Quick Facts About PHENIX, CCI APPS and Structure Refinement with phenix.refine

01 – February – 2008

The PHENIX project is a collaboration between several groups to develop highly integrated tools for automated crystallography:

- Los Alamos National Lab
  Tom Terwilliger, Li-Wei Hung (SOLVE / RESOLVE)
  Paul Langan, Marat Mustyakimov, Benno Schoenborn (Tools for Neutron crystallography) (separate funding, MNC)

- Cambridge University, UK
  Randy Read, Airlie McCoy (PHASER)

- Duke University
  Jane & David Richardson, Ian Davis (MolProbity)

- Lawrence Berkeley National Lab
  Paul Adams, Pavel Afonine, Ralf Grosse-Kunstleve, Nigel Moriarty, Nicholas Sauter, Peter Zwart (CCI APPS)

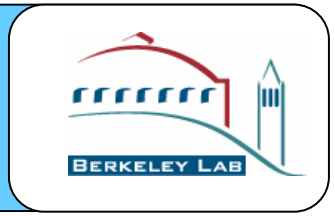
- Texas A&M University
  Tom Ioerger, Jim Sacchettini, Erik McKee (TEXTAL)

Tools for automated structure solution:

- ***Automated analysis of data quality and integrity***
  - phenix.xtriage (detect twinning and more)
- ***Rapid substructure determination*** – Hybrid Substructure Search (HySS)
- ***Scoring of substructure enantiomorphs*** – Analysis of maps calculated with each hand
- ***Phasing*** – SOLVE, PHASER
- ***Density modification*** – Statistical density modification (RESOLVE)
- ***Automated model building***
  - Pattern matching methods (RESOLVE or TEXTAL)
  - Building and refinement of multiple models
- ***Structure refinement***
  - phenix.refine (complete and automated structure refinement at any resolution)
- ***Ligand building and fitting*** – eLBOW (cif files), AutoLigand
- ***Comprehensive PDB file manipulations, models superposition***
  - phenix.pdbtools, phenix.superpose_pdbs
- ***Structure validation***
  - MolProbity, Reduce (conformations and contacts, hydrogen atoms)

CCI APPS: Subset of PHENIX components developed by the Computational Crystallography Initiative (CCI) at LBNL:

Computational Crystallography Initiative (LBNL)
-***Paul Adams*, Ralf Grosse-Kunstleve, Pavel Afonine**
-**Nigel Moriarty, Nicholas Sauter, Peter Zwart**



CCI APPS: Command-line tools only. “One click” installation on most of platforms (list of supported platforms: http://phenix-online.org/download/cci_apps/). Does not require installation of anything else. Can be obtained and installed separately from PHENIX package.

CCI APPS include:

phenix.refine – Highly-automated state-of-the-art structure refinement

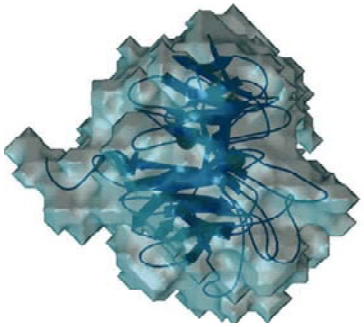
HySS – Highly-automated location of anomalous scatterers

eLBOW – Parameter and topology file builder for unknown ligands

Xtriage – Comprehensive analysis of experimental data

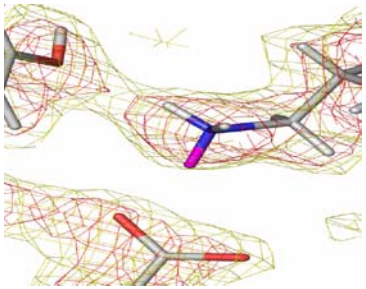
Below we mostly focus on phenix.refine. For information about other components please refer to separate documentation: http://phenix-online.org/

Low



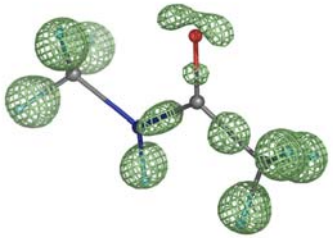
- Group ADP refinement
- Rigid body refinement
- Torsion Angle dynamics

Middle and High



- Restrained refinement (xyz, ADP: isotropic, anisotropic, mixed)
- Automatic water picking

Subatomic



- Bond density model
- Unrestrained refinement
- FFT or direct
- Explicit hydrogens

- Automatic NCS restraints
- Simulated Annealing
- Occupancies (individual, group, constrains for alternative conformation)
- TLS refinement
- Use hydrogens at any resolution
- Refinement with twinned data
- X-ray, Neutron, joint X-ray + Neutron refinement

Refinement flowchart

PDB model,
Any data format
(CNS, Shelx, MTZ, …)

Input data and model processing

Refinement strategy selection

Bulk-solvent, Anisotropic scaling, Twinning parameters refinement

Ordered solvent (add / remove)

