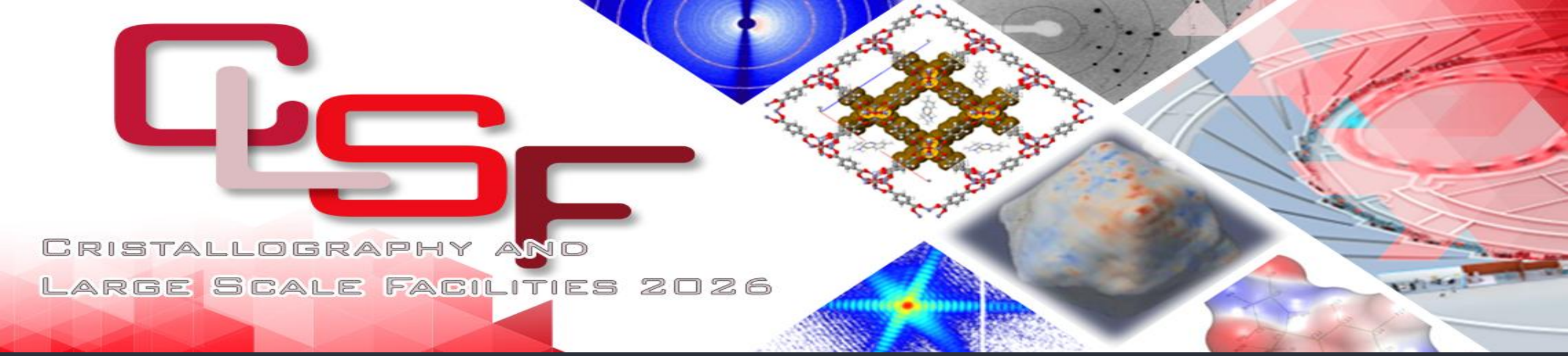


Powder Diffraction Data Analysis

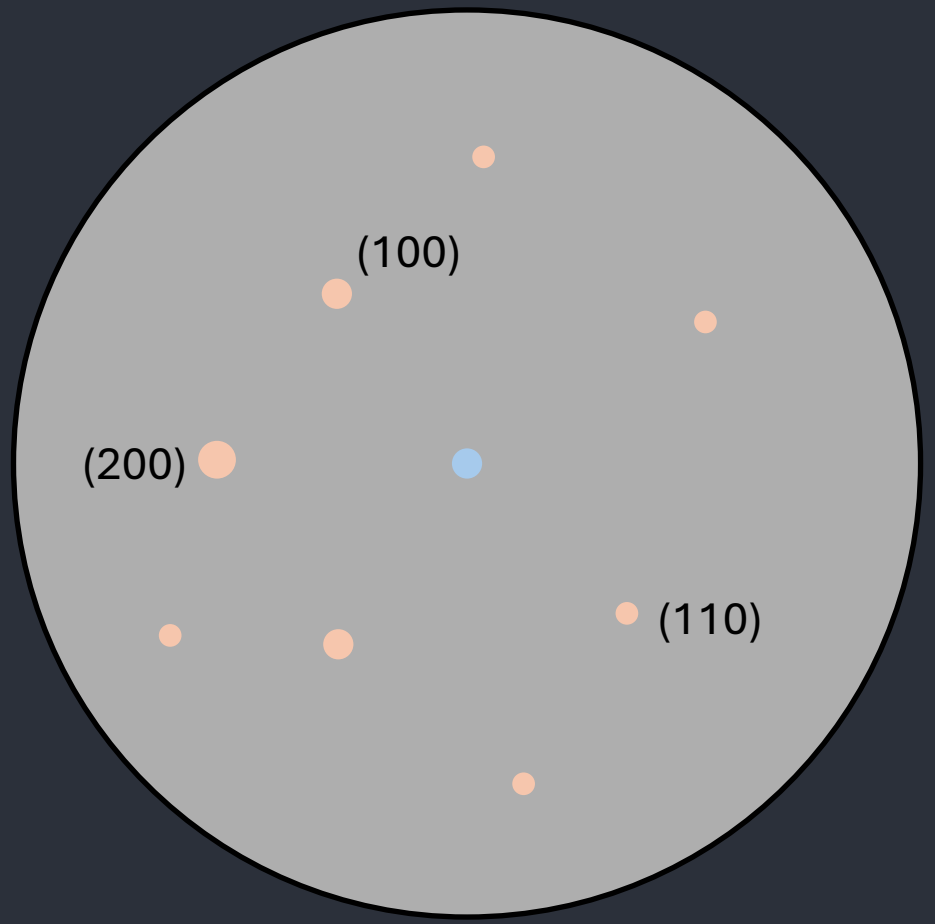
Hubert Chevreau - Synchrotron SOLEIL

hubert.chevreau@synchrotron-soleil.fr

OUTLINE

- Single Crystal *vs.* Powder
- Data Collection & Reduction
- Phase Identification
- *ab initio* Structure Determination (indexation, space group, intensity extraction, Rietveld refinement, final check)
- Fourier Map, Quantitative Phase Analysis, Microstructure, Crystallite Size
- Computational and data-driven Crystallography (DFT & AI)

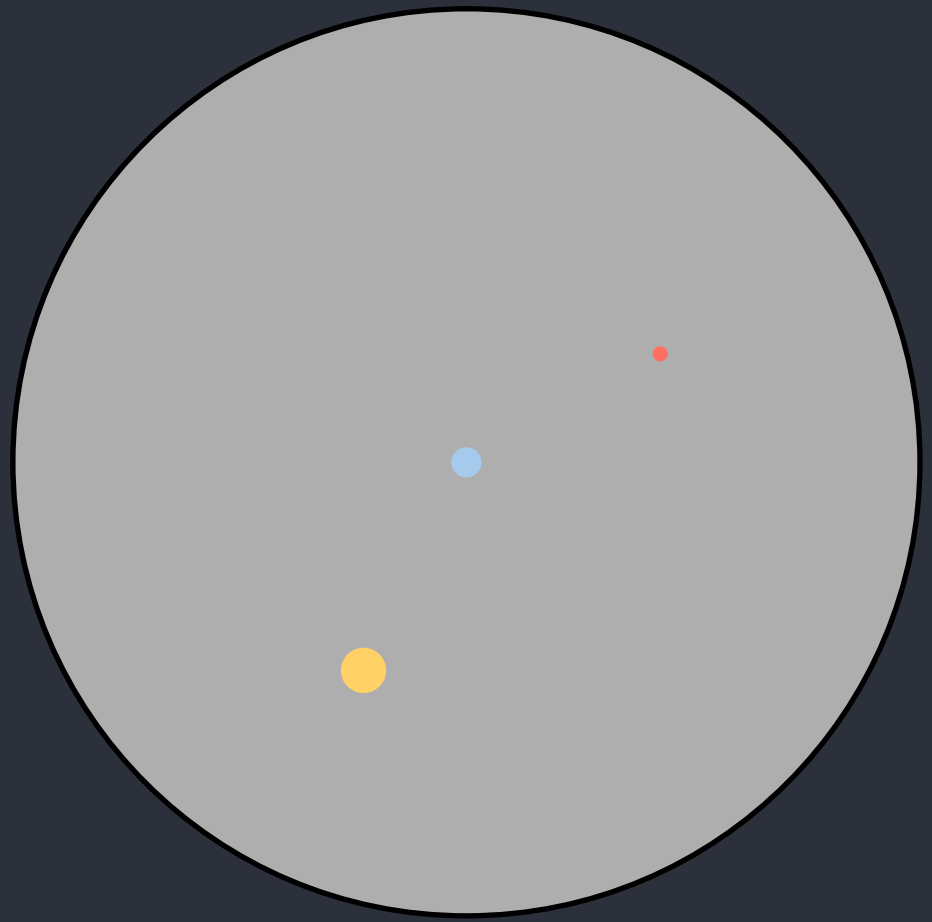
Single Crystal *vs.* Powder



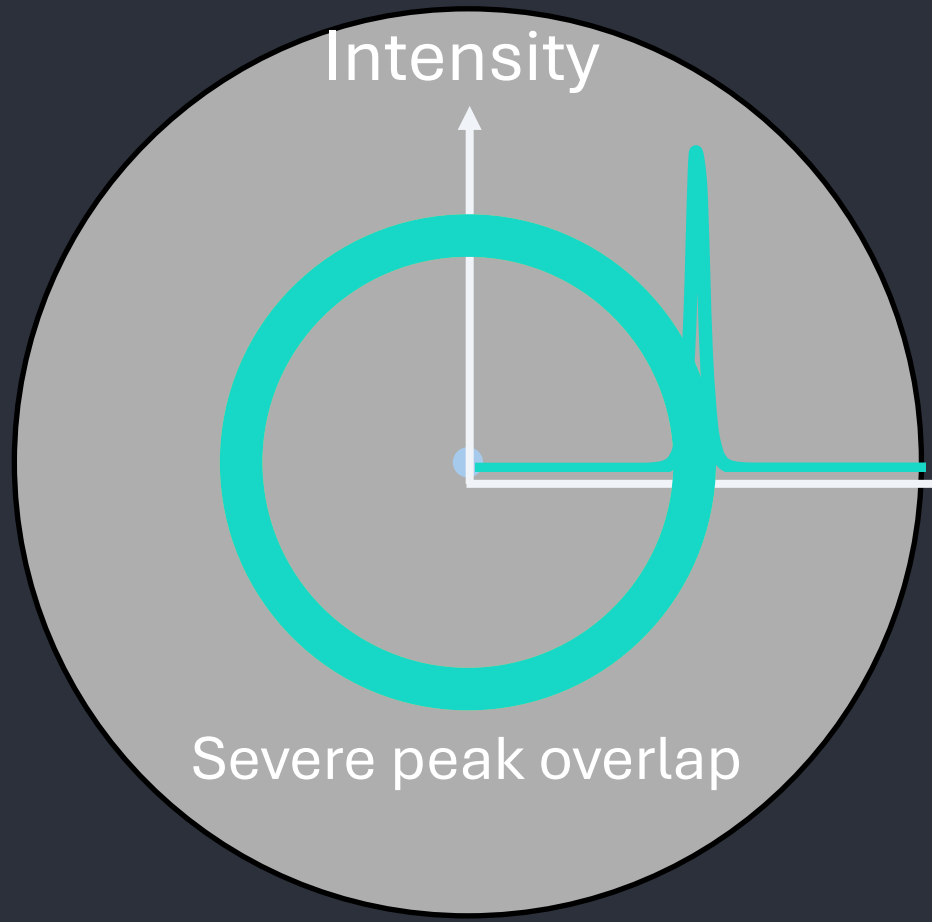
Debye-Scherrer Rings



Why powder is more complicated than single-crystal?



3D reciprocal space to 1D pattern



- Preferred orientation such as needle-like crystals
- Peak broadening in powders is not just structural; it is a convolution of instrumental optics, crystallite size, and microstrain.

Powder Diffraction Powder Content

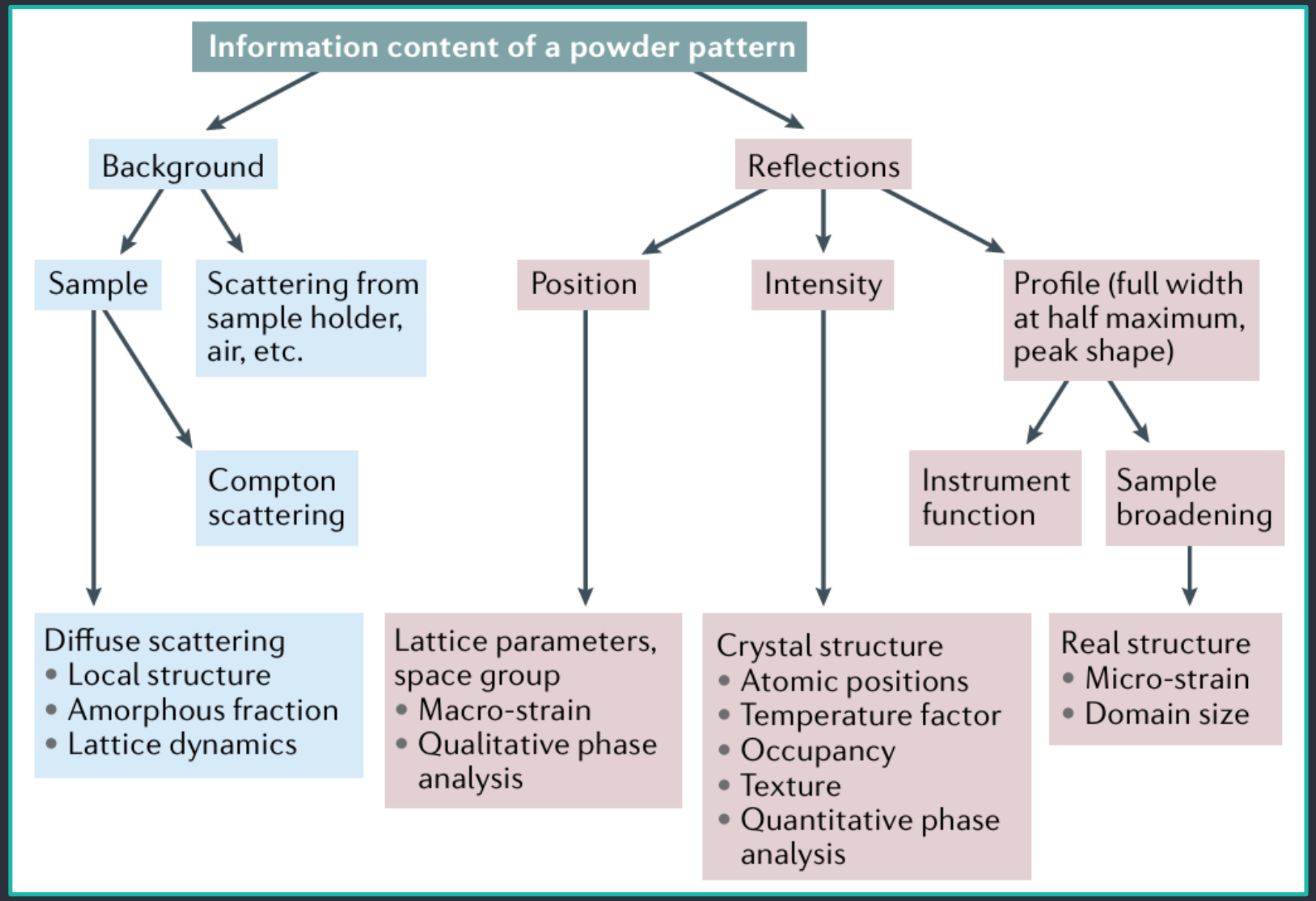


Figure taken from: Dinnebier, R. E. & Billinge, S. J. L. Powder Diffraction: Theory and Practice (RSC, 2008).

