Restraints in Phenix

Generating & modifying for all scenarios

Nigel W. Moriarty

User's Meeting
LBL, March 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

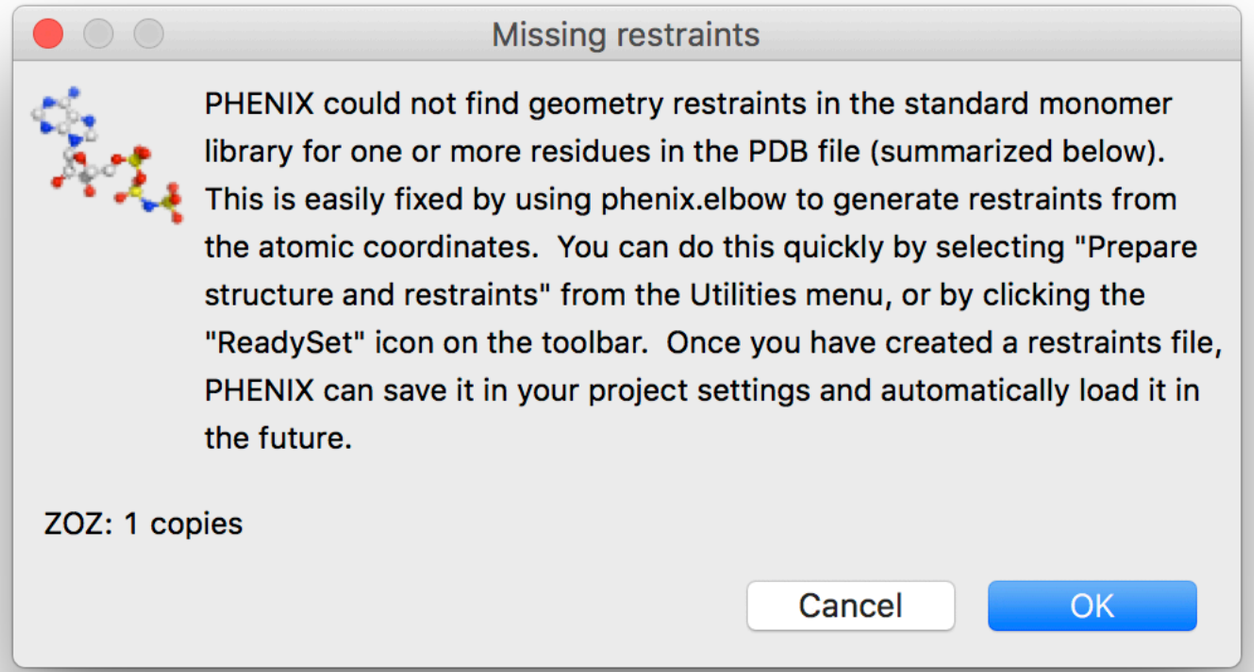
Restraints in Action

- 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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

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

Generate ligand restraints

Fit ligand to density

Refine macromolecule and ligand

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

eLBOW goals

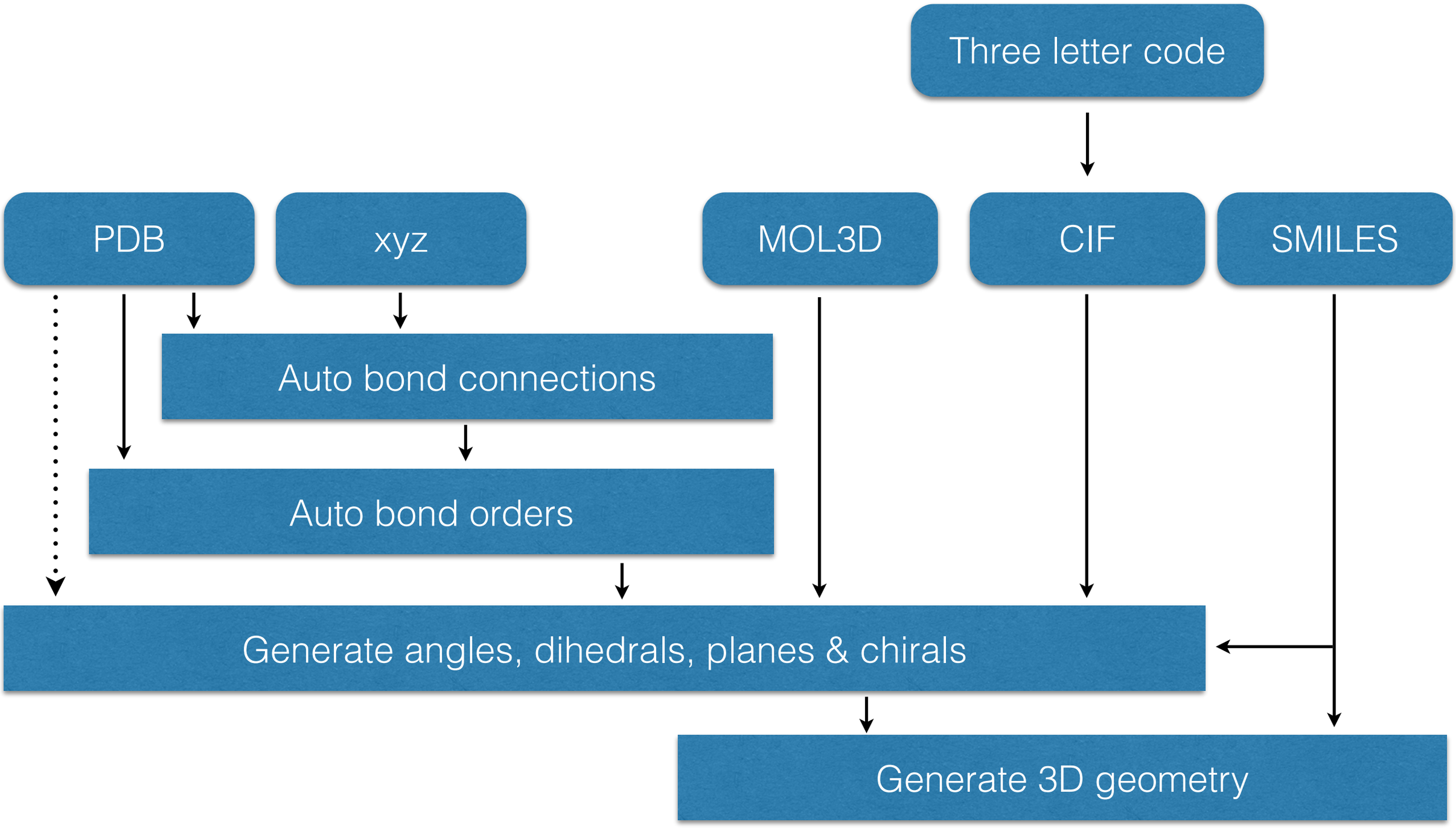
- Fast, simple and flexible procedure to include ligands
- Reduce the tedium of building 3D ligand models
- Automate generation of restraints for ligands
- Comparison of ligand structures

