Crystallography for Structural Biologists

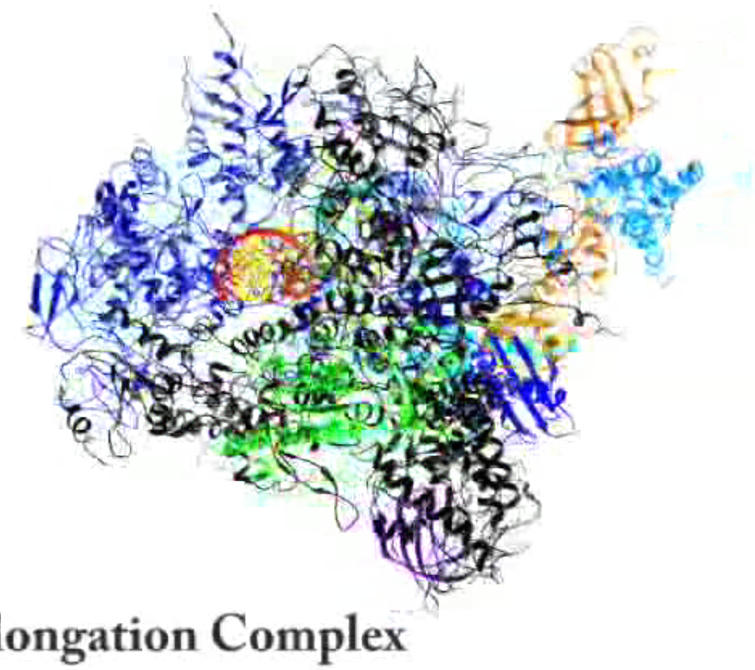
ADTB workshop, Grenoble, 19th November 2019

presented by:
Dr. Tim Grüne
Head of the Centre for X-ray Analysis
Faculty of Chemistry
University of Vienna

tim.gruene@univie.ac.at

Learning from Crystal Structures

RNA Polymerase II: Crystal “Snapshots”

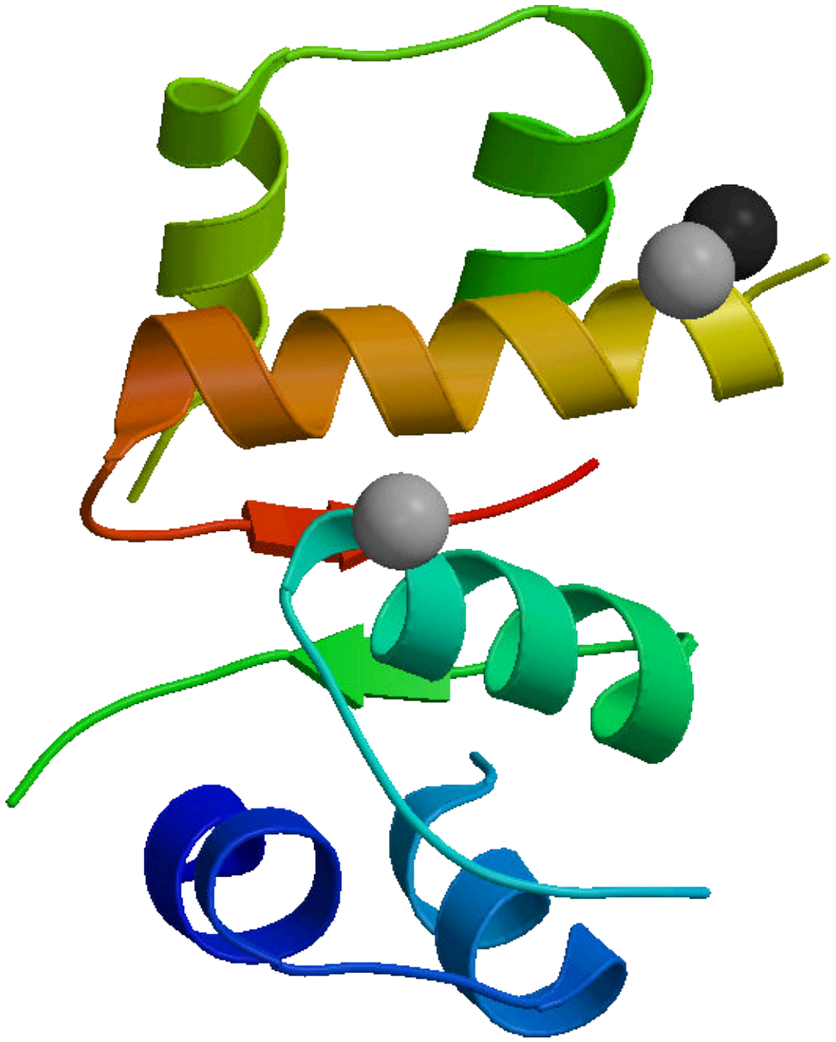


Several structures of RNA Polymerase II in different states of action lead to a concept of the mode of function.

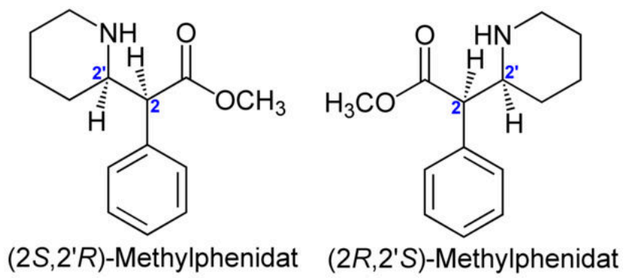
(Brueckner F et al.: “A movie of the RNA polymerase nucleotide addition cycle.” Curr Opin Struct Biol 19, 294-299 (2009).) Courtesy P. Cramer, MPI Göttingen

Insulin: Quality Control

- 1982: production of recombinant human insulin (improvement of tolerance compared to bovine insulin)
- recombinant and purified human insulin structurally identical
- structure based point-mutations of insulin improve functionality (*e.g.* rate of release). An extensive list can be found at http://de.wikipedia.org/wiki/Insulinpräparat (sorry, no translations provided).



Small Molecules: Handedness and Purity

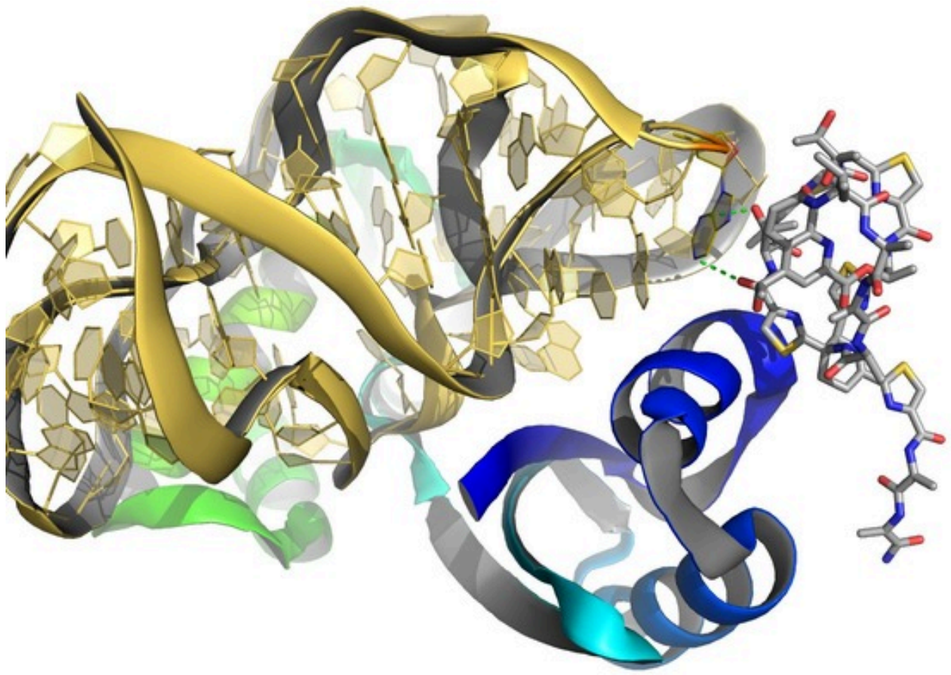


http://de.wikipedia.org/wiki/Methylphenidat

- Methylphenidate (*aka* Ritalin): drug to treat **attention-deficit hyperactivity disorder** (ADHD)
- Contains two stereochemical centres, *i.e.* there are four different forms
- Often only one form has the desired effect, others often contribute to (undesired) side-effects
- see *e.g.* E. J. Ariëns: *Stereochemistry, a basis for sophisticated nonsense in pharmacokinetics and clinical pharmacology*, European Journal of Clinical Pharmacology, **26** (1984), pp. 663–668.

Crystallography is the only experimental technique to determine chirality.

Structure Guided Drug Design

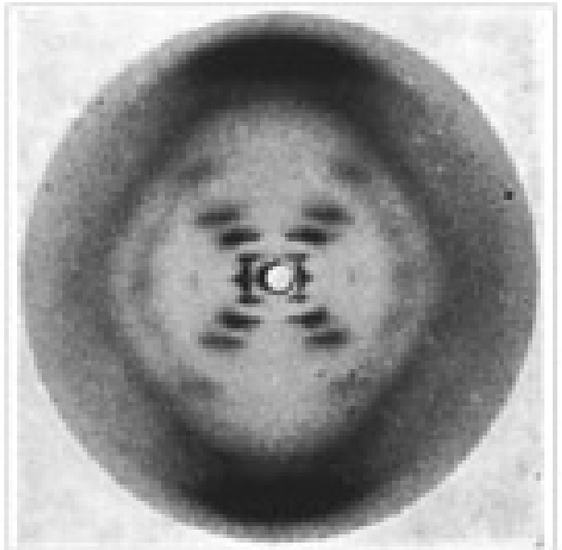


The antibiotic Thiostrepton in contact with its target DNA. Image courtesy Dr. K. Pröpper.

Atomic coordinates for ligand and target enable

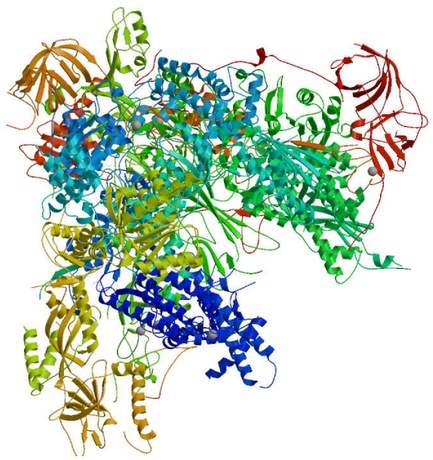
- fine-tuning of contact
- fine-tuning of shape: influence mode of function **and** access towards target.

DNA Double Helix

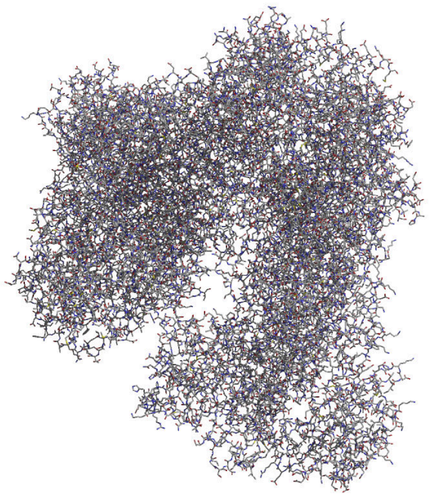


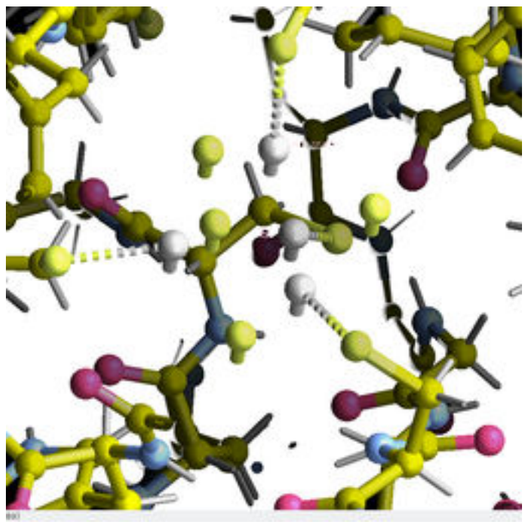
- X-ray image of fibrous, crystalline DNA by R. Franklin, which led her with co-workers and Watson/Crick to the double-helical structure of DNA
- The model is often considered the “birth of modern molecular biology” (Voet & Voet, *Biochemistry* (1995), Wiley & Sons).

