Data Collection

Overview

with a unit cell, possible space group and computer reference frame (orientation matrix); the location of diffracted x-rays can be calculated \((h k ℓ)\) and intercepted by something sensitive to x-rays (film or electronic devise)

integrated intensity: total number of photons from \((h k ℓ)\)

Methods

1. film Weissenberg and precession photos
   measure intensity relative to a calibration strip of film

2. automatic diffractometer
   a. single-point detector
      computer-controlled crystal orientation so diffracted beam goes into the detector – scintillation counter

      energy is transferred to Tl atoms that produce a flash of light; light intensity is proportional to x-ray intensity; photocathode converts the light to voltage

Counter Methods

Crystal Quality with \(ω\) - Scans

Crystal Quality with \(ω\) - Scans

good crystal partial twin twin not centered

Lorentzian
Data Collection

1. scan method usually $\varphi$ or $\omega - 2 \theta$ ($\theta - 2 \theta$).

2. scan rate usually variable depending on intensity
   weak reflections scanned more slowly
   fast pre-scan to determine best rate
   4.0 to 29.3°/min

3. background peak scanned on either side to fix baseline
   scan range 1° below $K_{\alpha 1}$ to 1° above $K_{\alpha 2}$

4. maximum 2$\theta$ depends on instrument, quality, and size of
crystal; intensities go down with 2$\theta$
   usually 45 – 60°

Resolution

protein crystallographers usually use resolution:

$$\frac{\lambda}{\sin \theta}$$

same information as maximum 2$\theta$, but refers to the resolution of
detail (atoms) in the final structure

for small-molecule studies the data are collected to $\lambda / \sin \theta$ of at
least 1.7 Å ($\theta_{\text{max}} > 25°$ for MoK$_\alpha$, $\theta_{\text{max}} > 67°$ for CuK$_\alpha$)

Intensities

intensity, $I$, is measured in counts (photons)

$$I = rs(RB)$$

$r$ = scan rate, °/min
$s$ = total scan count
$R$ = scan to background time ratio
$B$ = background count

Standard Deviation

can also calculate $\sigma I$:

estimated standard deviation

$$\sigma I^2 = r^2(s + R^2B)$$

used in weighing reflections
Decay Correction

x-rays can damage crystal while data collected minimize exposure by turning off X-ray beam when crystal is repositioned between data points when decay occurs, data usually adjusted with a linear correction 3 or more check reflections are remeasured periodically (every xxx reflections, every x hours of exposure) throughout collection fit to a decay curve and scale reflections accordingly

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>_diffrn_standards_number</td>
<td></td>
</tr>
<tr>
<td>_diffrn_standards_interval_count</td>
<td></td>
</tr>
<tr>
<td>_diffrn_standards_interval_time</td>
<td></td>
</tr>
<tr>
<td><em>diffrn_standards_decay</em>?</td>
<td></td>
</tr>
</tbody>
</table>

Data

each data point consist of a number of parameters:

- profile $I_\sigma_1, h, k, l, 2\theta, \omega, \chi, \phi$
- left and right background

measure 1000 – 4000 reflections depending on:

crystal size, contents, symmetry

Methods

1. film Weissenberg and precession photos
   measure intensity relative to a calibration strip of film
2. automatic diffractometer
   a. single-point detector
      computer-controlled crystal orientation so diffracted beam goes into the detector – scintillation counter
   b. area detector
      image plates, multi-wire proportional counters, and charge-coupled devises (CCD)
      detail of data collection depend on manufacturer

Image Plate

Multitwire Proportional Counter

CCD
Symmetry
amount of independent data limited by:
sphere of reflection
weak reflections at large 2θ
symmetry – not all data independent
1. Friedel’s Law – at least half the reflections redundant
   \[ I_{\text{hkl}} = I_{\text{Lkl}} \]
2. additional redundancies based on the crystal system
   monoclinic \[ I_{\text{hkl}} = I_{\text{Lkl}} = I_{\text{Lhl}} = I_{\text{Llh}} = I_{\text{hkl}} \]
   orthorhombic: \[ I_{\text{hkl}} = I_{\text{Lkl}} = I_{\text{Lhl}} = I_{\text{Llh}} = I_{\text{hkl}} = I_{\text{Lkl}} \]
   triclinic