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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

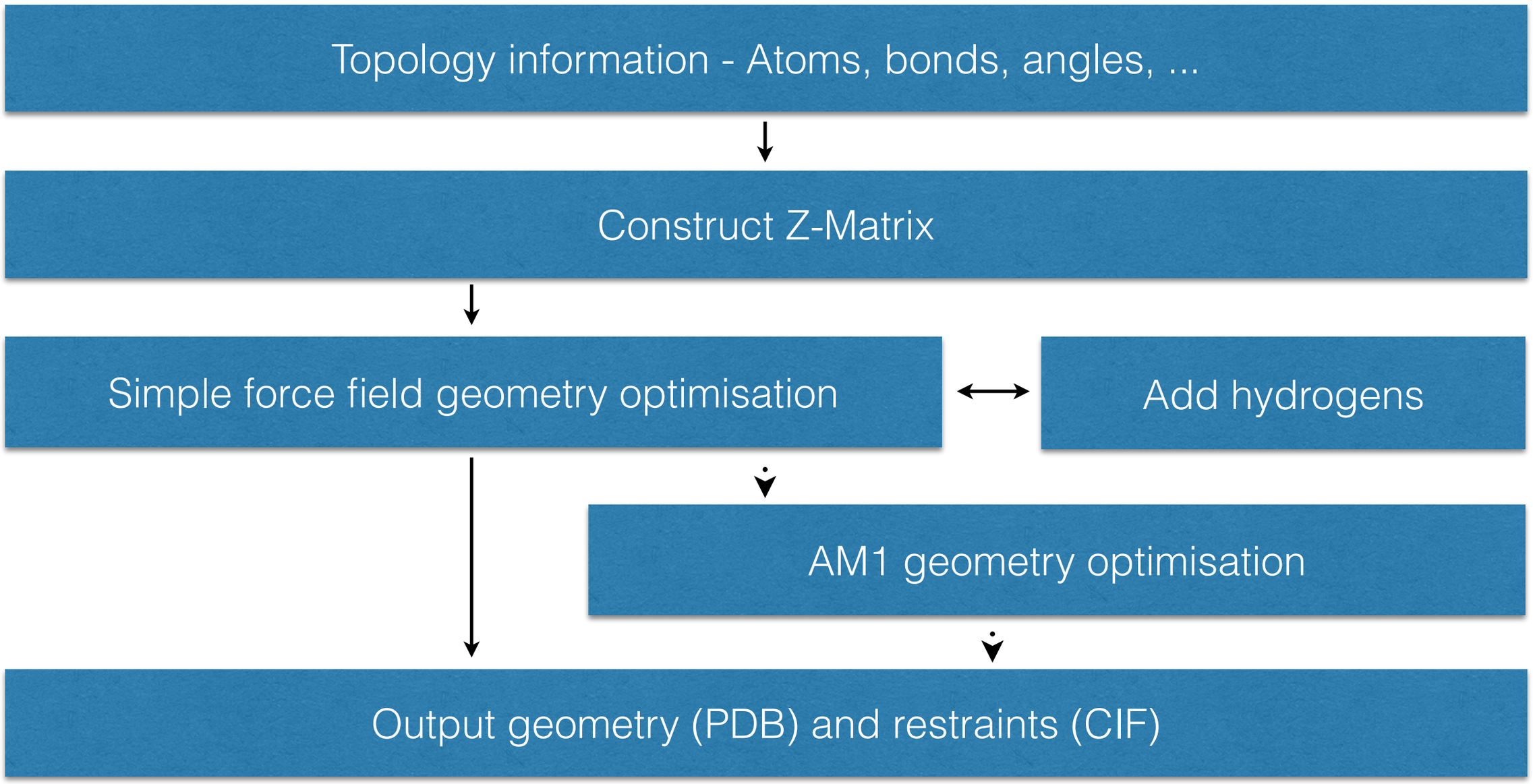
Topology



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Optimisation



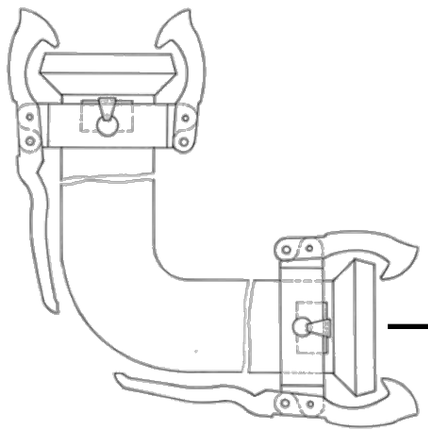
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Getting ready to refine

- Many details needed to prepare for structure refinement

Chemical input



Chemical restraints (CIF)
Cartesian coordinates (PDB)

Protein Information

ReadySet!

