Model Refinement

Pavel Afonine

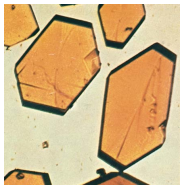
Lawrence Berkeley National Laboratory (LBNL)

September 26th, 2024
BNL

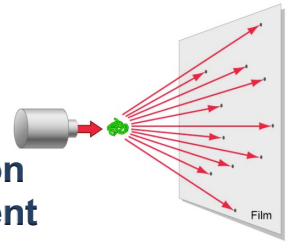
Solving structure by crystallography

Pure sample

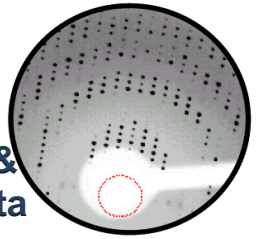
Crystals



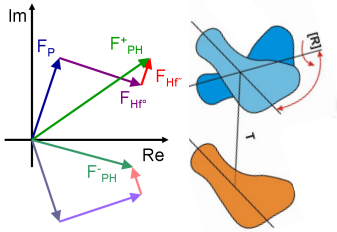
Diffraction experiment



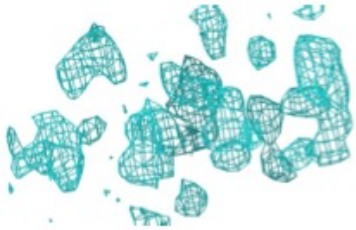
Process & check data



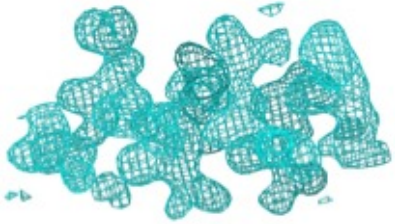
Solve phase problem



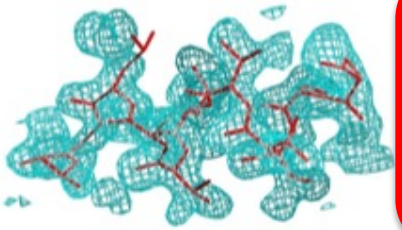
Initial map



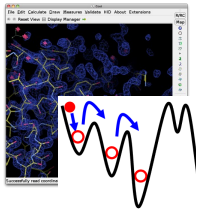
Improved map



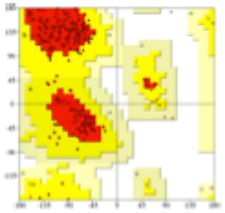
Build model



Refine model



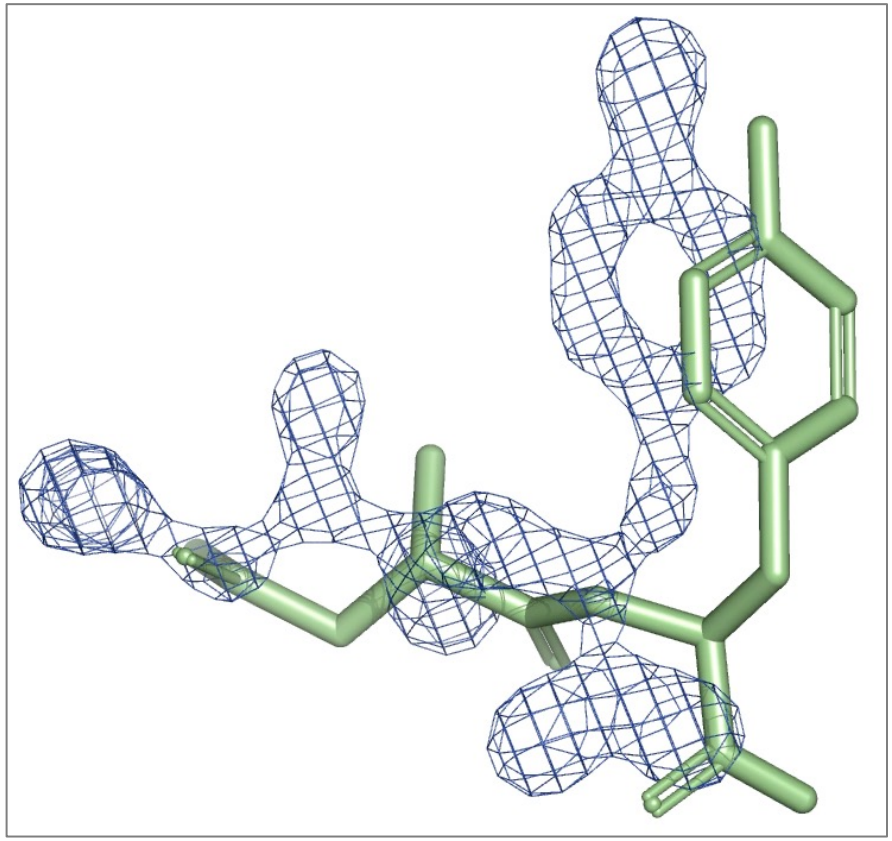
Validate model

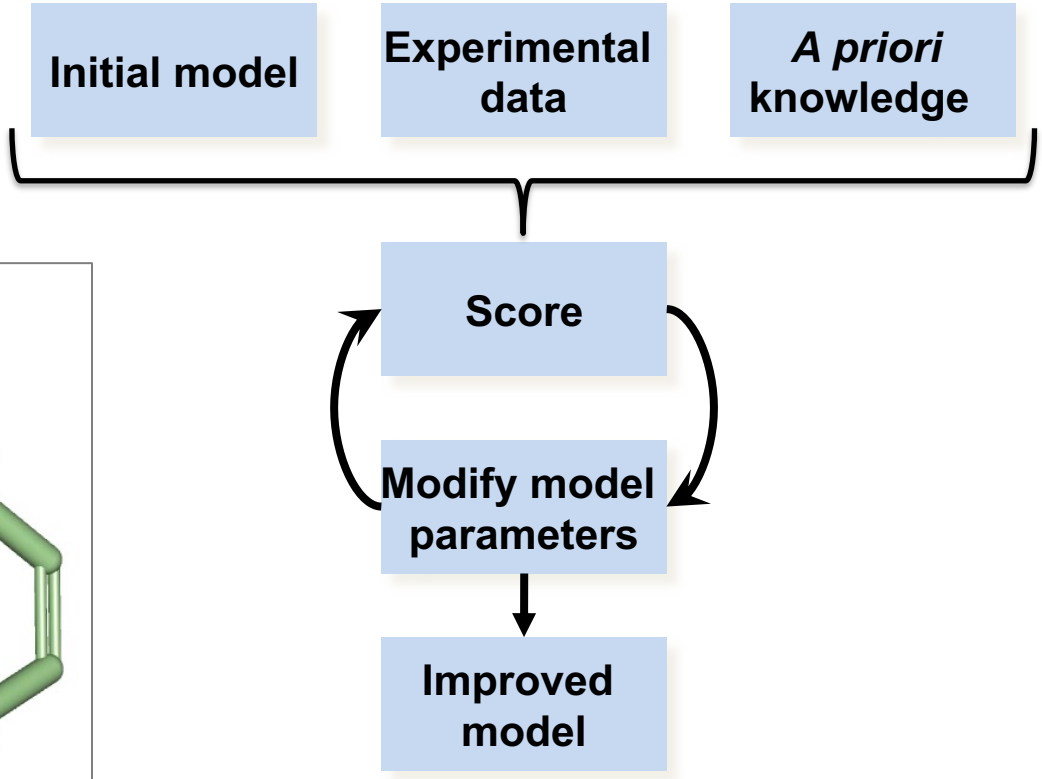


Publish & deposit

- Process is not as ‘linear’ as shown
- Each step has numerous sub-steps
- Crystals may not grow or exhibit pathologies
- Stuck solving phase problem

