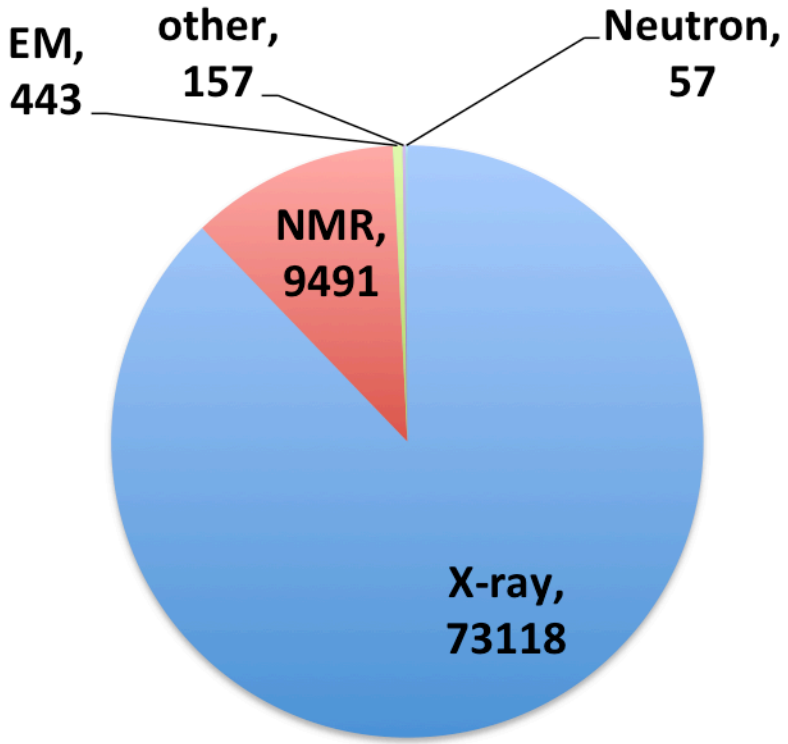
Hydrogens, X-rays, neutrons and all that…

Pavel Afonine

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

PDB overview

- Total number of structures in the Protein Data Bank (31 July 2012): 83266

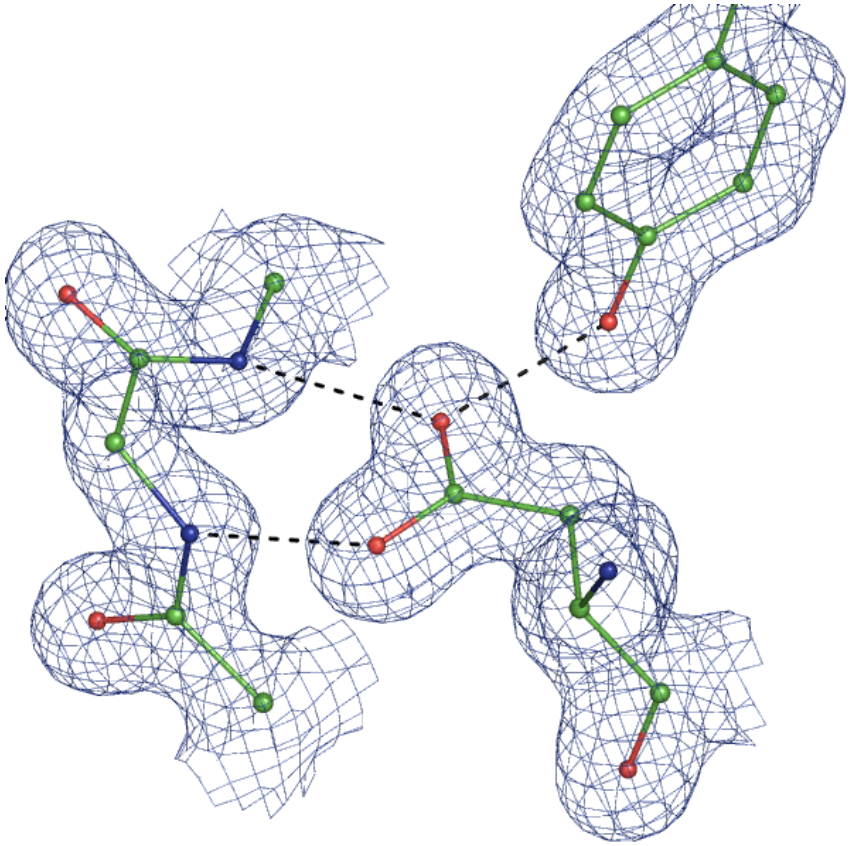


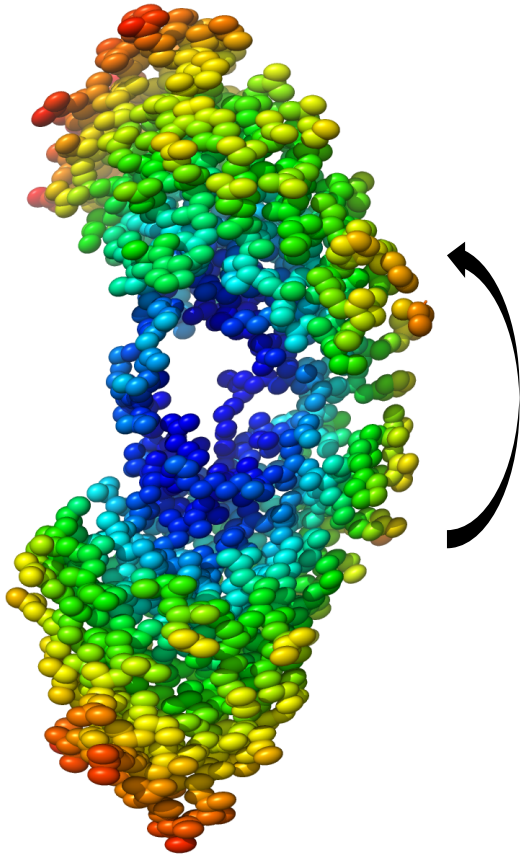
- Crystallography is clearly a leading tool for the determination of bio-macromolecular structure

X-ray crystallography of biomacromolecules

- X-ray diffraction gives information about **non-hydrogen** atoms

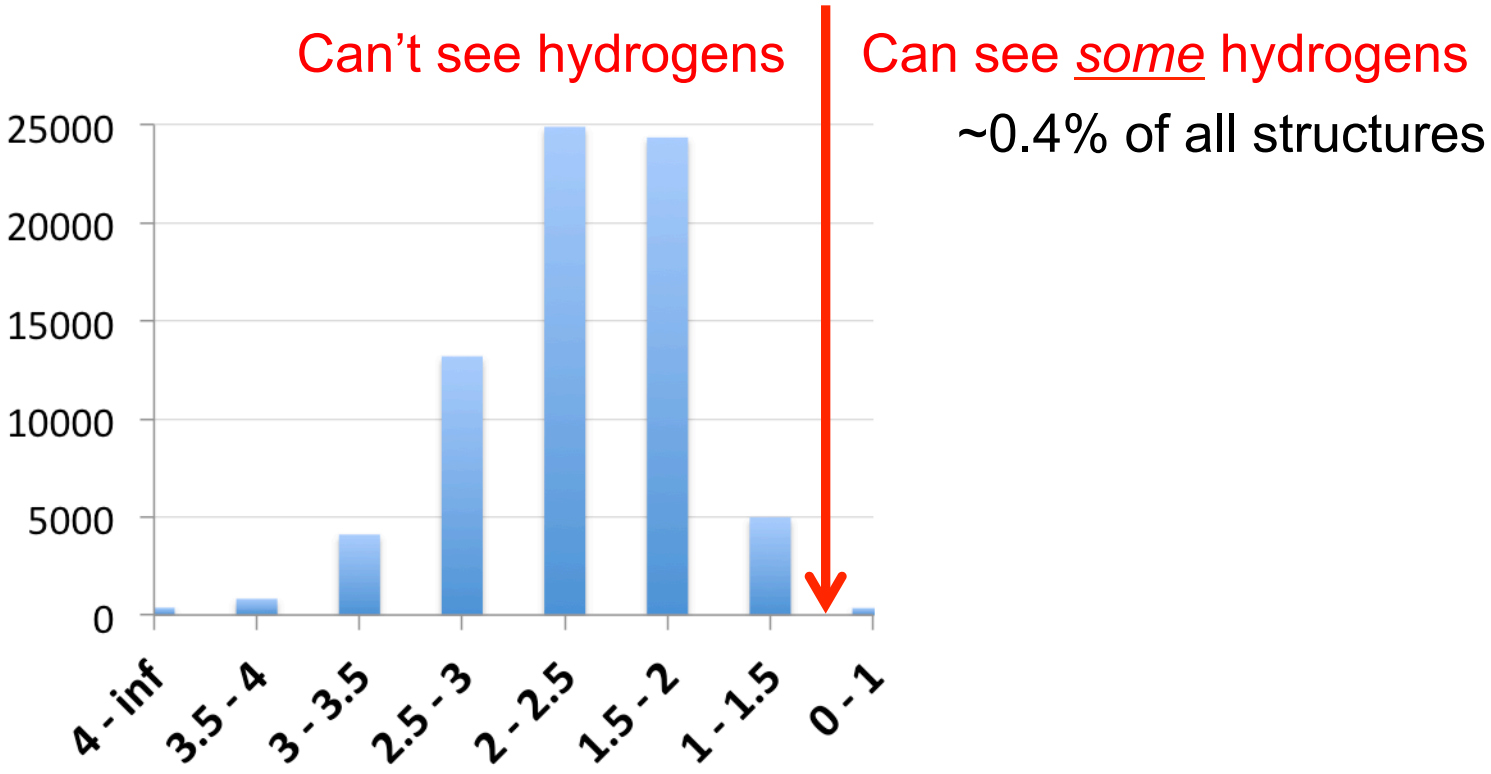
Positions and Disorder





X-ray crystallography and hydrogen atoms

- Number of structures in the PDB *vs* resolution (Å)