Experimental data

refinement

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

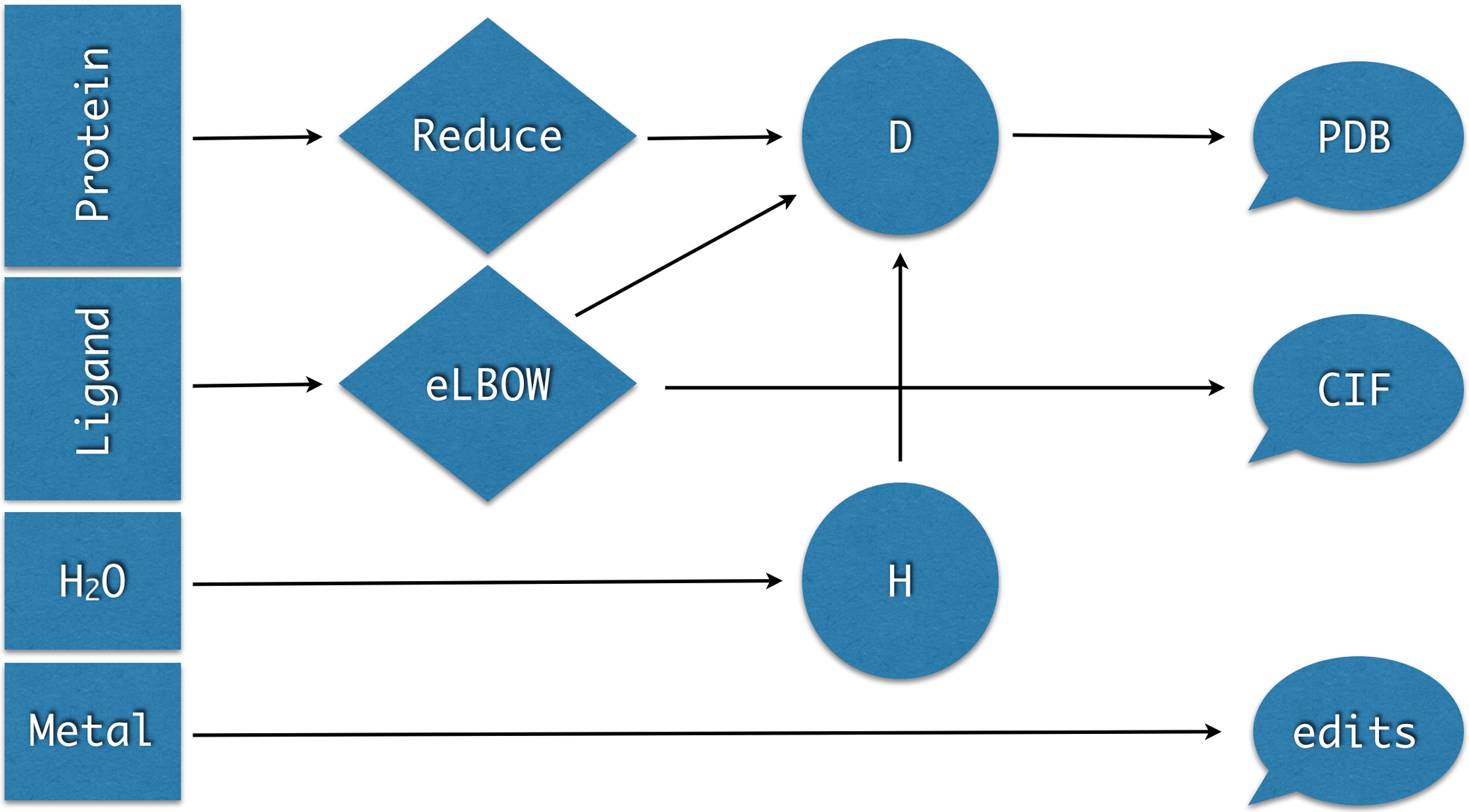
ReadySet!

- 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!



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

Ligand Fitting

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

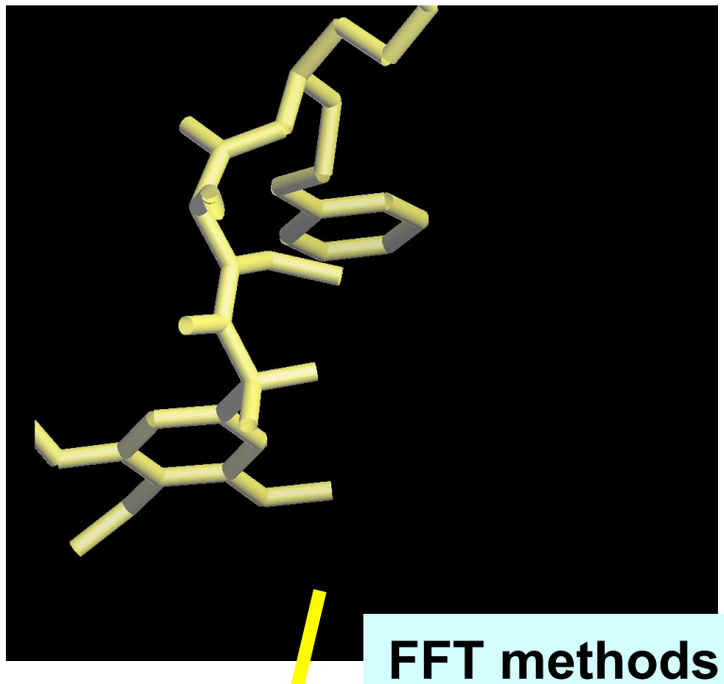
Approach

- 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

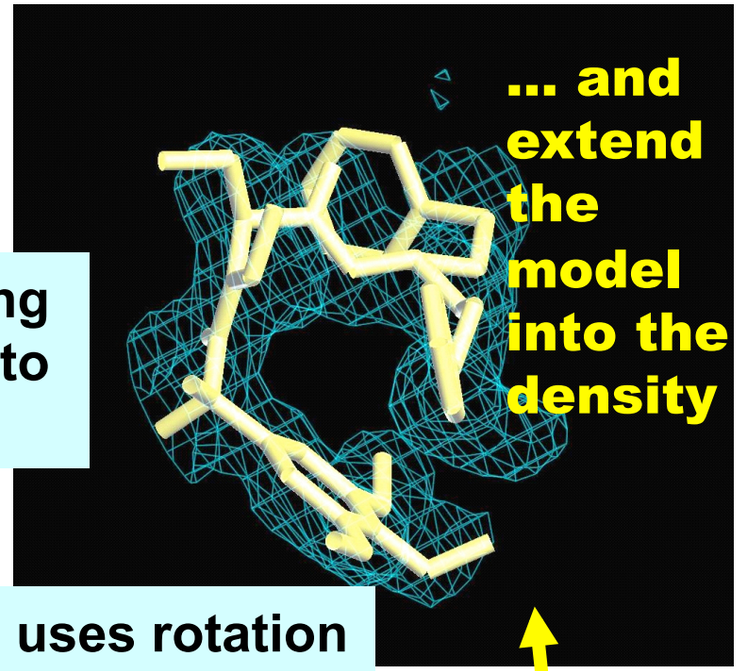
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Automated Ligand Fitting