Method Overview

Data Collection

Beam type, wavelength, instrument, time, sample environment

Data Reduction

2D-1D conversion, data merge, background, detector mask

Data Analysis

Qualitative Analysis

Structure determination

Rietveld

Peak profile

QPA

Microstructure

Data Visualization

Data Report & Publication

Data Collection... in the Lab





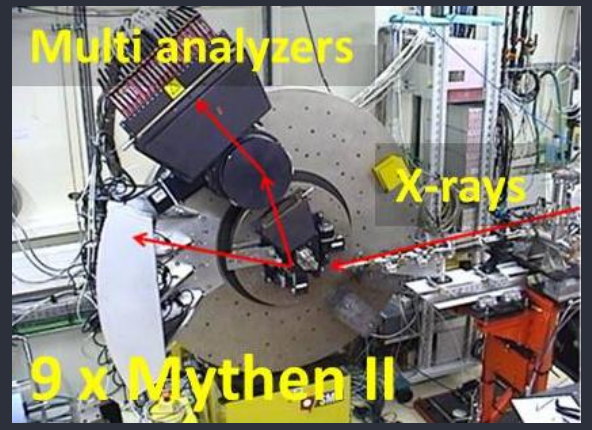


BRUKER

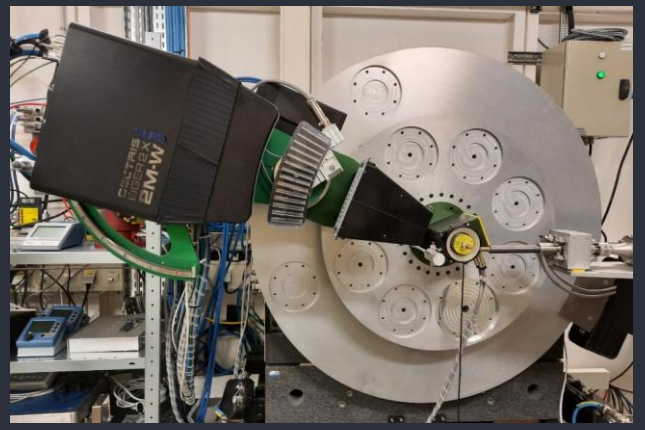
PANalytical
The Analytical X-ray Company

Rigaku

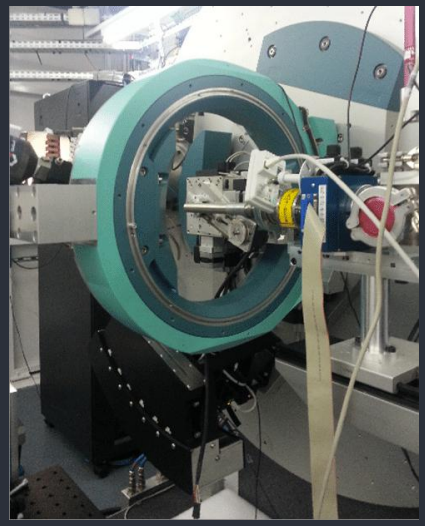
Data Collection… at Synchrotron



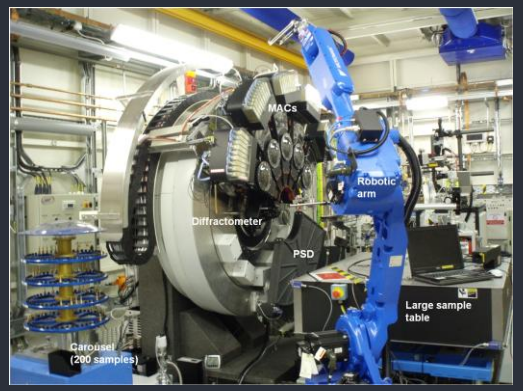
CRISTAL@SOLEIL



ID22@ESRF



MSPD@ALBA



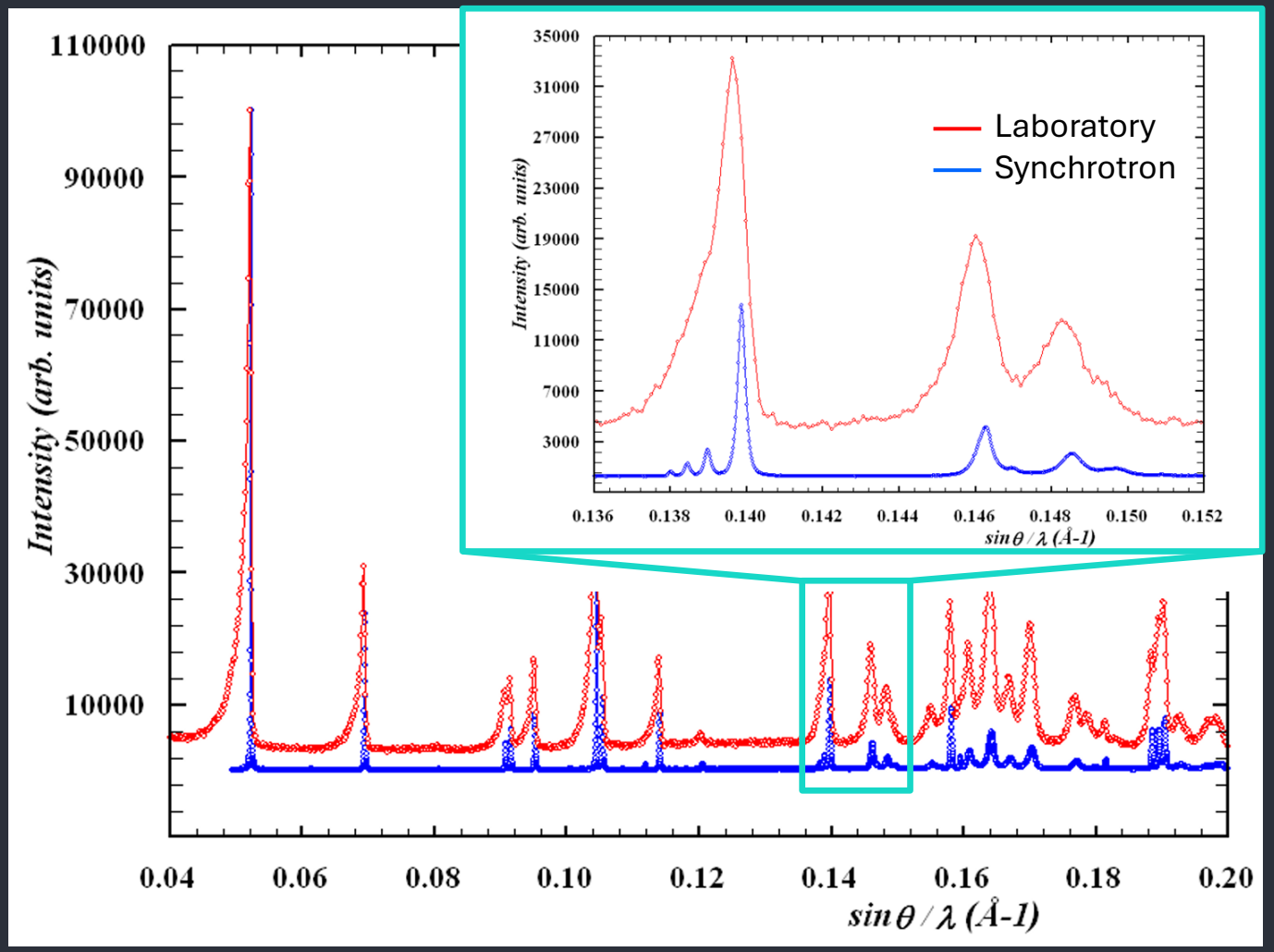
I11@DIAMOND



11BM@APS

1D@SLS
P02.1@DESY
BL02B2@SPring-8
BL10@Australian-Synchrotron
28-ID-2@NSLS-II
EMA@SIRIUS
.....

X-ray Labs *vs.* Synchrotron



Advantages of synchrotron:

- High Angular Resolution allow separation of overlapped reflections
- Fast acquisition for in situ and in operando experiments (real-time reactions, phase transition, battery charging/discharging)
- Tunable wavelength for anomalous scattering (Fe vs. Co) & avoid fluorescence
- High energy provides high penetration power through complex sample environments

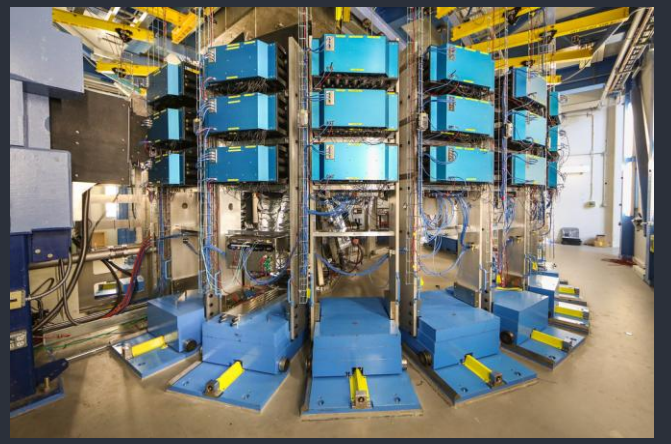
Data Collection… with Neutron Scattering



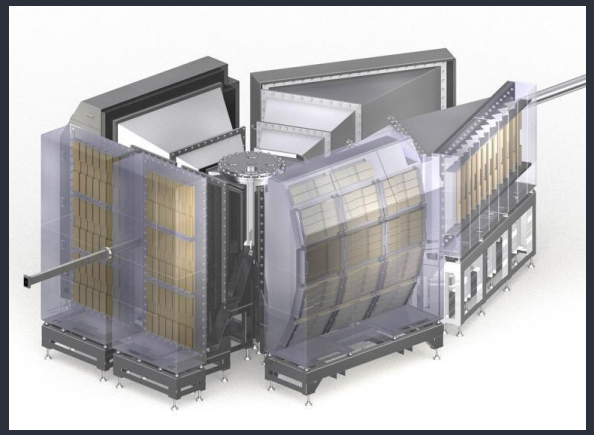
D2B, D1B, D20@ILL



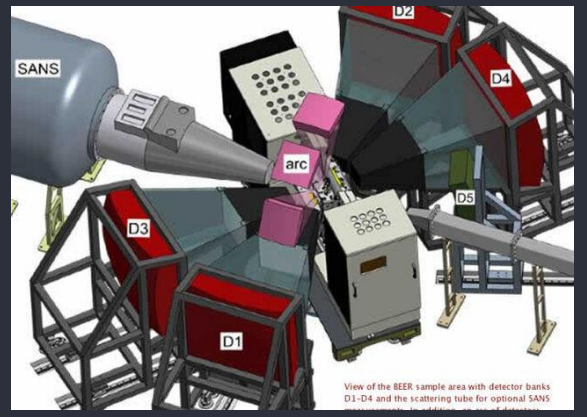
ECHIDNA, WOMBAT@ANSTO



POWGEN, HB-2A@ORNL



SuperHRPD, SPICA@J-PARC



DREAM,HEIMDAL@ESS

Data Reduction

➡ Gap between raw hardware signals (tiff/edf/raw files from detectors) and the clean 2θ vs. Intensity text files

- ***2D Integration & Calibration (Area Detectors):***
  - ***pyFAI:*** *Python-based library/GUI for fast azimuthal integration.*
  - ***Fit2D:*** *The historical pioneer for detector calibration and 2D-to-1D integration.*
  - ***Dioptas:*** *User-friendly GUI for interactive 2D data reduction and mask creation.*
  - ***FOXTROT:*** *Synchrotron-specific processing pipelines available at SOLEIL*
- ***Format Conversion & Manipulation:***
  - ***CMPR:*** *format conversion, plotting, and basic powder manipulations*
  - ***PDViPeR:*** *Data treatment, background stripping, smoothing, and peak picking*

➡ Ask your local contact

Identification: First thing First… Database



Inorganic Crystal Structure Database

CCDC

Cambridge Structural Database

COD

Crystallography Open Database

PDF-5+
ICDD PDF

AMCSD

American Mineralogist Crystal Structure Database

The Materials Project
materialsproject.org

Theorical database:
Materials Project, OQMS, Aflowlib

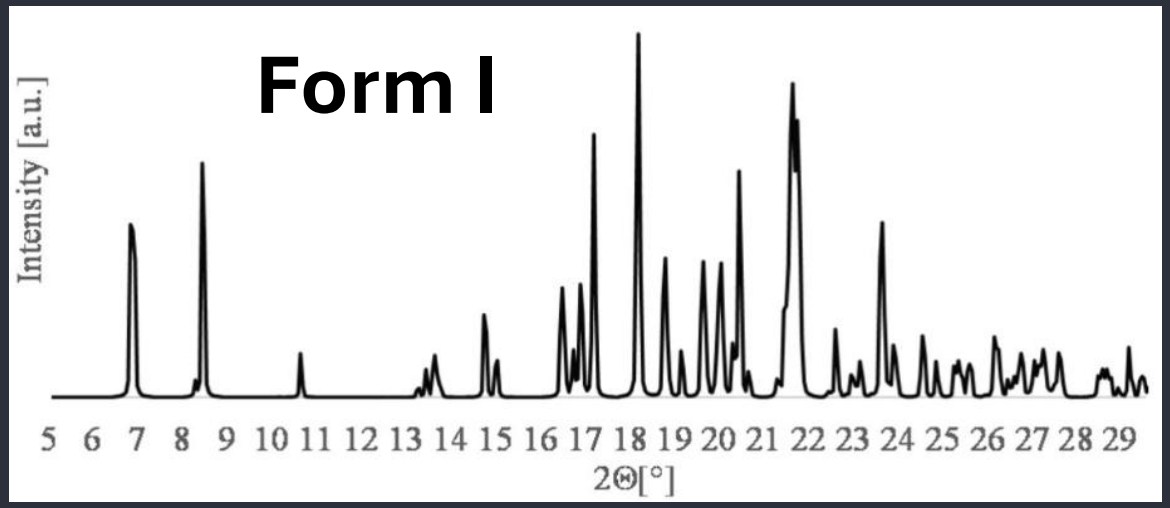
Example: Same molecule. Different crystal structure. Different drug performance: Polymorphism.