Target weights calculation

Coordinate refinement
(rigid body, individual) (minimization or Simulated Annealing)

ADP refinement
(TLS, group, individual iso / aniso)

Occupancy refinement (individual, group)

Repeated several times

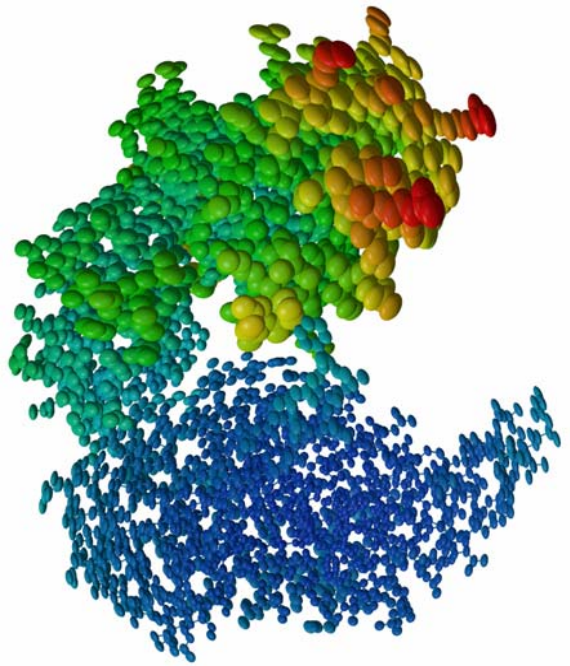
Output: Refined model, various maps, structure factors, complete statistics, ready for deposition PDB file

Files for COOT, O, PyMol

Refine any part of a model with any strategy: *all in one run*

Rigid body +
Overall ADP

Individual
coordinates +
TLS



Individual coordinates +
TLS + restrained isotropic
ADP

+ Automatic water picking

+ Simulated Annealing

+ Add and use hydrogens

Designed to be very easy to use:

Refinement of individual coordinates and B-factors:

```
% phenix.refine model.pdb data.hkl
```

Same as above plus Simulated Annealing (SA) and water picking:

```
% phenix.refine model.pdb data.hkl simulated_annealing=true ordered_solvent=true
```

Rigid-body + SA + grouped B-factors refinement:

```
% phenix.refine model.pdb data.hkl strategy=rigid_body+group_adp simulated_annealing=true
```

TLS + SA + individual coordinates and B-factors refinement:

```
% phenix.refine model.pdb data.hkl strategy=tls+individual_sites+individual_adp \
simulated_annealing=true
```

Refinement of individual coordinates and B-factors using neutron data:

```
% phenix.refine model.pdb data.hkl scattering_dictionary=neutron
```

To see all parameters (more than 200):

```
% phenix.refine model.pdb data.hkl  --show_defaults=all
```

See the documentation for more examples: http://www.phenix-online.org/

Running phenix.refine with a parameter file:

% phenix.refine model.pdb data.hkl parameters_file

where the parameter_file contains following lines:

```
refinement.main {
   high_resolution = 2.0
   low_resolution = 15.0
   simulated_annealing = True
   ordered_solvent = True
   number_of_macro_cycles = 5
}
refinement.refine.adp {
    tls = chain A
    tls = chain B
}
```

This is equivalent to typing from the command line:

% phenix.refine model.pdb data.hkl xray_data.high_resolution=2 xray_data.low_resolution=15 simmulated_annealing=true ordered_solvent=True adp.tls="chain A" adp.tls="chain B" main.number_of_macro_cycles=5

Shortcuts are allowed, as long as they are unambiguous:

% phenix.refine model.pdb data.hkl xray_data.high_res=2 xray_data.low_res=15 simulated_annealing=true ordered_s=True adp.tls="chain A" adp.tls="chain B" main.number_of_m=5

When running:

% phenix.refine model.pdb data.hkl

each item in model.pdb is matched against the CCP4 Monomer Library to extract the topology and parameters and to automatically build corresponding restraints.

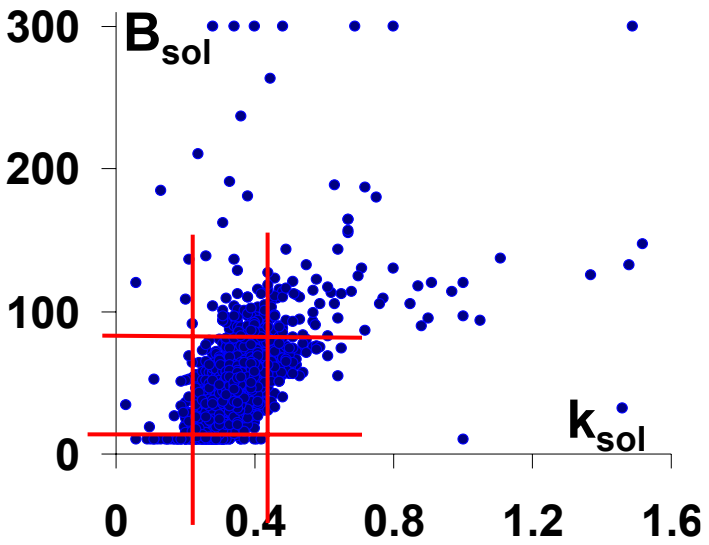
If model.pdb contains an item not available in CCP4 Monomer Library, e.g. a novel ligand, use eLBOW to generate the required topology and parameter definitions for refinement:

% phenix.elbow model.pdb --residue=LIG

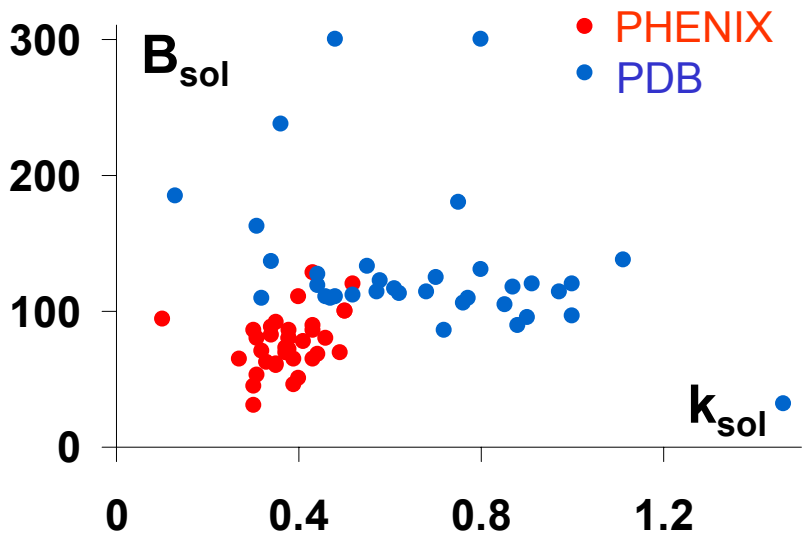
This will produce the file LIG.cif which can be used for refinement:

% phenix.refine model.pdb data.hkl LIG.cif

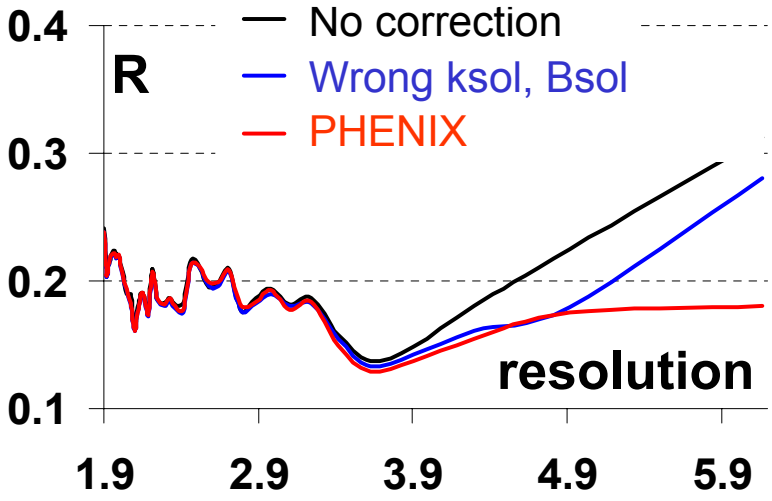
PDB survey



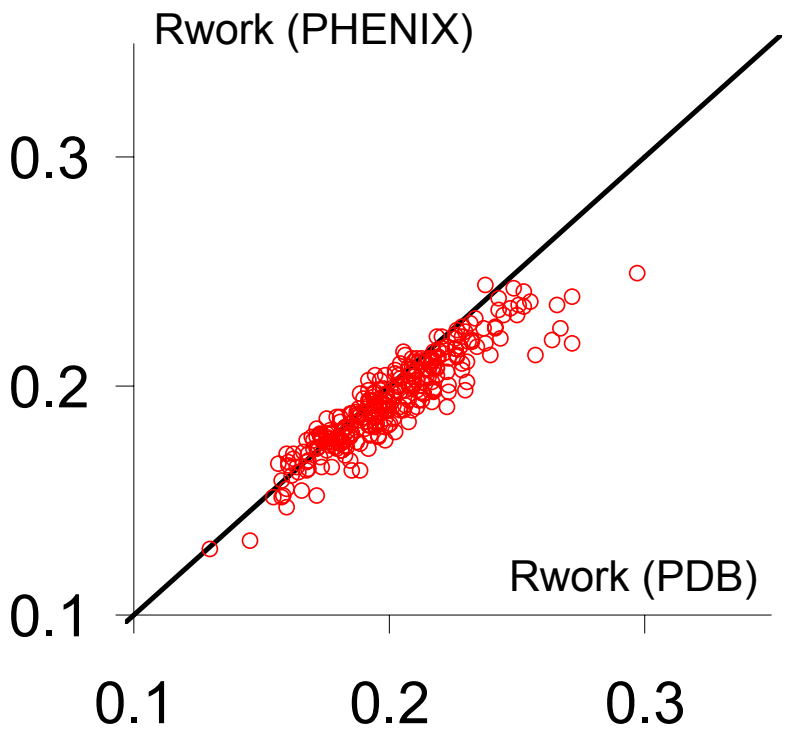
Fixing outliers with PHENIX

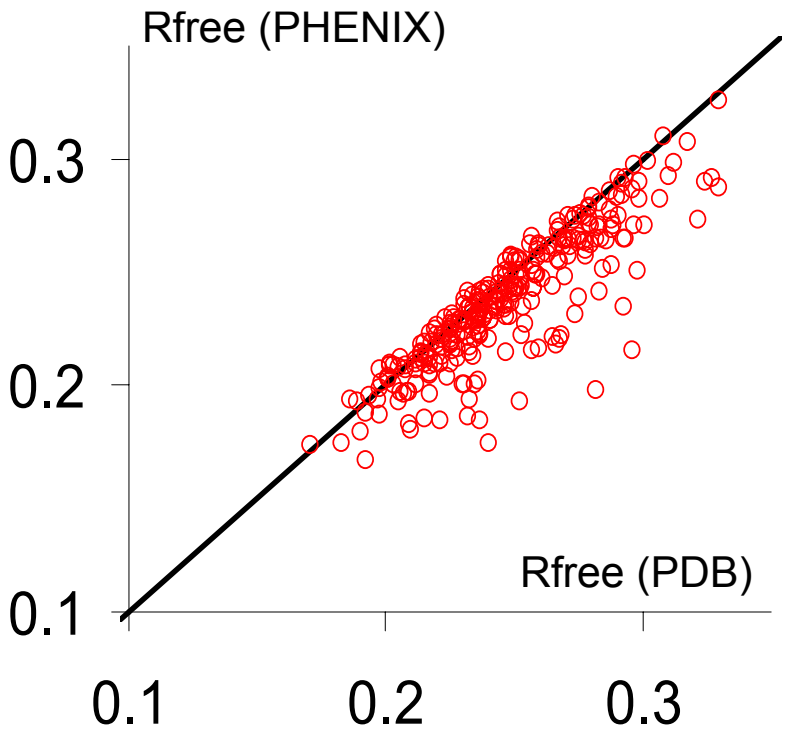


Effect on R-factors



- Re-refinement of 355 PDB models (with TLS parameters)
- In many cases the phenix.refine results are significantly better
- Never crashed or got “unstable”

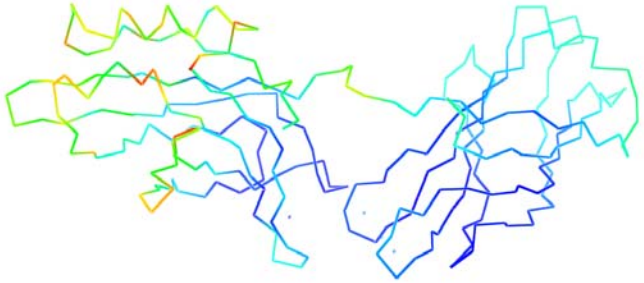




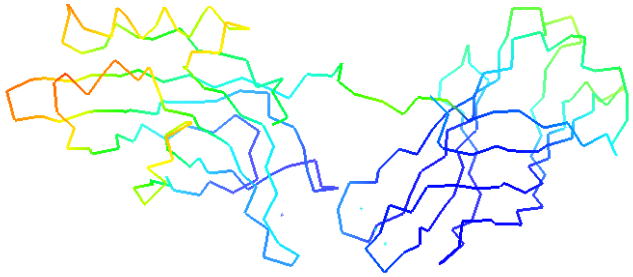
Restraints target for individual isotropic ADP refinement

$$E_{ADP} = \sum_{i=1}^{N_{atoms}} \left[\sum_{j=1}^{M_{atoms}} \frac{1}{r_{ij}^{distance\_power}} \frac{(U_i - U_j)^2}{\left(\frac{U_i + U_j}{2}\right)^{average\_power}} \right]_{sphereR}$$

