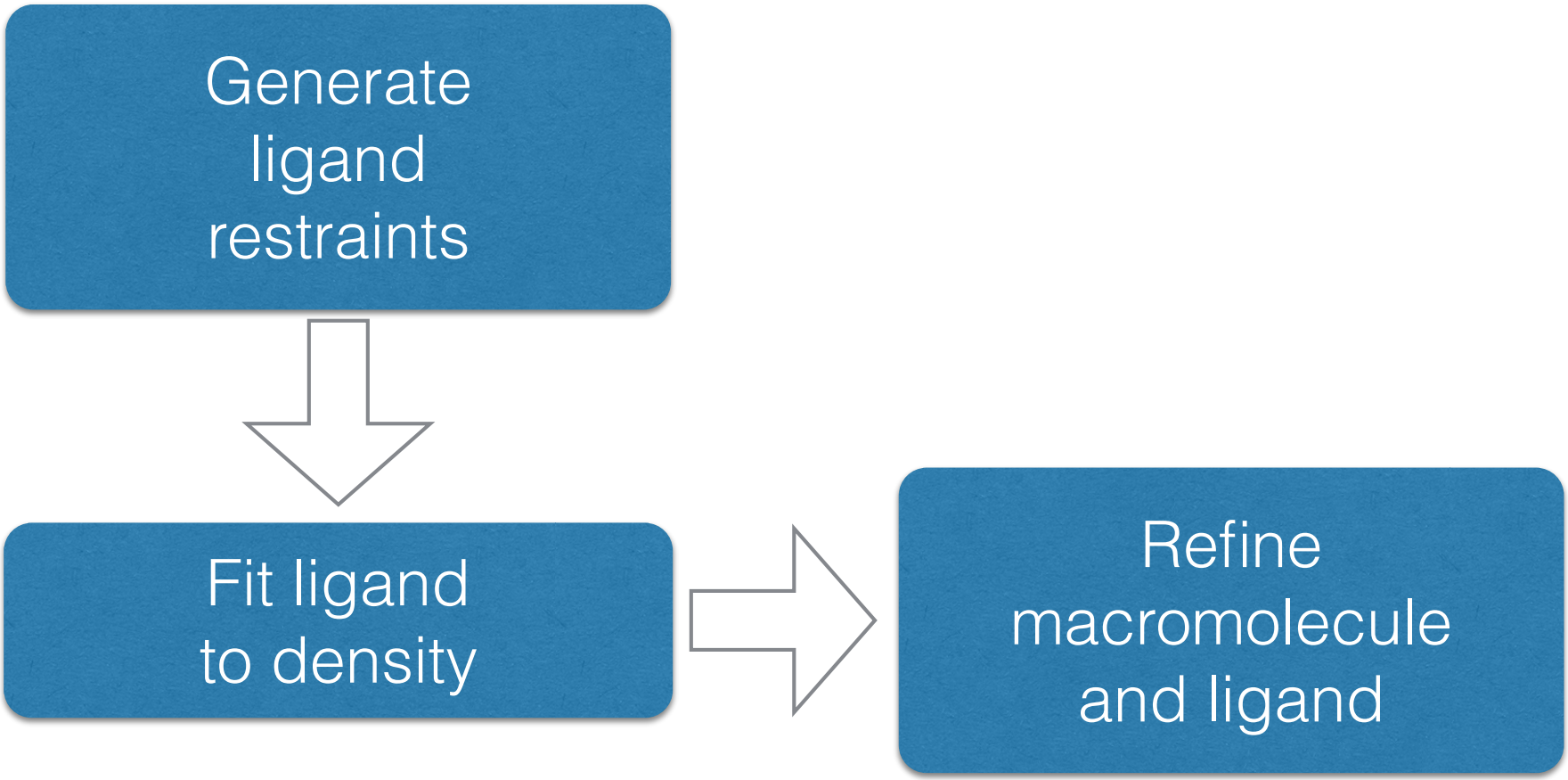
# Overview

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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Ligands in crystallography




BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# 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


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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# 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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Amino Acid

RCSB PDB PROTEIN DATA BANK

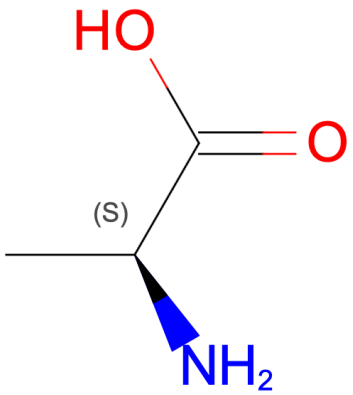
217,157 Structures from the PDB

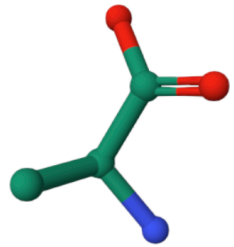
1,068,577 Computed Structure Models (CSM)

3D Structures | Enter search term(s), Entry ID(s), or sequence | Include CSM

Advanced Search | Browse Annotations | Help

PDB-101 | wwPDB | EMDataResource | NAKB | wwPDB Foundation | PDB-Dev



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

search

Find related ligands:

Similar Ligands (Stereospecific)
Similar Ligands (including Stereoisomers)
Similar Ligands (Quick Screen)
Similar Ligands (Substructure Stereospecific)
Similar Ligands (Substructure including Stereoisomers)

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Water

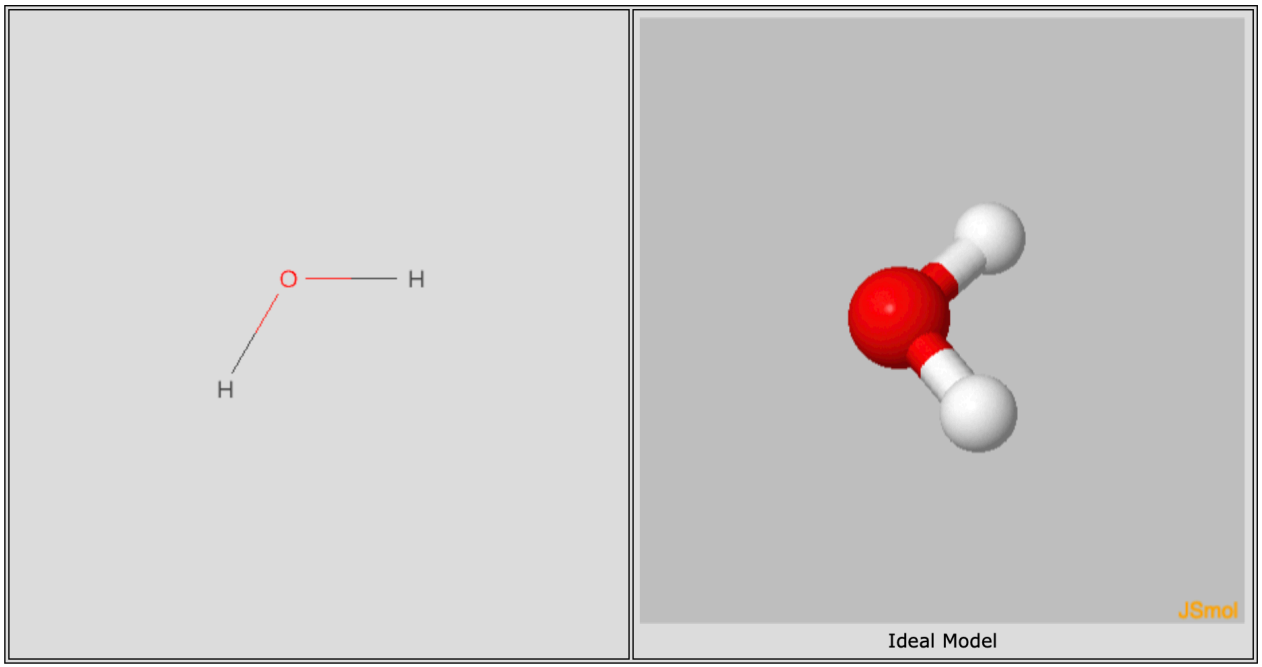
RCSB PDB
PROTEIN DATA BANK
RCSB PDB | Contact Us

