Macromolecular X-ray Crystallography

Amie Boal
Northwestern University
Rosenzweig Lab
References


• Hampton Research Catalog, volume 18


• http://skuld.bmsc.washington.edu/scatter
The real experts...

Amy Rosenzweig

Tom Lawton
Why X-ray Crystallography?

• Widely-used method for molecular-level 3-D structure determination of biomolecules (proteins, nucleic acids)
Why X-ray Crystallography?

• For bioinorganic chemists:
  • Visualize how biomolecules interact with metals
  • Structures of complex metallocofactors
  • Identification of metal centers
  • Stoichiometry

Why X-ray Crystallography?

- Limitations
  - Oxidation states often unknown
  - Can be affected by X-ray exposure
  - Occupancy of metal binding sites can be altered
  - An ensemble technique

A crystallographic experiment

Protein  Crystals  X-ray diffraction  Electron density map and Molecular model

Why use this approach?
Microscopy analogy

Object

Light

Diffracted light rays

Image

© 2006 Academic Press
Diffraction

• Bending or scattering of waves around obstacles

• Object and wavelength ($\lambda$) must be of similar size

• Scattered waves can interfere constructively or destructively

double slit experiment
Diffraction and structure

• In microscopy, light is diffracted by objects and is focused by lenses to create a magnified image

• $\lambda$ of incident light $\sim$ object

• Visible light is 400-700 nm

• Bonded atoms are 0.1 nm ($\sim$1 Å)

• Appropriate $\lambda$ is in X-ray range (0.1-100 Å)
Two problems:

1. Single molecules diffract X-rays weakly

2. X-rays can’t be focused by lenses
A crystallographic experiment

- Crystallization
- Diffraction data collection
- Structure solution
- Model building and refinement
An X-ray diffraction experiment

APS
Argonne National Lab

Rigaku in-house generator

X-ray tube

Direct X-ray beam

Crystal

Diffracted X-rays

Film

Reflections
I. Crystallization

- A macromolecular crystal is an ordered 3D array of molecules held together by weak interactions
- Made up of unit cells
- Cells are the repeating unit
I. Crystallization

- Lattice is array of unit cells
- Crystallographic experiment provides an image of average electron density in a unit
I. Growing crystals

- Theory: supersaturation, nucleation, growth
- Practice: slow controlled precipitation from solution without unfolding the protein
I. Growing crystals

• Supersaturation can be achieved by addition of precipitating agents
  • Salts (ammonium sulfate, NaCl)
  • Polymers (polyethylene glycol)
  • Organic solvents (alcohols, 2-methyl-2,4-pentanediol)
I. Growing crystals

• Nucleation and growth
  • An aggregate of sufficient size and order initiates crystal growth
  • Usually a random event but techniques exist to drive nucleation (i.e., seeding)
  • Additives can influence growth and nucleation
I. Growing crystals

- Other variables
  - Protein purification, preparation, stability
  - Structural homogeneity
  - Temperature
  - pH
I. Improved purification

N-terminal His-tagged *B. subtilis* NrdF
old prep

- 20-30% PEG 4000
- 0.2-0.4 M lithium sulfate
- 0.1 M HEPES pH 6.5-7.0

N-terminal His-tagged *B. subtilis* NrdF
new prep
+ MonoQ purification step

same crystallization conditions
I. Growing crystals

• Other variables
  • Protein purification, preparation, stability
  • Structural homogeneity
  • Temperature
  • pH
I. Growing crystals

- Other variables
- Protein purification, preparation, stability
- Structural homogeneity
- Temperature
- pH

M. capsulatus (Bath) pMMO
I. Growing crystals

- can’t predict crystallization conditions
- extensive screening by trial and error required
I. Growing crystals

Vapor diffusion is the most commonly used method

\[ \text{[ppt]}_{\text{drop}} = \frac{\text{[ppt]}_{\text{reservoir}}}{2} \]

**Figure 1**

**Figure 2**

Sitting drop

Hanging drop

low [ppt]

high [ppt]
I. Growing crystals

• Sparse matrix screens are typical first step

• Diverse sets of conditions based on database mining of published crystallization conditions

• Intentionally biased screen towards conditions that have worked previously

• Commercially available, 96-well format
I. Once you have crystals...

