



Computational Crystallography Initiative

Crystallographic structure refinement in PHENIX and more

Pavel Afonine

Computation Crystallography Initiative,
Physical Biosciences Division, Lawrence Berkeley National Laboratory

APS, May 6, 2008

PHYSICAL BIOSCIENCES DIVISION

What is PHENIX?

- PHENIX = Python-based Hierarchical Environment for Integrated Xtallography
- Actively developed package for automated structure solution
- Solid background:
 - Xplor/CNS:
- **New approaches:**
 - Modern programming concepts (Python, C++) and new algorithms
 - Modularization: accelerated development through reuse
 - Integration: combination of heterogeneous algorithms
- Designed to be used by both novices and experienced users
- Long-term development and support

What is PHENIX?

Collaboration between several groups:

- Los Alamos National Lab
Tom Terwilliger, Li-Wei Hung ([SOLVE / RESOLVE](#), [Ligandfit](#), ...)
Paul Langan, Marat Mustyakimov, Benno Schoenborn ([Tools for Neutron crystallography](#)) (separate funding, MNC)
- Cambridge University, UK
Randy Read, Airlie McCoy, Laurent Storoni ([PHASER](#))
- Duke University
Jane & David Richardson, Ian Davis, Vincent Chen ([MolProbity](#), [hydrogens](#))
- Lawrence Berkeley National Lab
Paul Adams, Pavel Afonine, Ralf Grosse-Kunstleve, Nigel Moriarty, Nicholas Sauter, Peter Zwart ([CCI Apps: phenix.refine, phenix.elbow, phenix.xtriage,...](#))
- Texas A&M University
Tom Ioerger, Jim Sacchettini, Erik McKee ([TEXTAL](#))

Paul Adams – project director

CCI Apps: facts

- Subset of PHENIX components developed by the Computational Crystallography Initiative (CCI) at LBNL:
Paul Adams, Pavel Afonine, Ralf Grosse-Kunstleve, Nigel Moriarty, Nicholas Sauter, Peter Zwart
- Command-line tools only. “One click” installation on most of platforms. Does not require installation of anything else. Can be obtained and installed separately from PHENIX package.
- CCI Apps include (list is not complete):
 - 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
 - Xtrriage** – Comprehensive analysis of experimental data
 - phenix.pdbtools** – Set of tools for PDB file manipulations

phenix.refine

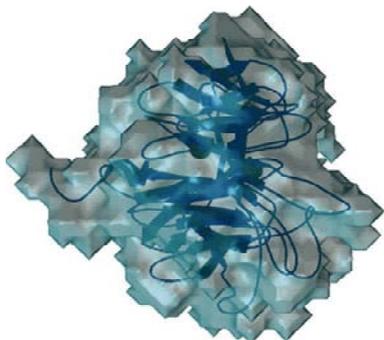
What is phenix.refine ?

phenix.refine

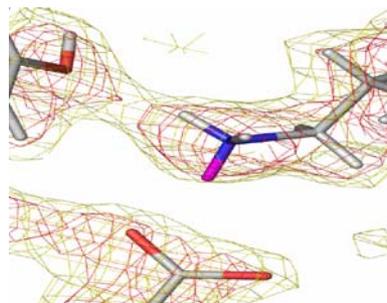
- Highly-automated state-of-the-art structure refinement part of PHENIX
- Under active development by *Paul Adams, Pavel Afonine, Ralf Grosse-Kunstleve, Nigel Moriarty, Peter Zwart*
- Works on most platforms (Linux, Mac, Windows)

phenix.refine: single program for a very broad range of resolutions

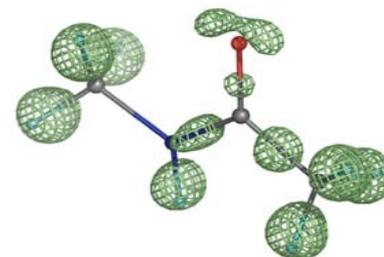
Low



Medium and High



Subatomic



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

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

- 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 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)

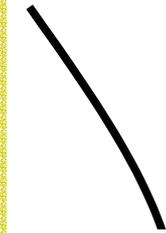
Target weights calculation

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

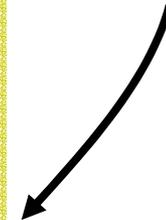
ADP refinement
(TLS, group, individual iso / aniso)

Occupancy refinement (individual, group)

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



Repeated
several times



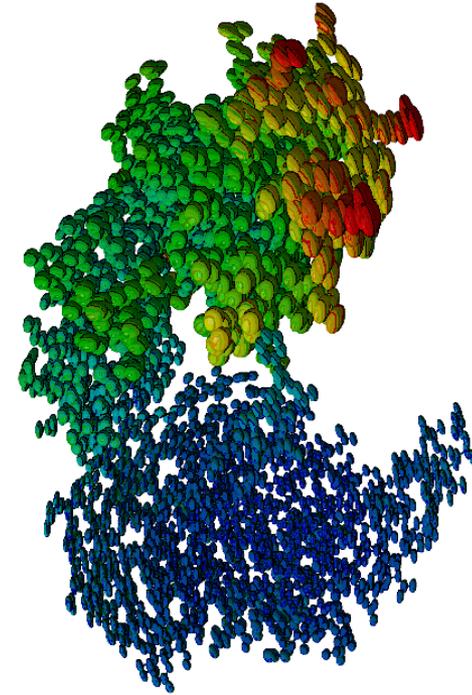
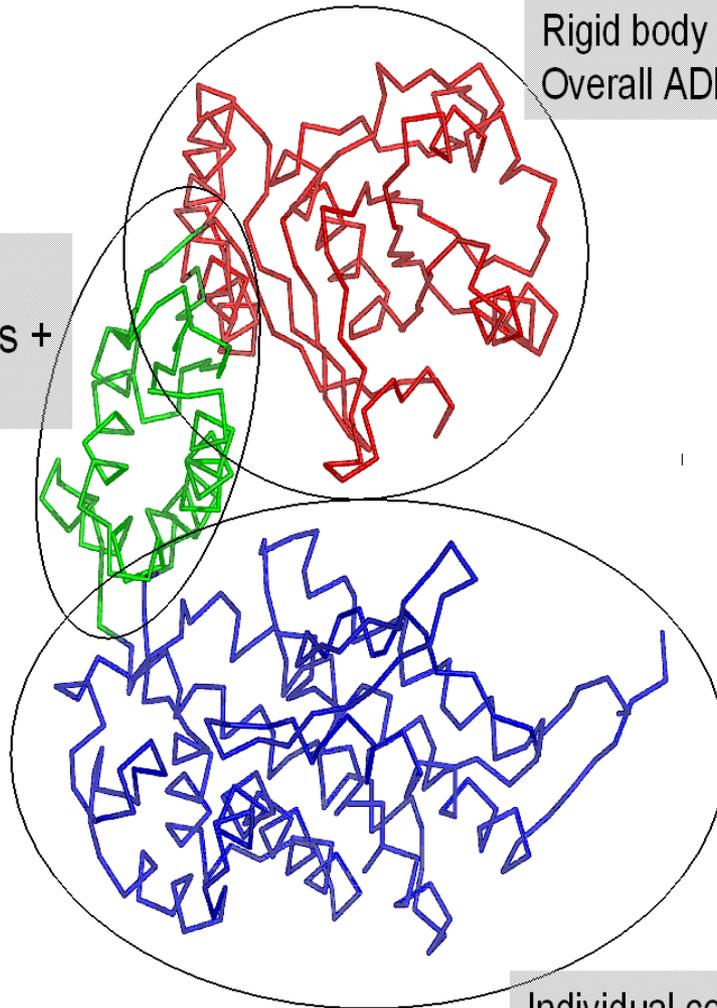
Files for
COOT, O,
PyMol



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

Individual
coordinates +
TLS

Rigid body +
Overall ADP



Individual coordinates +
TLS + restrained isotropic
ADP

- + Automatic water picking
- + Simulated Annealing
- + Add and use hydrogens

Running phenix.refine

Designed to be very easy to use:

Refinement of individual coordinates, B-factors, and occupancies for some atoms:

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

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

```
% phenix.refine model.pdb data.hkl simulated_annealing=true \  
ordered_solvent=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 --show_defaults=all
```

Running phenix.refine

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

where `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  
}
```

Equivalent command line run:

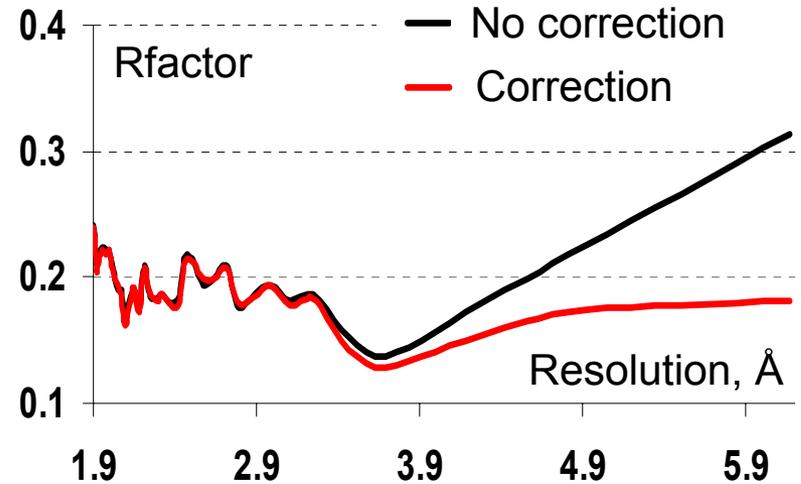
```
% phenix.refine model.pdb data.hkl xray_data.high_resolution=2  
xray_data.low_resolution=15 simulated_annealing=true  
ordered_solvent=True adp.tls="chain A" adp.tls="chain B"  
main.number_of_macro_cycles=5
```

Bulk Solvent : facts

- Macromolecular crystals contain ~20 - 80% of solvent, most of it is disordered and is called bulk solvent.

- Bulk solvent significantly contributes to low resolution reflections (~4-6Å and lower).

Effect on total R-factor: from invisible to several percents (function of data resolution).



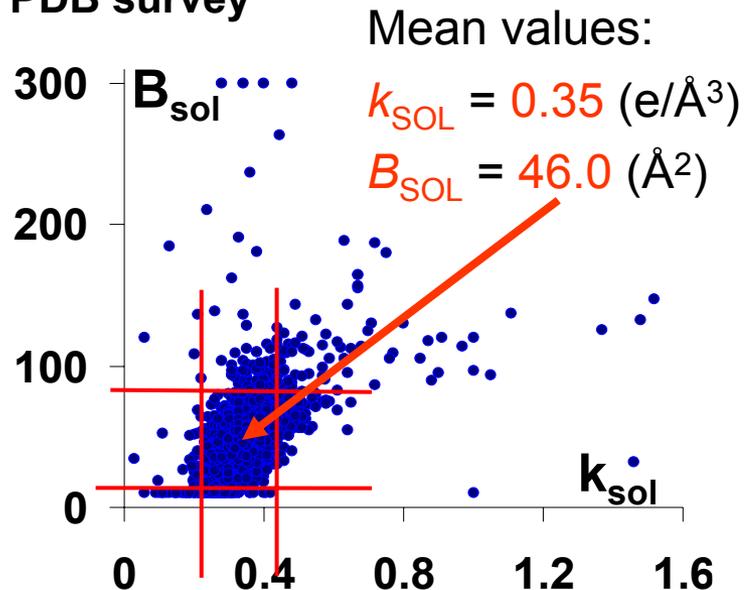
- Flat Bulk Solvent Model* is currently the best. It assumes the constant electron density distribution outside of macromolecular region with $k_{\text{SOL}} \sim 0.35e/\text{Å}^3$ and smearing factor $B_{\text{SOL}} \sim 50\text{Å}^2$.

- Total model structure factor used in refinement and map calculation:

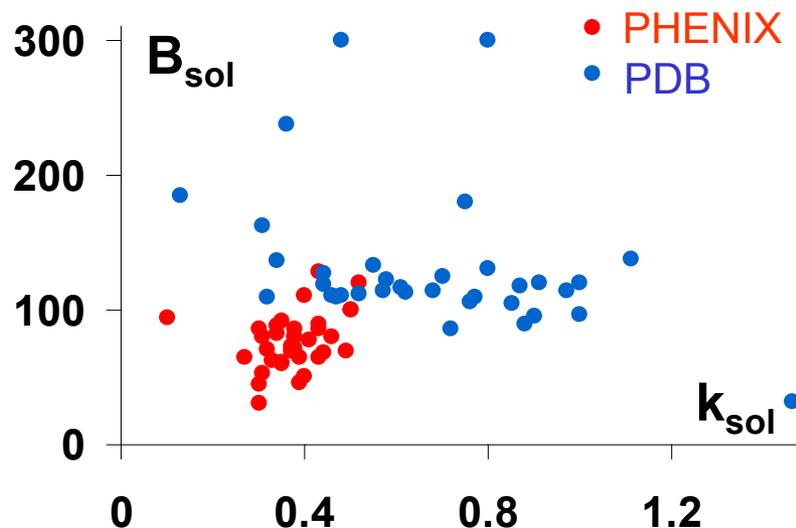
$$\mathbf{F}_{\text{MODEL}} = k_{\text{OVERALL}} e^{-sU_{\text{CRYSTAL}} s^t} \left(\mathbf{F}_{\text{CALC_ATOMS}} + k_{\text{SOL}} e^{-\frac{B_{\text{SOL}} s^2}{4}} \mathbf{F}_{\text{MASK}} \right)$$

Bulk-solvent: robust implementation combined with anisotropic scaling

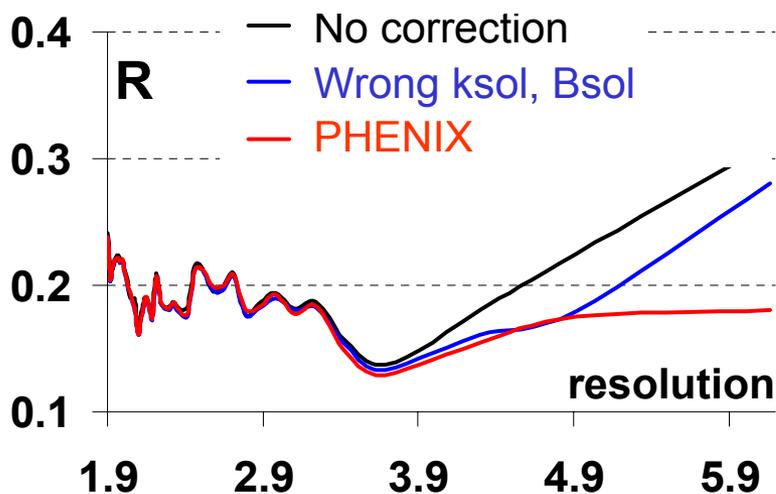
PDB survey