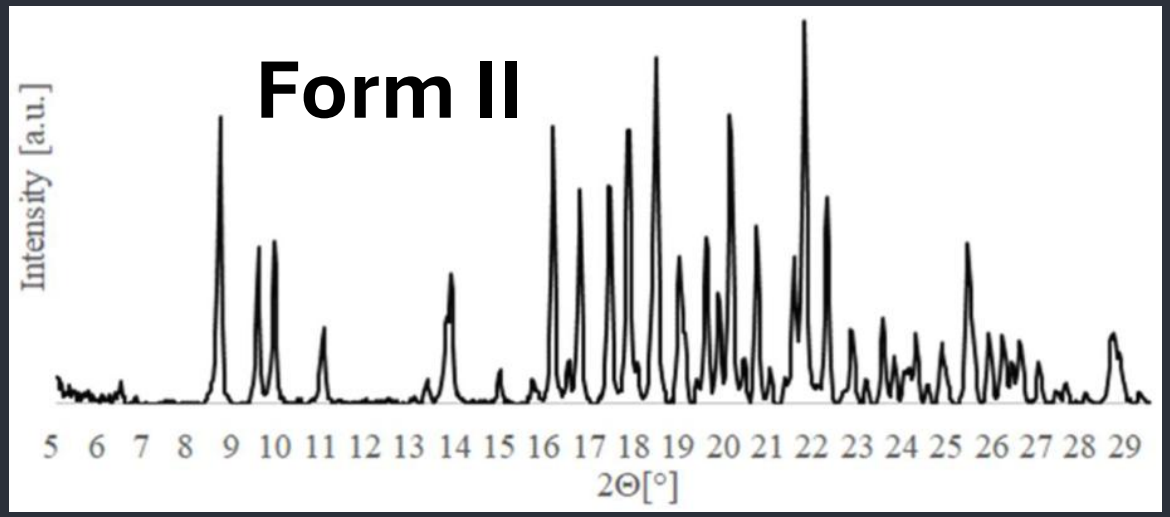
The case of Ritonavir

Treament of HIV-1 infection

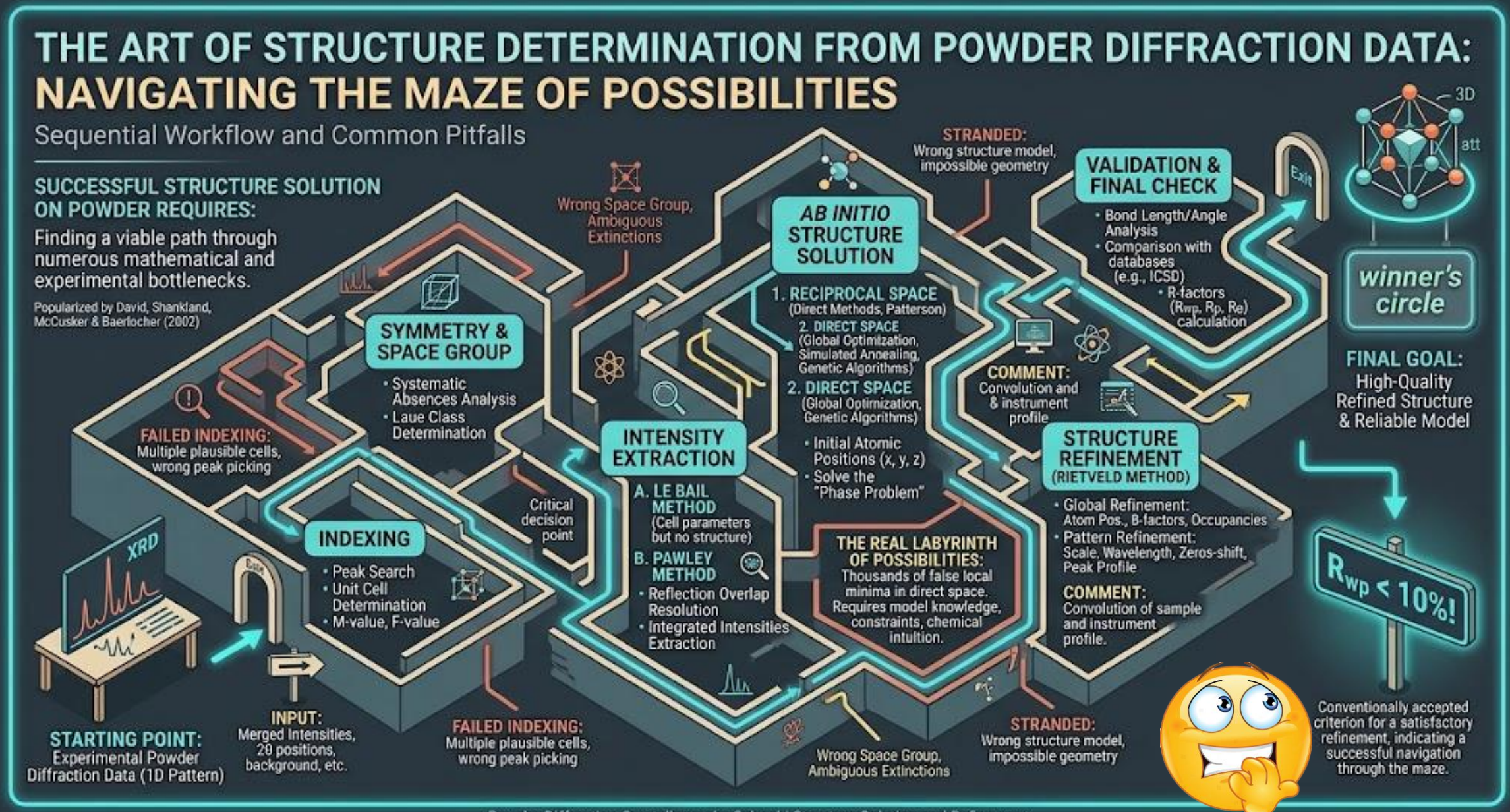


In 1998, a previously unknown Form II appeared during manufacturing. It was less soluble, causing the marketed capsules to fail dissolution tests. The product had to be withdrawn and reformulated.





New structure? No worries! CRISTAL is the answer!



Indexation

Indexing consists of solving this equation for each observed reflection

$$Q(hkl) = \frac{1}{d_{hkl}^2} = h^2a^{*2} + k^2b^{*2} + l^2\,c^{*2} + 2hk\,a^*b^*\cos\gamma + 2hl\,a^*c^*\cos\beta + 2kl\,b^*c^*\cos\alpha$$

a, b* and c* being the reciprocal lattice parameters*

Cubic $\frac{1}{d_{hkl}^2} = \frac{h^2 + k^2 + l^2}{a^2}$

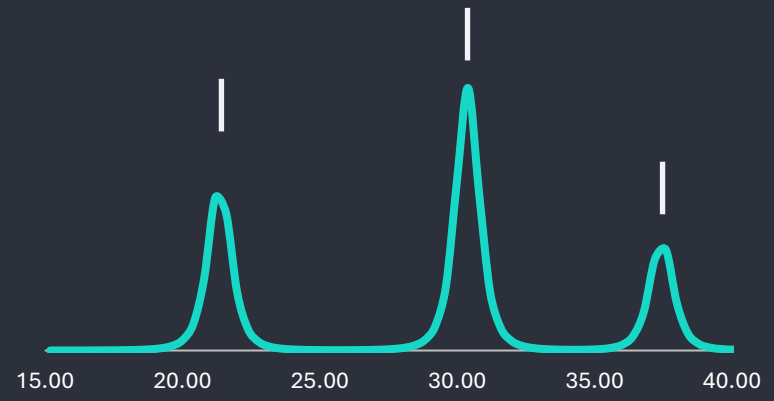
Rhombohedral $\frac{1}{d_{hkl}^2} = \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2}$

Tetragonal $\frac{1}{d_{hkl}^2} = \frac{h^2 + k^2}{a^2} + \frac{l^2}{c^2}$

Hexagonal $\frac{1}{d_{hkl}^2} = \frac{4}{3}\left(\frac{h^2 + hk + k^2}{a^2}\right) + \frac{l^2}{c^2}$



Indexation



1. Deterministic / Systematic Search

- **DICVOL** *(Boultif & Louër):* Dichotomy method. Systematically varies cell parameters between min/max bounds using constant steps.

2. Trial and Error

- **TREOR / N-TREOR** *(Werner; Altomare et al.):* Assigns small Miller indices to the first few reflections and solves linear equations to determine cell parameters.
- **ITO** *(Visser):* Optimized for **triclinic** and lower-symmetry systems.

Indexation

3. Stochastic / Random Search

- **MacMaille** *(A. Le Bail):* Monte Carlo random generation of cells, calculating Q_{hkl} and comparing them with observed line positions.

Figures of Merit - FOM:

- **De Wolff (M_{20}):** Evaluates precision of peak positions and the proportion of unindexed lines. ($M_{20} > 10$ usually indicates a reliable solution).
- **Smith-Snyder (F_N):** Evaluates the average absolute error in 2θ versus the number of observed and calculated lines.

« Space Group » (SG) Determination

1. Classical Analysis (Systematic Absences)

- Systematic Absences (Extinctions): Analyze reflection conditions (hkl, $0kl$, $h0l$, etc.) starting from low-angle data.
- **Lattice Centring & Laue Class:** Determine centring type (P, C, I, F, R) and crystal class.

2. Software-Assisted

- **Automated Extinction Testing:** Programs like *FullProf/FOX* use symmetry testing to perform Le Bail fits across all space groups, ranking them based on systematic absence agreement and R factors.

3. Rule of Thumb

- Always test the highest possible symmetry first

Intensity Extraction

Profile fitting without a structural model

The observed intensity (I_{hkl}) is directly linked to the **Structure Factor (F_{hkl})**, which is the Fourier transform of the **Electron Density (ρ)** inside the unit cell:

$$I_{hkl} \propto |F_{hkl}|^2 \quad \& \quad F_{hkl} = \int_V \rho(\mathbf{r}) e^{2\pi i \mathbf{G}_{hkl} \cdot \mathbf{r}} \, d\mathbf{r}$$

Form factor

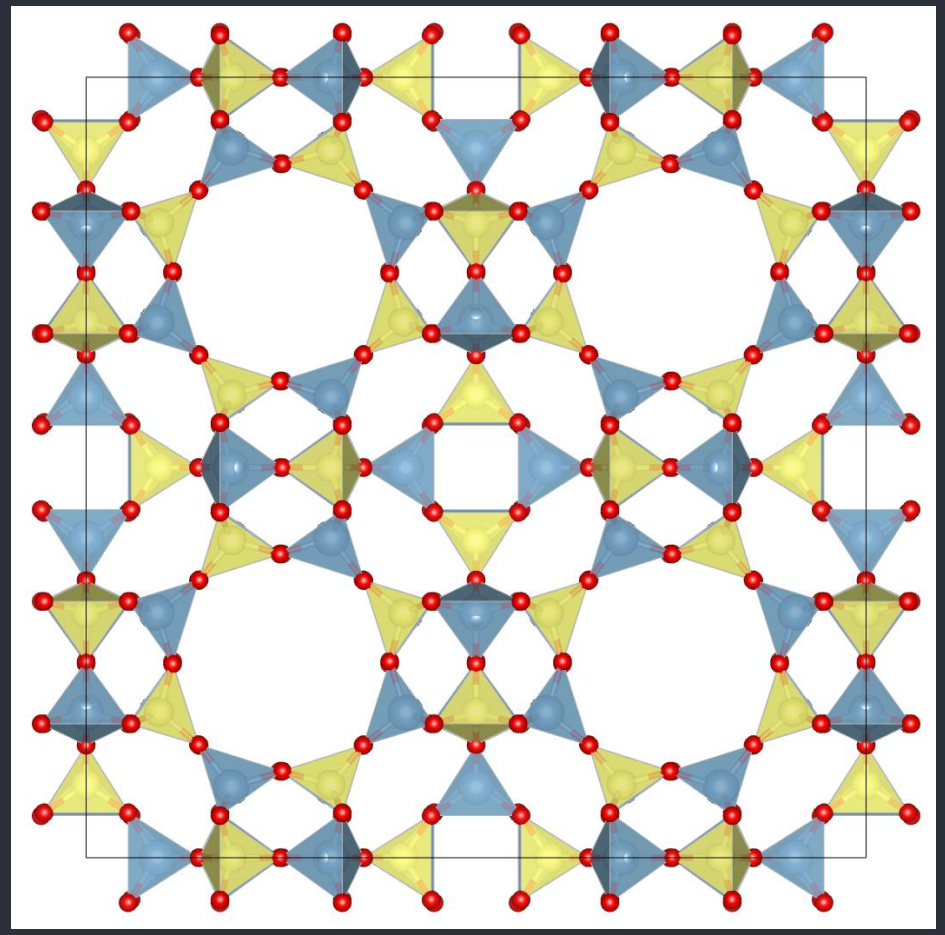
Electron density

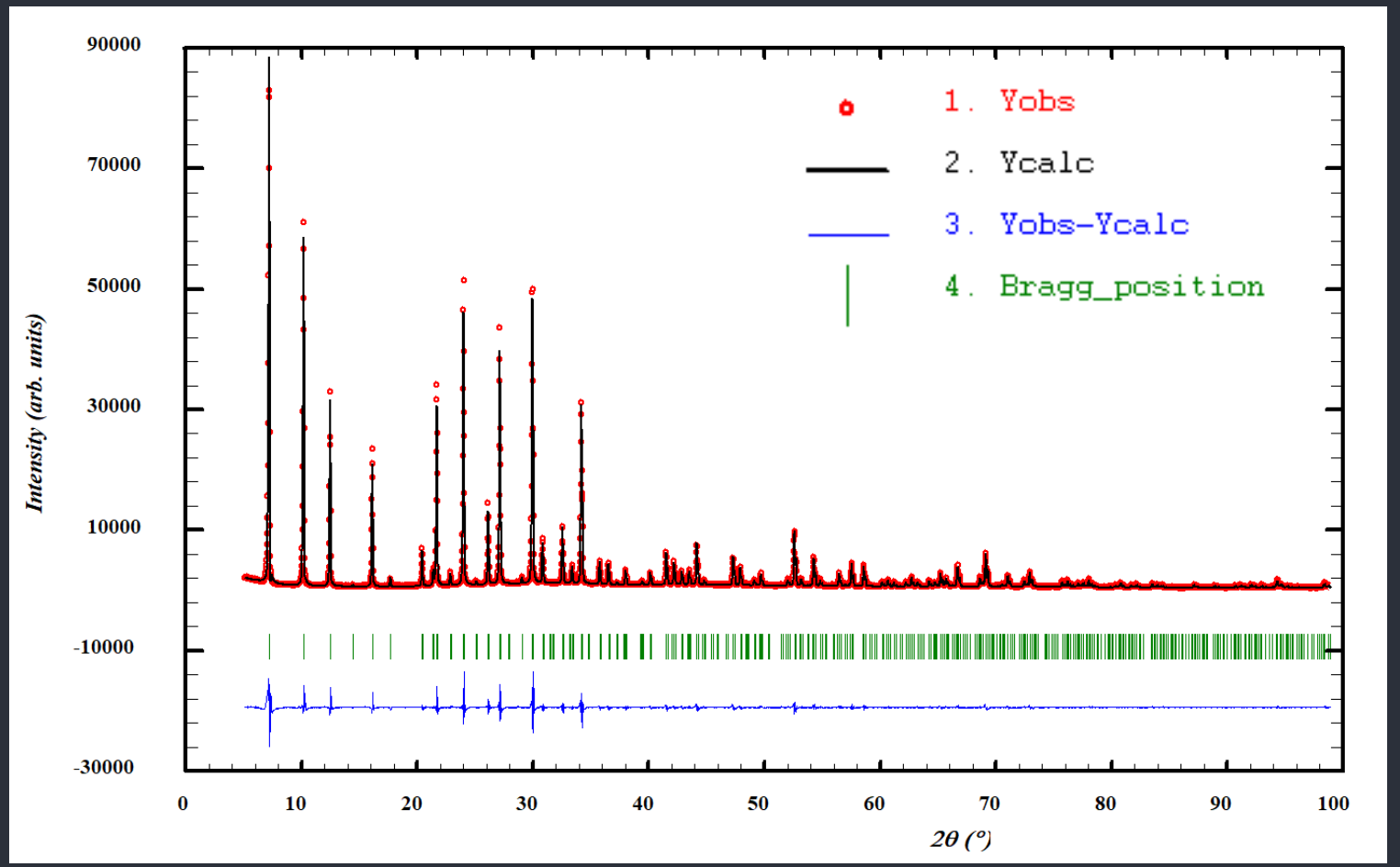
Le Bail: Iteratively partitions observed peak intensity among overlapping reflections based on calculated proportions.

