Chemical Crystallography and Structural Chemistry

VO 270287

Lecture № 7 — 6th May 2021

Dr. Tim Grüne

Centre for X-ray Structure Analysis

Faculty of Chemistry

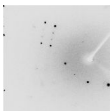University of Vienna

tim.gruene@univie.ac.at

Reminder: Start the recording!

Contents

Data Integration

Data collection ➡ **Data integration** ➡ Data Scaling


```
0 0 −1     2.7   0.9
0 0  1     4.0   1.0
0 0 −2 1'257.0  35.5
0 0 −2 1'600.0  42.7
```

```
0 0 −1     2.8   0.55
0 0  1     3.8   0.63
0 0 −2 1'432.0  95.7
0 0 −2 1'282.0  85.9
```

several GB   several files, 100's MB   1 "hkl"-file, 50MB

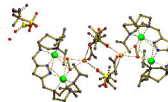➡ Phasing   ➡ Refinement
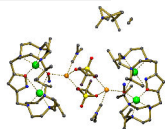
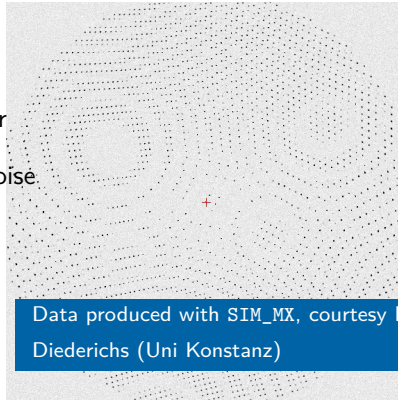Starting model   Chemically   sensible
                          model

1 Spot Intensity
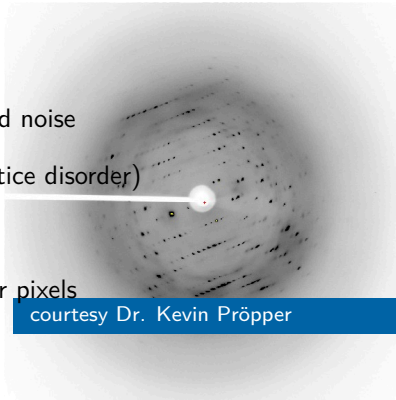
Ideal diffraction image

- no lattice disorder

- no background noise

- perfect beam

Data produced with SIM_MX, courtesy Kay Diederichs (Uni Konstanz)

Realistic diffraction image

- strong background noise

- smeary spots (lattice disorder)

- spot overlap

- saturated detector pixels

- finite resolution

courtesy Dr. Kevin Pröpper

Signal extraction

1. Calculate reflex positions

2. Determine local background

3. Differentiate background from signal (spot volume, shape)

4. Different approach for strong spots and weak spots
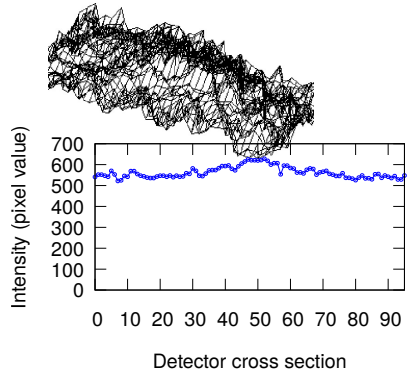
Signal extraction: strong spots

- High intensity $\Rightarrow$ Small error from noise

- Spot covers large detector area: noise approximated by average

- Good spot separation

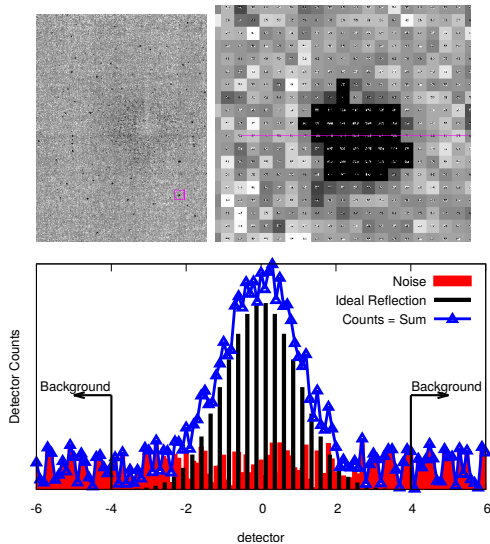- Good spot profile (shape)
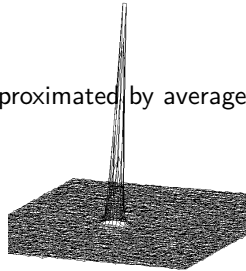
Signal extraction: weak spots

