Chemical Crystallography and Structural Chemistry

VO 270063-1

Lecture № 7 — 11th May 2023

Dr. Tim Grüne

Centre for X-ray Structure Analysis

Faculty of Chemistry

University of Vienna

tim.gruene@univie.ac.at

Course Schedule

2nd	March	Lecture № 1	9th	March	Lecture № 2
16th	March	Lecture № 3	23rd	March	Exercise № 1
30st	March	Lecture № 4	20th	April	Lecture № 5
27th	April	Exercise № 2	4th	May	Lecture № 6
11th	May	Lecture № 7	25th	May	Exercise № 3
1st	June	**no lecture**	15th	June	Lecture № 8
22nd	June	Lecture № 9	29th	June	Exercise № 4

Previous Lecture

- Refinement and Model Building

- Validation: Data quality

- Validation: Model quality

Contents

1 Structure Refinement

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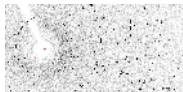

```
0 0 -1      2.7  0.9
0 0  1      4.0  1.0
0 0 -2 1'257.0 35.5
0 0 -2 1'600.0 42.7
```

```
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
```
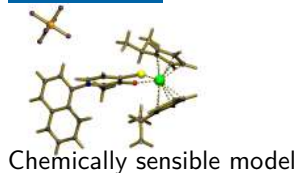
several GB

several   files,
100's MB

1 "hkl"-file, 50MB

➡ **Phasing**   ➡ **Refinement**

Starting model   Chemically sensible model
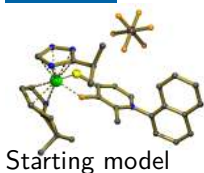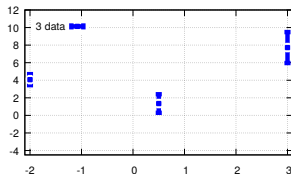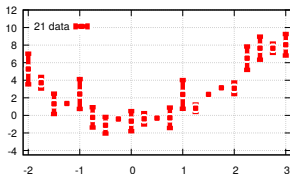
Example: Stabilisation through restraints

Two hypothetic measurements:

Experiment 1: high resolution, 21 pairs of measurements $(x_1, y_1), \ldots, (x_{21}, y_{21})$ and errors $\sigma_1, \ldots, \sigma_{21}$

Experiment 2: low resolution, 3 pairs of measurements $(x_1, y_1), \ldots, (x_3, y_3)$ and errors $\sigma_1, \ldots, \sigma_3$

Example: Stabilisation through restraints

Testing two models:

$$\textbf{Model 1:}\ g(x) = g_2 x^2 + g_1 x + g_0$$

$$\textbf{Model 2:}\ h(x) = h_3 x^3 + h_1 x + h_0$$

Either model has three parameters, g_0, g_1, g_2 and h_0, h_1, h_3 respectively. These parameters correspond to *e.g.* the model coordinates (x_i, y_i, z_i), or the ADPs U_i.

We will fit both models to the data to find out which model better represents the data.

Example: Stabilisation through restraints

Least-squares-minimisation:

$$\text{minimise: } \sum_{i=1}^{N} \frac{1}{\sigma_i^2}(y_i - g(x_i))^2 \qquad \textcolor{magenta}{\text{model 1}}$$

$$\text{minimise: } \sum_{i=1}^{N} \frac{1}{\sigma_i^2}(y_i - h(x_i))^2 \qquad \textcolor{green}{\text{model 2}}$$
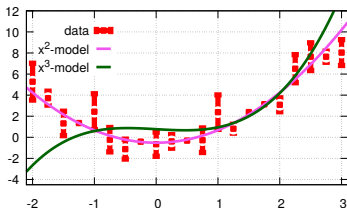
- Experiment 1: $N = 21$ data points

- Experiment 2: $N = 3$ data points

We will start with the high resolution experiment 1

Example: Stabilisation through restraints

experiment 1: high resolution; high data to parameter ratio = 21:3=7

Model 1: $1.2x^2 + 0.0x - 0.5$
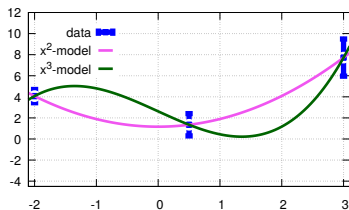rmsd: 1.07
Model 2: $0.5x^3 - 0.3x - 0.8$
rmsd: 4.74

The *root mean square deviation* rmsd between model and data corresponds to the crystallographic $R1$ value.