Pawley: Treats every individual reflection intensity (I_{hkl}) as an independent free variable in least-squares refinement.

Le Bail refinement

Zeolithe LTA - $Na_{12}[Al_{12}Si_{12}O_{48}]_8$





Space Group $Fm\bar{3}c$ (n°226)
a = 24.58 Å

- Intensities set free
- Peaks position constrained to lattice parameters and SG

ab initio structure determination

0. Single Crystals: Get back to the lab and improve the crystallinity

1. Direct Methods (Reciprocal Space)

Relies on mathematical probability relationships (e.g., triplet phase invariants) to directly determine the phases of structure factors, followed by a Fourier transform to reveal electron density.

Pros: No need to know the exact asymetric unit.

Software: EXPO, (SHELXT, SHELXS, SIR).



EXPO – Direct Methods in a nutschell

The « famous » **Phase problem**: Diffraction measures intensities, but phases are lost.

Principle of Direct Methods: Statistical estimation of missing structure-factor phases using relationships between reflections.

Key concept: Triplet invariants connect reflections whose Miller indices satisfy R1 + R2 + R3 = 0. For sufficiently strong normalized structure factors, their phase sum is statistically likely to be close to zero, allowing missing phases to be estimated.

For example:

$$(100) + (010) + (\bar{1}\bar{1}0) = (000)$$

$$\phi_{100} + \phi_{010} + \phi_{\bar{1}\bar{1}0} \approx 0$$

ab initio structure determination

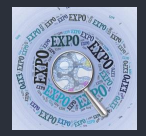
2. Direct-Space Methods (Global & Local Optimization)

Moves atoms, molecules, or rigid fragments randomly within the unit cell (Monte Carlo / Simulated Annealing), calculates the simulated pattern, and minimizes the difference against observed data.

Software: FOX, dash, Topas, Expo, GALLOP.







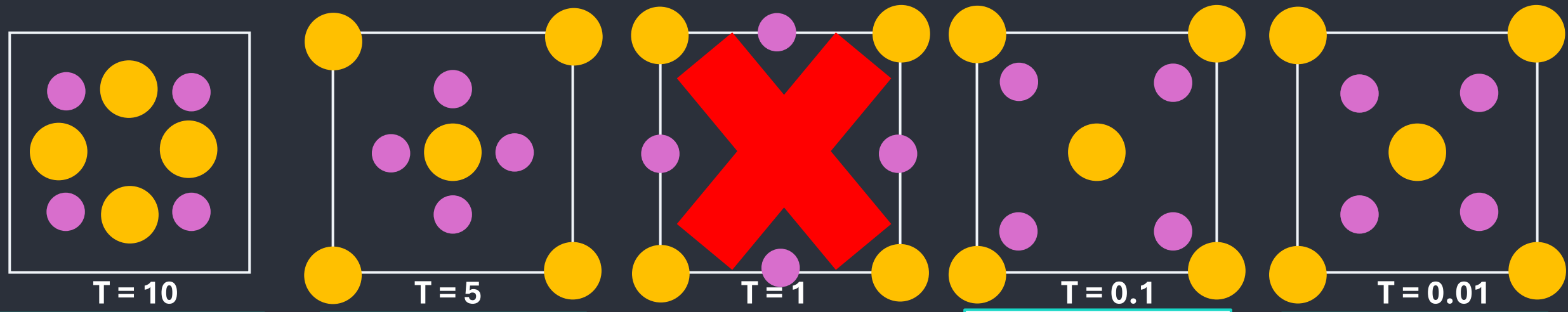
Simulated Annealing & Monte Carlo

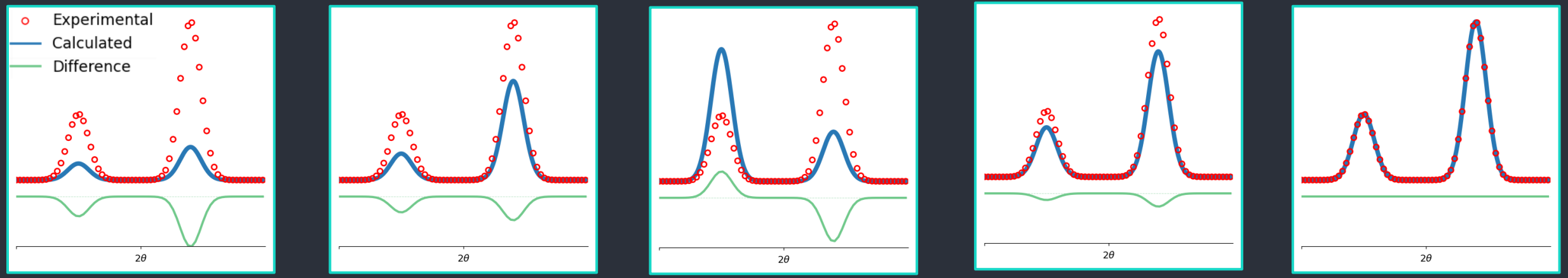
Principle: Stochastic search of structural models to find the structure that best fits the experimental diffraction data.

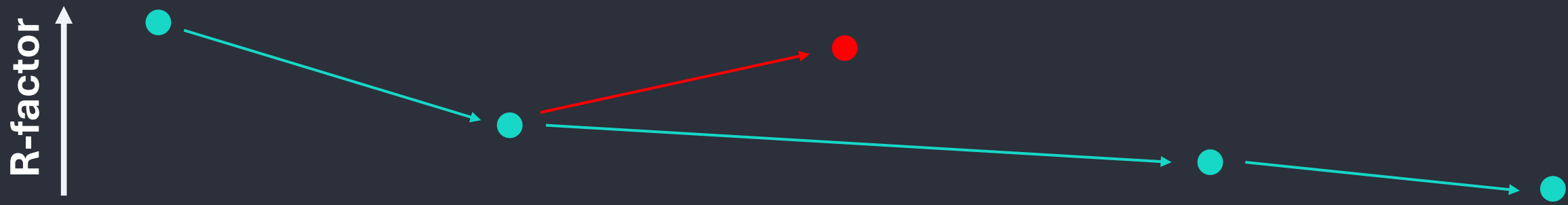
Monte Carlo: Randomly move atoms then accepts/rejects trial.

Simulated Annealing: Fictitious temperature to escape local minima.

Simulated Annealing & Monte Carlo







ab initio structure determination

3. Patterson Method

Software: WingX

4. Charge Flipping

Software: Superflip

The Rietveld method

Full-Pattern Fitting Using a Structural Model

Concept:

- Unlike Pawley/Le Bail, it **uses a complete structural model** (atomic positions, coordinates, occupancy, thermal displacement factors).
- Minimizes the difference between the **entire observed powder pattern** ($y_{i,\mathrm{obs}}$) and the **calculated pattern** ($y_{i,\mathrm{calc}}$) via Least-Squares.

$$S = \sum_i w_i \left(y_{i,\mathrm{obs}} - y_{i,\mathrm{calc}}\right)^2$$

$w_i = 1/y_{i,\mathrm{obs}}$ *is the statistical weight*

What goes into $y_{i,\mathrm{calc}}$?

- Unit cell, space group, zero, U, V, W, X, …
- Atom positions, site occupancy, atomic displacement B_{eq}/U_{iso}

Rietveld *et al.*, J. Appl. Cryst., 1969

Rietveld refinement - Strategy & Validation

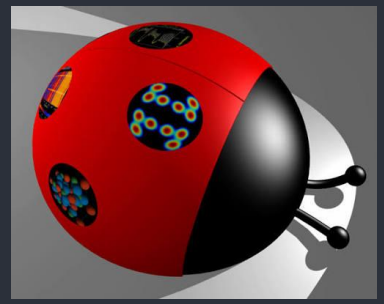
Step-by-step approach & quality indicators

Refinement Strategy - The Golden Rule: "Slow and Steady"

- Profile function (Pseudo-Voigt, Pearson VII, …) & Scale factor ***
- Background, unit cell , zero-shift.
- Caglioti U, V, W: FWHM$^2 = U\,tan^2\,\theta + V\,tan\,\theta + W$ #Gaussian
- X (and Y) #Lorentz
- Atomic coordinates (x, y, z), Occupancy, thermal parameters (B_{eq}/U_{iso})

Validation with Statistical R-factors:

- Profile residual: R_p < 10%
- Weighted profile residual: R_{wp} < 10%
- Goodness of Fit (GoF or χ^2): $\chi^2 = \left(\frac{R_{wp}}{R_{exp}}\right)^2$

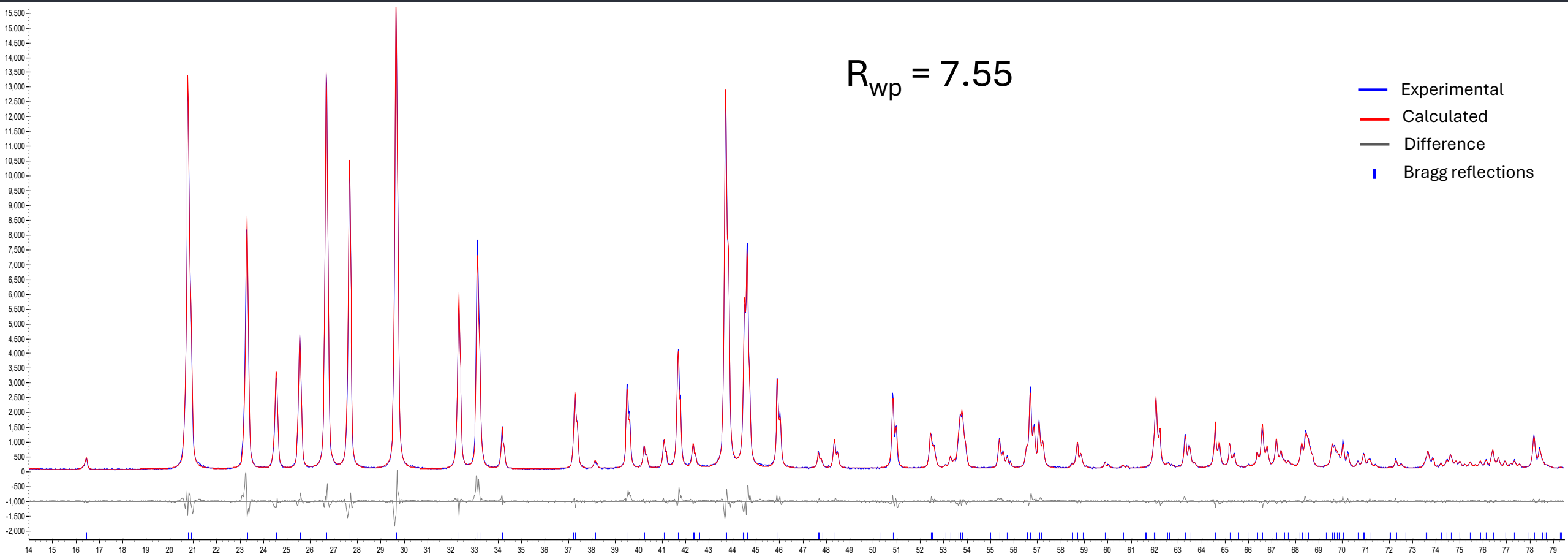


MAUD



TOPAS

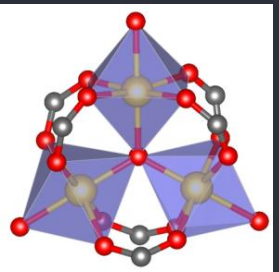
Rietveld refinement of $PbSO_4$



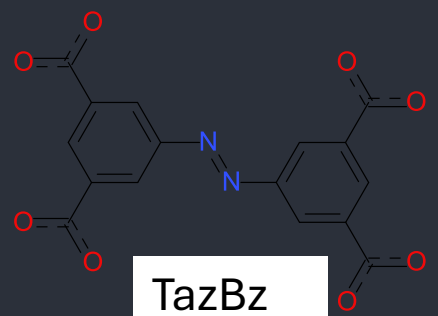
X-ray data – Durham University

Example 1: MIL-127(Fe)