- Low intensity: $\Rightarrow$ high error from noise

- Spot covers small detector area: large effect from noise

- Background difficult to determine

Cross section of a spot on the detector

Reflection profile

Most integration programs create a set of reflection profiles from strong and reliable reflections, *e.g.* a 3D Gauss function.

The profiles depend on the region on the detector and on the crystal orientation.

Advantages:

1. Measurement of weak reflections (fitting data for profile)

2. takes non-isotropic crystal shape into accoung

3. takes regions of varying detector sensitivity into account

4. produces a standard deviation of the reflection intensity: (h, k, l, I, σ_I)
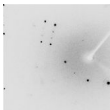
Summary Data Integration

- Starts with indexing: crystal orientation, unit cell

- Look at all images per run

- Look only at calculated spot positions on detector

- Strong spots: sum pixel values, substract background

- Strong spots: determine average reflection profile

- Weak spots: extract data based on profiles

2 Scaling

Scaling

Data collection ➡ **Data integration** ➡ **Data Scaling**


```
0 0 -1      2.7    0.9
0 0  1      4.0    1.0
0 0 -2 1'257.0   35.5
0 0 -2 1'600.0   42.7
```

```
0 0 -1      2.8    0.55
0 0  1      3.8    0.63
0 0 -2 1'432.0   95.7
0 0 -2 1'282.0   85.9
```

several GB   several files, 100's MB   1 "hkl"-file, 50MB
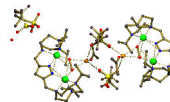
➡ Phasing
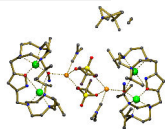
Starting model

➡ Refinement

Chemically   sensible
model

Objective of the diffraction experiment

Structure elucidation of a chemical compounds

- chemical composition (*e.g.* purity after chromatography)

- Connectivity, distances between (non-) bonded atoms

- Configuration of stereochemical centres (R,S)

The values should be independent from the experimental setup.
Scaling makes the raw intensities (from data integration) independent from
the experimental setup

Calculation of Intensities

Under consideration of the experimental, non-idealised setup, intensities are calculated as [1, Ch. Diffraction Intensities]

$$I_{exp}(hkl) \quad = \frac{e^4}{m_e{}^2 c^4} \frac{\lambda^3 V_{crystal}}{V_{u.c.}^2} I_0 LPTE |F_{theor.}(hkl)|^2 \qquad (1)$$

- I_0 incoming intensity (may vary with time)

- L Lorentz factor describes trajectory through the Eqwald sphere

- P Polarisation correction; $P = (1 + \cos^2 2\theta)/2$ for unpolarised source

- T Absorption correction (esp. heavy elements)

- E extinction correction (crystal defects, mosaicity)

- $F_{theor.}(hkl)$ structure factor (calculated from structure)
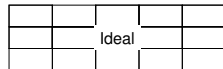
Scaling = Idealisation and correction

In order to make data as independent from the experiment as possible, data are "standardised". Some corrections are of numerical nature (polarisation), others are sample dependent (extinction: depends on elements in compounds). Two examples: absorption and angle of incidence
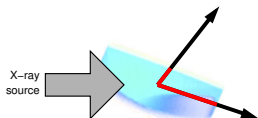
Absorption in the crystal depends on the path
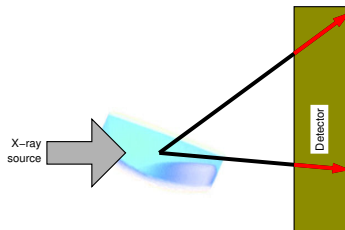
Higher angle of incidence w.r.t. detector surface leads longer path through detector phosphor and thus to stronger signal

Basis for Scaling: Symmetry and multiple measurements

- Some corrections depend on the instrument and can be calibrated (polarisation, angle of incidence, Lorentz factor)