Resolution: Details of Crystal Structure

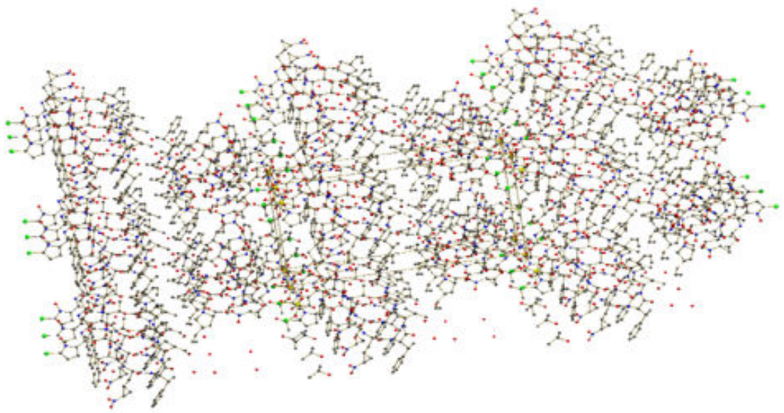


PDB 4C3I: RNA Polymerase I, 3Å





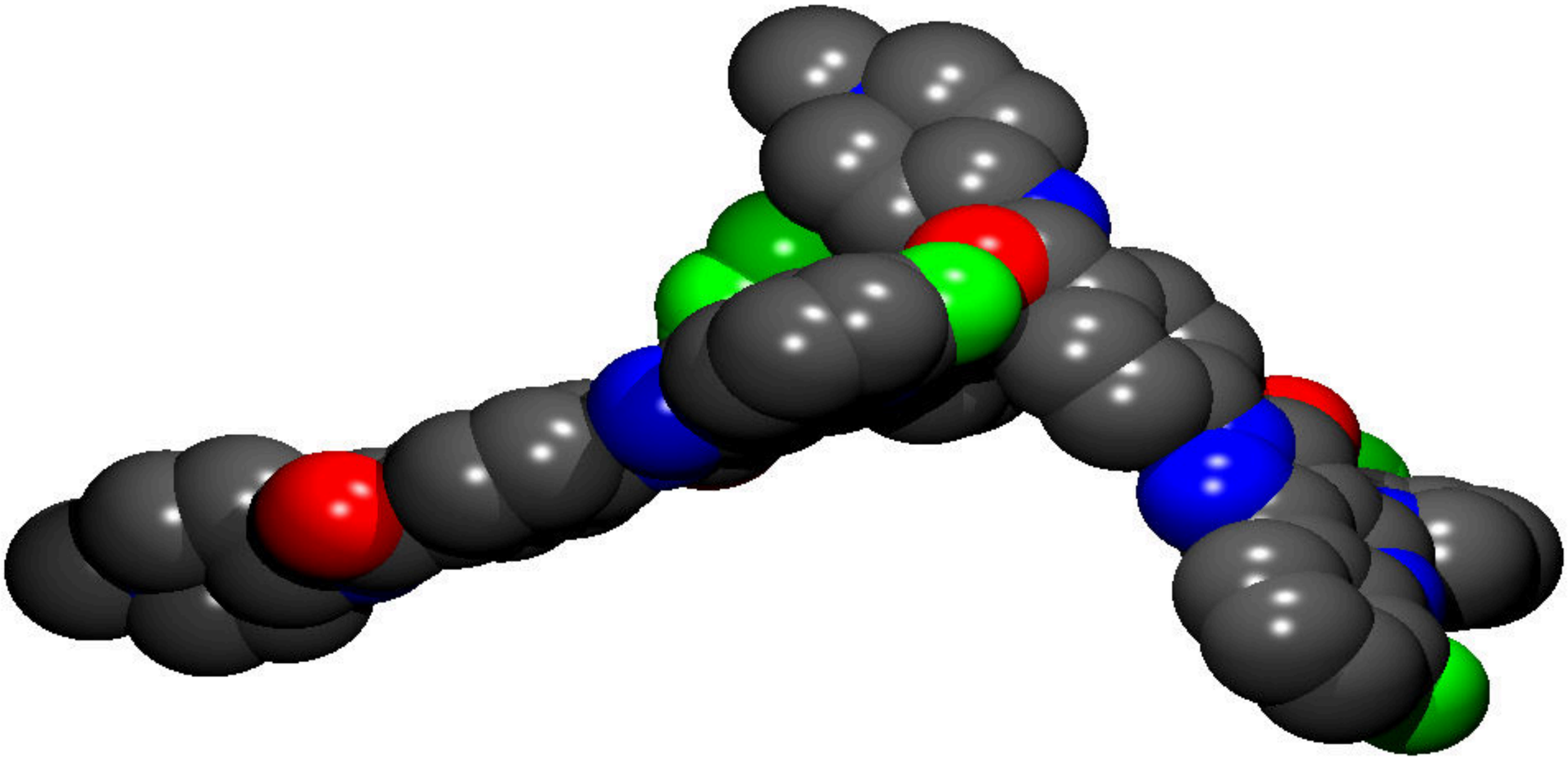
PDB 5WQQ: FeS cluster, 0.8Å



CSD TUCMEL: Hormaomycin, 1Å

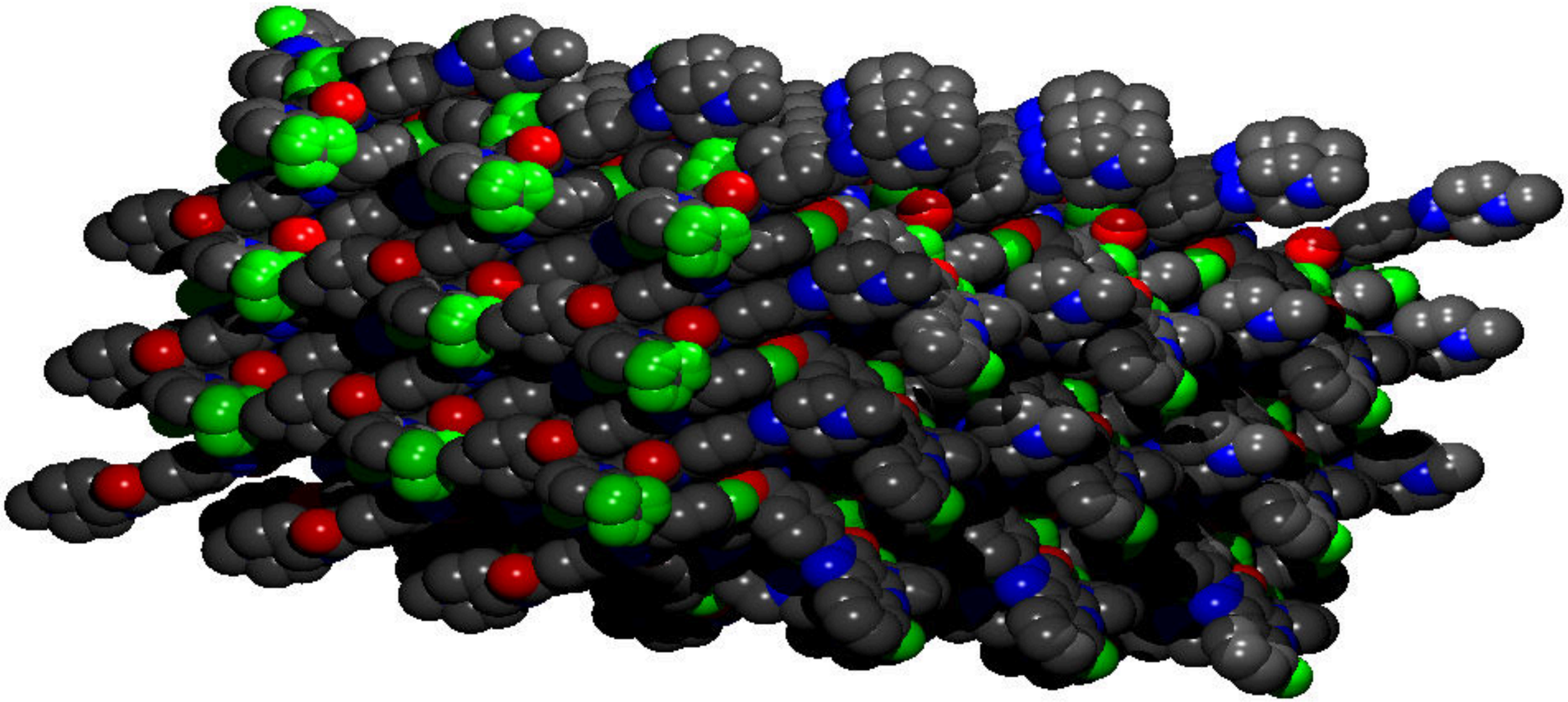
What are Crystals?

Starting point: Molecule / Compound



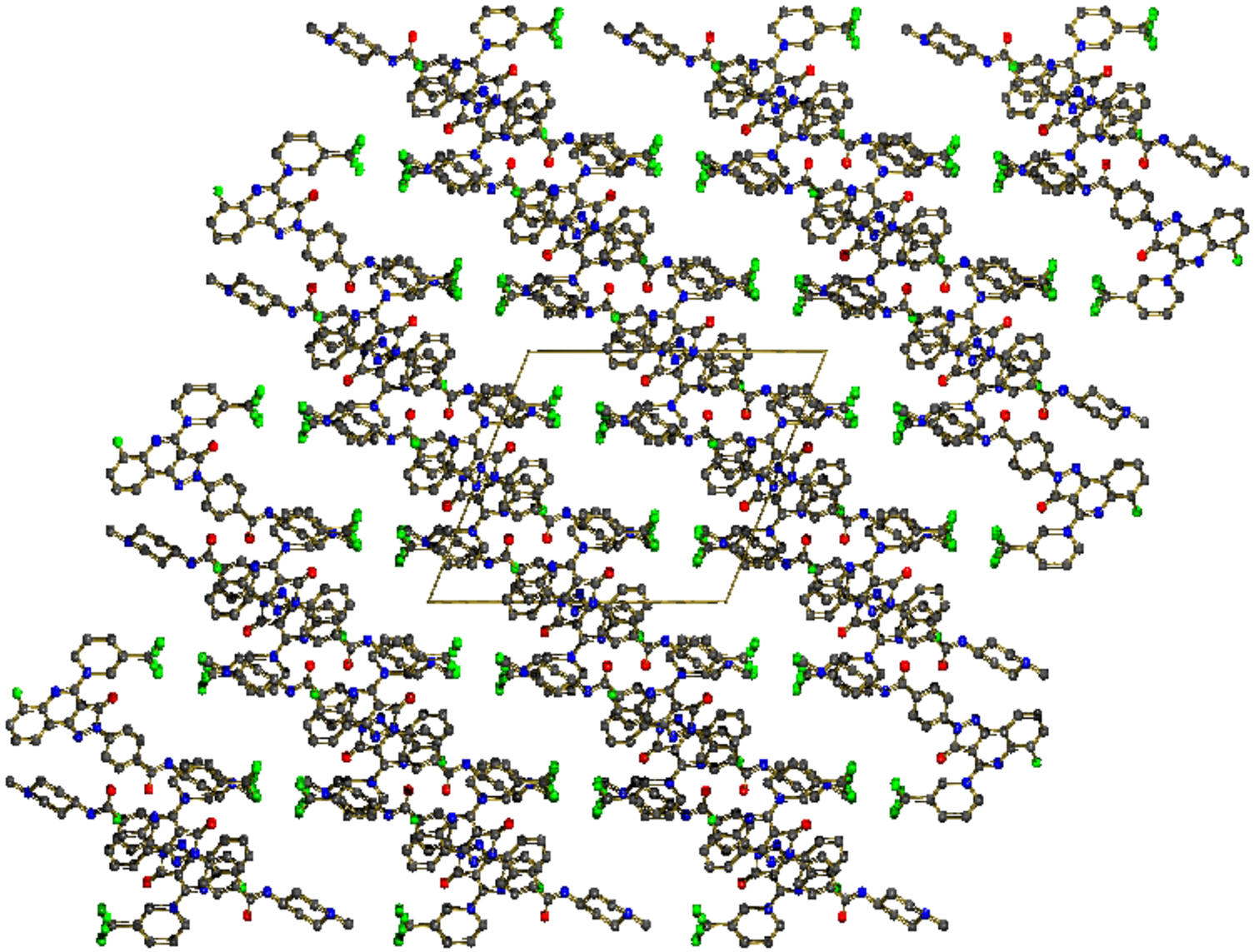
M. Taddio *et al.*, DOI 10.1021/acs.jmedchem.9b00858, CSD SOXZAJ; Figure: PLATON