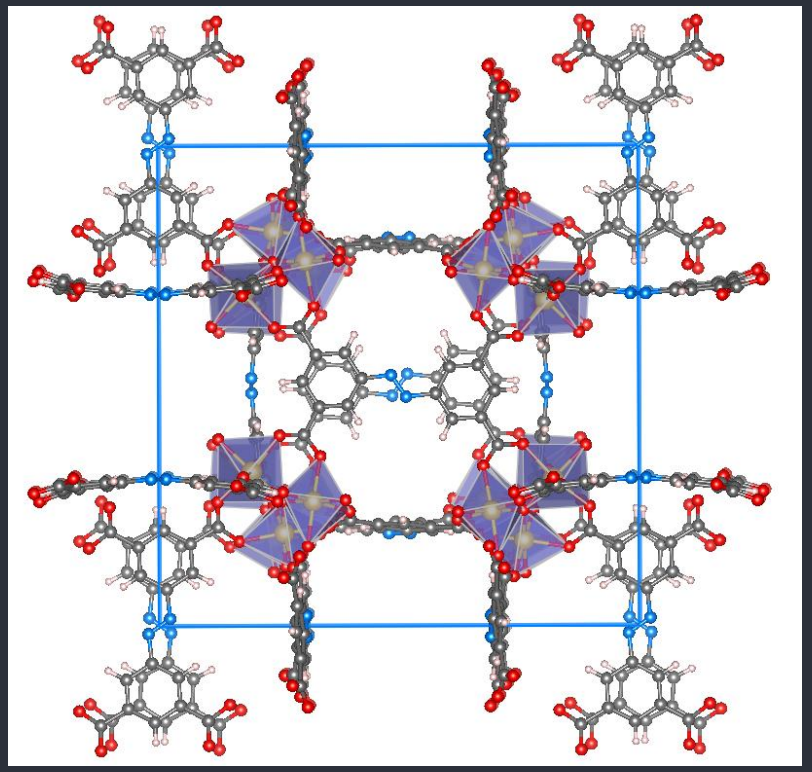
C. Serre, T. Devic, P. Horcajada



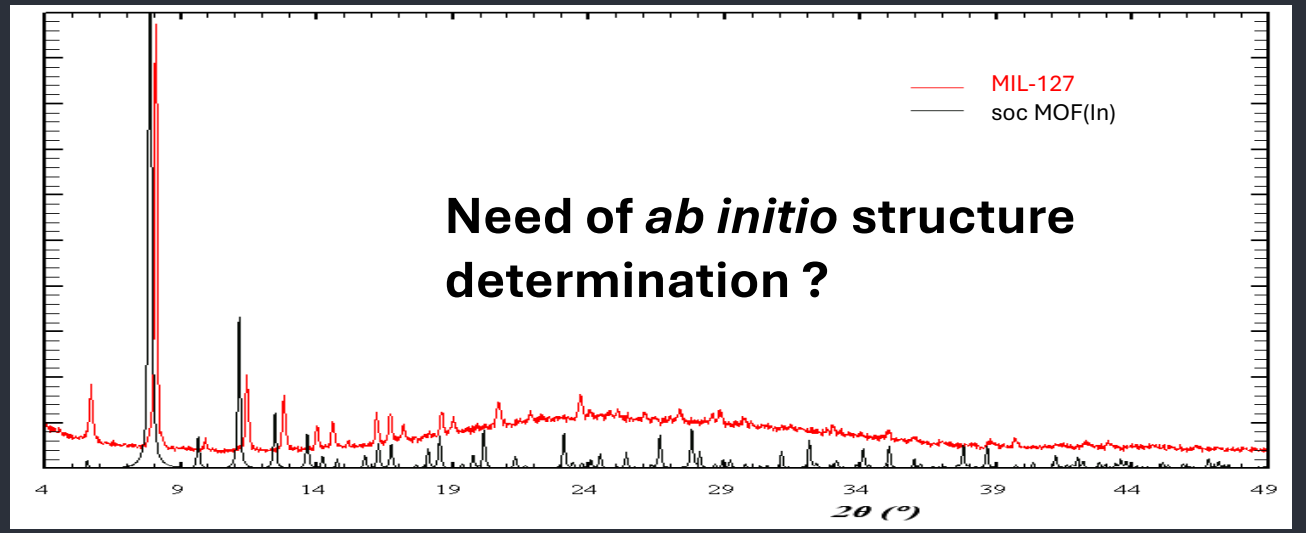
Fe trimers

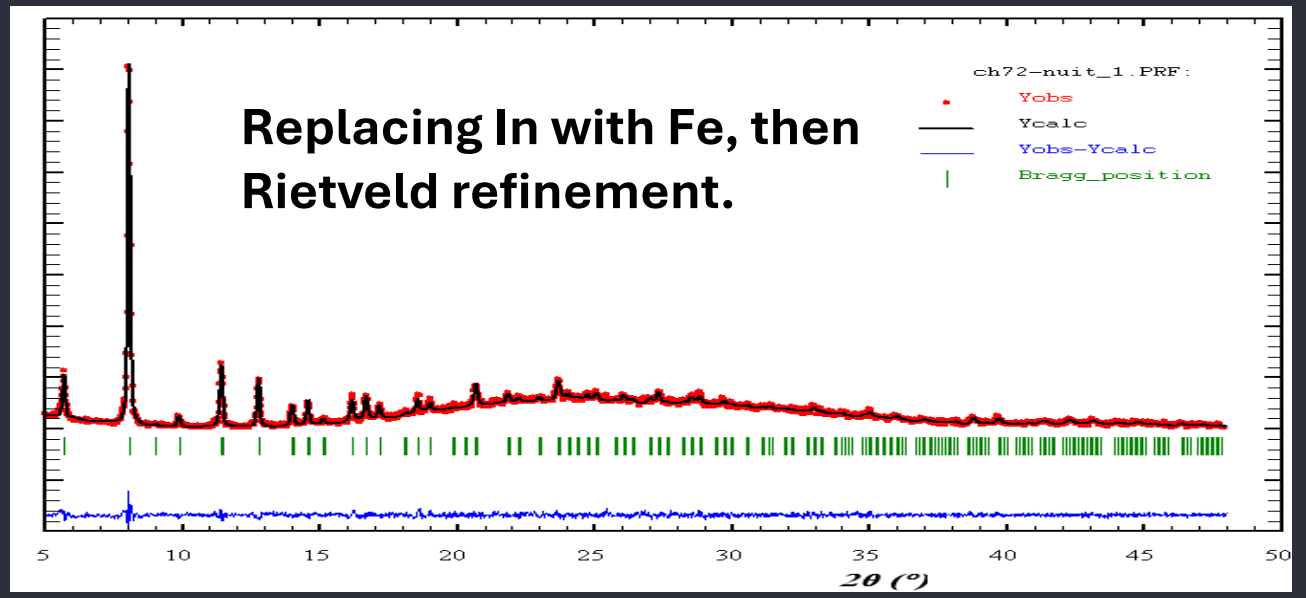


TazBz



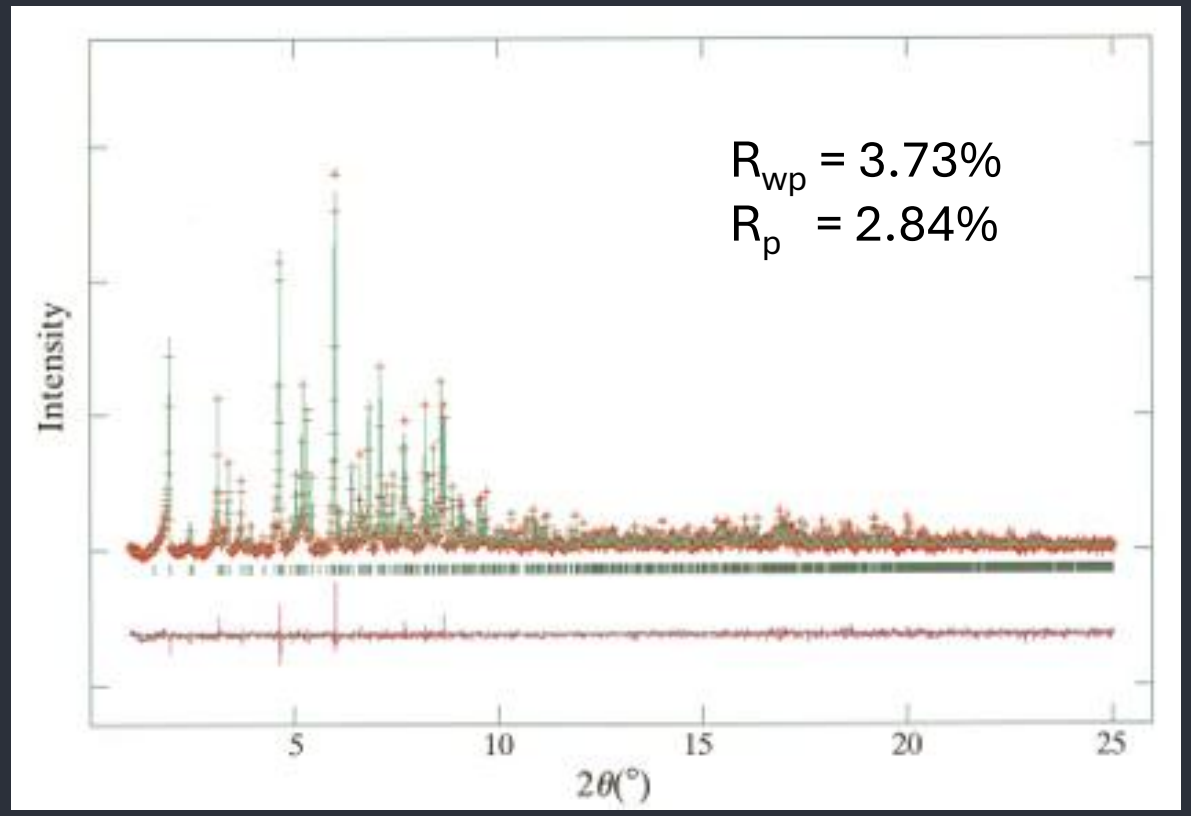
Chevreau *et al.*, Cryst. Eng. Comm., 2016



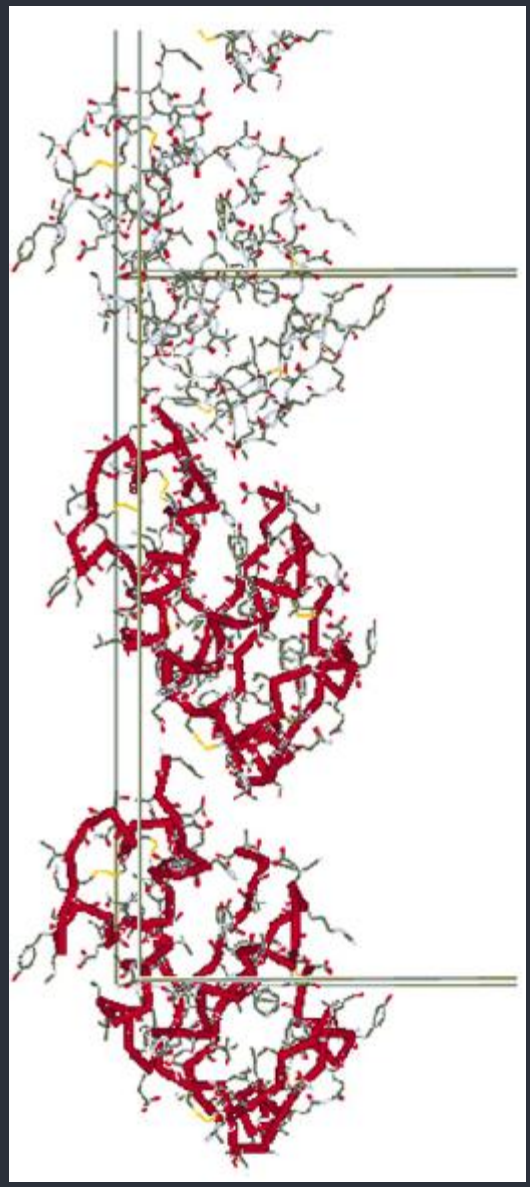


Example 2: Protein crystal structure determination

human insulin-zinc complex



High-resolution X-ray powder diffraction profille from the final Rietveld refinement of the T_3R_3DC complex