The lower rmsd 1.07 clearly favours model 1. The pink curve also visually fits the data better than the green curve.

Example: Stabilisation through restraints

experiment 2: low resolution, low data to parameter ration = 3:3 = 1

model 1: $0.7x^2 + 0.0x + 1.2$
rmsd: 0
model 2: $0.5x^3 - 2.7x - 2.6$
rmsd: 0

When there are as many parameters as data points, any model can be fitted
perfectly to the data. We cannot distinguish between the two models

Example: Stabilisation through restraints

experiment 2: low resolution with constraint

For some reason we know that the data must pass through the point $(0, 0)$.
For the two models this means

$$0 = g(0)$$
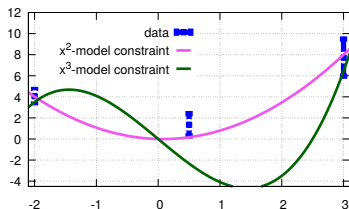$$= g_2 * 0^2 + g_1 * 0 + g_0$$
$$\Rightarrow g_0 = 0$$

and analogously

$$h_0 = 0$$

The constraint reduced the number of parameters, only two parameters per model

Example: Stabilisation through restraints

experiment 2: low resolution with constraint

model 1: $0.9x^2 - 0.1x$
rmsd: 1.13
model 2: $0.8x^3 - 4.9x$
rmsd: 3.7

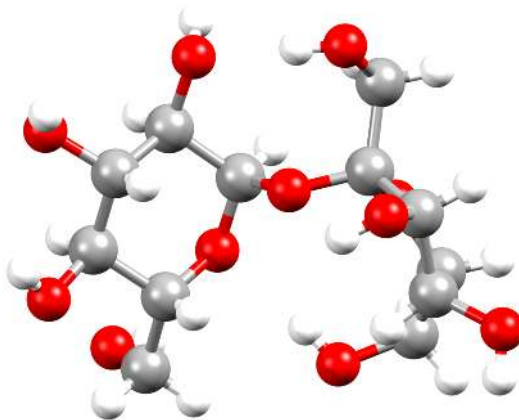Due to the constraint, data to parameter ratio = 3:2 = 1.5. Now there is an *rmsd*, and it favours (again) the first model.

Summary Stabilisation through Restraints / Constraints

- Low resolution data means low data to parameter ratio

- high data to parameter ratio: trustworthy parameters

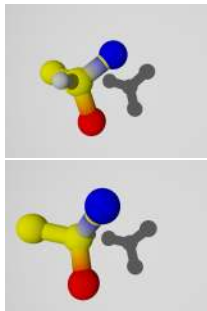- low data to parameter ratio: risk of overfitting

2 Absolute Structure — Chirality

Sucrose

Crystal structure of sucrose

2.1 Cahn-Ingold-Prelog *R/S*-System

- Heavy atoms arranged around chiral centre with lightest one pointing towards viewer

- Three atoms arranged heavy to light anti-clockwise: *R*

- Three atoms arranged heavy to light clockwise: *S*

Chirality in Pharmaceutics

- chiral compounds are optically active: turn the plane of plane-polarised light

- Usually only one stereoisomer pharmaceutically active. Others add to side effects. (*cf.* E. J. Ariëns: *Stereochemistry, a basis for sophisticated nonsense in pharmacokinetics and clinical pharmacology*, European Journal of Clinical Pharmacology, **26** (1984), pp. 663–668).
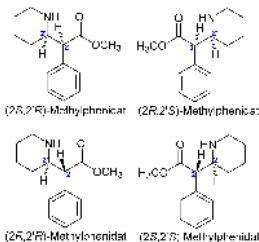
Technical blindness to chirality

- Most technologies are "blind" to chirality

- All angles and all bond-distances are identical between enantiomers

- Some methods determine chirality, but not structure (polarised light)

- Only (X-ray) crystallography can determine absolute structure

Friedel's Law

Calculation of the structure factor from atom coordinates (lecture No. 4):

$$F(hkl) = \sum_{\substack{\text{in u.c.}}}^{\text{atoms } j} f_j(\theta) e^{-8\pi^2 U_j(\theta,\lambda)} e^{2\pi i(hx_j + ky_j + lz_j)} \tag{1}$$

This results in Friedel's law, $I(hkl) = I(\bar{h}\bar{k}\bar{l})$: The diffraction pattern is centrosymmetric, and therefore, the diffraction pattern is blind to chirality.

We have to look more closely!