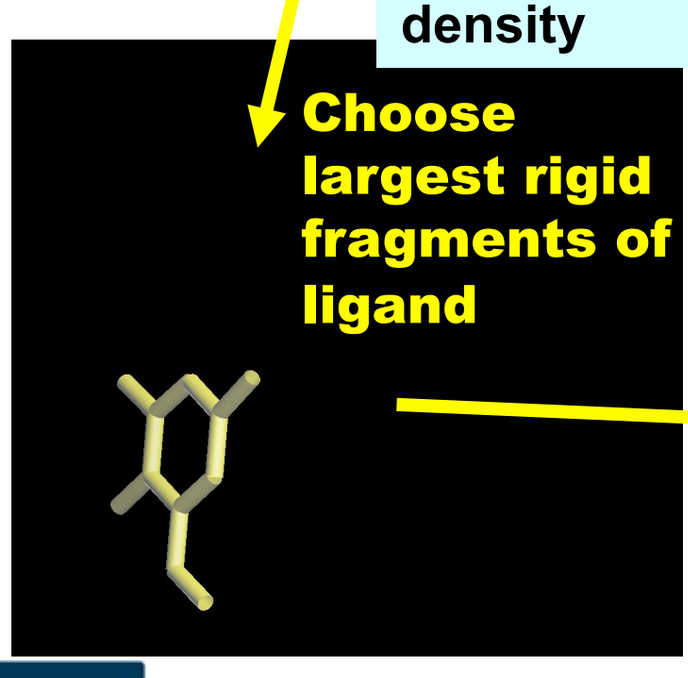


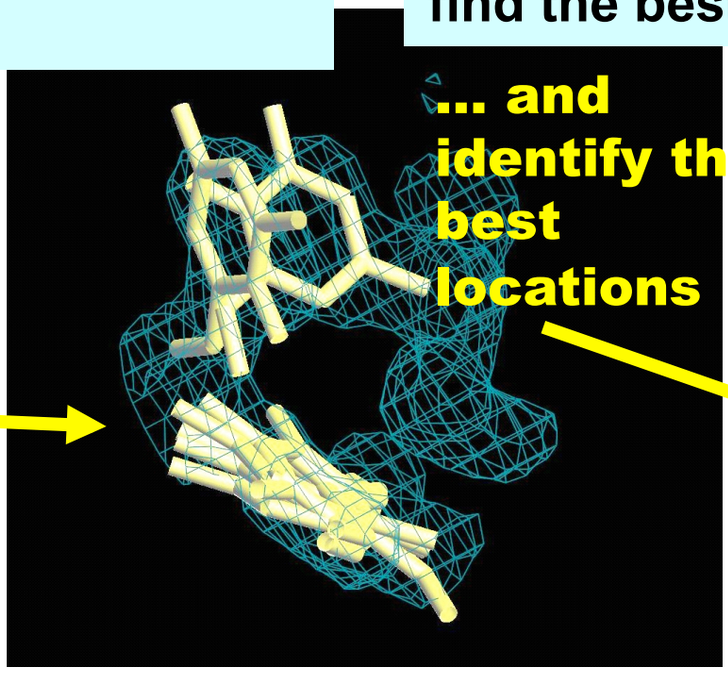
The best fit is measured using the correlation of the model to the electron density

