Restraints in Phenix

Generating & modifying for all scenarios

Nigel W. Moriarty

User’s Meeting
University of Montana, June 2024
Ligands in Phenix
Generating & modifying for all scenarios
Nigel W. Moriarty

User’s Meeting
University of Montana, June 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

PHENIX could not find geometry restraints in the standard monomer library for one or more residues in the PDB file (summarized below). This is easily fixed by using phenix.elbow to generate restraints from the atomic coordinates. You can do this quickly by selecting "Prepare structure and restraints" from the Utilities menu, or by clicking the "ReadySet" icon on the toolbar. Once you have created a restraints file, PHENIX can save it in your project settings and automatically load it in the future.

ZOZ: 1 copies
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

Generate ligand restraints

Fit ligand to density

Refine macromolecule and ligand
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

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

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 heterogenous 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

**Chemical Description**
- **Name**: WATER
- **Formula**: H2 O
- **Formal charge**: 0
- **Molecular weight**: 18.015 g/mol
- **Component type**: NON-POLYMER
Metal clusters
5-letter codes

- 49k combinations for 3-letter codes
Human readable

<table>
<thead>
<tr>
<th>Standard</th>
<th>Human readable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Uppercase</td>
</tr>
<tr>
<td>10I0</td>
<td>10i0</td>
</tr>
<tr>
<td>1IJJ</td>
<td>1ijj</td>
</tr>
<tr>
<td>40CL</td>
<td>4ocl</td>
</tr>
<tr>
<td>5SS2</td>
<td>5ss2</td>
</tr>
</tbody>
</table>

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

PDB → Auto bond connections → Auto bond orders → Generate angles, dihedrals, planes & chirals → Generate 3D geometry

xyz → Auto bond connections

MOL3D

CCL code

CCL

SMILES

Generate 3D geometry
Optimisation

- Topology information - Atoms, bonds, angles, ...
- Simple force field geometry optimisation
- Advanced geometry optimisation
- Output geometry (PDB) and restraints (CIF)
- Add hydrogens
Getting ready to refine

- Many details needed to prepare for structure refinement

- Chemical input
  - Chemical restraints (CIF)
    - Cartesian coordinates (PDB)
  - Experimental data
    - Protein Information
      - ReadySet!
      - 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

Chemical input → REEL → Protein Information

Chemical restraints (CIF)
Cartesian coordinates (PDB)

Reflection data → ReadySet!

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

**ATP**

**View preferences loaded**
<table>
<thead>
<tr>
<th>Atom ID</th>
<th>Atom ID</th>
<th>Restrained To</th>
<th>E/Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>C03</td>
<td>C02</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>Code</td>
<td>Name</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>HOH WATER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1CU COPPER ION, 1 WATER COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20F FERROUS ION, 2 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3OF HYDRATED FE (III) ION, 2 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>543 CALCIUM ION, 6 WATERS PLUS ETHANOL COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>CD1 CADMIUM ION, 1 WATER COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>CD3 CADMIUM ION, 3 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CD5 CADMIUM ION, 5 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>CO5 COBALT ION, 5 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>DOD DEUTERATED WATER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>HC0 2 IRON/2 SULFUR/6 CARBONYL/1 WATER INORGANIC CLUST...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>HC1 2 IRON/2 SULFUR/5 CARBONYL/2 WATER INORGANIC CLUST...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>HCN 2 IRON/2 SULFUR/3 CARBONYL/2 CYANIDE/WATER/METHYLER...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>KO4 POTASSIUM ION, 4 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>MN5 MANGANESE ION, 5 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>MN6 MANGANESE ION, 6 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>MO1 MAGNESIUM ION, 1 WATER COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>MO2 MAGNESIUM ION, 2 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>MO3 MAGNESIUM ION, 3 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>MO4 MAGNESIUM ION, 4 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>MO5 MAGNESIUM ION, 5 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>MO6 MAGNESIUM ION, 6 WATERS COORDINATED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>MTO BOUND WATER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>MW1 MANGANESE ION, 1 WATER COORDINATED</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

Choose largest rigid fragments of ligand

FFT methods are used to rapidly fit fragments to density

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

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

... and identify their best locations

Take each best location...

... and extend the model into the density

Tom Terwilliger, LANL
Fitting Over a Range of Resolutions

FAD (0.95 Å, INIP)

ATP (3 Å, INBM)

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

(1-(4-IODOBENZOYL)-5-METHOXY-2-METHYL-INDOLE-3-ACETIC ACID (4.5 Å, IPGF)
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
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 \textit{in situ}

• There are two ways of using QMR
  • In \texttt{phenix.refine}
  • In a standalone program \texttt{mmtbx.quantum_interface}

• Python3 installers from the bottom of the download page

• For ORCA, set $\texttt{PHENIX_ORCA}$
**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

BER  C1    C2  aromatic  1.378  0.020
BER  C1    N1  aromatic  1.404  0.020
BER  C1    C3  aromatic  1.396  0.020
BER  C2    C4  aromatic  1.383  0.020
BER  C2    C1  aromatic  1.404  0.020
BER  N1    C2  aromatic  1.365  0.020
BER  C2    C3  single   1.474  0.020
BER  C2    C8  aromatic  1.391  0.020
BER  C3    N8  single   1.080  0.020
BER  C4    O0  single   1.404  0.020
BER  C4    C0  single   1.481  0.020
BER  C7    C1  aromatic  1.407  0.020
BER  C7    N51 single   1.081  0.020
BER  C6    C12 aromatic  1.359  0.020
BER  C6    N61 single   1.009  0.020
BER  C7    C10 single   1.459  0.020
```
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
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

<table>
<thead>
<tr>
<th>Configuratio</th>
<th>Energy (kcal/mol)</th>
<th>ΔE</th>
<th>H bonds</th>
<th>r.m.s.d.</th>
<th>Rotamer</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>HD1, HE2</td>
<td></td>
<td></td>
<td></td>
<td>m90</td>
</tr>
<tr>
<td>1</td>
<td>HD1, HE2</td>
<td>-1019.6</td>
<td>2.6</td>
<td>14</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>HD1 only</td>
<td>-1022.2</td>
<td>0.0</td>
<td>14</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>HE2 only</td>
<td>-1003.0</td>
<td>19.3</td>
<td>13</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>HD1, HE2</td>
<td>-1004.0</td>
<td>18.3</td>
<td>12</td>
<td>0.29</td>
</tr>
<tr>
<td>5</td>
<td>HD1 only</td>
<td>-1004.4</td>
<td>17.9</td>
<td>12</td>
<td>0.38</td>
</tr>
<tr>
<td>6</td>
<td>HE2 only</td>
<td>-1009.4</td>
<td>12.9</td>
<td>11</td>
<td>0.32</td>
</tr>
</tbody>
</table>

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.
QM Energies

- Strain energy
  - Relaxation of ligand to local minimum
- Bound energy
  - Includes binding energy
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