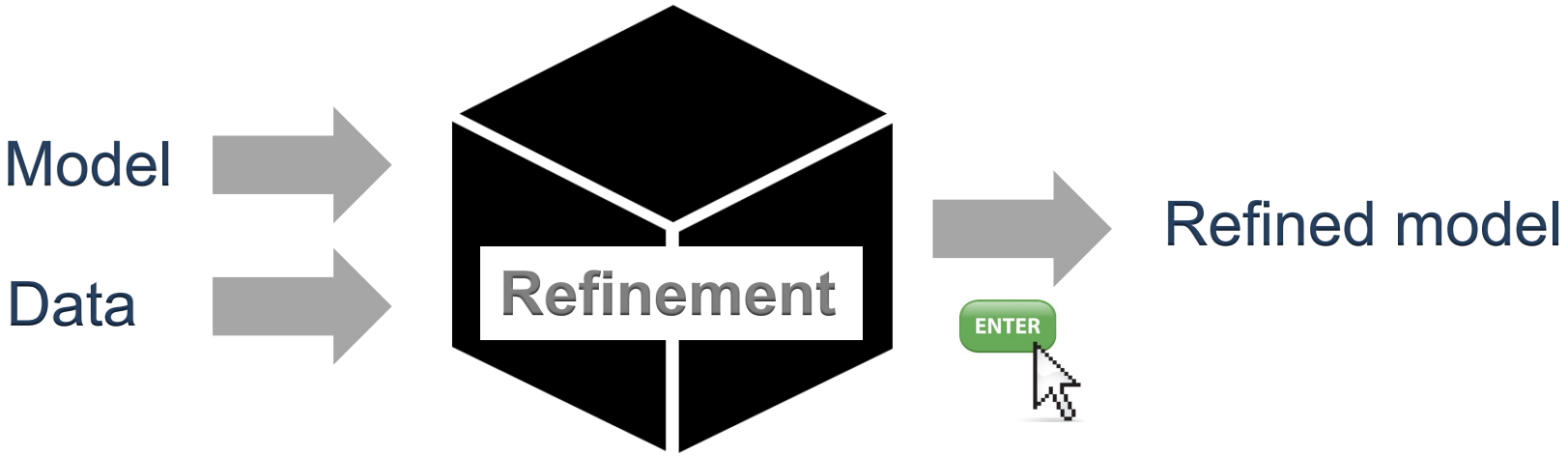
Model refinement





Optimization process of fitting atomic model parameters to experimental data

Model refinement: black box



- Does it always work?
- Is it always as easy as *poor model in, better model out*?

Model refinement: black box

- No. Because:
  - Refinement parameterization isn't easy
  - Default settings suit most common scenario
    - Typical resolution data, model reasonably fits data
  - Less typical situations need customizations
    - Low or high resolution data
    - Incomplete models
    - Final models
    - AlphaFold predicted models
    - Novel ligands

Model refinement: lot of stuff to know…

Reference model?

TLS?

Rotamer fixing?

AltLocs?

Group B vs individual?

Local minima?

ADP?

Clashes?

NCS?

IAS?

tNCS?

Grid search?

Weights?

CDL?

SA?

Minimization?

Rama plot restraints?

f’ & f”?

Hydrogens?

Restraints?

Bulk-Solvent?

Rigid body?

Rama-Z?

Anisotropy?

NQH flips?

SS restraints?

Twinning?

Model refinement: black box

- What to do when the ‘black box’ does not work?
  - **Your decision-making** is needed (and it is not always easy!)

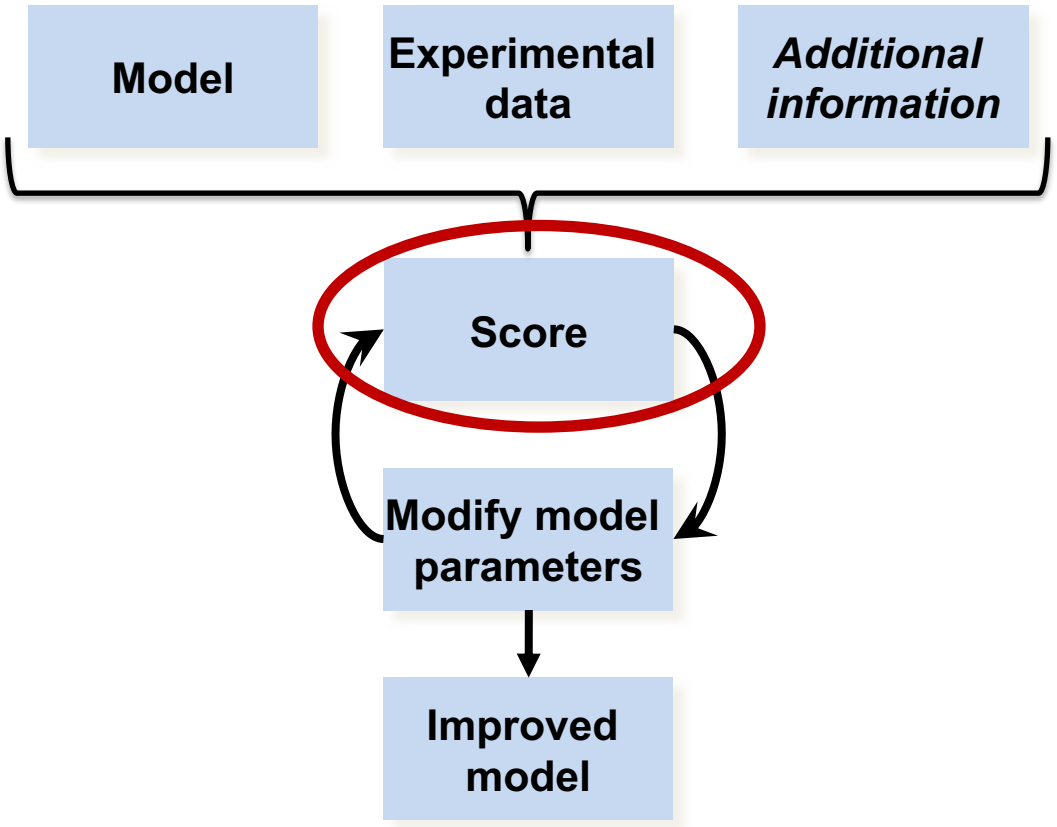
How you know…

- … refinement worked ?
- … you did it correctly ?
- … the model you got is good enough to publish ?

- Do validation!

