Chemical Crystallography and Structural Chemistry

VO 270287

Lecture № 7 — 9th June 2022

Dr. Tim Grüne
Centre for X-ray Structure Analysis
Faculty of Chemistry
University of Vienna

tim.gruene@univie.ac.at

Course Details

3rd	March	Lecture № 1	10th	March	Lecture № 2
17th	March	**no lecture**	24th	March	Exercise № 1
31st	March	Lecture № 3			
7th	April	Discussion Ex. 1	14th	April	**Easter break**
21st	April	**Easter break**	28th	April	Lecture № 4
5th	May	Power cut	12th	May	**no lecture**
19th	May	Lecture № 6	26th	May	**Ascension Day**
2nd	June	Exercise № 2	9th	June	Lecture № 7
16th	June	**Corpus Christi**	23th	June	Lecture № 9
30th	June	**no lecture**			

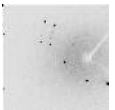
Previous Lecture

-

Contents

From Data Collection to Structure

Data collection → **Data integration** → **Data Scaling**



several GB

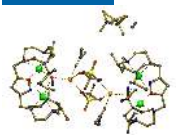
```
0 0 -1      2.7   0.9
0 0  1      4.0   1.0
0 0 -2 1'257.0   35.5
0 0 -2 1'600.0   42.7
```

several files, 100's MB

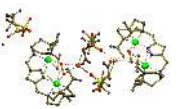
```
0 0 -1      2.8   0.55
0 0  1      3.8   0.63
0 0 -2 1'432.0   95.7
0 0 -2 1'282.0   85.9
```

1 "hkl"-file, 50MB

→ **Phasing** → **Refinement**



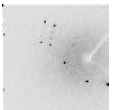
Starting model



Chemically sensible model

1 Phasing

Data collection → Data integration → Data Scaling



several GB

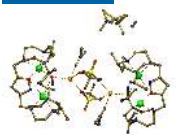
```
0 0 -1      2.7   0.9
0 0  1      4.0   1.0
0 0 -2 1'257.0   35.5
0 0 -2 1'600.0   42.7
```

several files, 100's MB

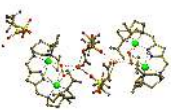
```
0 0 -1      2.8   0.55
0 0  1      3.8   0.63
0 0 -2 1'432.0   95.7
0 0 -2 1'282.0   85.9
```

1 "hkl"-file, 50MB

→ **Phasing** → Refinement



Starting model



Chemically sensible model

The phase problem

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

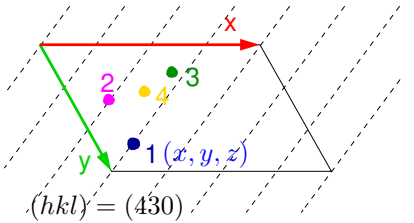
an amplitude $|F(hkl)| = \sqrt{I(hkl)/c}$

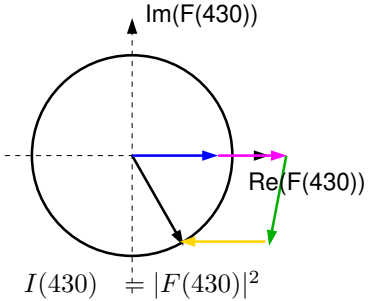
a phase $\phi(hkl) =$???

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

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

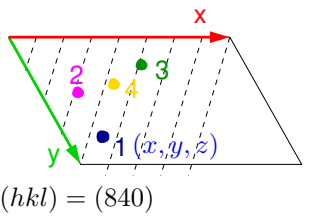
The phases are related to the chemical structure

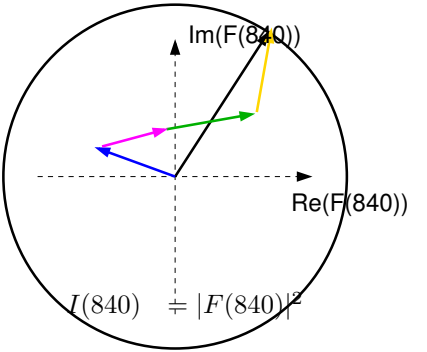




Measurement $I(430) = |F(430)|^2$: loss of phase information

The phases are related to the chemical structure





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

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"

Solving the structure = Solving the phase problem

> In the beginning, crystallographers worked on the structures of simple molecules and they could often make a good guess of the conformation of a molecule and even how it might pack in the crystal lattice. The guesses could be tested by calculating a diffraction pattern and comparing it to the observed one. If a guess places the atoms in about the right place, then the calculated phases will be approximately correct and a useful electron density map can be computed by combining the observed amplitudes with the calculated phases. If the model is reasonably accurate, such a map will show features missing from the model so that the model can be improved. You can remind yourself how this works by looking at Kevin Cowtan's cats.
>
> Randy Read, [1]

Phasing and then?

- "phasing" estimates the phase $\phi(hkl)$ for each reflection of the dataset
- the estimates are often quite poor
- the initial average phase errors can be several 10s of degrees $(10, 20, 30, \ldots^\circ)$
- initial phases means
  1. finding **some** atom positions
  2. some correct, some incorrect element types
  3. often fixed, isotropic ADP values
- model building and refinement improves these phases by correcting these approximations

Phasing methods [2, 3, 4]

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.

2 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 ($(uvw) = \vec{u}$):

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

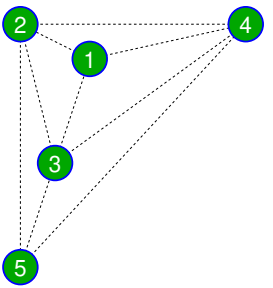
Meaning of the Patterson map

It can be shown that the Patterson map

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

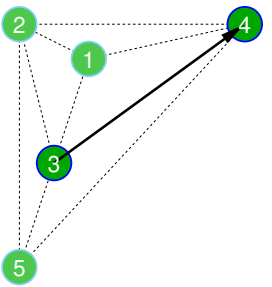
has its peaks at vector (positions) $\vec{u}$ that corresponds to the connecting vector between two atoms in the molecule in the unit cell.

Illustration of the Patterson map in 2D

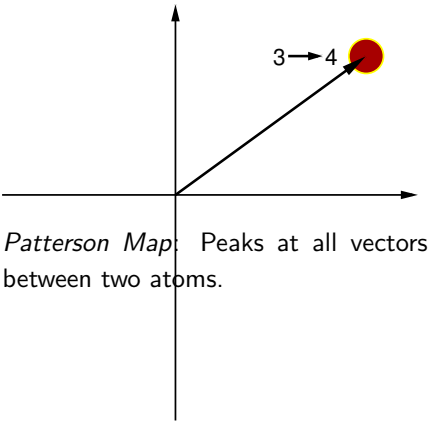


"2D molecule, 5 atoms" peaks of Patterson map

Illustration of the Patterson map in 2D



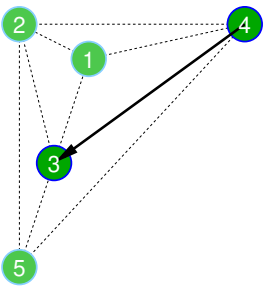
"2D molecule, 5 atoms"



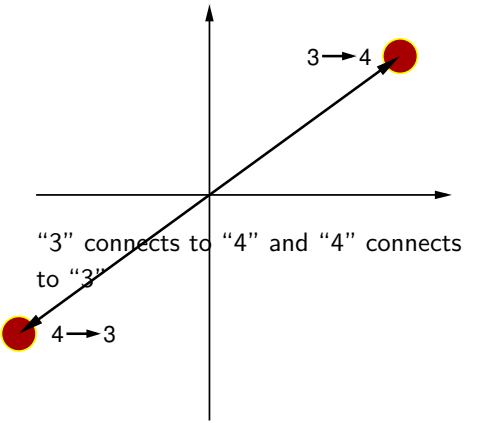
Patterson Map: Peaks at all vectors between two atoms.

peaks of Patterson map

Illustration of the Patterson map in 2D

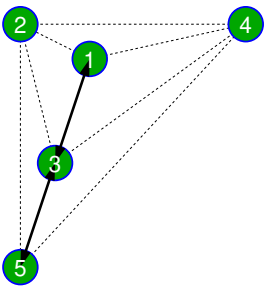


"2D molecule, 5 atoms"

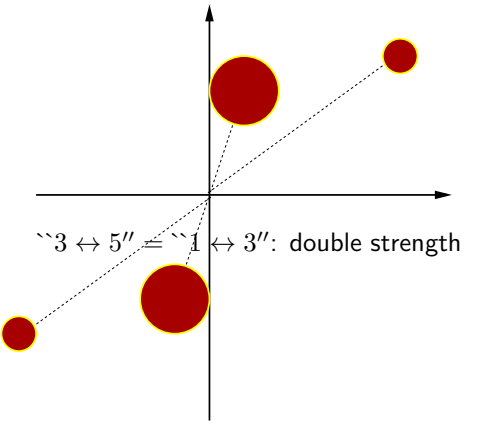


peaks of Patterson map

Illustration of the Patterson map in 2D

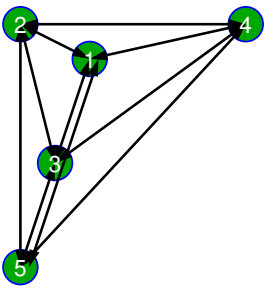


"2D molecule, 5 atoms"

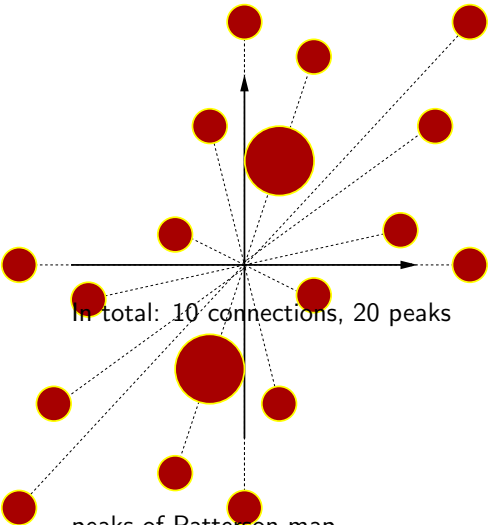


peaks of Patterson map

Illustration of the Patterson map in 2D

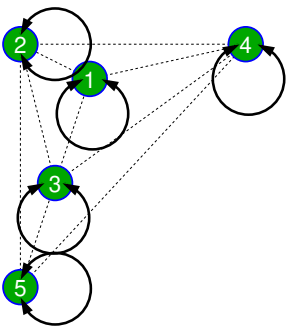


"2D molecule, 5 atoms"

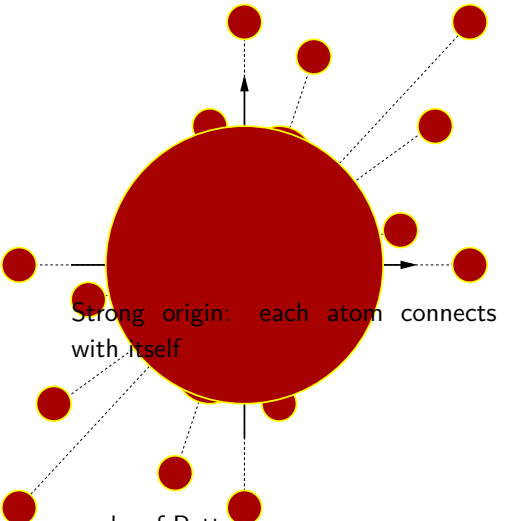


peaks of Patterson map

Illustration of the Patterson map in 2D



"2D molecule, 5 atoms"



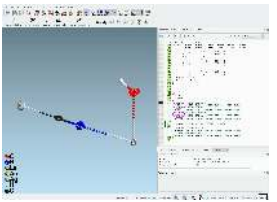
Strong origin: each atom connects with itself

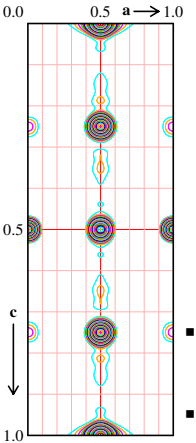
peaks of Patterson map

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

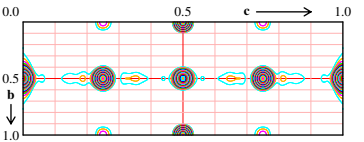
Patterson map for $La[Au(CN)_2]_3 \cdot 3H_2O$



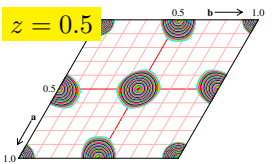


$y = 0.0$

$x = 0.0$



$z = 0.5$



- 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

3 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, [5])

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

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

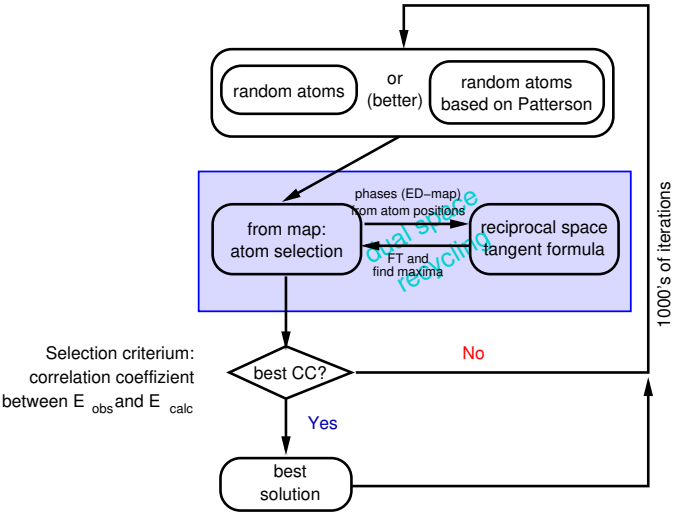
creates a network of phase relationships between (hkl) and $(h-h', k-k', l-l')$. Historically based on Sayre-Equation (1952)

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

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

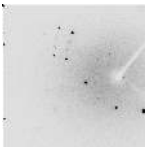
Example: Sucrose

Sucrose, solved with automated atom assignment (SHELXT)

Final structure

4 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
```

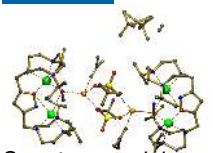
```
0 0 -1      2.8 0.55
0 0  1      3.8 0.63
0 0 -2 1'432.0 95.7
0 0 -2 1'282.0 85.9
```

several GB

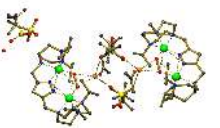
several files, 100's MB

1 “hkl”-file, 50MB

➡ **Phasing** ➡ **Refinement**



Starting model



Chemically sensible model

Model Building & Refinement

- Refinement optimises computationally the parameters of the structure with respect to the data
- Model building make modifications that are too large for computer optimisation, e.g.
  1. Addition or removal of atoms
  2. correction of atom types
  3. modelling of disorder and multiple conformations

Structure parameters

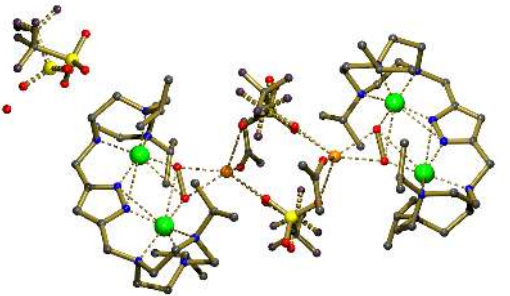
A "structure" consists of a set of parameters, *i.e.* numbers. Refinement improves these numbers for make the structure better correspond to the data.

```
       X        Y        Z    occ.
N2A  2 0.8142   0.9066   0.8201 11.00000 =
       0.0497   0.0413   0.0363 -0.0136  -0.0041 -0.0063
       U11      U22      U33      U23       U13      U12
```

1. 3 atom coordinates x, y, z
2. 6 atomic displacement parameters ADP $\begin{pmatrix} U_{11} & U_{12} & U_{13} \\ & U_{22} & U_{23} \\ & & U_{33} \end{pmatrix}$
3. possibly 1 occupancy parameter for disorder

9 Parameters per atom of the asymmetric unit are being refined, plus extra parameters in case of disorder or other special circumstances

Example for parameters [6]



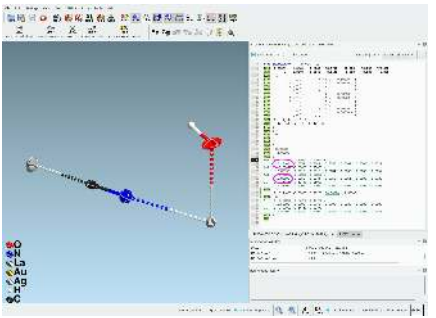
- $C_{34}H_{63}Cu_2F_6N_8NaO_{9.5}S_2$
- 62.5 non-hydrogen atoms
- 724 parameters
- hydrogen atoms are “special”

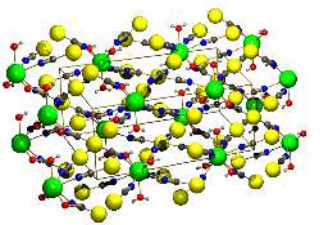
Atom occupancy — symmetry

- Asymmetric unit: average of all asymmetric units of the crystal
- Molecules do not always strictly follow symmetry
- Some atoms sit on "special position", i.e. fix points of symmetry elements. Their occupancy is divided by the multiplicity of the symmetry element
- e.g. atom on three-fold axis: occupancy 33 %

Atom occupancy — alternative elements

Example: $La[Ag_{0.39}Au_{0.61}(CN)_2]_3 \cdot H_2O$ [7] 39% of all unit cells contain Ag, 61% contain Au at the same position

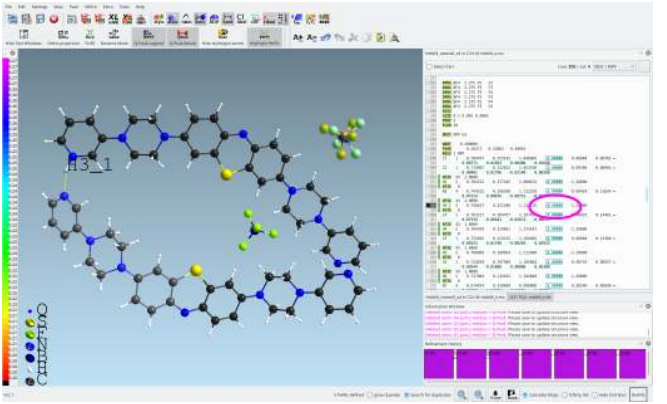




occ(Au1) + occ(Ag1) = 0.154 + 0.096 = 1/4 with 4-fold multiplicity

Atom occupancy

Example: Disordered BF_4^- and one H-atom on special position [8]



Refinement = improvement of parameters

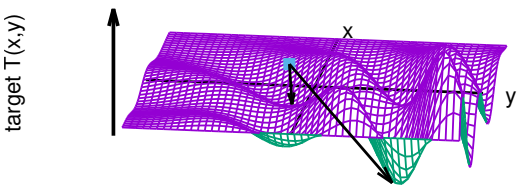
Computationally, refinement minimises the discrepancy between the observed data I_{obs} and the calculated data I_{calc}. I_{calc} is calculated from the model parameters, mainly atom coordinates x, y, z and atomic dispersion parameters ADPs

$$T(\vec{x}_i, U_i, (\text{occupancies}, \ldots)) = \sum_{(hkl)} w(hkl) |I_{obs}(hkl) - I_{calc}(hkl)|^2$$

$w(hkl)$ downweights untrusted reflections, typically $w(hkl) = 1/\sigma_I(hkl)$.
Note: different refinement programs use different target functions.

Least-square-minimisation

The shape of the target function $T = \sum_{(hkl)} w(hkl)|I_{\text{obs}}(hkl) - I_{\text{calc}}(hkl)|^2$ enables optimisation based on least-squares method (`L.S.` command in SHELXL).

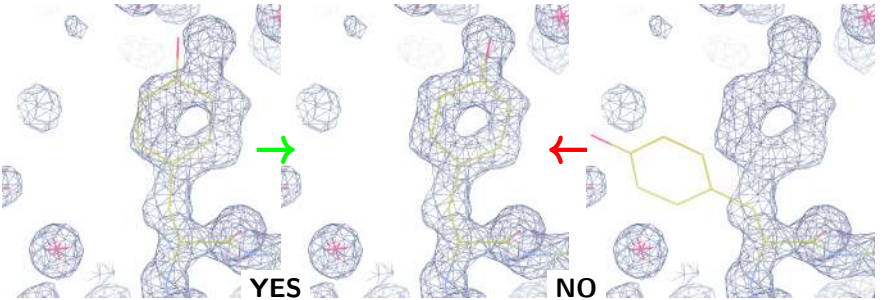


The algorithm finds the next minimum, but cannot jump across humps.

Advantages and limitations of refinement

- Refinement finds the "next" local minimum
- only small changes in the structure
- does not add or remove atoms
- no change of element types
- one never knows whether the optimum is reached. However, for small molecules, the starting model usually converges to a good model.

The “next” local minimum



Model building

- manual modifications "help" refinement cross local humps
  - large movements of individual atoms (out of local traps)
  - delete wrong atoms
  - add missing atoms
  - correct atom type
- model building = add chemical understanding to the model
- graphic programs (Olex2, shelXle, Crystals, JANA2020 ...[9, 10, 11, 12])
- guided by the electron density map

Iterative process: improve model -> refine -> improve model -> refine -> ...

Electron density map and difference map

$$\rho(x,y,z) = FT(|F_{\text{obs}}(hkl)|, \phi_{\text{calc model}}(hkl))$$

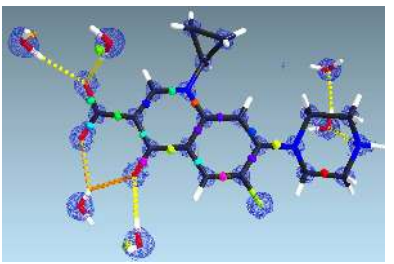
Fourier transformation from measured structure factor amplitudes $|F_{\text{obs}}(hkl)|$ and calculated phases $\phi_{\text{calc model}}(hkl)$
This model should follow this map.
The map

$$\Delta\rho(x,y,z) = FT(|F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)|, \phi_{\text{calc}}(hkl))$$

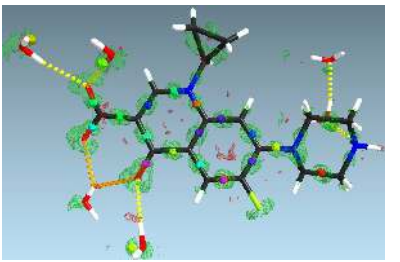
is called **difference map**. It reveals discrepancies between the model and the data.
Model building and refinement aim at reducing these discrepancies.

Example map: Ciprofloxacin

Structure of Ciprofloxacin, [13], ultra high resolution 0.43 Å



$\rho(x, y, z)$ (usually blue mesh)



$\Delta\rho(x, y, z)$ (usually green / red mesh)

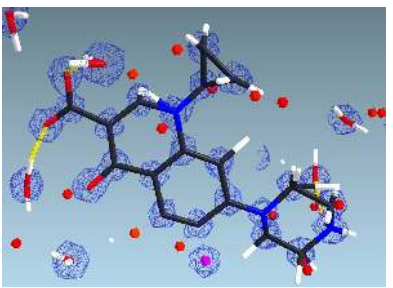
positive $\Delta\rho$: Model misses something. SHELXL places **Q-peaks**

negative $\Delta\rho$: model contains too much

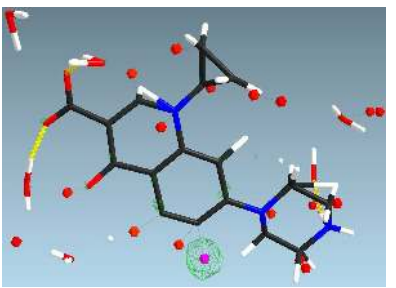
Example map: Ciprofloxacin

Structure of Ciprofloxacin, [13], ultra high resolution 0.43 Å

- data resolution truncated to 0.9 Å
- Fluorine atom F removed from model



$\rho(x, y, z)$ (blue mesh)

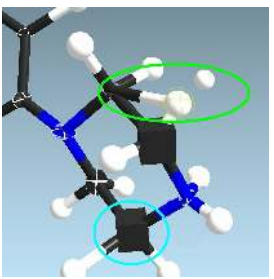


$\Delta\rho(x, y, z)$ (red / green mesh)

Refinement without restraints

$$T(\vec{x}_i, \mathbf{U}_i) = \sum_{(hkl)} w(hkl)(I_{\text{data}}(hkl) - I_{\text{model}}(hkl))^2$$

This formula carries out **unrestrained refinement**, purely taking experimental data into account. With poor data, this **can** cause



- unrealistic bond distances and bond angles
- negative ADPs (cubes) are physically meaningless
- refinement can produce non-sense results

Unrestrained refinement, example

Unrestrained refinement of protein structure with 1.4 Å resolution

Data to parameter ratio

Example Ciprofloxacin, ($a = 9.5$Å, $b = 9.9$Å, $c = 11.0$Å, $\alpha = 94.2°$, $\beta = 100.2°$, $\gamma = 91.3°$)

- $FC_{17}N_3O_9H_{30}$: $60 \times 9 = 540$ parameters

0.43 Å resolution 26'308 reflections. $26'308 : 540 = 48.7$ data points per parameter: very high data to parameter ratio, data sufficient to produce chemically sensible structure

0.8 Å resolution 2'926 reflections. $2'926 : 540 = 5.4$ data points per parameter: low data to parameter ratio, data insufficient to produce chemically sensible structure

Chemically sensible part needs to be **restrained** -> restrained refinement

Restrained refinement

Except for at very high resolution, the refinement program has to be told some chemistry to make sure the model remains chemically meaningful. There are two different types how this can be accomplished:

Constraints Express an equality and permit no deviation from fixed value

Restraints Express similarity and provide some flexibility from target value.

Restraints are much more common than constraints

Constraints

- The structure of $La[Ag_{0.39}Au_{0.61}(CN)_2]_3 \cdot H_2O$ has either gold or silver at one location.
- In every unit cell there is always one atom at this location

$$occ(Au) + occ(Ag) = 1$$
$$occ(Au) = 1 - occ(Ag)$$

- Only the occupancy of silver has to be determined. The occupancy of gold can be calculated (or *vice versa*)
- remark: the program SHELXL uses the command FVAR ("free variables") to realise constraints.

Each constraint reduces the number of parameters by 1

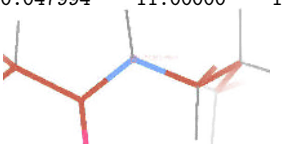
Important constraints

negative ADP value, mainly for hydrogen atoms: U(HA)= 1.2*U(CA)

```
CA    1    0.673087    0.878303    0.111632    11.00000     0.31129
HA    6    0.679625    0.855075    0.095775    11.00000    -1.20000
```

hydrogen positions: `AFIX`

```
N     3    0.611916    1.012005    0.052456    11.00000     0.18165
AFIX  43
H     6    0.628491    1.011598    0.033498    11.00000    -1.20000
AFIX   0
CB    1    0.622779    1.076653    0.067974    11.00000     0.18216
AFIX  23
HB1   6    0.608063    1.103479    0.072220    11.00000    -1.20000
HB2   6    0.641195    1.080130    0.047994    11.00000    -1.20000
AFIX   0
```



AFIX: riding atom model

- Except for at very high resolution ($d \ll 0.8$ Å), hydrogen atoms are invisible to X-rays
- the positions of most hydrogen atoms can be calculated: bond distances are known from spectroscopy, positions are determined by reducing steric clashes
- Advantages: hydrogen atoms do not add parameters, the contribute to VdW repulsion (BUMP command), they have a small, but non-zero contribution to the scattering.

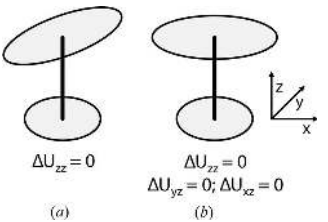
Restraints: Geometry

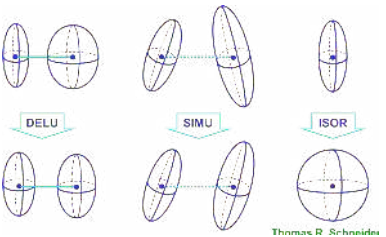
- restraints can be expressed as inequality "$\leq$"
- best known restraints: R. A. Engh, R. Huber, *Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement*, Acta Crystallogr. (1991), **A**47, pp. 392–400; *e.g.*

$$|d(N, C_\alpha) - 1.458\text{Å}| \leq 0.02 \qquad |d(C_\alpha, C_\beta) - 1.521\text{Å}| \leq 0.02$$

Restraints: ADP values [14]

- restraints for ADPs: chemical bond affects thermal vibrations





DELU, SIMU, ISOR (Thomas R. Schneider)

Restraints resemble data

Restraints are treated with additional terms to the target function:

$$T(\vec{x}_i, \mathbf{U}_i) = \sum_{hkl} w_{hkl}(I_{\text{data}}(hkl) - I_{\text{model}}(hkl))^2 + W \sum_{\text{N.B. } i} w_i(T_i^{\text{data}} - \langle T_i \rangle)^2$$

Restraints act like additional data points

- W weights restraints and observed data
- the higher the resolution, the lower weight W
- the expected mean values $\langle T_i \rangle$ can be derived statistically from high resolution structures, or sometimes can be computed quantum chemically

Summary refinement & model building

- model building improves the model in large steps
- refinement optimises the model against the data
- constraints and restraints are used to ensure a chemically reasonable model
- constraints reduce the number of parameters, restraints act like data: both increase the data to parameter ratio

5 Model quality and data quality: structure validation

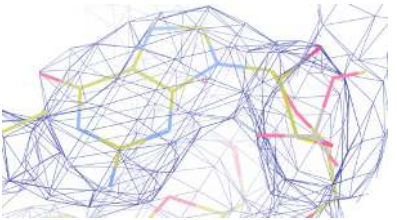
Atom coordinates $\neq$ model accuracy

Guanine model in ribosome, data resolution 3.1Å

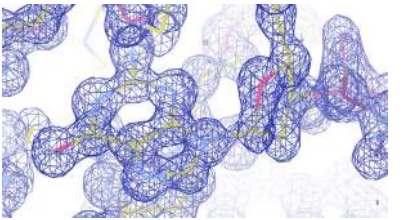
Guanine model in Z-DNA, at resolution 1.0 Å

The coordinates of the model do no reveal the data quality, nor the model quality.

Model coordinates = interpretation of data



Guanine model **with map** in ribosome, data resolution 3.1Å



Guanine model **with map** in Z-DNA, at resolution 1.0 Å

Only in combination with the data can we judge the model quality

Once more: data to parameter ratio

Example Ciprofloxacin ($a = 9.5$Å, $b = 9.9$Å, $c = 11.0$Å, $\alpha = 94.2°$, $\beta = 100.2°$, $\gamma = 91.3°$)

- $FC_{17}N_3O_9H_{30}$: $60 \times 9 = 540$ Parameter

data resolution 0.43 Å: 26'308 reflections $\hat{=}$ 48.7 data points per parameter: very high, reliable refinement

data resolution 0.8 Å: 2'926 reflections $\hat{=}$ 5.4 data points per parameter: medium, refinement needs checking

Once more: data to parameter ratio

Example Ribosome ($a = 401.4$Å, $b = 401.4$Å, $c = 175.9$Å, $\alpha = \beta = \gamma = 90°$, $P4_12_12$)

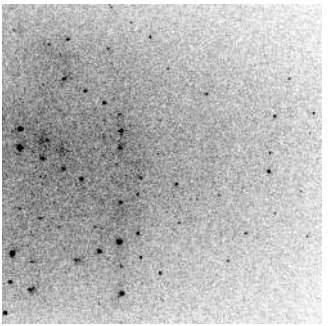
- PDB ID 1J5E: 51'atoms atoms = 207'768 parameters
- data resolution 3.05 Å 238'205 reflections

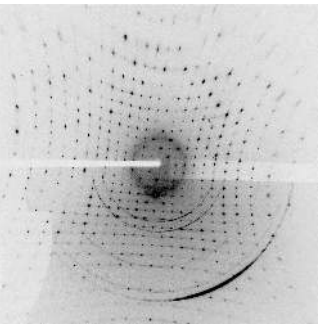
$$\frac{238'205}{207'768} = 1.15$$

Even at such low data to parameter ratio can a reasonable model be built and refined. It is important to be aware of differences in the interpretation of the data

6 Indicators for data quality

Example data quality





Important quality indicators

R_{meas} relative difference between symmetry equivalent reflections and their mean value

data completeness : fraction of measured data w.r.t. theoretically possible data

multiplicity (*alias*: *redundancy*): how often every unique reflection was measured (on average)

signal strength $I(hkl)/\sigma_{I(hkl)} < 1$: noise

$CC_{1/2}$
1. split data set into two random halves
2. calculated correlation coefficient between symmetry equivalent reflections

R-values for data

The classic data quality indicator is R_{int}, alias R_{merge} or R_{sym}:

$$R_{\text{int}} = \sum_h \sum_j \frac{|I_{hj} - \langle I_h \rangle|}{\langle I_h \rangle}$$

R_{int} mathematically increases with multiplicity, although data quality improves with multiplicity

R_{int} is typically shown in publications. It is, however, obsolete and should not be published. R_{meas} *alias* $R_{\text{r.i.m.}}$ should be published instead:

$$R_{\text{meas}} = \sum_h \frac{n_h}{n_h - 1} \sum_j \frac{|I_{hj} - \langle I_h \rangle|}{\langle I_h \rangle}$$

Example data statistics (XPREP)

Resolution	#Data	#Theory	%Complete	Redundancy	Mean I	Mean I/s	Rmerge
Inf - 2.46	196	197	99.5	39.27	215.01	110.27	0.0300
2.46 - 1.13	1762	1825	96.5	14.86	75.32	42.01	0.0453
1.13 - 0.89	1972	2123	92.9	8.71	25.52	19.00	0.0895
0.89 - 0.77	2007	2258	88.9	6.81	10.84	10.39	0.1425
0.77 - 0.69	1864	2499	74.6	3.37	5.66	5.76	0.1885
0.69 - 0.62	2108	3360	62.7	2.24	2.88	3.29	0.2890
0.62 - 0.57	1929	3542	54.5	1.44	1.51	1.79	0.4191
0.57 - 0.54	1123	2367	47.4	1.10	0.90	1.14	0.5593
0.64 - 0.54	3720	7014	53.0	1.43	1.47	1.76	0.4170
Inf - 0.54	12961	18171	71.3	5.08	20.64	13.61	0.0514

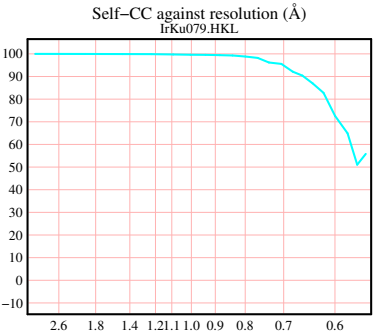
Merged [A], lowest resolution = 11.49 Angstroms

CC1/2, and resolution cut-off

A good quality crystal diffracts beyond the theoretical limit $d_{\text{min}} = \lambda/2$. Resolution cut-off is not an issue, one can use all data. Large complexes, supramolecular structures, low quality crystals reach the diffraction limit before the theoretical limit. One has to decide where to cut the diffraction data.

- CC1/2 should be close to 100% throughout resolution range
- where CC1/2 drops below 70%, noise becomes significant, and data at higher resolution can be excluded from refinement
- $I/\sigma(I)$ should be about 2, where CC1/2 about 70%
- $I/\sigma(I)$ should be about 1, where CC1/2 about 40% (in cases very resolution cut-off is critical)

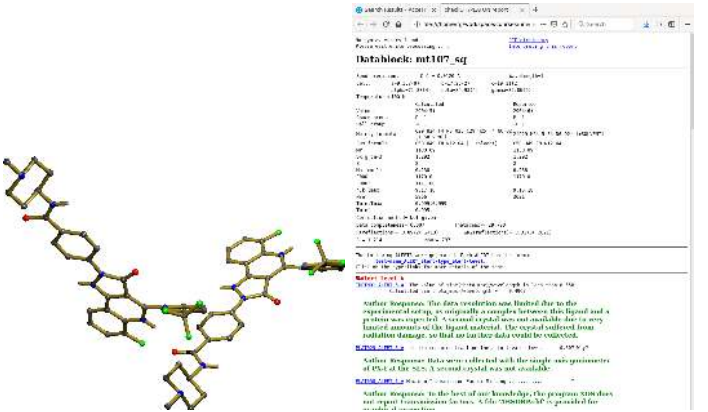
Example $CC_{1/2}$, and resolution cut-off



CC1/2 *vs.* data resolution; plot generated with XPREP

7 Indicators for model quality

Model quality [15]



R-values for the model

$$R = R1 = \sum_h \frac{||F_h(data)| - |F_h(model)||}{|F_h(data)|}$$

weighted intensity based R-value:

$$wR2 = R_B = \sqrt{\sum_h \frac{|w_h(I_h(data) - I_h(model))^2|}{w|I_h(data)|^2}}$$

small molecules $R1$ of the refined model 2-5 %.

supramolecules compounds, MOFs, ... $R1$ of the refined model **can** be highter, 2-15 %

macromolecular compounds $R1$ of the refined model 15-25 %

To a great extent, this discrepancy is due to the unmodelled solvent region in the latter two types of compounds

Goodness of Fit — GooF

$$GooF = \sqrt{\frac{\sum_h w_h \left({F_h}^2(data) - {F_h}^2(model)\right)^2}{n-p}}$$

- Takes number of parameters (p) and number of data (n) into account
- Ideally ≈ 1, increases with worse model

model: residual density

SHELXL calculates the "highest peak" and "deepest hole" in the electron density map. Units are electrons, e.g. at the **beginning** of model building:

```
Electron density synthesis with coefficients Fo-Fc

Highest peak     4.95  at  0.5434  0.9981  0.3231  [  0.04 A from RU01
]
Deepest hole    -3.34  at  0.0057  0.4976  0.3299  [  0.99 A from RU02
]
^^^^^^
Mean =    0.00,   Rms deviation from mean =    0.34 e/A^3
^^^^^^^^^^^^
```

model: residual density

SHELXL calculates the "highest peak" and "deepest hole" in the electron density map. Units are electrons, e.g. for the **refined** model:

```
Electron density synthesis with coefficients Fo-Fc

Highest peak    0.50  at  0.6610  0.1969  0.4278  [  0.69 A from C006
]
Deepest hole   -1.22  at  0.2635  0.6156  0.2132  [  0.04 A from P003
]
^^^^^^
Mean =    0.00,   Rms deviation from mean =    0.06 e/A^3
^^^^^^^^^^^^
```

checkCIF `https://checkcif.iucr.org/`

Every published structure *should* have a checkCIF report. There are different alert levels of decreasing severity. Reviewers typically require that a structure should **not** contain A- or B-alerts.

Summary Validation

- A model without data does not reflect data quality
- Data quality: data resolution, multiplicity, R-values, I/σ_I, $CC_{1/2}$
- Model quality: R1-values, GooF, residual density
- available for everyone: checkCIF `http://checkcif.iucr.org` (with or without data)
- *ALERT levels* A, B, ...
- (Analogously for macromolecular structures: `http://molprobity.biochem.duke.edu/`)

8 Overview of additional topics in crystallography

Anomalous dispersion and chirality	Ariëns [16] and Spek [17]
Twinning	Sevvana et al. [18] and Nespolo, Ferraris and Souvignier [19]
Polymorphism, crystal engineering	Bernstein [20], Desiraju [21] and Hilfiker and Raumer [22]
Incommensurate crystals and quasicrystals	Janssen, Chapuis and Boissieu [23] and Steurer and Deloudi [24]
High-pressure crystallography	Katrusiak [25]
Quantum crystallography (charge density refinement)	Grabowsky, Genoni and Bürgi [26]
Neutron Crystallography	Blakeley [27]
Electron Crystallography	Gemmi et al. [28] and Gruene et al. [29]

References

[1] R. Read. *Protein Crystallography Course. The phase problem: introduction to phasing methods*. URL: https://www-structmed.cimr.cam.ac.uk/Course/Basic_phasing/Phasing.html (visited on 07/06/2022).

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

[3] 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.

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

[5] 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.

[6] K. Dalle et al. 'A weakly coupled biologically relevant $Cu_2^{II}(\mu-\eta^1:\eta^1-O_2)$ *cis*-peroxo adduct that binds side-on to additional metal ions'. In: *J. Am. Chem. Soc.* 136 (2014), pp. 7428–7434. DOI: 10.1021/ja5025047.

[7] Julie Clarissa F. Colis et al. 'Structural studies of lanthanide ion complexes of pure gold, pure silver and mixed metal (gold–silver) dicyanides'. In: *Dalton Trans.* (2005), pp. 675–679. DOI: 10.1039/B413967D.

[8] Tim Gruene et al. 'Rapid structure determination of microcrystalline molecular compounds using electron diffraction'. In: *Angew. Chem., Int. Ed.* 57 (2018), pp. 16313–16317. DOI: 10.1002/anie.201811318.

[9] *Olex2*. Version 1.5. URL: https://www.olexsys.org/olex2 (visited on 07/06/2022).

[10] C. Hübschle. *ShelXle*. URL: https://www.shelxle.org (visited on 07/06/2022).

[11] Václav Petříček, Michal Dušek and Lukáš Palatinus. *JANA2020*. URL: http://jana.fzu.cz (visited on 07/06/2022).

[12] Paul W. Betteridge et al. *CRYSTALS*. URL: http://www.xtl.ox.ac.uk/crystals.html (visited on 07/06/2022).

[13] Julian Jacob Holstein, Christian Bertram Hübschle and Birger Dittrich. 'Electrostatic properties of nine fluoroquinolone antibiotics derived directly from their crystal structure refinements'. In: *CrystEngComm* 14 (2012), pp. 2520–2531. DOI: 10.1039/C1CE05966A.

[14] Andrea Thorn, Birger Dittrich and George M. Sheldrick. 'Enhanced rigid-bond restraints'. In: *Acta Crystallogr* A68 (2012), pp. 448–451. DOI: 10.1107/S0108767312014535.

[15] Marco F. Taddio et al. 'Synthesis and Structure–Affinity Relationship of Small Molecules for Imaging Human CD80 by Positron Emission Tomography'. In: *J. Med. Chem.* 62 (2019), pp. 8090–8100. DOI: 10.1021/acs.jmedchem.9b00858.

[16] E. J. Ariëns. 'Stereochemistry, a basis for sophisticated nonsense in pharmacokinetics and clinical pharmacology'. In: *Eur. J. Clin. Pharmacol.* 26 (1984), pp. 663–668.

[17] A. L. Spek. 'Absolute structure determination: pushing the limits'. In: *Acta Crystallogr.* B72 (2016), pp. 659–660. DOI: 10.1107/S2052520616014773.

[18] Madhumati Sevvana et al. 'Non-merohedral twinning: from minerals to proteins'. In: *Acta Crystallogr.* D75 (2019), pp. 1040–1050. DOI: 10.1107/S2059798319010179.

[19] Massimo Nespolo, Giovanni Ferraris and Bernd Souvignier. 'Effects of merohedric twinning on the diffraction pattern'. In: *Acta Crystallogr.* A70 (2014), pp. 106–125. DOI: 10.1107/S2053273313029082.

[20] J. Bernstein. *Polymorphism in Molecular Crystals*. Oxford University Press, 2007. ISBN: 9780199236565.

[21] Gautam R. Desiraju. 'Cryptic crystallography'. In: *Nat. Mater.* 1 (2002), pp. 77–79. DOI: 10.1038/nmat726.

[22] R. Hilfiker and M. v. Raumer, eds. *Polymorphism in the Pharmaceutical Industry. Solid Form and Drug Development*. Wiley-VCH, 2019. ISBN: 978-3-527-34040-8.

[23] T. Janssen, G. Chapuis and M. de Boissieu. *Aperiodic crystals*. Oxford University Press, 2018. ISBN: 9780198824442.

[24] Walter Steurer and Sofia Deloudi. 'Fascinating quasicrystals'. In: *Acta Crystallogr.* A64 (2008), pp. 1–11. DOI: 10.1107/S0108767307038627.

[25] Andrzej Katrusiak. 'High-pressure crystallography'. In: *Acta Crystallogr.* A64 (2008), pp. 135–148. DOI: 10.1107/S0108767307061181.

[26] Simon Grabowsky, Alessandro Genoni and Hans-Beat Bürgi. 'Quantum crystallography'. In: *Chem. Sci.* 8 (2017), pp. 4159–4176. DOI: 10.1039/C6SC05504D.

[27] M. P. Blakeley. 'Neutron macromolecular crystallography'. In: *Crystallogr. Rev.* 15 (2009), pp. 157–218. DOI: 10.1080/08893110902965003.

[28] M. Gemmi et al. '3D Electron Diffraction: The Nanocrystallography Revolution'. In: *ACS Cent. Sci.* 5 (2019), pp. 1315–1329.

[29] T. Gruene et al. 'Establishing electron diffraction in chemical crystallography'. In: *Nat. Rev. Chem.* accepted (2021), n/a. DOI: 10.1038/s41570-021-00302-4.