Collecting Symmetry Related Reflections
if the crystal allows it (decay not a problem) and you have time,
symmetry related reflections can be measured and averaged
improves statistics and gives an idea of crystal quality
\[ \Delta(I)/ \langle I \rangle \] for symmetry-equivalent reflections
\[ \Delta(I) \] is the average absolute difference between \( I \) (average I) and the individual symmetry-equivalent \( I \)'s

equivalent reflections are averaged to during Data Reduction

Bottom Line on Data Collection
obtain the most, good (usable), unique data possible
more data means better statistical determination of structure

Data Reduction
solving a structure means to rephase the scattered x-rays (lens),
which requires the amplitudes and phases of the scattered x-rays
for any wave:
amplitude is proportional to the square root of the intensity
amplitudes derived from x-ray data are called:
Structure Factors \( F_{\text{calc}} \)
\[ |F_{\text{calc}}| \propto \sqrt{T} \]

\( F \)'s are converted to \( F_{\text{obs}} \) or \( F_{\text{e}} \) with several corrections based on decay, absorption and the geometry of data collection

Data Reduction
Decay

using check reflections, the decay rate can be estimated:

\[ y = -3.7791x + 1233.8 \]

Empirical Absorption Correction – \( \psi \) scans

find several reflections with \( \chi \) near 90°; rotate crystal around \( \phi \) and measure how the intensity changes

Absorption Correction Cif Definitions

<table>
<thead>
<tr>
<th>absorption</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>analytical</td>
<td>analytical from crystal shape</td>
</tr>
<tr>
<td>cylinder</td>
<td>cylindrical</td>
</tr>
<tr>
<td>empirical</td>
<td>empirical from intensities</td>
</tr>
<tr>
<td>gaussian</td>
<td>Gaussian from crystal shape</td>
</tr>
<tr>
<td>integration</td>
<td>integration from crystal shape</td>
</tr>
<tr>
<td>multi-scan</td>
<td>symmetry related measurements</td>
</tr>
<tr>
<td>none</td>
<td>no absorption correction applied</td>
</tr>
<tr>
<td>numerical</td>
<td>numerical from crystal shape</td>
</tr>
<tr>
<td>psi-scan</td>
<td>( \psi ) scan corrections</td>
</tr>
<tr>
<td>refdelf</td>
<td>refined from ( \Delta F )</td>
</tr>
<tr>
<td>sphere</td>
<td>spherical</td>
</tr>
</tbody>
</table>

Geometric Corrections – \( L_p \)

\[ |F_{\text{abs}}| = \sqrt{\frac{K |F_{\text{mat}}|}{L_p}} \]

K scale factor: depends on crystal size and x-ray intensity
estimated early and refined at the end

L. Lorentz factor: time for a reflection to pass through Ewald’s Sphere depends on the 2\( \theta \) angle

Lorentz Factor - \( L \)

for a single-point diffractometer:

\[ L = \frac{1}{\sin 2\theta} \]
Polarization - $p$

$$p = \frac{(1 + \cos^2 \theta)}{2}$$

Geometric Corrections – $L_p$

$$|F_{sat}| = \sqrt{\frac{K_I}{L_p}}$$

$K$ scale factor. depends on crystal size and x-ray intensity. estimated early and refined at the end

$L$. Lorentz factor time for a reflection to pass through Ewald’s Sphere depends on the 2θ angle

$p$. polarization reflected x-ray is partially polarized usually called the $1/L_p$ correction

Structure Factors

the idea in solving a crystal structure:

1. measure reflections and reduce data to $F_o = \sqrt{\frac{K_I}{L_p}}$

   $h\ k\ t\ I\ \sigma_I$ $\Rightarrow$ $h\ k\ t\ F\ \sigma_F$

2. “guess” structure; calculate structure factors that would result from that structure: $F_e$

3. compare $F_o$ and $F_e$; improve “guess”, calculate new structure factors ($F_e$) that result from improved structure

4. continue process until $F_o$ and $F_e$ match within statistically acceptable error and structure is chemically reasonable

Scattering Factors

to calculate $F_e$, it is necessary to know the scattering power of each atom in the cell and how it contributes to structure factor

Scattering Factors

function of:

1. number of $e^-$ in atom
2. $\theta$ scattering angle (Bragg)

Scattering Factors – Number of $e^-$

x-ray hitting $e^-$ causes it to oscillate at same frequency wavelet emitted with amplitude $\alpha$ to oscillation magnitude each $e^-$ wave adds to scattered x-ray from atom

gives a scattered wave with an amplitude $\alpha$ to total $e^-$

Scattering Factor Curve of Carbon

$sin\theta/\lambda = 1/2d$
Scattering Factors – \( \theta \) Dependence

Scattering factor would always be atomic number of the atom if all e\(^-\) at a point; however, a real atom has volume.