- Reasons:
  - Hydrogen is a weak X-ray scatterer
  - Radiation damage: hydrogen abstraction

(*Meents, et al., 2009, J.Sync.Rad., 183-190*)

X-ray crystallography and hydrogen atoms

- Fraction of observed H for some ultra-high resolution X-ray structures

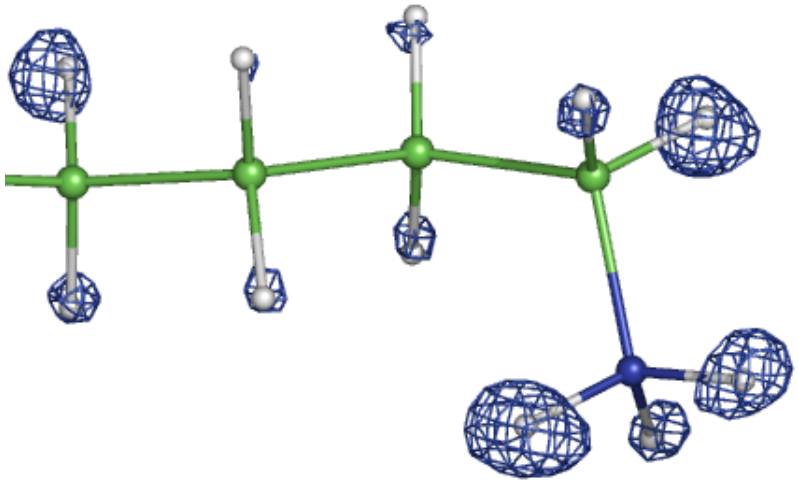
Structure	PDB code	Resolution, Å	% of observed H
Aldose reductase	1US0	0.66	**54**
Lysine-49 phospholipase A2	1MC2	0.8	**38**
Beta-lactamase Tem-1	1M40	0.85	**70**

Petrova & Podjarny, 2004. Rep. Prog. Phys., 1565-1605

- Hydrogen density is often weak for some atoms

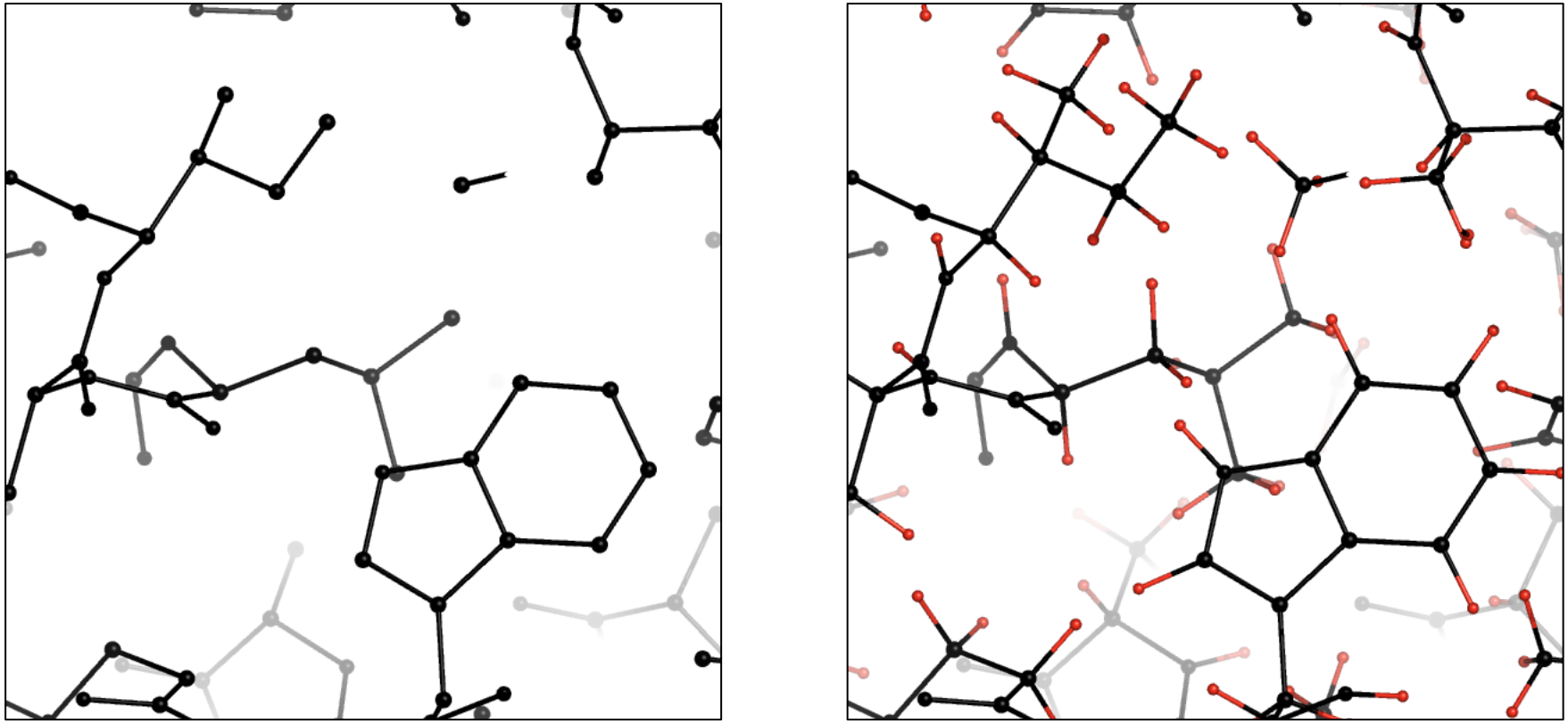
1US0: H-omit mFo-DFc map

(Lys 77, countered at 3σ)



Hydrogen atoms: why care about?

- Hydrogen atoms:
  - Make up half of the atoms in a protein molecule
  - Make most interatomic contacts
  - Positions of most can be inferred from positions of other atoms
  - 10-15% have rotational d.o.f. => can't be unambiguously placed



Identical view of a macromolecular structure portion without (left) and with (right) hydrogen atoms

Hydrogen atoms: significance

- Electrostatic interactions (geometry of hydrogen bonding involved in stabilizing molecules).
- Determination of the protonation states of catalytic groups.
- Orientation of key water molecules in enzymatic reactions, molecular recognition and protein folding.
- Knowledge of hydrogen positions facilitates unambiguous ligand docking.
- Parameters of hydrogen bonds can be used in the calculation of hydrogen bond energies to be used in molecular dynamics simulations.
- Selective H/D exchange (in polarized bonds like N—H and O—H) may be used for studying protein flexibility and packing.
- Account for scattering from H

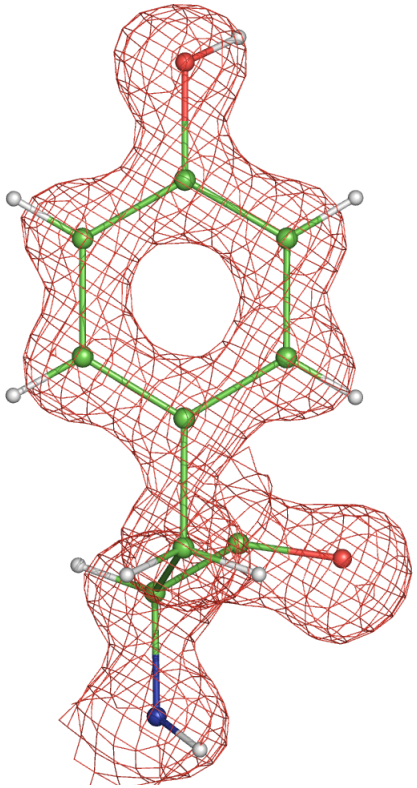
Literature:

Kossiakoff & Spencer (1981),
Wlodawer & Sjölin (1982),
Kossiakoff (1985, 1986),
McDowell & Kossiakoff (1995),
Shu et al. (2000),
Habash et al. (2000),
Engler et al. (2003),
Fenimore et al. (2004),
Kurihara et al. (2004)
Niimura et al. (2004),
Bennett et al. (2006),
Katz et al. (2006),
Chatake et al. (2008),
Blum et al. (2009),
Fisher et al. (2010),
Kovalevsky et al. (2010),
Sukumar et al. (2010)