Crystal: Regular Packing in three Dimensions



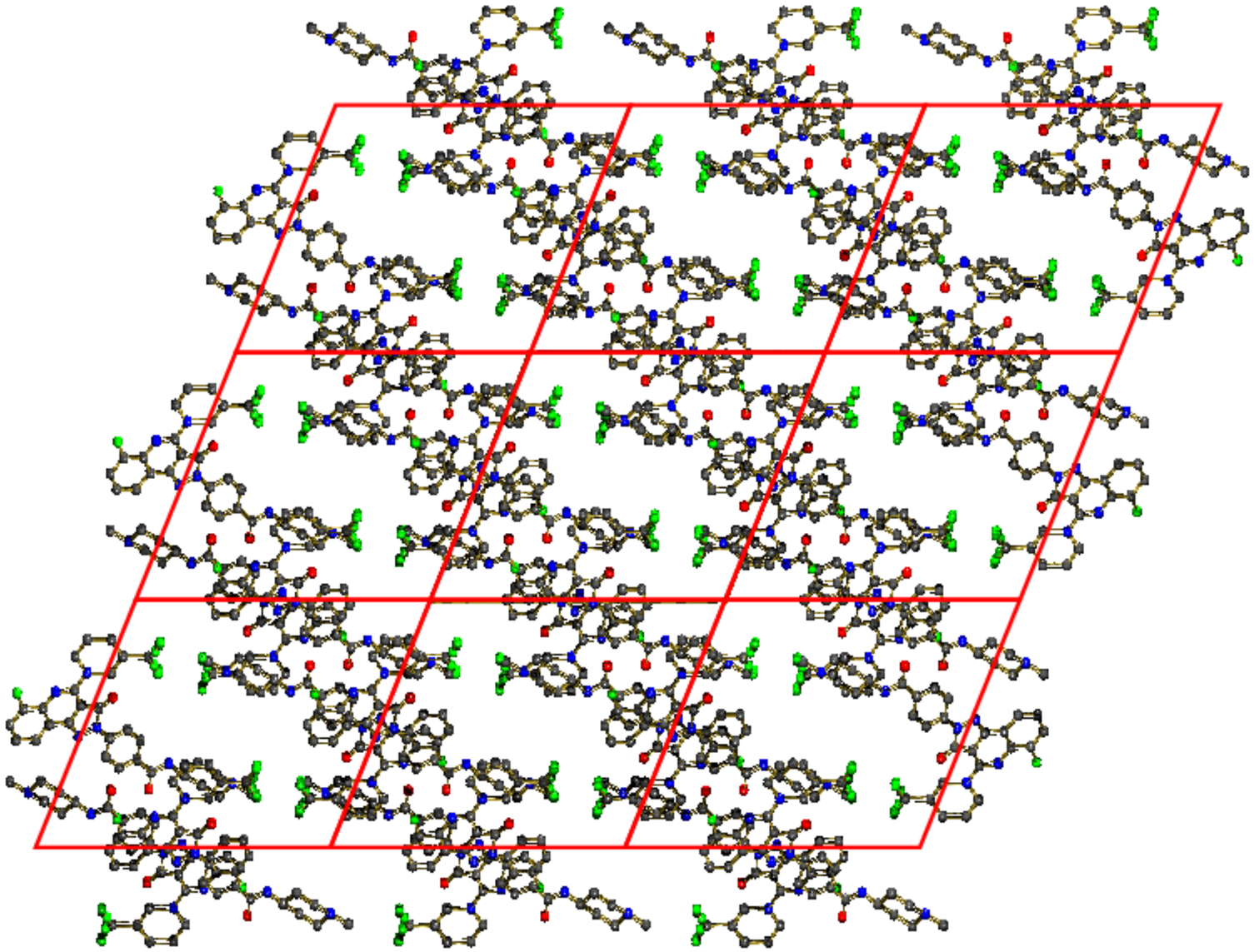
M. Taddio *et al.*, DOI 10.1021/acs.jmedchem.9b00858, CSD SOXZAJ; Figure: PLATON

Crystal: Regular Packing in three Dimensions



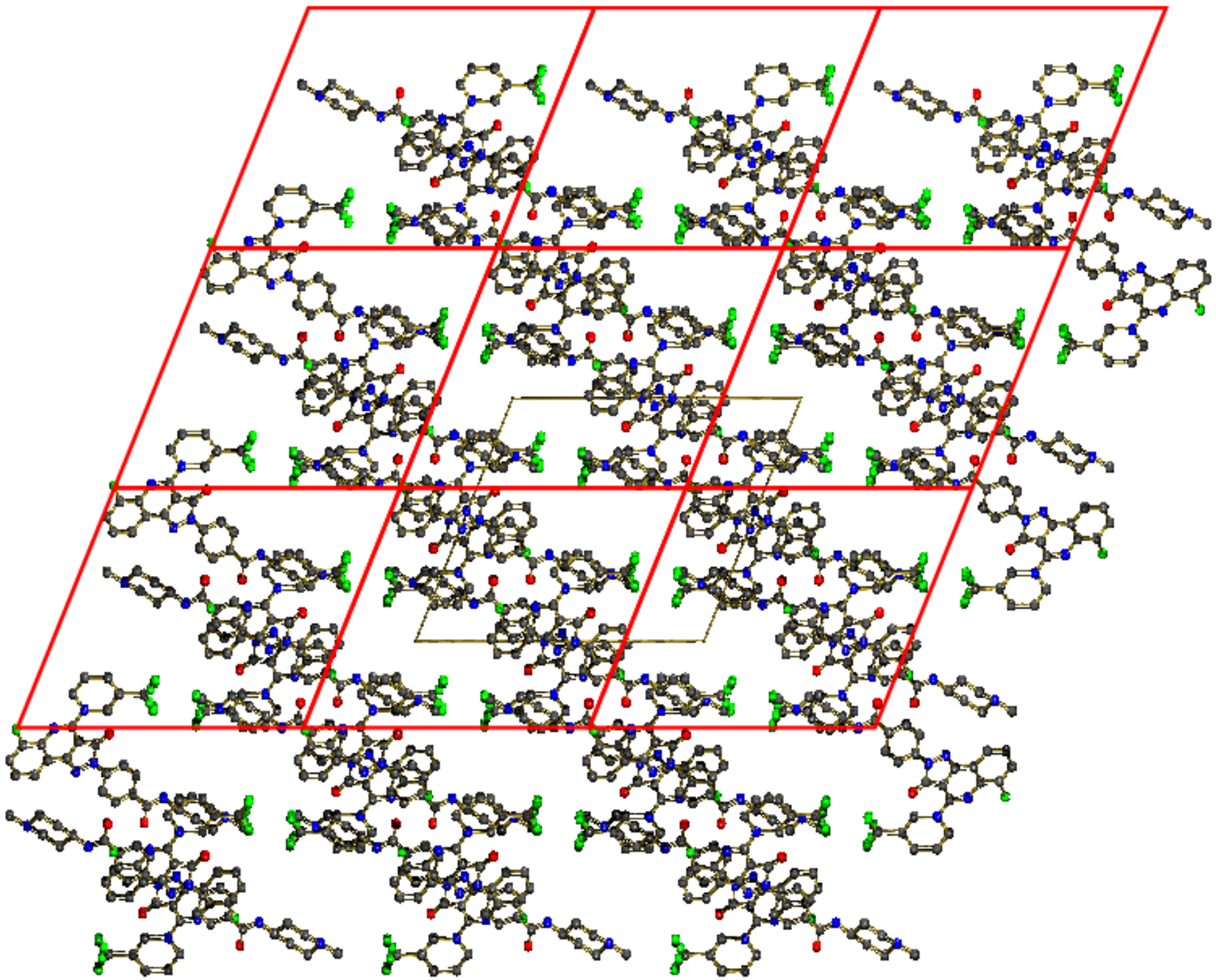
M. Taddio *et al.*, DOI 10.1021/acs.jmedchem.9b00858, CSD SOXZAJ; Figure: PLATON

Regularity: Repetition of the Unit Cell



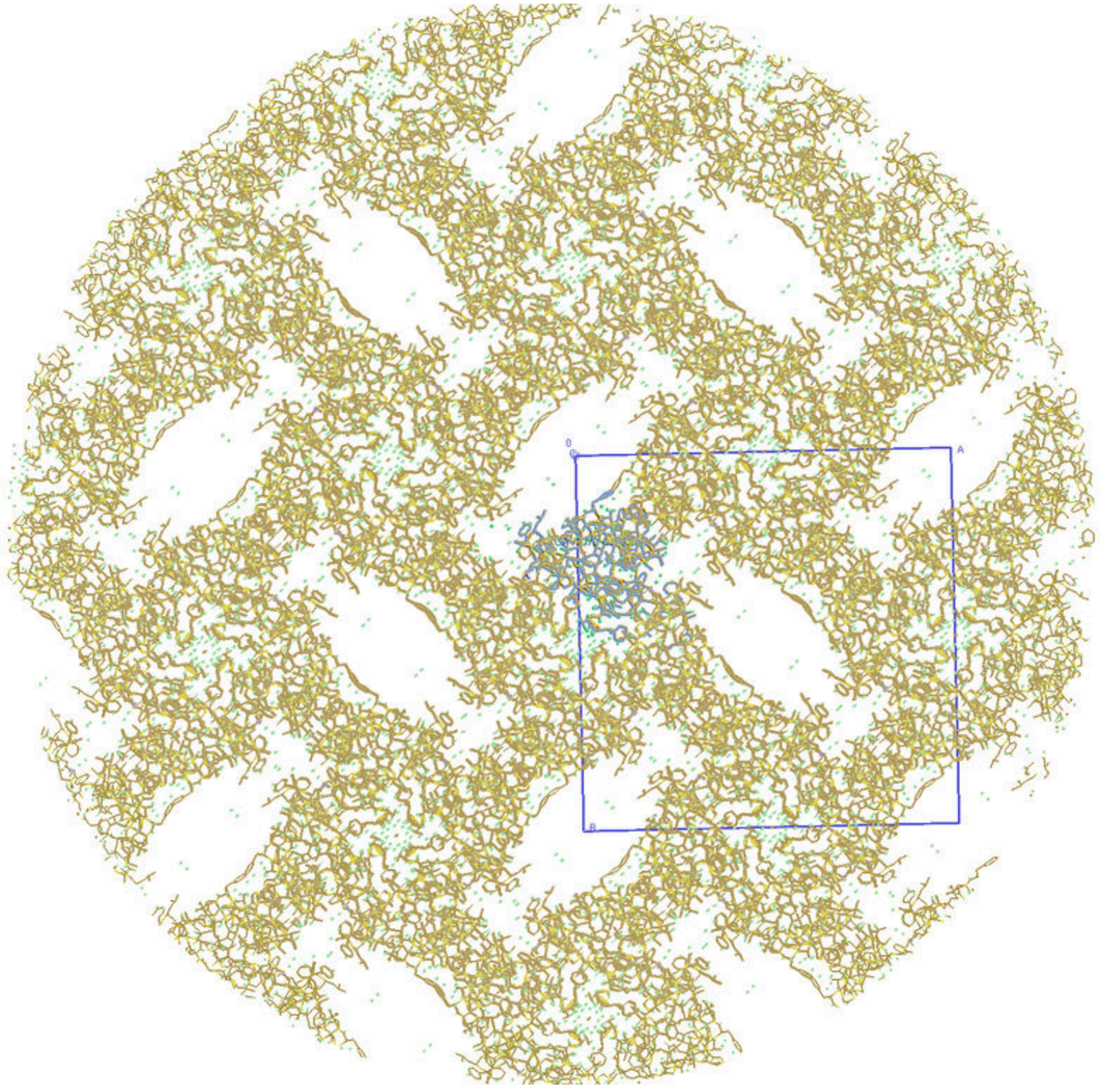
M. Taddio *et al.*, DOI 10.1021/acs.jmedchem.9b00858, CSD SOXZAJ; Figure: PLATON

