Chemical Crystallography and Structural Chemistry

VO 270287

Lecture № 10 — 24th June 2021

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

tim.gruene@univie.ac.at

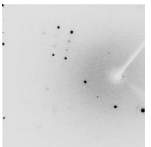
Reminder: Start the recording!

Contents

1 Model quality and data quality: structure validation

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
```

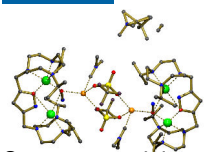
```
0 0 -1      2.8 0.55
0 0  1      3.8 0.63
0 0 -2 1'432.0 95.7
0 0 -2 1'282.0 85.9
```

several GB

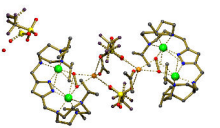
several files, 100's MB

1 “hkl”-file, 50MB

➡ **Phasing** ➡ **Refinement**

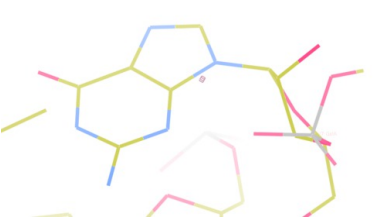


Starting model

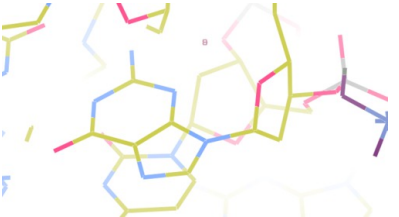


Chemically sensible model

Atom coordinates $\neq$ model accuracy



Guanine model in ribosome, data resolution 3.1Å

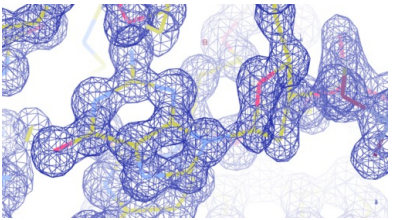


Guanine model in Z-DNA, at resolution 1.0 Å

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, $\alpha = \beta = \gamma = 90°$, $P4_12_12$)

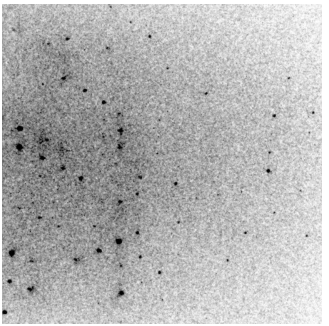
- PDB ID 1J5E: 51'atoms atoms = 207'768 parameters
- data resolution 3.05 Å 238'205 reflections

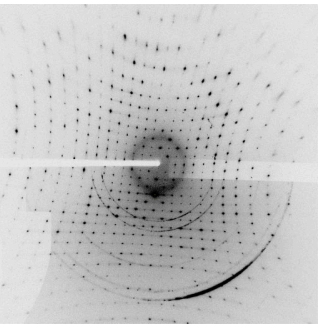
$$\frac{238'205}{207'768} = 1.15$$

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

2 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* $R_{r.i.m.}$ 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

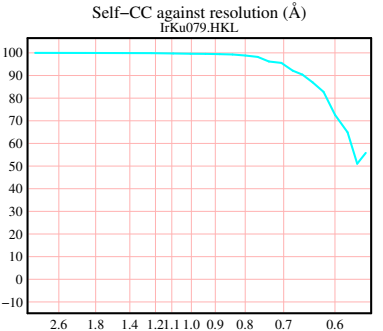
Merged [A], lowest resolution = 11.49 Angstroms

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)

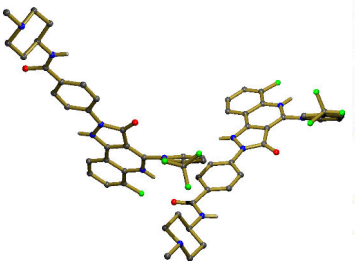
Example $CC_{1/2}$, and resolution cut-off

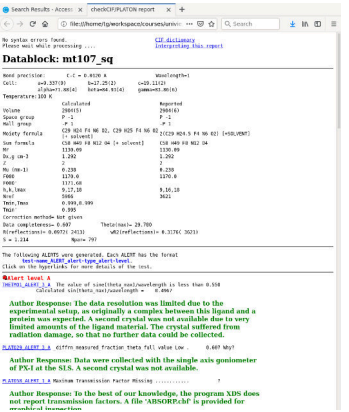


CC1/2 *vs.* data resolution

3 Indicators for model quality

Model quality [1]





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 %.

GooF

Goodness of Fit

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

- 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
^^^^^^^^^^^^
```