Standard validation protocols are designed to answer these questions

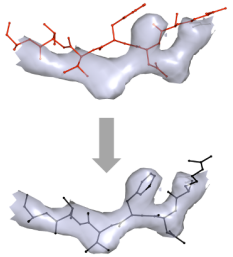
Refinement target function (score)

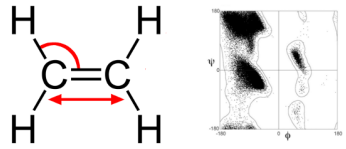


Restraints and data resolution

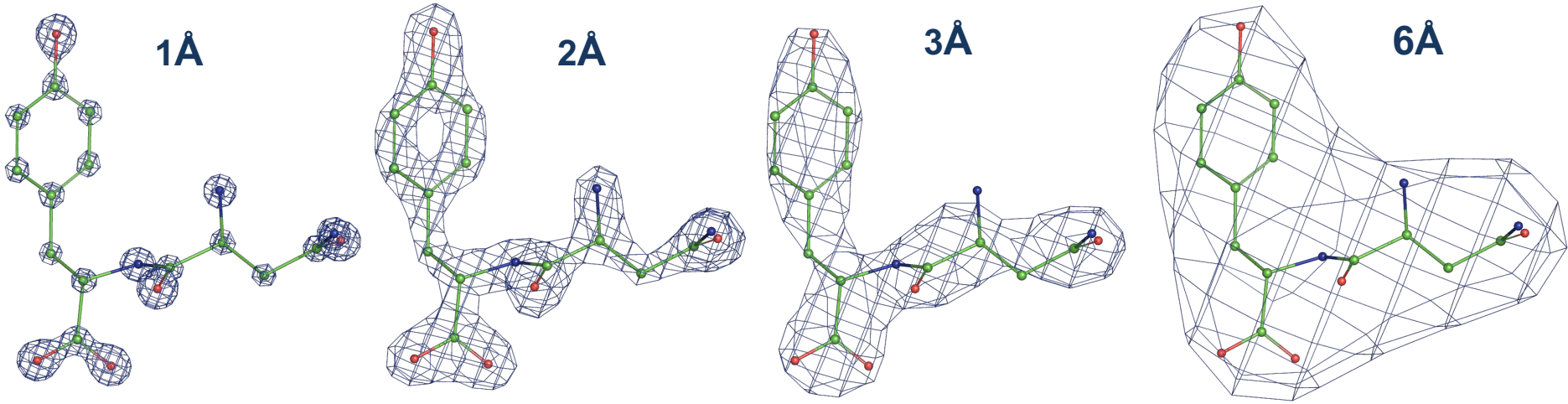
$T = T_{DATA} + w * T_{RESTRAINTS}$

Optimize consensus between model-to-data fit and... common sense



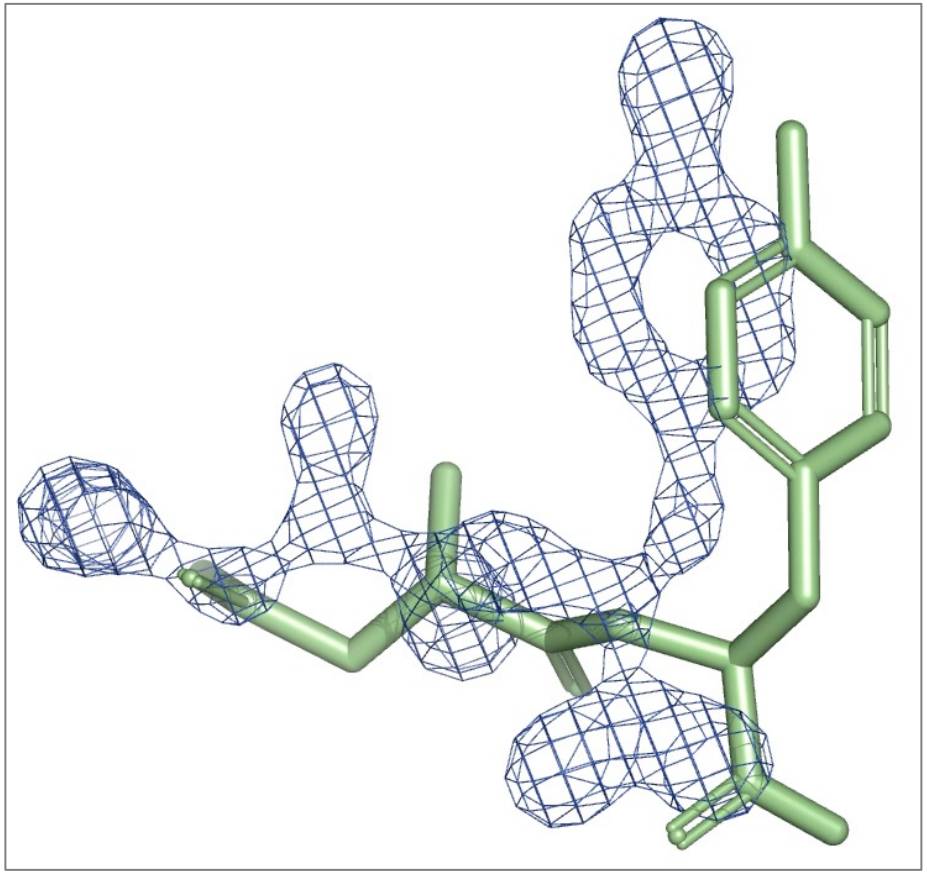


Bonds, angles, planes, torsions, chirality, non-bonded repulsion

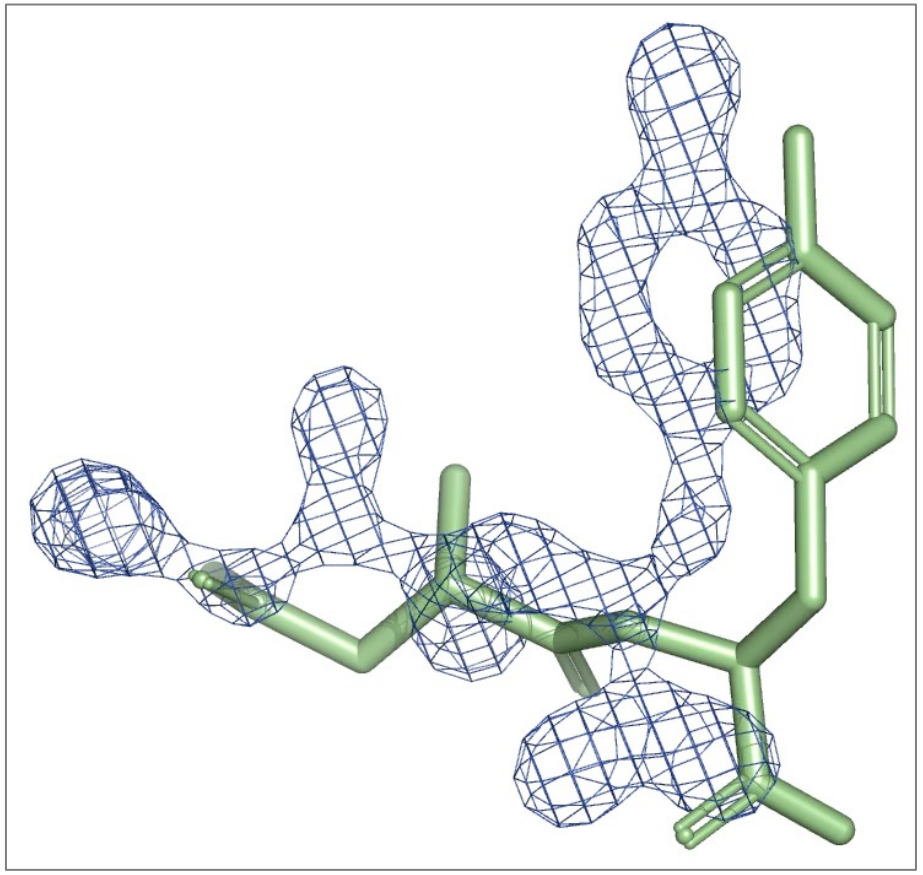


Model refinement with *vs* no restraints

$$T = T_{DATA} + w * T_{RESTRAINTS}$$



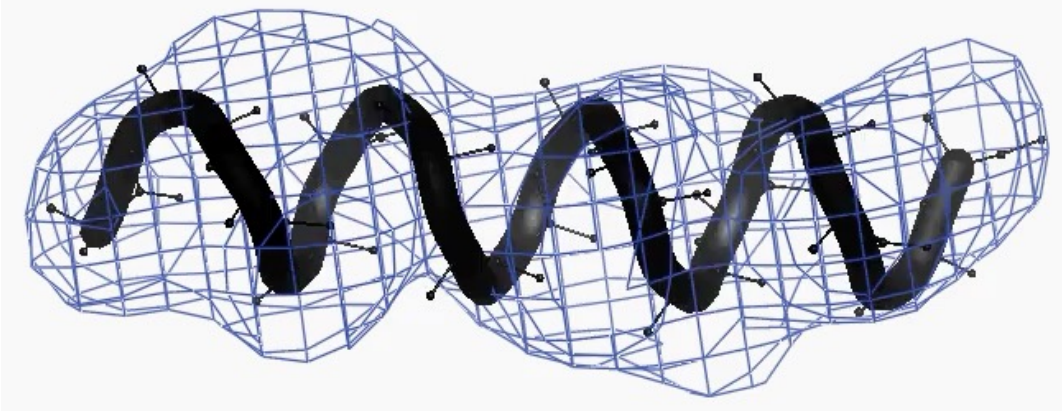
Using restraints



No restraints

