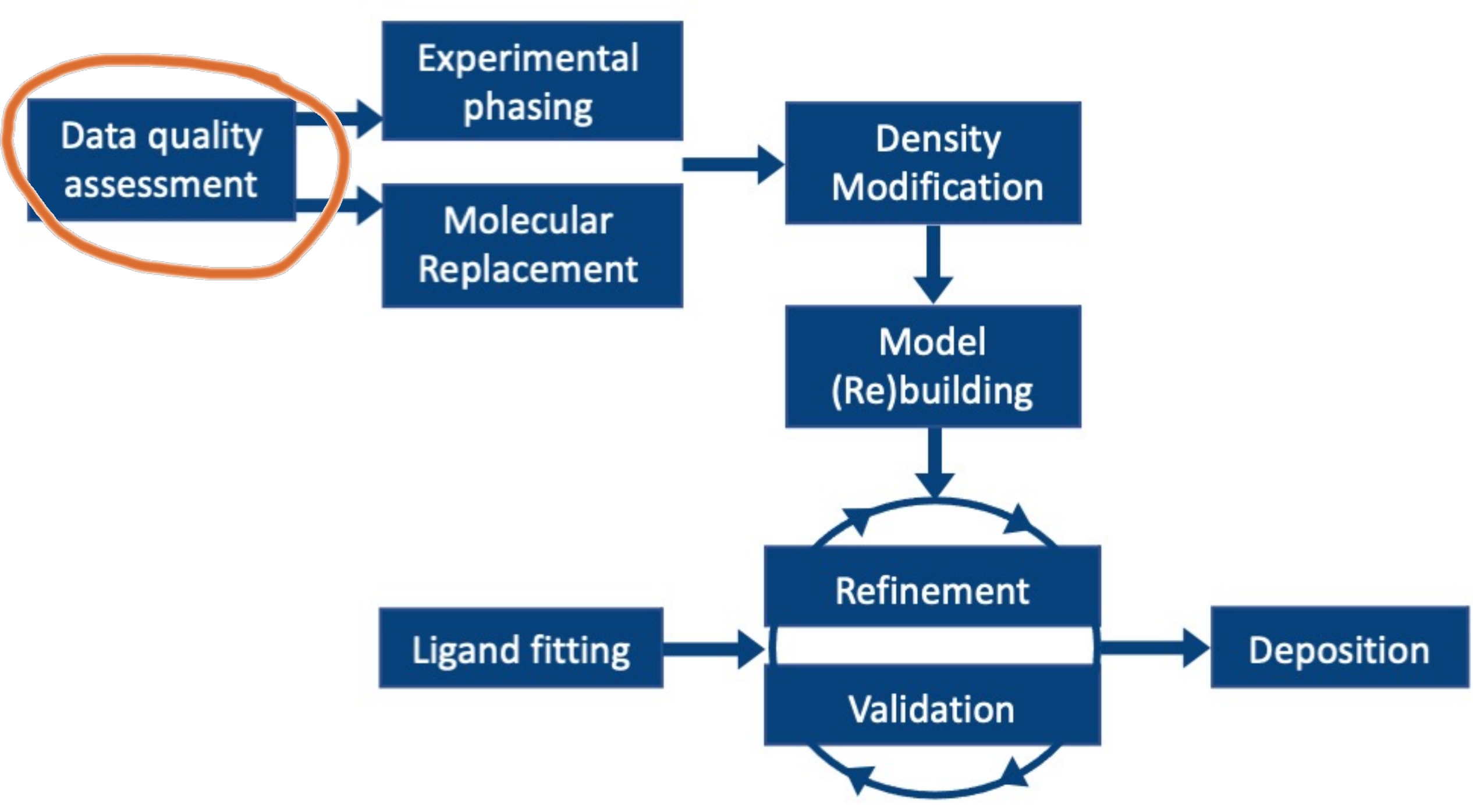
CBMS workbench (virtual), Oct 13 2021

Tools for crystallography

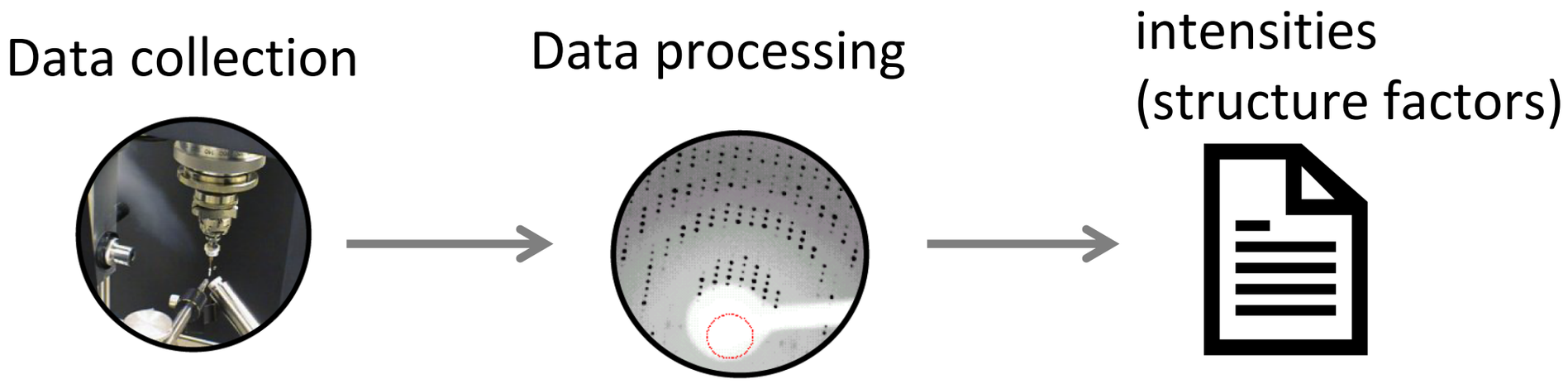
Dorothee Liebschner

Lawrence Berkeley Laboratory

Steps in crystallography



Data Quality Assessment



Macromolecular crystals are prone to pathologies:

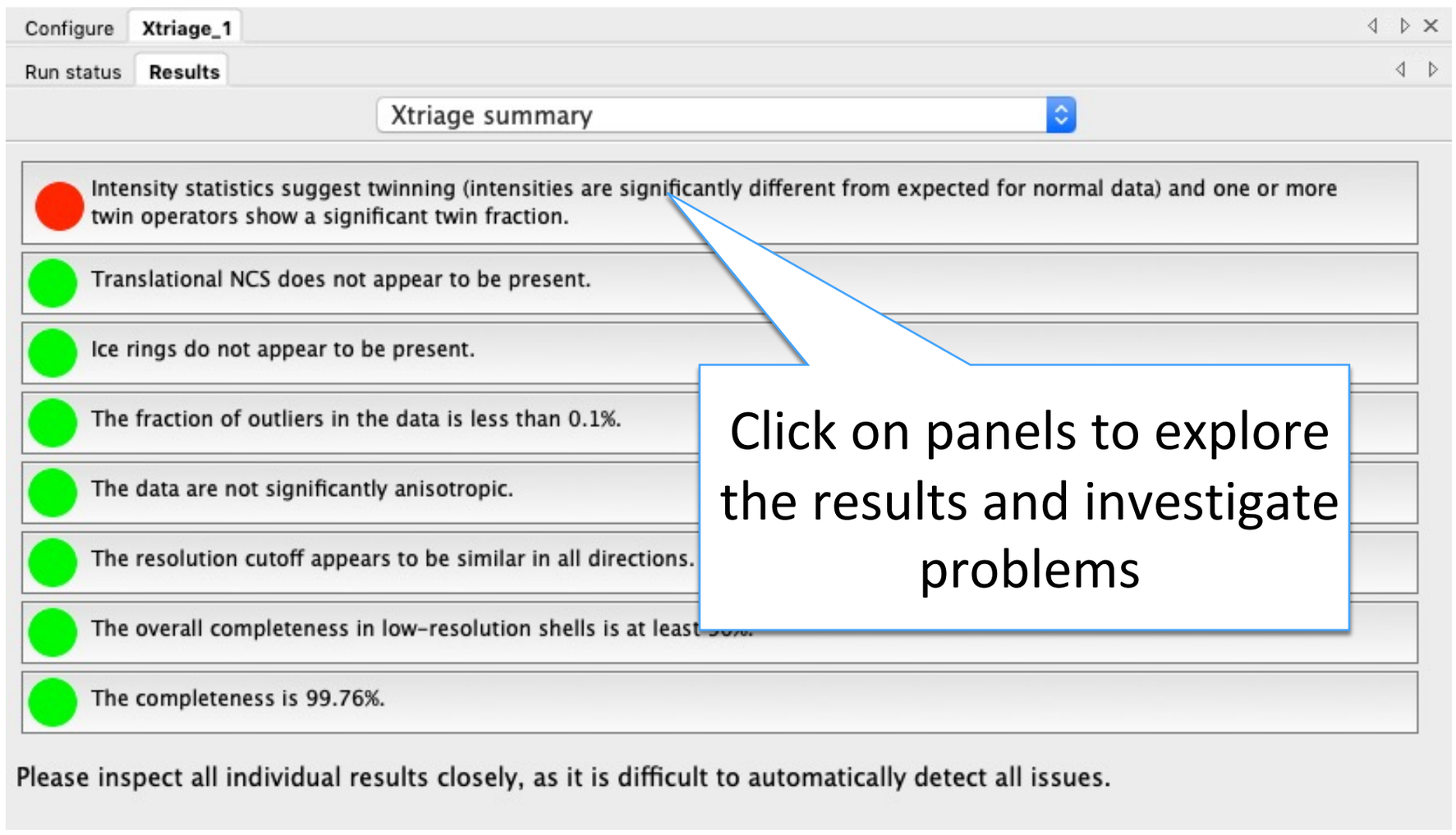
- Twinning: two or more crystals are intergrown (orientations are related by twin operation)
- tNCS: more than one copy of a molecule is in a similar orientation in the asymmetric unit

Data Quality Assessment

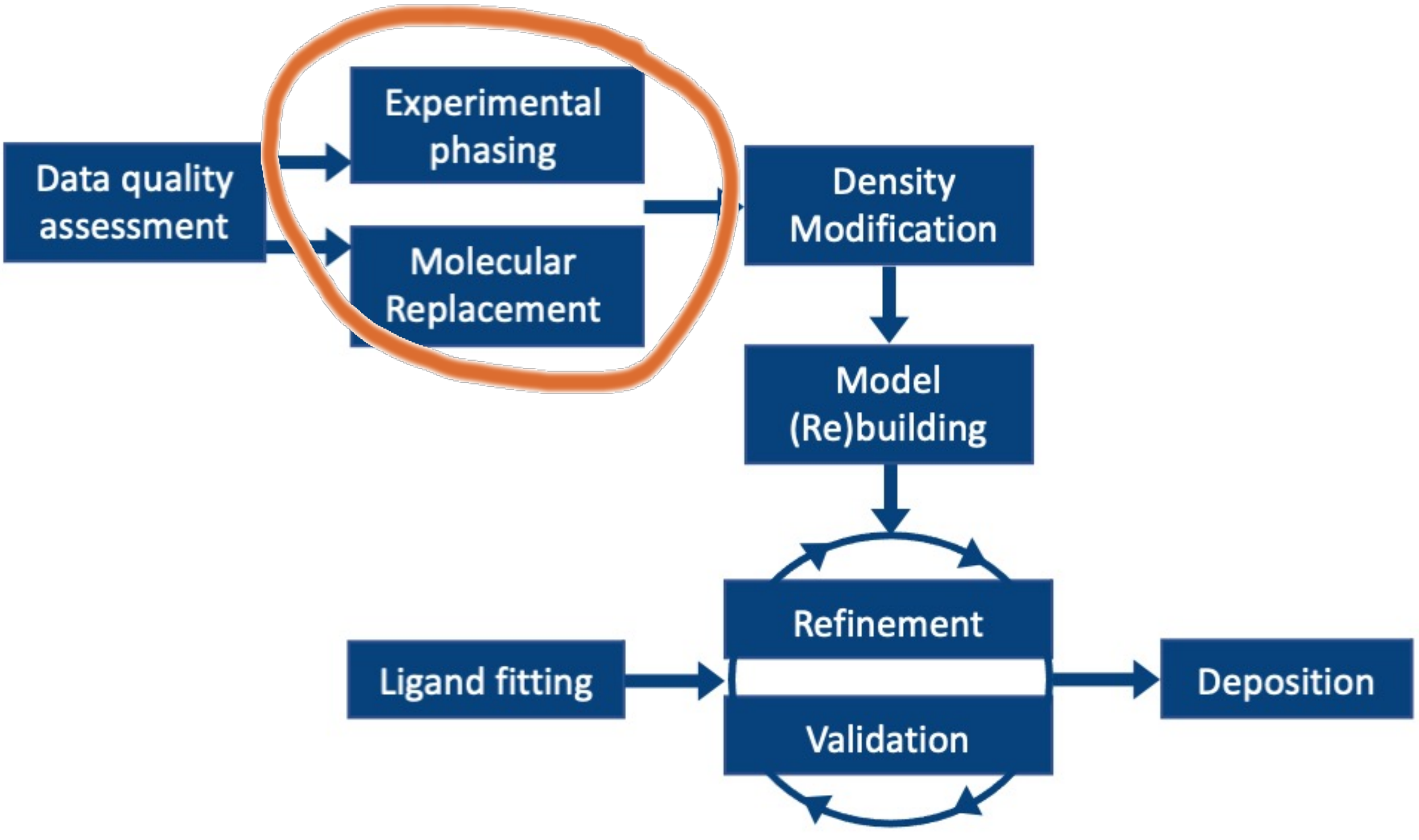
Data anomalies can prevent structure solution!

→ It is important to check your data before phasing, model building and refinement.

Xtriage does diagnostics for major pathologies and data properties (Wilson plot, completeness, symmetry).

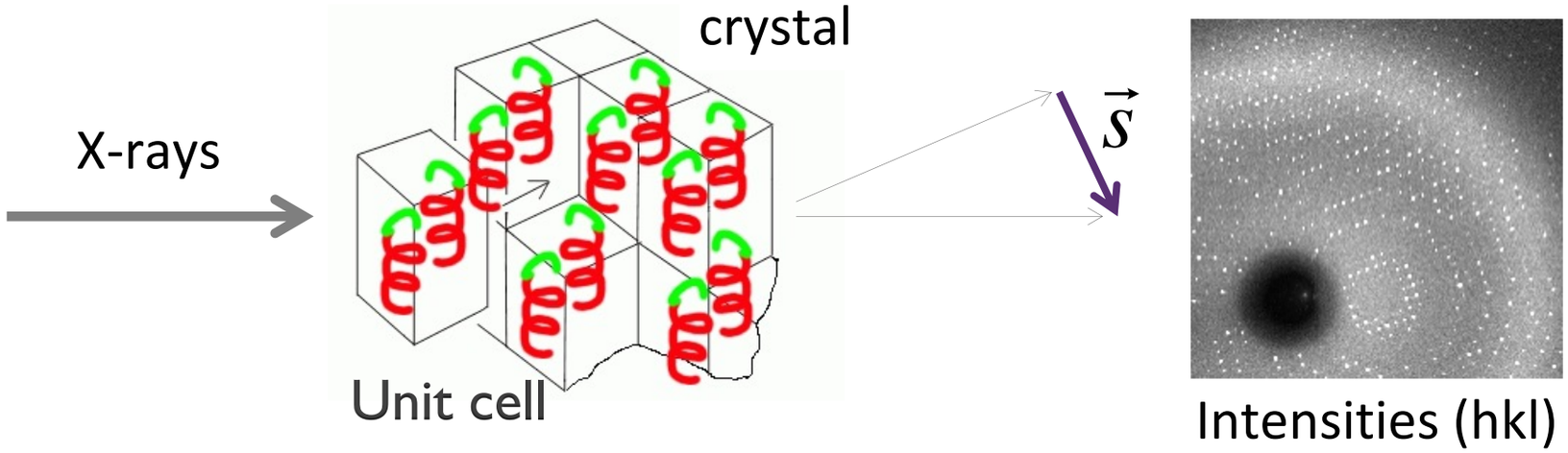


Steps in crystallography



Goal of crystallographic experiment

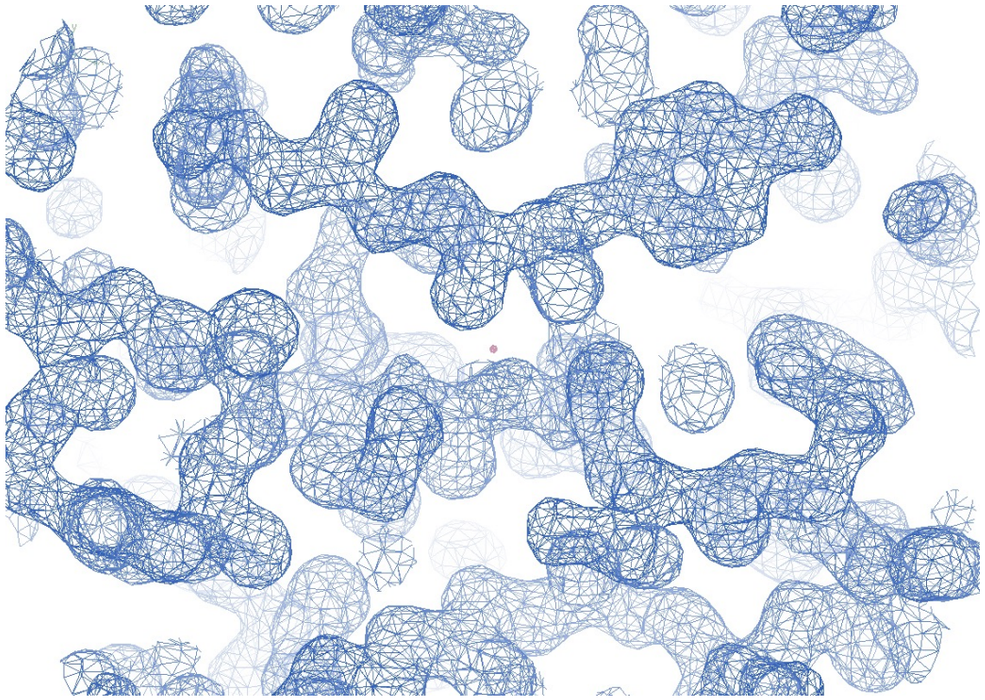
Typically, the goal is to determine the **structure**.
(arrangement of atoms in space)



The electron density in the unit cell is related to the Fourier transform of the **amplitude and phase of the scattered X-rays**.

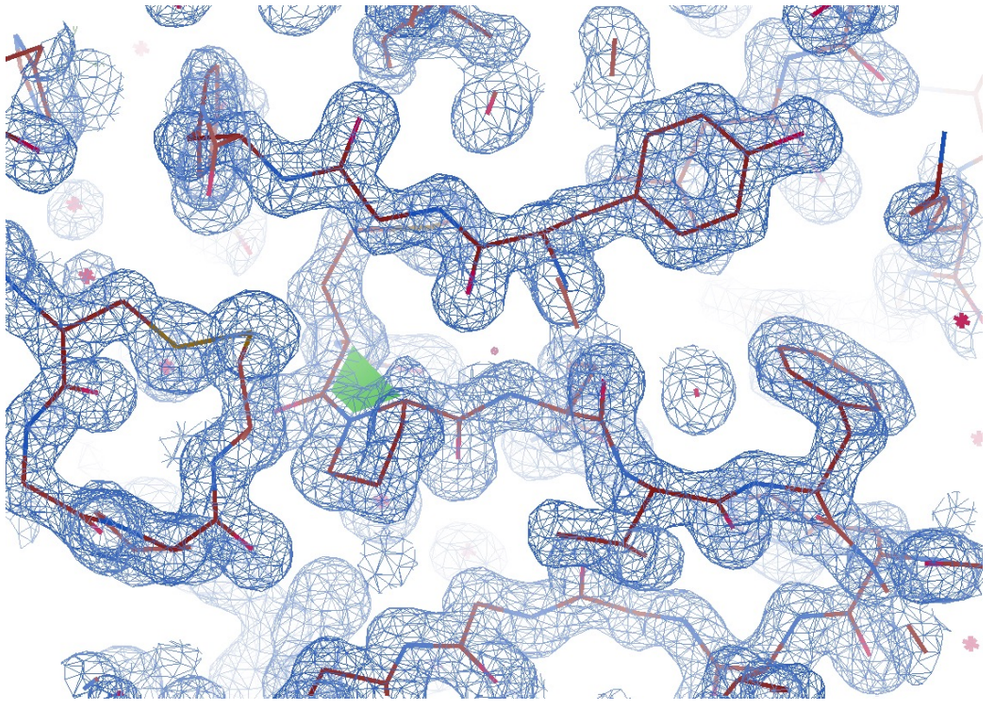
$$\rho(\vec{r}) = FT\left(\vec{F}(\vec{H})\right) = \frac{1}{V}\int \vec{F} \cdot e^{-2i\pi\vec{H}\cdot\vec{r}}$$

Goal of crystallographic experiment



If we know the density…

Goal of crystallographic experiment



If we know the density…

… then we can
determine the structure

Unfortunately:

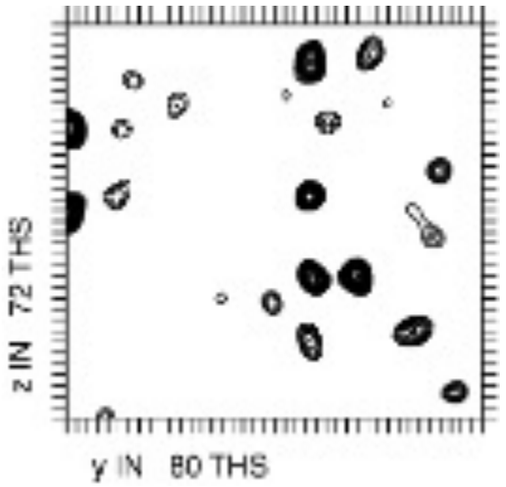
ϕ is lost: phase problem

$$\rho(\vec{r}) = FT\left(\vec{F}(\vec{H})\right) = \frac{1}{V}\int |F| e^{i\phi} \cdot e^{-2i\pi\vec{H}\cdot\vec{r}}$$

obtained from the experiment: $I \propto |F|$

→ We need to recover the phases.

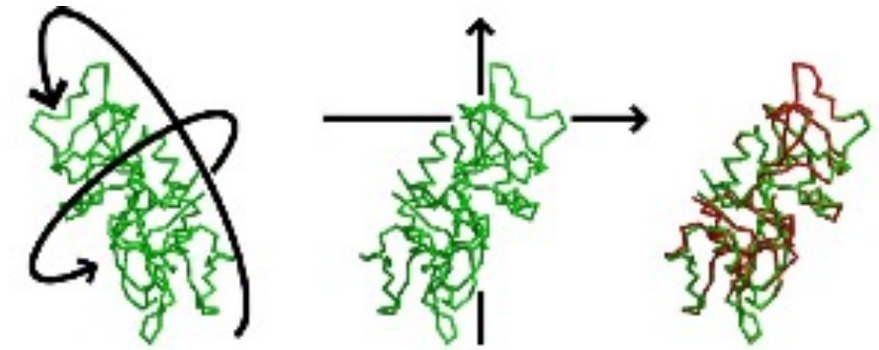
How to recover phases



Experimentally

Exploit the properties of a few special atoms:

- anomalous scattering
- a large number of electrons



Computationally

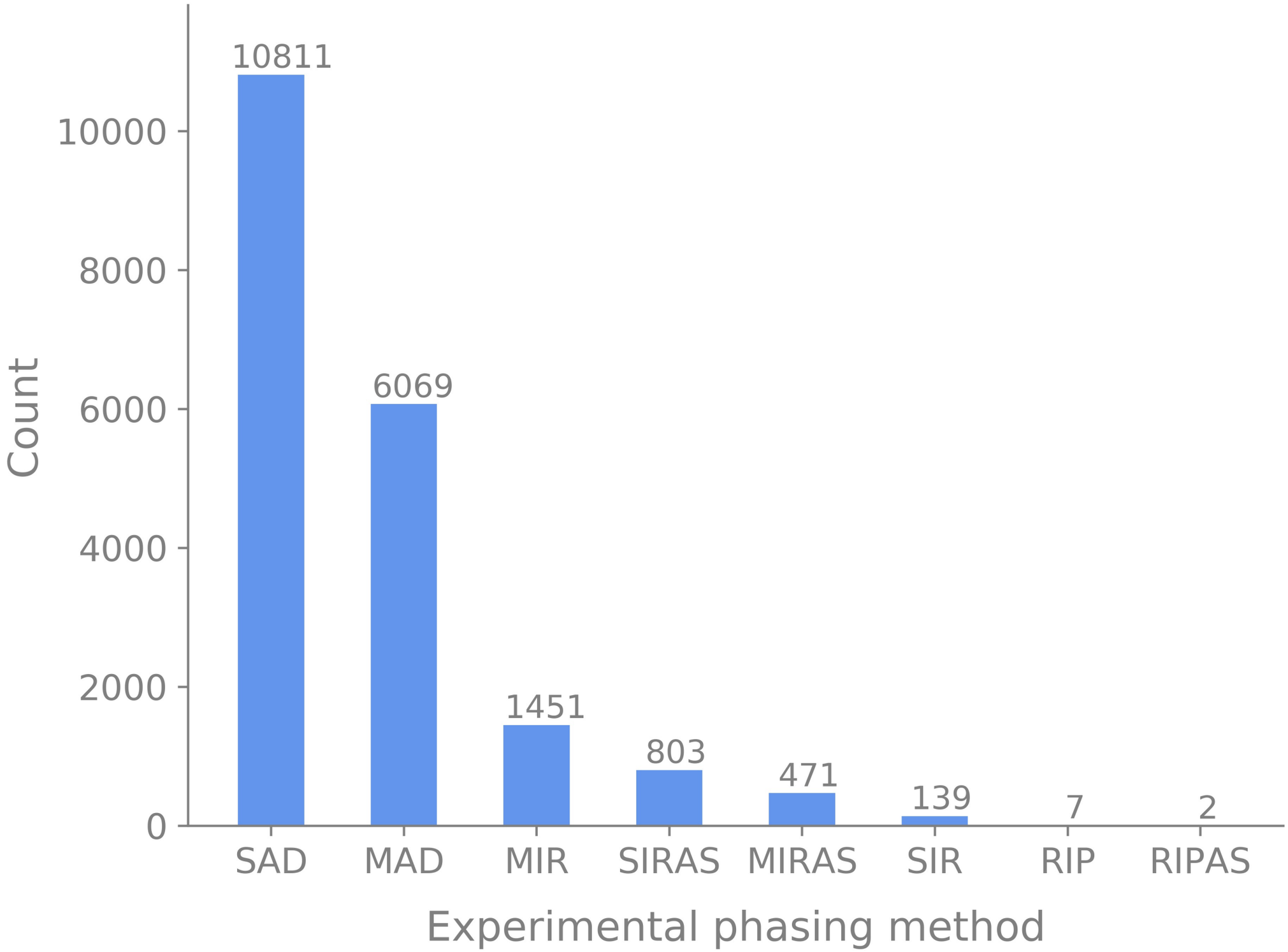
- *Molecular Replacement (MR)*
  A previously known structure can provide initial phase estimates for a new structure

- *Direct Methods*
  Phase relationships can be formulated by assuming the positivity and atomicity of the electron density

Experimental phasing methods

Method		Phasing information
SIR	Single isomorphous replacement	A few electron-rich atoms
MIR	Multiple isomorphous replacement	Several derivatives
SAD	Singe-wavelength anomalous diffraction	A few anomalous scatterers
native SAD	SAD based on native sulfurs	Special case of SAD
MAD	Multiple-wavelength anomalous diffraction	A few anomalous scatterers; Collect at several wavelengths
SIRAS	Single isomorphous replacement with anomalous scattering	SAD and SIR
MIRAS	Multiple isomorphous replacement with anomalous scattering	MAD and MIR
RIP	Radiation damage induced phasing	Special case of SIR

Experimental phasing methods: count



Retrieved Oct 2021

Experimental Phasing with AutoSol

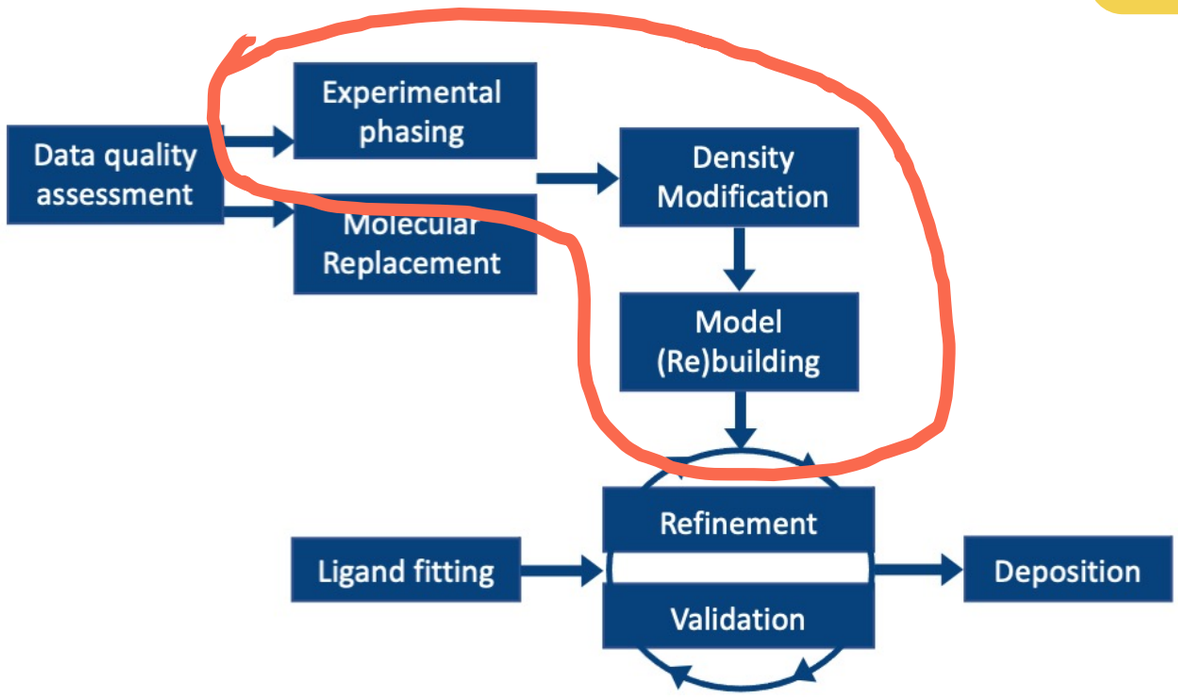
Experimental data
Sequence
Special atom
Wavelength(s)

1. Determine the substructure

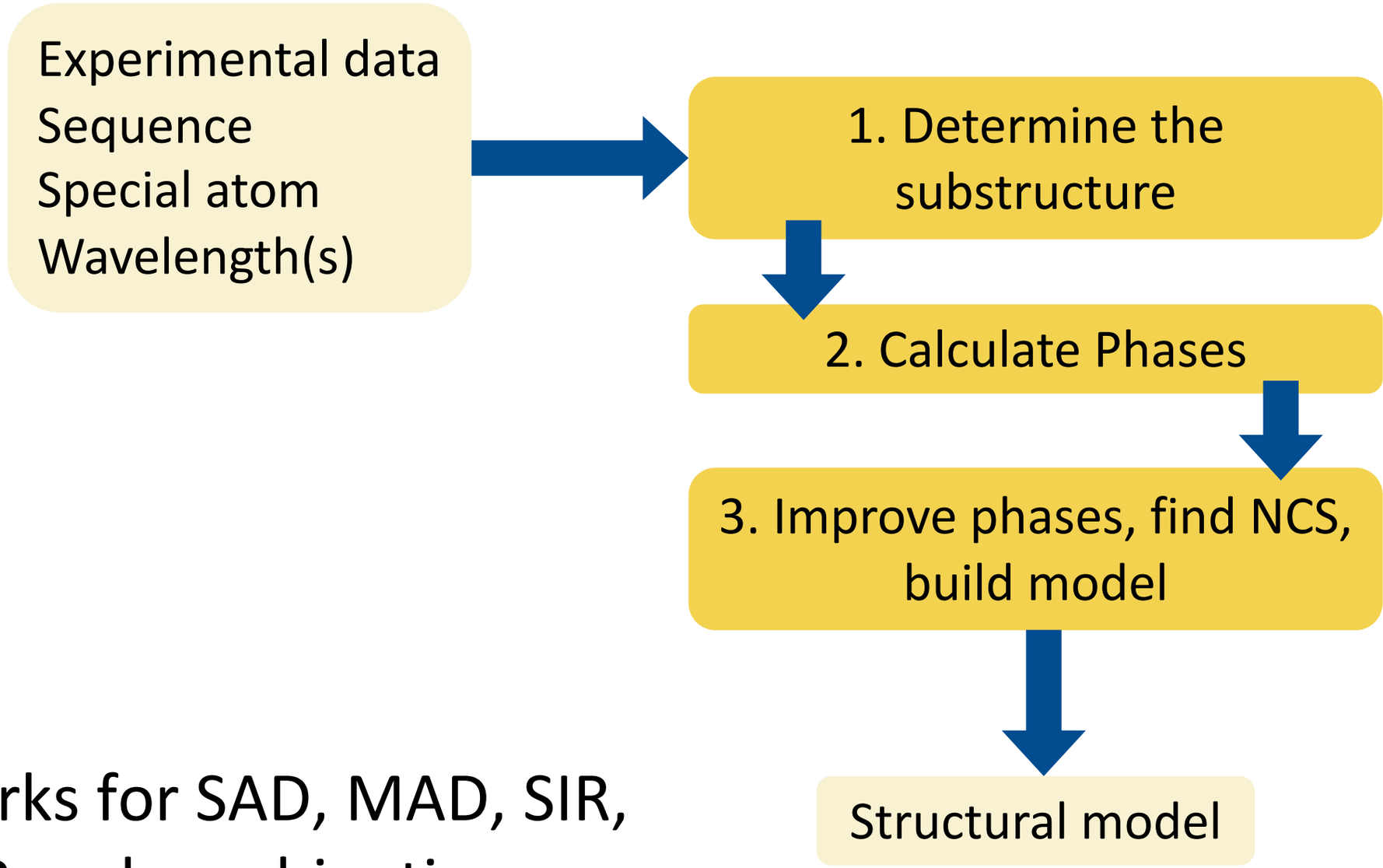
2. Calculate Phases

3. Improve phases, find NCS, build model

Structural model



Experimental Phasing with AutoSol

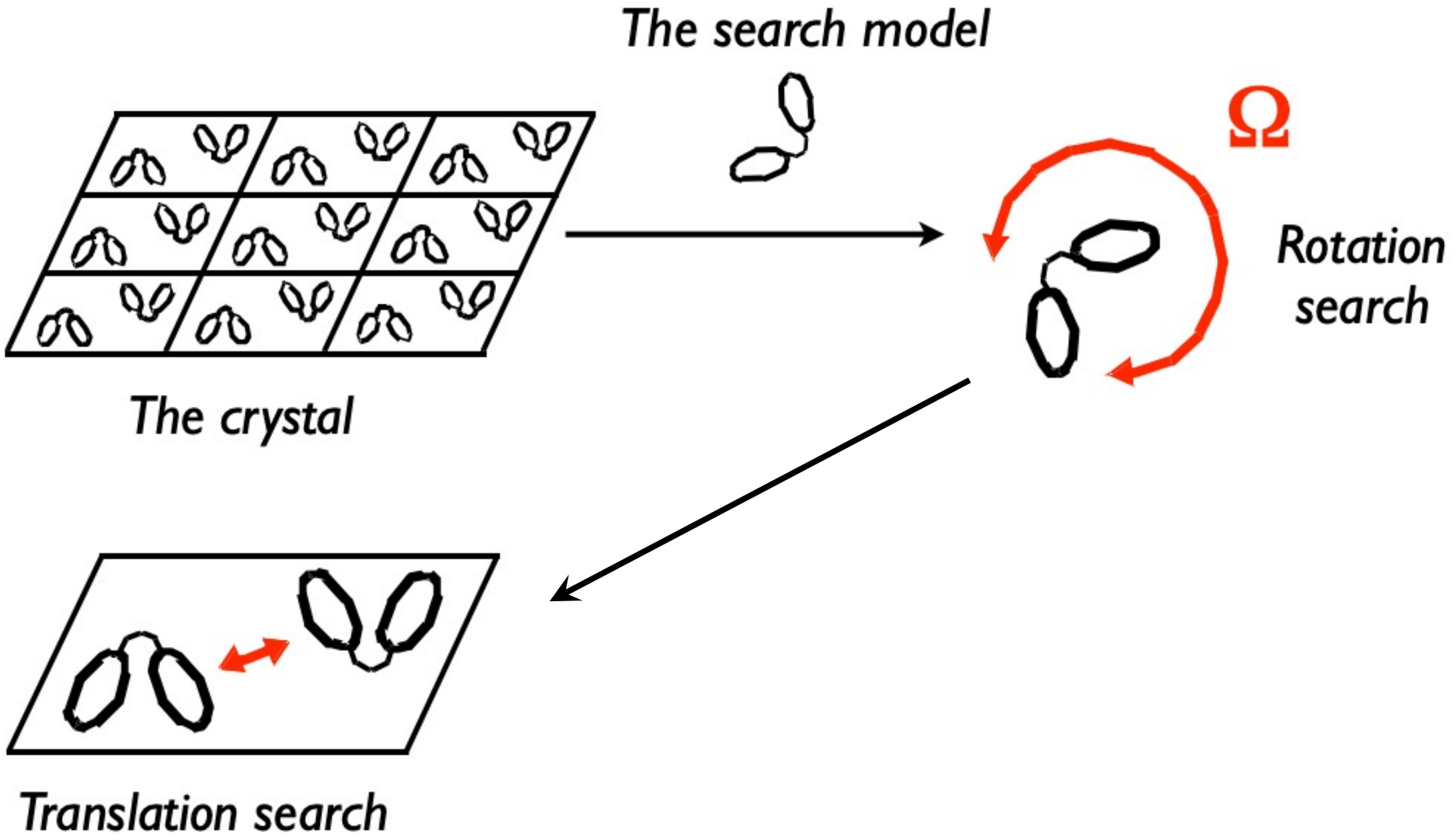


Works for SAD, MAD, SIR, MIR and combinations.

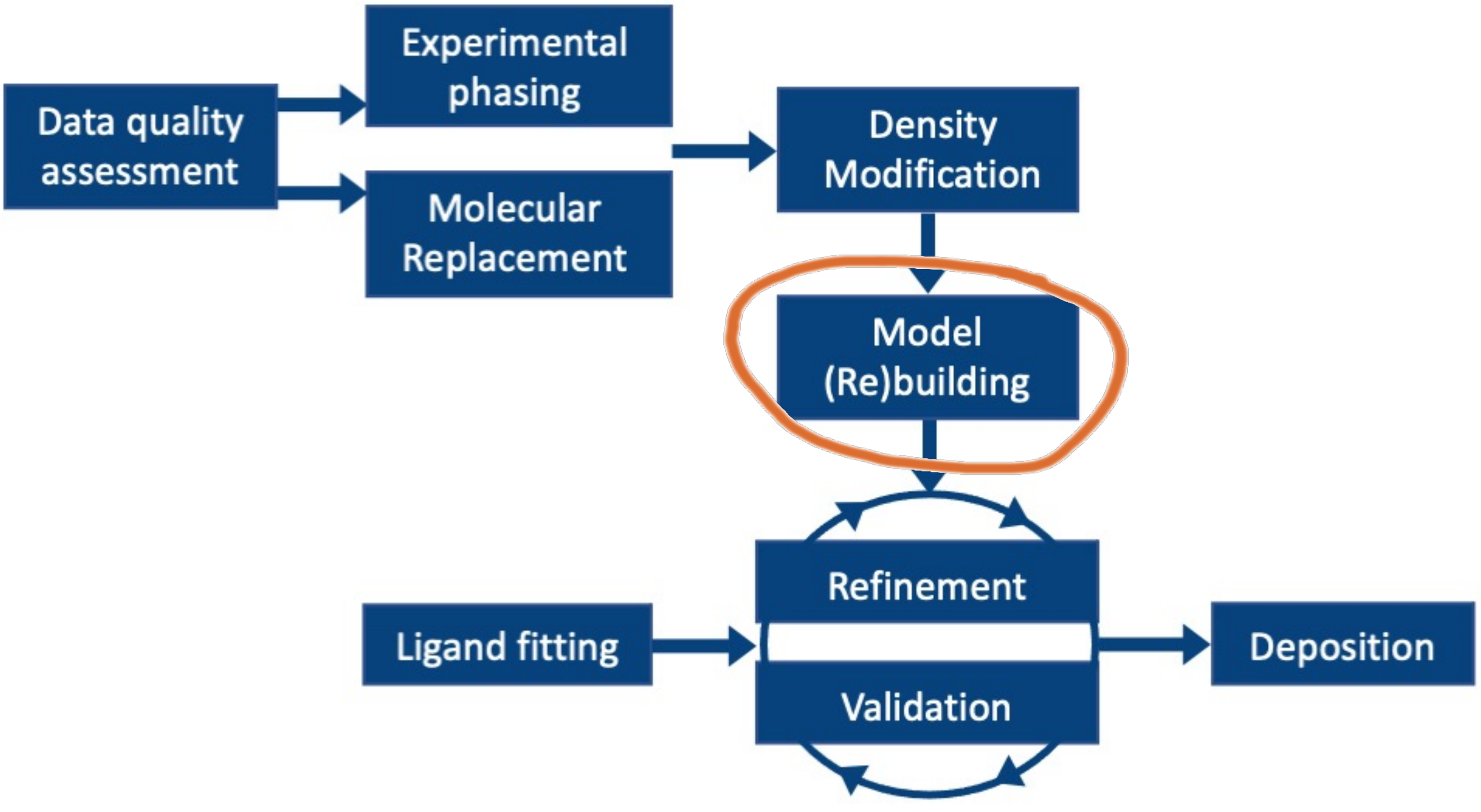
This procedure is fully automatic!

Molecular replacement (MR)

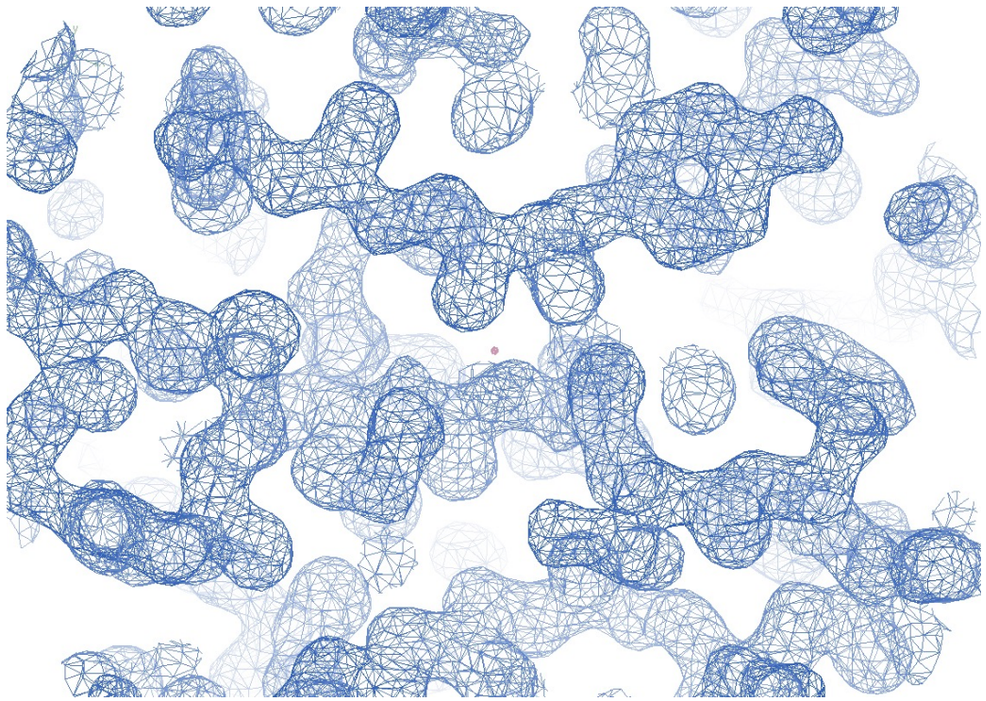
Use a previously known structure to get phase estimates



Steps in crystallography



Automated Model Building



Goal:
Build a model into an (interpretable) density map

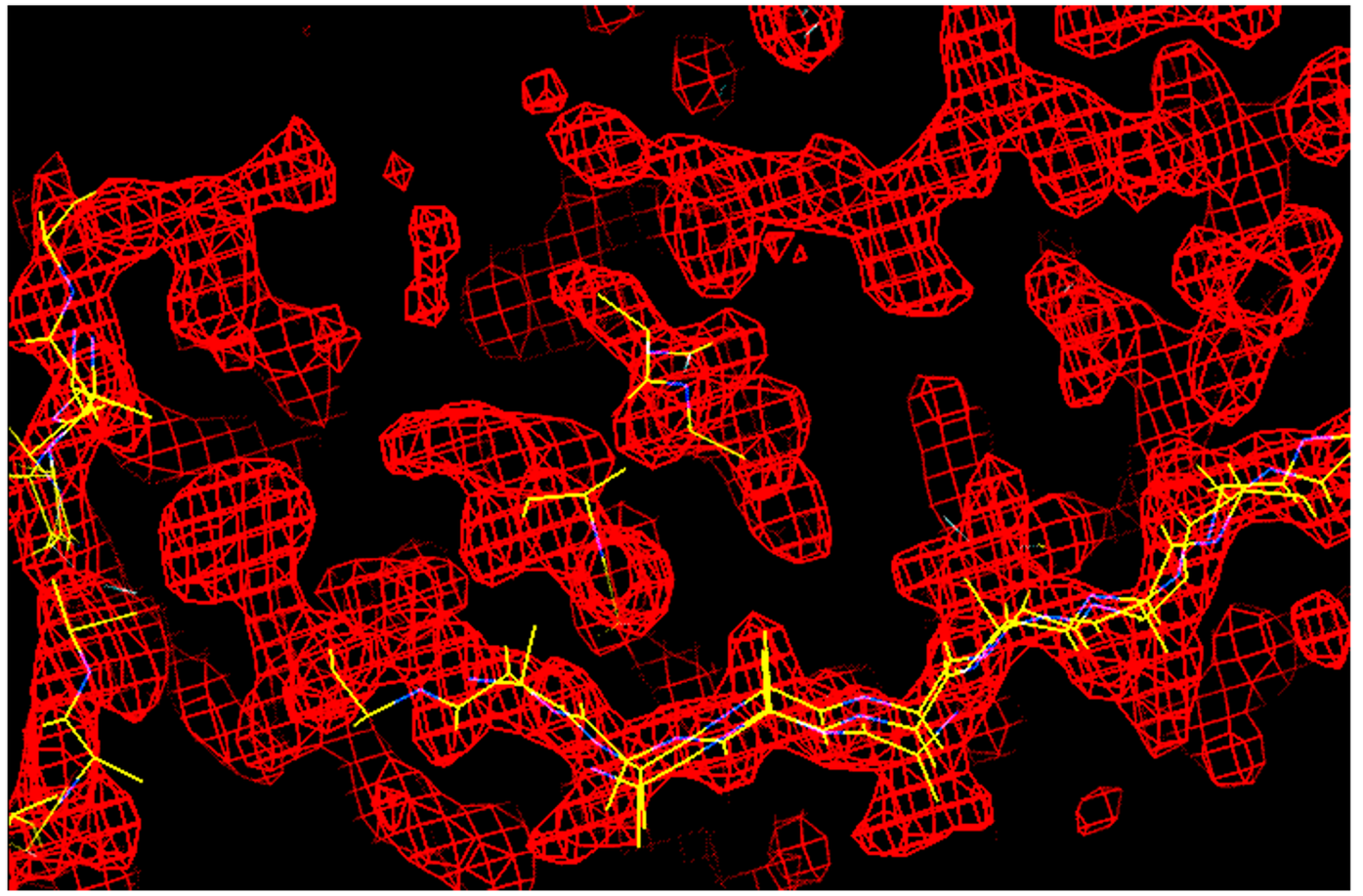
“by hand”: tedious, time-consuming, prone to errors.

→ Task can be automated: **AutoBuild**

Automated Model Building: AutoBuild

Multi-step procedure:

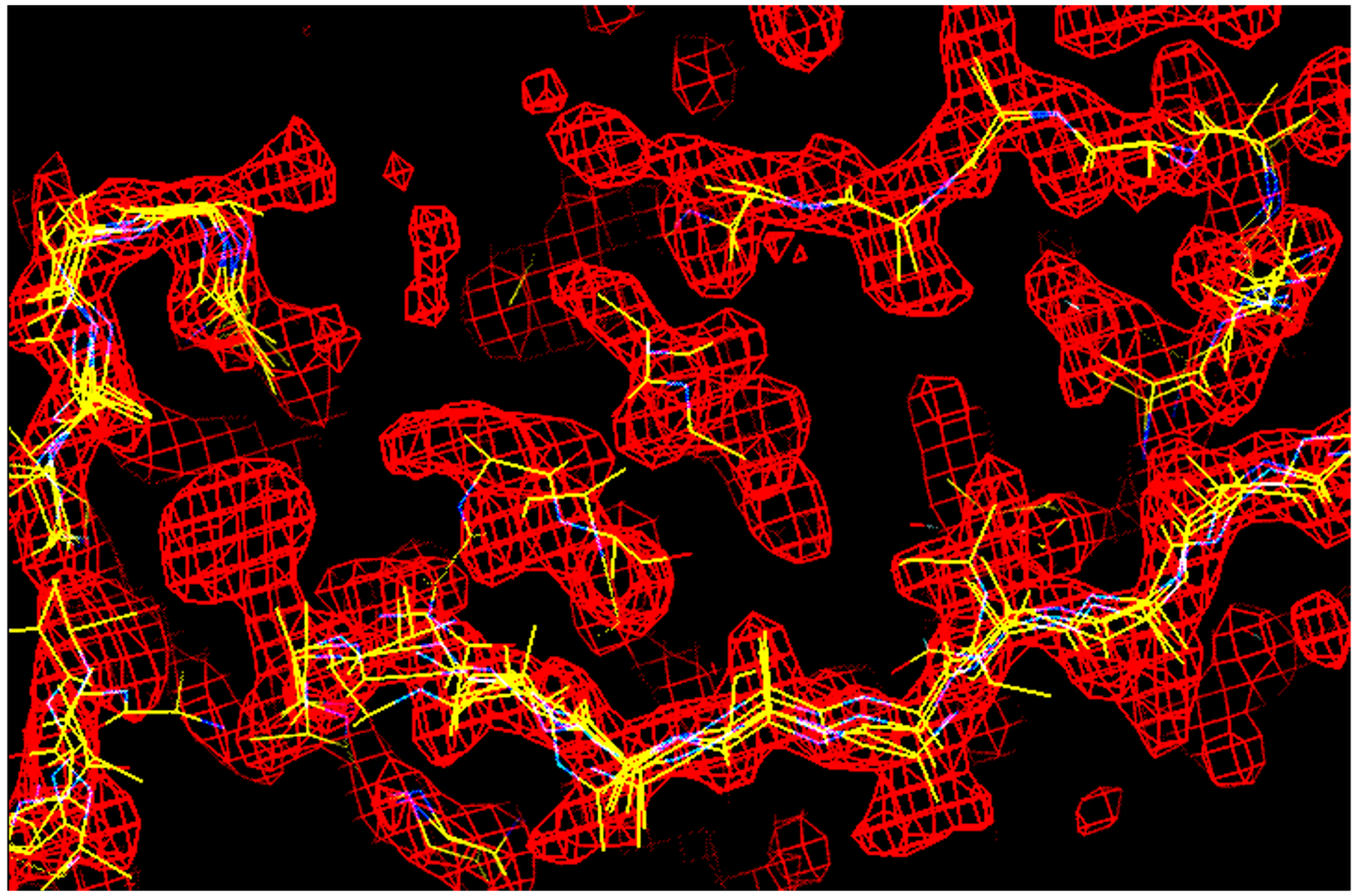
- Locate helices and strands



Automated Model Building: AutoBuild

Multi-step procedure:

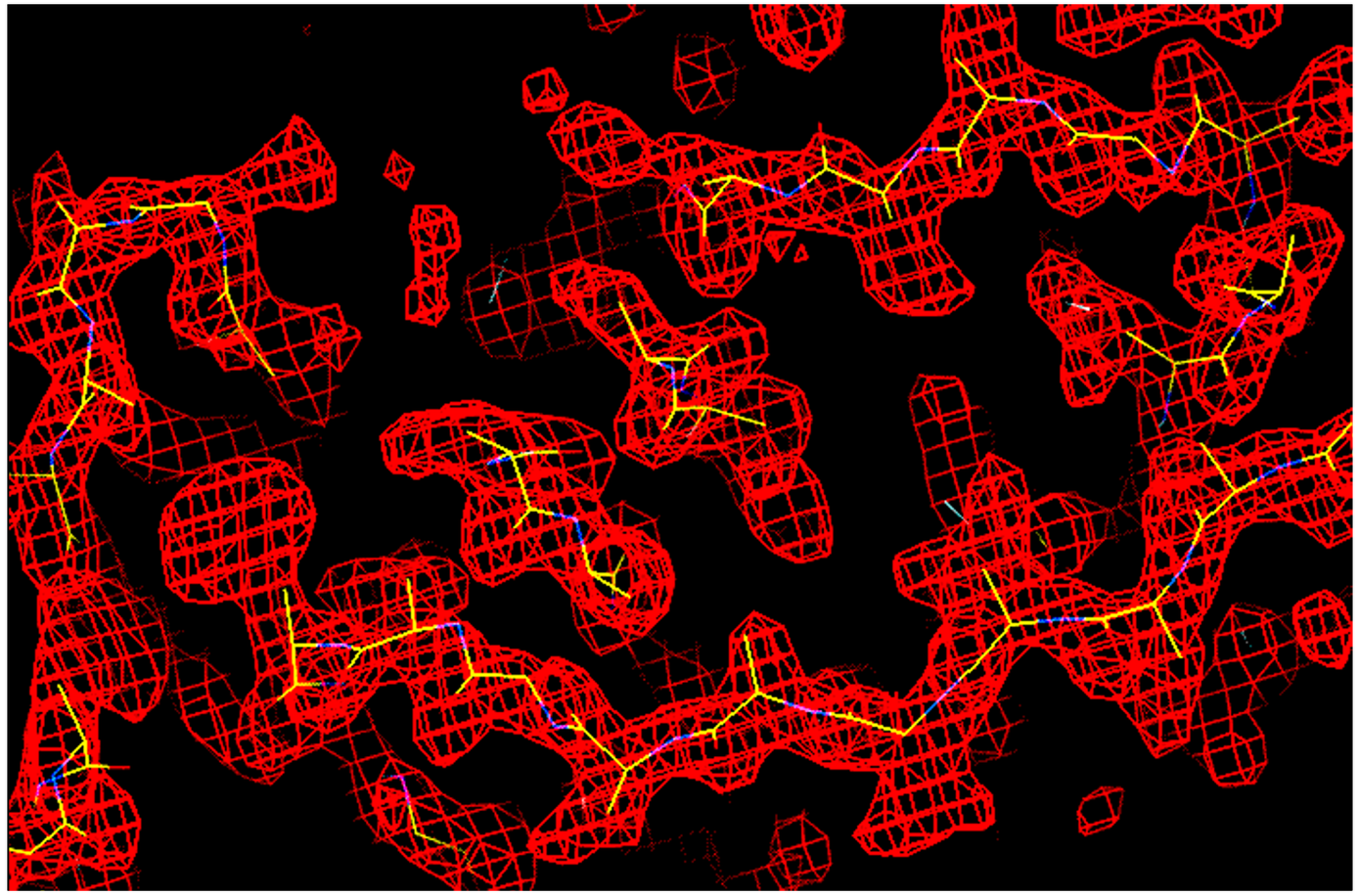
- Locate helices and strands
- Extend helices and strands iteratively with tripeptides from libraries



Automated Model Building: AutoBuild

Multi-step procedure:

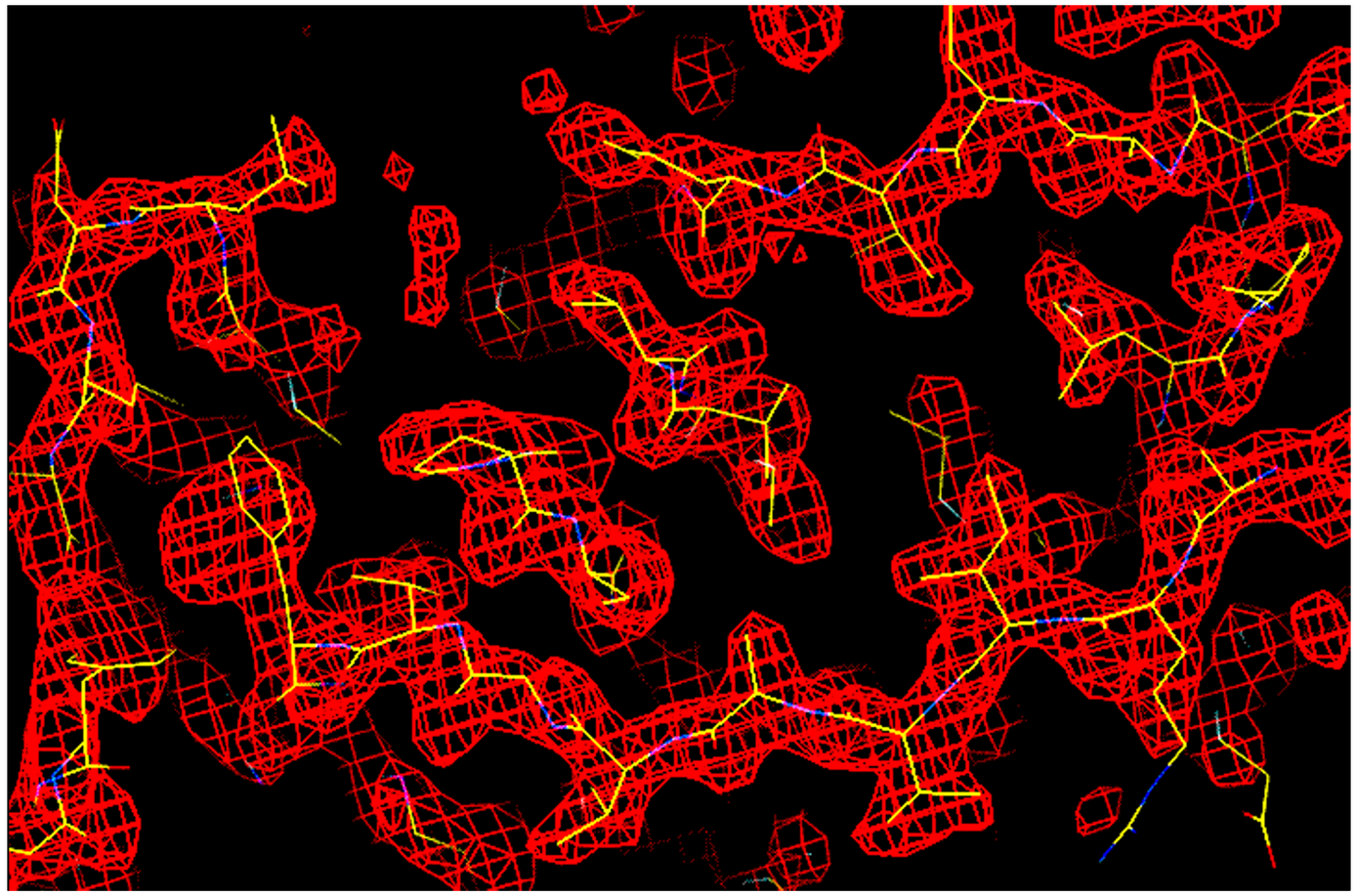
- Locate helices and strands
- Extend helices and strands iteratively with tripeptides from libraries
- Assemble fragments into a poly-ala chain



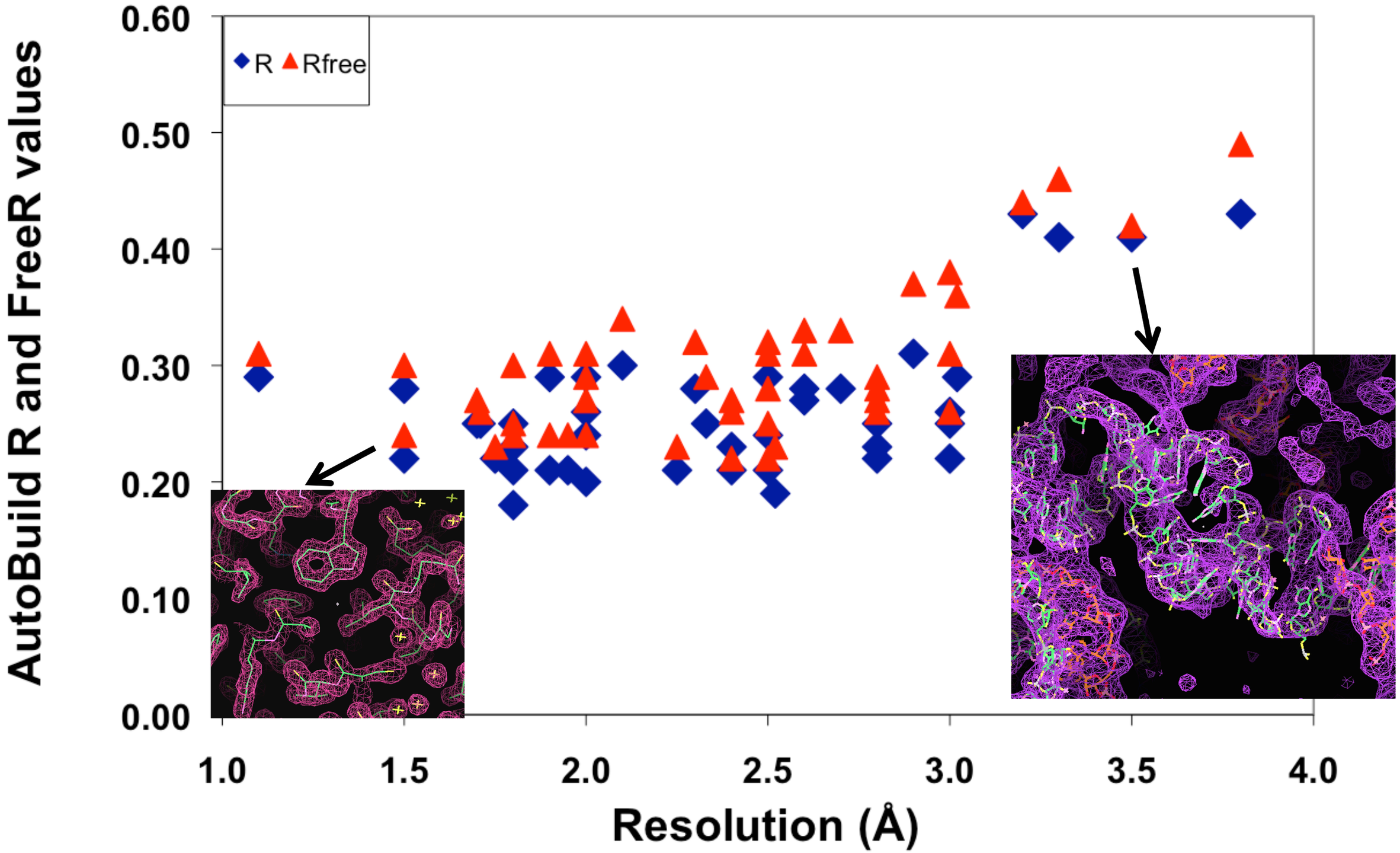
Automated Model Building: AutoBuild

Multi-step procedure:

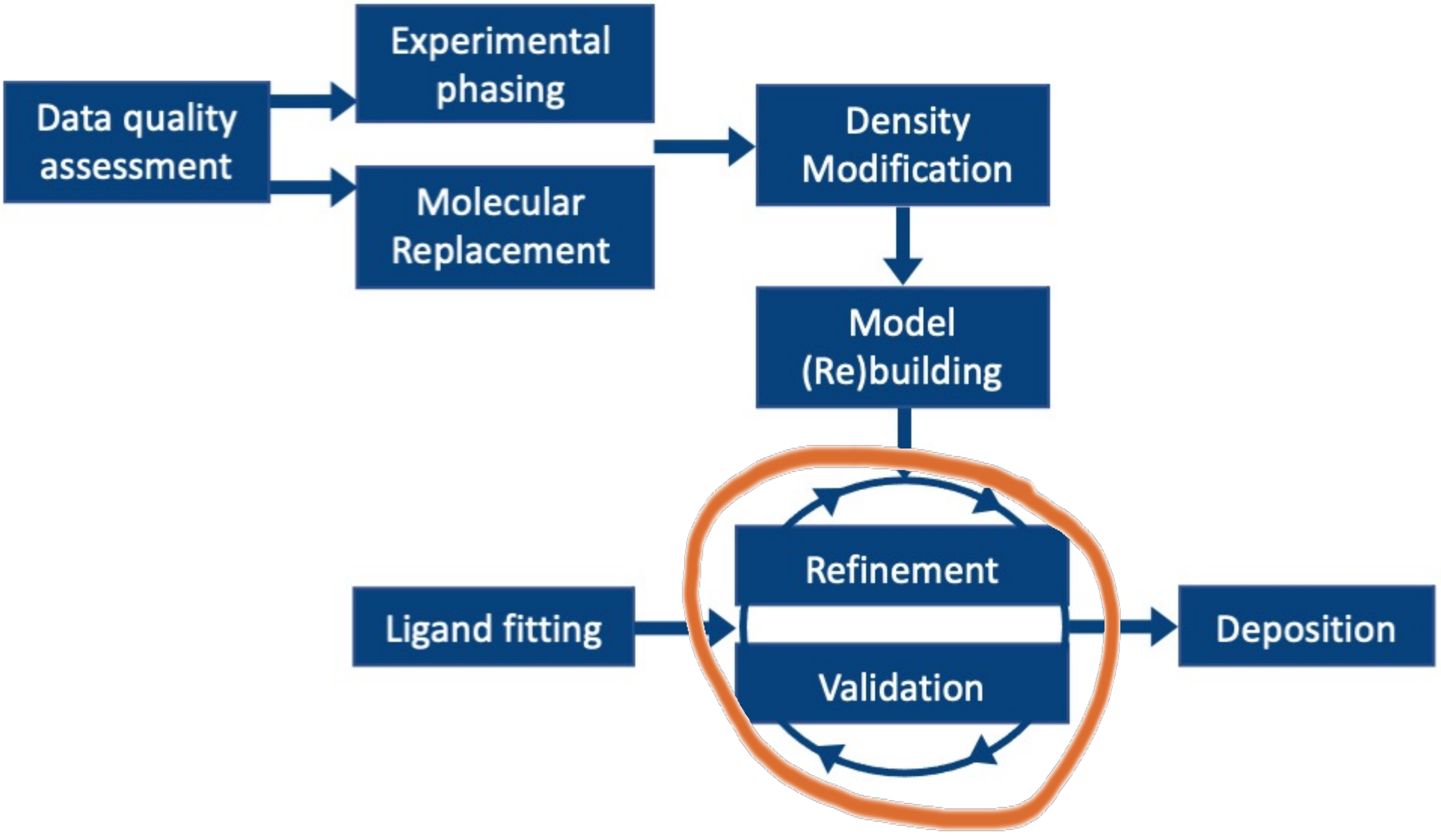
- Locate helices and strands
- Extend helices and strands iteratively with tripeptides from libraries
- Assemble fragments into a poly-ala chain
- Build side chains and align them to the protein sequence



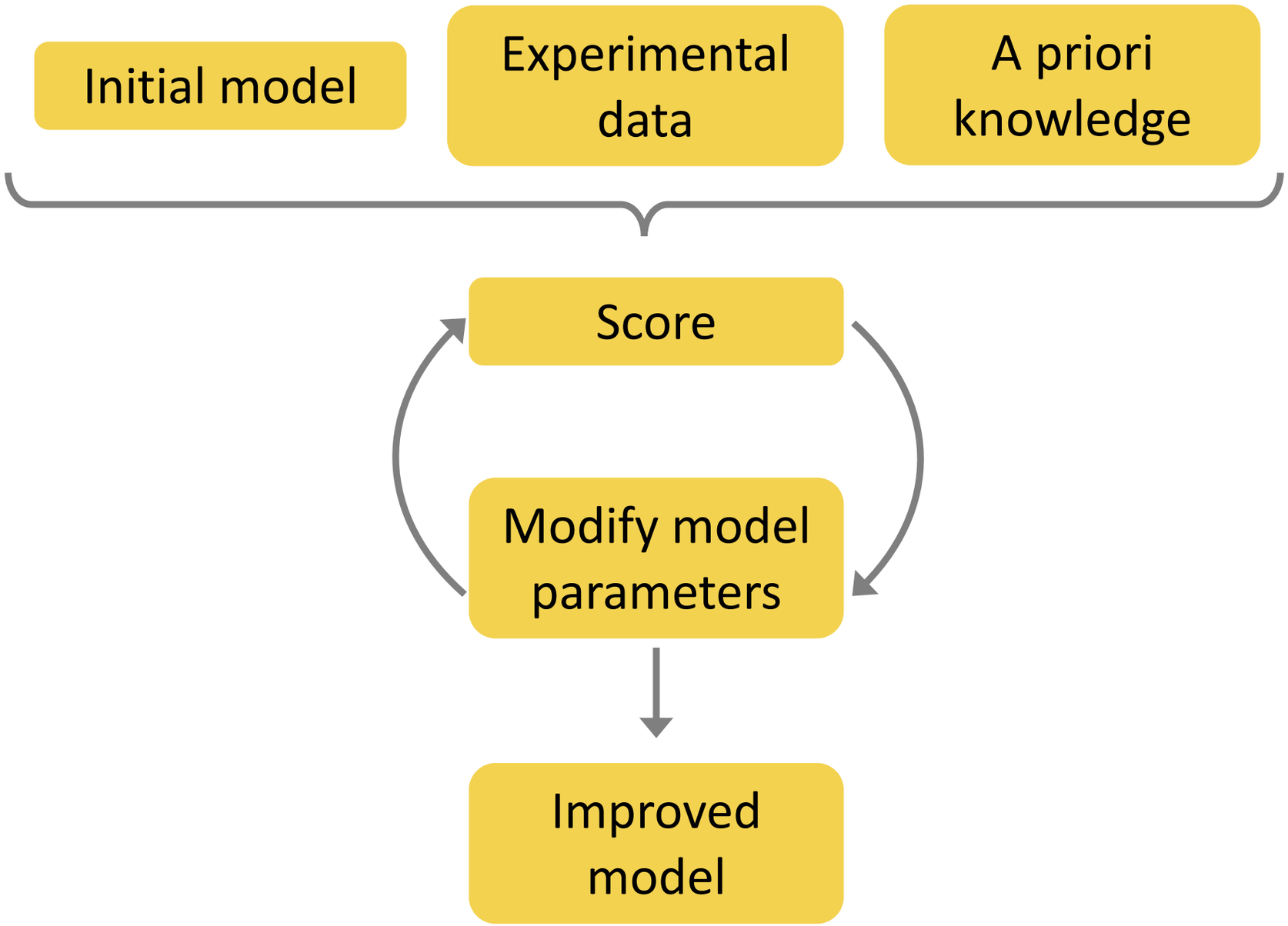
Automated Model Building: AutoBuild



Steps in crystallography



Refinement

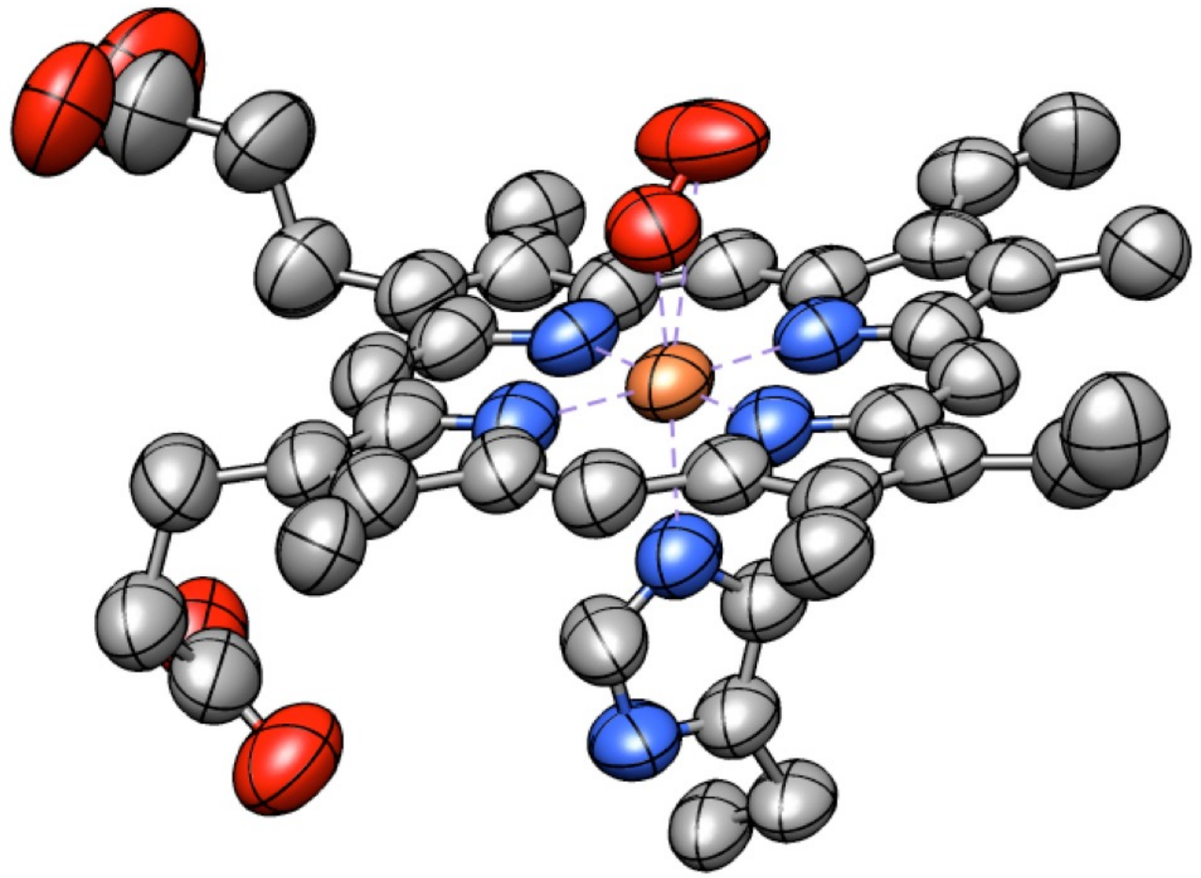


Refinement = Use an *optimization* algorithm to minimize a *target function* by changing the *parameters* of the model

Refinement: Model parameters

Parameters that describe the crystal and its content.

- **Atomic model:**
  - coordinates
  - B-factors
  - occupancies
- **Non-atomic model**
  - bulk-solvent
  - anisotropy
  - twinning



$$\mathbf{F} = k\{\mathbf{F}_{\text{calc}} \exp[-\Delta B(\sin\theta/\lambda)^2] + d_{\text{solv}}\mathbf{F}_{\text{solv}} \exp[-B_{\text{solv}}(\sin\theta/\lambda)^2]\}$$

$$\mathbf{F}_{\text{CALC (ATOMS)}}(h,k,l) = \sum_{n=1}^{Natoms} q_n f_n(s) \exp\left(-\frac{B_n s^2}{4}\right) \exp(2i\pi \mathbf{r}_n \mathbf{s})$$

Refinement: Target function

Objective function

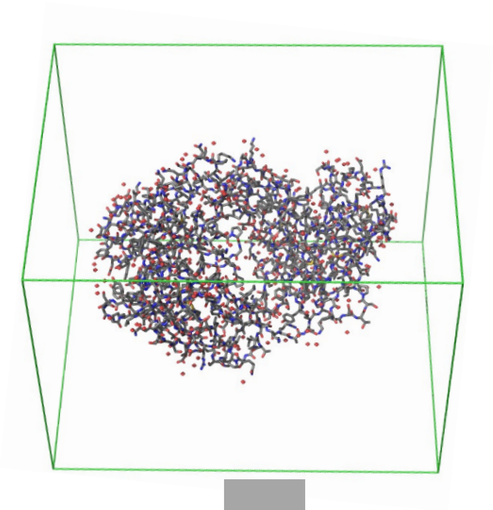
- assesses the fit to experimental data
- relates model parameters and data
- Least-squares (LS), Maximum-Likelihood (ML), R-factor, …

- Based on structure factors (reciprocal-space refinement)
- Based on electron density (real-space refinement)

$$T = T_{\mathrm{DATA}}(F_{\mathrm{OBS}}, F_{\mathrm{MODEL}})$$

Refinement: Target function

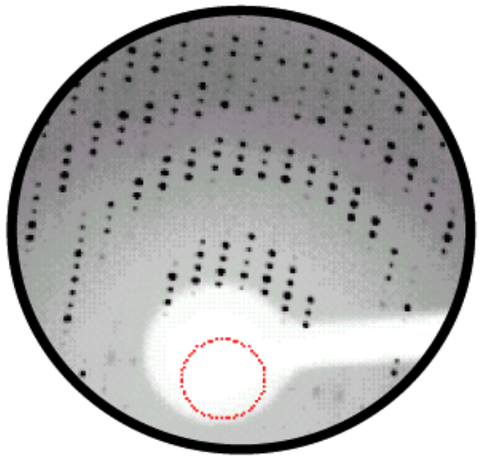
Model



Model parameters:
Coordinates, B-factors,
Bulk-solvent…

$\mathbf{F}_{\mathrm{MODEL}}$

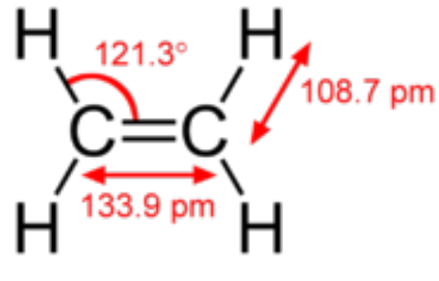
Experimental data



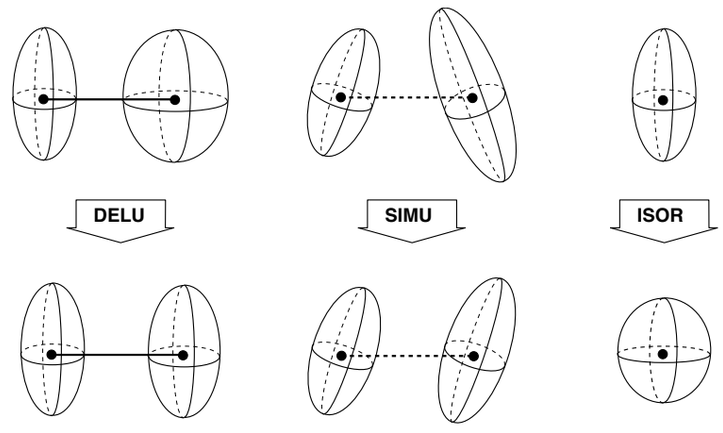
$\mathbf{F}_{\mathrm{OBS}}$

A *priori* knowledge about model

Covalent geometry:



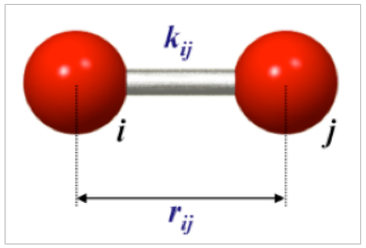
Similarity of B-factors



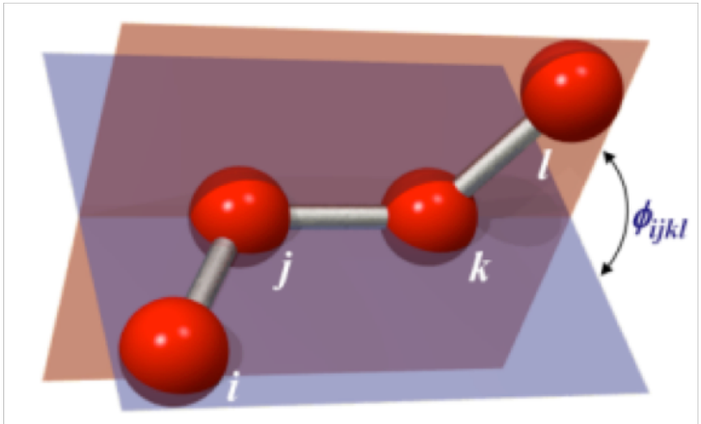
$$T = T_{\mathrm{DATA}}(F_{\mathrm{OBS}}, F_{\mathrm{MODEL}}) + wT_{\mathrm{RESTRAINTS}}$$

Restraints: *a priori* knowledge

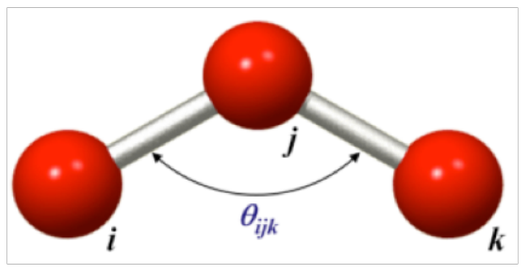
Chemistry



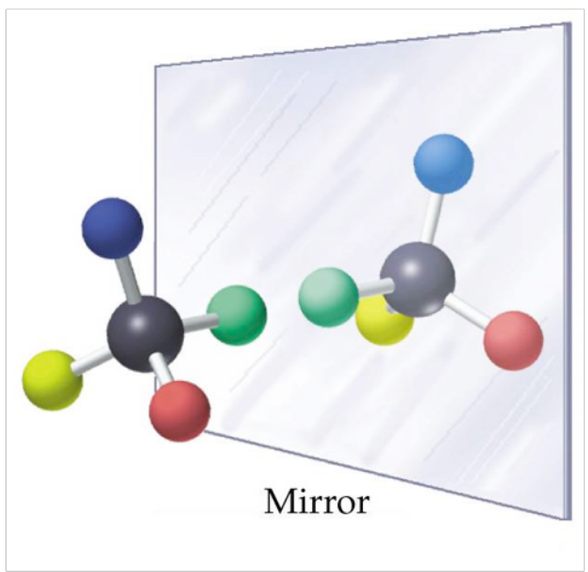
$\Sigma_{bonds}\ \omega(d_{model} - d_{ideal})^2$



$\Sigma_{dihedrals}\ \omega(1 + \cos(n\chi_{model} + \chi_{shift}))$

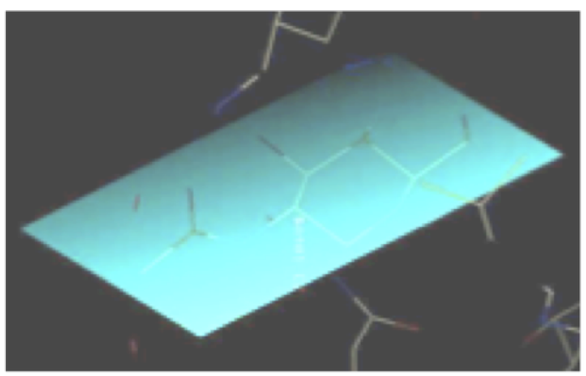


$\Sigma_{angles}\ \omega(\theta_{model} - \theta_{ideal})^2$



Volume (V) = $(r_N - r_{CA}) \bullet [(r_C - r_{CA}) \times (r_{CB} - r_{CA})]$

Images from PumMa web site (http://www.pumma.nl)



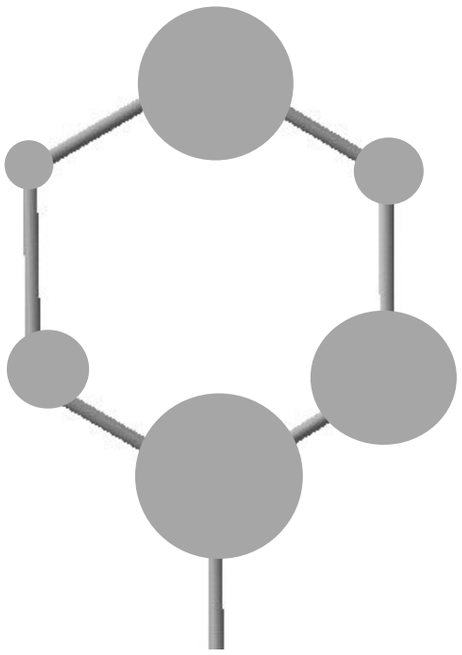
$E = \Sigma_{planes} \Sigma_{atoms}\ w\ (m \bullet r - d)^2$

Used automatically (no need to activate)

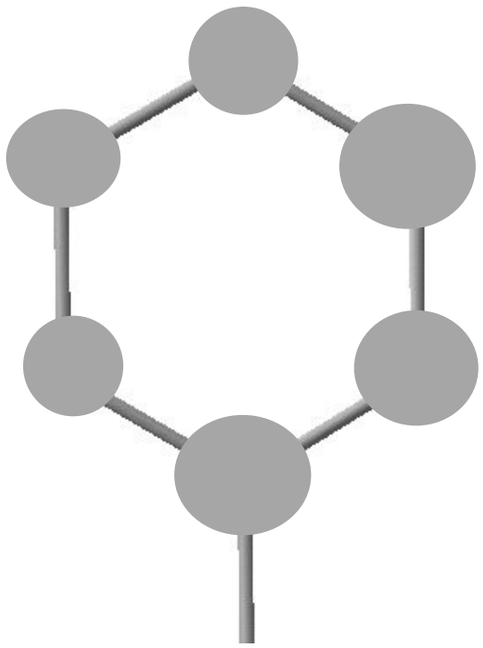
Restraints: ADP

Isotropic ADPs

Unlikely



Reasonable

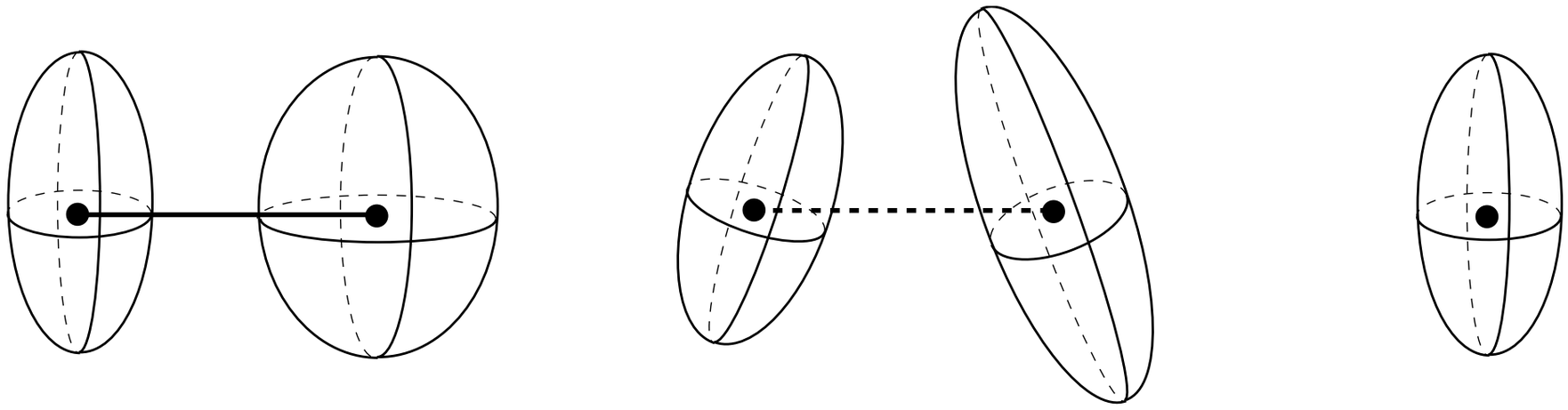


Used automatically (no need to activate)

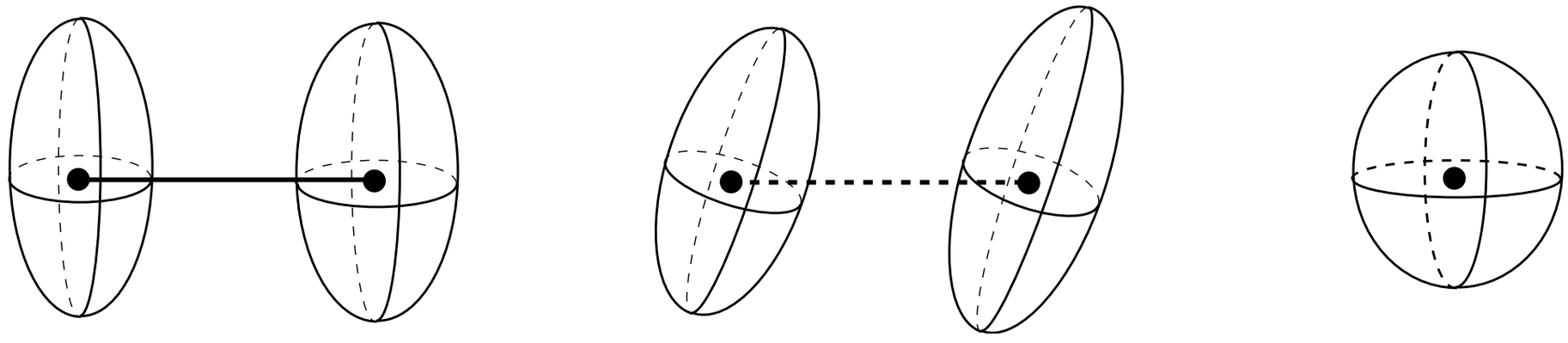
Restraints: ADP

Anisotropic ADPs

Unlikely



Reasonable

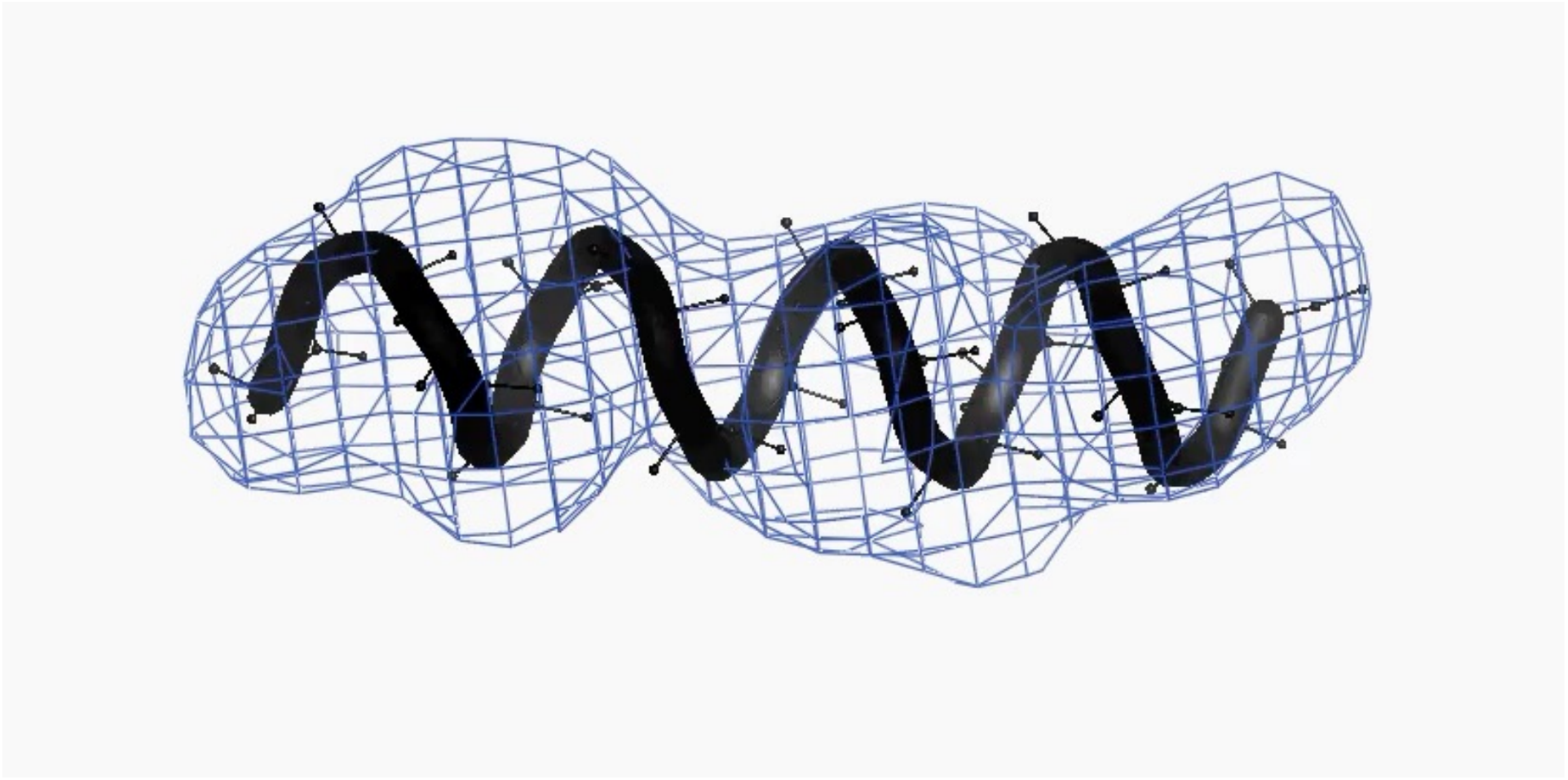


Pictures from Thomas Schneider

Restraints: secondary structure

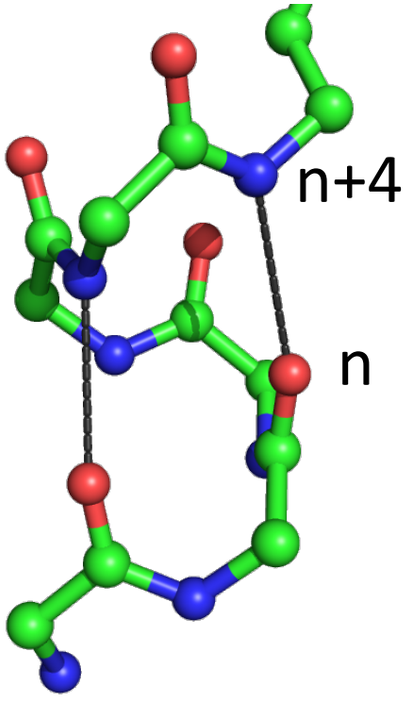
Example: refine a perfect α-helix into low-res map

- standard restraints on covalent geometry are insufficient
- Model geometry deteriorates

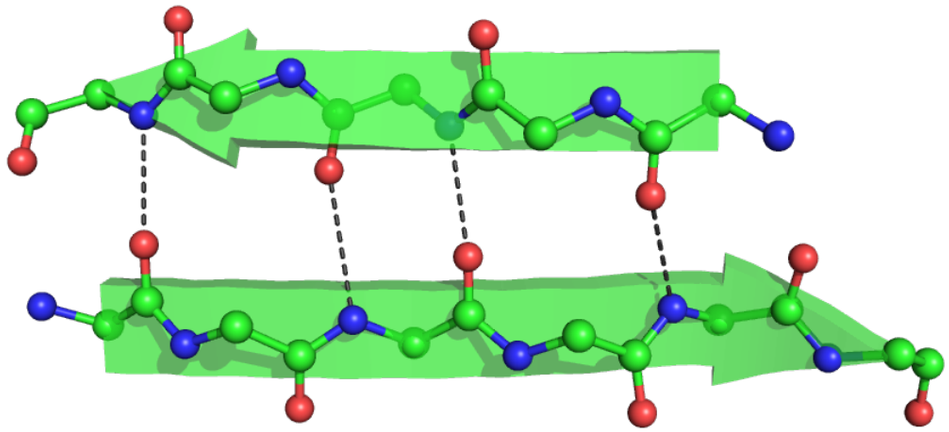


Restraints: secondary structure

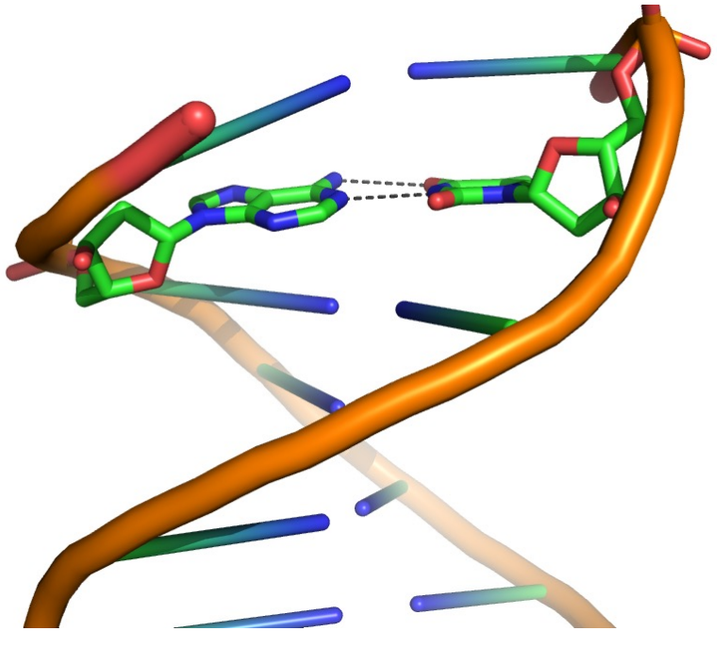
Restrain the geometry of H-bonds and stacking pairs:



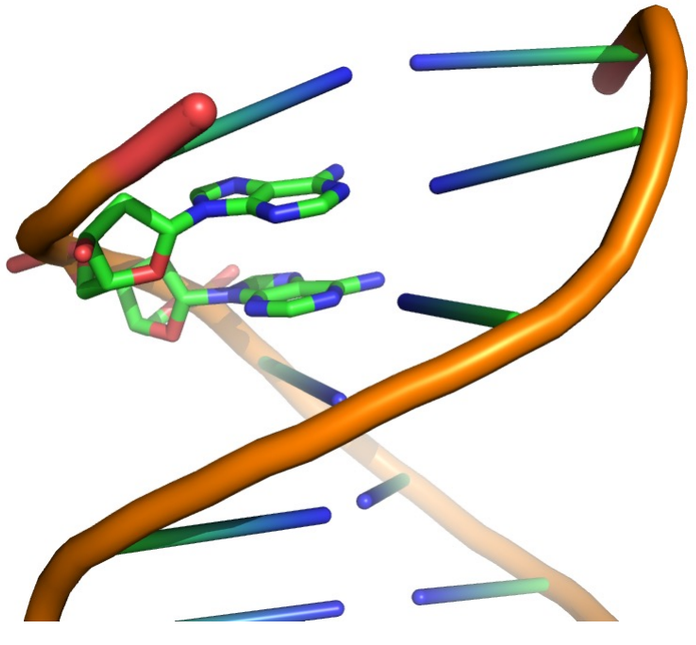
helix



beta sheet



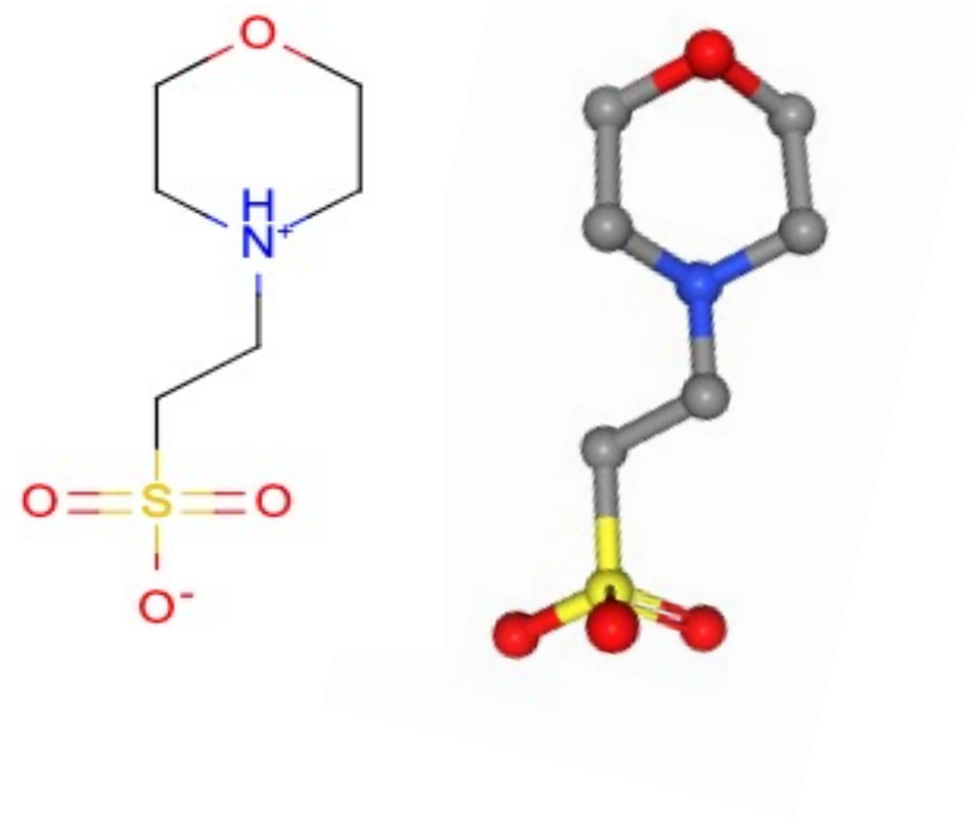
Base pairs



Stacking pairs

Restraints: Ligands

Restraints of common ligands are included in libraries.

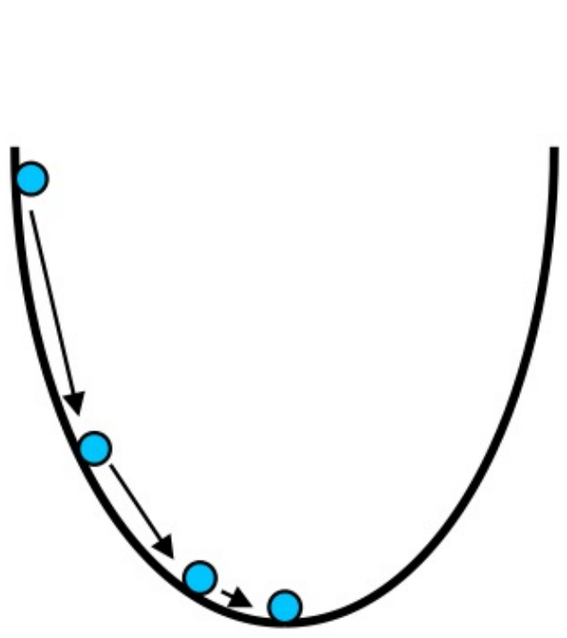


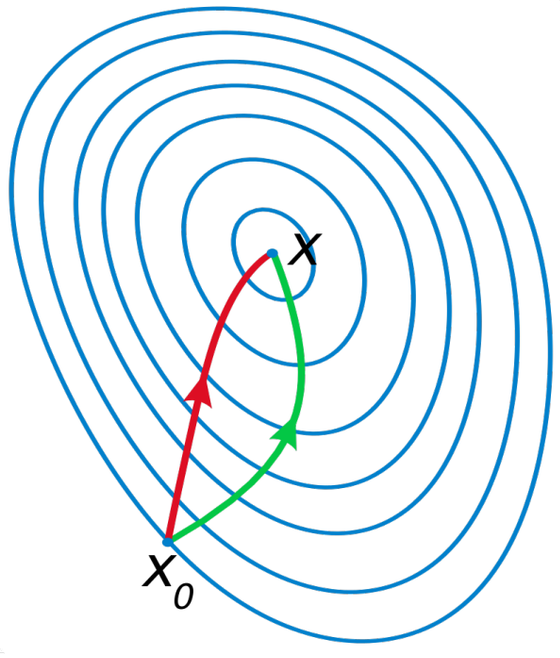
If novel ligand:

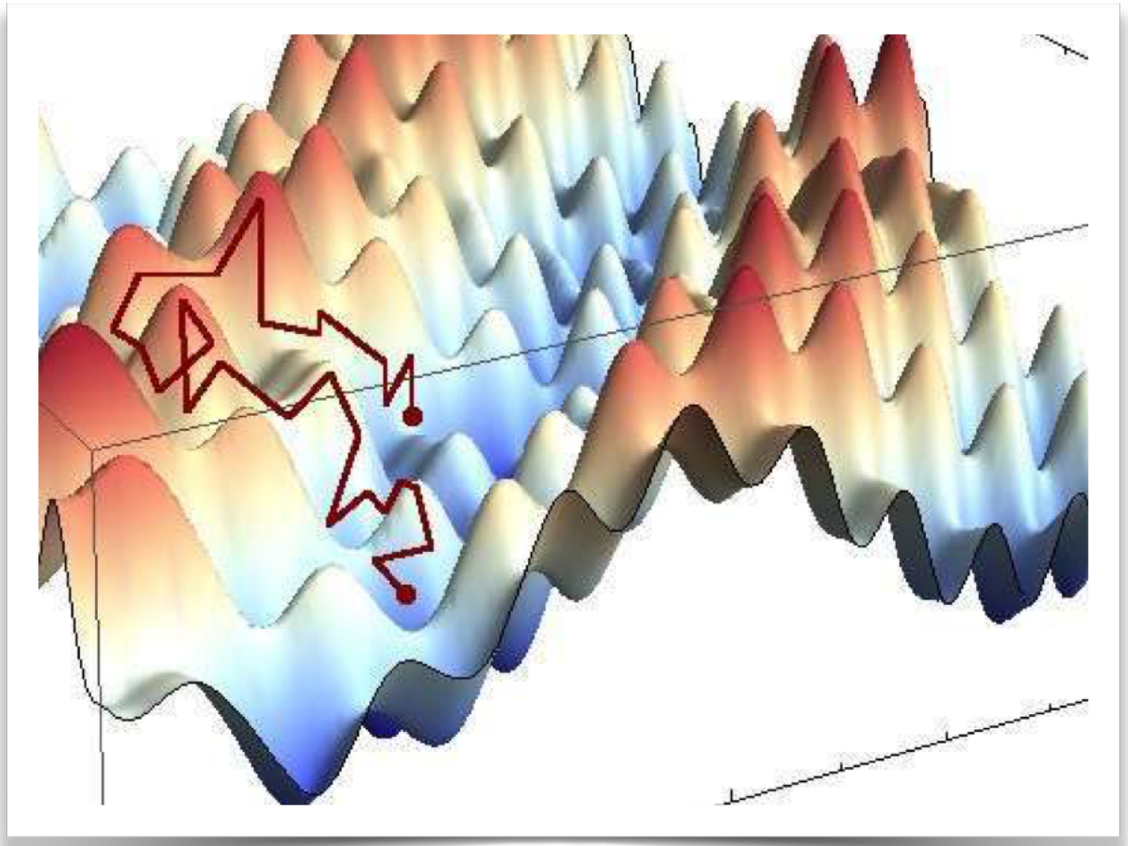
- **eLBOW**
  ligand builder
- **ReadySet**!
  Wrapper for eLBOW with additional features (add H atoms)
- **REEL**
  GUI for editing restraints

Refinement: optimization algorithm

Minimization:
Follow the local gradient



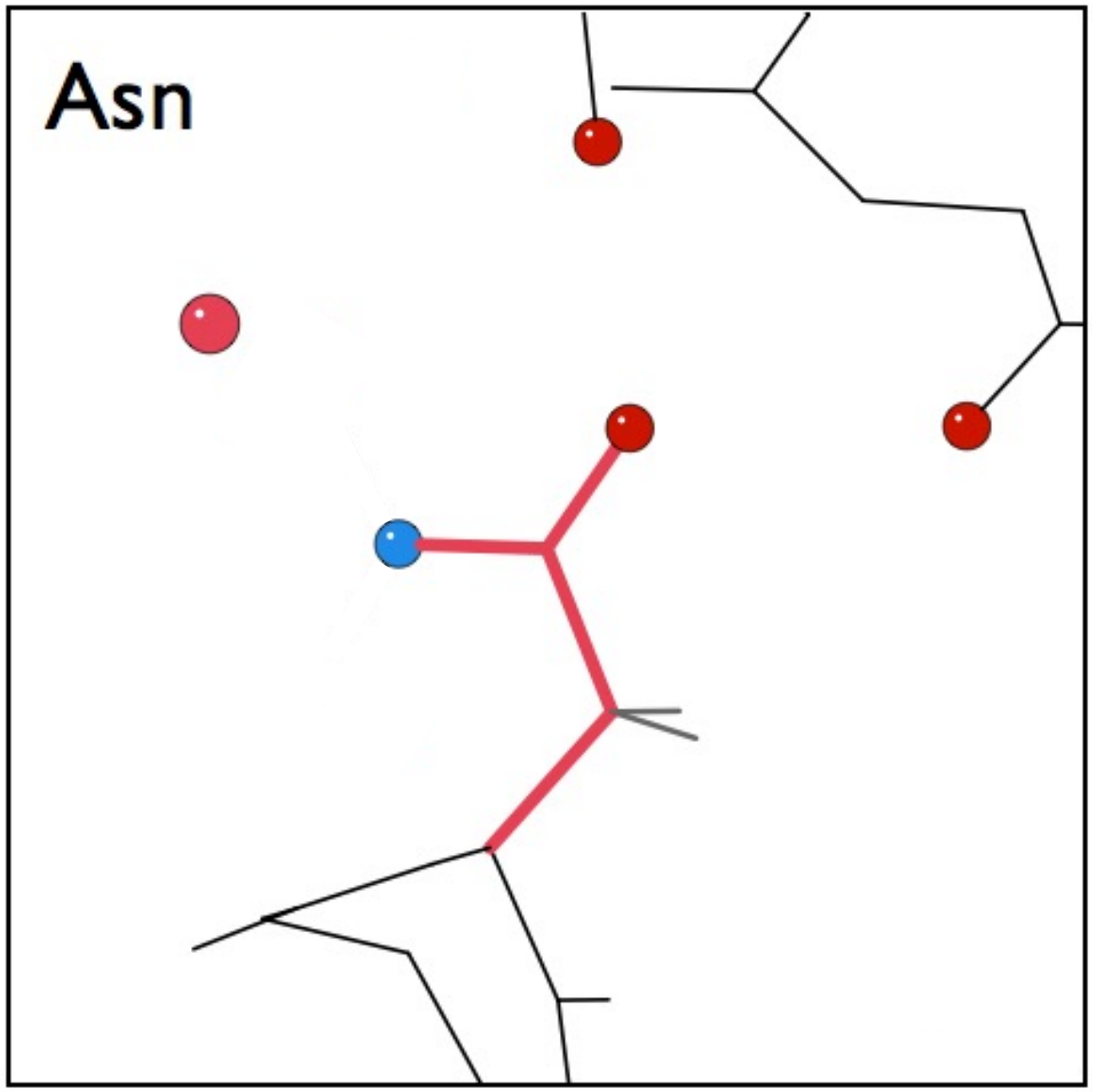


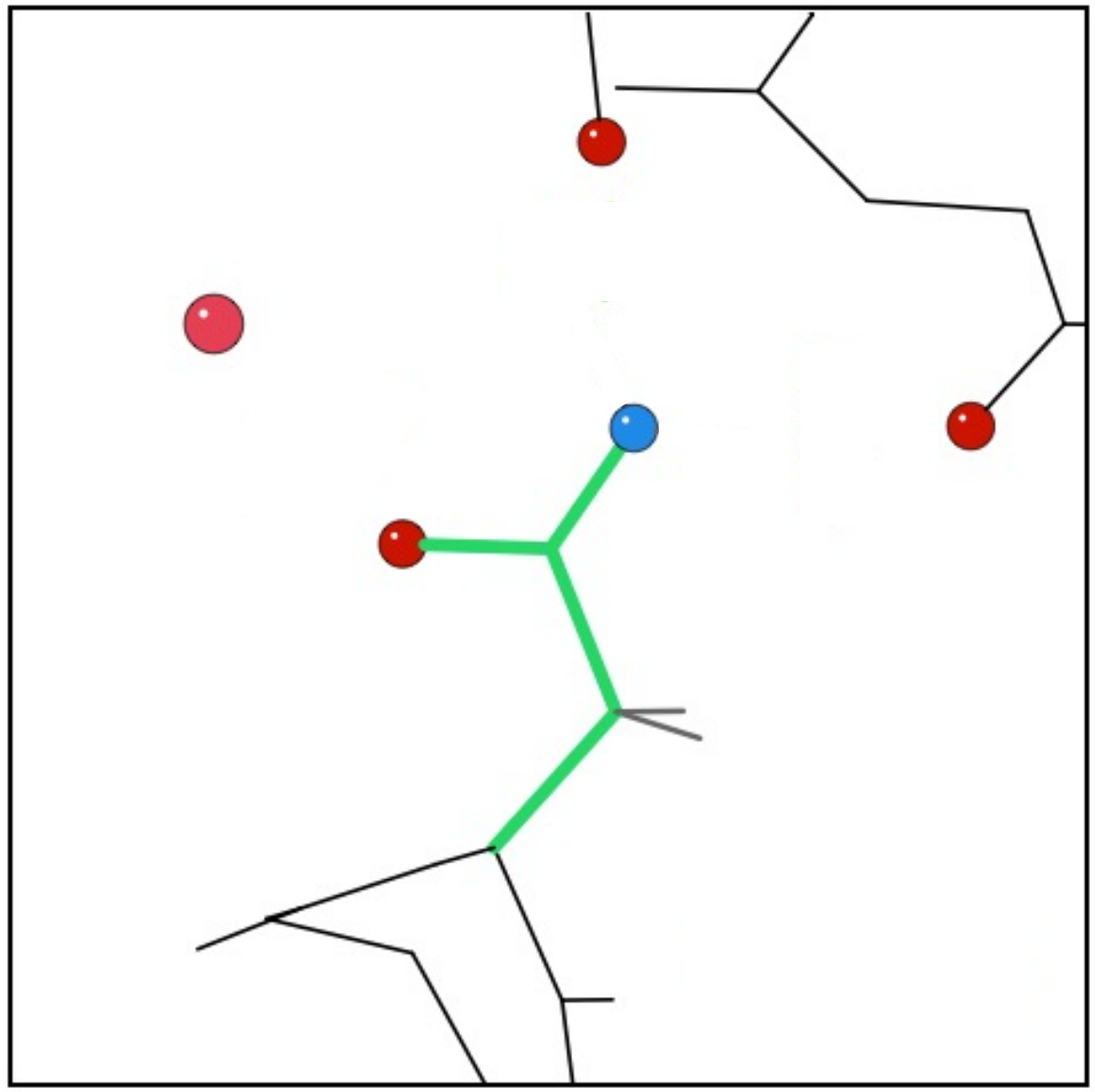


Simulated annealing:
Simulates heating up a system and slowly cooling it down, as a way of escaping local energy minima trapping

Other features

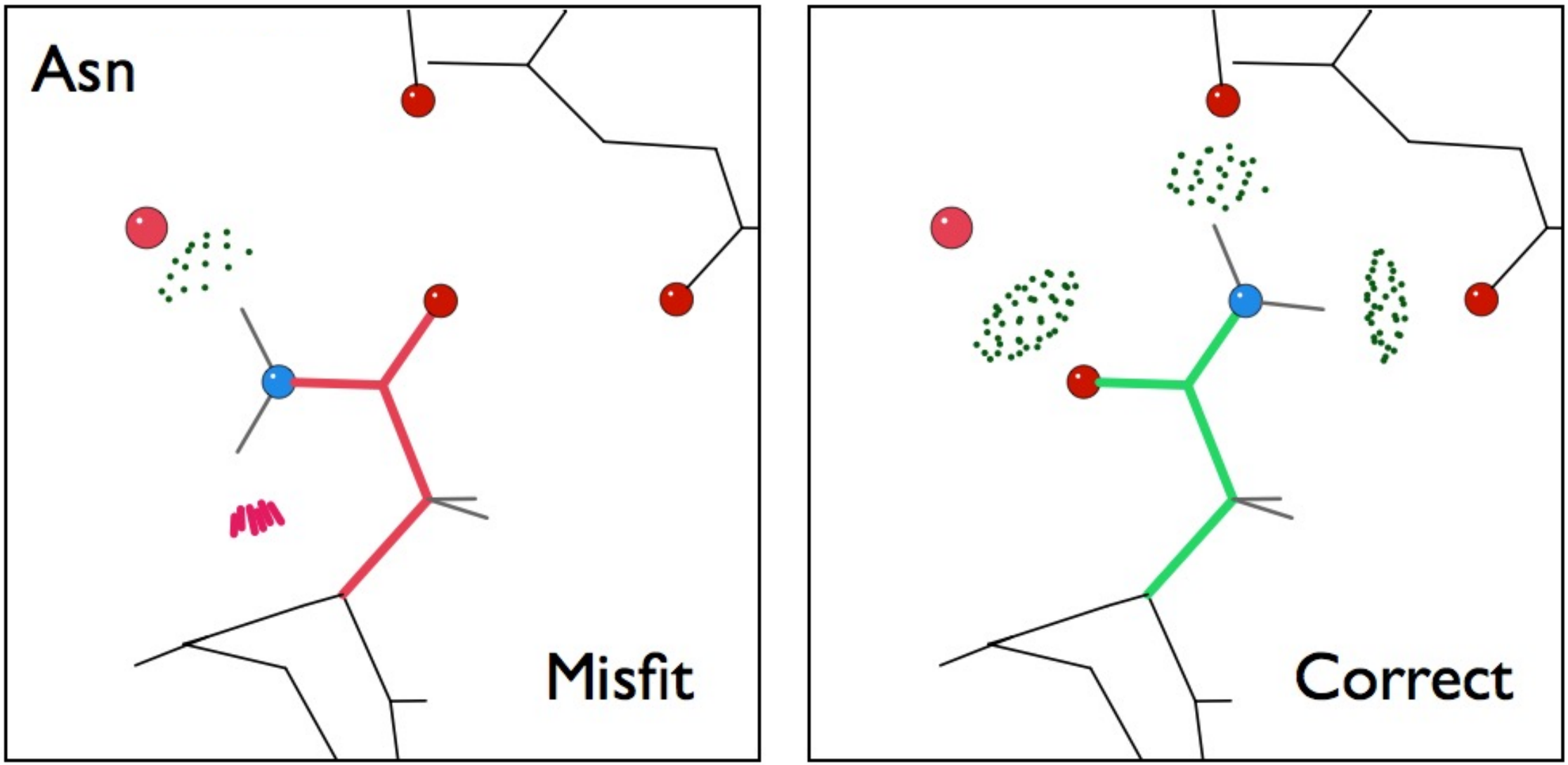
Asn/Gln/His flips





Other features

Asn/Gln/His flips



- Automatically detect and correct flipped N/Q/H residues
- Uses Molprobity and Reduce methodology

Other features

Automatic water update

Add/remove/refine automatically as part of refinement cycle

→ No need to do it as a separate step using external tools

Remove “bad” water:

- 2mFo-DFc (peak height)
- distances
- map correlation
- B-factors and anisotropy
- occupancy

Add new:

- difference map
- distances

Needs to be activated!

Validation

Should be done after **each** refinement run.

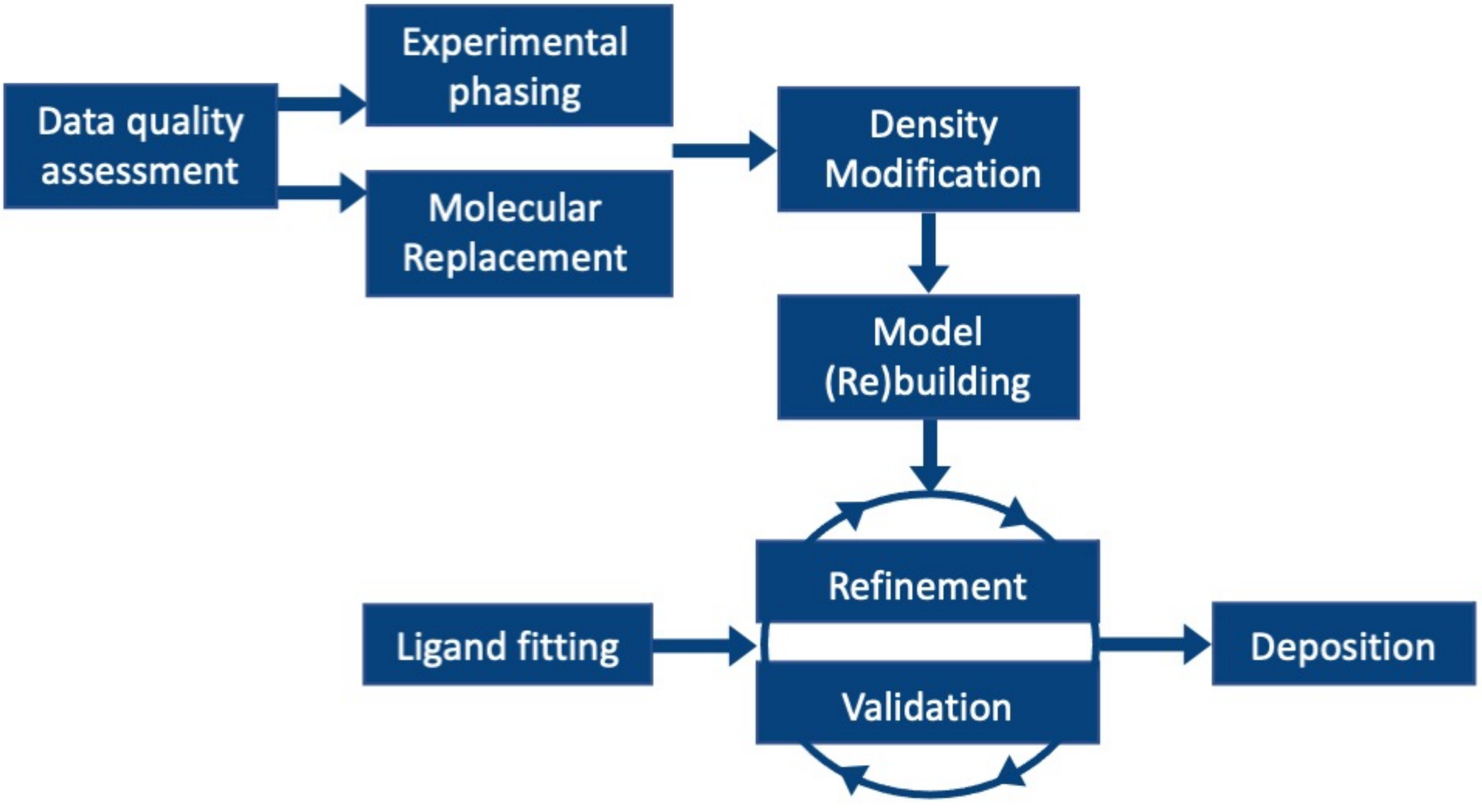
Why?

- Minimization did not converge to global minimum
- Software is not perfect (bugs)
- Double-check parameterization
- Look at problematic areas

Phenix.refine GUI makes the task easy: integration with COOT

(see demo of phenix.refine)

Steps in crystallography



The Phenix Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Dorothee Liebschner, Nigel Moriarty, Billy Poon, Christopher Schlicksup, Oleg Sobolev

Los Alamos National Laboratory
New Mexico Consortium

Tom Terwilliger, Li-Wei Hung

UTHealth

Matt Baker, Corey Hyrc

University of Cambridge

Randy Read, Airlie McCoy, Tristan Croll, Claudia Millán Nebot, Rob Oeffner

Duke University

Jane & David Richardson, Christopher Williams, Vincent Chen

An NIH/NIGMS funded Program Project

Liebschner D, *et al*., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Cryst. 2019 **D75**:861–877