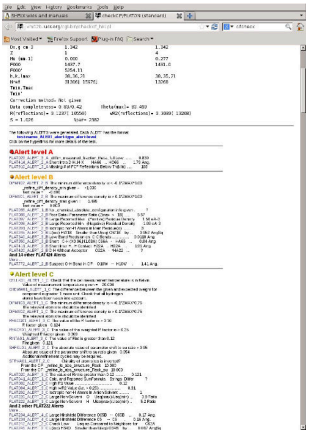
model: residual density

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/`)

4 Overview of additional topics in crystallography

Anomalous dispersion and chirality	Ariëns [2] and Spek [3]
Twinning	Sevvana et al. [4] and Nespolo, Ferraris and Souvignier [5]
Polymorphism, crystal engineering	Bernstein [6], Desiraju [7] and Hilfiker and Raumer [8]
Incommensurate crystals and quasicrystals	Janssen, Chapuis and Boissieu [9] and Steurer and Deloudi [10]
High-pressure crystallography	Katrusiak [11]
Quantum crystallography (charge density refinement)	Grabowsky, Genoni and Bürgi [12]
Neutron Crystallography	Blakeley [13]
Electron Crystallography	Gemmi et al. [14] and Gruene et al. [15]

5 Electron Crystallography

3D Electron Crystallography

- $>$ late 1990s, as opposed to 2D electron crystallography
- confusingly many terms (ADT , RED, EDT, PEDT, MicroED, ...)
- Historical "dispute", Ute Kolb, Mainz University, $\approx$ 2007 (ADT), Xiaodong Zou & Sven Hovmøller, Stockholm University $\approx$ 2011 (RED)
- technical term: "3D Electron Diffraction", Enrico Mugnaioli (PSI 2017; IUCrJ (2019), 6, 178–188)
- "3D":
  1. Collection of 3D reciprocal space
  2. 3D crystals: $\geq 10-15$ unit cell in each direction; typically 200–1000nm per dimension

Spot Position

- Spots positions according to Laue Conditions and orientation of Unit Cell:

$$(\vec{S}_o - \vec{S}_i).\vec{a} = h$$
$$\textbf{and } (\vec{S}_o - \vec{S}_i).\vec{b} = k$$
$$\textbf{and } (\vec{S}_o - \vec{S}_i).\vec{c} = l$$

- Monochhromatic wave: $\vec{S} = (\vec{S}_o - \vec{S}_i)$ depends on wavelength λ and experimental geometry
- Spot **position** $\Leftrightarrow$ Crystal lattice, independent from radiation type
- Resolution d_{hkl} of a spot from position on detector *via* Bragg's law, $\lambda = 2d_{hkl}\sin(\theta)$

Spot Intensity

- Spots **intensity** depends on physics of interaction

  X–rays interact with electrons, crystallographic map corresponds to electron density (number of electron per volume, e^-/A^3).

  Electrons interact with electrostatic potential from electrons + nucleus ($\varphi(\vec{r})$)

  Neutrons interact with nucleus *via* weak interaction, and magnetic moment.