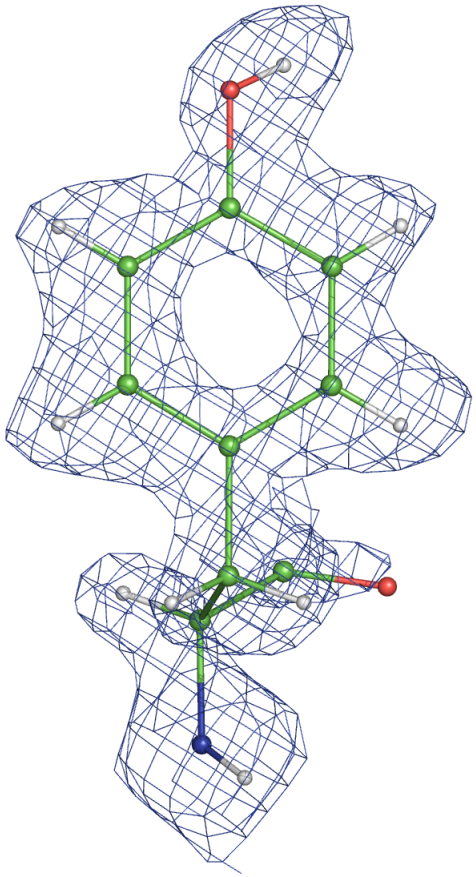
Neutron crystallography

- Nuclear maps show H (D) atoms at typical macromolecular resolutions (~2Å)

X-ray (1.1 Å)



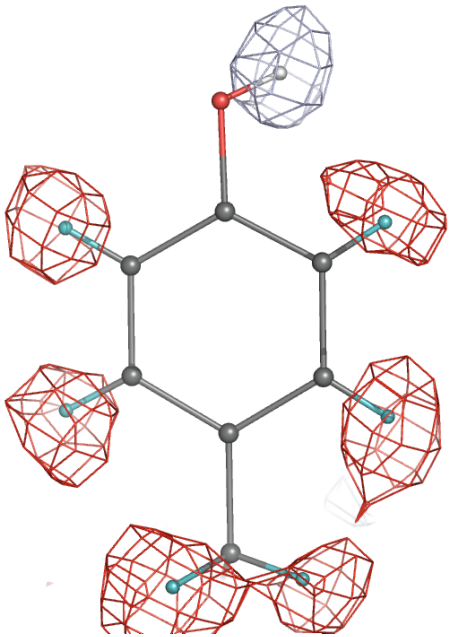
Neutron (1.7 Å)



2mFo-DFc maps at 1.5σ (Rubredoxin, PDB code: 3KKY)

Neutron crystallography

- Partially and fully deuterated samples:
  - H atoms have a negative scattering length => appear as negative peaks
  - Peaks from D atoms are very similar to C or O atom peaks
  - Peaks at exchangeable H/D sites are reduced due to mutual cancellation and may completely vanish at H:D ratio ~ 0.6:0.4



Rubredoxin (PDB code: 1IU6;1.6Å)

mFo-DFc, H/D-omit map, neutron data

positive (blue, 2.6σ, D atoms)

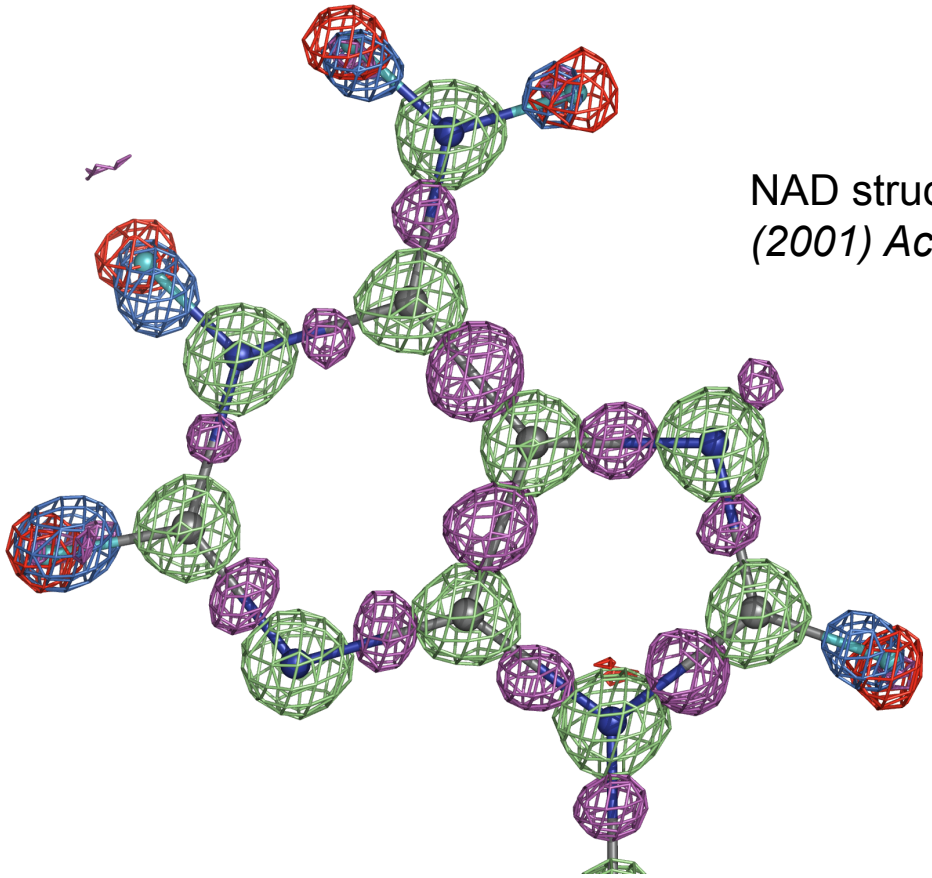
negative (red, -2.9σ, H atoms)

X-ray and Neutron Crystallography: Complimentary Methods

- Still complimentary even at subatomic resolution (NAD structure):

Neutron 2Fo-Fc map (0.65 Å), ±2.4σ, green: positive, red: negative

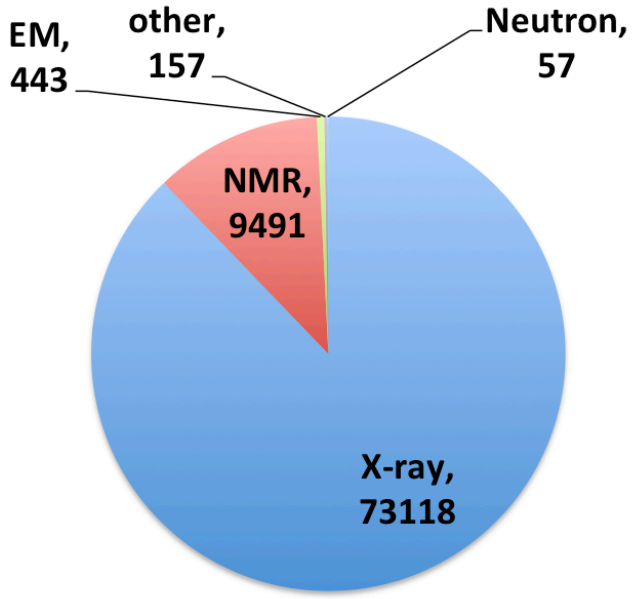
X-ray Fo-Fc map (0.6 Å), blue: H omit, 5σ, magenta: 2.8σ all atoms included



NAD structure: *Guillot et al. (2001) Acta Cryst., D57, 981-989*

Afonine et al., Acta Cryst. (2010). D66, 1153-1163

If it's so good - why so few neutron structures?



- Challenges:
  - Experimental
    - Requirements for large crystals
    - Limited amount of data collection facilities
    - Large data collection times
    - Specifics of sample preparation
  - Structure solution, refinement and completion:
    - Software
    - Methods

… It’s important to use the right tool for the job…

Image: http://www.deadissue.com/archives/2007/01/31/right-tool-for-the-job/

… It’s important to use the right tool for the job…

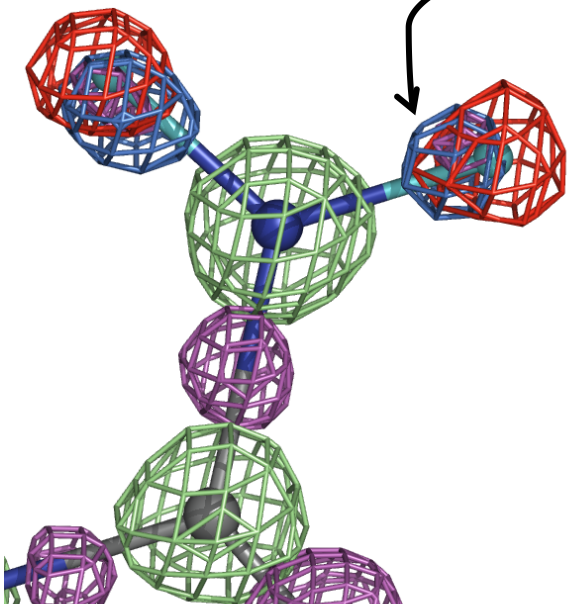
Image: http://www.deadissue.com/archives/2007/01/31/right-tool-for-the-job/

Neutron Crystallography Challenges

- **Software challenges:** programs are designed and optimized to work with X-ray data:
  - Manual (often tedious = error prone) work to customize software to handle neutron data
    - adopt neutron scattering tables
    - handle H, D, H/D atoms (nomenclature naming and geometry)
    - add D or exchangeable H/D
  - Could not use all software features (e.g. TLS, real-space refinement)
  - Reporting results (statistics for PDB deposition)
  - Validation tools are designed to work with X-ray structures
  - Reporting results (PDB files)