Fixing outliers with PHENIX



Effect on R-factors



Acta Cryst. (2005). D61, 850-855

A robust bulk-solvent correction and anisotropic scaling procedure

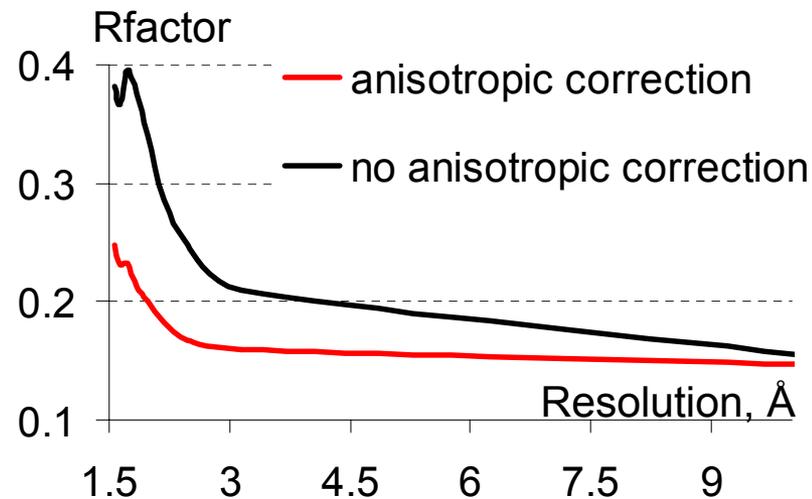
P.V. Afonine, R.W. Grosse-Kunstleve & P.D. Adams

Effect of anisotropic scaling ($\mathbf{U}_{\text{CRYSTAL}}$)

- Total model structure factor used in refinement and map calculation:

$$\mathbf{F}_{\text{MODEL}} = k_{\text{OVERALL}} e^{-s\mathbf{U}_{\text{CRYSTAL}}s^t} \left(\mathbf{F}_{\text{CALC_ATOMS}} + k_{\text{SOL}} e^{-\frac{B_{\text{SOL}} s^2}{4}} \mathbf{F}_{\text{MASK}} \right)$$

- 2MHR model from PDB



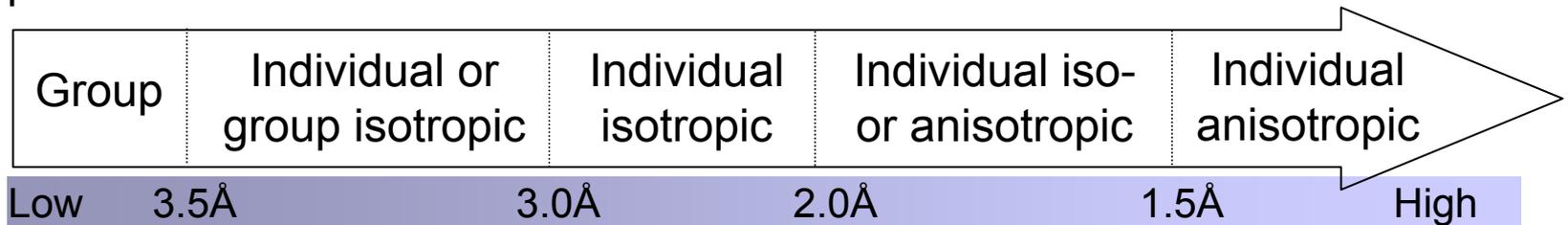
Significant impact on total R-factors:

no correction: R ~ 25%

correction: R ~ 17% , $\mathbf{U}_{\text{CRYSTAL}} = (6.5 \ -9.1 \ 3.8 \ 0 \ 0 \ 0)$

Atomic Displacement Parameters (ADP or “B-factors”)

- Total atomic ADP $U_{\text{TOTAL}} = U_{\text{CRYSTAL}} + U_{\text{TLS}} + U_{\text{INTERNAL}} + U_{\text{ATOM}}$
 - U_{CRYSTAL} - overall anisotropic scale w.r.t. cell axes (6 parameters).
 - U_{TLS} - rigid body displacements of molecules, domains, secondary structure elements. $U_{\text{TLS}} = T + ALA^t + AS + S^tA^t$ (20 TLS parameters per group).
 - U_{INTERNAL} - arising from normal modes of vibration (not modeled in current refinement software packages).
 - U_{ATOM} - vibration of individual atoms. Should obey Hirshfeld’s rigid bond postulate.



TLS refinement in PHENIX: robust and efficient

$$U_{\text{TOTAL}} = U_{\text{CRYSTAL}} + U_{\text{TLS}} + U_{\text{ATOM}}$$

Get start TLS parameters:

- Group isotropic B-factor refinement (one B per residue)
- Split U_{TOTAL} into U_{ATOM} and U_{TLS} (U_{CRYSTAL} is part of scaling):

$$U_{\text{TOTAL}} = U_{\text{TLS}} + U_{\text{ATOM}} + U_{\text{CRYSTAL}}$$

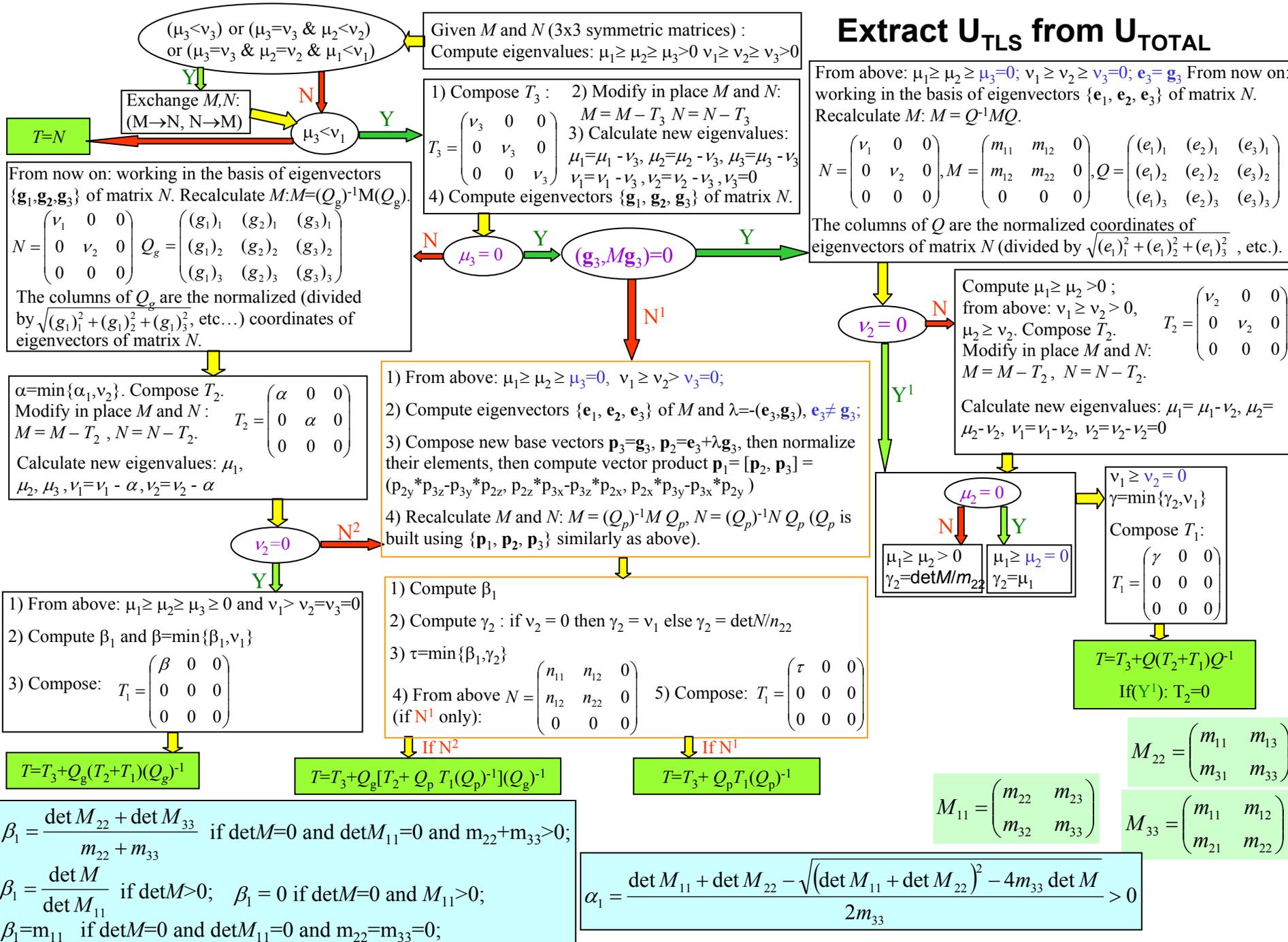
Refine U_{TLS} through refinement of T, L and S:

$$U_{\text{TOTAL}} = U_{\text{ATOM}} + U_{\text{TLS}} + U_{\text{CRYSTAL}}$$

Refine U_{ATOM} (restrained individual isotropic or group):

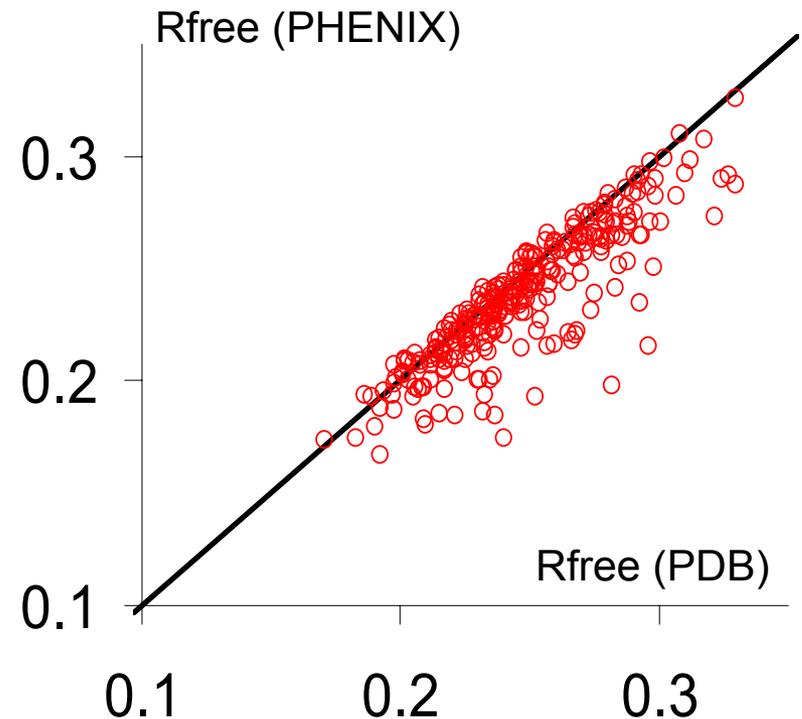
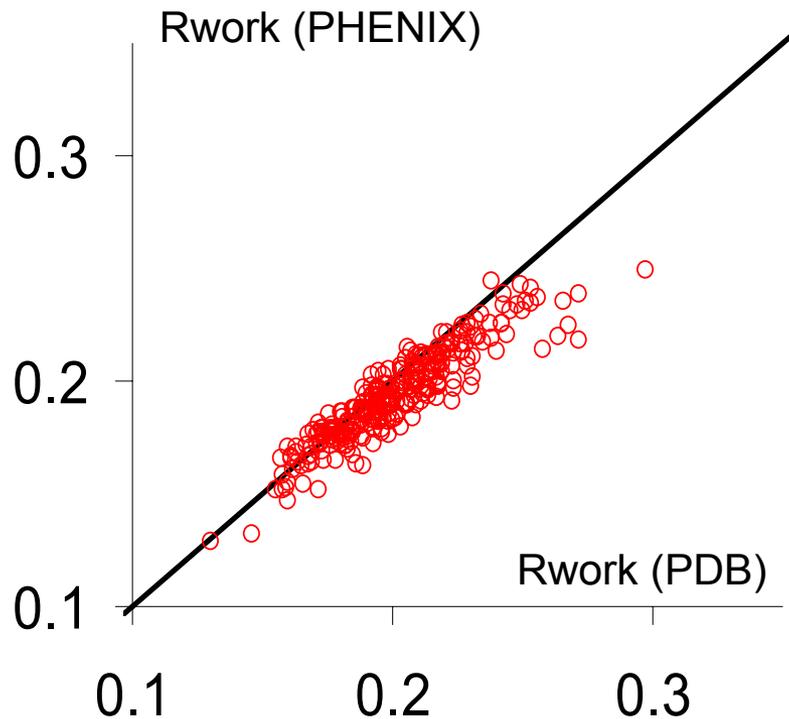
$$U_{\text{TOTAL}} = U_{\text{ATOM}} + U_{\text{TLS}} + U_{\text{CRYSTAL}}$$

Extract U_{TLs} from U_{TOTAL}



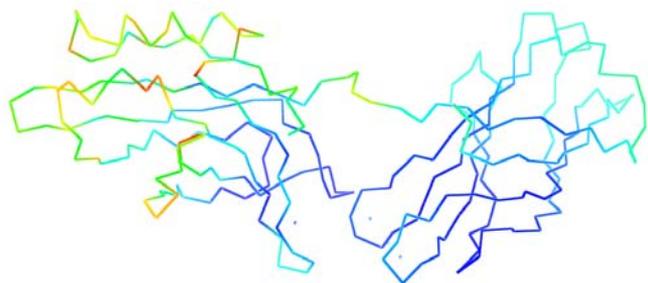
TLS refinement in PHENIX: robust and efficient

- Highly optimized algorithm based on systematic re-refinement of ~350 PDB models
- In most of cases phenix.refine produces better R-factors compared to published
- Never crashed or got “unstable”



ADP refinement : from group B and TLS to individual anisotropic

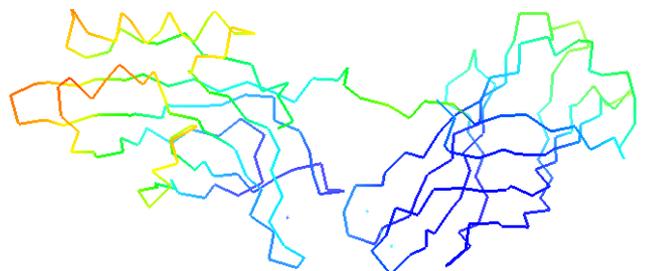
Synaptotagmin refinement at 3.2 Å



CNS

R-free = 34.%

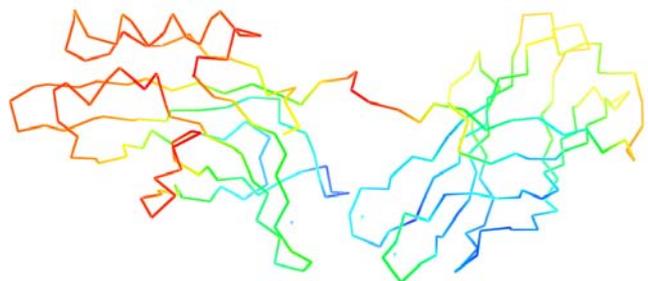
R = 29.%



PHENIX – Isotropic restrained ADP

R-free = 27.7%

R = 24.6%



PHENIX – TLS + Isotropic ADP

R-free = 24.4%

R = 20.7%

ADP refinement: what goes to PDB

phenix.refine outputs TOTAL B-factor (iso- and anisotropic):

$$U_{\text{TOTAL}} = U_{\text{ATOM}} + U_{\text{TLS}} + U_{\text{CRYST}}$$

Isotropic equivalent

ATOM	1	CA	ALA	1	37.211	30.126	28.127	1.00	26.82	C	
ANISOU	1	CA	ALA	1	3397	3397	3397	2634	2634	2634	C

$$U_{\text{TOTAL}} = U_{\text{ATOM}} + U_{\text{TLS}} + U_{\text{CRYST}}$$

Stored in separate record in PDB file header

Atom records are self-consistent:

- ✓ Straightforward visualization (color by B-factors, or anisotropic ellipsoids)
- ✓ Straightforward computation of other statistics (R-factors, etc.) – no need to use external helper programs for any conversions.

Refinement with twinned data

- Two steps to perform twin refinement:

- 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,-1"
```

- Taking twinning into account makes difference:

Interleukin mutant (PDB code: 1l2h)

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

Occupancy refinement

- Default:

- constraints for alternative conformations are built automatically based on input PDB file;
- refine individual occupancies for atoms with partial occupancies (zero occupancies are ignored).

- Optional:

- can do group occupancy refinement (one occupancy per selected group; no constraints or restraints are applied);
- can do individual occupancy refinement.

Restraints and novel ligands in phenix.refine

- 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 topology and parameter definitions for refinement:

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

Or

```
% phenix.elbow model.pdb --do-all
```

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

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

Automatic Water Picking

- **Built into refinement:**

Loop over refinement macro-cycles:

- bulk-solvent and anisotropic scale
- **water picking**
- refinement (XYZ, ADP, occupancies,...)

- **Water picking steps:**

- remove “dead” water:

2mFo-DFc, distances: water-other, water-water, Bmax/Bmin,
anisotropy, occupancy max/min

- add new: mFo-DFc, distances: water-other, water-water

- refine ADP (always) and occupancy (optional) for water only

- remove “dead” water:

2mFo-DFc, distances: water-other, water-water, Bmax/Bmin,
anisotropy, occupancy max/min

- **Very flexible:** there are ~39 parameters available to adjust (if really wanted)

- **Limitation:** no peak sphericity or connectivity analysis (ligand density can be filled)

Neutron and joint X-ray/Neutron refinement

Macromolecular Neutron Crystallography Consortium (MNC)



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



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

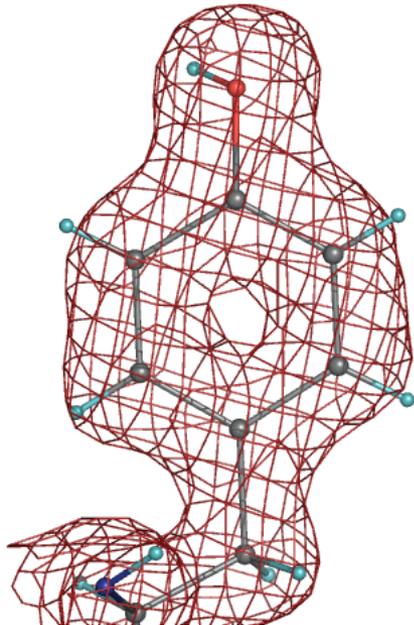
<http://mnc.lanl.gov/>

Maps: X-ray and neutron

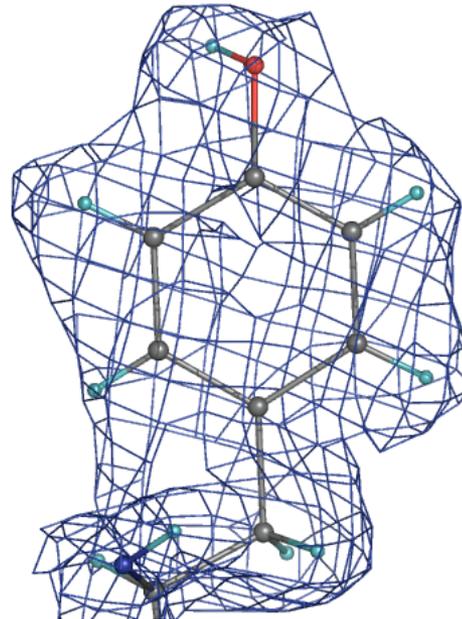
- Different techniques – different information

2mFo-DFc maps (Aldose Reductase)

X-ray (1.8 Å)



Neutron (2.2 Å)

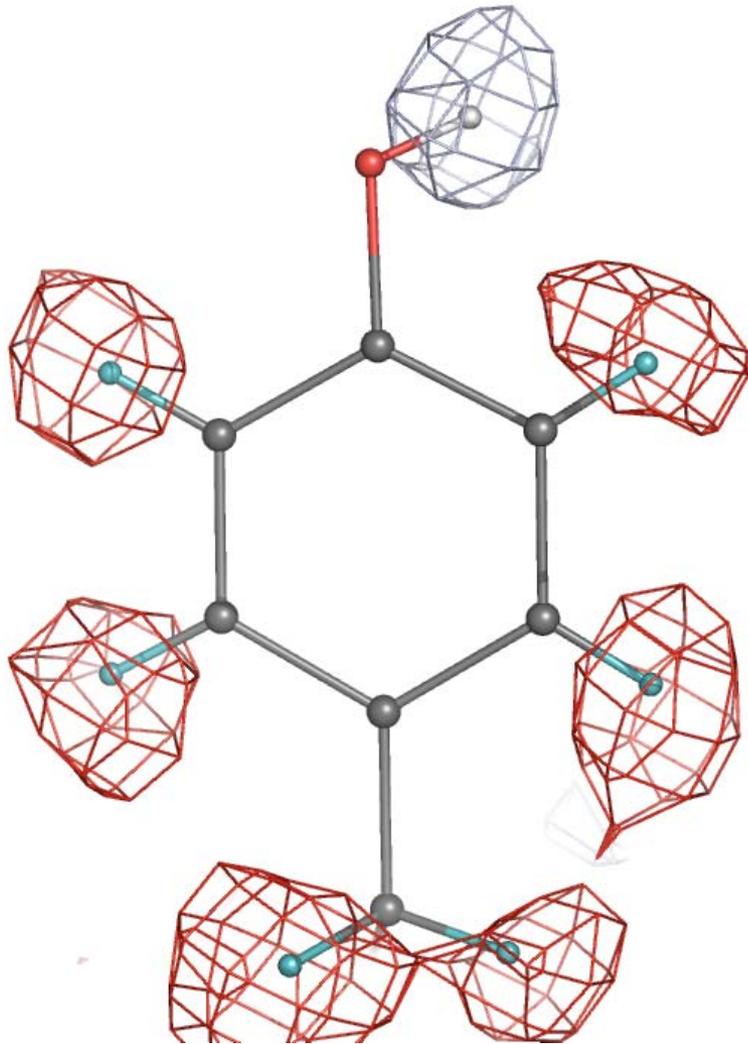


Quantum model of catalysis based on a mobile proton revealed by subatomic x-ray and neutron diffraction studies of h-aldose reductase

PNAS, 2008; 105(6): 1844 - 1848.

Maps: X-ray and neutron

- Different techniques – different information (Automatic determination of H/D state)



PDB: 1iu6 and 1iu5 (resolution $\sim 1.6\text{\AA}$)

joint XN refinement

Fo-Fc map, (H and D omitted), neutron data

positive (blue, 2.6σ , D atoms)

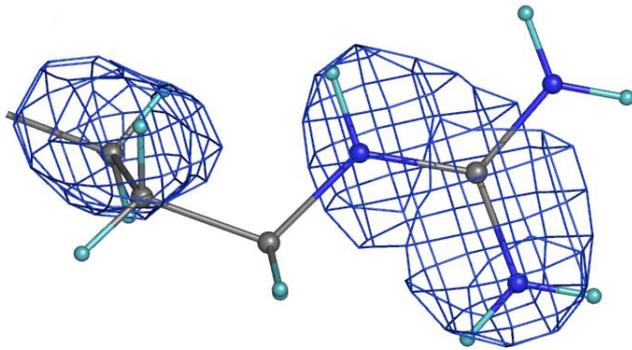
negative (red, -2.9σ , H atoms)

Individual neutron and joint X+N refinement

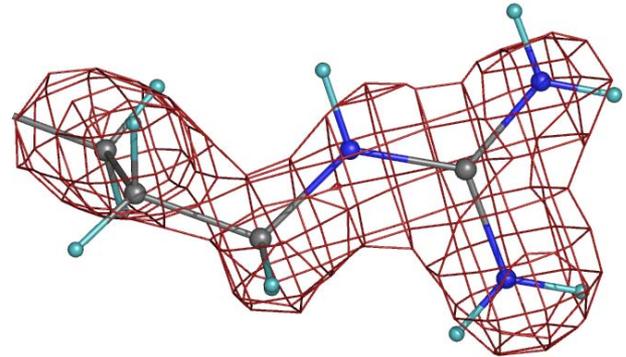
- Maps are improved after joint refinement compared to refinement with neutron data only:

2mFo-DFc, neutron data, 2σ , 2.2 Å resolution (Aldose Reductase)

Refinement (neutron data only)



Refinement (X-ray and neutron data)



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

$$\text{Target}_{\text{JOINT}} = E_{\text{XRAY}} * w_{\text{XC}} + E_{\text{NEUTRON}} * w_{\text{NC}} * w_{\text{XN}} + E_{\text{GEOM}}$$

- Running joint X-ray + neutron refinement in PHENIX

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

Hydrogen atoms in refinement

- phenix.refine offers various options for handling H atoms:
 - Riding model (low-high resolution)
 - Individual atoms (ultrahigh resolution or 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 (remove bad clashes)
 - 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)
1akg	1.1	1.9
1byp	1.75	1.41
1dkp	2.3	0.93
1rgv	2.9	0.50

- Build hydrogens:

```
%phenix.reduce model.pdb > model_H.pdb
```

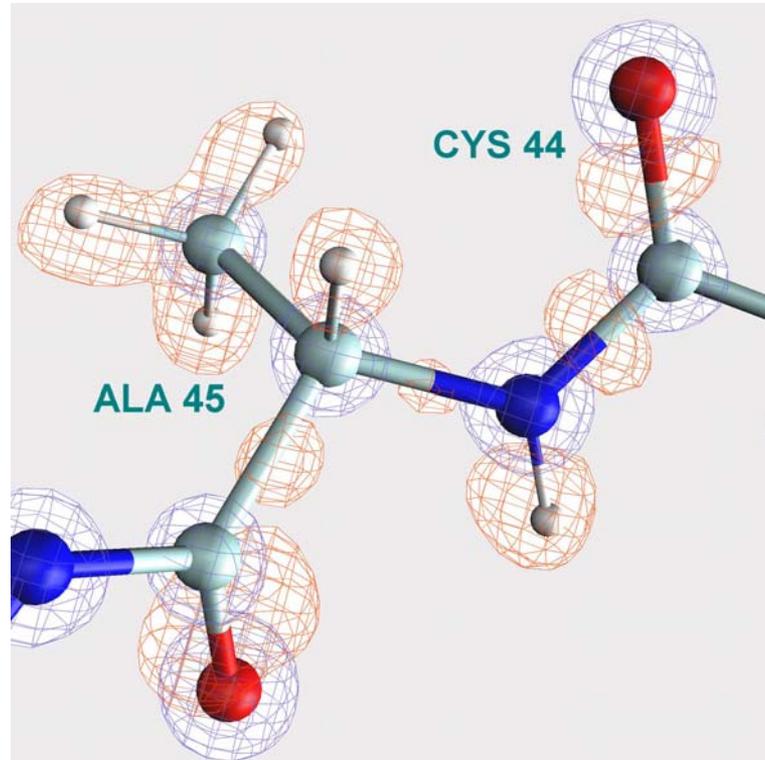
or

```
%phenix.build_hydrogens model.pdb
```

Refinement at subatomic resolution

- Subatomic resolution (higher than $\sim 0.9 \text{ \AA}$): bond densities and H atoms

Aldose Reductase (0.66 \AA resolution)



Fo-Fc (orange)

2Fo-Fc (blue)

Modeling at subatomic resolution: IAS model

- Basics of IAS model:

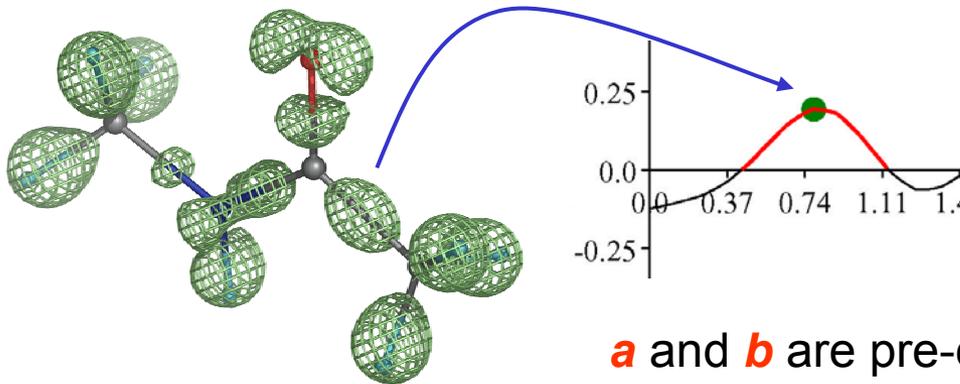
Afonine et al, Acta Cryst. D60 (2004)

- First practical examples of implementation and use in PHENIX:

Afonine et al, Acta Cryst. D63, 1194-1197 (2007)



IAS modeling in PHENIX



Simple Gaussian is good enough:

$$f_{bond_scatterer}(\mathbf{s}) = \mathbf{a} \exp(\mathbf{b} \mathbf{s}^2)$$

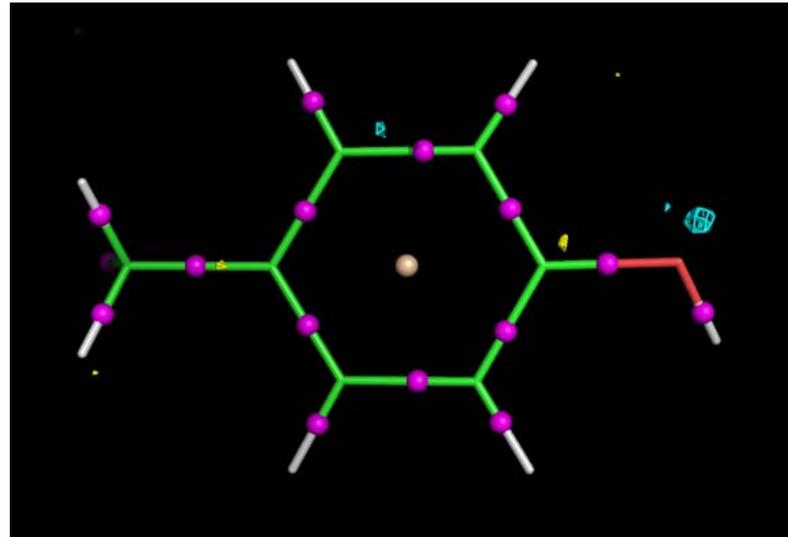
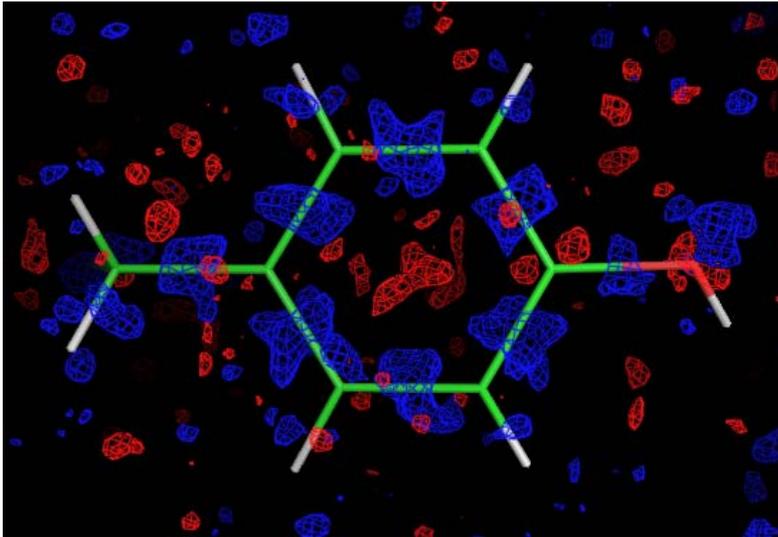
a and **b** are pre-computed library for most bond types

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

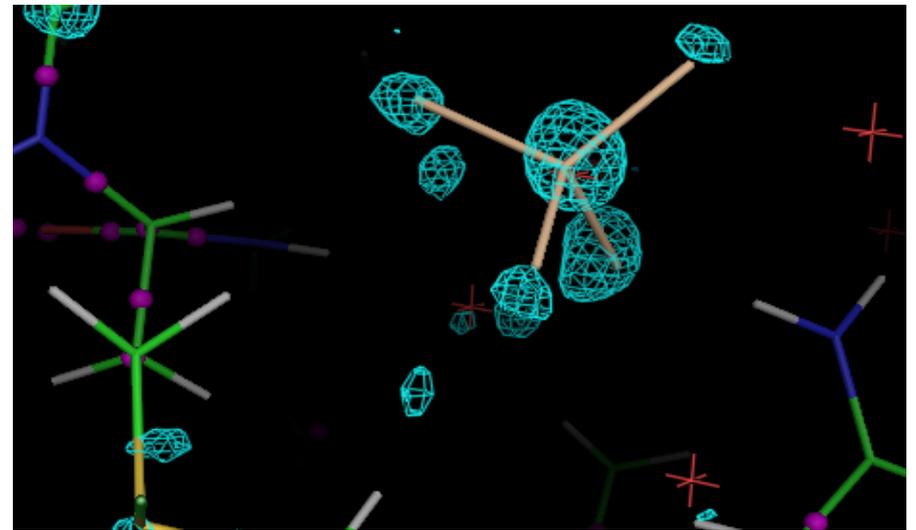
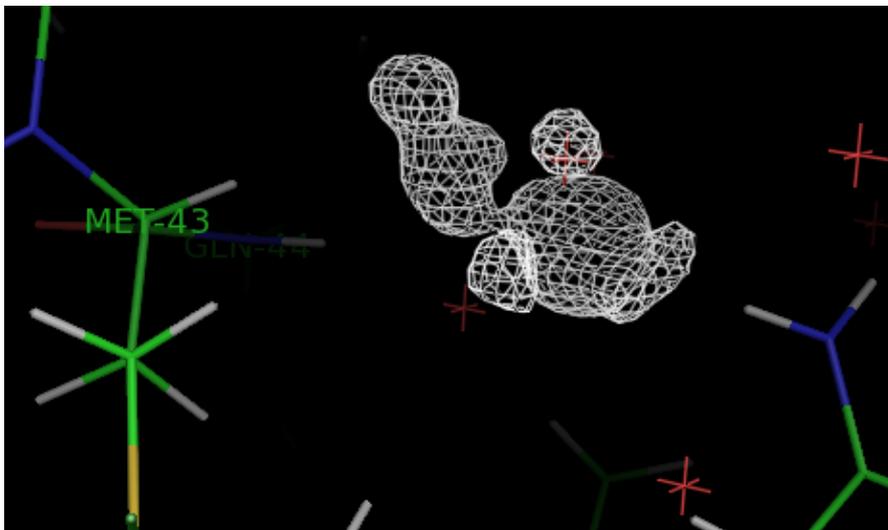
- faster and much simpler computations,
- less or no risk of overfitting,
- similar results as Multipolar model (R-factors, ADP, maps)

IAS modeling: benefits

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

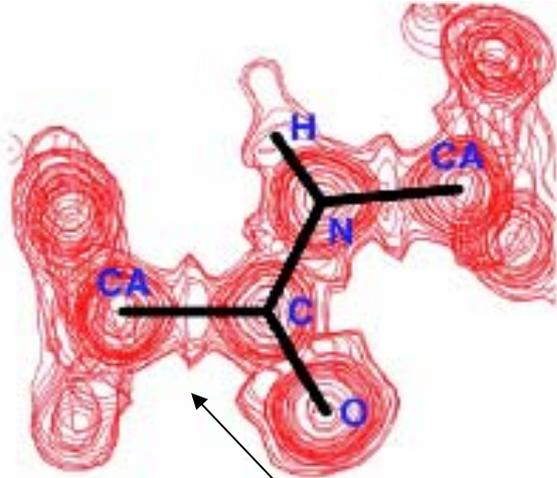


- Find new features: originally wrong water (left) replaced with SO₄ ion (right) clearly suggested by improved map after adding IAS

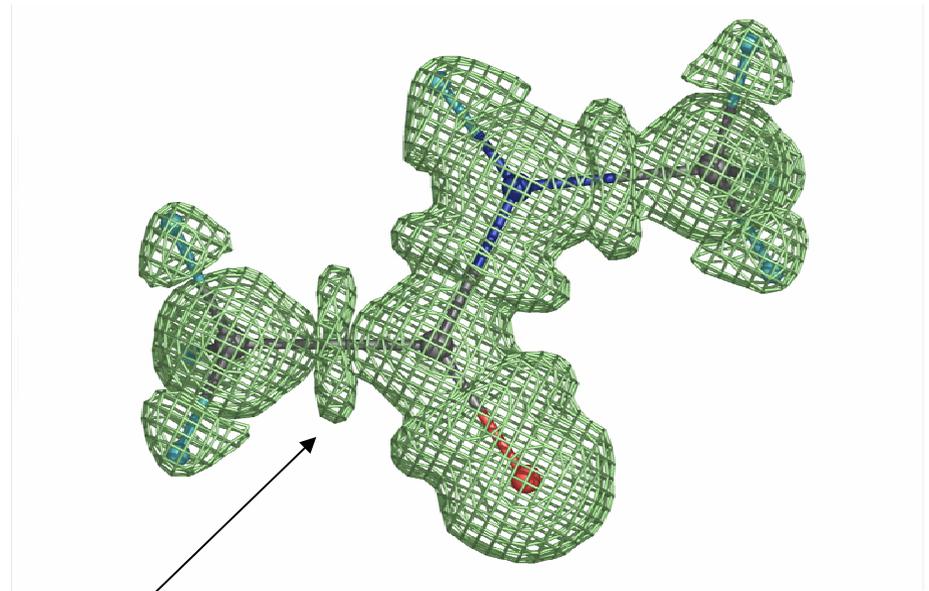


Maps at subatomic resolutions: dangers

- “Experimental Observation of Bonding Electrons in Proteins”, JBC, 1999, Vol. 274.



- $(F_{\text{CALC}}, \phi_{\text{CALC}})$ synthesis at 0.6 Å :



This is not bonding electrons! This is Fourier series truncation ripples !

Shocking examples (or why automation is important...)

- Structure from PDB: **1eic** (resolution = 1.4Å)

PUBLISHED: Rwork = 20% Rfree = 25%

- Clear problems:

- No H atoms;
- All atoms isotropic;

- Potential problems

- Inoptimal weights, refinement is not converged, incomplete solvent model

- Fixing the model with PHENIX:

- Add and refine H as riding model
- Update ordered solvent
- Refine all atoms as anisotropic (except H and water)
- Optimize Xray/Restraints weights

FINAL MODEL: Rwork = 14% Rfree = 17%

Shocking examples (or why automation is important...)

- Structure from PDB: **1ejg** (resolution = 0.54Å)

PUBLISHED: Rwork = 9.0% Rfree = 9.4% (Multipolar model used)

- Clear problems:

- No water molecules;
- No free-R flags;

- Statistics computed using files from PDB “as is”:

Rwork = 20.81%

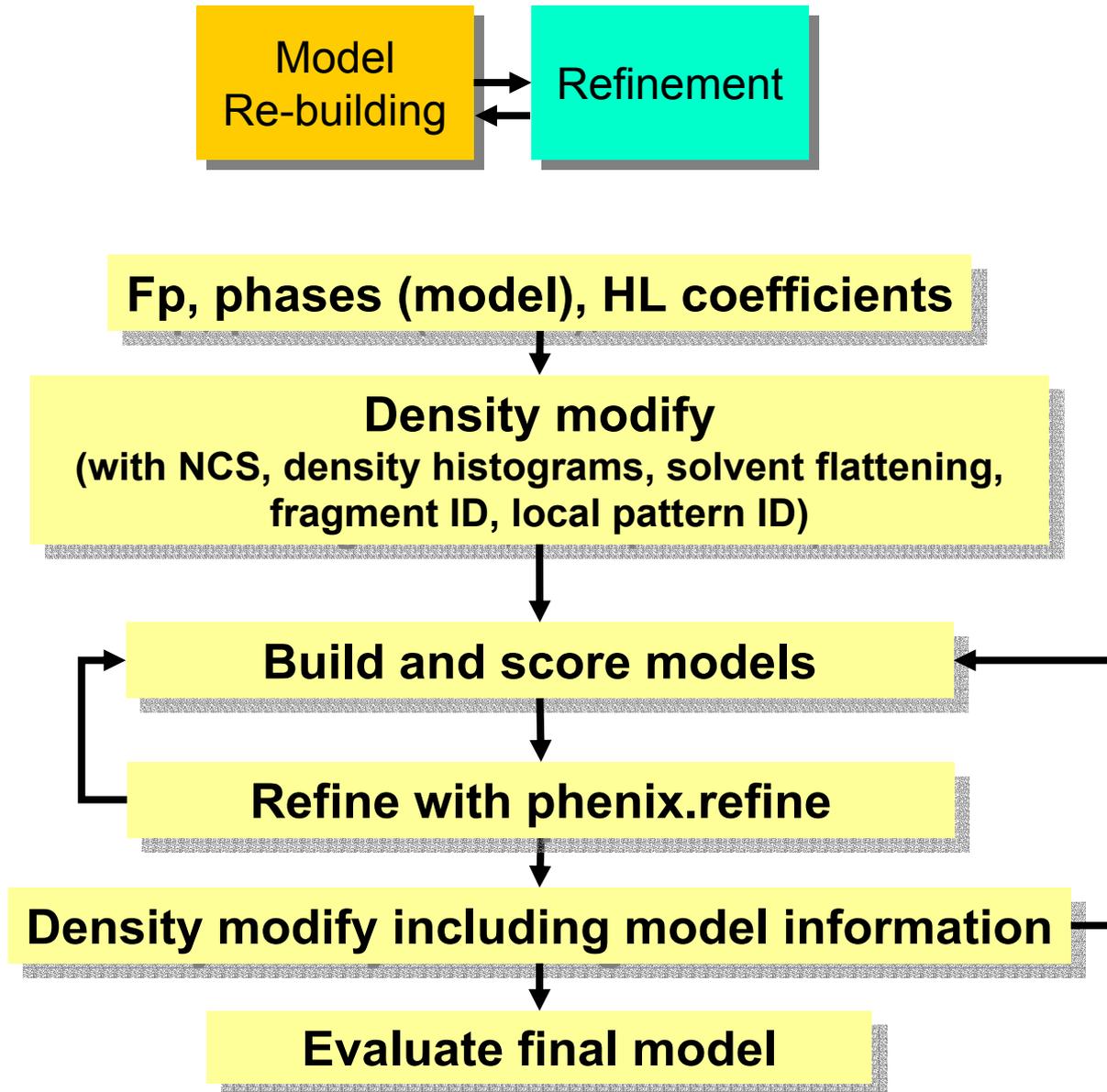
- Fixing the model with PHENIX:

- Add water;
- Add H atoms to water

FINAL MODEL: Rwork = 8.0% (No multipolar model used)

- Using IAS can make it even better

Autobuild wizard in PHENIX: phenix.refine + (SOLVE & RESOLVE)



phenix.pdbtools

- phenix.pdbtools – set of tools for PDB file manipulations
- For any selected model part:
 - shake coordinates, ADP, occupancies
 - rotation-translation shift of coordinates
 - shift, scale, set ADP (add, multiply, assign a constant)
 - converting to isotropic / anisotropic
 - removing selected part of a model
- Easy to run:
% `phenix.pdbtools model.pdb rotate="10 20 30" selection="chain A"`
- Also:
 - complete model statistics (geometry, B-factors)
 - geometry regularization
 - output MTZ with Fcalc (or Fmodel) computed as:

```
Fmodel = scale * exp(-h * bcart * ht) * (  
    Fcalc_atoms + ksol * exp(-bsol * s^2) * Fmask)
```

phenix.superpose_pdbs

- Usage:

- uses alignment if atoms not 100% matching:

```
% phenix.superpose_pdbs fixed.pdb moving.pdb
```

- superpose using selected parts (must exactly match):

```
% phenix.superpose_pdbs fixed.pdb moving.pdb \  
selection_fixed="chain A and name CA" \  
selection_moving="chain B and name CA"
```



NEW [PHENIX 1.3 beta rc6 available](#) ; [Phenix user meeting](#)

Python-based Hierarchical ENvironment for Integrated Xtallography

PHENIX is a new software suite for the automated determination of macromolecular structures using X-ray crystallography and other methods.

Citing PHENIX:

PHENIX: building new software for automated crystallographic structure determination P.D. Adams, R.W. Grosse-Kunstleve, L.-W. Hung, T.R. Ioerger, A.J. McCoy, N.W. Moriarty, R.J. Read, J.C. Sacchettini, N.K. Sauter and T.C. Terwilliger. *Acta Cryst.* D58, 1948-1954 (2002)

Download the latest release (1.3 beta rc6) [First [request download password](#)]

Help: [FAQ](#) [Mailing List Subscription](#) [List Archives](#) [Report a Bug](#) [Email for Help](#)

Using PHENIX (release 1.3 beta rc6):

[Full Documentation](#) [PDF](#)

- Assessing data quality with [phenix.xtriage](#)
- Automated structure solution with [AutoSol](#)
- Automated molecular replacement with [AutoMR](#)
- Automated model building and rebuilding with [AutoBuild](#)
- Automated ligand fitting with [LigandFit](#)
- Structure refinement with [phenix.refine](#)
- Generation of ligand coordinates and restraints with [elbow](#)
- The [PHENIX Graphical User Interface](#)

The PHENIX system also includes SOLVE/RESOLVE, Phaser, Textal, the CCI Applications (phenix.xtriage, phenix.refine, elbow and many more), components from Molprobit, and the Computational Crystallography Toolbox in a Python framework.

Funding for PHENIX: [Protein Structure Initiative](#) ([NIH General Medical Sciences](#))

The PHENIX Industrial Consortium

For-profit groups can obtain access to PHENIX through a Consortium agreement. This provides a license to use PHENIX and research funds to develop new features in PHENIX tailored to the needs of commercial users.

Groups developing PHENIX:

[Paul Adams](#)

[Randy Read](#)

[Jane & Dave Richardson](#)

[Tom Terwilliger](#)

[Tom Ioerger & Jim Sacchettini](#)



[Introduction to PHENIX](#)

[Using PHENIX](#)

[Platforms](#)

[Licensing](#)

[Download](#)

[Recent Changes](#)

[Publications](#)

[Presentations](#)

[Computational Crystallography Toolbox](#)

[Contact Us](#)

[The PHENIX Team](#)

[Acknowledgments](#)

[Intranet](#)

[Information](#)

[Members](#)

[Download](#)

[Contact Us](#)

Reporting bugs, problems, asking questions

- **Something didn't work as expected?... program crashed?... missing feature?...**
 - **Bad:** silently give up and run away looking for alternative software.
 - **Good:** report us a problem, ask a question, request a feature (explain why it's good to have), ask for help (send data).
- **Reporting a bug / problem:**
 - **Bad:** "Hi! phenix.refine crashed and I don't know why and what to do."
 - **Good:** "Hi! phenix.refine crashed. Here are:
 - 1) PHENIX version;
 - 2) The exact command I used;
 - 3) Input and output files (at least logs)."

PHENIX: www.phenix-online.org

- Computational Crystallography Initiative
 - Paul Adams
 - Nigel Moriarty
 - Nick Sauter
 - Peter Zwart
 - Ralf Grosse-Kunstleve
- Los Alamos National Laboratory
 - Tom Terwilliger
 - Li-Wei Hung
- Cambridge University
 - Randy Read
 - Airlie McCoy
 - Laurent Storoni
- Texas A&M University
 - Tom Ioerger
 - Jim Sacchettini
 - Erik McKee

- Others
 - Axel Brunger
 - David Abrahams
 - CCP4 developers
 - Alexei Vagin & Garib Murshudov
 - Kevin Cowtan
 - Sasha Urzhumtsev
 - Vladimir Lunin
- Duke University
 - Jane and David Richardson
 - Ian Davis
 - Vincent Chen
 - Bob Immormino

Funding:

NIH / NIGMS [*P01GM063210, R01GM071939, P01GM064692*]

LBNL [*DE-AC03-76SF00098*]

PHENIX industrial consortium