Ligand Expo

Home Search Browse Download Ligand Expo Help

Chemical Details | Geometry | Atom Nomenclature | Downloads | Related Resources

## PDB Chemical Component HOH




Ideal Model

## Chemical Description

| | |
|---|---|
| **Name** | WATER |
| **Formula** | H2 O |
| **Formal charge** | 0 |
| **Molecular weight** | 18.015 g/mol |
| **Component type** | NON-POLYMER |

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Metal clusters

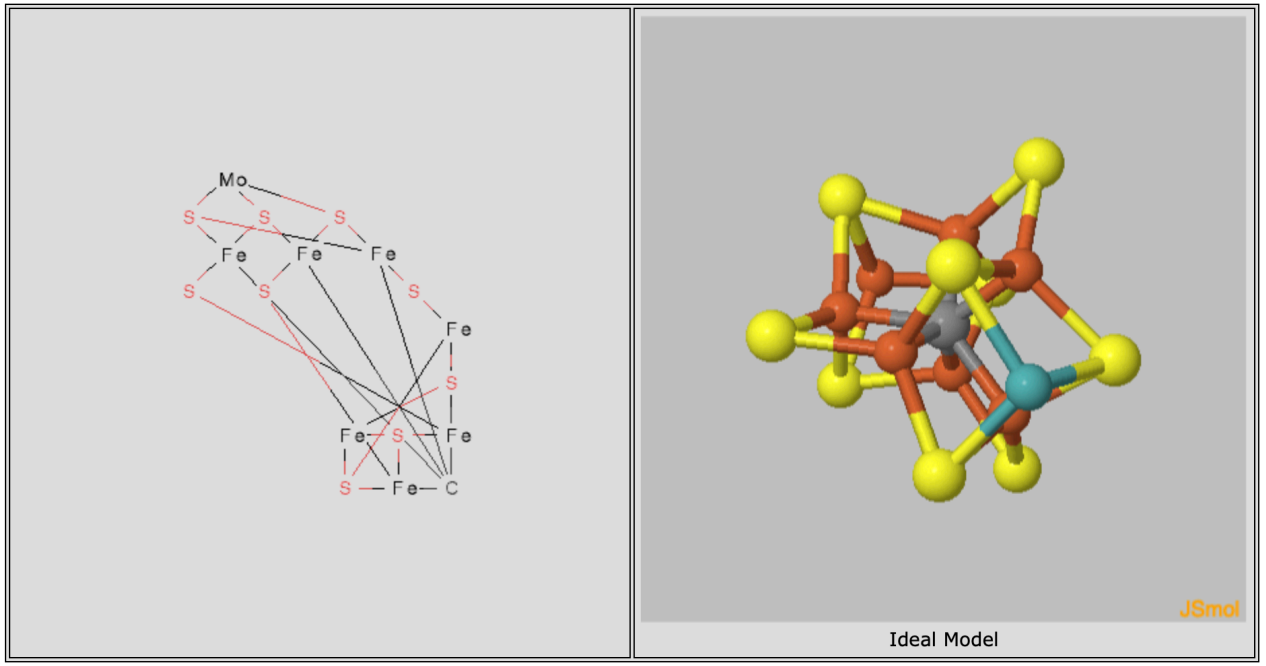
RCSB PDB
PROTEIN DATA BANK
RCSB PDB | Contact Us

Ligand Expo

Home Search Browse Download Ligand Expo Help

Chemical Details | Geometry | Atom Nomenclature | Downloads | Related Resources

**PDB Chemical Component ICS**




Ideal Model

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# 5-letter codes

RCSB PDB PROTEIN DATA BANK

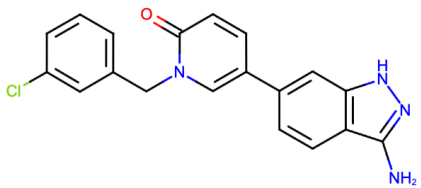
217,157 Structures from the PDB

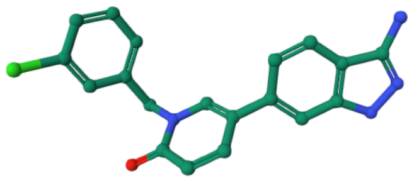
1,068,577 Computed Structure Models (CSM)

3D Structures | Enter search term(s), Entry ID(s), or sequence | Include CSM

Advanced Search | Browse Annotations | Help

PDB-101 | wwPDB | EMDataResource | NAKB | wwPDB Foundation | PDB-Dev



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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Human readable

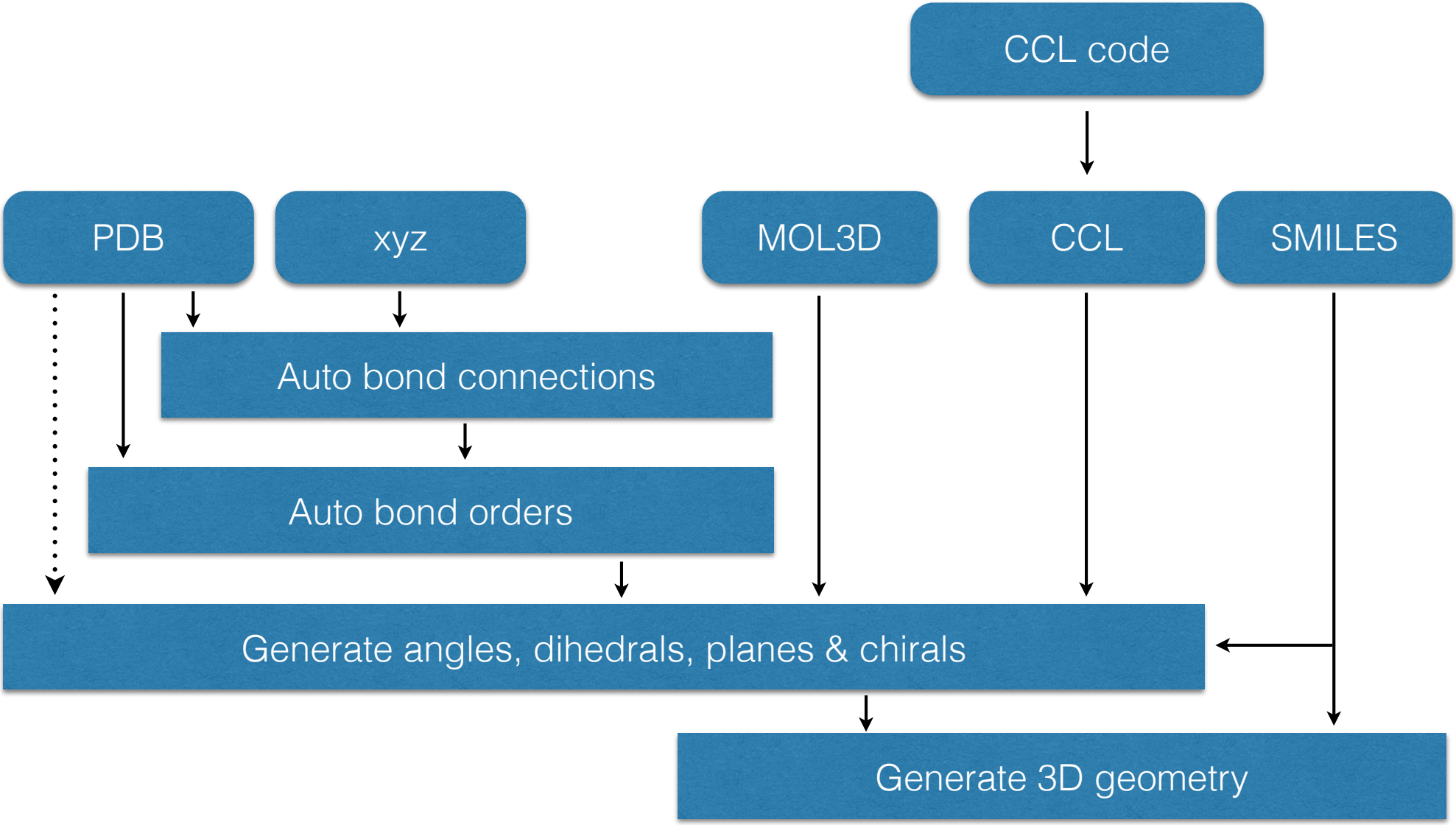
| Standard | | Human readable | |
|---|---|---|---|
| Uppercase | Lowercase | Uppercase | Lowercase |
| 1OI0 | 1oi0 | 1oi0 | 1oi0 |
| 1IJJ | 1ijj | 1iJJ | 1ijj |
| 4OCL | 4ocl | 4oCL | 4ocL |
| 5SS2 | 5ss2 | 5ss2 | 5ss2 |