Neutron Crystallography Challenges: X-H geometry

- "Short" (X-ray) *vs* "long" (neutron) X-H distances
  - Refinement and validation software are not consistent

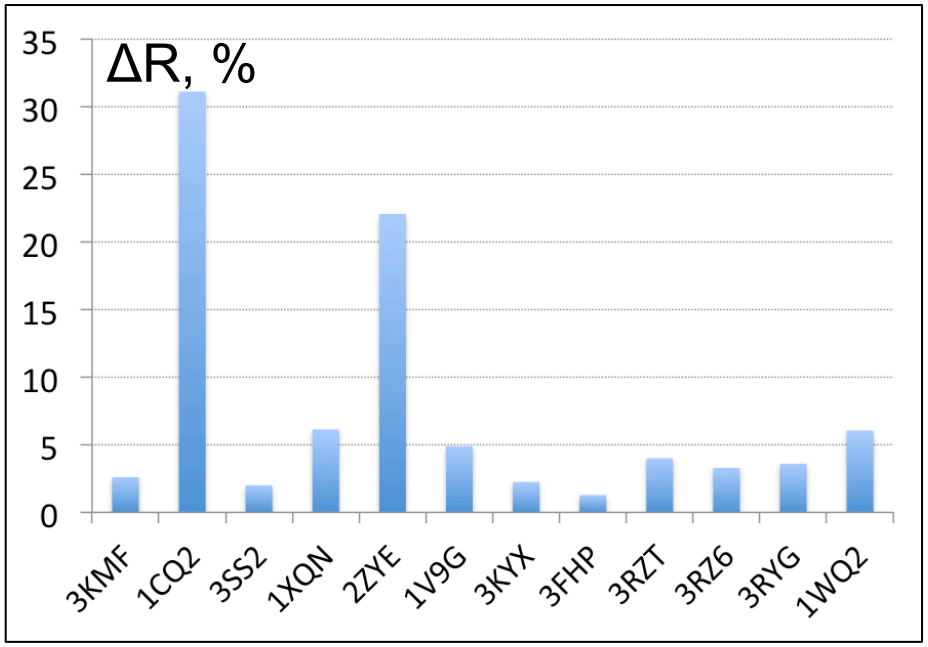


X-ray position of H (density peak is shifted along X-H bond; e.g. used in SHELX)

Neutron (nuclear) position of H; library values *(Allen, Acta Cryst. (1986). B42, 515-522)*

Neutron structures in PDB

- ΔR = R(**recalculated**)-R(**published**)



- **Identified problems** (at least two apply to each structure):
  - Bad or absent free-R flags
  - Negative occupancies
  - Geometry problems
  - Unaccounted twinning

- Sum of exchangeable H/D does not add up to 1
- Iobs *vs* Fobs mismatch in input data file
- Bad or missing information in PDB file header
- H/D exchange is not modeled or incomplete
- Atoms with unknown scattering type

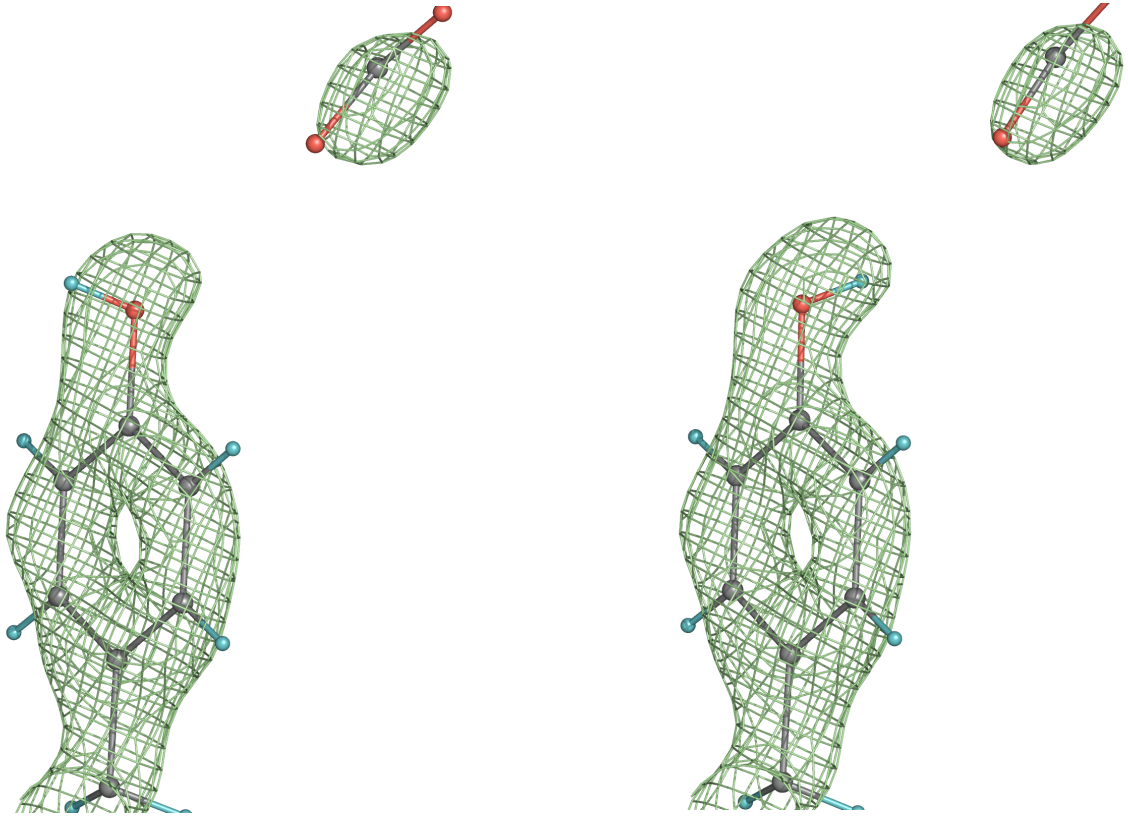
Afonine et al., J. Appl. Cryst. 43, 677-685 (2010)

Neutron Crystallography Challenges

- **Methodological challenges:**
  - Build H, D or H/D in model, including water or ligands
  - Optimize fit of water (DOD) into the density
  - Optimize fit of rotatable X-H/D bonds into the density
  - Less data, more parameters to refine individually (H/D ~50% of the atoms)
  - Cancellation effects make X-H species poorly defined in density.
  - Constrained occupancy refinement of H/D sites
  - Data quality (typically low overall and resolution bin completeness)

Rotatable X-H(D)