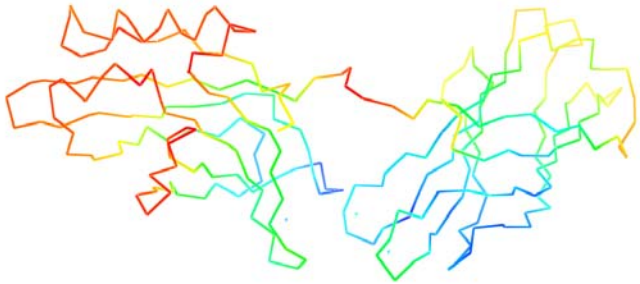
Synaptotagmin refinement at 3.2 Å



CNS
R-free = 34%
R = 29%



PHENIX – Isotropic restrained ADP
R-free = 27.7%
R = 24.6%



PHENIX – TLS + Isotropic ADP
R-free = 24.4%
R = 20.7%

- Currently LS X-ray refinement target:

$$E = \sum_j w_j \left(F_{\text{obs}} - G_{\text{model}}\right)^2$$

$$G_{\text{model}} = \sqrt{(1-\alpha)kF^2_{\text{model,h}} + \alpha kF^2_{\text{model,Rh}}}$$

$$\mathbf{F}_{\text{model}} = \mathbf{F}_{\text{atoms}} + f(k_{sol}, B_{sol})\mathbf{F}_{\text{sol}}$$

- Running twin twin refinement is easy and includes two steps:

  - run phenix.xtriage to get twin operator (twin law):

    % phenix.xtriage data.mtz

  - run phenix.refine:

    % phenix.refine model.pdb data.mtz twin_law="-h-k,k,-l"

- Taking twinning into account makes difference:

PDB code: 1l2h

	R/R-free (%)
Deposited (SHELXL):	15.4 / 19.3
PHENIX (no twinning):	24.9 / 27.4
PHENIX (twin refinement):	15.3 / 19.2

PDB code: 1p7g

	R/R-free (%)
Deposited (CNS):	16.1 / 21.7
PHENIX (twin refinement):	16.2 / 21.6
PHENIX (twin + TLS refinement):	14.6 / 19.6

Macromolecular Neutron Crystallography Consortium (MNC)

Los Alamos National Laboratory
Paul Langan, Marat Mustyakimov, Benno Schoenborn

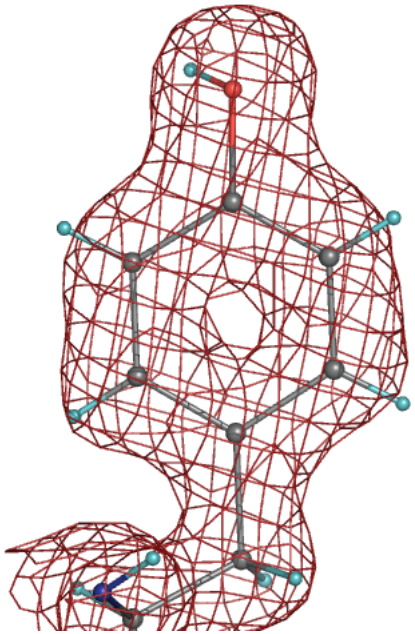
Lawrence Berkeley National Lab (LBNL)
Paul Adams, Pavel Afonine

http://mnc.lanl.gov/

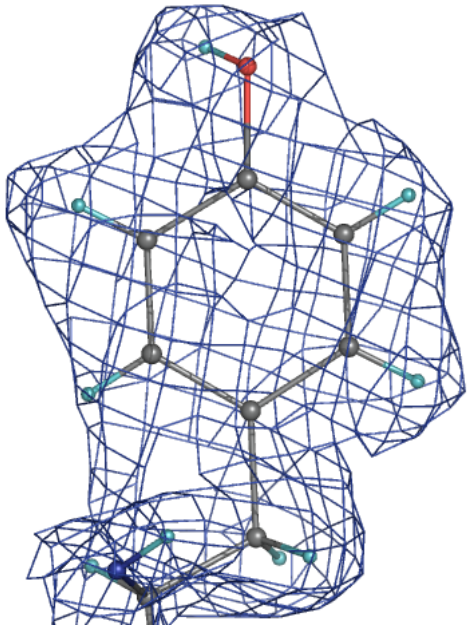
- Different techniques – different information

2mFo-DFc maps (Aldose Reductase)

X-ray (1.8 Å)



Neutron (2.2 Å)

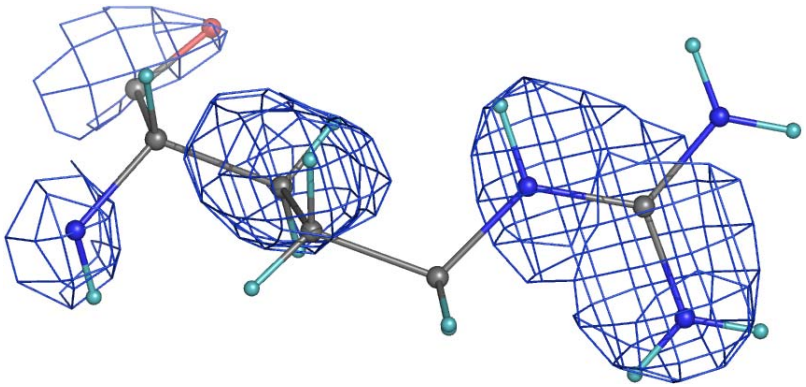


- Target used for joint X-ray + neutron refinement:

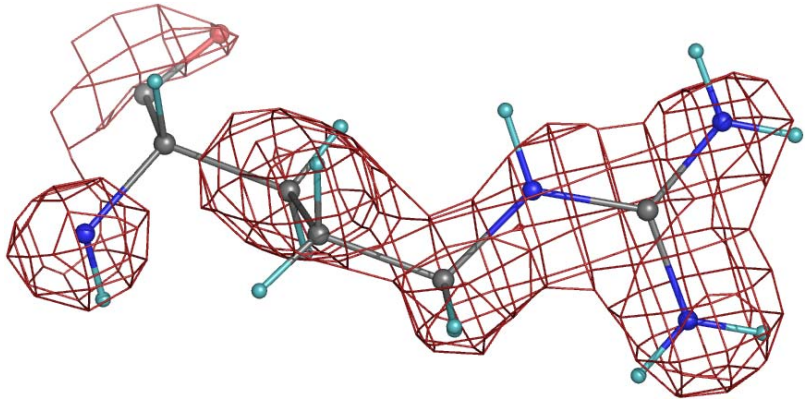
$$Target_{JOINT} = E_{XRAY} * w_{XC} + E_{NEUTRON} * w_{NC} * w_{XN} + E_{GEOM}$$

2mFo-DFc, neutron data, 2σ, 2.2 Å resolution

Refinement (neutron data only)



Refinement (X-ray and neutron data)



- Neutron maps are improved after joint refinement comparing to refinement with neutron data only

Running joint X-ray + neutron refinement

```
% phenix.refine model.pdb data_xray.hkl neutron_data.file_name=data_neutron.hkl
input.xray_data.labels=FOBSx input.neutron_data.labels=FOBSn
```

- phenix.refine offers various options for handling H atoms:

  - Refine as riding model at any resolution
  - Refine as individual atoms (at ultrahigh resolution or using neutron data)
  - Account for scattering contribution or just use to improve the geometry

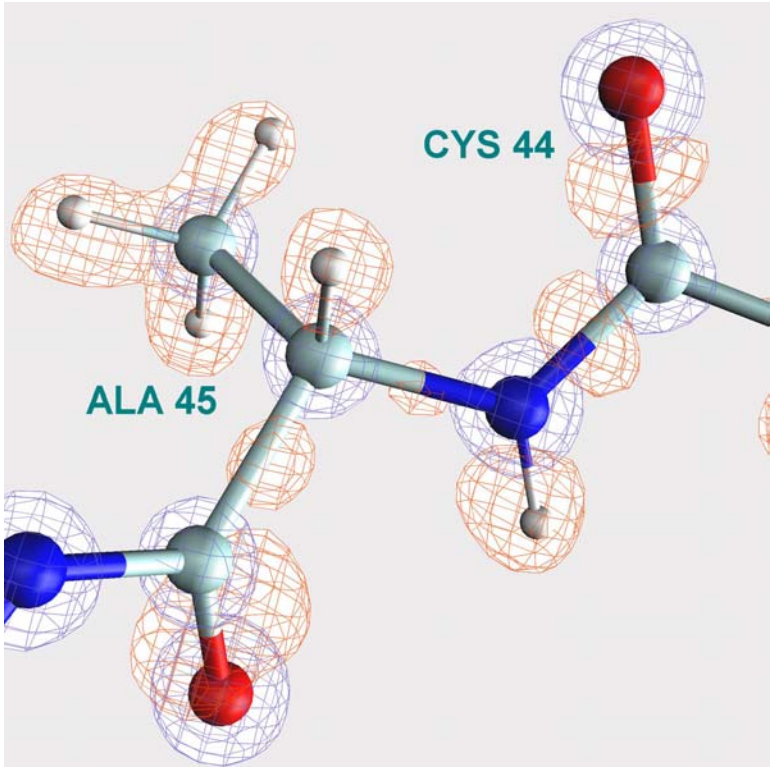
- Expected benefits from using the H atoms in refinement:
  - Improve R-factors
  - Improve model geometry
  - Model residual density at high resolution or in neutron maps

- Example from automatic re-refinement of 1000 PDB models with and without H

pdb	resolution	Rfree(no H) - Rfree(with H)
1dv7	1.3	2.47
1byp	1.8	1.51
1dkp	2.3	1.36
1rgv	2.9	2.01

- Subatomic resolution (higher than ~ 0.9 Å): bond densities and H atoms

Aldose Reductase (0.66 Å resolution)



Fo-Fc (red)

0.28 e/A^3

2Fo-Fc (blue)

