Chemical Crystallography and Structural Chemistry

VO 270287

Lecture № 8 — 27th May 2021

Dr. Tim Grüne

Centre for X-ray Structure Analysis

Faculty of Chemistry

University of Vienna

tim.gruene@univie.ac.at
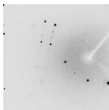
Reminder: Start the recording!

Contents

1 Where we are

Solving the Structure —Phasing

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
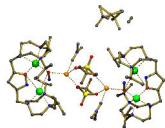
➡   **Phasing**   ➡   Refinement
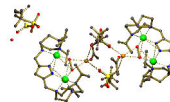

Starting model   Chemically   sensible model

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 $I(hkl)$.

- Previous lecture:

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

- **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 $\phi(hkl) =$???

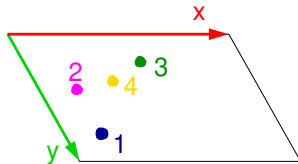$F(hkl) = \sqrt{I(hkl)/c} \times e^{-i\phi(hkl)}$

We can measure the amplitudes $|F(hkl)| = \sqrt{I(hkl)/c}$.

The values $\phi(hkl)$ cannot be directly measured.

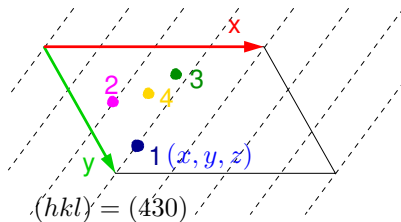This is known as the **phase problem of crystallography**.

2 Relation between Phases and Molecule

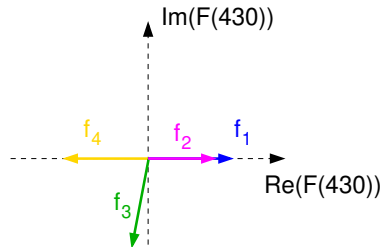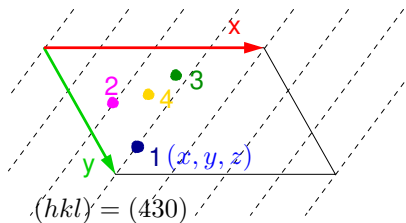## The phases are related to the chemical structure

Unit cell with four atoms inside.

The phases are related to the chemical structure

Example reflection $F(430) = \sum_{j=1}^{4} f_j(\theta) e^{-8\pi^2 U_j(\theta, \lambda)} e^{2\pi i (4x_j + 3y_j + 0z_j)}$

universität
wien

The phases are related to the chemical structure

$$\phi_j = 2\pi(4x_i + 3y_i + 0z_i)$$

$$= \text{rel. distance to origin}$$

$$= \text{rel. distance to lattice plane}$$

$$F(430) = \sum_{j=1}^{4} f_j(\theta) e^{-8\pi^2 U_j(\theta,\lambda)} e^{2\pi i(4x_j + 3y_j + 0z_j)}$$
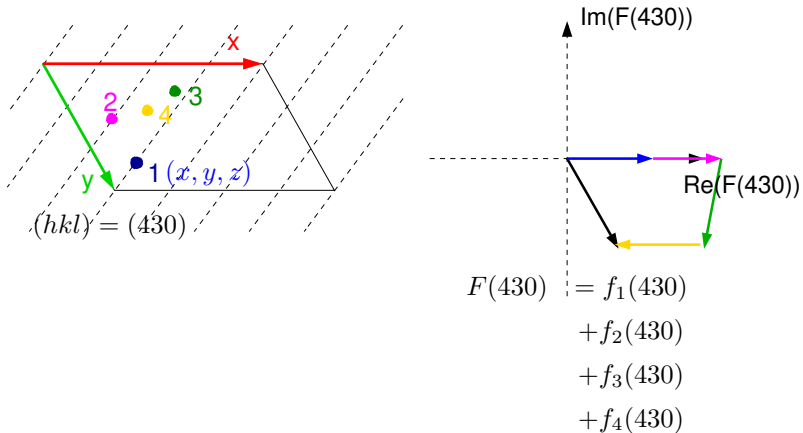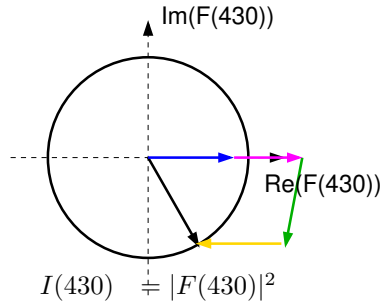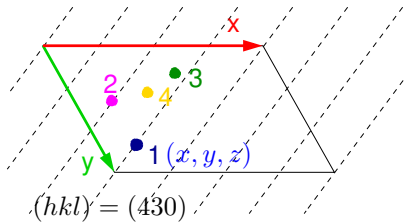
The phases are related to the chemical structure

$$F(430) = f_1(430)$$
$$+ f_2(430)$$
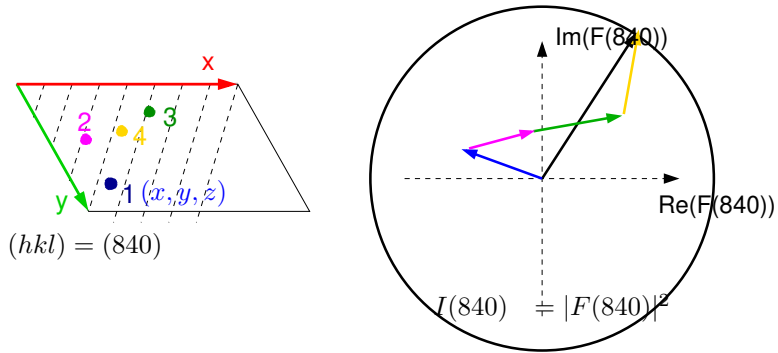$$+ f_3(430)$$
$$+ f_4(430)$$

Total $F(430)$ = sum of individual vectors

The phases are related to the chemical structure

Measurement $I(430) = |F(430)|^2$: loss of phase information

The phases are related to the chemical structure

$(hkl) = (840)$

$I(840) \doteq |F(840)|^2$

Same coordinates, *different* contributions per atom to $F(840)$

Phases $\phi(hkl)$

- The phase problem arises, because $F(hkl)$ is a complex number.

- $F(hkl) = \sum_{\text{all atoms}} f_j(\theta) e^{-8\pi^2 U_j(\theta,\lambda)} e^{2\pi i(hx_j+ky_j+lz_j)}$

- The length f_j for each atom is independent of (hkl).

- The phase contribution $e^{2\pi i(hx_j+ky_j+lz_j)}$ varies for each reflection.

- the phases $\phi(hkl)$ is an angle between $0°$ and $360°$

- The total phase $\phi(hkl)$ contains convoluted information from each atom. Each contribution depends on the position of the atom **relative** to the lattice planes for the reflection (hkl), *i.e.*, there are as many phases as there are reflection spots.

3 Solving the structure = Solving the phase problem

Solving the structure = Solving the phase problem

- The phase problem prevents us from calculating the electron density map directly from our data

- Phases can be calculated from a chemical model (its coordinates)

- "Phasing" means to find a model close enough to the proper model

- Once a good enough molecule has been found, it needs to be improved: "model building" and "refinement"

Phasing methods [1, 2, 3]

There are several methods to solve the phase problem. This lecture will cover to most popular ones

1. Patterson map

2. Direct methods

Patterson maps are common for small molecules, which contain a mixture of heavy and light atoms.

Direct methods are particularly useful for structure with similar elements, *e.g.* organic compounds.

4 The Patterson map

The Patterson map

- since 1934, Arthur Lindo Patterson (1902–1966)

- good for very small structures with some heavy elements

- direct determination of atom positions

Calculation of the Patterson map

The Patterson map ignores phases and calculates the Fourier transformation from the intensities:

$$P(uvw) = \sum_{(hkl)} I(hkl)e^{-2\pi i(hu+kv+lw)}$$

This can be calculated without knowing the phases $\phi(hkl)$, only from the measured intensities.

It turns out this map is the "auto-convolution" of the electron density with itself $((uvw) = \vec{u})$:

$$P(uvw) = \rho(\vec{x}) \star \rho(\vec{x} - \vec{u})$$
$$= \int_{\text{unit cell}} \rho(\vec{x})\rho(\vec{x} - \vec{u})d^3x$$

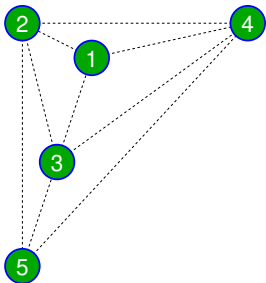## Meaning of the Patterson map

It can be shown that the Patterson map

$$P(uvw) = \rho(\vec{x}) \star \rho(\vec{x} - \vec{u})$$

$$= \int_{\text{unit cell}} \rho(\vec{x})\rho(\vec{x} - \vec{u})d^3x$$

has its peaks at vector (positions) $\vec{u}$ that corresponds to the connecting vector between two atoms in the molecule in the unit cell.
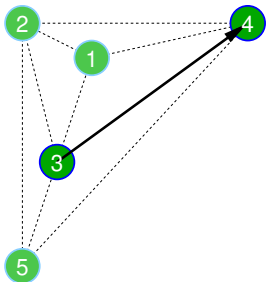
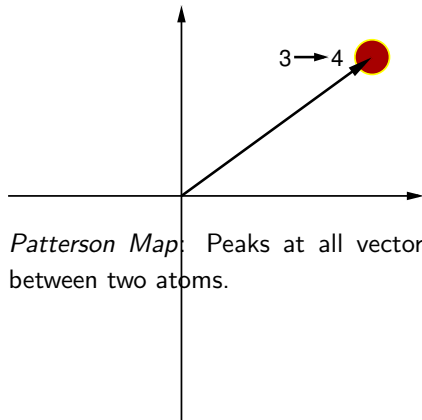## Illustration of the Patterson map in 2D

"2D molecule, 5 atoms"   peaks of Patterson map

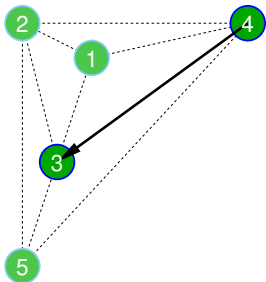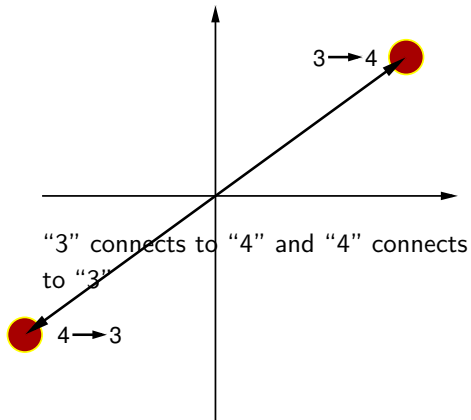## Illustration of the Patterson map in 2D

"2D molecule, 5 atoms"

Patterson Map: Peaks at all vectors between two atoms.

peaks of Patterson map
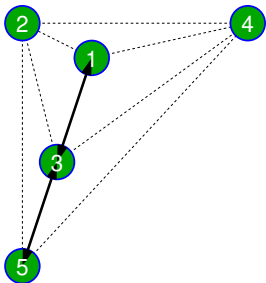
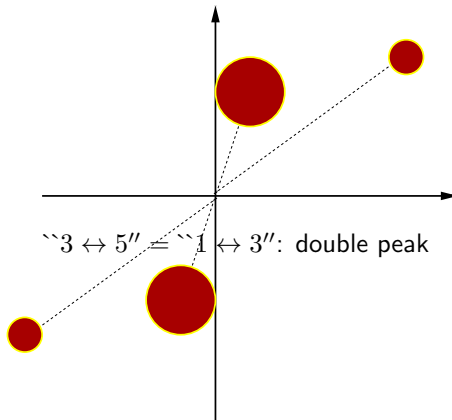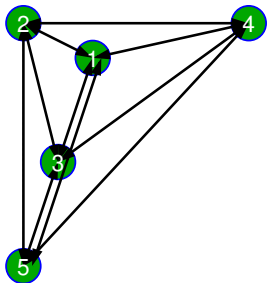## Illustration of the Patterson map in 2D

"2D molecule, 5 atoms"

"3" connects to "4" and "4" connects to "3"

peaks of Patterson map

Illustration of the Patterson map in 2D

"2D molecule, 5 atoms"

``3 ↔ 5'' = ``1 ↔ 3'': double peak

peaks of Patterson map

Illustration of the Patterson map in 2D

"2D molecule, 5 atoms"   In total: 10 connections, 20 peaks

                                peaks of Patterson map

Illustration of the Patterson map in 2D

"2D molecule, 5 atoms"

Strong origin: each atom connects with itself

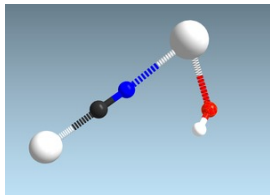peaks of Patterson map

Patterson map observations

- heavy elements have stronger peaks (high density $\rho(x, y, z)$)

- with too many atoms: origin peak overwhelms: non-interpretable

- with too many atoms: too many peaks, overlap

- Patterson map always centro-symmetric (peak at $(x, y, z) \Leftrightarrow$ peak at $(-x, -y, -z)$)

- the Patterson map does not **directly** reveal the molecule shape

Patterson map for $La[Au(CN)_2]_3 \cdot 3H_2O$
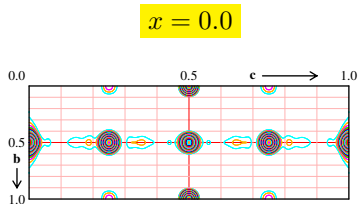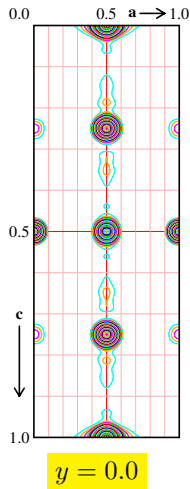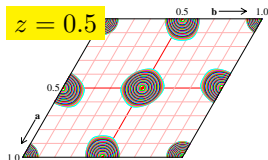
- La(III): $54e^-$, Au(I): $78e^-$, O^{2-}: $10e^-$

- $54*78 = 4'212 \gg 78*10 = 780$: 1 dominating peak

- Patterson maps: typically "origin peak removed"

Summary Patterson map

- Patterson map calculated from intensities, without phases

- Patterson map corresponds to convolution of density $\rho(x, y, z)$

- Peaks correspond to connecting vectors between atoms

- Peak height corresponds to product of number of electrons

- Atom coordinates can be deduced from map in case of few atoms, or few heavy atoms

- The more atoms (of similar weight), the harder to interpret

5 Direct methods

Direct methods

- Well suited with molecules of similar atom types (organic compounds with $C, N, O, \ldots$)

- Can work with thousands of atoms

- Requires atomic resolution, better than 1.2 Å (Sheldrick's rule, [4])

Concept of direct methods

1. Generate roughly the number of expected atoms at arbitrary positions

2. Calculate phases of this pseudo-molecule

3. Improve phases based on **tangent formula**

4. Improved phases produce an improved electron density map

5. Peak picking from improved map

6. Repeat

7. Best solution: assign atom types

Direct methods: the tangent formula

Tangent formula[1] was derived by H. A. Hauptman and J. Karle — chemistry Nobel prize 1985

$$\tan(\phi_{\mathbf{h}}) \approx \frac{\sum_{\mathbf{h'}} |E_{\mathbf{h'}} E_{\mathbf{h-h'}}| \sin(\phi_{\mathbf{h'}} + \phi_{\mathbf{h-h'}})}{\sum_{\mathbf{h'}} |E_{\mathbf{h'}} E_{\mathbf{h-h'}}| \cos(\phi_{\mathbf{h'}} + \phi_{\mathbf{h-h'}})}$$

creates a network of phase relationships between (hkl) and $(h-h', k-k', l-l')$. Historically based on Sayre-Equation (1952)

$$F(hkl) = q(\sin\theta/\lambda) \sum_{(h'k'l')} F(h'k'l') * F(h-h', k-k', l-l')$$

Sayre equation is exact for cases of only one atom type in crystal (diamond, silicon, *etc*). That is why the tangent formula works best for similar-atoms-compounds.
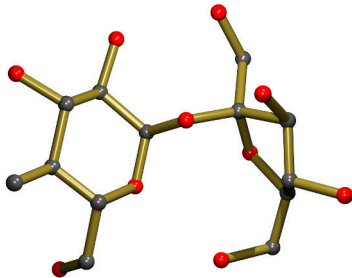
---
[1] $E(hkl)$: normalised structure factors, derived from measured $F(hkl)$
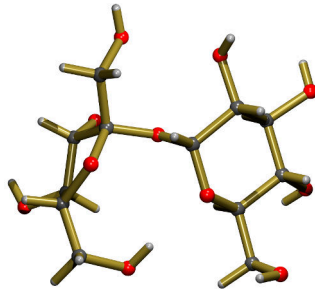
Direct methods: dual space recycling
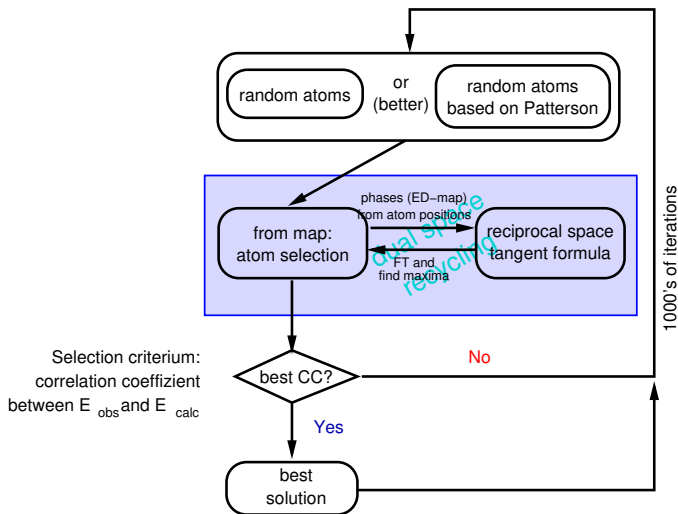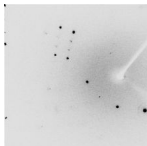

Amazingly, it works

Example: Sucrose

Sucrose, solved with automated atom assignment (SHELXT)

Final structure

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

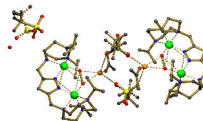several files, 100's MB

1 "hkl"-file, 50MB
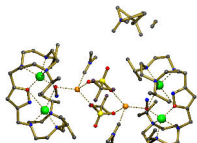
➡ **Phasing**

Starting model

➡ **Refinement**

Chemically sensible model

References

[1]   C. Giacovazzo. *Phasing in Crystallography. A Modern Perspective*. Oxford University Press, 1998.

[2]   Lukáš Palatinus and Gervais Chapuis. 'SUPERFLIP – a computer program for the solution of crystal structures by charge flipping in arbitrary dimensions'. In: *J. Appl. Crystallogr.* 40 (2007), pp. 786–790. DOI: 10.1107/S0021889807029238.

[3]   George M. Sheldrick. 'SHELXT – Integrated space-group and crystal-structure determination'. In: *Acta Crystallogr* A71 (2015), pp. 3–8. DOI: 10.1107/S2053273314026370.

[4]   Richard J. Morris and Gérard Bricogne. 'Sheldrick's 1.2Å rule and beyond'. In: *Acta Crystallogr* D59 (2003), pp. 615–617. DOI: 10.1107/S090744490300163X.