Anomalous (X-ray) dispersion

M
L
K
atomic nucleus
Lα
Kα
Kβ
kicked-out electron
external stimulation
radiation energy
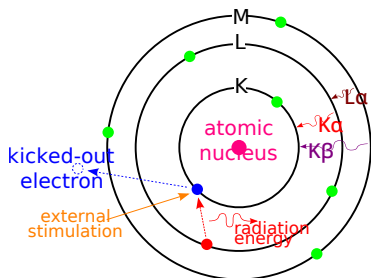
EDX-scheme.svg, commons.wikimedia.org

- Incident X-rays can kick out an inner-shell electron

- higher-shell electrons fill the hole

- they emit characteristic radiation

- This is called "anomalous dispersion" in crystallography

Crystallographic description

Anomalous dispersion can be described with a modified atomic scattering factor:

$$f(\lambda, \theta) = f_0(\theta) + f'(\lambda) + i f''(\lambda)$$

- $f_0(\lambda)$ "normal" form factor, resolution dependent *cf.* Lecture No. 5

- $f'(\theta)$ absorptive component; wavelength dependent

- $i f''(\theta)$ imaginary component, *i.e.* phase shift; wavelength dependent
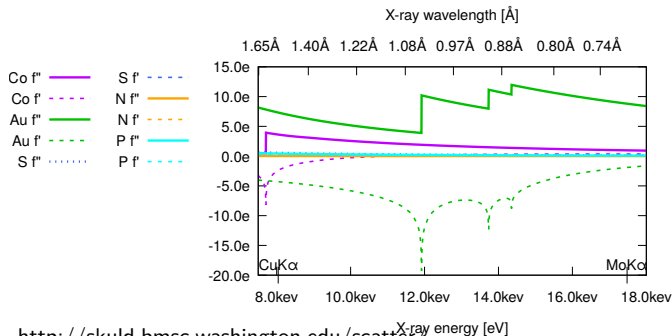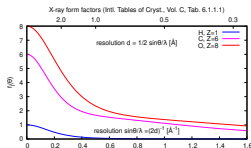
Anomalous scattering

The scattering factor of all atoms, $F(hkl)$, consists of non-anomalous contributions and anomalous contributions:

$$
\begin{aligned}
F(hkl) \quad = \quad & \underbrace{\sum_{\substack{\text{atoms } \nu}}^{\text{normal}} f_\nu e^{2\pi i \mathbf{h} \mathbf{r}_\mu}}_{F_P} \\
& + \underbrace{\sum f_X e^{2\pi i \mathbf{h} \mathbf{r}_X}}_{F_A} \\
& + \underbrace{\sum f'_X e^{2\pi i \mathbf{h} \mathbf{r}_X}}_{F'} \\
& + \underbrace{i \sum f''_X e^{2\pi i \mathbf{h} \mathbf{r}_X}}_{iF''}
\end{aligned}
$$

NB: f' usually negative, pointing backwards
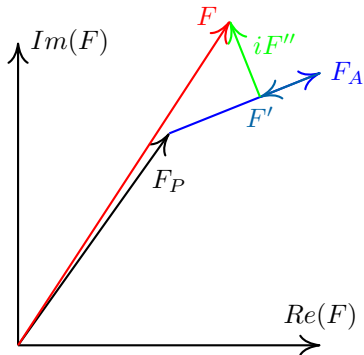
Form factor components $f_0(\theta) + f'(\lambda) + i f''(\lambda)$
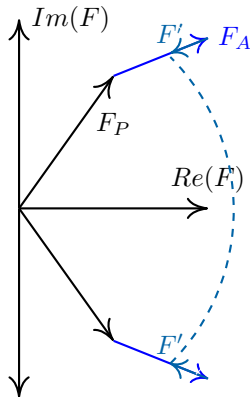

$f'' > 0e$

$f' < 0e$

http://skuld.bmsc.washington.edu/scatter/

iF'' breaks Friedel's law

Compare $F(hkl)$ and $F(\bar{h}\bar{k}\bar{l})$:

$$F(\bar{h}\bar{k}\bar{l}) = \underbrace{\overset{\text{normal}}{\underset{\text{atoms } \nu}{\sum}} f_\nu e^{-2\pi i \mathbf{h}\mathbf{r}_\mu}}_{F_P}$$