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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

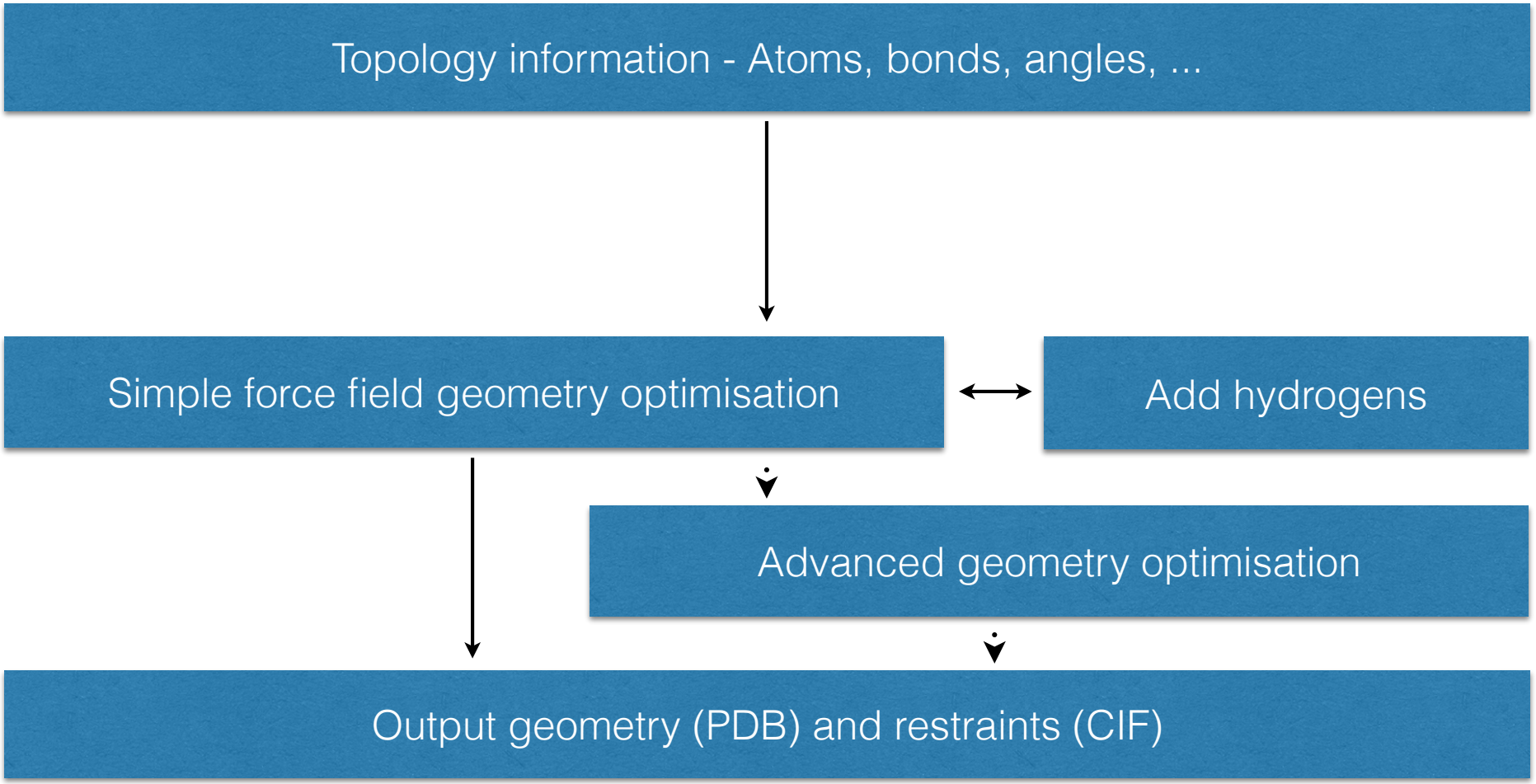
# Topology




BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Optimisation

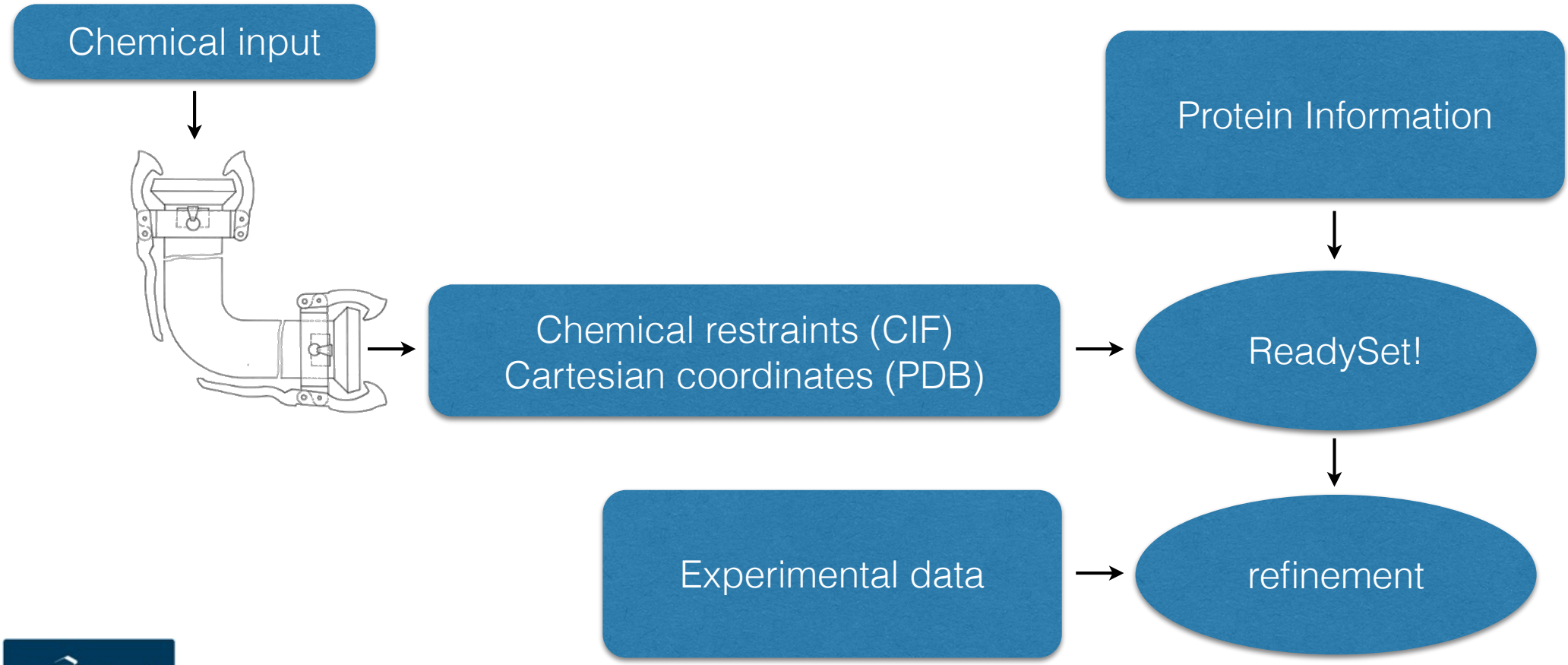



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Getting ready to refine

- Many details needed to prepare for structure refinement




BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# 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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# ReadySet!