- Spot intensity $\Leftrightarrow$ Unit cell content: where are the atoms, what type of atoms are they

3D Electron Crystallography (3D ED)

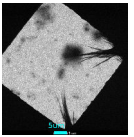
- Electrons interact with electrostatic potential
- Electrons interact **much,much** stronger with matter than X-rays

$\Rightarrow$ Much smaller crystals

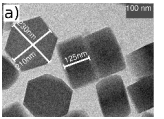
$\Rightarrow$ problematic: dynamic diffraction, $|F| \neq \sqrt{I}$

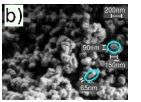
- Electron optics enable some special applications and tiny beam (5 nm diameter)

3D ED: small crystals

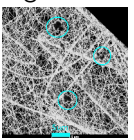


organic compound

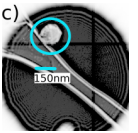


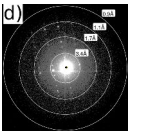


Silicalite–1 / ZSM–5 (Teng Li)

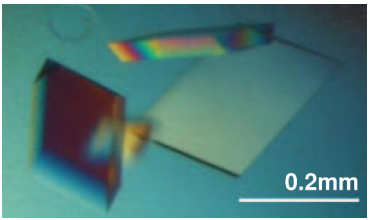


sucrose

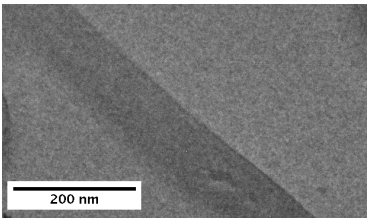




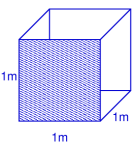
How small is “nano”?

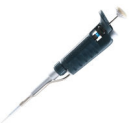


typical protein crystal size for X-rays



typical protein crystal size for electrons, 100x140x1,700 nm^3





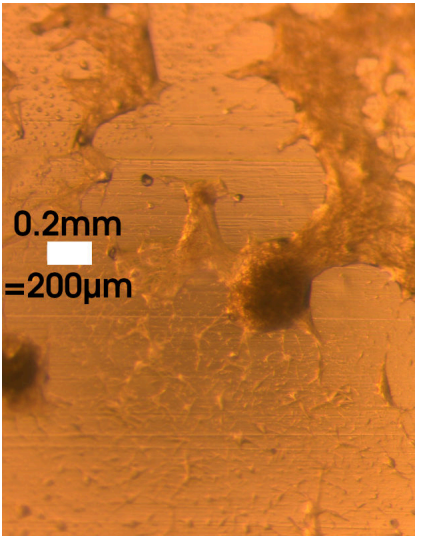
volumes compare like $1m^3$ or 6 bath tubs of water *vs.* $10\mu l$

6 Example Structures

6.1 Nd-MOF

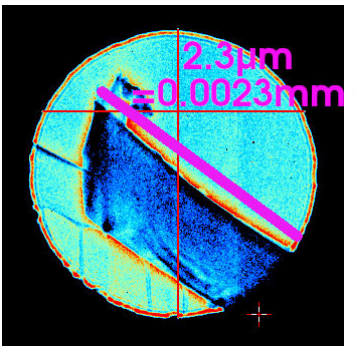
Prof. Jia Min Chin & Prof. Michael Reithofer, University of Vienna

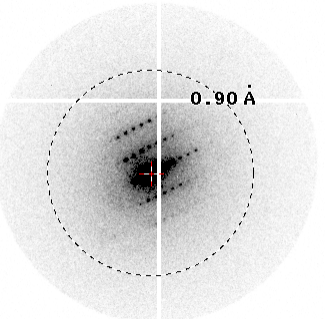
Jewels in the mud



photographs courtesy Dipl.-WIng. A. Roller

Powerful electron diffraction

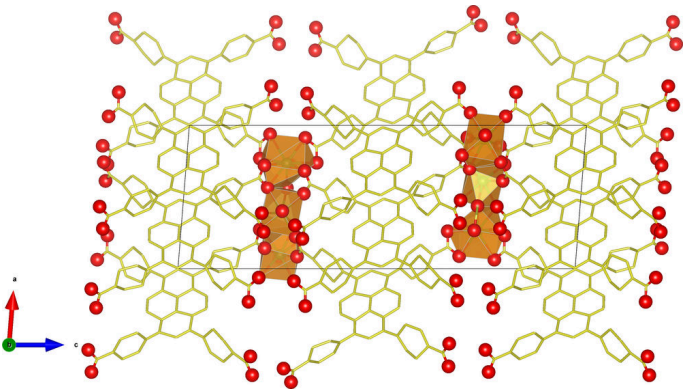




Sample preparation: A. Roller & N. Gajic

At DESY, the strongest X-ray source in the world, this crystal would probably not show any diffraction.