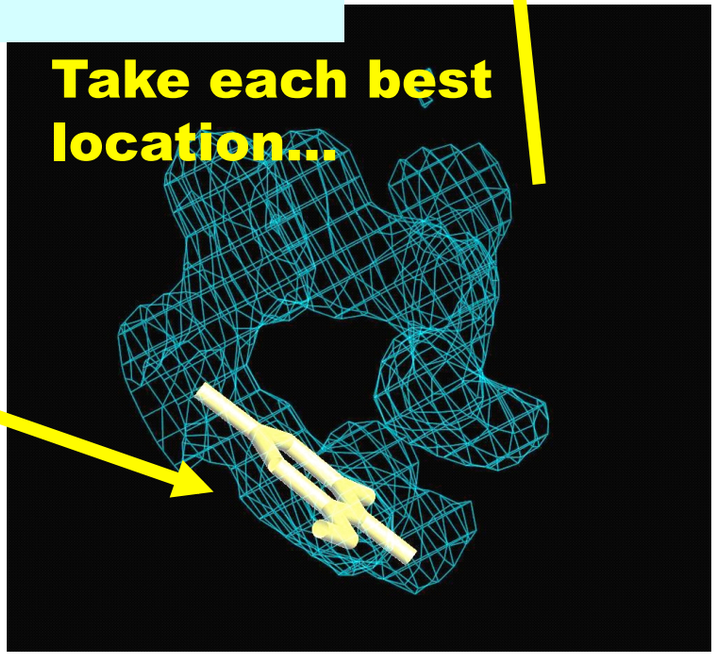


FFT methods are used to rapidly fit fragments to density

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







BERKELEY LAB
Lawrence Berkeley National Laboratory

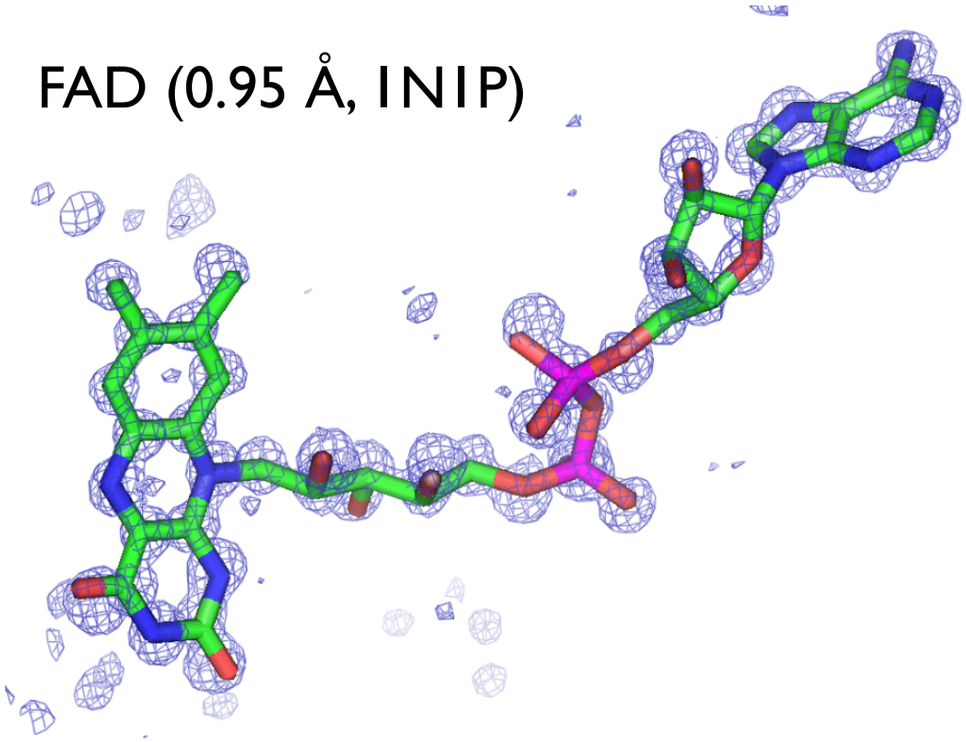
Tom Terwilliger, LANL

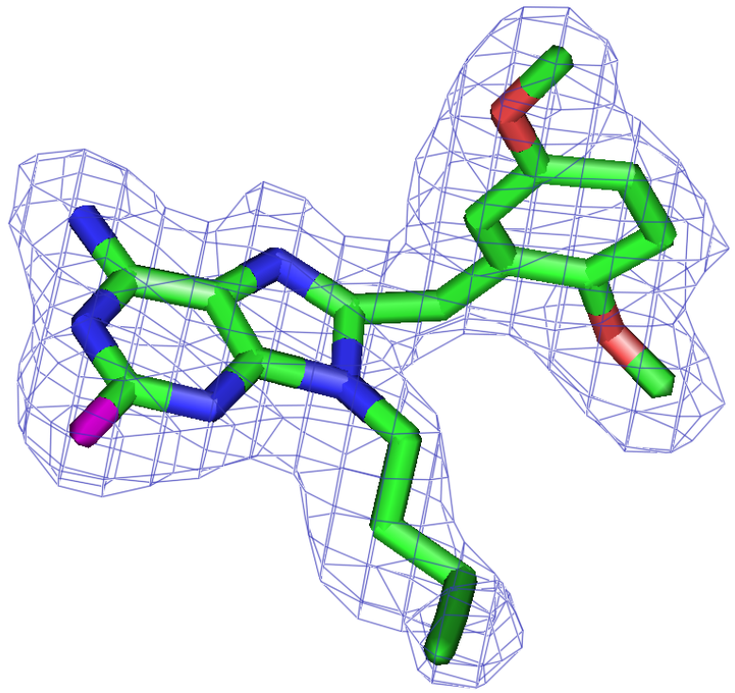
Phenix

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

Fitting Over a Range of Resolutions

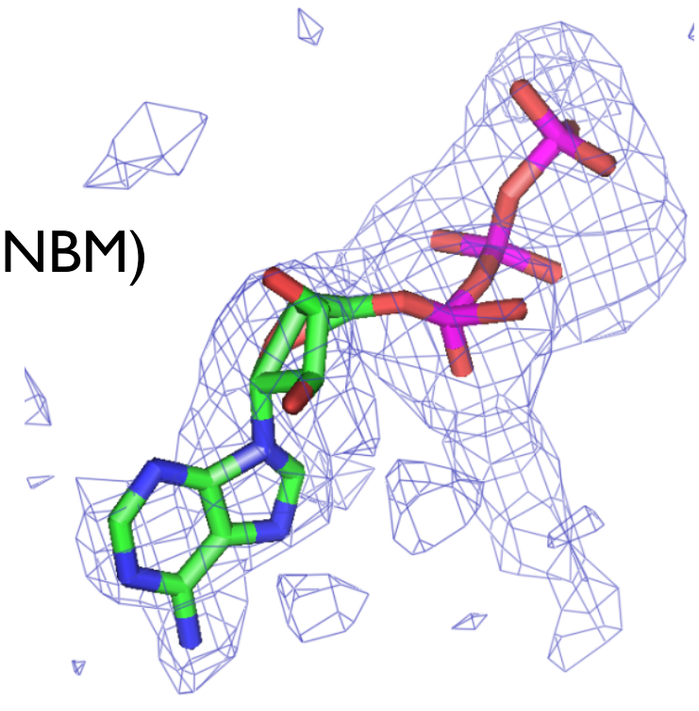
FAD (0.95 Å, 1NIP)

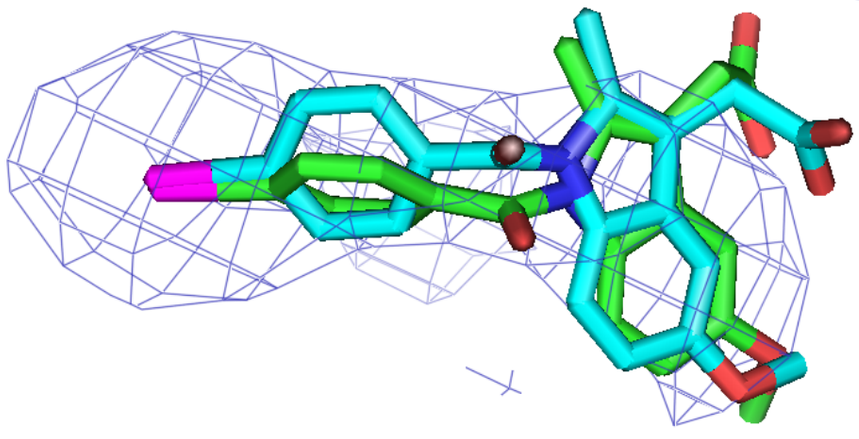




8-(2,5-DIMETHOXY-BENZYL)-2-FLUORO-9-PENT-9H-PURIN-6-YLAMINE (2.2 Å 1UYI)

ATP (3 Å, 1NBM)





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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Results

- Based on 6209 PDB entries with 9319 ligands

	Ligands in PDB fitting Fo-Fc map with CC ≥ 0.75	All Ligands
Number of ligands	6590	9319
Mean CC of ligand in PDB	0.85	0.78
Mean CC of fitted ligand (unrefined)	0.76	0.72
Rmsd ≤ 1.0 Å	2715 (41%)	3004 (32%)
Rmsd ≤ 2.0 Å	4666 (71%)	5421 (58%)
Rmsd > 10 Å	310 (5%)	1417 (15%)