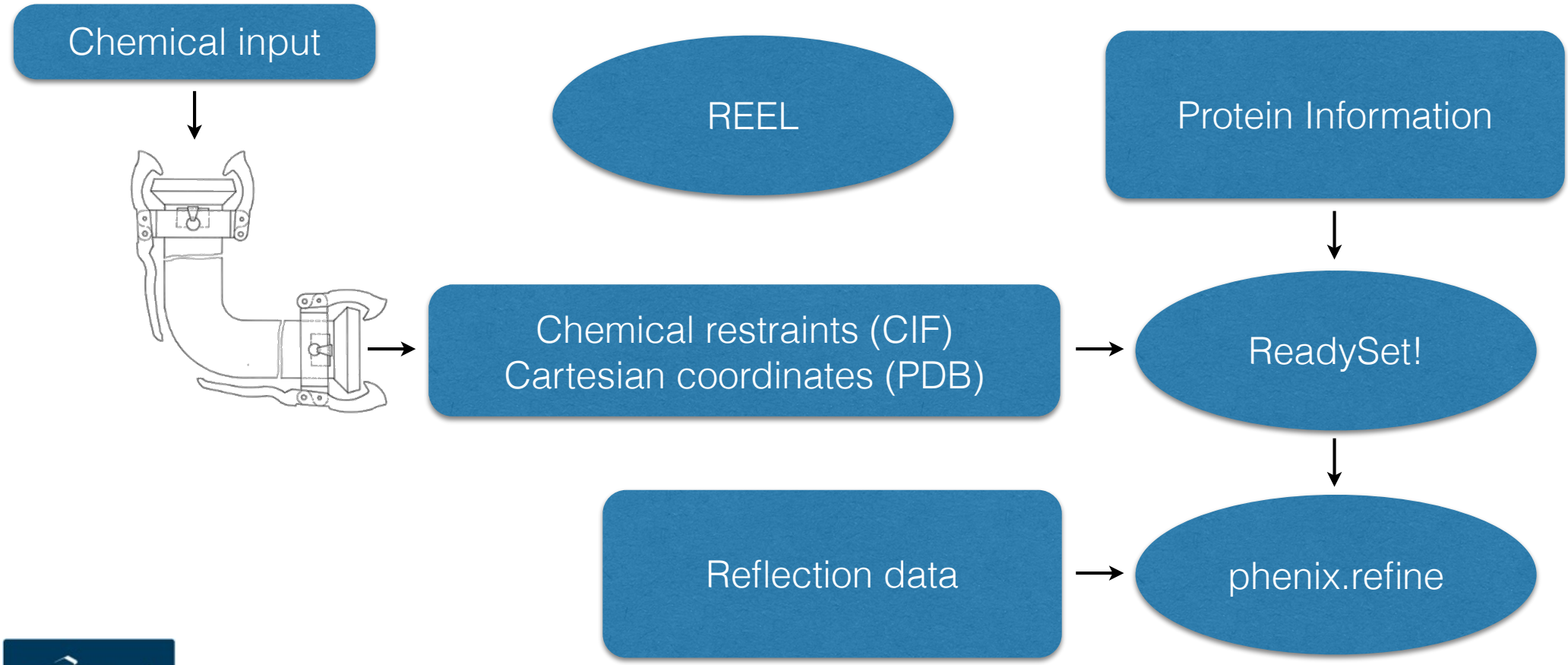
- Restraints CIF filename
- Restraints CIF directory
- LINKS to “edits”
- --dry-run to show ligand process pathway
- Metal coordination

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Restraints editing

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




BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Ligand Fitting


Phenix

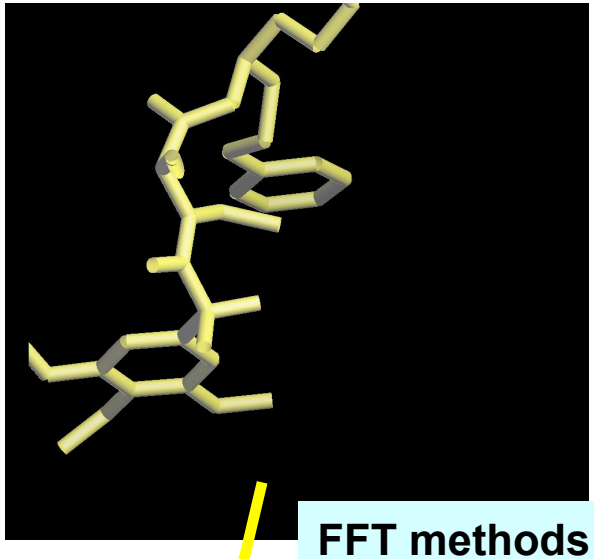
# Approach

- 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

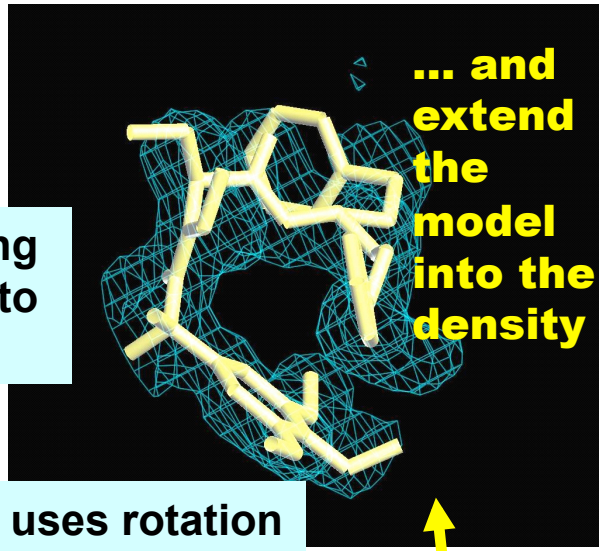
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# Automated Ligand Fitting

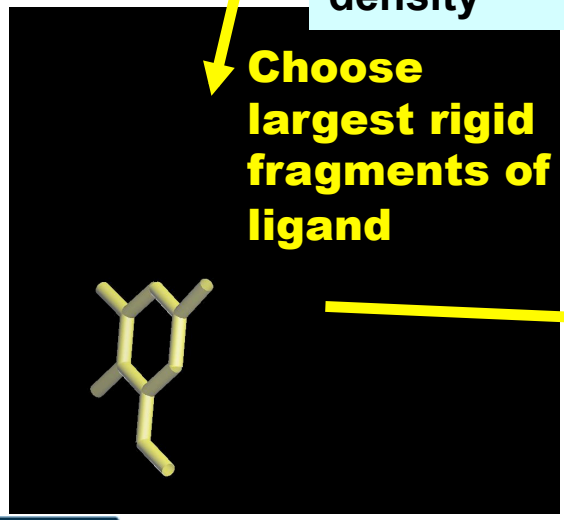


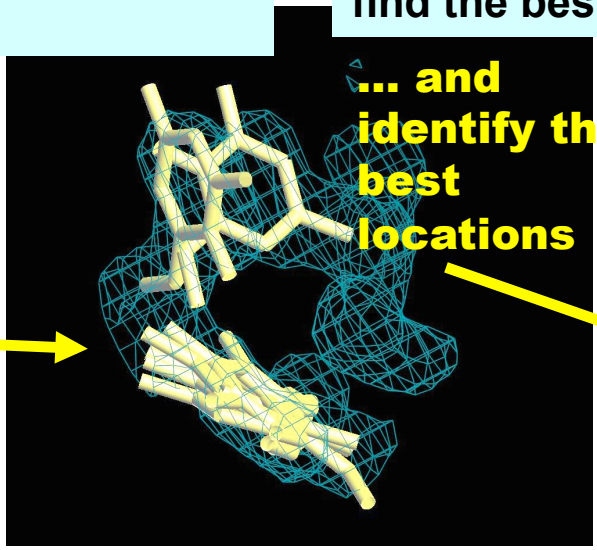
The best fit is measured using the correlation of the model to the electron density