Regularity: Repetition of the Unit Cell



M. Taddio *et al.*, DOI 10.1021/acs.jmedchem.9b00858, CSD SOXZAJ; Figure: PLATON

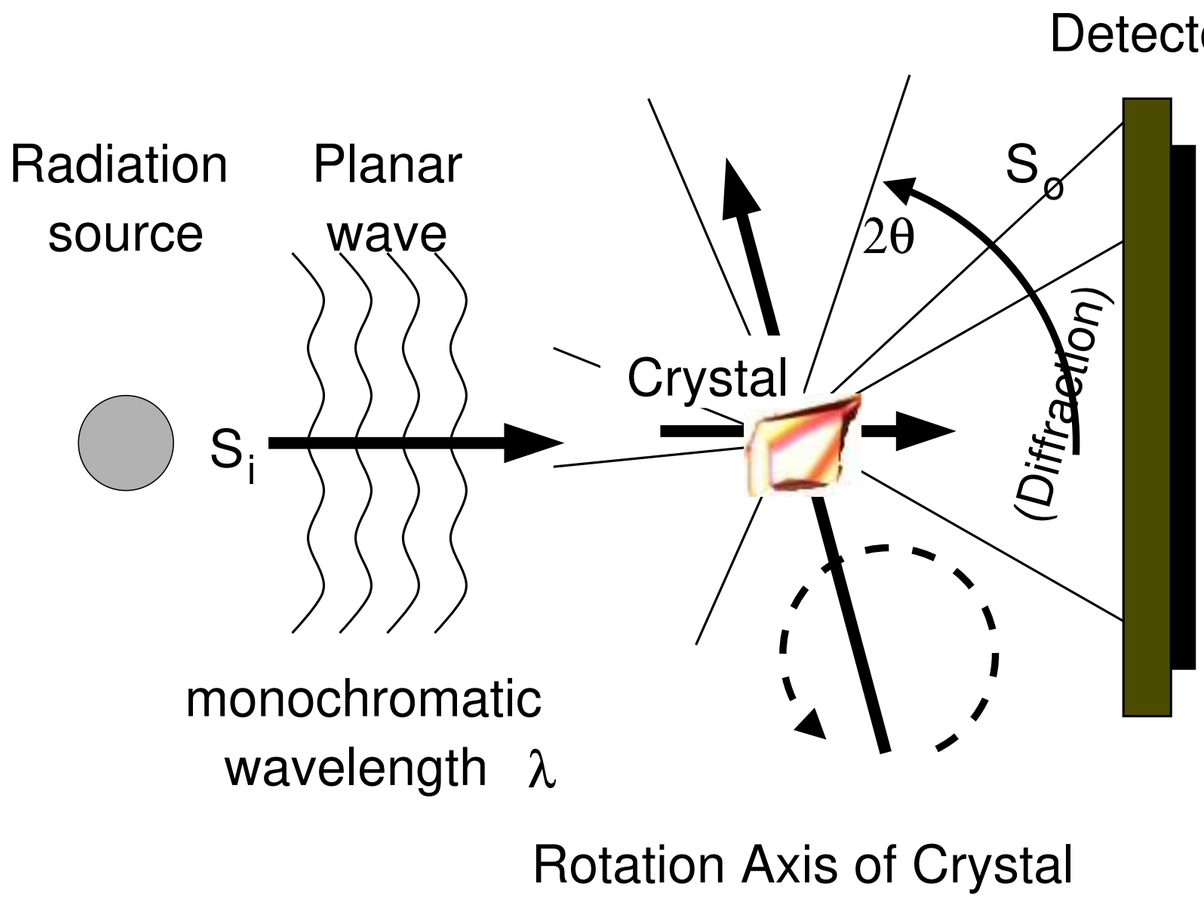
Protein Crystals with Solvent Channels

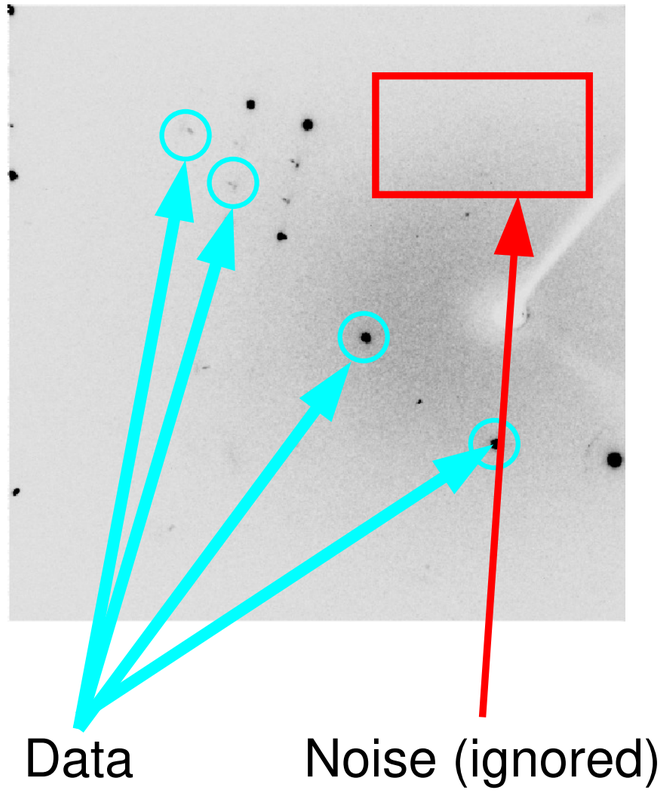


PDB ID 3LDC, image prepared with COOT

Crystals and (X-ray) Diffraction

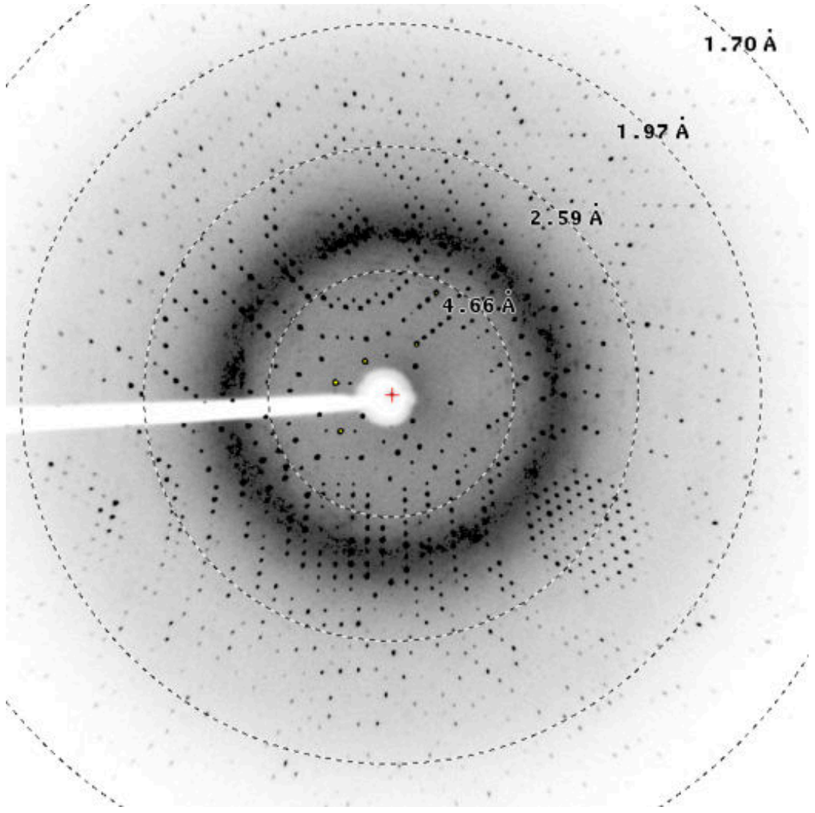
The Diffraction Experiment



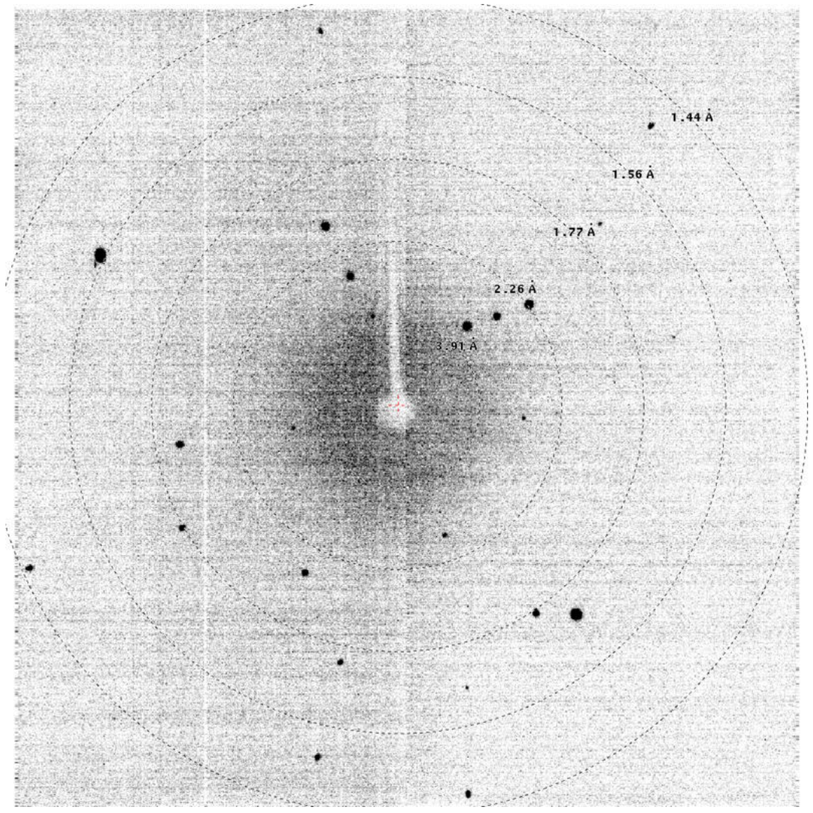


- Diffraction spots: interaction between **wave** and **crystal**
- Experimental result: **Position** and **Intensity** for each spot
- Data recorded as **frames** with an area detector
- Scattering vector $\vec{S} = \vec{S}_o - \vec{S}_i$ describes every spot.