in phase if \( \theta = 0^\circ \)

increasingly out of phase at \( \theta > 0^\circ \)

---

Scattering Factor Calculation

\[
 f(\sin\theta/\lambda) = \sum_{i=1}^{6} a_i e^{-b_i(\sin\theta/\lambda)^2}
\]

Atom
\[
\begin{align*}
 a_1 & \quad a_2 & \quad a_3 & \quad a_4 & \quad a_5 & \quad a_6 \\
 b_1 & \quad b_2 & \quad b_3 & \quad b_4 & \quad b_5 & \quad b_6
\end{align*}
\]

Carbon
\[
\begin{align*}
 2.09921 & \quad 1.80832 & \quad 1.26159 & \quad 0.56775 & \quad 0.26303 & \quad 0.00010 \\
 13.18997 & \quad 30.37956 & \quad 0.69255 & \quad 0.16381 & \quad 68.42774 & \quad 0.44083
\end{align*}
\]

---

Anomalous Dispersion

if an atom absorbs x-rays before re-emitting, the phase changes; intensity is also changed

\[
f_{\text{atom}} = f_{\text{anom}} = f_o + \Delta f_{\text{atom},\lambda} + i\Delta f'_{\text{atom},\lambda}
\]

Freidel’s Law violated for noncentrosymmetric space groups numbers for every atom are part structure solution software

Multiwavelength Anomalous Diffraction (MAD) phasing

---

Temperature Factor

scattering factor based on a spherical, stationary, atoms but atoms are in constant motion; magnitude depends on mass, bond strength and temperature

thermal motion spreads e\(^-\) density over a larger volume, causing scattering to drop off more rapidly

\[
f = f_o e^{-B(\sin\theta/\lambda)^2}
\]

\[
B = 8\pi\bar{u}^2
\]

\(B\) (in Å\(^2\)) is the temperature factor or thermal parameter

---
**Temperature Factor**

\[ f = f_o e^{-B \sin(\theta)/\lambda^2} = f_o e^{-8\pi^2 U_{hk\ell}} \]

mean-square amplitude \( U_{hk\ell} \)

\[ U = B / 8\pi^2 \]

typical values for \( U \):
- isotropic: \( 0.05 \text{Å}^2 \)
- anisotropic: \( 4 \text{Å}^2 \)

**Wilson Plot**

A method to estimate overall temperature factor and scale factor

Arthur J. C. Wilson

Canadian

1914-1995

\[ I_{\text{relative}} = (F_{\text{relative}})^2 \]

\[ I_{\text{absolute}} = \sum_{i=1}^{N} f_i^2 \text{N atoms in cell} \]

Average intensity depends on what is in cell, not where it is.

\[ I_{\text{absolute}} = \sum_{i=1}^{N} f_i^2 e^{-2B\sin(\theta)/\lambda^2} \]

**Wilson Plot**

\[ I_{\text{abs}} = e^{-2B\sin(\theta)/\lambda^2} \sum_{i=1}^{N} f_i^2 \]

\[ I_{\text{rel}} = C I_{\text{abs}} \]

\[ I_{\text{rel}} = C e^{-2B\sin(\theta)/\lambda^2} \sum_{i=1}^{N} f_i^2 \]

\[ \sum_{i=1}^{N} f_i^2 = C \sum_{i=1}^{N} f_i^2 e^{-2B\sin(\theta)/\lambda^2} \]

\[ \ln \left( \frac{\sum_{i=1}^{N} f_i^2}{\sum_{i=1}^{N} f_i^2} \right) = \ln C - 2B\sin(\theta)/\lambda^2 \]

\[ y = b + m x \]

**Other Statistics**

\[ E_{\text{rel}}^2 = \frac{F_{\text{rel}}^2}{\sum_{j=1}^{N} f_j^2} \]

\( e \) = integer (usually 1) related to symmetry

- centrosymmetric
- noncentrosymmetric

average \( |E| \):
- 1.0000
- 0.968
- 0.798

average \( |E| - 1 |\)
- 0.968
- 0.798
- 0.886

|\( |E| > 1 \)| | \( |E| > 2 \)| | \( |E| > 3 \) |
|---|---|---|
|32.0% | 5.0% | 0.3% |
|36.8% | 1.8% | 0.01% |

\( Z = 4 \)

\( x, y, z \) - crystalllographic

\( x, y, z \) - approximate