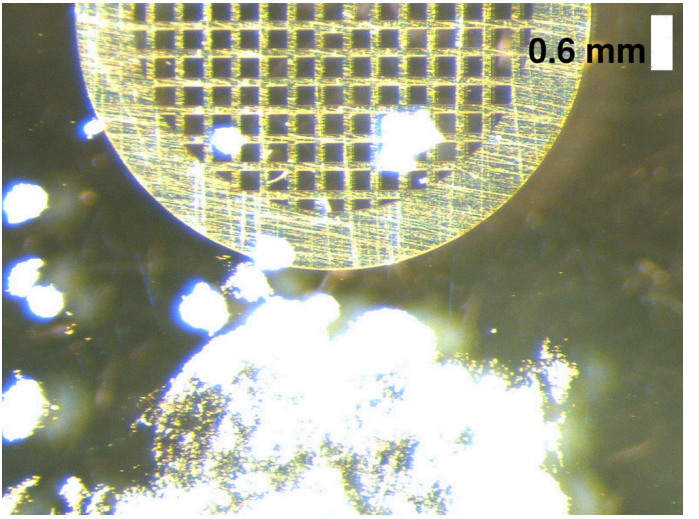
Nd-MOF structure from 5 crystals



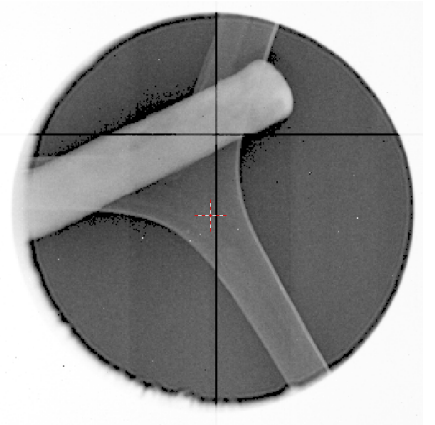
Courtesy Jia Min Chin & Michael Reithofer, unpublished data

Room temperature measurement, under vacuum

6.2 Oseltamivir (Roche)

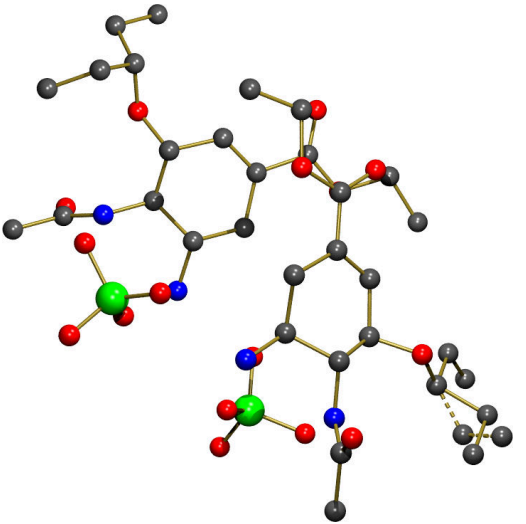


Oseltamivir (Roche)



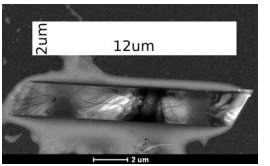
Diameter field of view: 4 µm

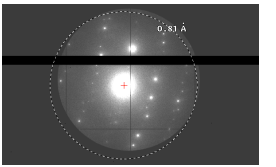
Oseltamivir (Roche)