- Some corrections (*e.g.* absorption, extinction) are (also) sample dependent

Idea: symmetry equivalent reflections, or multiply measured reflections, should have the same intensity

Scaling means

1. Determination of the measured intensity I_{obs} of a set of equivalent reflections

2. Determination of their standard uncertainties

3. Result: idealised data set

Examples for corrections
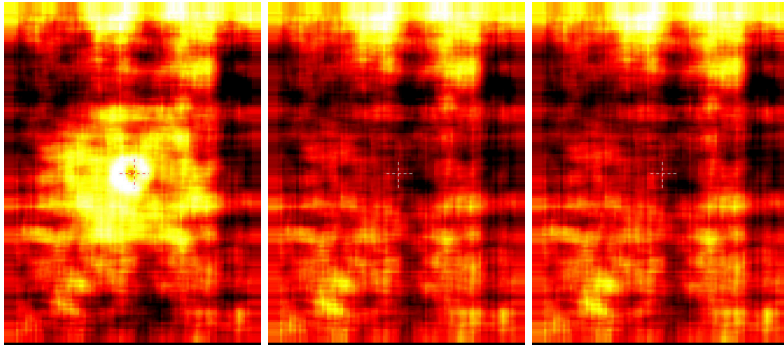

Detector "gain": dark regions of the detector are more sensitive (factor between 0.97 and 1.3). Direct beam (left): indicates non-linear response at high intensity

Scaling details

Options to find the mean intensity $< I >$ of a group of equivalent reflections

1. Statistical average: exaggerates outliers

2. Scaling of each group of equivalents independently from other groups: neglects systematic errors.

3. Instead: **One** scale factor for several groups of equivalent reflections (*e.g.* 100 groups) which are close together on the detector

5 equiv. reflections
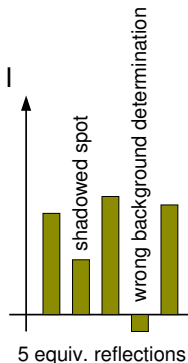
Scaling plots (program SADABS)
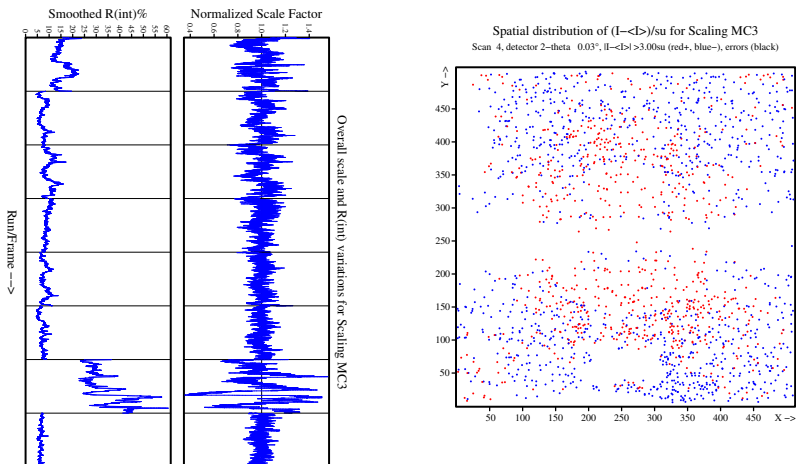
Idealised intensities

Before Scaling   After Scaling

$$I(hkl) \quad = \frac{e^4}{m_e{}^2 c^4} \frac{\lambda^3 V_{\text{crystal}}}{V_{\text{u.c.}}^2} I_0 LPTE|F(hkl)|^2 \qquad I(hkl) \quad = c|F(hkl)|^2 \qquad (2)$$

```
0 -1  5  1.379E+03  2.516E+02        0 -1  5  7.014E+01  1.208E+01
0 -1 -5  1.367E+03  2.726E+02        0 -1 -5  6.812E+01  1.274E+01
0  1  5  1.184E+03  2.610E+02        0  1  5  5.987E+01  1.231E+01
0  1 -5  1.347E+03  2.674E+02        0  1 -5  6.753E+01  1.258E+01
0 -1  6  1.090E+04 -1.229E+03           outlier removed
0 -1 -6  4.677E+03  5.733E+02        0 -1 -6  2.365E+02  2.856E+01
0  1  6  4.286E+03  5.488E+02        0  1  6  2.145E+02  2.689E+01
0  1 -6  9.065E+03 -1.034E+03           outlier removed
0 -1  7  0.204E+02  0.571E+01        0 -1  7  1.404E+02  2.271E+01
```
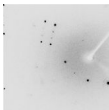
Summary Scaling