$$+ \underbrace{\sum f_X e^{-2\pi i \mathbf{h}\mathbf{r}_X}}_{F_A}$$

$$+ \underbrace{\sum f'_X e^{-2\pi i \mathbf{h}\mathbf{r}_X}}_{F'}$$

$$+ \underbrace{i \sum f''_X e^{-2\pi i \mathbf{h}\mathbf{r}_X}}_{iF''}$$
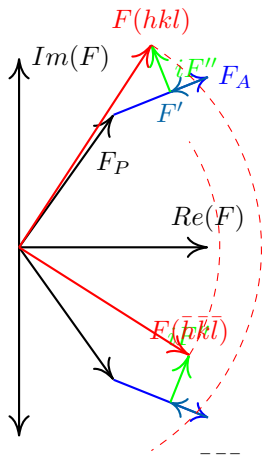

without iF': Friedel's law holds

iF'' breaks Friedel's law

Compare $F(hkl)$ and $F(\bar{h}\bar{k}\bar{l})$:

$$F(\bar{h}\bar{k}\bar{l}) = \underbrace{\sum_{\substack{\text{normal} \\ \text{atoms } \nu}} f_\nu e^{-2\pi i \mathbf{h}\mathbf{r}_\mu}}_{F_P}$$

$$+ \underbrace{\sum f_X e^{-2\pi i \mathbf{h}\mathbf{r}_X}}_{F_A}$$

$$+ \underbrace{\sum f'_X e^{-2\pi i \mathbf{h}\mathbf{r}_X}}_{F'}$$

$$+ \underbrace{i \sum f''_X e^{-2\pi i \mathbf{h}\mathbf{r}_X}}_{iF''}$$

$iF''(hkl)$ breaks the symmetry with $iF''(\bar{h}\bar{k}\bar{l})$

Measuring the anomalous difference

- Anomalous effect can be very small (organic compounds)

- requires very accurate data

- Consider choice of wavelength (near, but below absorption edge)

- In presence of anomalous signal: $F(hkl)$ and $(F\bar{h}\bar{k}\bar{l})$ are called "Bijvoet pair", instead of "Friedel pair"
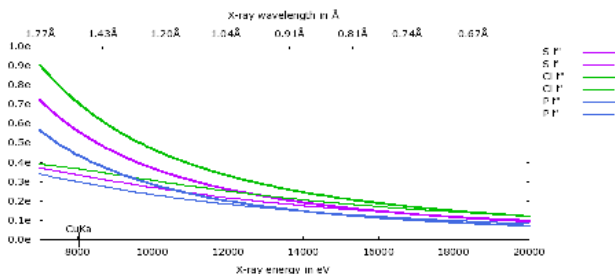
Chirality from anomalous data

Both model and inverted model are compared with the data

1. Refine structure "model 1"

2. Calculates $|F_{\mathsf{calc}}^1(hkl)|$

3. Invert the structure "model 2" and calculate $|F_{\mathsf{calc}}^2(hkl)|$

4. Calculate $R1$ and $wR2$ *via*

$$|F_{\mathsf{calc}}(hkl)| = (1-k)|F_{\mathsf{calc}}^1(hkl)| + k|F_{\mathsf{calc}}^2(hkl)|$$

Chirality from anomalous data (cont'd)

4. Calculate $R1$ and $wR2$ *via*

$$|F_{\text{calc}}(hkl)| = (1-k)|F_{\text{calc}}^1(hkl)| + k|F_{\text{calc}}^2(hkl)|$$

5. Optimise k by minimising $R1$

$k \approx 0$ correct hand

$k \approx 1$ incorrect hand: invert model

between 0 and 1 mixture of both hands or poor data

k is called the **Flack parameter** (Howard D. Flack, University of Geneva).

Flack parameter & Parsons' coefficient

- Parsons' quotient

$$Q(hkl) = \frac{I(hkl) - I(\bar{h}\bar{k}\bar{l})}{I(hkl) + I(\bar{h}\bar{k}\bar{l})}$$

- same as before: compare calculated with observed data

- Parsons' coefficient more sensitive than "conventional" Flack parameter

- *cf*: S. Parsons, H.D. Flack, T. Wagner, *Acta Crystallogr.* (2013) B69, S. 249–259

- enables absolute structure determination from light-atom only structures

Chirality of light-atom only structures

All atoms exhibit an anomalous signal. For C, H, N, O, it is very weak, requires very good data.

Options in case the anomalous signal is too weak, data too poor:

- soak single heavy (salt): same structure, amplification of anomalous signal

- Co-crystallisation of molecule with known chirality (*cf.* NMR)

$\rightarrow$ X-ray data determine chirality relative to the known molecule

$\rightarrow$ If know molecule inverted: also invert the model.

Summary absolute structure

- (physical) origin of anomalous signal

- anomalous signal breaks Friedel's law

- Flack parameter to determine chirality