Spacegroup $P2_12_12$, $a = 23.72$ Å, $b = 24.36$ Å, $c = 7.39$ Å

6.3 Grippostad®, STADA [16]



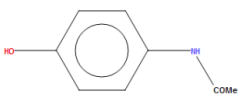


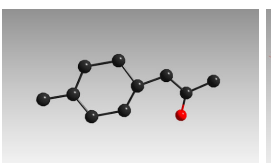
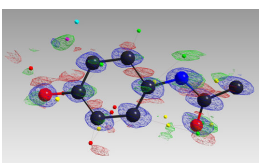
active compounds	non-active compounds
paracetamol	gelatine
ascorbic acid	glycerol tristearate
caffeine	lactose monohydrate
chlorphenamine maleate	quinoline yellow (E104)
	erythrosine (E127)
	titanium dioxide (E171)

- powder from capsule deposited on sample grid
- Crystal dimensions $2\mu m \times 12\mu m \times \approx 300nm$
- $d_{min} < 0.8$Å

Single Crystal Structure from a Pharmacy Powder

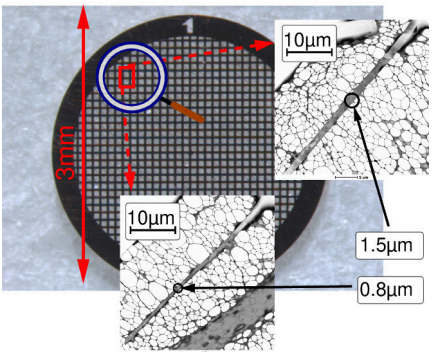
1. Data from single crystal: Completeness $< 40\%$
2. Cell parameters: $a = 6.9$, $b = 9.4$, $c = 11.6$, $\alpha = 90.6$, $\beta = 98.4$, $\gamma = 89.8$
   CSD search $a = 7.1$, $b = 9.3$, $c = 11.7$, $\alpha = 90.0$, $\beta = 97.7$, $\gamma = 90.0$
   CCDC HXACAN04, $P2_1/n$, Paracetamol,
3. SHELXT solves structure
4. Difference map reveals hydrogen atoms: data sensitivity



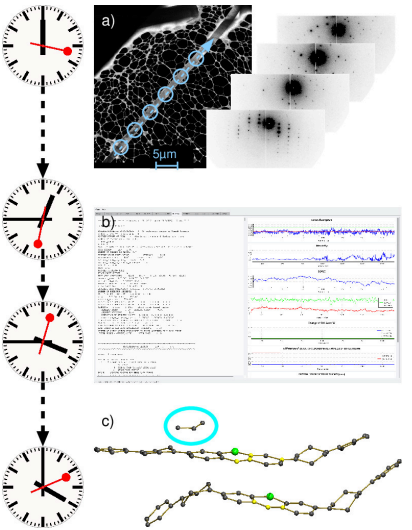
Future: Complete crystallographic analysis from powder blends

6.4 Drug Design: Structure of a New Methylene Blue Derivative $MBBF_4$ [16]



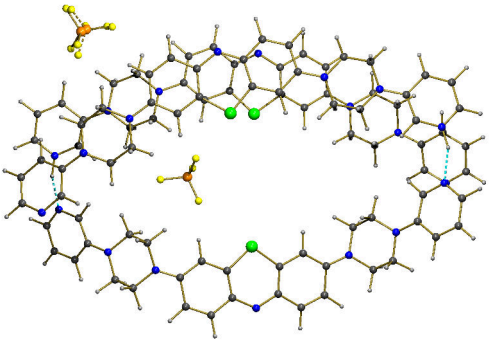
Tip of thin $MBBF_4$ needle on a TEM sample grid

$MBBF_4$ — EIGER and a TEM make a Synchrotron



- $60 - 120°$ @ $3°/s = 40\ s$ / data set
- 45 min for 16 data sets on both grids
- manual processing $\approx$ 4h to structure solution

Structure of $MBBF_4$ (Refinement J. Holstein, TU Dortmund)



- $R1 = 22.7\%(2941 F_o > 4\sigma_F)$
- $R1 = 27.2\%(4832 F_o)$
- GooF = 1.5

References

[1] 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.

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

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

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

[5] 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.

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

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

[8] 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.

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

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

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

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

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

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

[15] 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.

[16] 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.