- Scaling finds the mean intensity $< I >$ of symmetry equivalent reflections

- Scaling puts one or more data sets on a common scale

- Scaling removes outliers

- The result of scaling is a single (hkl)-file, which is used for all subsequent steps:

  1. Phasing, *i.e.* structure solution
  2. Model refinement

3 Phasing

Phasing

Data collection ➡ **Data integration** ➡ **Data Scaling**


```
0  0 -1      2.7    0.9        0  0 -1      2.8   0.55
0  0  1      4.0    1.0        0  0  1      3.8   0.63
0  0 -2 1'257.0   35.5        0  0 -2 1'432.0   95.7
0  0 -2 1'600.0   42.7        0  0 -2 1'282.0   85.9
```

several GB   several files, 100's MB   1 "hkl"-file, 50MB
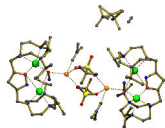
➡ **Phasing** ➡ **Refinement**

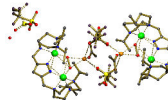Starting model   Chemically   sensible
                          model

The Structure Factor $F(hkl)$

- Context between atoms and diffraction intensities

- Describing the electron density with the independent atom model ("IAM")

- Formfactor and the "fudge factor" ADP

The Structure Factor

The structure factor $F(hkl)$ is related to the **electron density** $\rho(x, y, z)$, i.e. the distribution of electrons inside the unit cell:

$$F(hkl) = \int_{\text{unit cell}} \rho(x, y, z) e^{2\pi i(hx + ky + lz)} d^3 x \qquad (3)$$

This equation is the Fourier transformation of the electron density.

Note: The term "Fourier transformation" is important mainly because computers are very fast in calculating Fourier transformations.

The Structure Factor

The *Independent Atom Model* (IAM, alias *isolated atom model*) is a powerful method to calculate the atomic structure factor $F(hkl)$

$$F(hkl) = \sum_{\substack{\text{atoms } j \\ \text{in u.c.}}} f_j(\theta) e^{-8\pi^2 U_j(\theta,\lambda)} e^{2\pi i(hx_j + ky_j + lz_j)} \tag{4}$$

f_j atomic *form* factor. Dependent on atom element, decreases with decreasins scattering angle $\theta = \theta(hkl)$

$U_j(\theta, \lambda)$ atomic displacement parameter (ADP, alias Debye-Waller factor): models thermal vibration of atoms

$e^{2\pi i(hx_j + ky_j + lz_j)}$ phase shift of the atom relative to the origin of the unit cell

The form factor $f_j(\theta)$

Intensity of the scattered X-rays decreases with increasing scattering angle θ.

X-ray form factors (Intl. Tables of Cryst., Vol. C, Tab. 6.1.1.1)
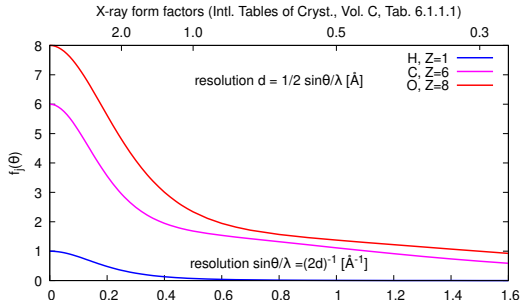
The wavelength λ is of the same order of magnitude as the size of the atoms: photons "see" the shape of atoms.