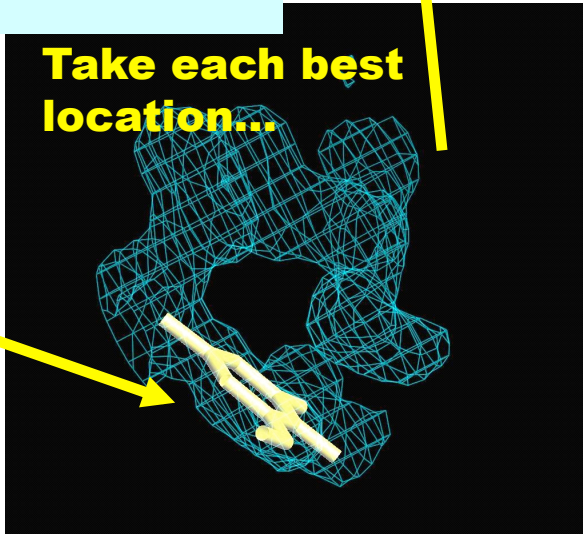



FFT methods are used to rapidly fit fragments to density

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




BERKELEY LAB
Lawrence Berkeley National Laboratory

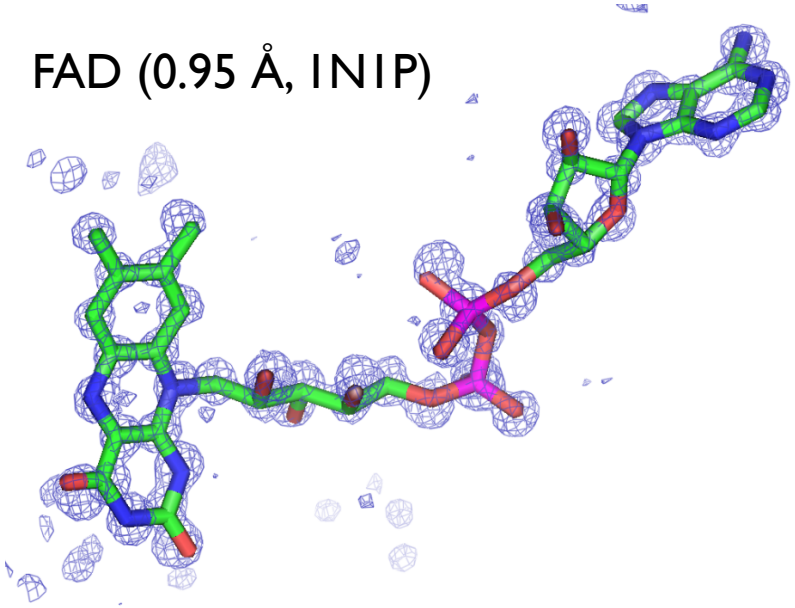
*Tom Terwilliger, LANL*

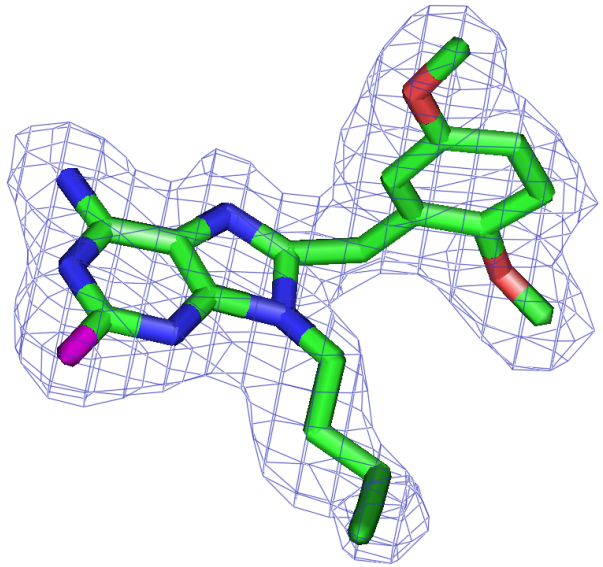
Phenix

*Acta Cryst.* 2006, **D62**:915-922.

# Fitting Over a Range of Resolutions

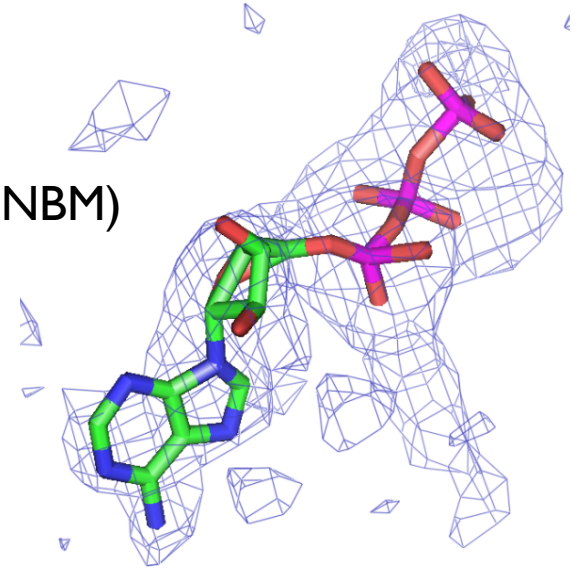
FAD (0.95 Å, 1N1P)

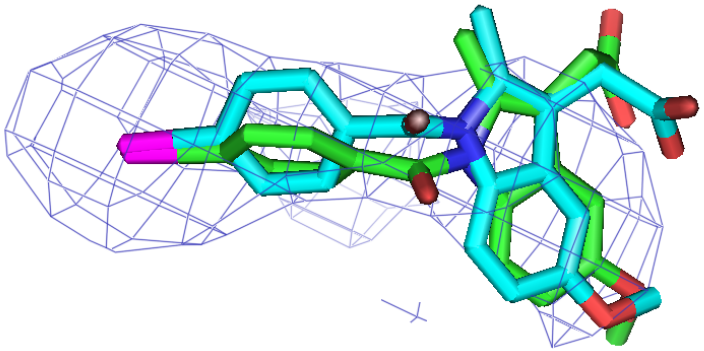




8-(2,5-DIMETHOXY-BENZYL)-2-FLUORO-9-PENT-9H-PURIN-6-YLAMINE (2.2 Å 1UYI)

ATP (3 Å, 1NBM)





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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# 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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