The Reflection Spots



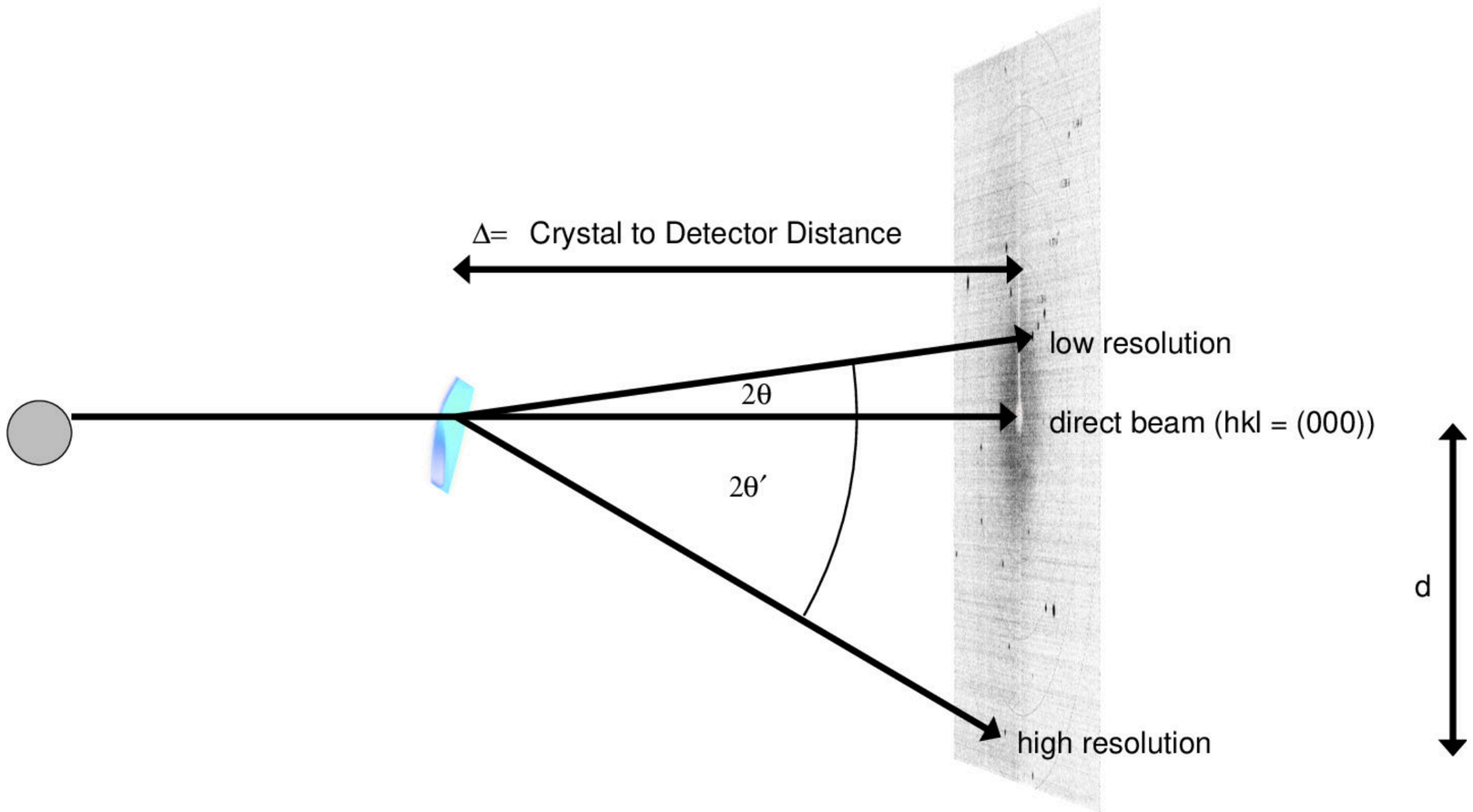
Macromolecule



Small Molecule

- Macromolecule: large unit cell = many spots with small separation. typical unit cell lengths 50Å to a few hundred Å
- Small molecule: small unit cell (20'ish Å) = few strong spots, large sepration
- Spot intensity decreases from central beam outwards

The Resolution $d(hkl)$ of a Reflection Spot



- 2θ = Angle between direct beam and spot on detector
- $\tan(2\theta) = d/\Delta$
- Spot resolution $d(hkl)$ by Bragg equation: $\lambda = 2d\sin\theta$
- λ: Wavelength of radiation

The Resolution d of a Data Set

Data set statistics from XDS

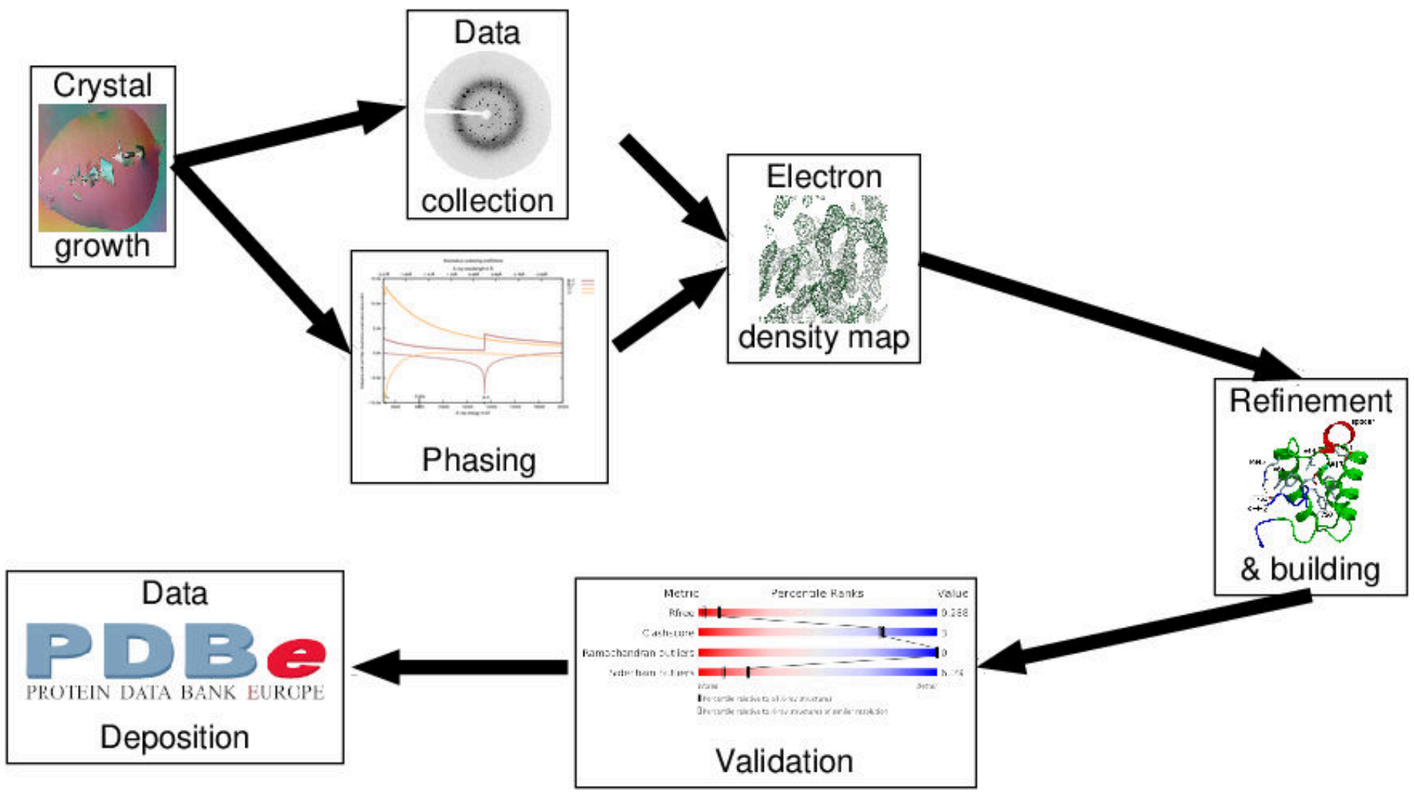
d	OBS	UNIQ	COMPLET.	I/SIG	Rmeas	CC1/2
4.57	568	98	96.1%	35.70	5.4%	99.9*
2.04	6182	959	97.2%	21.86	8.5%	99.8*
1.52	4951	1454	95.2%	8.88	11.6%	98.8*
1.27	5458	1707	93.5%	5.73	21.4%	94.7*
1.11	6005	2022	92.4%	4.24	30.8%	88.5*
0.99	4864	1658	61.9%	2.69	55.2%	70.8*
total	28028	7898	84.9%	7.62	11.0%	99.8*

- Resolution refers to a full data set
- Where $I/\sigma(I)$ falls below 2: $\approx$ optical resolution
- Refinement programs downweight unreliable reflections: Do include more: $I/\sigma(I) \approx 1.0$, or where CC1/2 is NOT marked (*) anymore
- D include more for refinement

Karplus & Diederichs, Science (2012), 336, 1030–1033

From Crystal to Structure

The Crystallography Flowchart



1. Crystallisation
2. Data Collection
3. Phasing
4. Model Building & Refinement
5. Interpretation & Validation
6. Deposition & Publication

The use of structural data requires (basic) understanding of the (crystallographic) workflow from sample (here: crystal) *via* data collection to model muilding & refinement

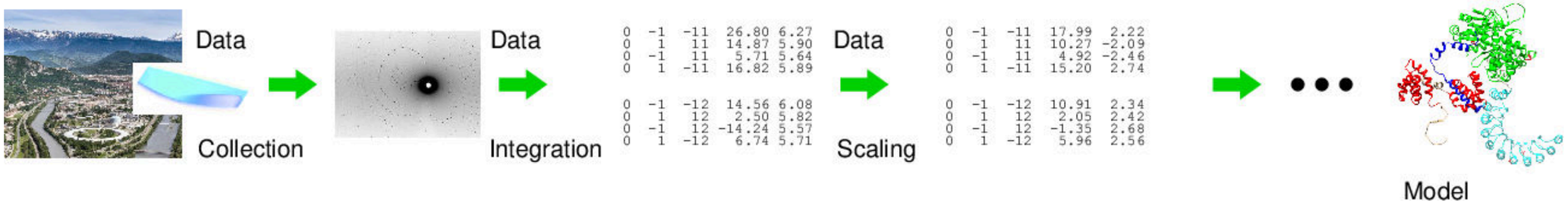
Data Collection

Data collection is your last experiment

Good Data $\neq$ High Resolution Data

- Always good: complete data
- Experimental phasing: accurate data $\rightarrow$ data redundancy
- Molecular replacement: accurate at low(!) resolution: avoid overloads, caveat beam stop
- Refinement: high resolution data

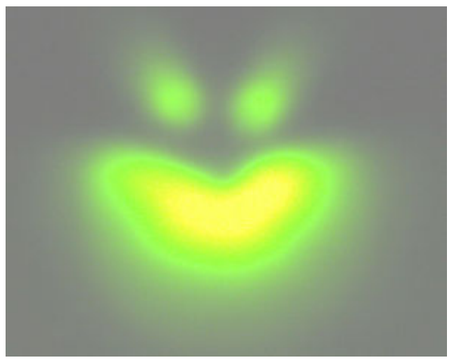
Data Reduction



- Purpose of crystal structure determination: Molecular Model
- Refinement of model against experimental data $h,k,l,F_{\text{exp}}(hkl),\sigma_F$ or $h,k,l,I_{\text{exp}}(hkl),\sigma_I$
- Experimental date should be independent of machine, of wavelength, of crystal shape and size