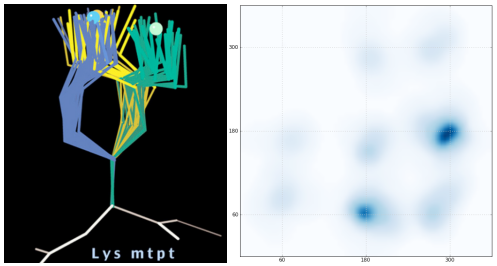
Model refinement with insufficient restraints

- Refinement of a perfect α-helix into low-res map
  - Using simplistic (standard) restraints on covalent geometry
    - Model geometry deteriorates as result of refinement

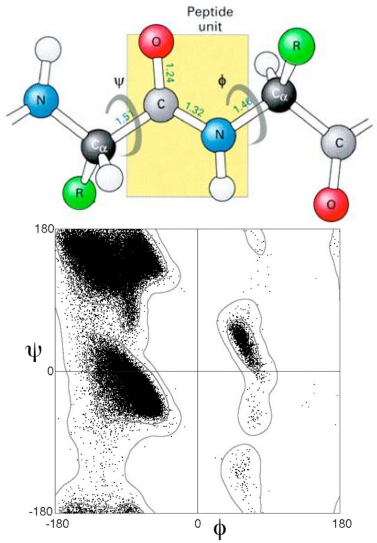


More restraints for low resolution

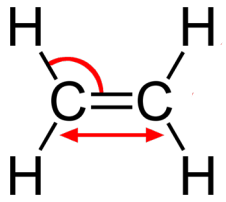
Side chain distributions



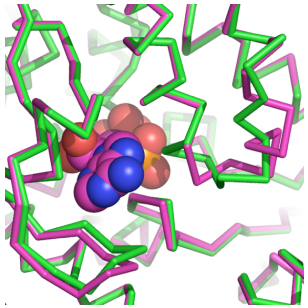
Main chain distributions



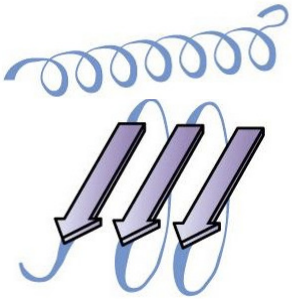
Covalent geometry



Internal symmetry (NCS)

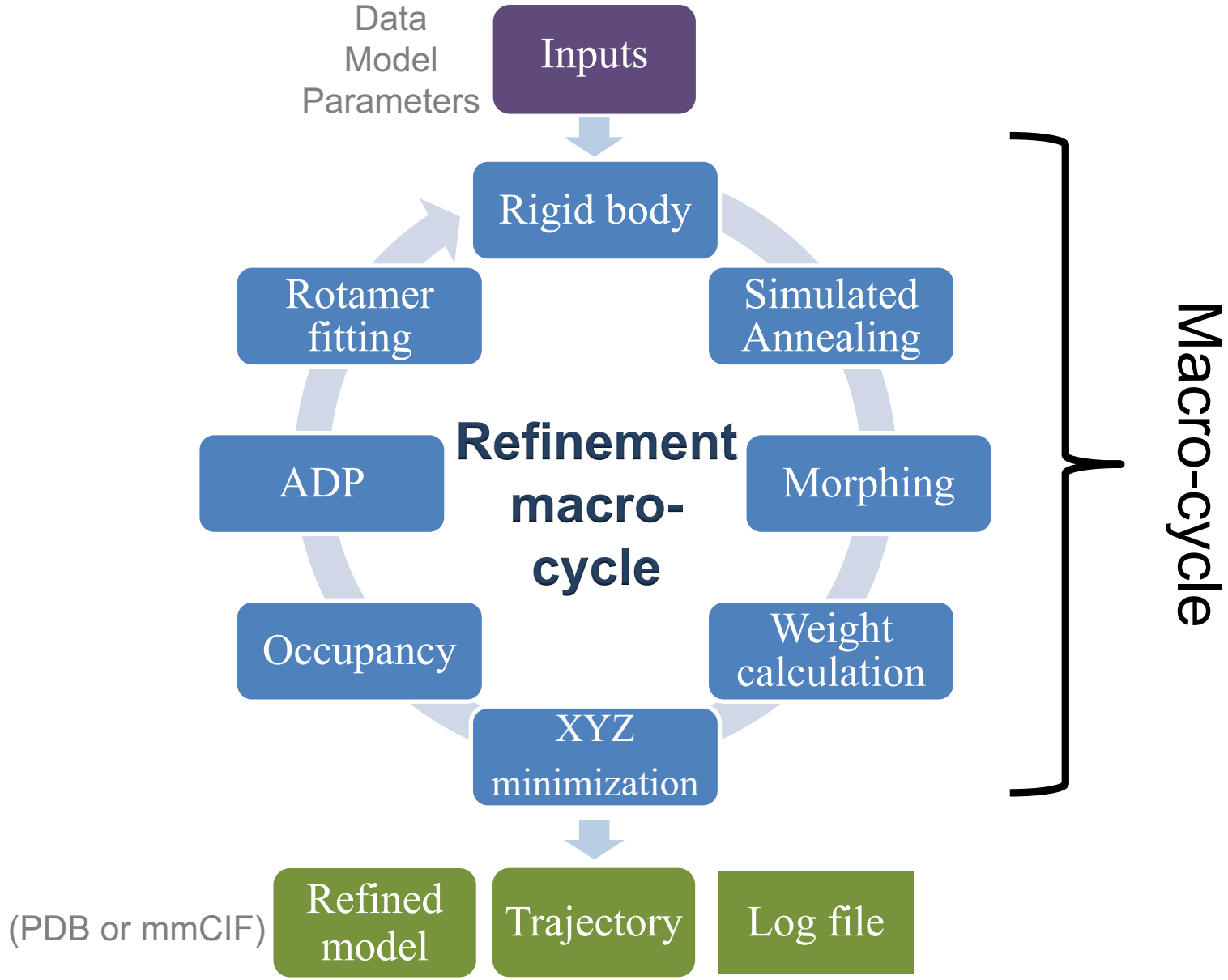


Similar (homologous) structures (reference model restraints)