- Are they protein crystals?
- Are they crystals of the protein of interest?
- Is it a single crystal of sufficient size?
- Can the crystals be harvested and frozen?
- Do the crystals diffract X-rays?
- Can enough data be collected to solve the structure?
I. Optimization

Secondary screens:
24-well format
Homemade screens
pH vs. [ppt] is a typical starting point
Other variables come into play
1. Crystal lattices

- A crystal lattice is made up of unit cells
- Unit cells have a defined shape
- $x, y, z$ coordinate system
I. Lattice types

14 different lattice systems

Defined by unit cell shape

- $a = b = c$ relationships are identities

If $a = b$, unit cell contents along axes are identical

Can have internal symmetry
  - multiple copies of protein in unit cell
  - higher order oligomerization state related by symmetry

Final model only describes structure of asymmetric unit
I. Symmetry operations

Symmetry operations classify lattices into space groups

230 different space groups

Chiral molecules restricted to 65 possible space groups

Assigning the correct space group is really important!
A crystallographic experiment

- Crystallization
- Diffraction data collection
- Structure solution
- Model building and refinement
II. A diffraction experiment

- Harvest and freeze crystals
- Take to X-ray source
- Screen for diffraction
- Collect datasets
II. A diffraction experiment
II. Diffraction data

• Each spot is a reflection with an index
• Indices tell us lattice type and unit cell dimensions

(indices increase)

(0,0,0) ➔ (h,k,l) index, intensity
II. X-ray diffraction

• In single crystal diffraction, only diffracted X-rays with strong constructive interference are observed
II. X-rays are waves

- Have an amplitude, phase, frequency

- Constructive interference increases amplitude

\[ f(x) = F \cos 2\pi(hx + \alpha) \]
II. Bragg’s law

• Parallel planes of atoms in crystals produce diffracted X-rays with strong constructive interference

• Conditions that satisfy Bragg’s law lead to diffraction

\[ 2d_{hkl} \sin \theta = n\lambda \text{ in phase} \]
II. Planes??

- A crystal lattice can be divided into regularly spaced sets of parallel planes
- Indexed based on number of times planes intersect the unit cell on each a,b,c edge
- Index notation is h, k, l and is the same as the index of corresponding reflection in the diffraction pattern
II. Examples of lattice planes
II. Planes II

• Each set of planes gives rise to a single reflection

• Reflections form a reciprocal lattice

• The reciprocal lattice is what we see in diffraction data - but related to crystal lattice via indices of lattice planes.
II. Real vs. reciprocal space

In crystals...

\[ a = \frac{1}{a^*} \]

\[ b = \frac{1}{b^*} \]

\[ c = \frac{1}{c^*} \]
II. Diffraction patterns

• Reciprocal lattice spacing is related real lattice spacing

• Can calculate unit cell dimensions from h,k,l indices alone

• Position of spots in diffraction pattern contains information about unit cell and its symmetry
II. What about intensities?

- Intensity reports on amount of electron density associated with given set of planes
- Electron density associated with every atom in the unit cell contributes to the intensity of a single reflection
II. Diffraction data

- What do we look for in a diffraction pattern?
  - resolution limit
  - spot shape
  - single lattice

(h,k,l) index, intensity
Multiple vs. single lattice
II. Data collection

• What is a dataset?
  • Set of diffraction pattern images collected while rotating the crystal
  • Goal is to collect a complete and redundant set of reflections and their intensities
Crystal rotation during data collection captures unrecorded reflections
II. Data collection variables

• X-ray exposure time
  • Avoid overexposure/saturation
• Crystal-to-detector distance
  • Use resolution limit as guide
• Frame width ($\omega$)
  • Partial vs. full reflections and overlaps
• How much data to collect?
II. How much data to collect?

- # of unique reflections depends on internal symmetry of lattice
- Higher symmetry = less data
- $180^\circ$ is a complete set
- Friedel’s law: $I_{hkl} = I_{\overline{h}k\overline{l}}$
- exceptions...
II. Data processing

• Peak search
• Indexing
• Integration
• Scaling
• Postrefinement

Final output is a list of reflections \((h, k, l)\), and intensities \((I)\)
III. Structure solution

- Structure factors
- Each reflection can be described by a structure factor \( (F_{hkl}) \) function
- \( F_{hkl} \) is the sum of diffractive contributions of all atoms in the unit cell

\[
F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)}
\]
- \( f_j \) is the atomic scattering factor
- \( hx_j + ky_j + lz_j \) is the phase of the diffracted ray
III. Structure solution

- Structure factors can be expressed as sum of contributions from volume elements in crystal

- Electron density of volume element centered on \((x, y, z) = \rho(x, y, z)\)

\[
F_{hkl} = \int_{V} \rho(x, y, z) e^{2\pi i (hx_j + ky_j + lz_j)} dV
\]
III. Structure solution

- Electron density map is a 3D plot of $\rho(x, y, z)$
- Fourier sum

\[
\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F_{hkl} e^{-2\pi i(hx_j + kx_j + lx_j)}
\]

$F_{hkl}$ is the structure factor, which is a sum over all reflections in the diffraction pattern.
III. Structure solution

• How do we compute $\rho(x, y, z)$ from diffraction data?