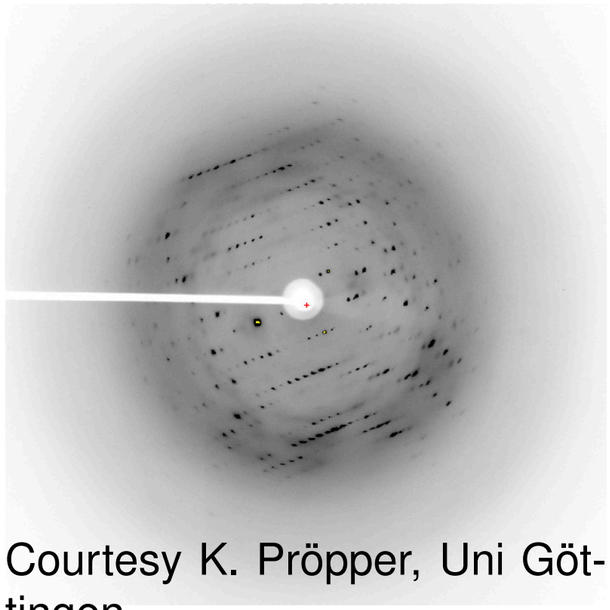
Experimental Errors

Unfocused Beam



Courtesy N. Sanshvili, & S. Corcoran, APS Chicago

Poorly diffracting crystals



Courtesy K. Pröpper, Uni Göttingen

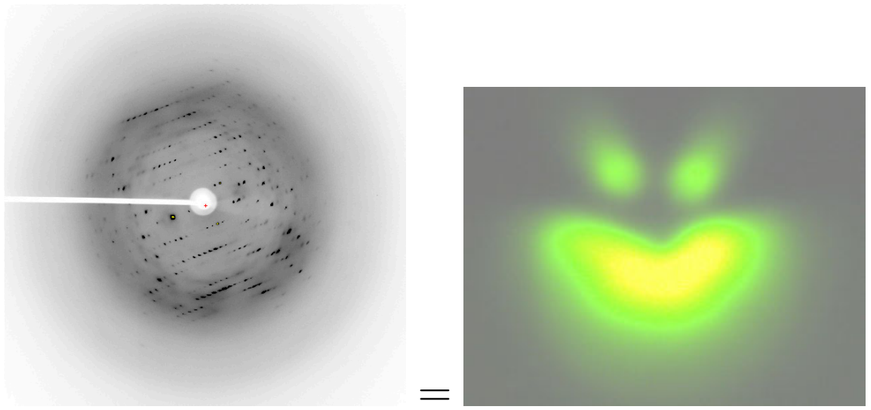
Sources of Errors

- Crystal imperfections
- Radiation damage
- Overloads
- Detector Background Noise
- Missettings (Note down wavelength and distance !!!)

Data Scaling

Intensities $I_{\text{obs}}(hkl)$ are *observed* quantities on the detector. Scaling corrects for various systematic and random errors to yield the experimental intensities $I_{\text{exp}}(hkl)$. $I_{\text{exp}}(hkl)$ are used for all subsequent steps (phasing, refinement, ...)

$$I_{\text{obs}}(hkl) = \frac{e^4}{{m_e}^2 c^4} \frac{V_{\text{crystal}}}{V_{\text{u.c.}}^2} \underbrace{\lambda^3 I_0 L P T E}_{\text{experimental factors}} I_{\text{exp}}(hkl)$$



Data Processing:

Data integration: Determination of intensities $I_{\text{obs}}(hkl)$ from frames

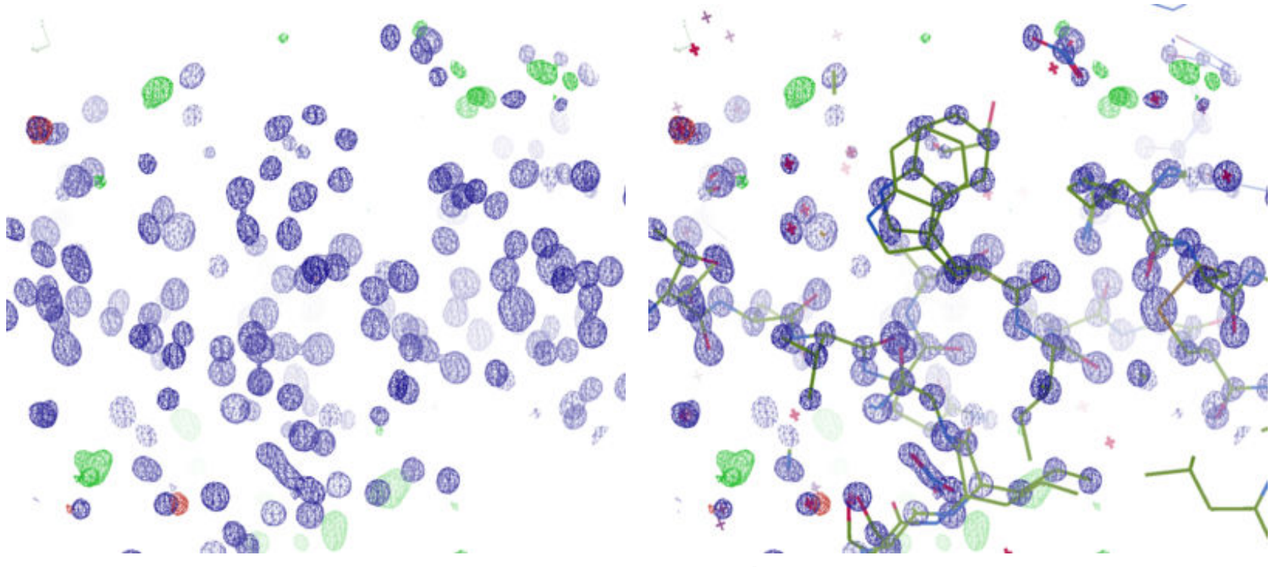
Scaling / Merging: Determination of amplitudes $I_{\text{exp}}(hkl)$ from $I_{\text{obs}}(hkl)$ and experimental settings

Giacovazzo et el., "*Fundamentals of Crystallography*" (IUCr Texts on Crystallography), 1985

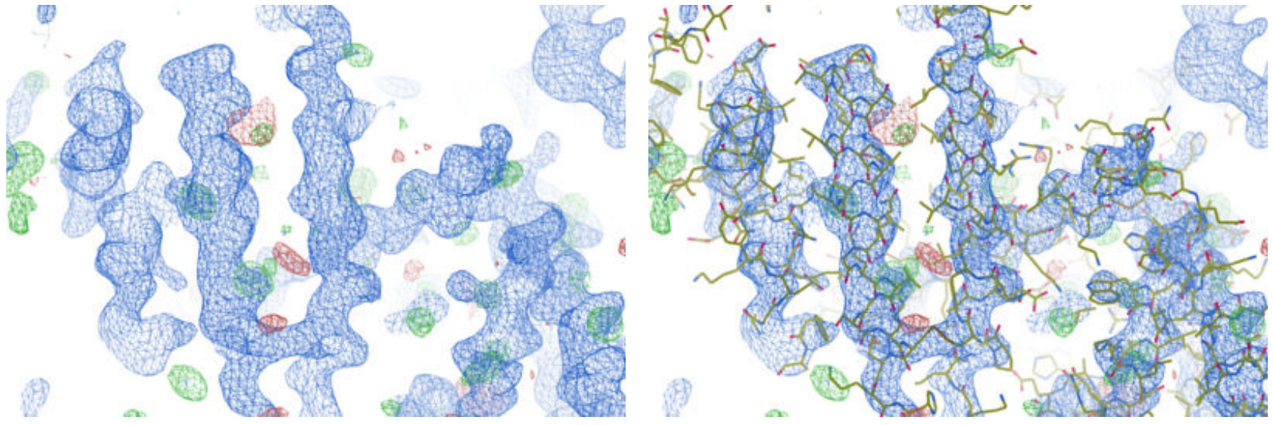
Model, Map, and Data

Giacovazzo et el., "*Fundamentals of Crystallography*" (IUCr Texts on Crystallography), 1985

Model = Interpretation of Map



Lysozyme, 2VB1, 0.65Å resolution



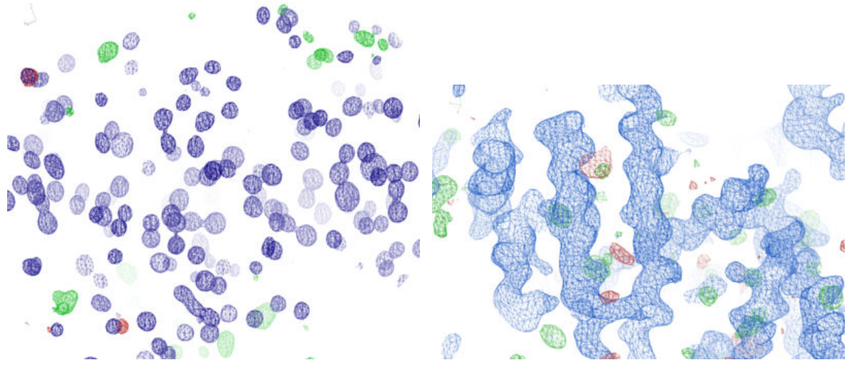
Human Serum Albumin, 6JE7, 3.9Å resolution

The structural model is created by fitting atoms into the experimental map.

Maps with COOT, and PDB2INS, A. Luebben & G. M. Sheldrick, DOI 10.1107/S1600576719005478

Map and Data

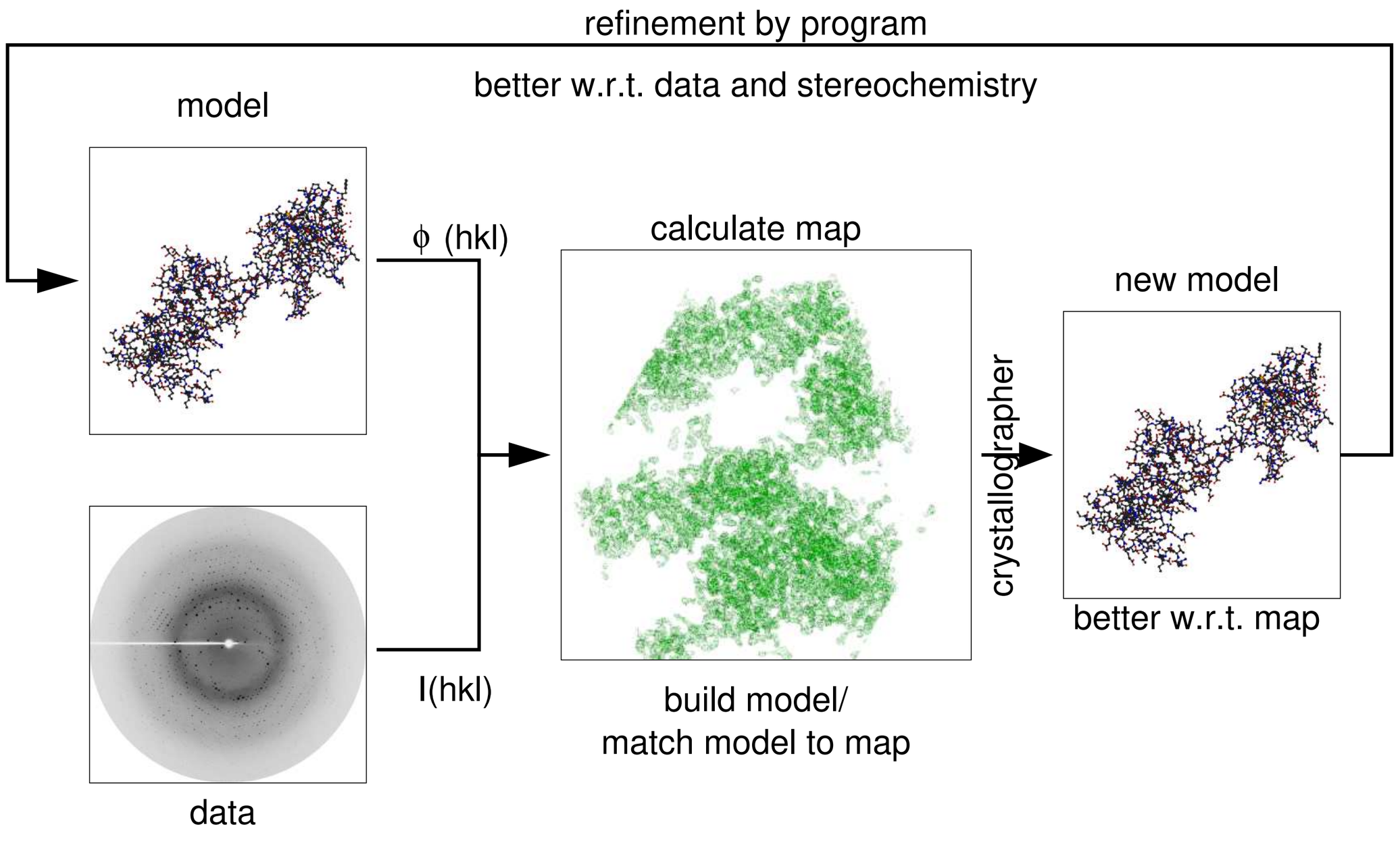
COOT displays a **map** $\rho(x,y,z)$ calculated as



$$\rho(x,y,z) = \sum_{h,k,l} \sqrt{I_{\text{exp}}(hkl)} e^{i\phi(hkl)} e^{-2\pi i(hx+ky+lz)}$$
$$= \sum_{h,k,l} |F_{\text{exp}}(hkl)| e^{i\phi(hkl)} e^{-2\pi i(hx+ky+lz)}$$