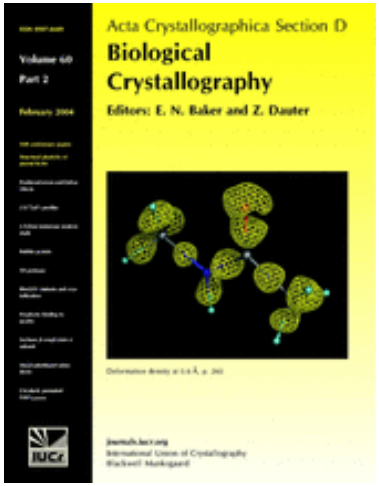
3.85 e/A^3

- Basics of IAS model:

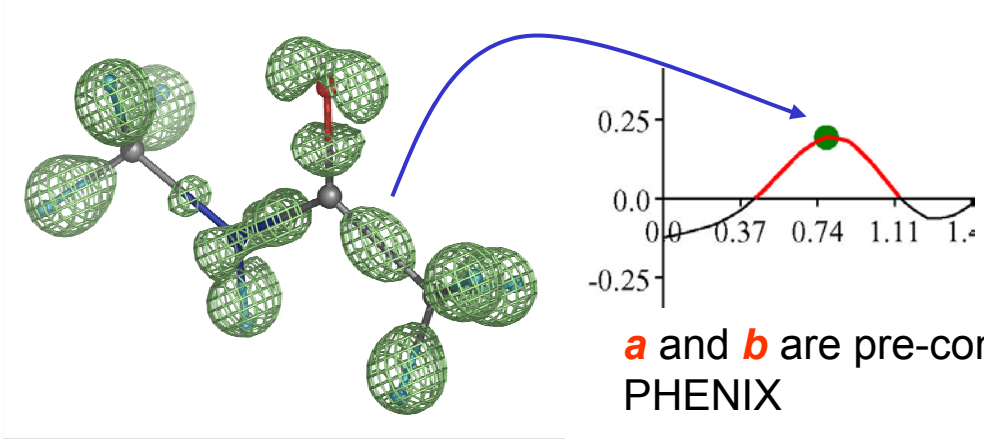
  Afonine *et al*, Acta Cryst. D60, 2004

- First practical examples of implementation and use in PHENIX:

  Afonine *et al*, Acta Cryst. D63, 1194-1197



IAS modeling in PHENIX



Simple Gaussian is good enough:

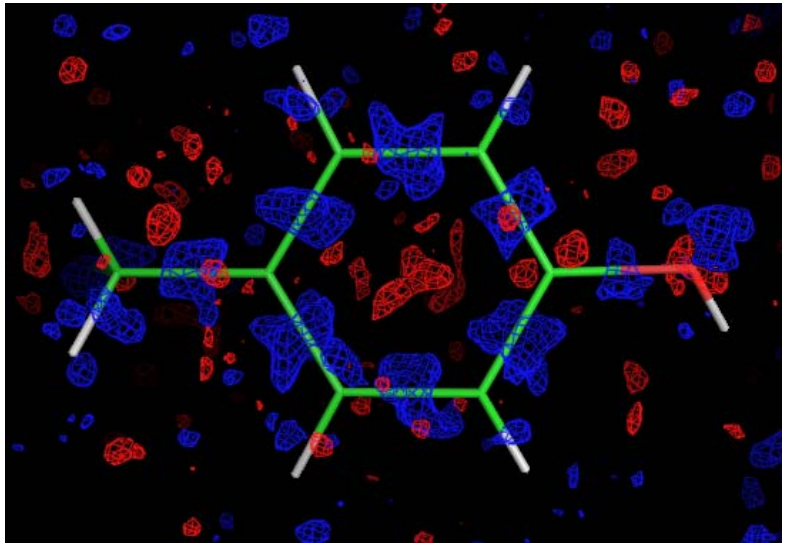
$$f_{bond\_scatterer}(\mathbf{s}) = \mathbf{a}\exp(-\frac{\mathbf{b}}{4}\mathbf{s}^2)$$

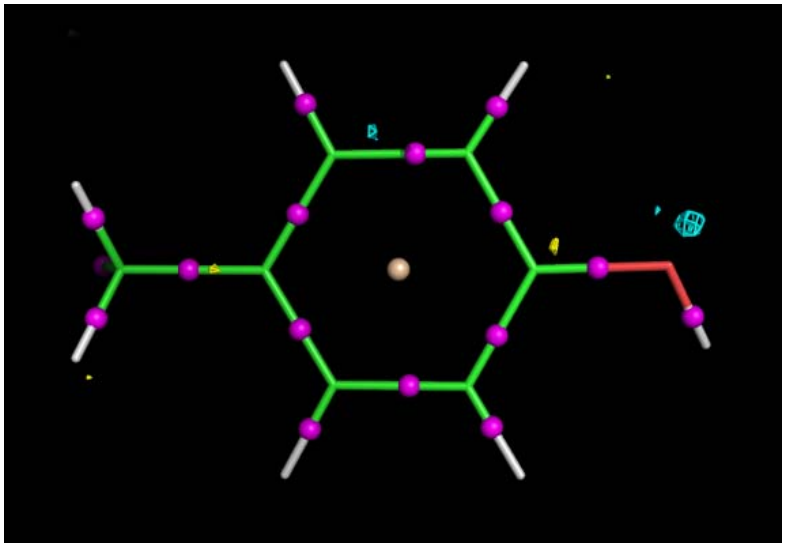
a and ***b*** are pre-computed for most of bonds and implemented in PHENIX