Secondary structure

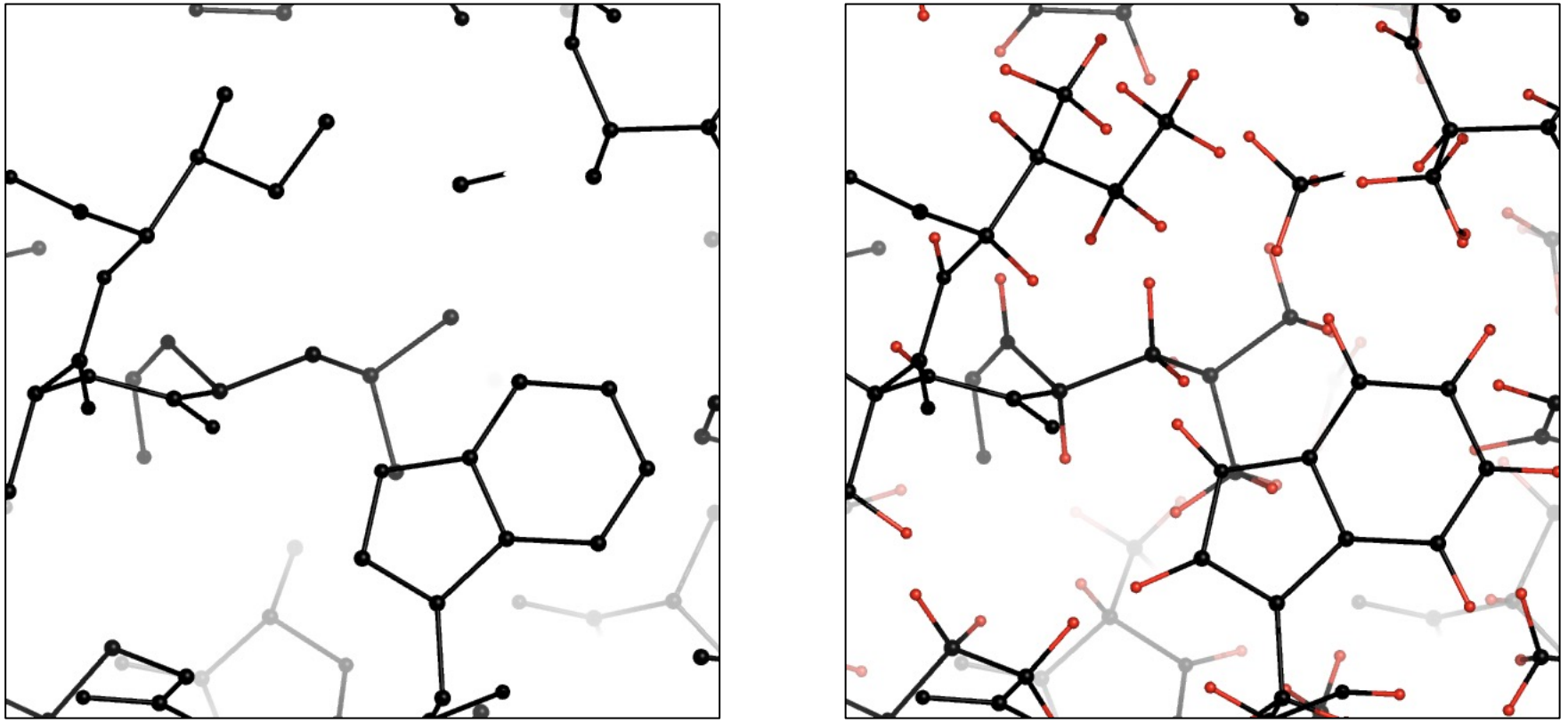
Refinement protocol



Refinement: practical considerations

Use Hydrogen atoms

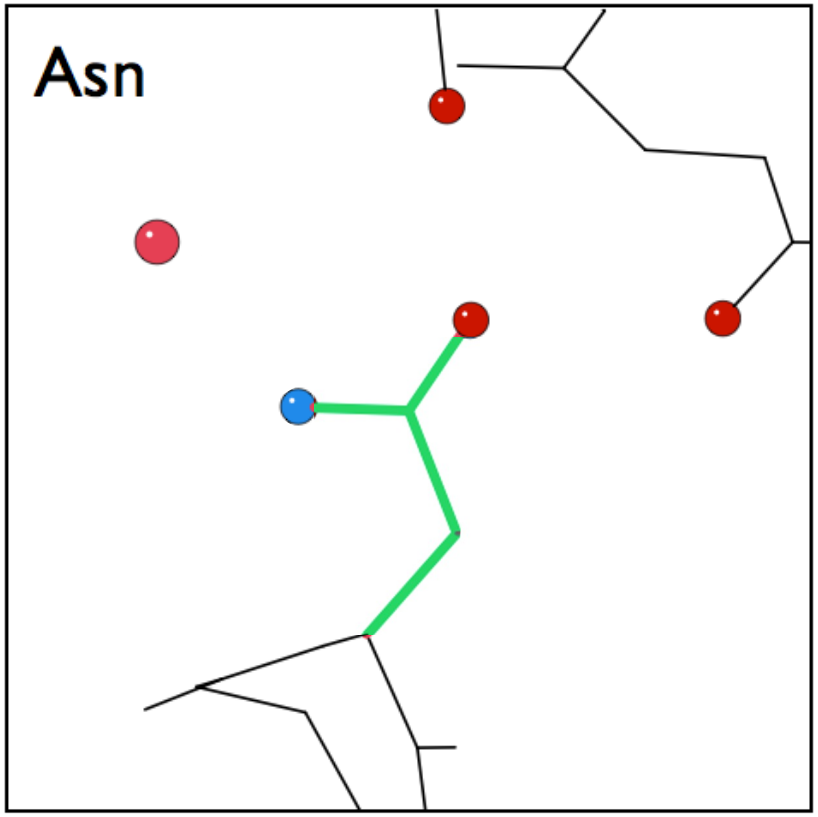
- Half of the atoms in a protein molecule
- Make most interatomic contacts
- Add to model towards the end, data resolution does not matter
- Once added, do not remove before the PDB deposition
- H do contribute to R-factors (expect 0.1-2% drop in R)