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

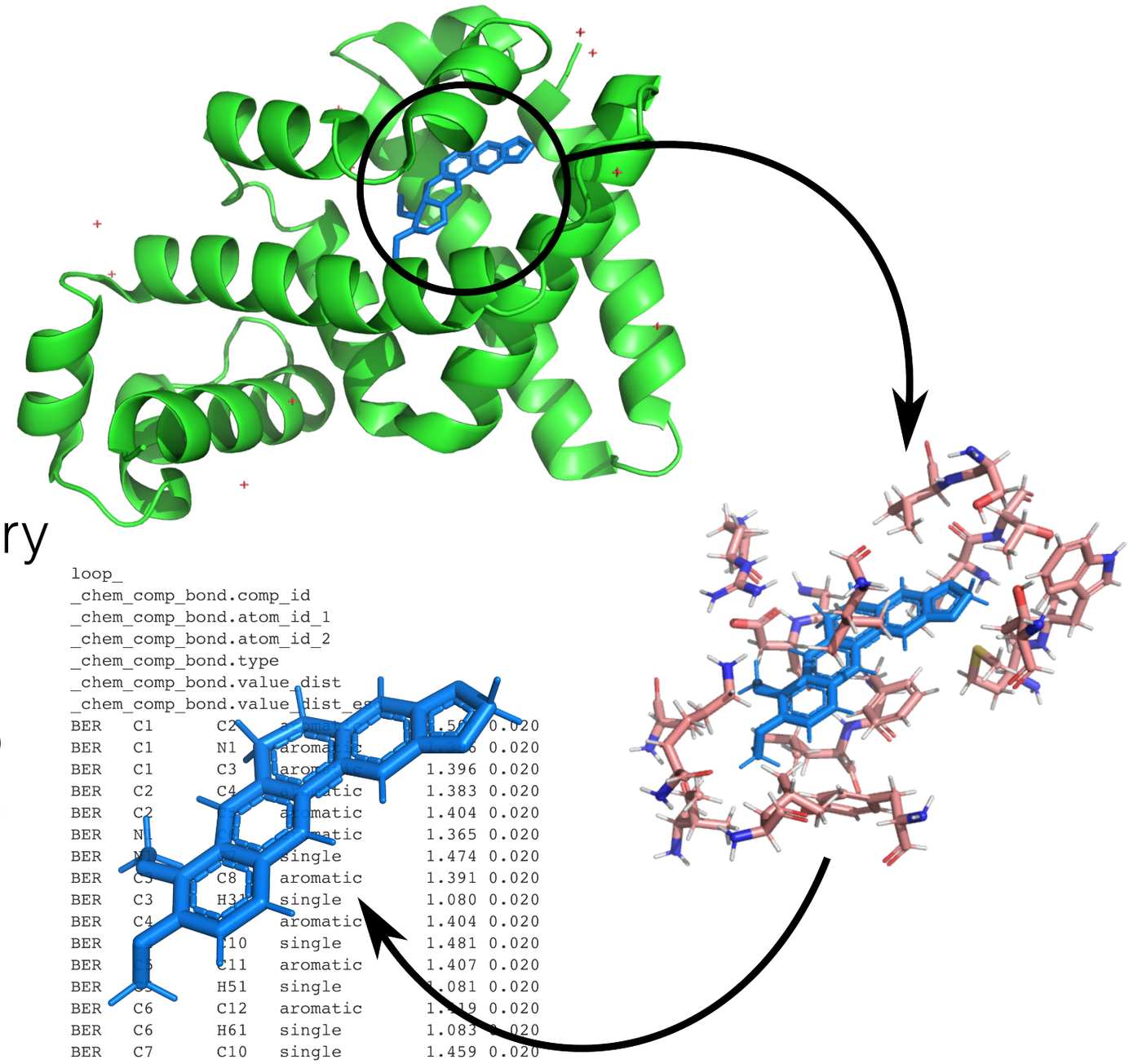
QM Restraints

- 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*
- Python2 needs $PHENIX_MOPAC
- 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



QI

- phenix.fetch_pdb 4W53 --mtz
- phenix.ready_set 4W53.pdb
- mmtbx.quantum_interface 4W53.updated.pdb format=qi write_qmr_phil=True

The result is a PHIL scope for MBN QMR restraints:

- 4W53.updated_A_200_MBN.phil

QMR phil

```
qi.qm_restraints {
  selection = "chain A and resid  200  and resname MBN"
  run_in_macro_cycles = *first_only first_and_last all last_only test
  buffer = 3.5

  calculate = *in_situ_opt starting_energy final_energy starting_strain final_strain starting_bound
final_bound starting_higher_single_point final_higher_single_point

  write_files = *restraints pdb_core pdb_buffer pdb_final_core *pdb_final_buffer

package {
  program =  *mopac test
  charge = Auto
  multiplicity = Auto
  method = Auto
  basis_set = Auto
  solvent_model = None
  nproc = 1
  read_output_to_skip_opt_if_available = True
  ignore_input_differences = False
  view_output = None}}
```