- Compared to Multipolar model that is commonly used at ultra-high resolutions, the new IAS model features:

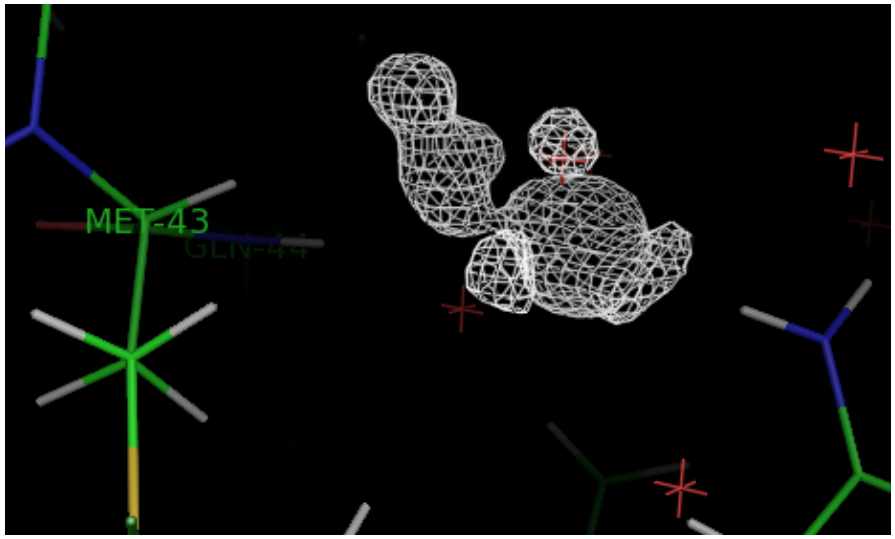
  - faster computations,
  - no risk of overfitting,
  - similar results as Multipolar model (R-factors, ADP, maps)

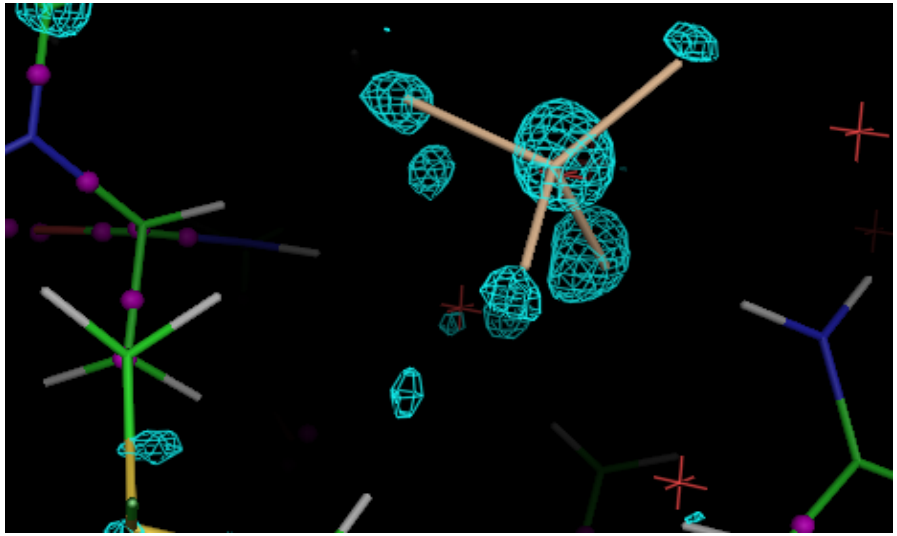
- Improve maps: reduce noise. Before (left) and after (right) adding of IAS.





- Find new features: originally wrong water (left) replaced with SO4 ion (right) clearly suggested by improved map after adding IAS





Future Projects (some of them)

- Real space refinement: fast ligand fitting, reduce amount of manual work
- Normal mode refinement: model more flexibility with general tools
- Torsion angle refinement: efficient low resolution refinement
- Automatic rigid groups search: fully automated TLS refinement
- GUI: ease navigation through refinement process
- Automatic H/D location at maps: enhance neutron and subatomic resolution refinement
- More automation

Availability

`http://www.phenix-online.org/`

Funding

- NIH / NIGMS [*P01GM063210, R01GM071939*, *P01GM064692*]
- LBNL [*DE-AC03-76SF00098*]
- PHENIX industrial consortium

Questions and comments about this document: PAfonine@lbl.gov (Pavel Afonine)