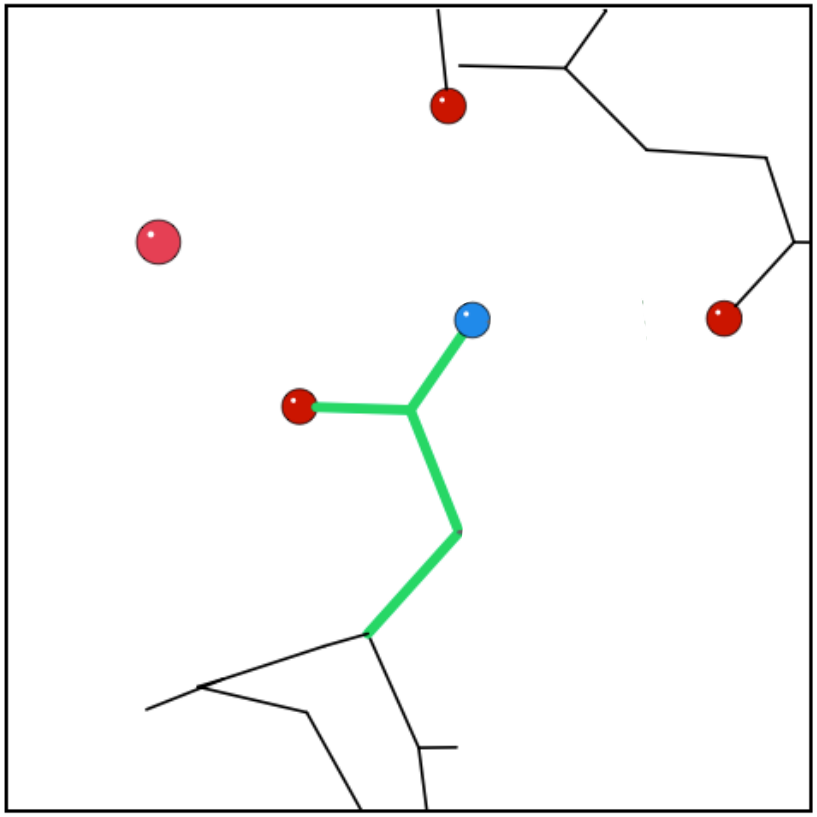


A structure without (left) and with (right) hydrogen atoms

Use Hydrogen atoms

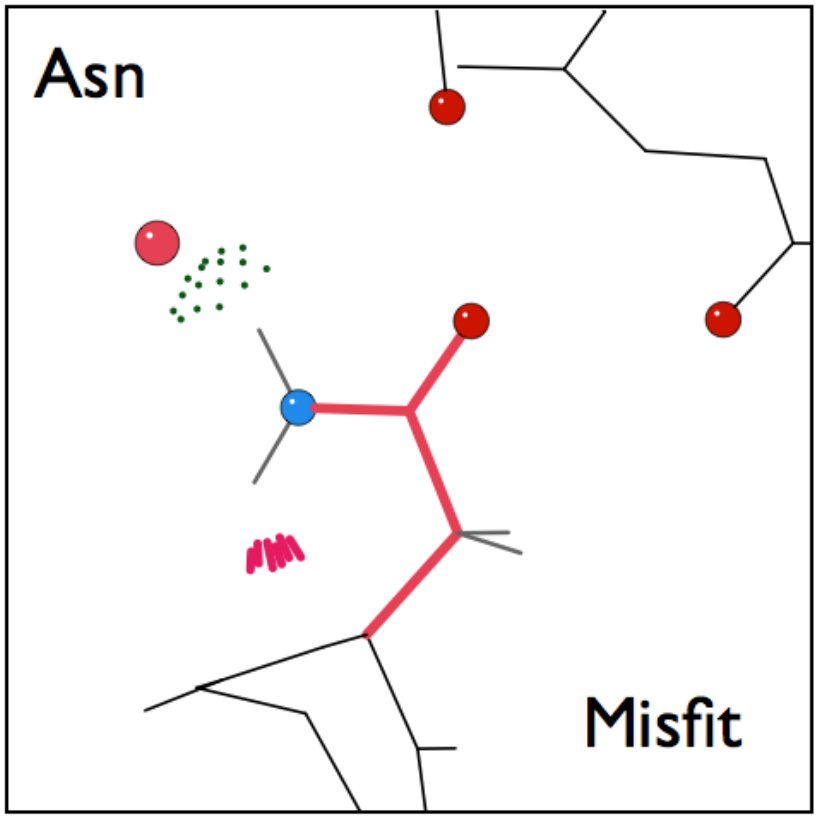
- N/Q/H flips (asparagine/glutamine/histidine)
  - Based on clash analysis
  - Requires H present

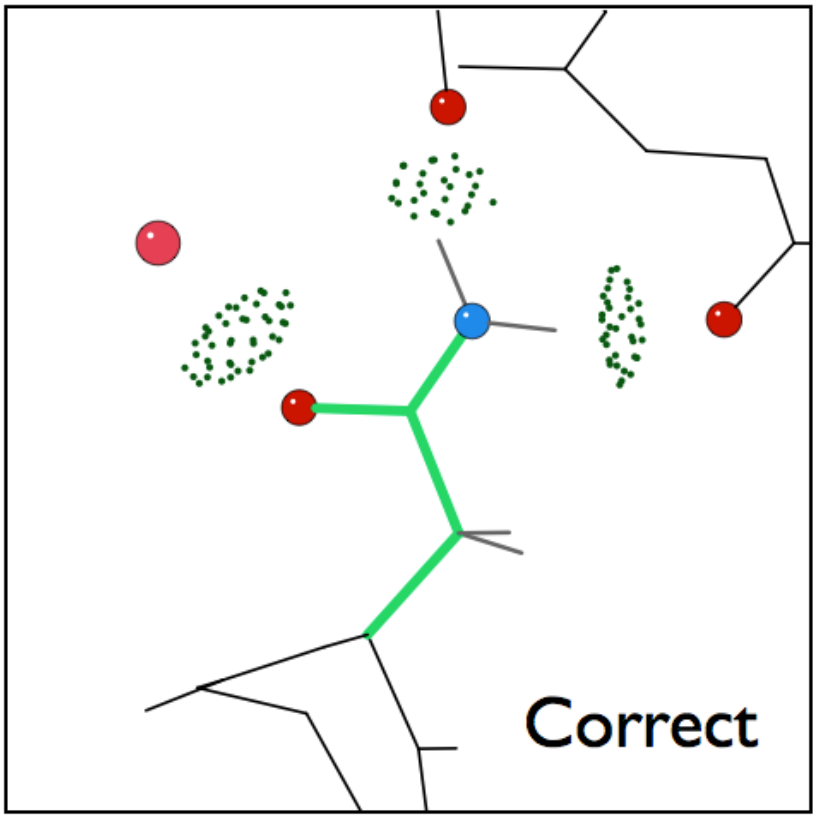




Use Hydrogen atoms

- N/Q/H flips
  - Based on clash analysis
  - Requires H present

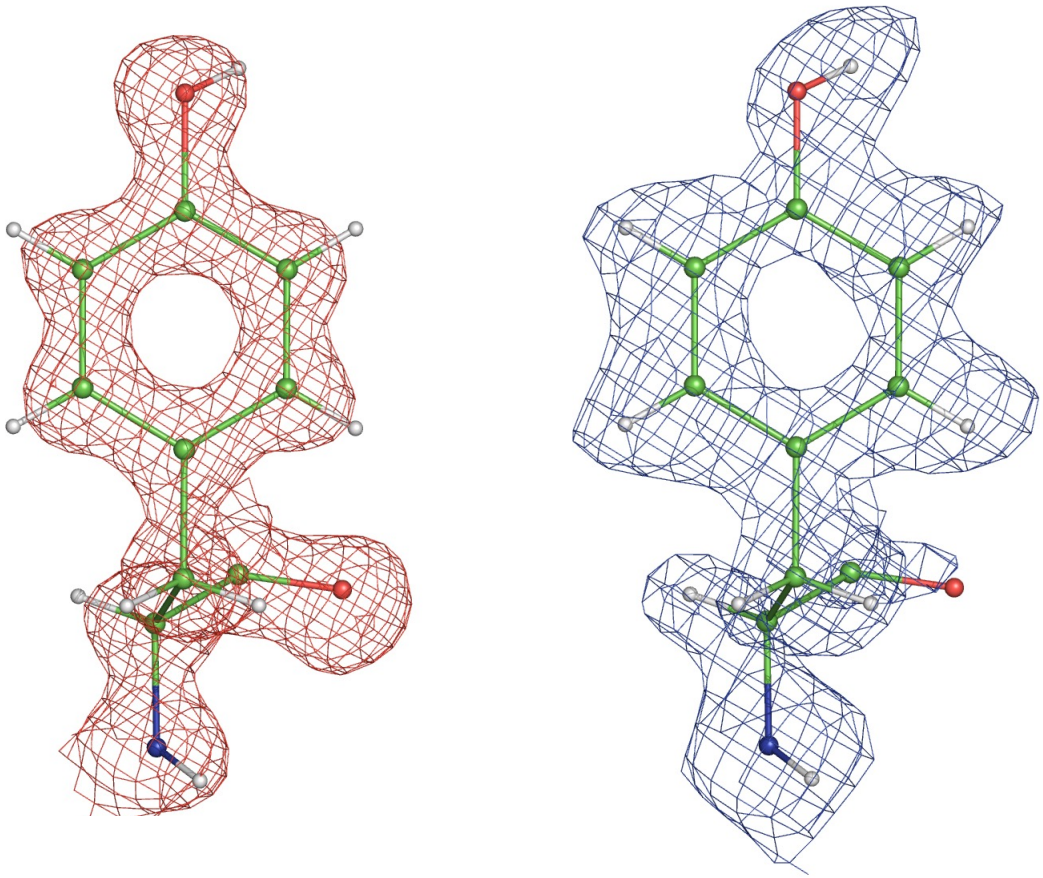




Hydrogens are best revealed by neutrons!

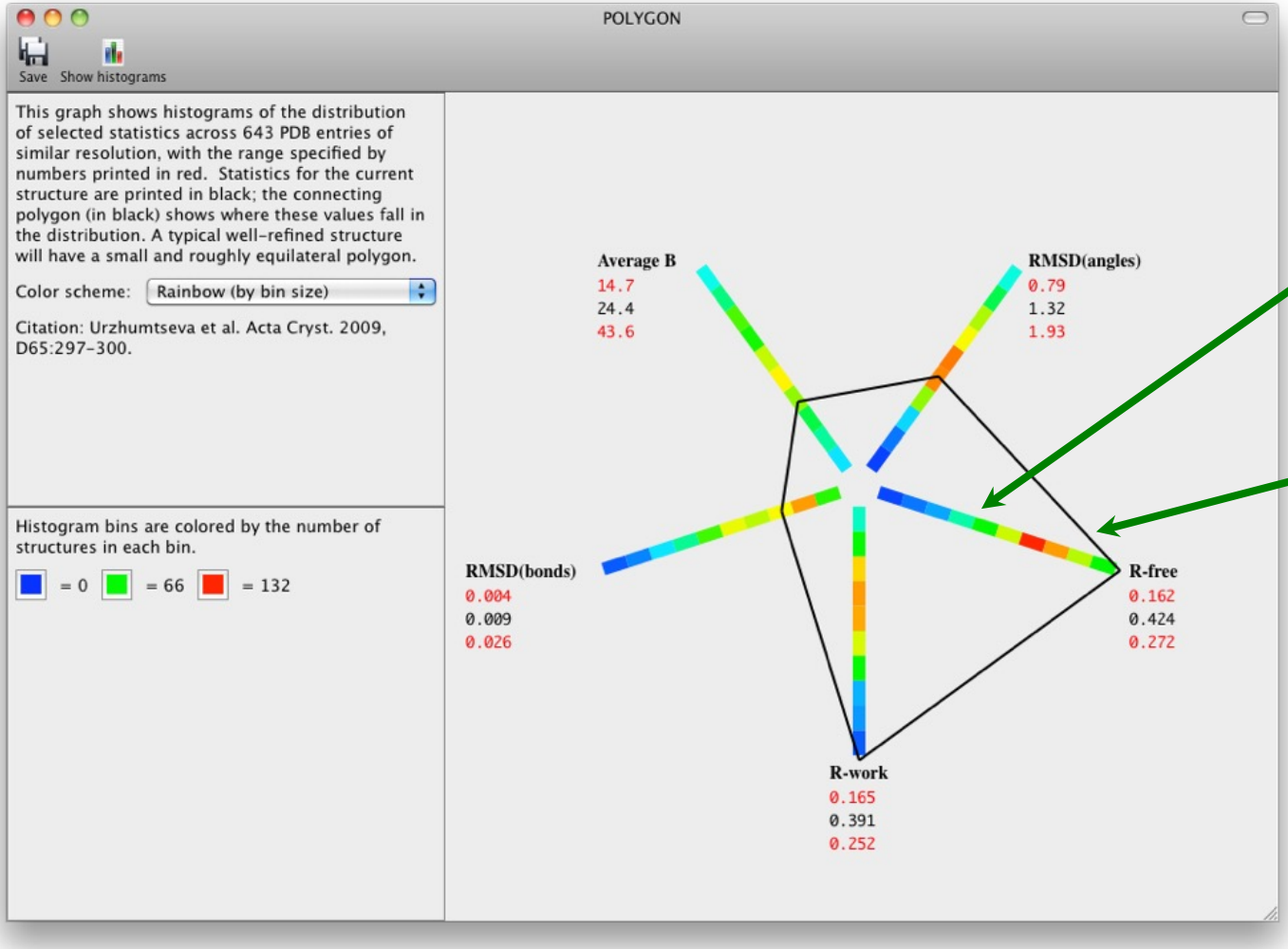
Nuclear density maps show H (D) at typical macromolecular resolutions (~2Å)

X-ray (1.1 Å) **Neutron (1.7 Å)**



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

Know when to stop



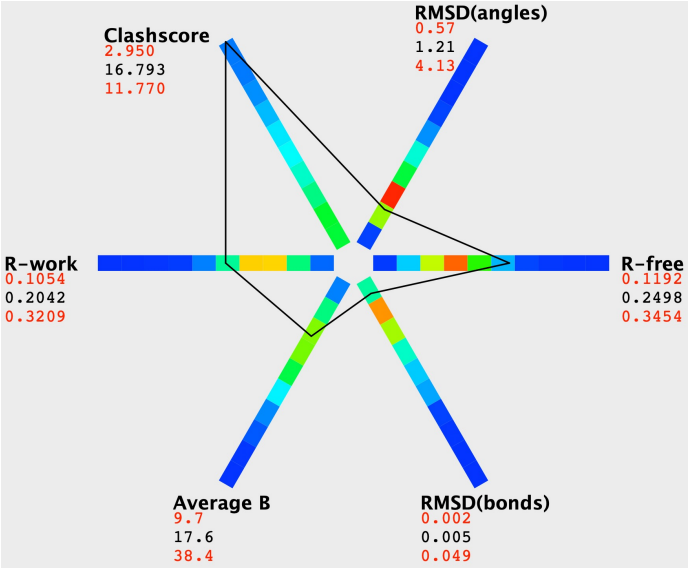
Colored bars are histograms showing distribution of values for structures at similar resolution

The black polygon shows where the statistics for the user's structure fall in each histogram

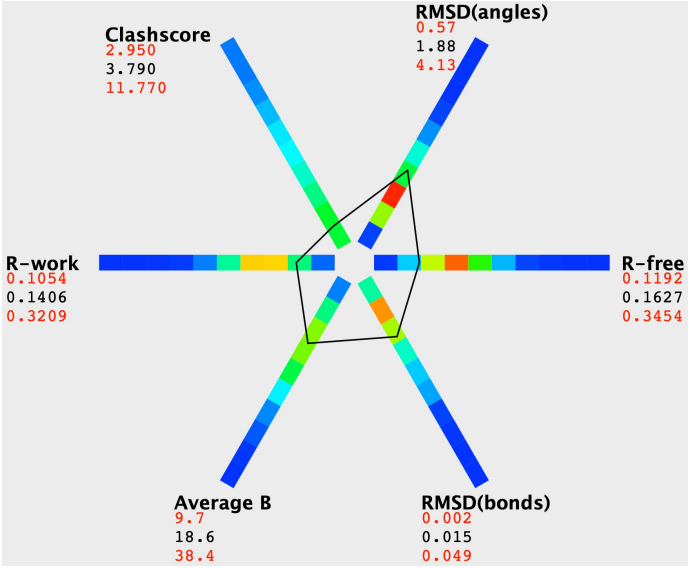
Crystallographic model quality at a glance.
L.Urzhumtseva, P.V.Afonine, P.D.Adams & A.Urzhumtsev. *Acta Cryst*. **D**65, 297-300 (2009)

Know when to stop

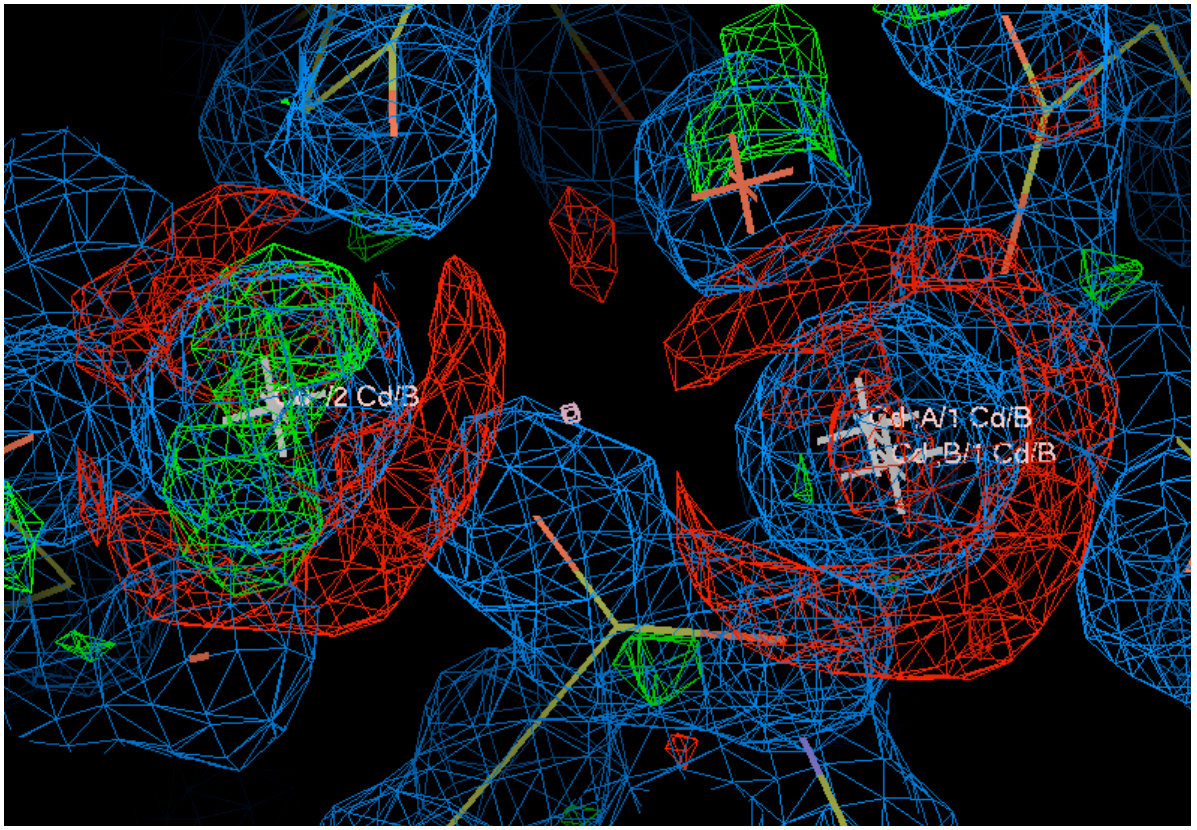
Clearly there are problems



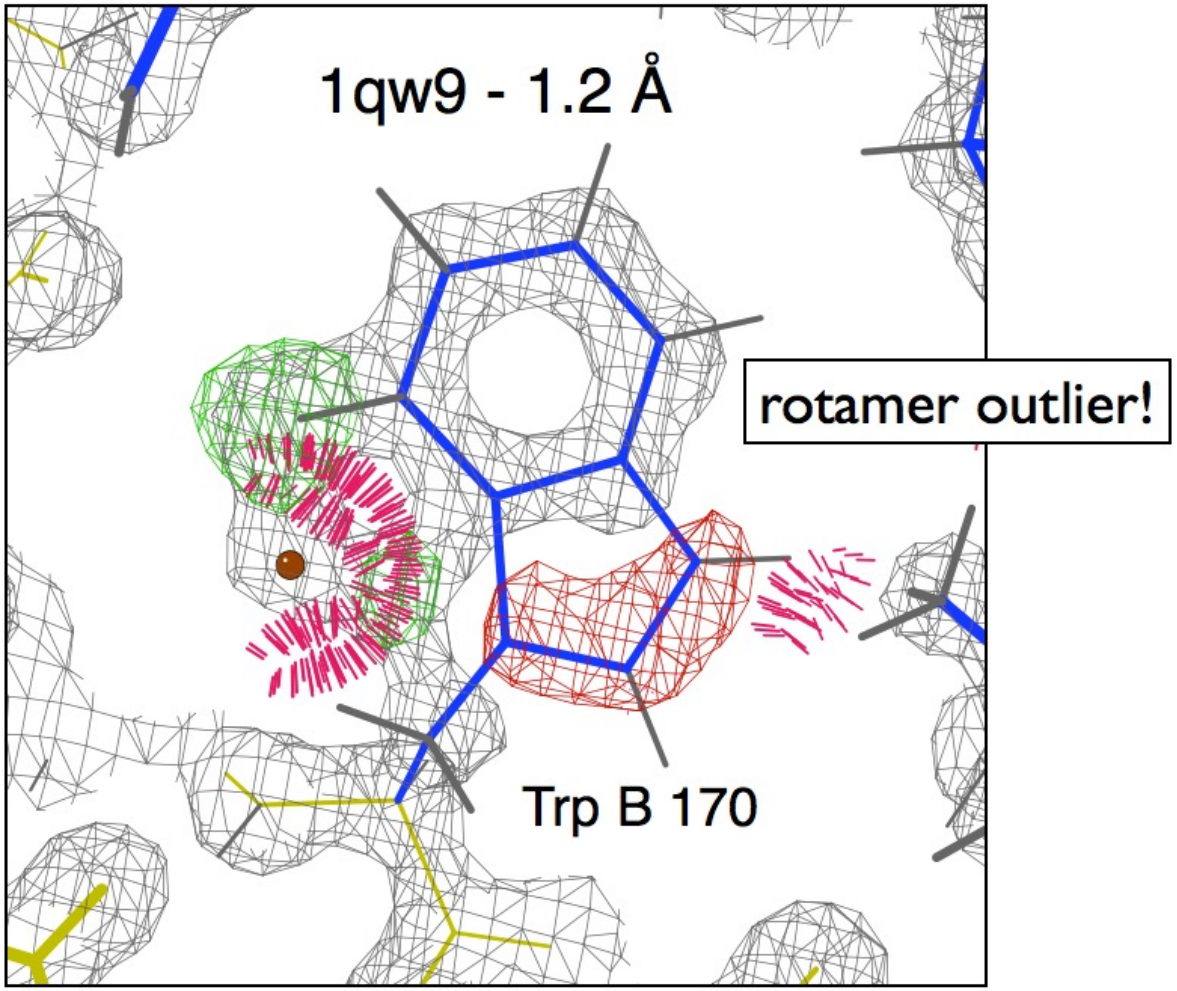
Likely overall good model



Map and model errors



Not all modeling errors can be fixed by refinement



Low resolution (3Å or worse)

- Use:
  - Ramachandran plot restraints
  - Secondary structure restraints
  - Reference model restraints (if quality homology model is available)
  - NCS (restraints or constraints)

Aggressive optimization methods

- Simulated annealing (SA)
- Model morphing
  - Only use if model has gross errors (correction requires large movements)
  - Do not use if model is relatively good and only needs small corrections