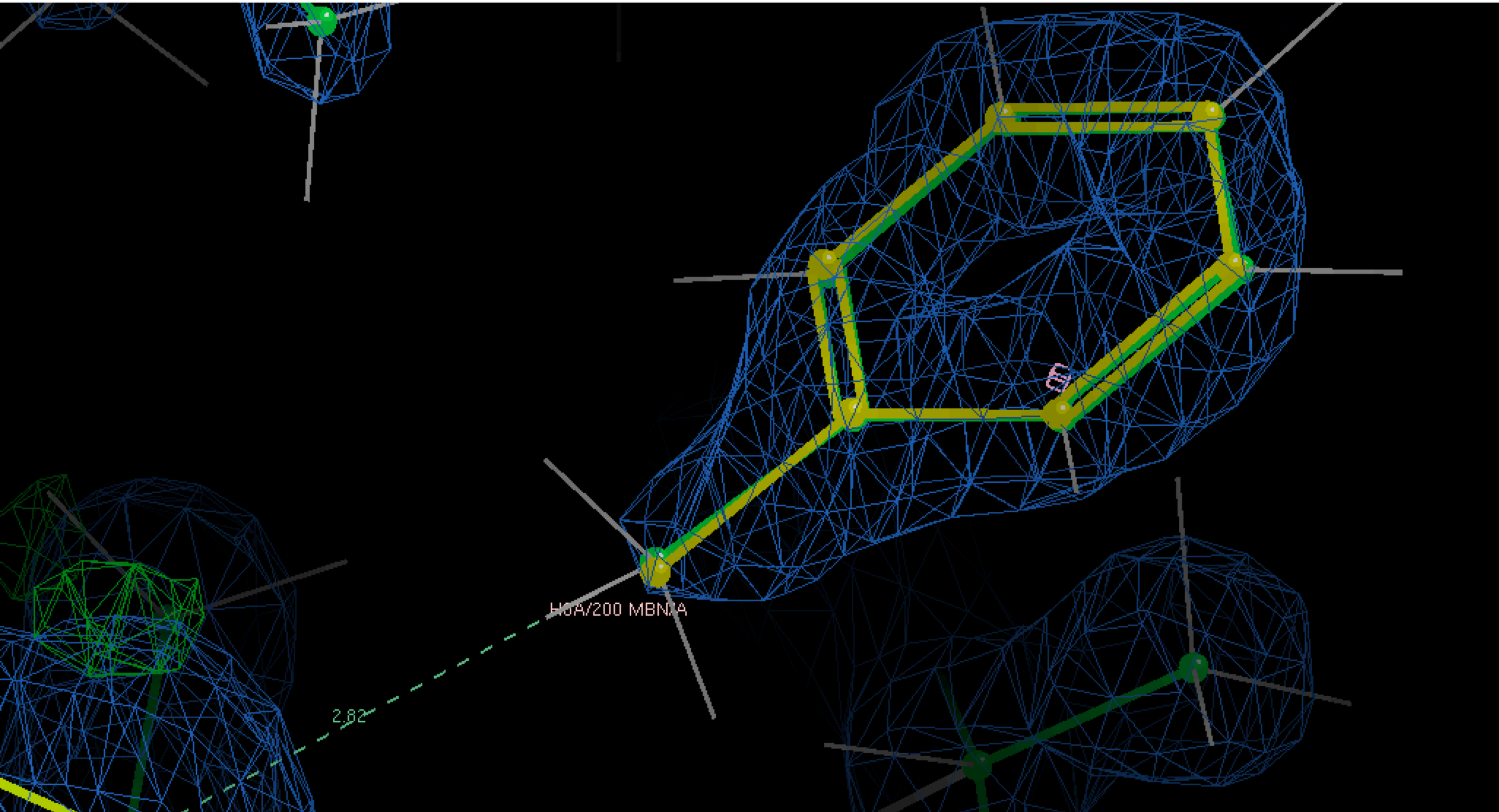
Run

- mmtbx.quantum_interface 4W53.updated.pdb 4W53.updated_A_200_MBN.phil run_qmr=True

```
QM energies

  "chain A and resid  200  and resname MBN"
    Macro cycle 1
      strain                              2.073 kcal/mol (atoms    15)
      opt                           -405946.999 kcal/mol (atoms   246)
```