Note: hydrogen atoms do not contribute to data higher than 1 Å

The atomic displacement parameter $U_j(\theta, \lambda)$

- Atoms vibrate at $T > 0K$

- Vibration leads to reduction of spot intensities, but not to change in spot shape

- At medium resolution: 1 parameter

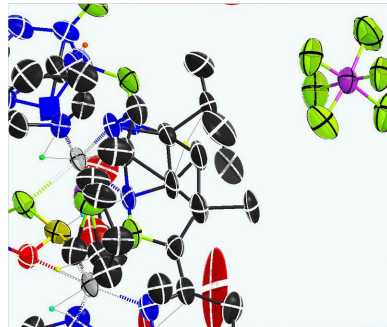- At high resolution: anisotropic description with 6 parameters as ellipsoids

Isotropic	**Anisotropic**
$U_j(\theta, \lambda) = 4U_{\text{iso}} \frac{\sin^2 \theta}{\lambda^2}$	$U_j(\theta, \lambda) = (hkl) \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{pmatrix} \begin{pmatrix} a^* \\ b^* \\ c^* \end{pmatrix}$
1 parameter per atom	6 parameters per atom

- Name : ADP = isotropic or anisotropic *atomic displacement parameter*
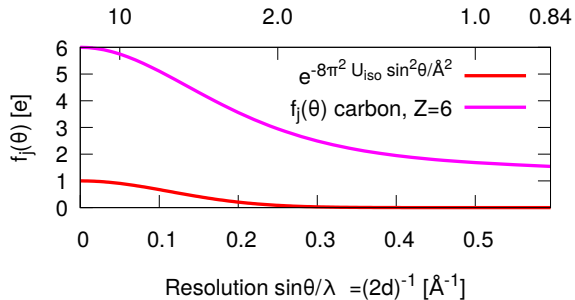
Example images for ADP
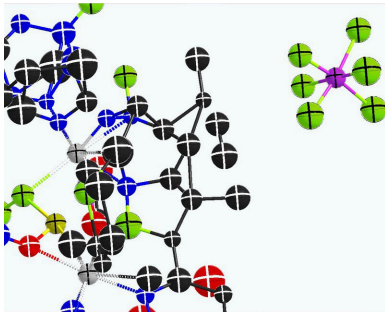
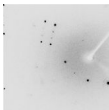Refinement with isotropic ADPs   Refinement with anisotropic ADPs

The ADP U: a fudge factor

- Sharp drop-off with resolution: can make wrongly placed atoms disappear.

- Similarity with form factor: confusion of atom types

- Risk of overfitting

Phasing: First look at the chemical structure

Data collection ➡ **Data integration** ➡ **Data Scaling**


```
0 0 -1      2.7    0.9          0 0 -1      2.8   0.55
0 0  1      4.0    1.0          0 0  1      3.8   0.63
0 0 -2 1'257.0   35.5          0 0 -2 1'432.0   95.7
0 0 -2 1'600.0   42.7          0 0 -2 1'282.0   85.9
```

several GB   several files, 100's MB   1 "hkl"-file, 50MB

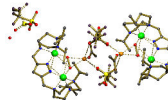➡ **Phasing**   ➡ **Refinement**

Starting model   Chemically   sensible
                        model

Phasing *alias* Solving the structure

- Ideally, we would measure data, run a single calculation, and get the molecular structure as result.

- our data: thousands of measured intensities $I(hkl)$.

- From Eq. 2 (p. 23) and Eq. 4 (page 29):

$$I(hkl) = c|F(hkl)|^2$$
$$= c \left| \int_{\text{unit cell}} \rho(x, y, z) e^{2\pi i(hx+ky+lz)} d^3x \right|^2$$

- **If** we could invert this equation, we could calculate the coordinates x, y, z in one go. But we cannot …

The phase problem

The inverse of the Fourier transformation

$$F(hkl) = \int_{\text{unit cell}} \rho(x, y, z) e^{2\pi i(hx+ky+lz)} d^3x$$

reads

$$\rho(x, y, z) = \sum_{(h,k,l)} F(hkl) e^{-2\pi i(hx+ky+lz)}$$

The phase problem

The structure factor $F(hkl)$ is a complex number. Therefore, it has

an amplitude $|F(hkl)| = \sqrt{I(hkl)/c}$

a phase $\phi(hkl) =$???

$F(hkl) = \sqrt{I(hkl)/c} \times e^{-i\phi(hkl)}$

We can measure the amplitude, but we cannot measure the phase. This is known as the **phase problem of crystallography**.

References

[1] C. Giacovazzo, ed. *Fundamentals of Crystallography*. Oxford University Press, 1985.