- $|F_{\text{exp}}(hkl)| = \sqrt{I_{\text{exp}}(hkl)}$ Structure factor amplitudes
- $\phi(hkl)$ Phase of reflection (hkl)
- $|F_{\text{exp}}(hkl)|e^{i\phi(hkl)}$: structure factor (complex number)
- x,y,z coordinates for points within unit cell

Model Building & Refinement



The Refinement Cycle: Words of Caution

- Only the intensities $I_{\text{exp}}(hkl)$ are **experimental data**.
- The initial map comes from the estimates from phasing.
- The improvement of the model means an improvement of the phases.
- One requires a model to produce a map and a map to improve the model.

The dependency model$\rightarrow$map$\rightarrow$model$\rightarrow$map . . . is error prone. It is important to understand the risks in order to produce a good model..

Validation: how good is a model

- Several validation tools available
  - PDB chart
  - Molprobity

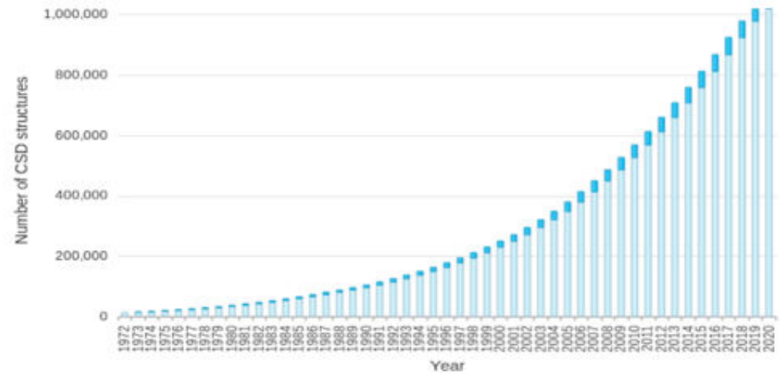
COOT: Take your own look at model *and* the map *and* the difference map

Crystallography with Neutron Radiation

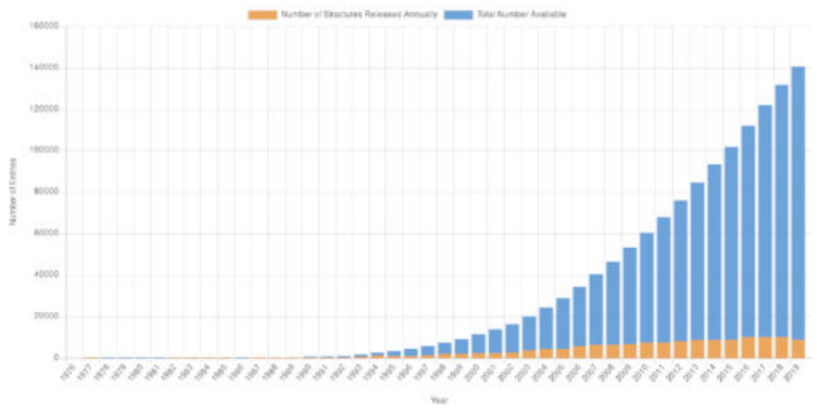
Dominating X-ray Crystallography

- most advanced (pipelines from data collection to structure refinement)
- typical wavelength: λ =0.8–1.9Å
- standard structure determination

	PDB (MX)	CSD (SX)
X-ray	>140,000	>1,000,000
neutron	161	1,500
electron	111	10'ish

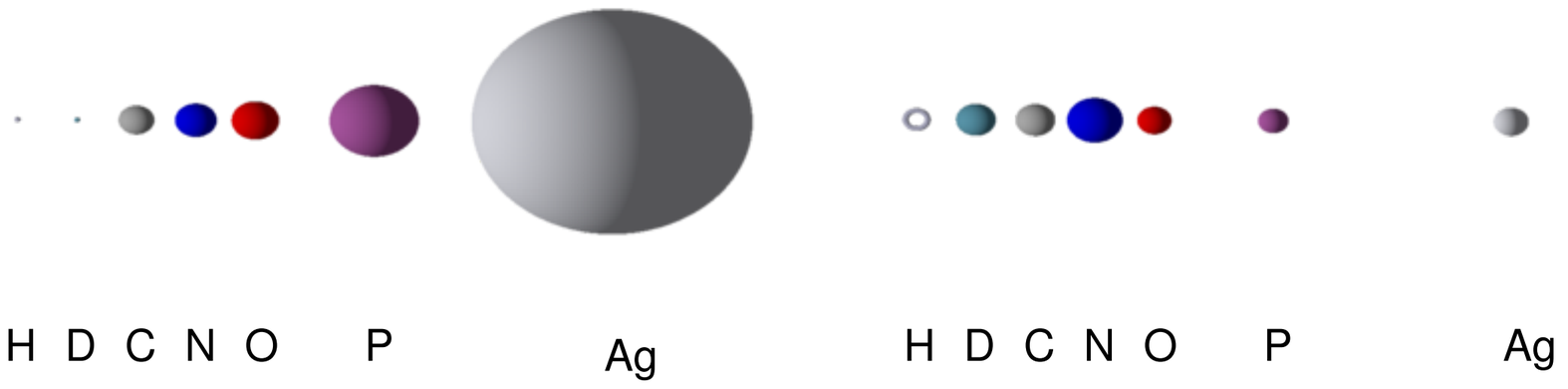


Annual Growth of the CSD



Annual Growth of the PDB (X–ray)

Neutron Crystallography



- X-rays: scattering power increases with atom number Z

$\Rightarrow$ Hydrogen nearly invisible

$\Rightarrow$ Neighbouring (heavy) atoms indistinguishable

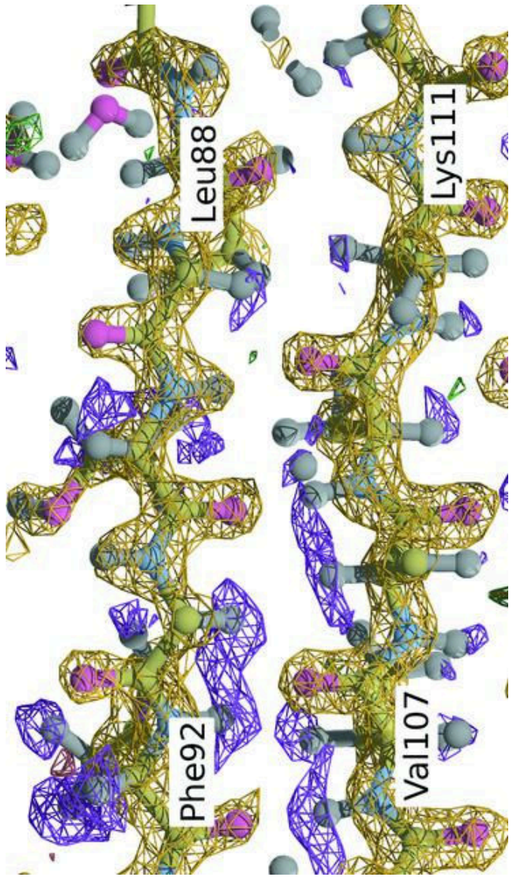
- neutons: scattering power varies irregularily

$\Rightarrow$ Hydrogen as strong as O, N, C

$\Rightarrow$ Some elements become distinguishable (semi-conductors, transition elements)

Applications of Neutron Crystallography

1. visualisation of hydrogen atoms
2. adjacent elements (e.g. K^+ vs. Cl^-, Zn^{2+} vs. Cu^+)
3. requires large crystals ($\geq 1\text{mm}^3$)
4. (virtually) no radiation damage
5. structure determination from radiation sensitive samples (Photosystem II)



PDB ID 2ZOI: D/H exchange in β–strand

Crystallography with Electron Radiation

3D Electron Crystallography (3D ED)

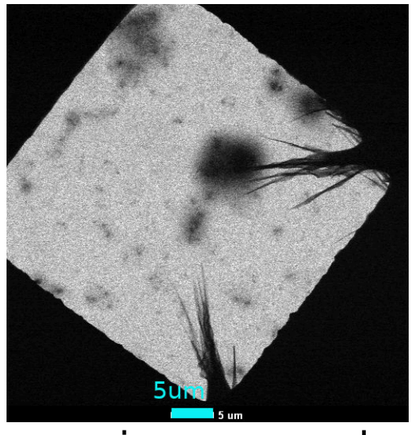
- Electrons interact with electrostatic potential
- Electrons interact **much,much** stronger with matter than X-rays

$\Rightarrow$ Much smaller crystals

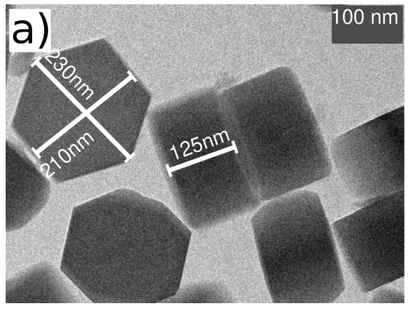
$\Rightarrow$ problematic: dynamic diffraction, $|F| \neq \sqrt{I}$

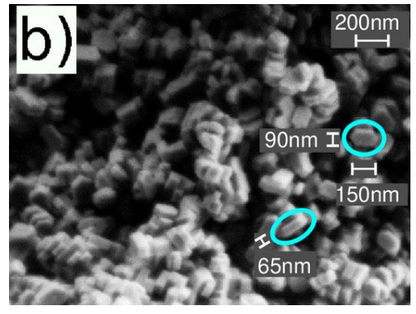
- Electron optics enable some special applications and tiny beam (5 nm diameter)