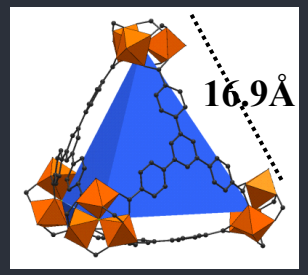
846 non-hydrogen atoms

Rhombohedral R3
a = 81.2780(7) Å
c = 73.0389(9) Å
V = 417860(8)Å^3

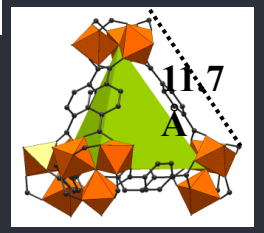
GSAS

Von dreele *et al.*, Acta. Cryst. D, 2000

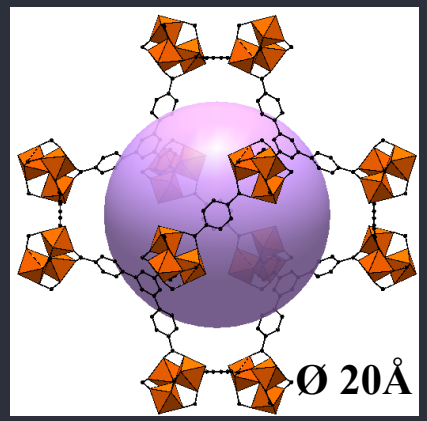
Example 3: MIL-142

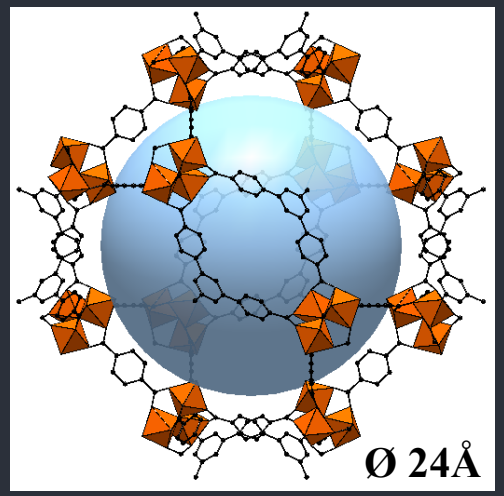


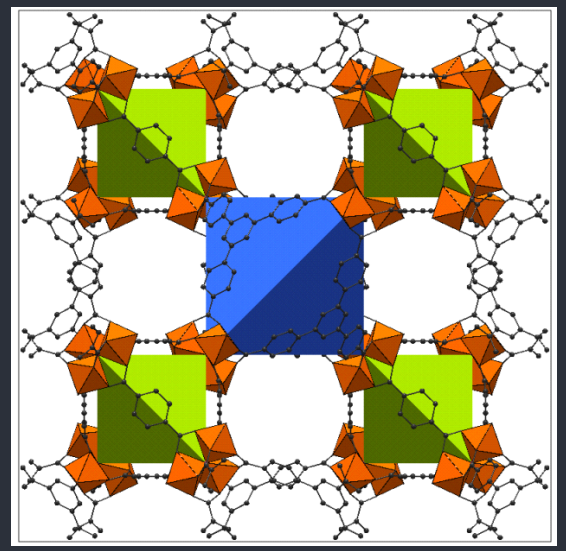
Extended ST of MIL-100



ST of MIL-101

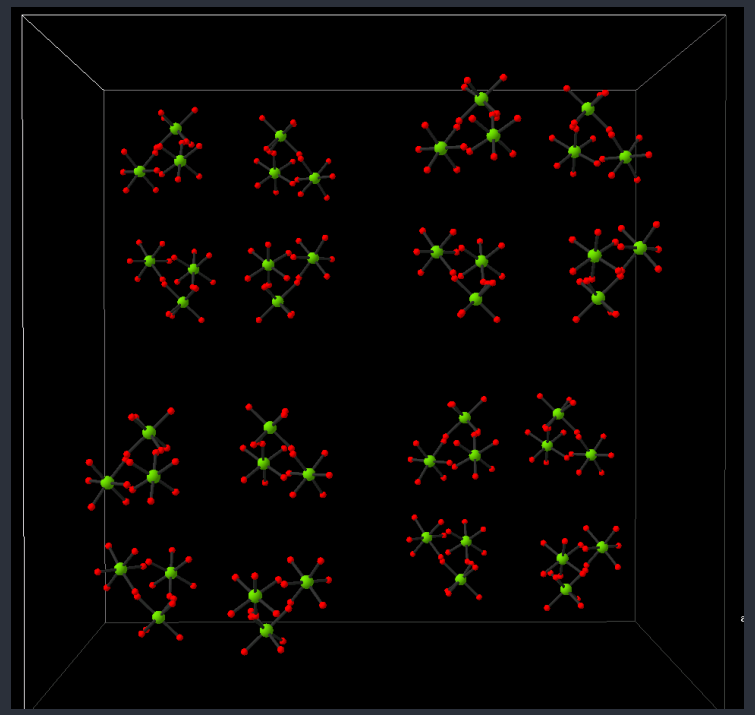


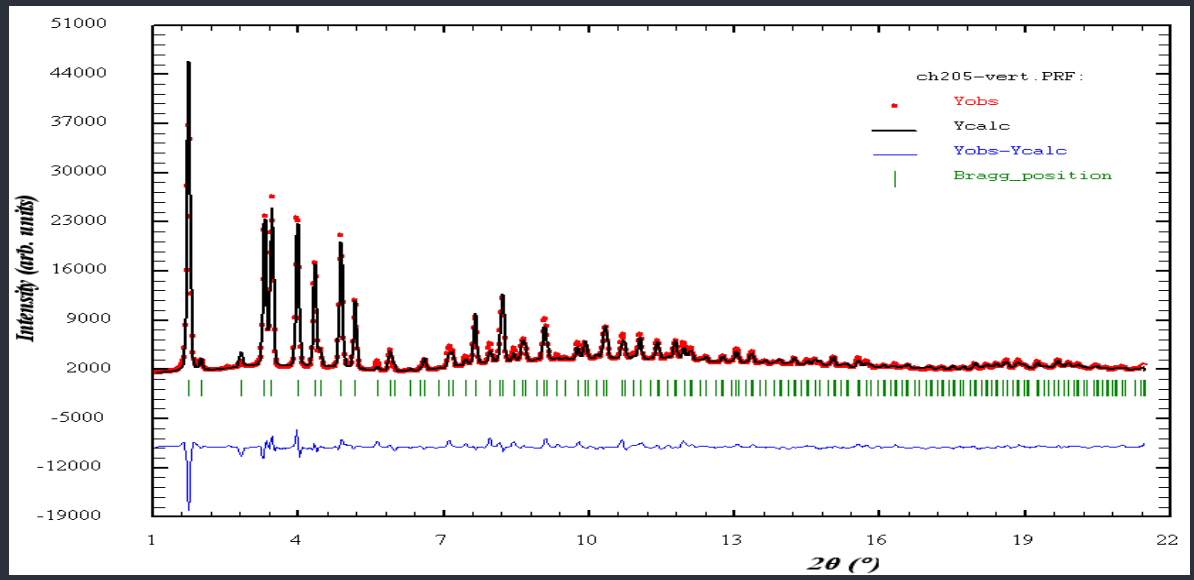




Space Group F23
a= 40.8635(1) Å

FOX to locate Iron trimers
+
Simulation to build the trimers and linkers

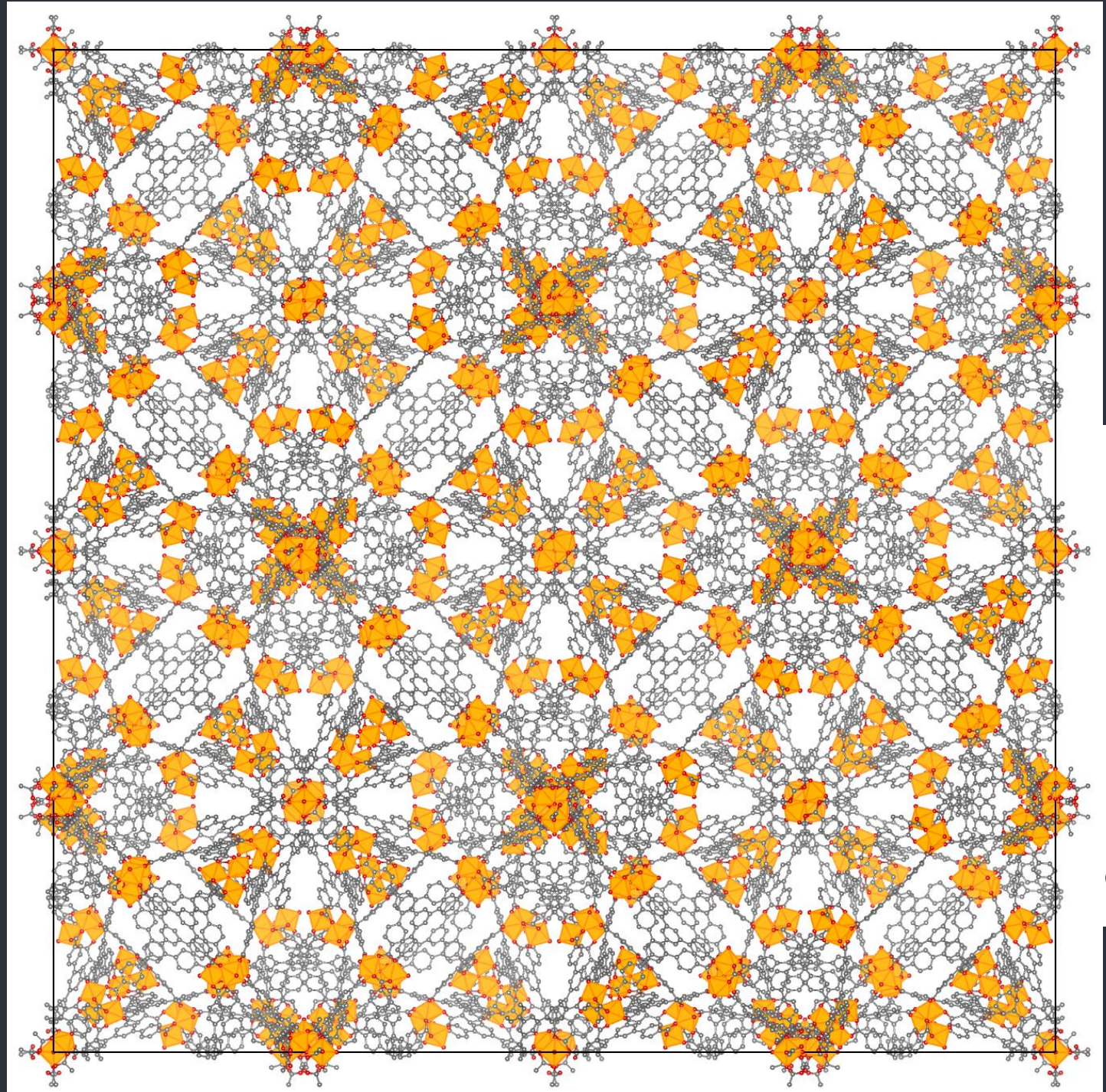




Chevreau *et al.*, Angew. Chem, 2013

Rietveld Refinement (R_p=5.40, R_{wp}= 5.97). BM1A@ESRF

Example 4: MIL-100 Extended



Cubic, $Fd\bar{3}m$ (n°227)

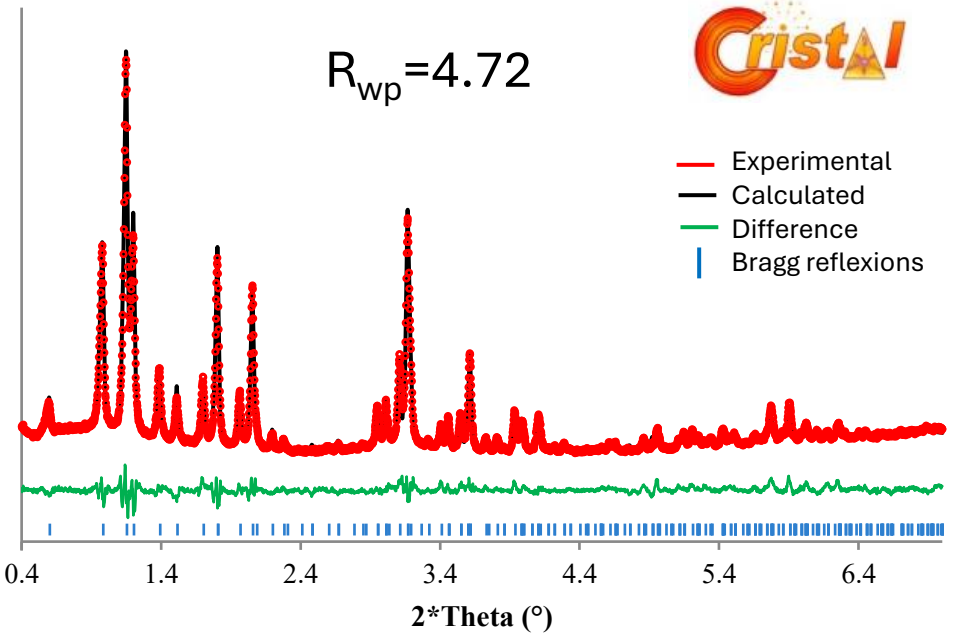
a = 130.005(5)Å

V = 2197253(6) Å^3

124 independent atoms

Ø pores: 44 & 51 Å

ρ = 0.2 g.cm^{-3}



Rietveld refinement of extended MIL-100(Fe)

Férey *et al.*, Science, 2005

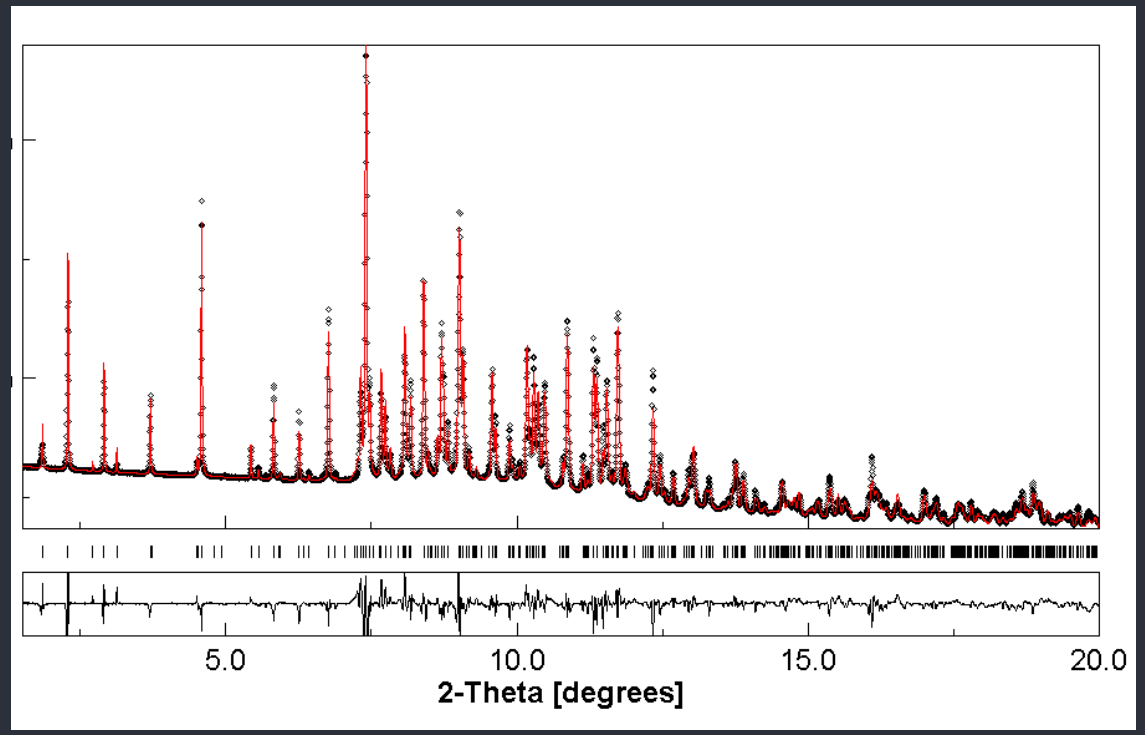
Horcajada *et al.*, Chem. Comm., 2014

Example 5: Co-Crystals

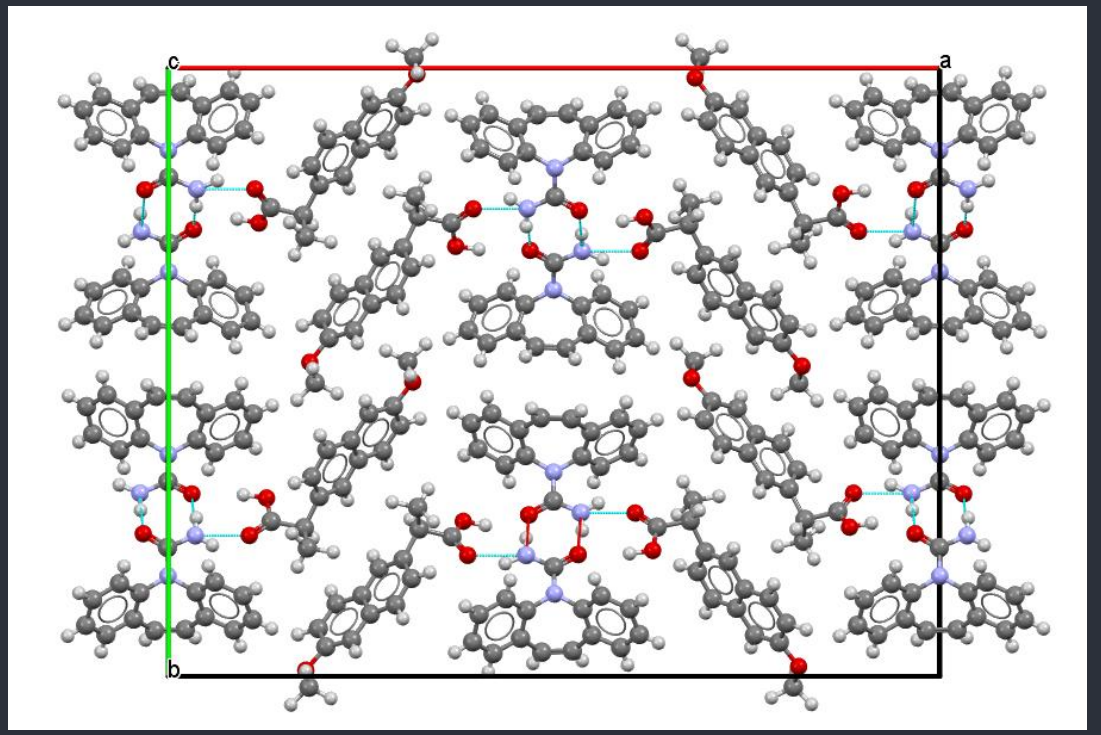
UMET
Unité Matériaux Et Transformations

M. Guerain

Optimizing the dissolution properties of poorly soluble pharmaceutical compounds while preserving their therapeutic properties



Rietveld refinement of CBZ-S-NAP cocrystals



Projection of the unit cell of carbamazepine and S-Naproxen cocrystal obtained along the c axis

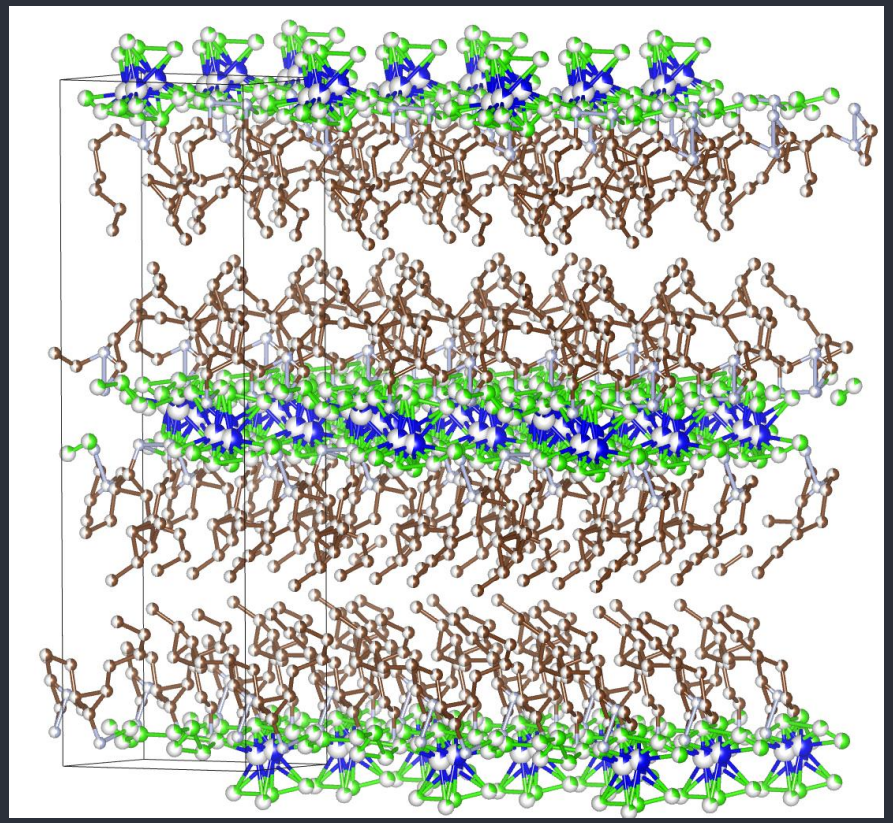
Guerain *et al.*, Acta Cryst. C., 2024.