• $F_{hkl}$ is a periodic function and has an amplitude, frequency, and phase

  • amplitude $\sim \sqrt{I_{hkl}}$
  
  • frequency $= \frac{1}{d_{hkl}}$
  
  • phase = ??
III. Solving the phase problem

• Three common methods
  • Molecular replacement
  • Isomorphous replacement
  • Anomalous scattering
III. Solving the phase problem

- Molecular replacement
  - Use similar protein of known structure to compute phases
  - Structure can be solved from single, native dataset
III. Solving the phase problem

- Isomorphous replacement and anomalous scattering
- Make use of heavy atom sites (i.e., metals) in crystals (derivative or native)
- Typically requires collection of multiple datasets
III. Anomalous Scattering

- X-ray absorption by heavy atoms alters diffraction
- Friedel’s law does not hold \( I_{hkl} \neq I_{hkl} \)
- Must be able to tune \( \lambda \) to absorption edge
- Intensity difference in Friedel pairs locates heavy atom sites in unit cell - used to compute phases
- Additional datasets are collected near heavy atom absorption edge
- Requires tunable X-ray source and data must be highly redundant with minimal X-ray damage
III. Anomalous scattering

• Can also be used to identify metals in crystal structures
IV. Model generation

• 3D plot of $\rho(x, y, z)$ function produces electron density map

• Phase improvement

• Map interpretation

• Refinement of coordinates

\[
\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i (hx_j + ky_j + lz_j - \alpha'_{hkl})}
\]

phases

amplitudes (from intensities)
Map Interpretation

Blue = 2F_o-F_c map
Red/Green = F_o-F_c map

1.8 Å resolution map
Map Interpretation

Blue = $2F_o - F_c$ map

Red/Green = $F_o - F_c$ map

1.8 Å resolution map
IV. Map interpretation

- 3D plot of $\rho(x, y, z)$ function produces electron density map
- Phase improvement
- Map interpretation
- Refinement of coordinates

$$
\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| e^{-2\pi i (hx_j + ky_j + lz_j - \alpha'_{hkl})}
$$

*phases*

*amplitudes (from intensities)*
IV. Model evaluation

- Data collection statistics
  - Resolution
  - $R_{\text{sym}}/R_{\text{merge}} < 0.1$
  - $I/\sigma I > 2$
  - Completeness/redundancy ~ 100%
# IV. Data collection statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Pt-Atox1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td><em>P3</em>$_2$1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>54.07, 54.07, 55.58</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (°)</td>
<td>90.00, 90.00, 120.00</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.00-1.60 (1.63-1.60)</td>
</tr>
<tr>
<td>$R_{\text{sym}}$ or $R_{\text{merge}}$</td>
<td>0.118 (0.503)</td>
</tr>
<tr>
<td>$I / \sigma I$</td>
<td>24.9 (3.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7 (99.7)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>19.5 (9.5)</td>
</tr>
</tbody>
</table>
IV. Model evaluation

• Model refinement statistics
  • R-factor
  • $R_{\text{free}}$ statistic
  • RMS deviation (bond lengths, angles)
• Ramachandran statistics
## IV. Data collection statistics

<table>
<thead>
<tr>
<th>Refinement</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>46.83-1.60</td>
</tr>
<tr>
<td>No. reflections</td>
<td>12060</td>
</tr>
<tr>
<td>$R_{\text{work}} / R_{\text{free}}$</td>
<td>0.179/0.210</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>496</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>21</td>
</tr>
<tr>
<td>Water</td>
<td>92</td>
</tr>
<tr>
<td>$B$-factors</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>17.0</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>33.9</td>
</tr>
<tr>
<td>Water</td>
<td>39.6</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>1.252</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.010</td>
</tr>
</tbody>
</table>
IV. Model evaluation

- Considerations for structures with transition metals
  - Metal ion occupancy
  - Oxidation state and influence by X-rays
  - Ligand order and occupancy
  - Anomalous scattering data

Evaluate by inspecting electron density maps
Three case studies...

- Crystallographic site assignment in the Mn/Fe cofactor of class Ic RNR
Class I $\beta$2 structures

Class Ia

Class Ib

Class Ic
Three case studies...

- Crystallographic site assignment in the Mn/Fe cofactor of class Ic RNR
Mn/Fe anomalous scattering

Initial crystal conditions

New crystal conditions

Mn edge 1.85 Å
site 1 binds Mn, Fe (or Pb)

Fe edge 1.7 Å
site 2 binds Fe

Site occupancy and loading

- Apo protein loaded with 1.5 eq Mn$^{ll}$ and < 0.5 eq Fe$^{ll}$ + O$_2$
- <1 eq Mn$^{ll}$ and 1.5 eq Fe$^{ll}$ + O$_2$

Three case studies...

- Structure of FeMoCo in nitrogenase
2.75 Å resolution

1.16 Å resolution
Three case studies...

- Identification of the metal sites in pMMO
pMMO X-ray structures

1 Cu

2 Cu → XAS

1 Zn?

M. caps

sp. M

0.2 M ZnAc

CH₄ 2-15 Cu? → CH₃OH

O₂

OB3b

1 Cu

Tom Lawton
Acknowledgements

Amy Rosenzweig  Tom Lawton  Jason Xu
Elle Moore  Paula Ham  Gaby Budzon
Sara Sperling  Kate Kurgan  Tianlin Sun