# 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

BERKELEY LAB
Lawrence Berkeley National Laboratory

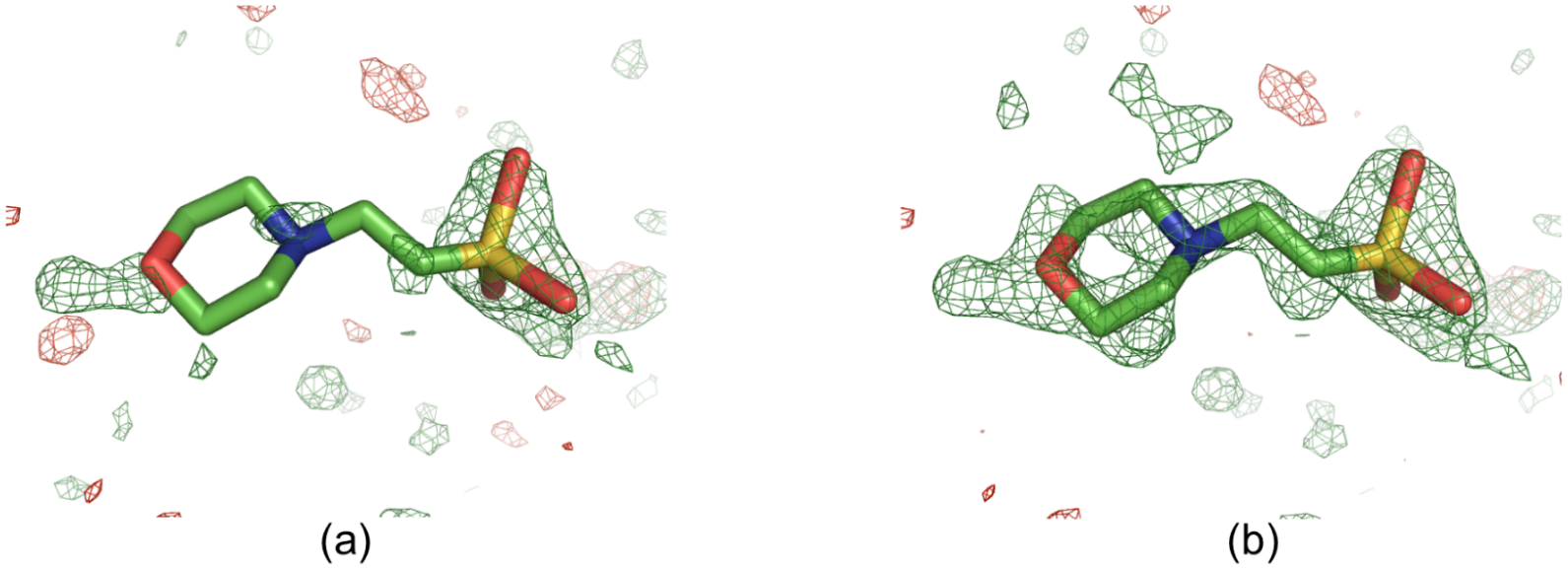
Phenix

# 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 +/-3σ. (b) Polder map contoured at +/-3σ. 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