3D ED: small crystals

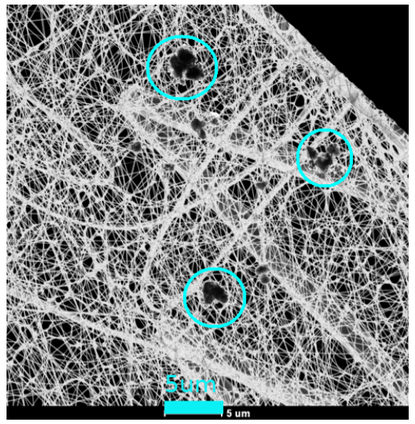


organic compound

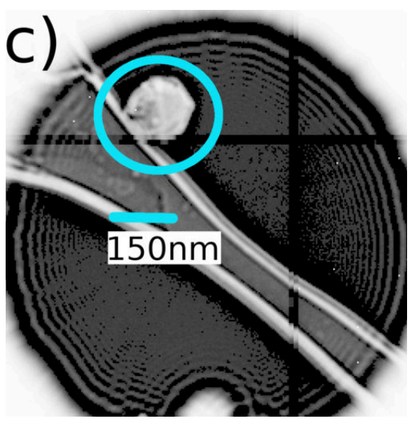


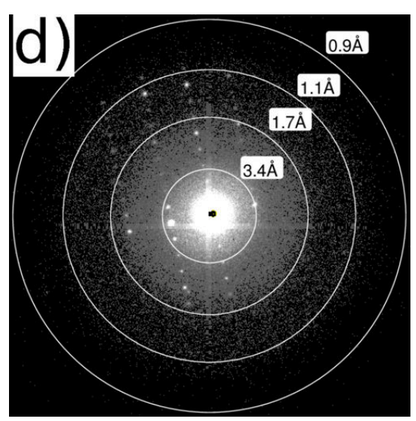


Silicalite–1 / ZSM–5 (Teng Li)
Gruene *et al.*, Chem. EurJ (2018), 24, 2384–2388

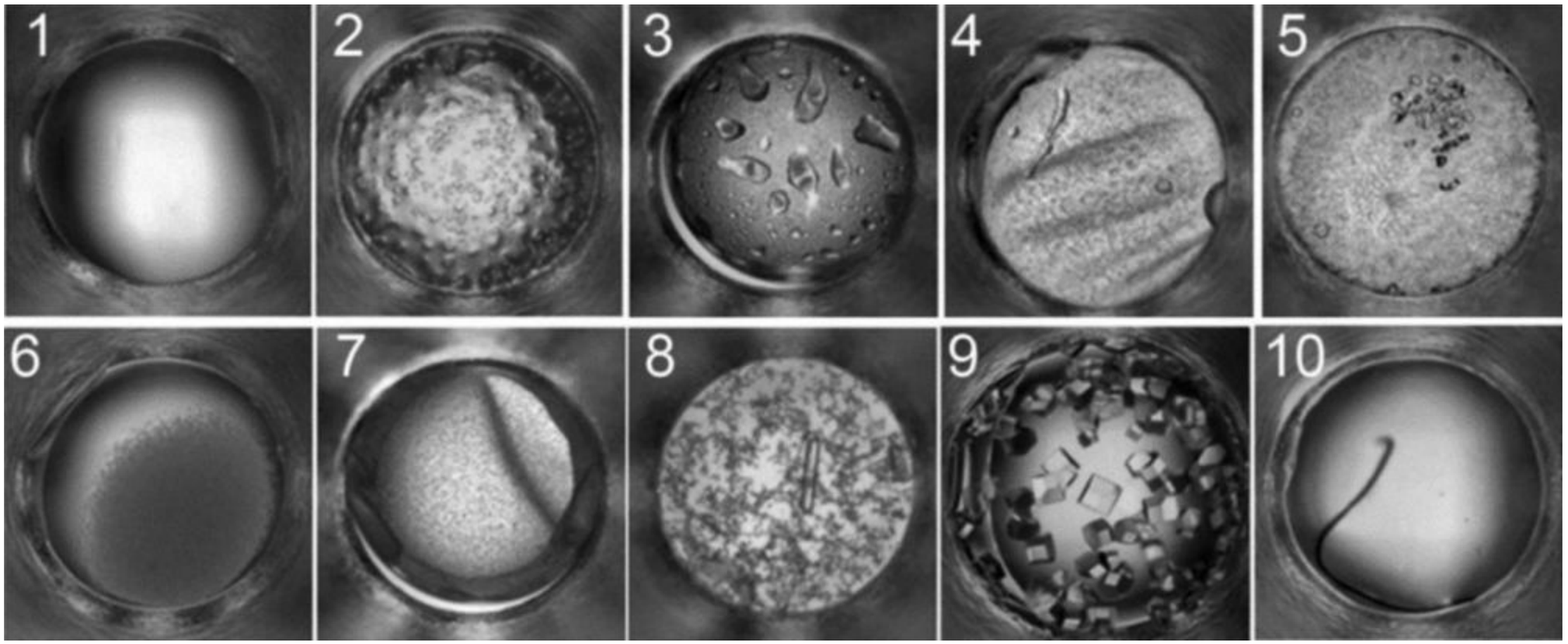


sucrose (ETH coffee bar)



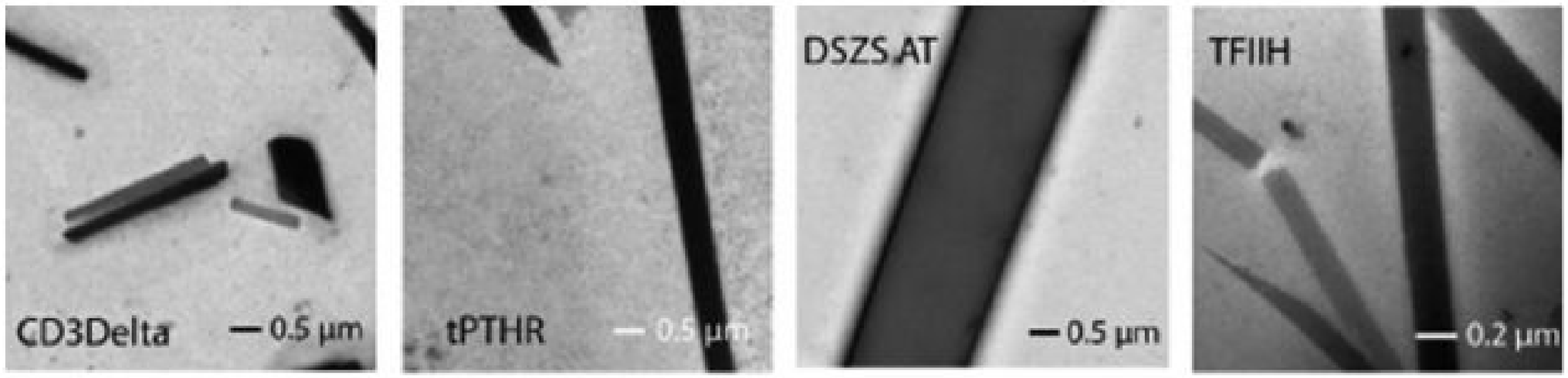


Protein Crystallisation



Luft, Wolfley, Snell, Crystal Growth& Design (2011), 11, 651–663

Drops viewed through TEM



Negative staining of crystallisation drops reveal submicrometer sized crystals in **about 1/3** of all crystallisation trials.

Stevenson,. . . , Calero, PNAS (2014) 111, 8470–8475 / Calero, . . . , Snell, Acta Cryst (2014) F70, 993–1008

3D ED for MX: Status quo

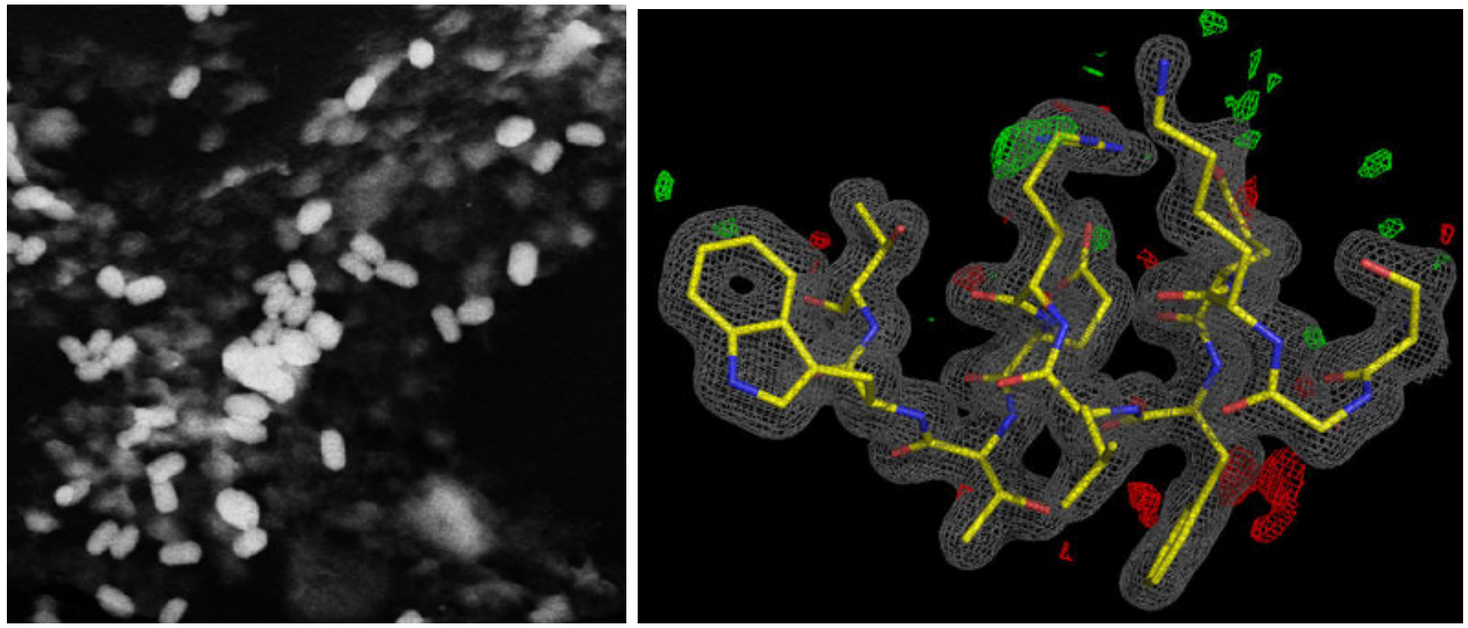
- Electron diffraction complements X-ray crystallography for tiny crystals
- Chemical compounds, when powder available: useable technology
- MX: Sample handling difficult
- MX: Resolution does not match X-ray experience

	e^-		X–ray		resol. ratio
	d_{min}	PDB-ID	d_{min}	PDB-ID	$d(e^-)/d(\text{X-rays})$
Lysozyme	1.80	5K7O	0.94	1IEE	1.9
Lysozyme	2.80	6HU5	0.94	1IEE	1.9
Catalase	3.20	5GKN	1.50	1DGF	2.1
Proteinase K	1.75	5I9S	0.83	2PWA	2.1
Xylanase	2.30	5K7P	0.97	3AKQ	2.3
Thaumatin	2.11	5K7Q	0.90	5X9L	2.3
Trypsin	1.70	5K7R	0.75	4I8H	2.2
Thermolysin	2.50	5K7T	1.12	5JVI	2.2

Promising: Serial ED

- I. Schlichting (early 2017), K. Diederichs (independently, autumn 2017): Combine serial crystallography with rotation method

Closest so far: Bückner, . . . , D. Miller, “Serial protein crystallography in an electron microscope”, ECM 2019 (*cf.* https://doi.org/10.1101/682575) granulovirus, $I23$, 103Å, 1.55Å



See also: Wang, Zou, and Smeets, “Automated serial rotation electron diffraction combined with cluster analysis: an efficient multi-crystal workflow for structure determination”, IUCrJ (2019), 6, 854–867

Resources for Crystallography — Further Reading

- `ccp4bb` CCP4 Mailing list, `CCP4BB-subscribe-request@JISCMAIL.AC.UK`
- B. Rupp (ed.), *Biomolecular Crystallography*, Garland Science, 2009
- J. Drenth, *Principles of Protein X-ray Crystallography*,Springer, 2007