- 2Fo-Fc nuclear maps: orientation of O-H/D

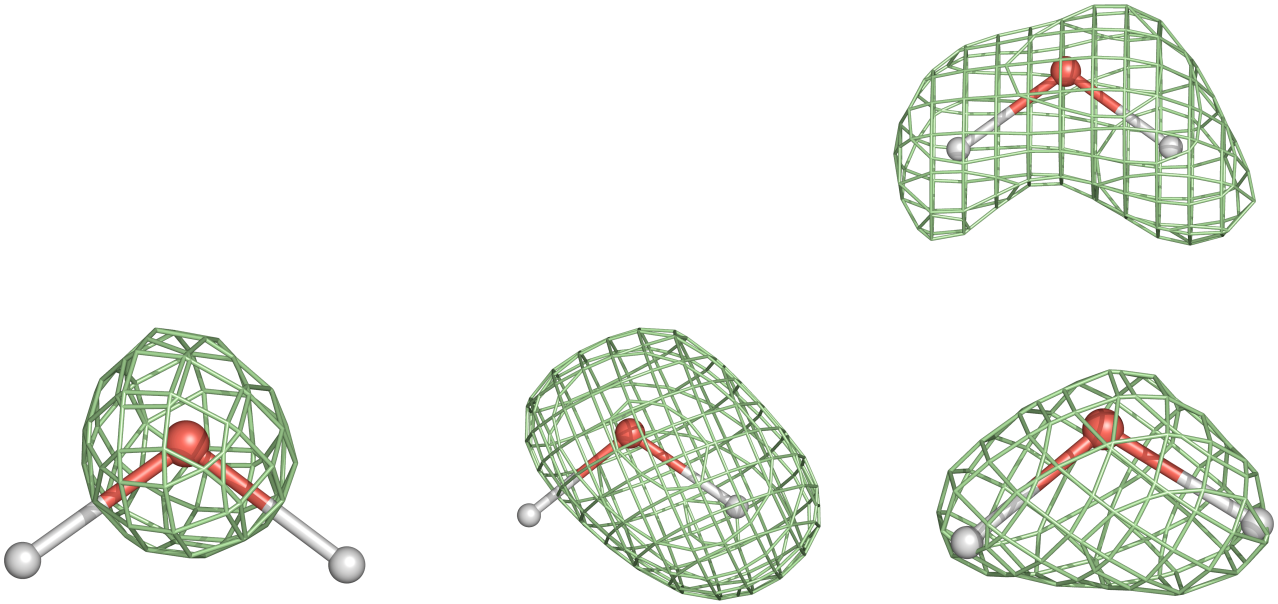


wrong

correct

Water

- Shapes of water molecules in nuclear maps

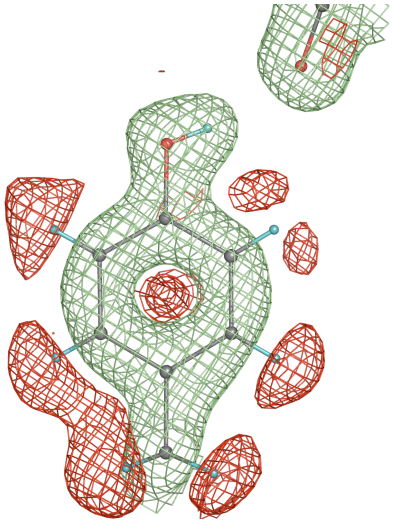


For example: *Chatake et al. (2004). Acta Cryst. D60, 1364-1373*

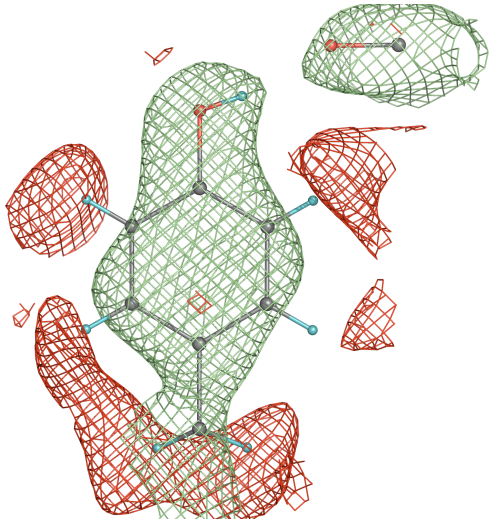
Cancellation effects

- 2Fo-Fc nuclear maps:
  - cancellation effects may perturb maps quite badly (partially deuterated samples)

1IU5 (1.6 Å)

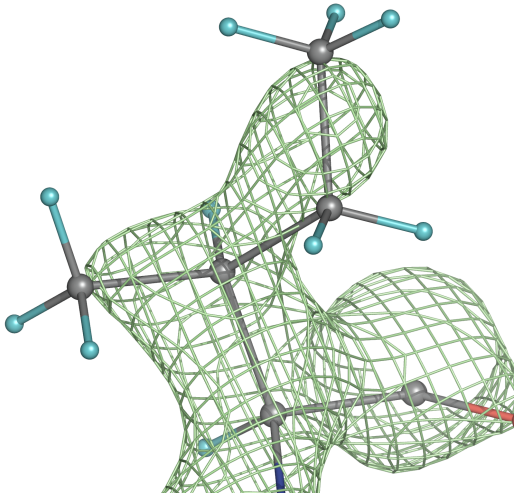


+1.2σ , -1.5σ



+1.2σ , -1.5σ

5RSA (2.0 Å)



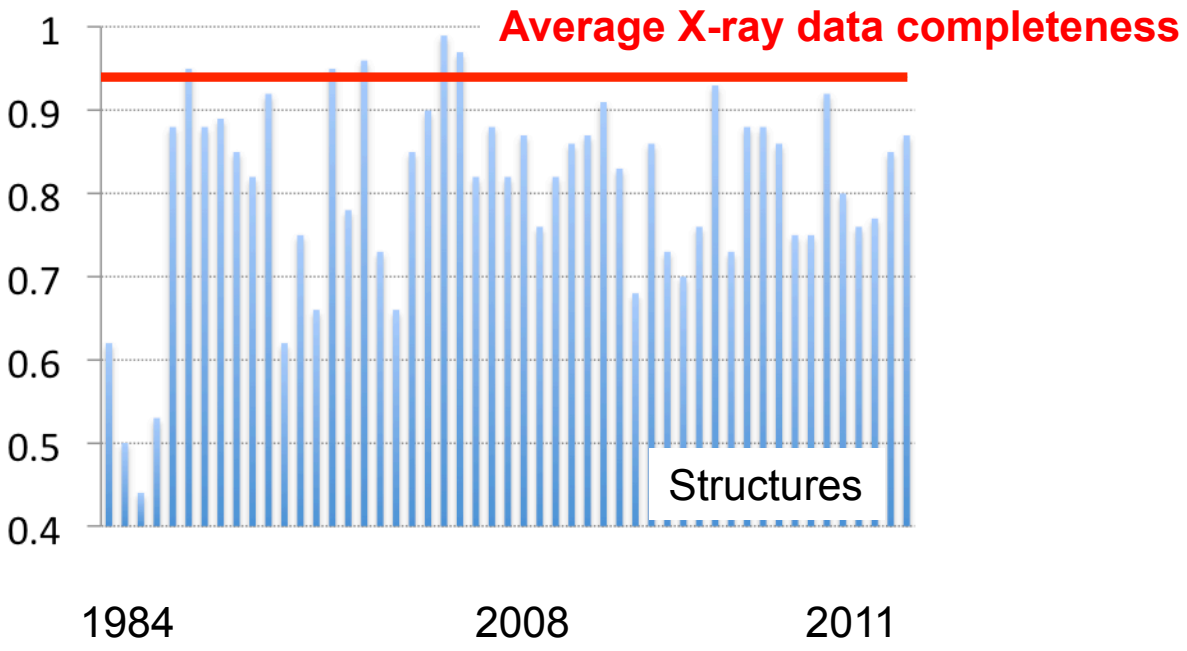
+2.0σ

Data quality (completeness)

- Average data completeness for all structures in the PDB:

  Neutron: 76%
  X-ray: 94%

- Completeness of neutron datasets (sorted by year)



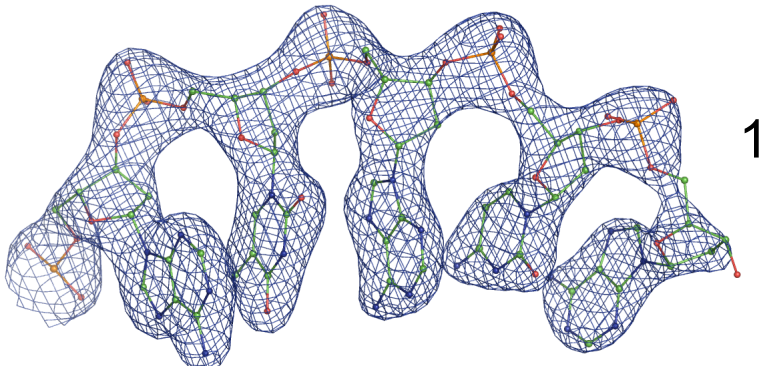
Comparison of F_{calc} maps for 1NH2 structure (1.9Å)

```
Completeness by resolution:
      19.9 - 3.2 0.80
       3.2 – 1.9 1.00
Overall completeness: 0.95
```

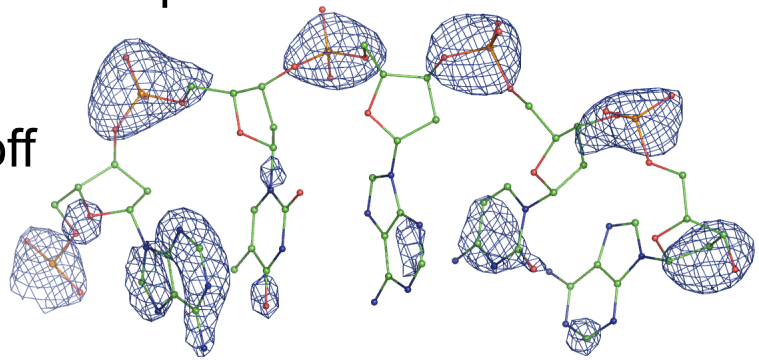
Fcalc maps:

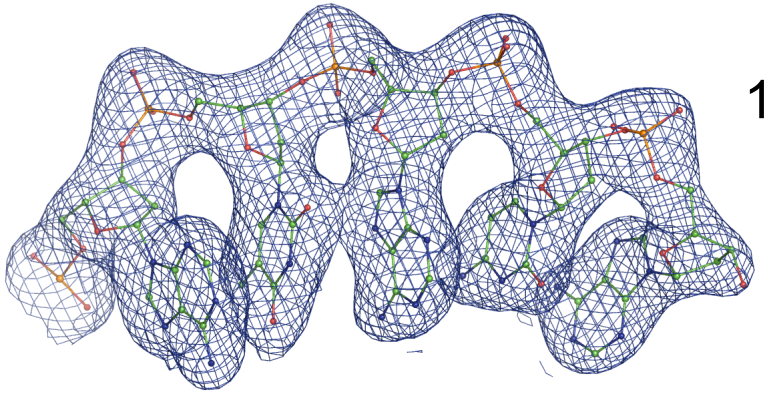
Full set 1.9-19.9Å

Incomplete set 1.9-19.9Å

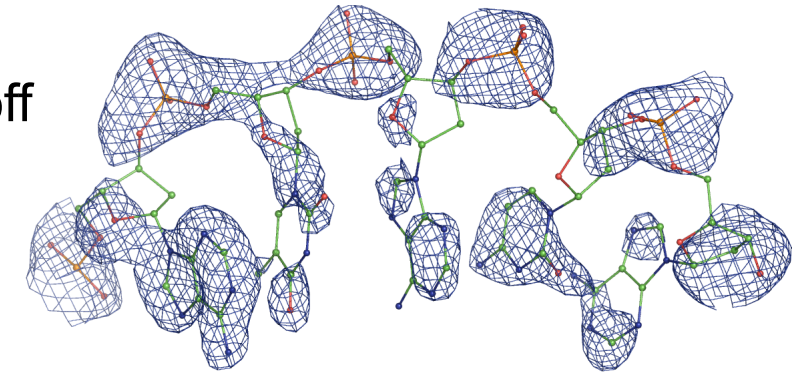


1.5σ map cutoff





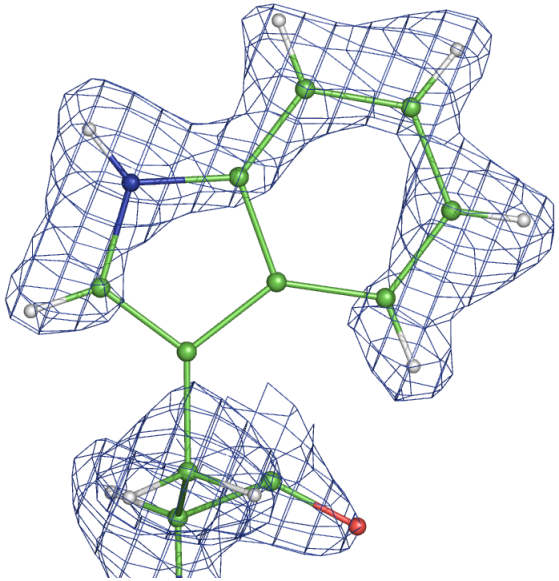
1σ map cutoff

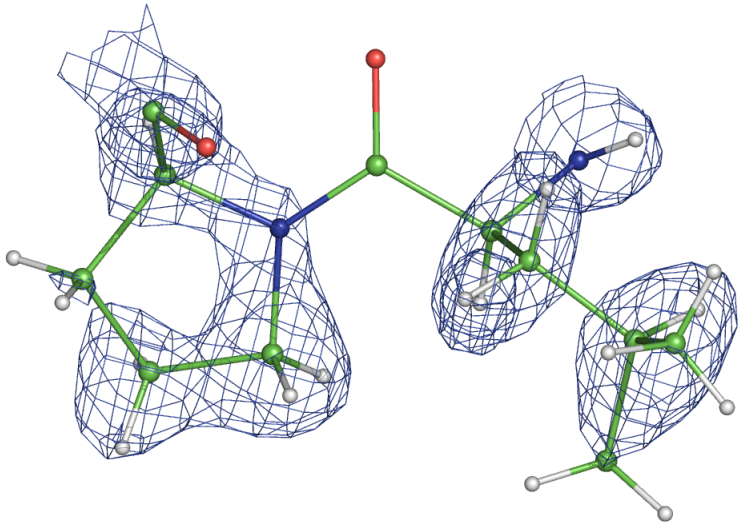


Data incompleteness distorts the maps

Reality (3KYY): 2Fo-Fc maps at 1.5σ

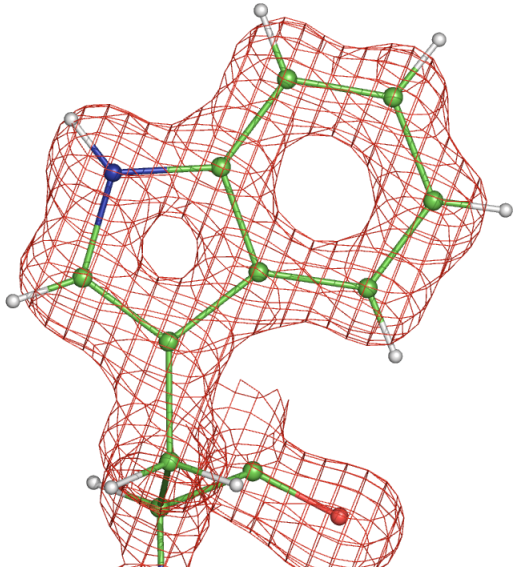
Neutron

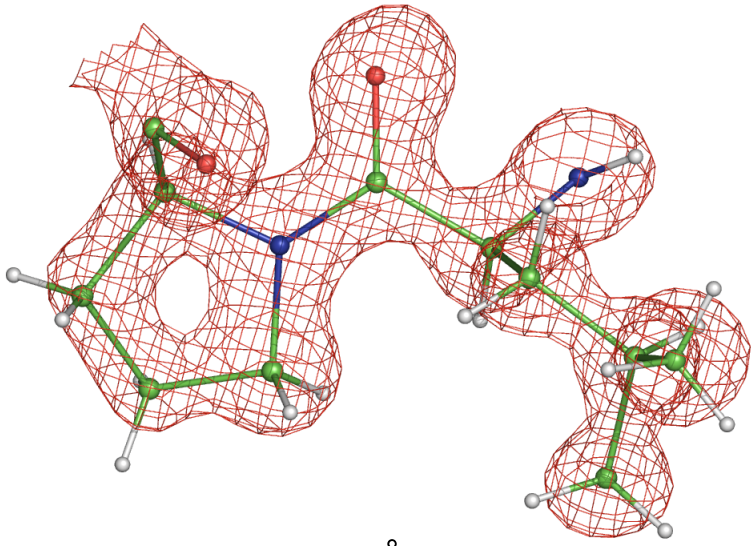




Resolution 1.7 Å, 70% complete

X-ray





Resolution 1.1 Å, 98% complete

Improving neutron macro-molecular crystallography

Macromolecular Neutron Crystallography Consortium (MNC)

Los Alamos National Lab (now at Oakridge National Lab)
Paul Langan, Marat Mustyakimov

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

Funded by NIH

Adams et al., Acta Cryst. 2009, D65, 567-573

What's required

- X-ray structure is available. Need tools to:
  - Prepare starting model for refinement against neutron data
    - Add H, D and H/D atoms
    - Create ligand geometry definitions (if applicable)
  - Refine model using neutron or both X-ray and neutron data sets
    - Handling H/D sites
  - Complete model (build DOD)
  - Validate model and communicate results
- **Approach**: implement these tools in PHENIX

What's PHENIX ?

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Nat Echols, Richard Gildea, Ralf Grosse-Kunstleve, Jeff Headd, Nigel Moriarty, Nicholas Sauter, Peter Zwart

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Cambridge University

Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner

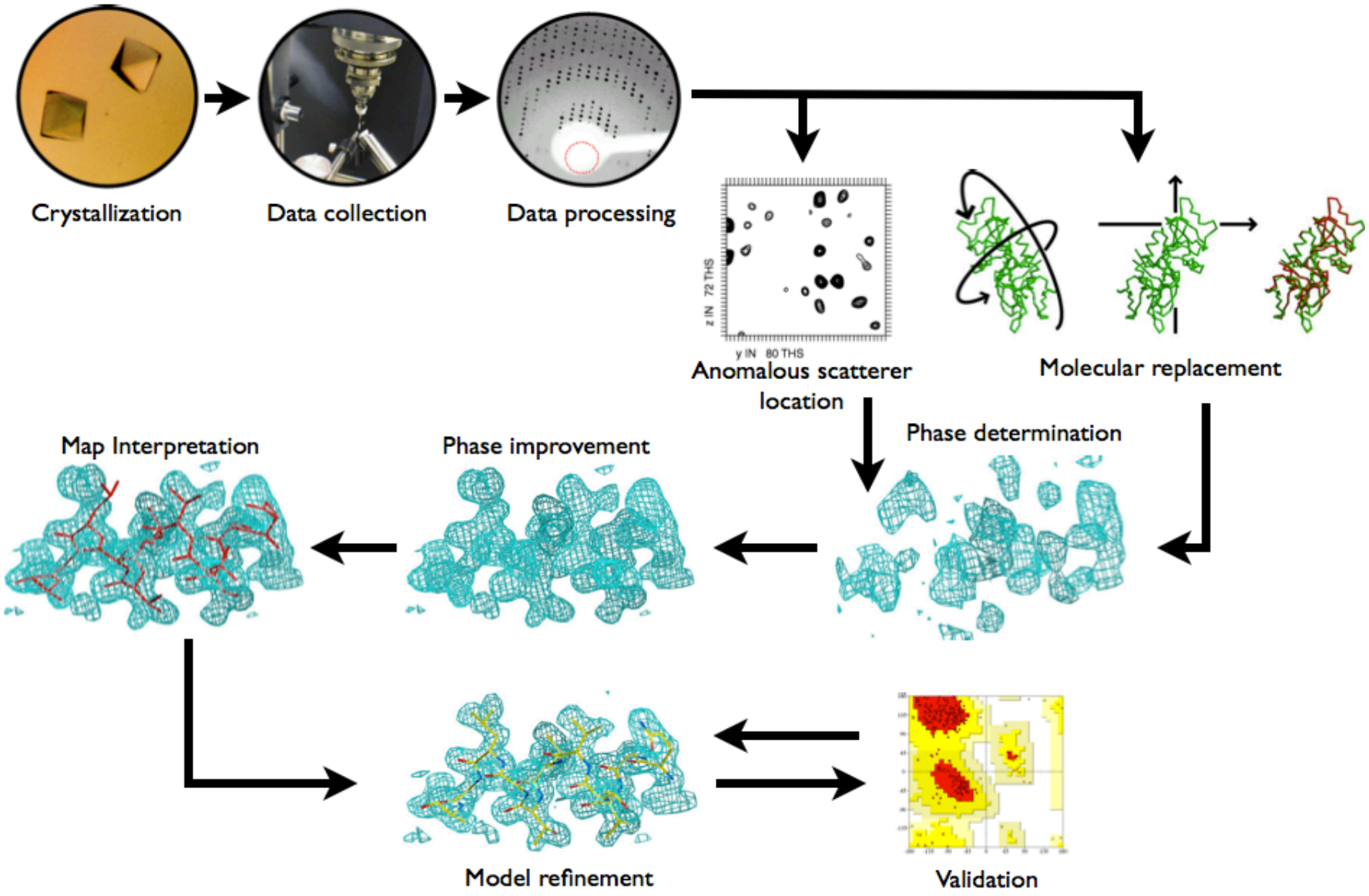
Duke University

Jane & David Richardson, Vincent Chen, Swati Jain, Gary Kapral, Chris Williams, Bryan Arendall, Bradley Hintze