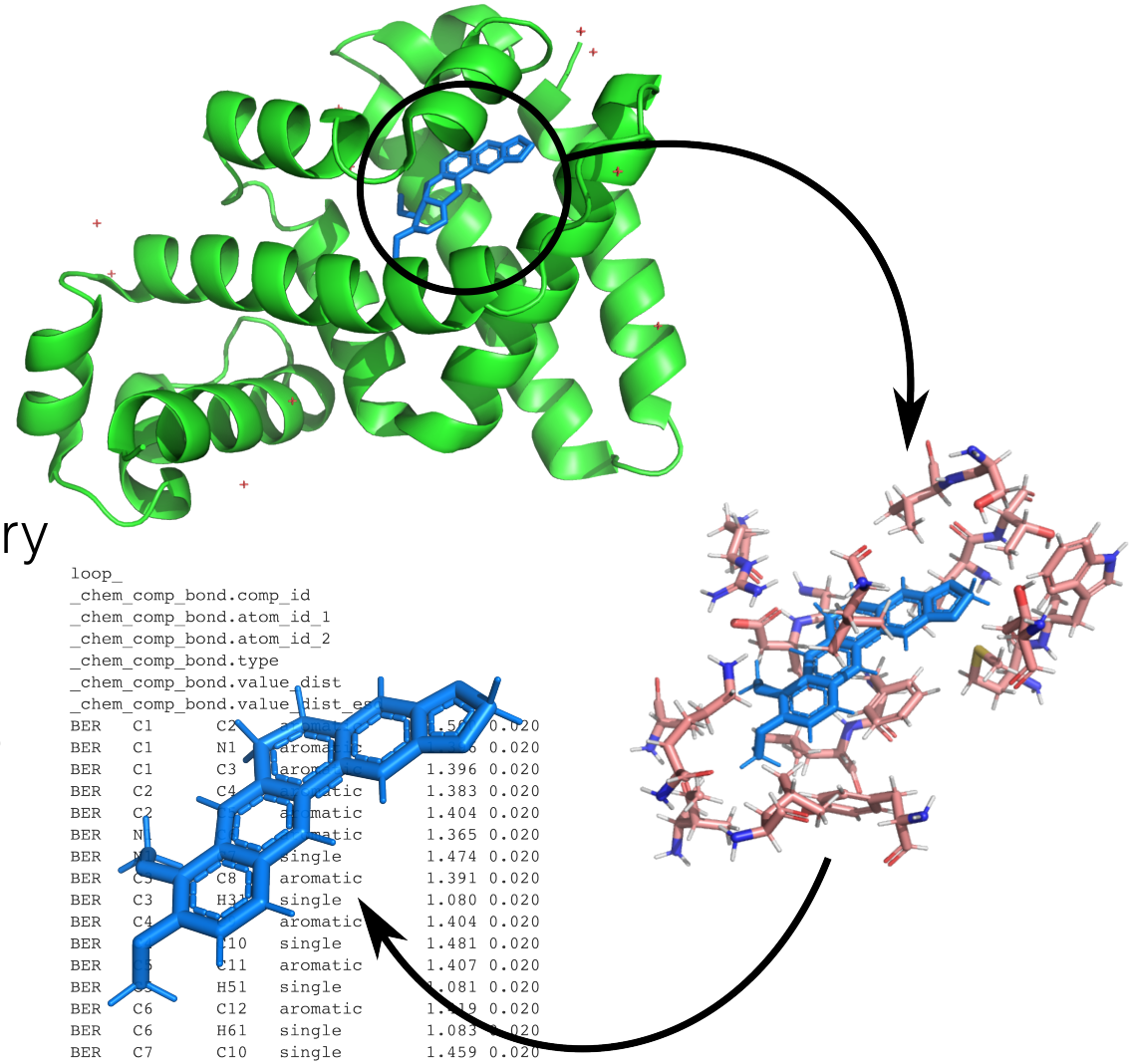
# QM Restraints

- 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

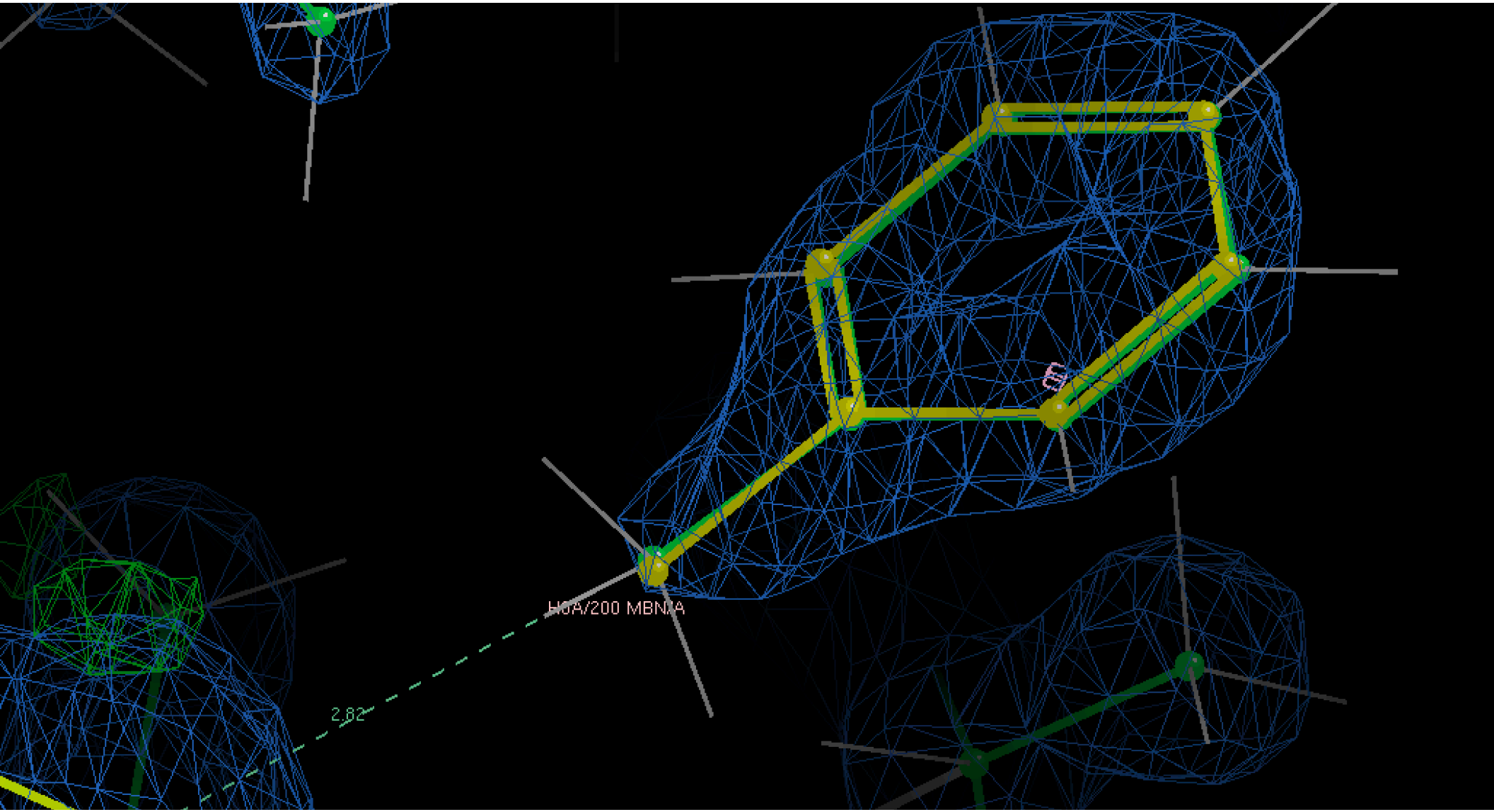
*In situ* restraints generation

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

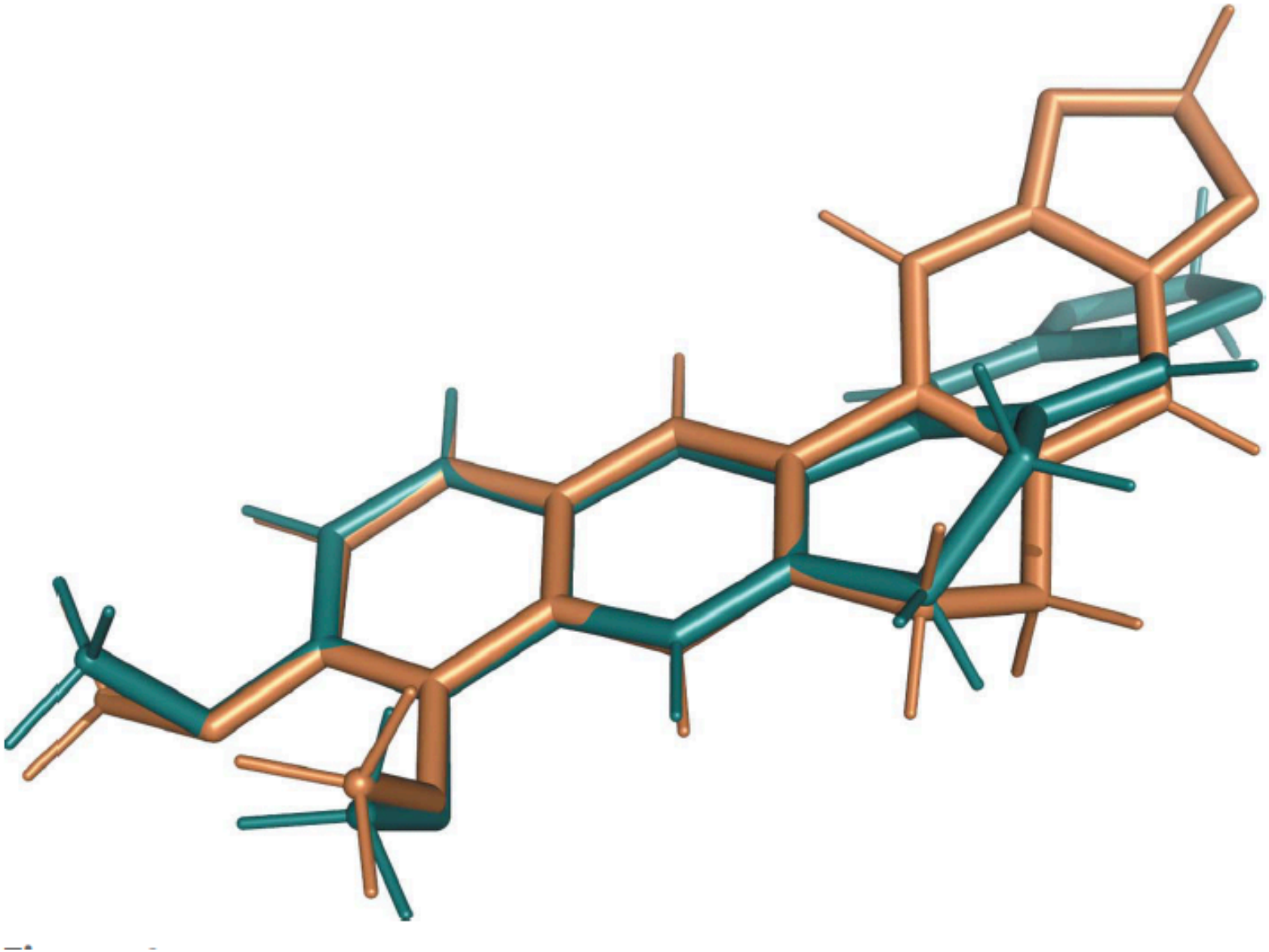


```
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_es
BER   C1      C2    aromatic      1.396 0.020
BER   C1      N1    aromatic      1.396 0.020
BER   C1      C3    aromatic      1.396 0.020
BER   C2      C4    aromatic      1.383 0.020
BER   C2      C5    aromatic      1.404 0.020
BER   N1      C4    aromatic      1.365 0.020
BER   N1      C7    single        1.474 0.020
BER   C3      C8    aromatic      1.391 0.020
BER   C3      H31   single        1.080 0.020
BER   C4      C9    aromatic      1.404 0.020
BER   C5      C10   single        1.481 0.020
BER   C5      C11   aromatic      1.407 0.020
BER   C5      H51   single        1.081 0.020
BER   C6      C12   aromatic      1.419 0.020
BER   C6      H61   single        1.083 0.020
BER   C7      C10   single        1.459 0.020
```