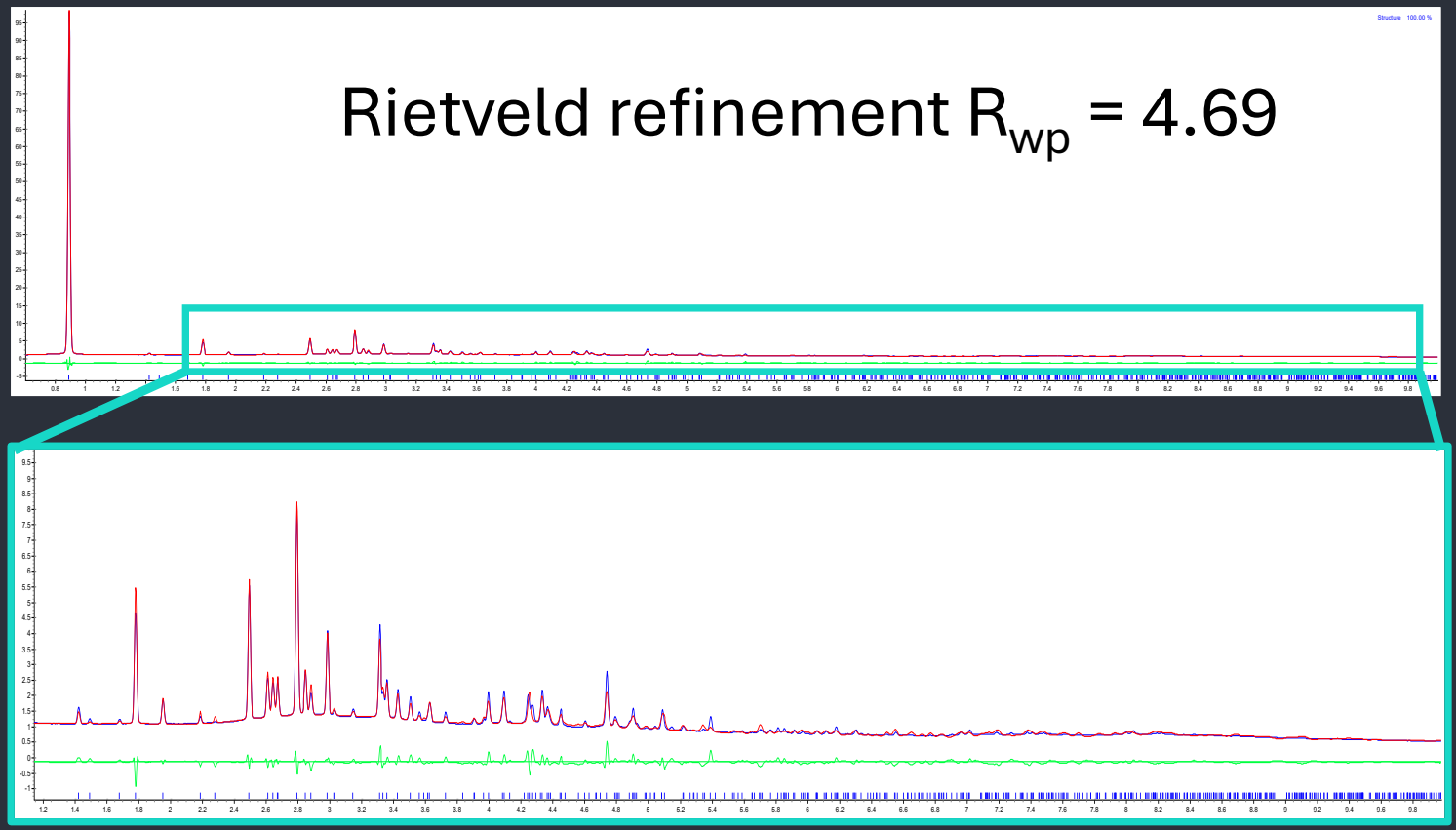
Guerain *et al.*, Acta Cryst. C., 2025.

Guerain *et al.*, Acta Cryst. C., 2026

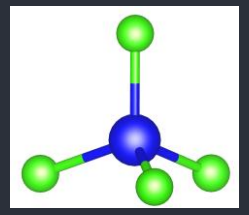
Example 6: Layered hybrid Materials for thermal storage

Solid – Solid Phase Change

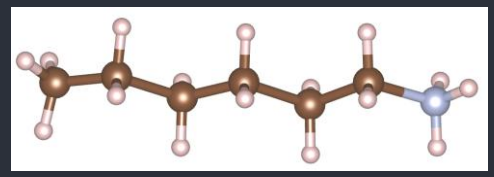




Rietveld refinement R_{wp} = 4.69



Tetrahedra $ZnCl_4$ — Disordered on 3 positions



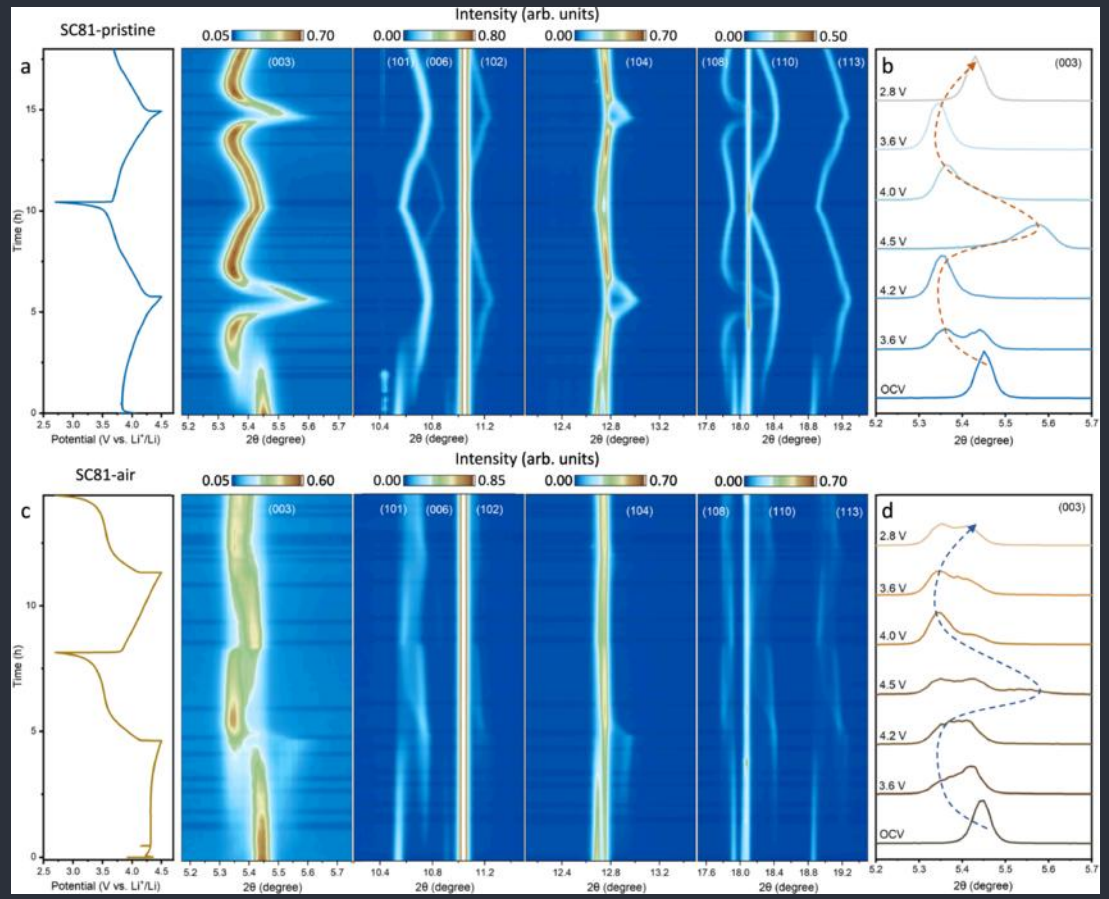
Alkylammonium C_6H_{13}-NH_3 — Disordered on 4 positions

Archinard *et al.*, Chem. Mater., 2026

Example 7: In operando studies of batteries

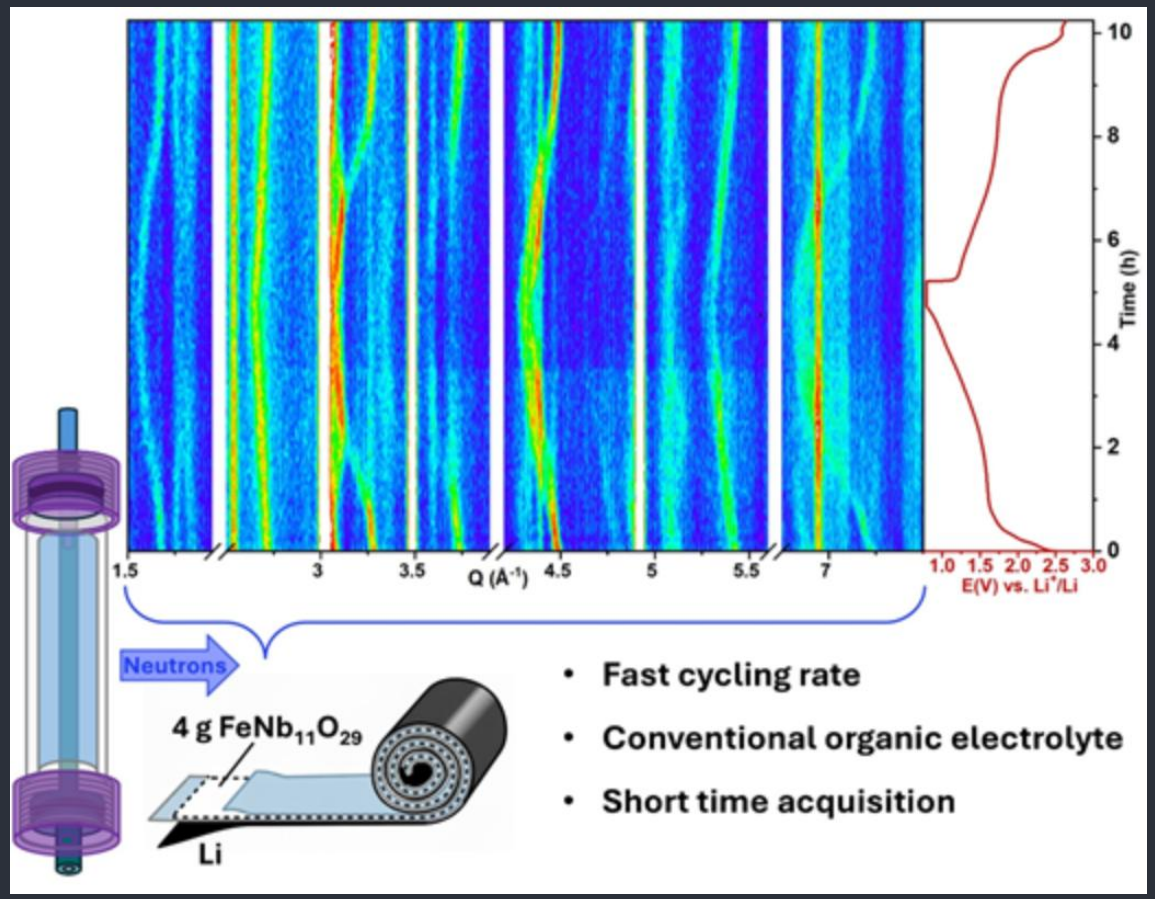
Sequential Rietveld refinement

X-ray Powder Diffration



Yu *et al.*, Nature Comm, 2025.

Neutron Powder Diffraction



Le Thanh *et al.*, J. Mater. Chem., 2026.