An NIH/NIGMS funded Program Project

Phenix

What's PHENIX for: Macromolecular crystallographic structure solution



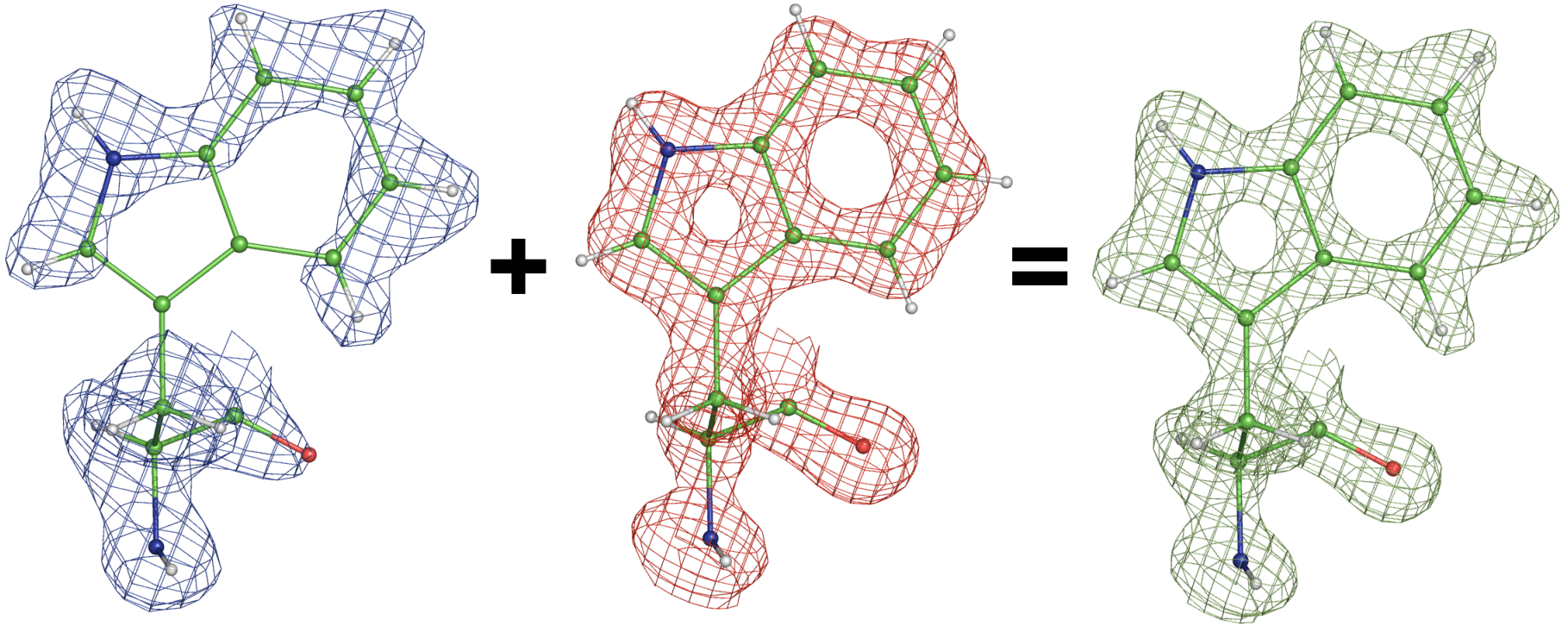
Adams et al., Acta Cryst. D66, 213-221 (2010).

Approaches

- Use all available information: joint X-ray and neutron refinement
- Joint X-ray + neutron refinement:
  - Classical
    - Suggested by Coppens *et al.* (1981), applied to macromolecules by Wlodawer & Sjölin (1982); Wlodawer *et al*. (1982, 1989)
    - $T_{JOINT} = E_{XRAY} * w_{XC} + E_{NEUTRON} * w_{NC} + w_C * E_{GEOM}$
    - Problems: requires comparable datasets collected from isomorphous crystals
  - Future (under development now):
    - Use X-ray model as a reference model in neutron refinement
    - Novel real-space approaches (fit multiple models into combined X+N maps)

Approaches – Joint XN (real-space)

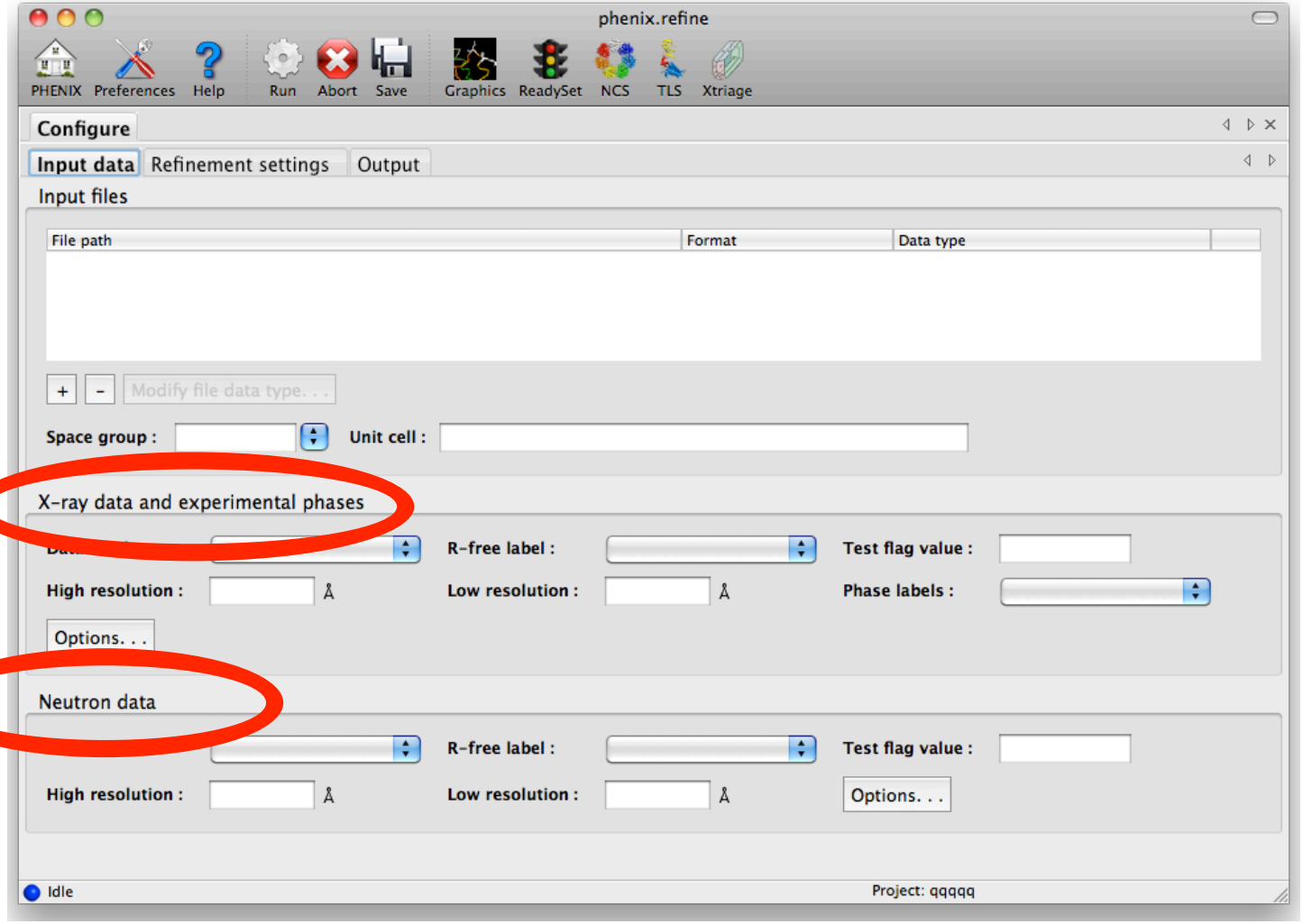
- Combining neutron and X-ray maps



Neutron X-ray Combined

phenix.refine

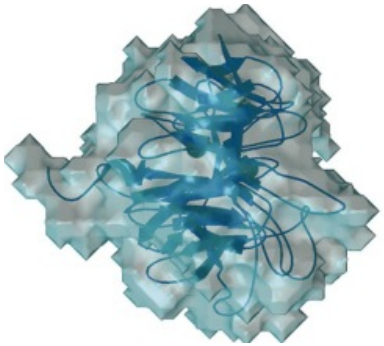
- Comprehensive system for structure refinement using X-ray, neutron or both jointly data sets



phenix.refine: *Afonine et al., Acta Cryst. (2012). D68, 352-367*
GUI: *Echols et al., J. Appl. Cryst. (2012). 45, 581-586*