# *In situ*



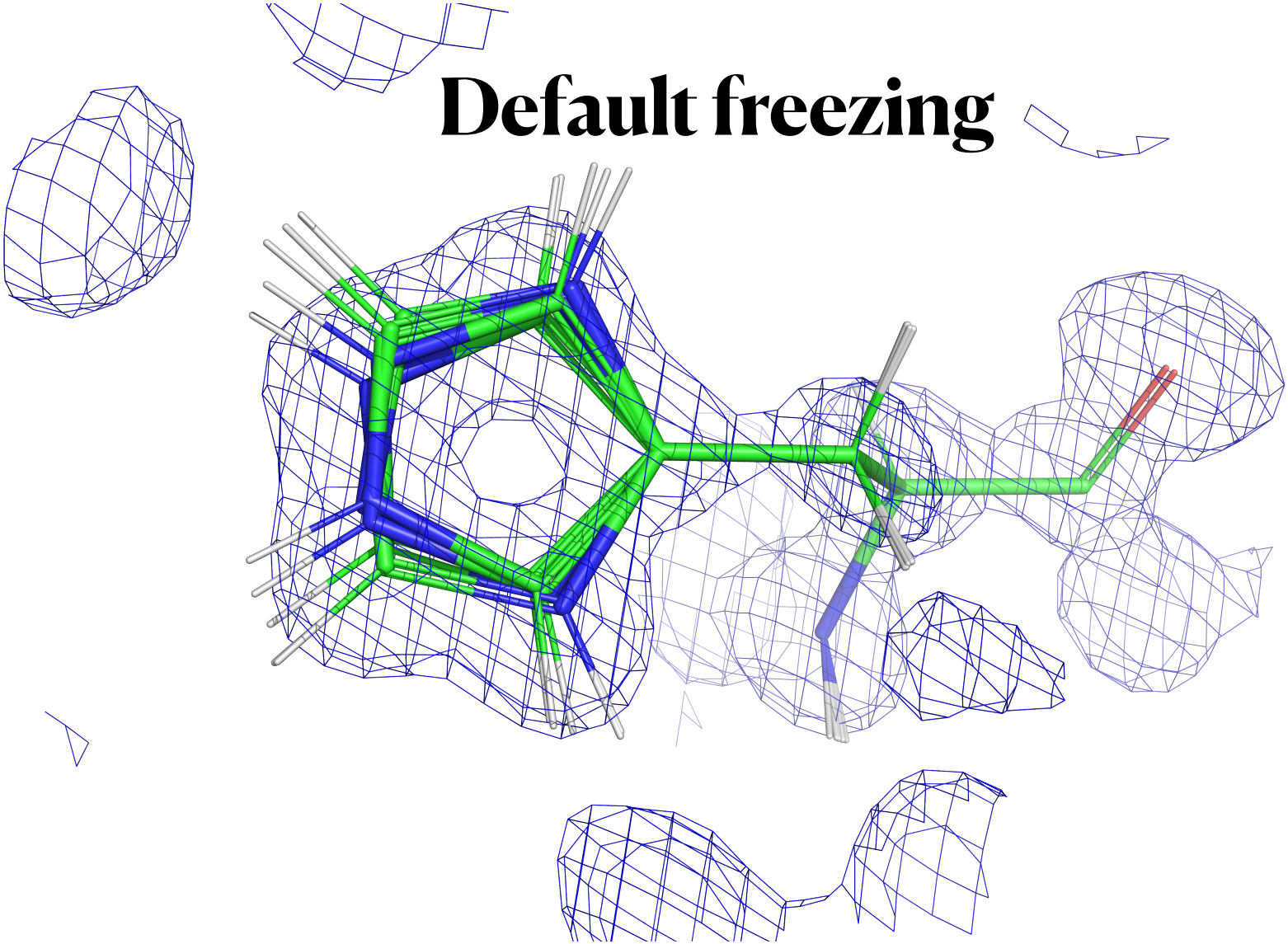
# BER in 3vw2



# 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

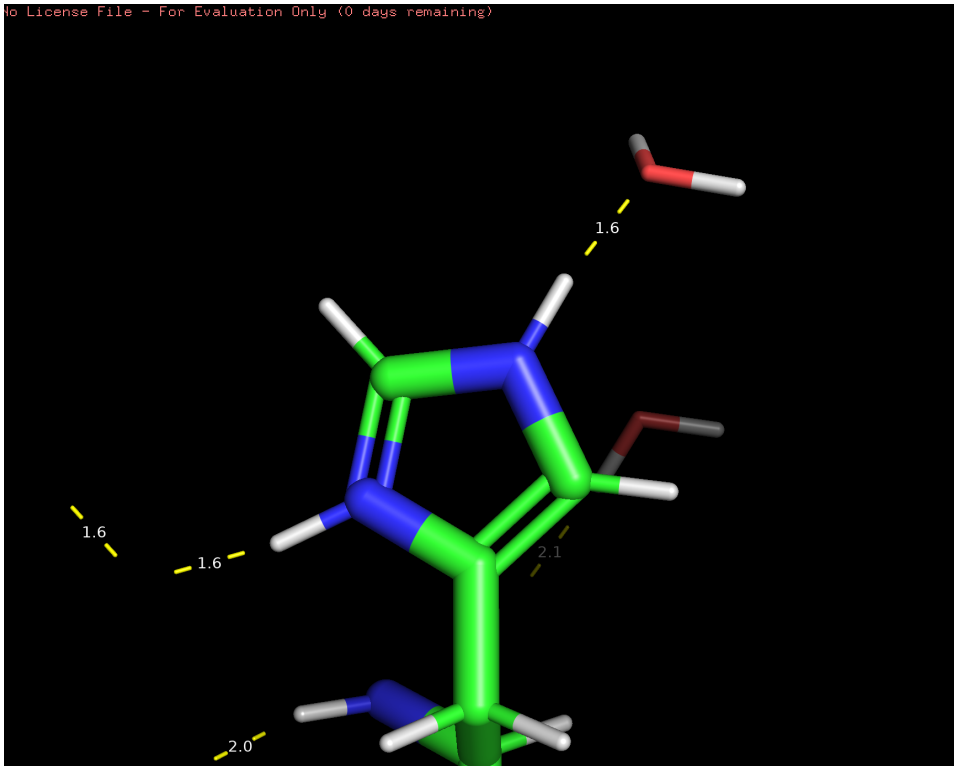


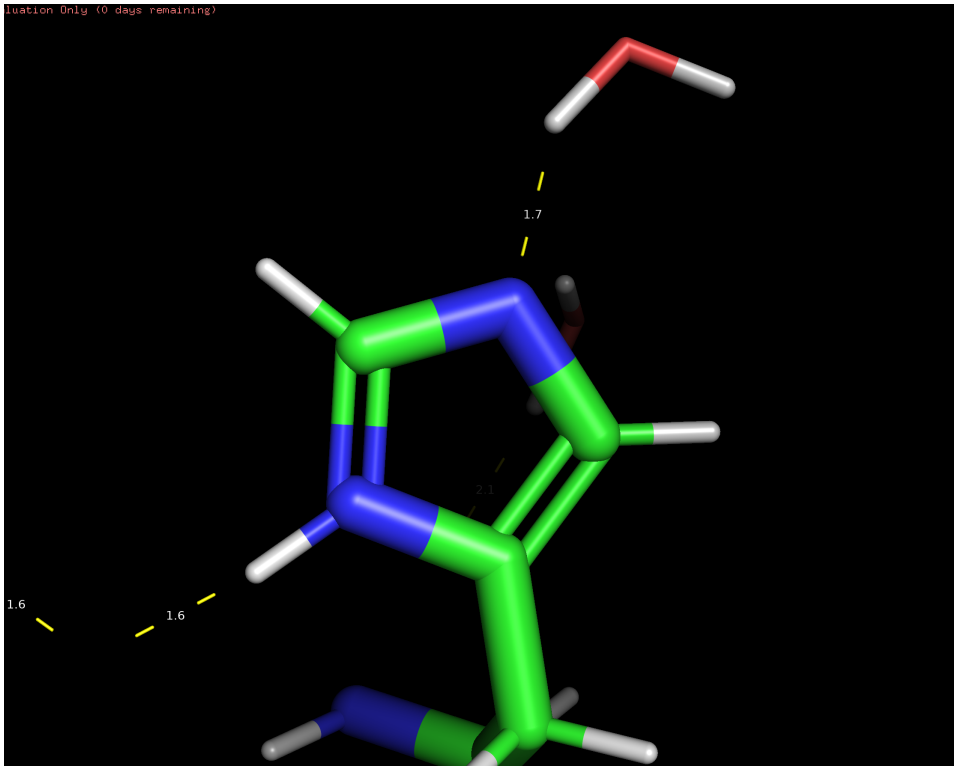
# Metrics

| | Configuratio | Energy (kcal/ | ΔE | 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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix