CBMS workbench (virtual), Oct 13 2021

Tools for crystallography

Dorothee Liebschner
Lawrence Berkeley Laboratory
Steps in crystallography

Data quality assessment → Experimental phasing → Molecular Replacement → Density Modification → Model (Re)building

Density Modification → Refinement

Refinement → Validation → Deposition

Ligand fitting
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).

Click on panels to explore the results and investigate problems.

Please inspect all individual results closely, as it is difficult to automatically detect all issues.
Steps in crystallography

Data quality assessment → Experimental phasing → Molecular Replacement → Density Modification → Model (Re)building → Refinement → Validation → Deposition
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.

\[
\rho(\mathbf{r}) = FT \left( \tilde{F} (\mathbf{H}) \right) = \frac{1}{V} \int \tilde{F} \cdot e^{-2i\pi \mathbf{H} \cdot \mathbf{r}}
\]
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:

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

\( \phi \) is lost: phase problem

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Phasing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIR</td>
<td>Single isomorphous replacement</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple isomorphous replacement</td>
</tr>
<tr>
<td>SAD</td>
<td>Single-wavelength anomalous diffraction</td>
</tr>
<tr>
<td>native SAD</td>
<td>SAD based on native sulfurs</td>
</tr>
<tr>
<td>MAD</td>
<td>Multiple-wavelength anomalous diffraction</td>
</tr>
<tr>
<td>SIRAS</td>
<td>Single isomorphous replacement with anomalous scattering</td>
</tr>
<tr>
<td>MIRAS</td>
<td>Multiple isomorphous replacement with anomalous scattering</td>
</tr>
<tr>
<td>RIP</td>
<td>Radiation damage induced phasing</td>
</tr>
</tbody>
</table>
Experimental Phasing with AutoSol

1. Determine the substructure
2. Calculate Phases
3. Improve phases, find NCS, build model

Structural model
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

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
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

![Graph showing the relationship between resolution (Å) and AutoBuild R and FreeR values. The graph includes markers for R and Rfree at various resolutions.]
Steps in crystallography

1. Data quality assessment
2. Experimental phasing
3. Molecular Replacement
4. Density Modification
5. Model (Re)building
6. Refinement and Validation
7. Deposition
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

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

\[
F_{\text{CALC (ATOMS)}}(h,k,l) = \sum_{n=1}^{\text{Natoms}} q_n f_n(s) \exp\left(-\frac{B_n s^2}{4}\right) \exp(2i\pi r_n s)
\]
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)

\[ T = T_{\text{DATA}}(F_{\text{OBS}}, F_{\text{MODEL}}) \]
Refinement: Target function

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

Experimental data

F

A priori knowledge about model
Covalent geometry:

$\text{C} \equiv \text{C}$

$121.3^\circ$

$108.7 \text{ pm}$

$133.9 \text{ pm}$

Similarity of B-factors

$T = T_{\text{DATA}}(F_{\text{OBS}}, F_{\text{MODEL}}) + wT_{\text{RESTRANTS}}$
Restraints: *a priori* knowledge

**Chemistry**

\[
\Sigma_{\text{bonds}} \omega (d_{\text{model}} - d_{\text{ideal}})^2
\]

\[
\Sigma_{\text{dihedrals}} \omega (1 + \cos(n\chi_{\text{model}} + \chi_{\text{shift}}))
\]

\[
\Sigma_{\text{angles}} \omega (\theta_{\text{model}} - \theta_{\text{ideal}})^2
\]

\[
E = \Sigma_{\text{planes}} \Sigma_{\text{atoms}} w \ (m \cdot r - d)^2
\]

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

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:
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

Data quality assessment

Experimental phasing

Molecular Replacement

Density Modification

Model (Re)building

Refinement

Validation

Ligand fitting

Deposition
An NIH/NIGMS funded Program Project

The Project

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

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

Los Alamos National Laboratory
New Mexico Consortium
Tom Terwilliger, Li-Wei Hung

UTHealth
Matt Baker, Corey Hyrc

Duke University
Jane & David Richardson, Christopher Williams, Vincent Chen