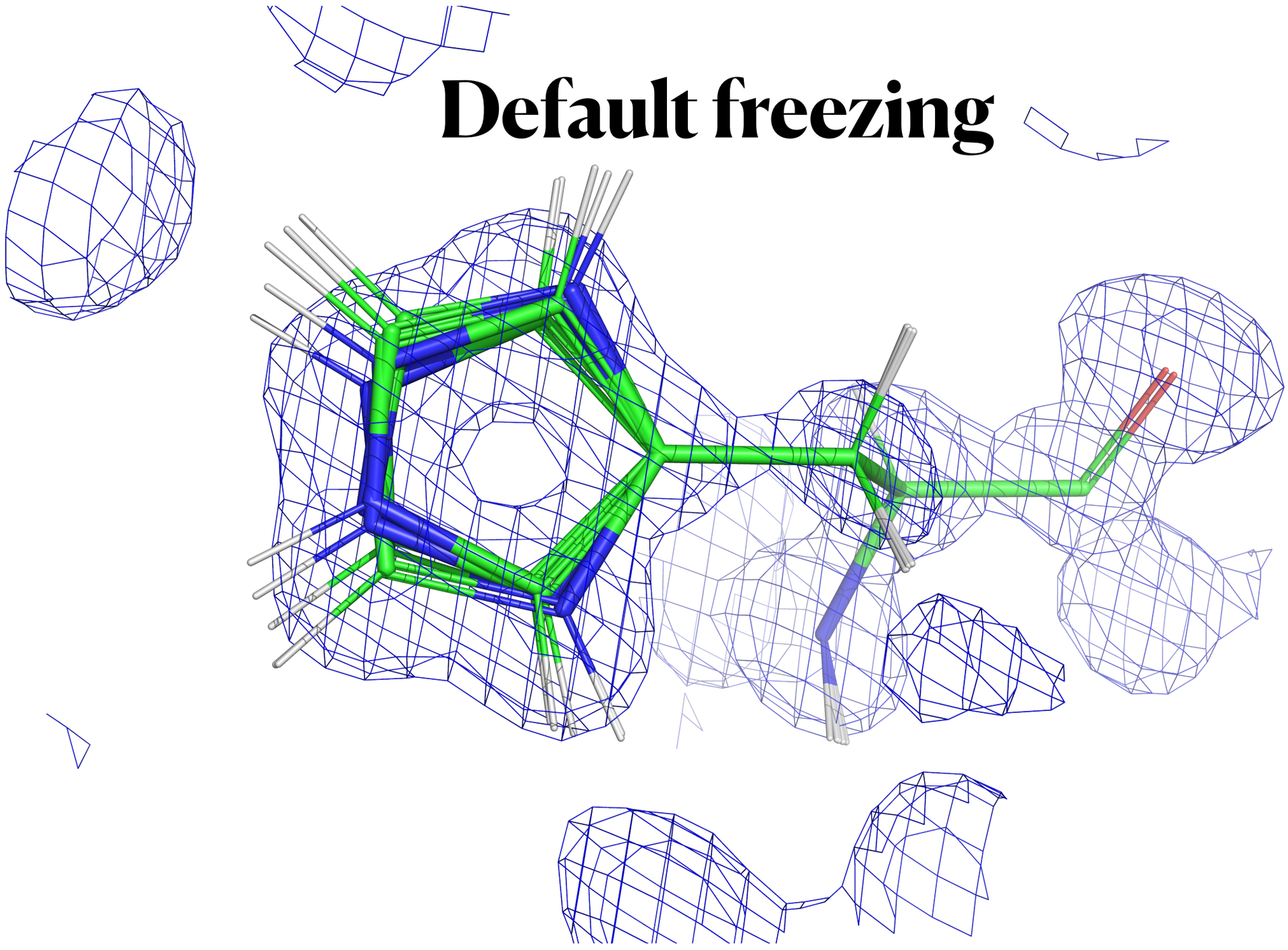
In situ



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

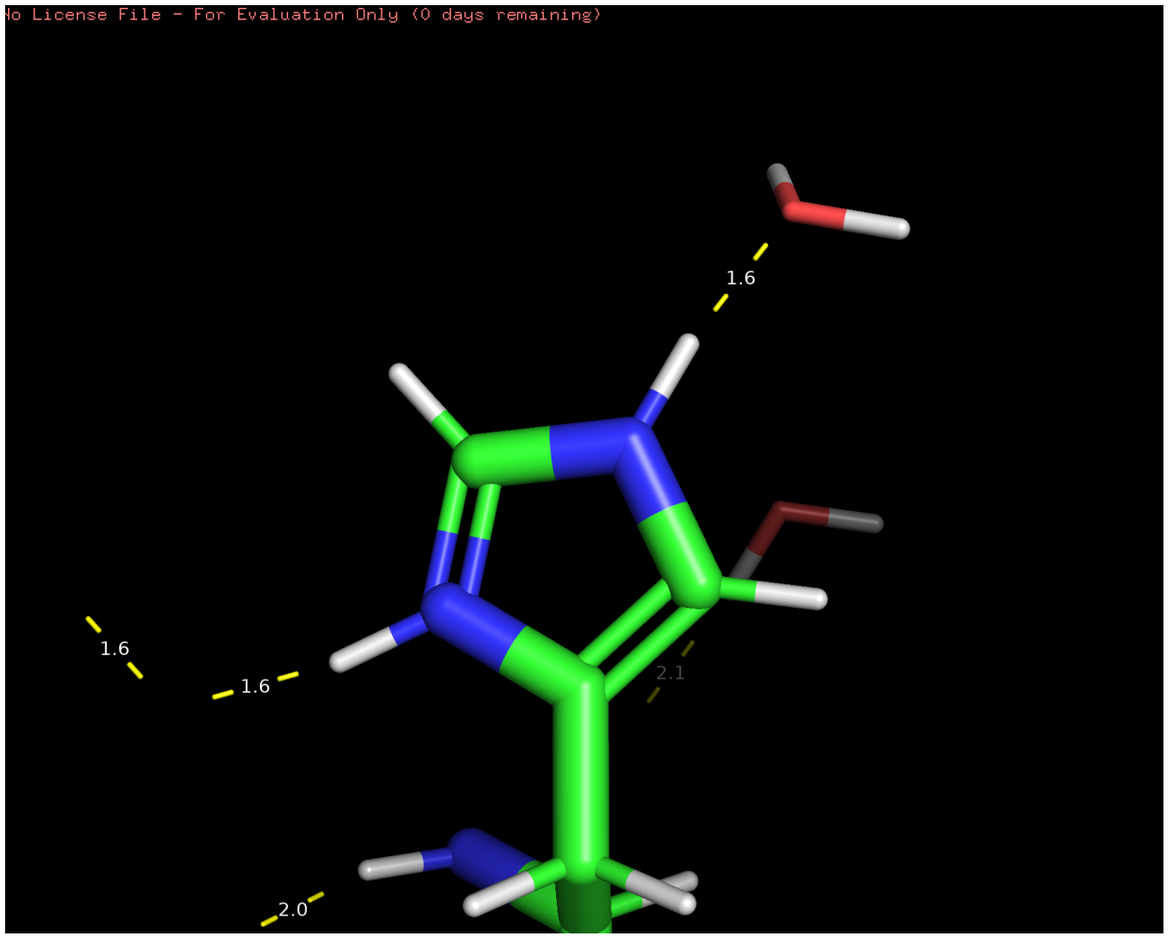


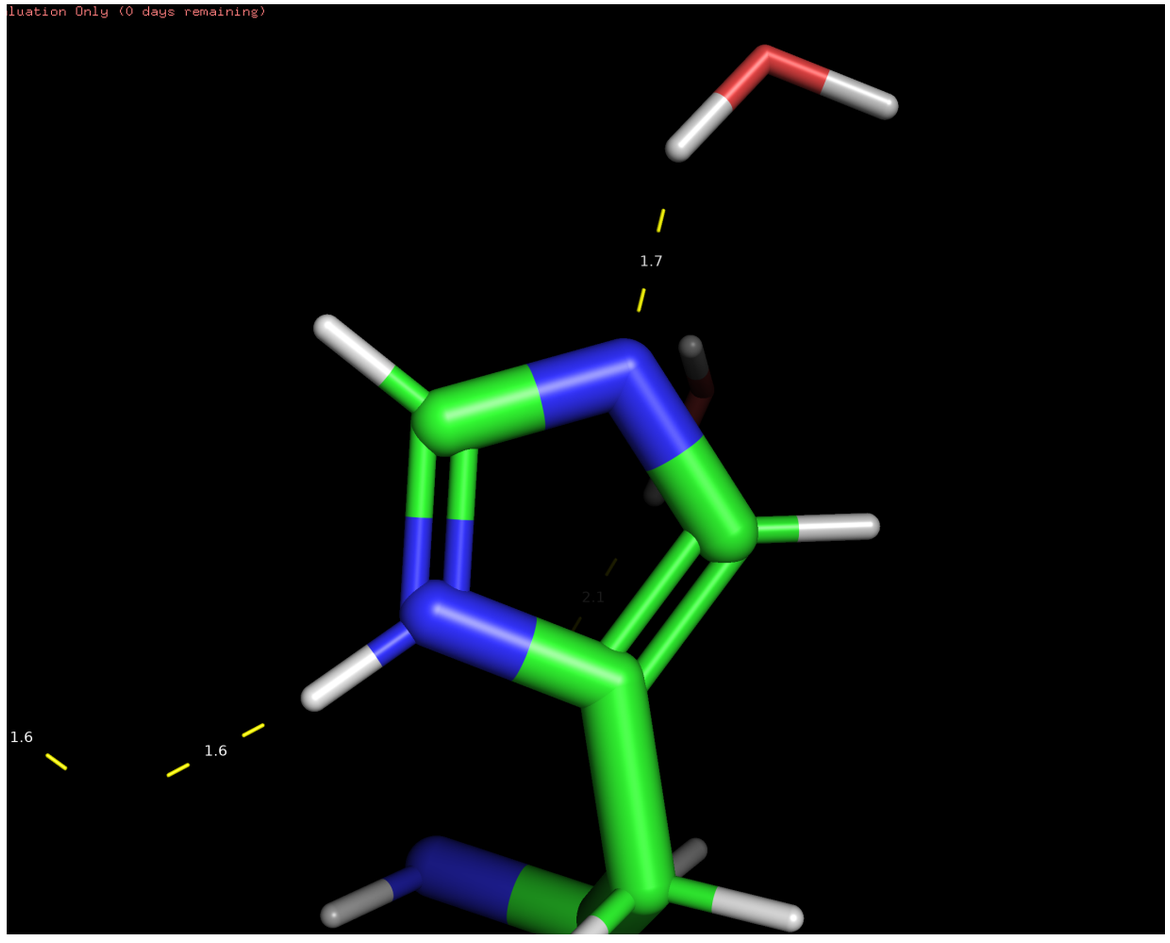
Metrics

	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