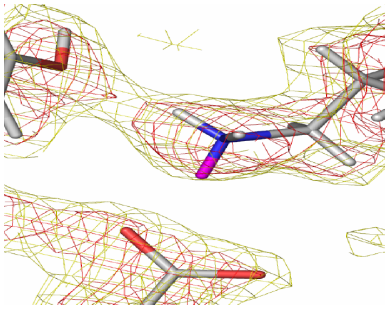
phenix.refine: single program for a very broad range of resolutions

Low



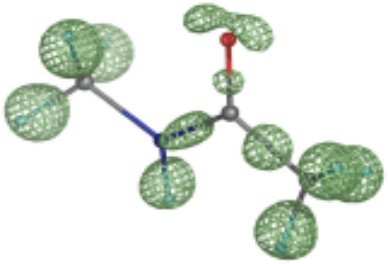
- Group ADP refinement
- Rigid body refinement
- Torsion Angle dynamics
- Reference model
- Ramachandran plot restraints
- Secondary structure restraints

Medium and High



- Restrained/constrained refinement of individual parameters
- Automatic water update

Subatomic



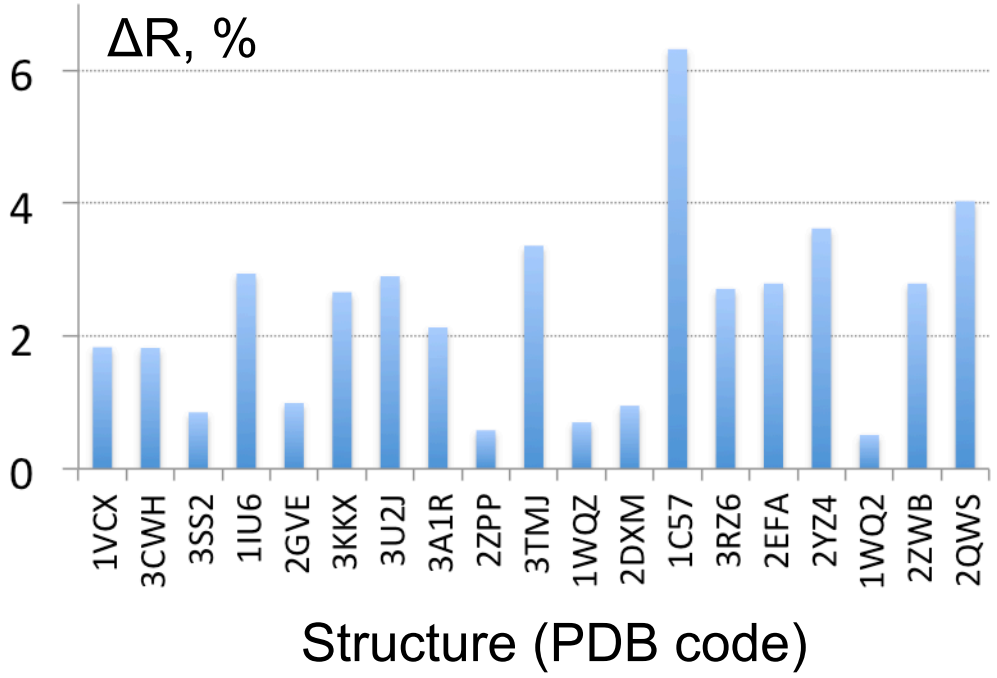
- Bond density model
- Unrestrained refinement
- FFT or direct
- Explicit hydrogens

- Automatic NCS restraints
- Simulated Annealing
- Automatic side chain rotamer fixing
- Occupancies (individual, group, automatic constrains for alternative conformations)
- Various targets: LS, ML, MLHL,…
- Dual (real/reciprocal) space refinement

- TLS refinement with automated TLS groups identification
- Use hydrogens at any resolution
- Refinement with twinned data
- X-ray, Neutron, joint X-ray + Neutron

Revisiting previously deposited structures

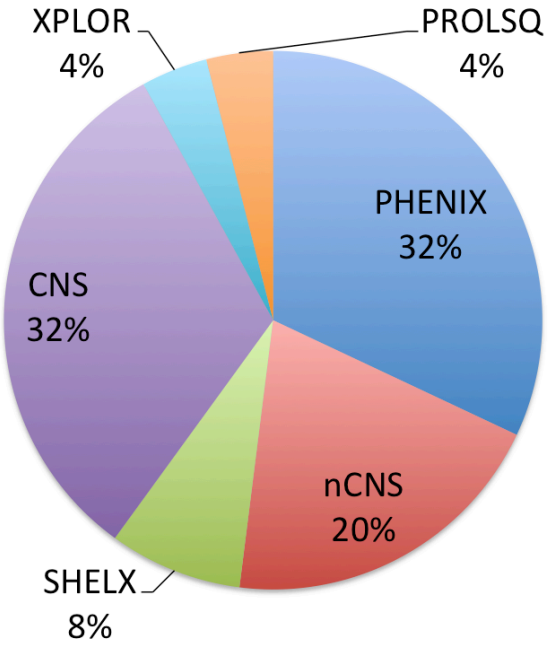
- ΔR = R (**published**) – R(**re-refined with phenix.refine**)



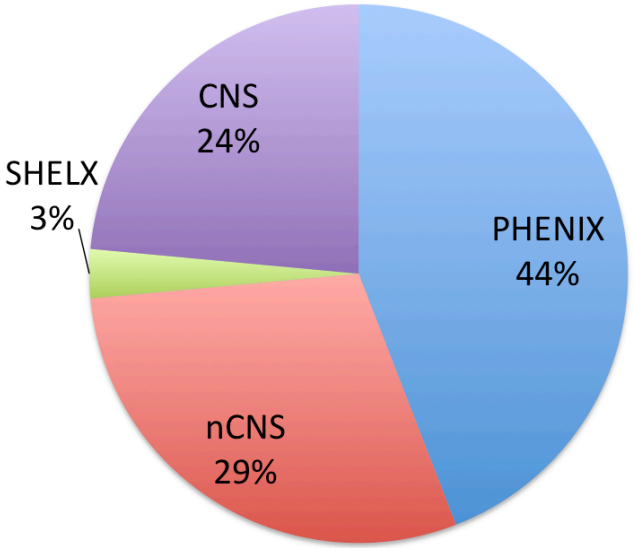
Afonine et al., Acta Cryst. (2010). D66, 1153–1163

Impact

Overall (1984 - 2011): 57 structures



After 2007: 35 structures



Summary

- X-ray crystallography is a leading tool for obtaining protein structures
  - Rarely identifies hydrogen atoms, leaving nearly half the atoms missing
- Hydrogen atoms play a key role in many aspects of biomacromolecules
- Neutron crystallography is complimentary to X-ray by providing experimental information about hydrogen atoms:
  - More complete models
- Limitations to neutron crystallography such as
  - Requirements for large crystals
  - Limited amount of data collection facilities
  - Large data collection times
  - Sample preparation
  - Absence of dedicated software and methodology

  have advanced significantly, making it more accessible for routine use
- Future:
  - Further automate neutron structure refinement and completion
  - Curate previously deposited PDB structures

Acknowledgments

- **Lawrence Berkeley Laboratory**
  - Paul Adams, Nat Echols, Richard Gildea, Jeff Headd, Ralf Grosse-Kunstleve, Nigel Moriarty, Nicholas Sauter, Peter Zwart
- **Los Alamos National Laboratory**
  - Tom Terwilliger, Li-Wei Hung
- **Cambridge University**
  - Randy Read, Airlie McCoy, Laurent Storoni, Gabor Bunkoczi, Robert Oeffner
- **Duke University**
  - Jane Richardson & David Richardson, Ian Davis, Vincent Chen, Jeff Headd, Chris Williams, Bryan Arendall, Laura Murray
- **Collaborators**
  - David Baker and Frank DiMaio
  - Andy Karplus and Dale Tronrud
  - Rhiju Das and Fang Chou
  - Paul Langan and Marat Mustyakimov
  - Tom Burnley and Piet Gros
  - Jaroslaw Kalinowski and Phillip Coppens
- **Others**
  - Garib Murshudov & Alexi Vagin
  - Kevin Cowtan, Paul Emsley, Bernhard Lohkamp
  - Alexandre Urzhumtsev & Vladimir Lunin
  - David Abrahams
  - PHENIX Testers & Users: James Fraser, Herb Klei, Warren Delano, William Scott, Joel Bard, Bob Nolte, Frank von Delft, Scott Classen, Ben Eisenbraun, Phil Evans, Felix Frolow, Christine Gee, Miguel Ortiz-Lombardia, Blaine Mooers, Daniil Prigozhin, Miles Pufall, Edward Snell, Eugene Valkov, Erik Vogan, Andre White, and many more
- **Funding:**
- NIH/NIGMS:
  - *P01GM063210, P50GM062412, P01GM064692, R01GM071939*
- Lawrence Berkeley Laboratory
- PHENIX Industrial Consortium

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix (www.phenix-online.org)