Improving low-resolution refinement of nucleic acids in *Phenix*

Oleg V. Sobolev\textsuperscript{a}, Pavel V. Afonine\textsuperscript{a}, Alexandre Urzhumtsev\textsuperscript{b,c}, Nigel W. Moriarty\textsuperscript{a} and Paul D. Adams\textsuperscript{a,d}

\textsuperscript{a}Lawrence Berkeley National Laboratory, Berkeley, CA  
\textsuperscript{b}Department of Integrated Structural Biology, IGBMC, Illkirch, France  
\textsuperscript{c}Physical Department, Université de Lorraine, France  
\textsuperscript{d}Department of Bioengineering, University of California at Berkeley, Berkeley, CA

WCPCW  
Monterey, 2015
Restraints and data resolution

• **Refinement target** - a weighted sum of experimental data ($E_{\text{data}}$) and *a priori* chemical knowledge terms (restraints; $E_{\text{restraints}}$): $E_{\text{total}} = w \times E_{\text{data}} + E_{\text{restraints}}$

• **Choice of restraints** depends on data quality (resolution):

  - **<1Å**: unrestrained refinement
  - **>3Å**: more restraints needed
    
    \[
    E_{\text{restraints}} = \ldots + E_{\text{Ramachandran}} + E_{\text{NCS}} + E_{\text{ReferenceModel}} + E_{\text{SecondaryStructure}} + \ldots
    \]

  - **1-3Å**: *standard* restraints are necessary
    
    \[
    E_{\text{restraints}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{nonbonded}} + E_{\text{planarity}} + E_{\text{chirality}}
    \]
Insufficiency of standard restraints at low resolution

- Example: real-space refinement with simulated annealing of 3gbi against 4Å 2mFo-DFc map with `phenix.real_space_refine`
- Refinement with standard restraints fits model into map well, but geometry is poor (no correct basepairing and stacking interactions)
Geometry restraints for DNA/RNA in Phenix

1. **Hydrogen bonds** between base pairs:
   - Bond length restraints
   - Bond angles restraints

2. **Planarity** of base-pairs:
   - Planarity restraint
   - Parallelity restraints

3. **Parallelity** of stacking nucleobases:
   - Parallelity restraints
Hydrogen bond restraints

The values for hydrogen bond lengths differ for different basepairing type and participating atoms:

- 122.8 (3.00)
- 119.1 (2.59)
- 122.2 (2.88)

The values for hydrogen bond lengths of the same type are essentially the same for DNA and RNA:

- 2.78 (0.10)
- 2.88 (0.07)
- 2.93 (0.10)

- 117.3 (2.86)
- 116.3 (2.66)
- 120.7 (2.20)
Stacking and basepairing restraints

Basepairing RMSDs from 0:
• parallelity: 14.87°
• planarity: 0.188Å

Stacking parallelity RMSD from 0°: 11.54°
DNA/RNA: example of low-resolution refinement

- Real-space refinements with simulated annealing against 4Å 2mFo-DFc map with `phenix.real_space_refine`

**Starting model**

- Nucleobases somewhat parallel with a number of outliers

**Standard restraints**

- Geometry becomes worse: nucleobases are not parallel, poor H-bonding

**Standard + stacking + basepair + H\textsubscript{bond} restraints**

- Nucleobases are parallel, correct H-bonding
Implementation in Phenix

• Available in all relevant tools (phenix.refine, phenix.real_space_refine, phenix.geometry_minimization, phenix.dynamics) via secondary_structure.nucleic_acid scope

• Turn on restraints:
  secondary_structure.enabled=True

• Generate phil file with NA definitions:
  phenix.secondary_structure_restraints <model.pdb>