Fourier Map

Computed using observed minus calculated structure factors:

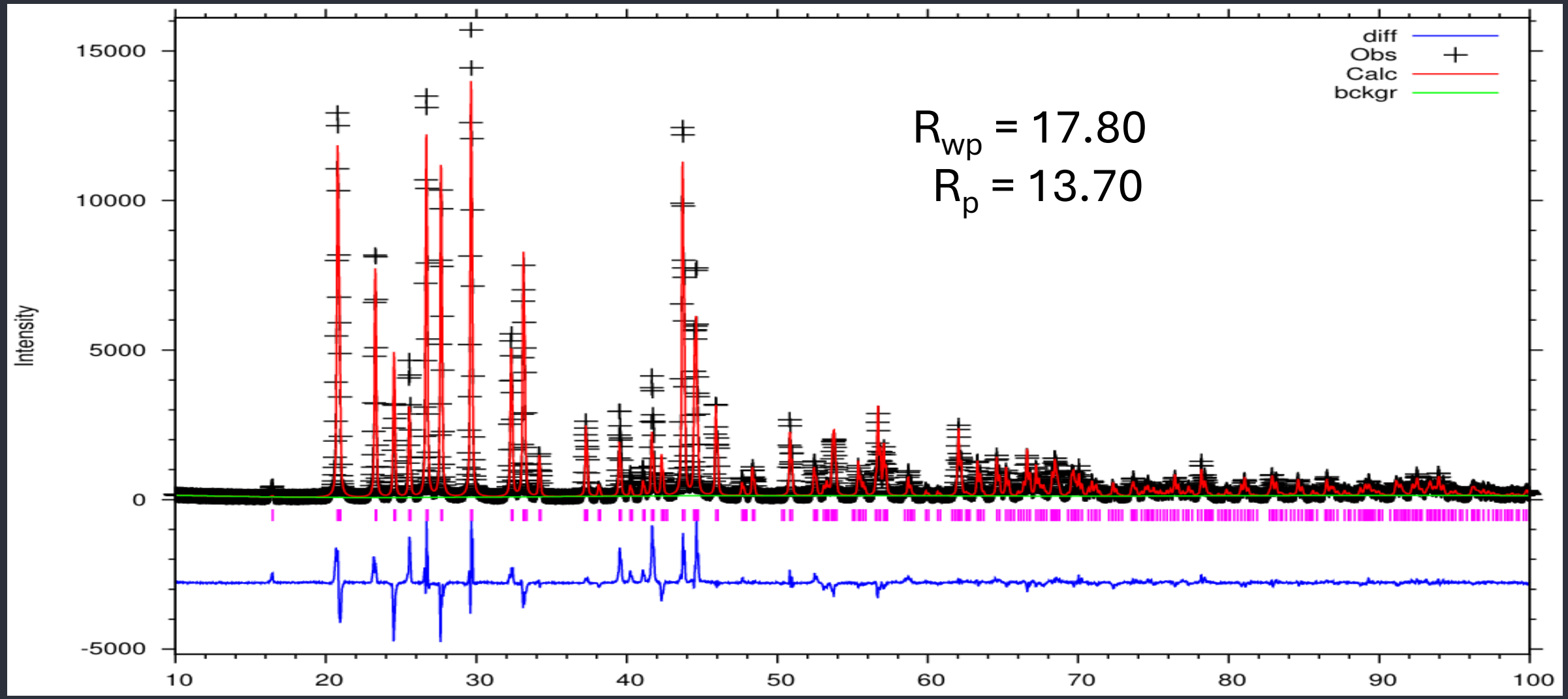
$$\Delta F = |F_{\text{obs}}| - |F_{\text{calc}}|$$

Use case:

- Reveals missing atoms
- Misplaced fragments
- Interstitial electron density (water molecules, partial substitutions, …)

Example 1: Completing structural model

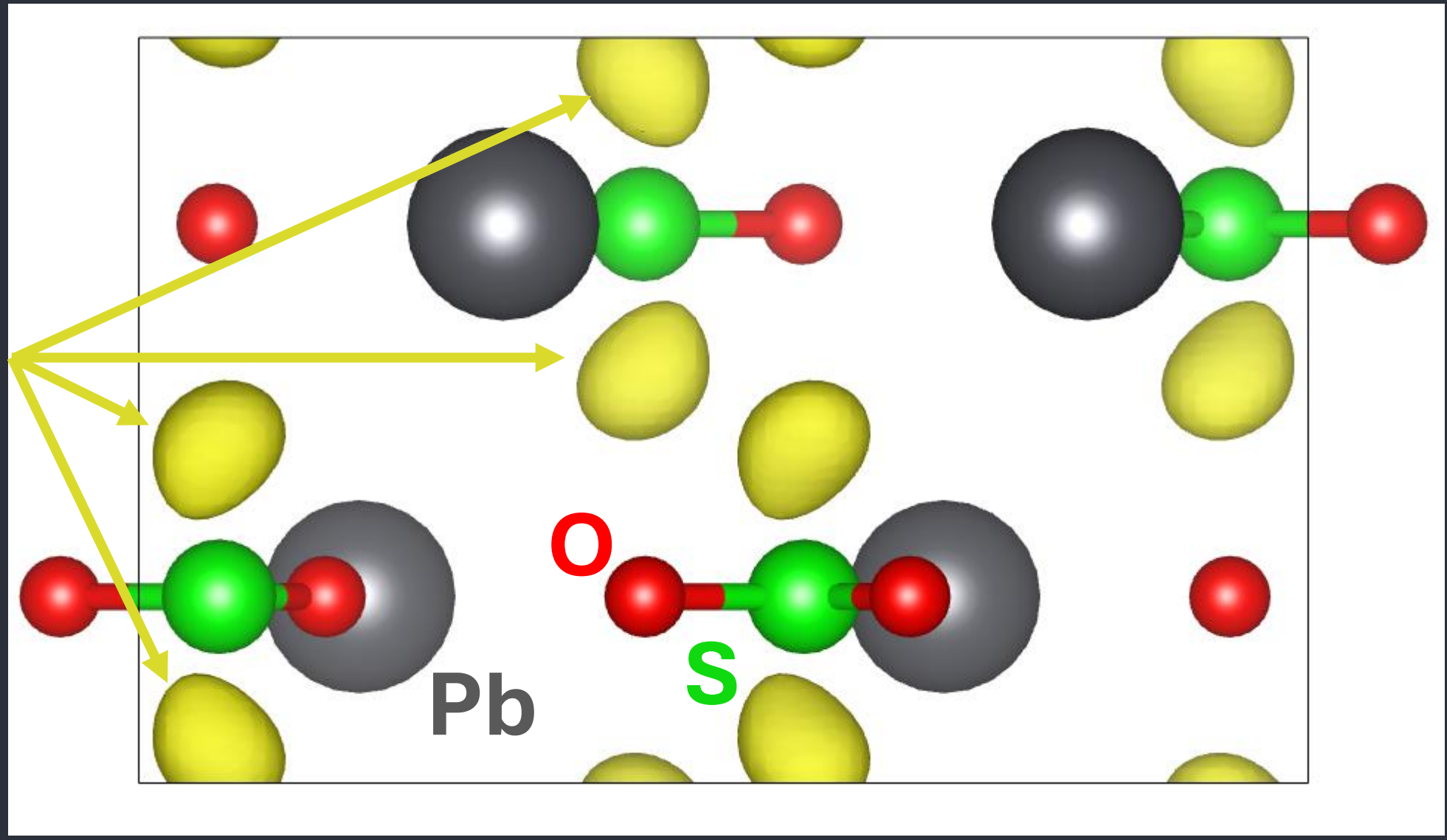
$PbSO_4$ – Heavy atom & Finding Light Atom



Example 1: Completing structural model

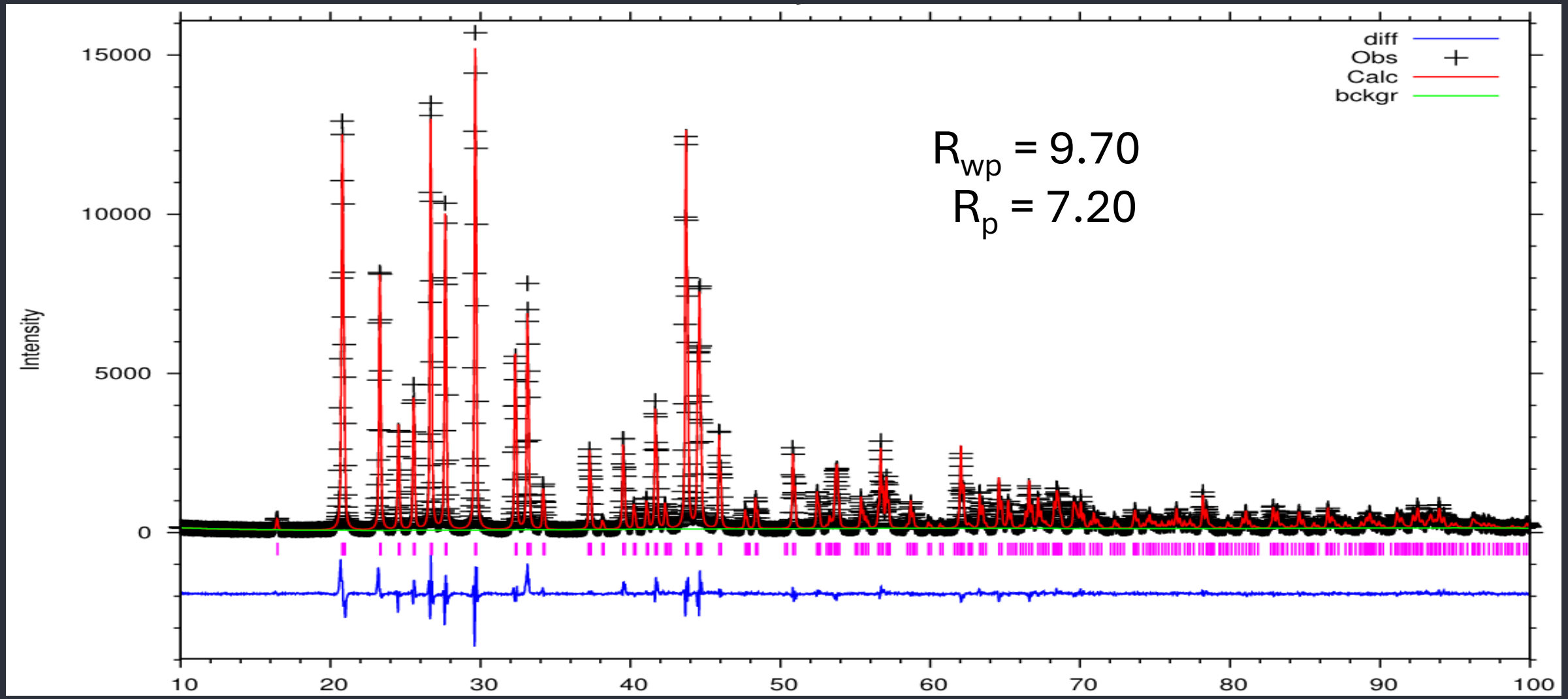
$PbSO_4$ – Heavy atom & Finding Light Atom

Residual Electron Density



Example 1: Completing structural model

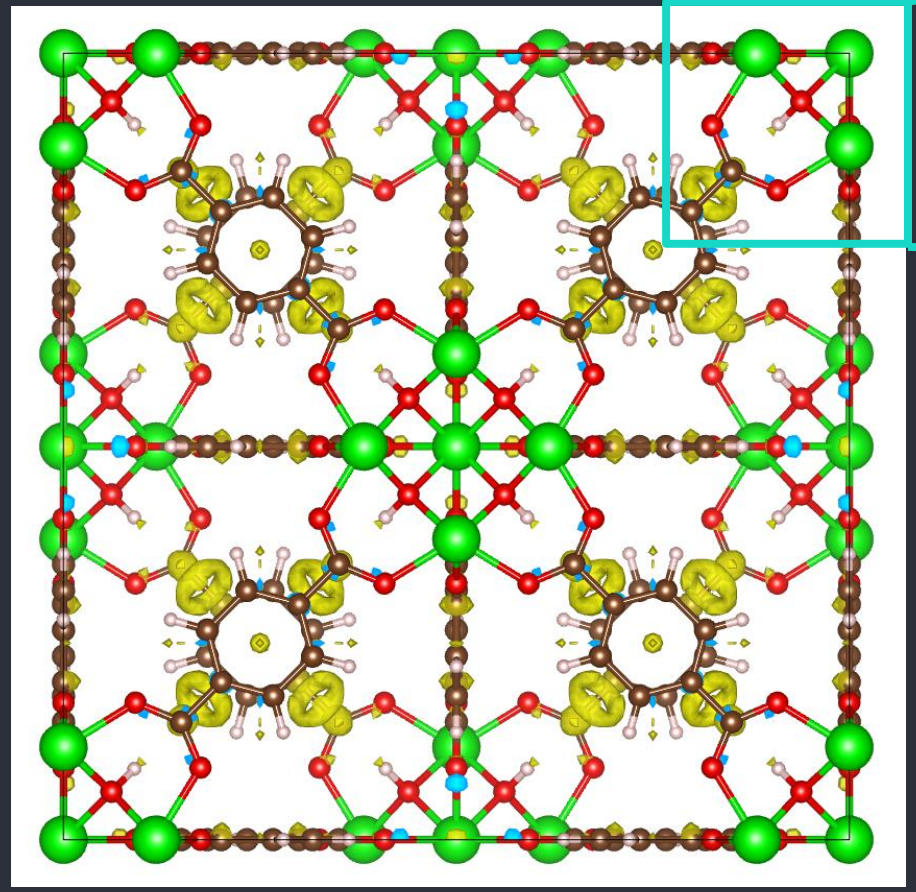
$PbSO_4$ – Heavy atom & Finding Light Atom

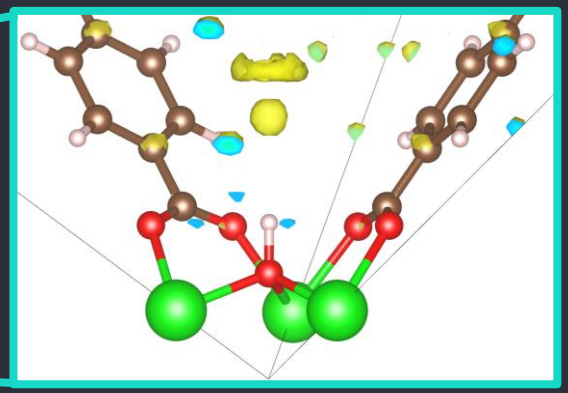


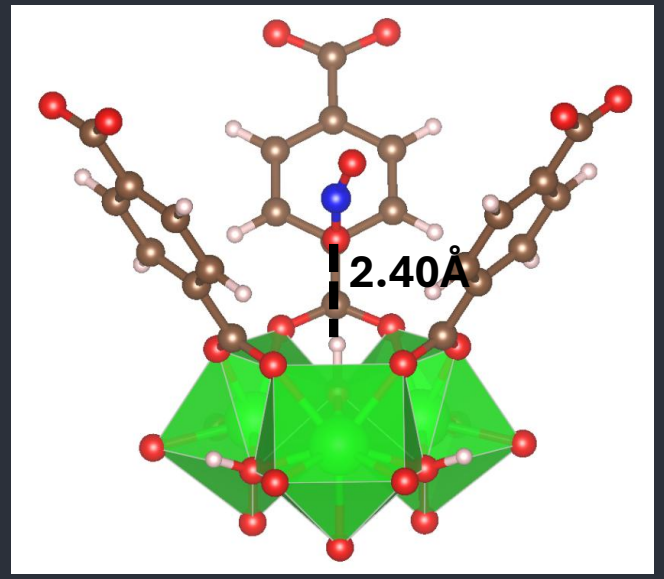
Example 2: Locating CO_2 molecules in UiO-66(Zr)

Using Neutron Powder Diffraction (NPD)

Echidna – High Resolution Powder Diffraction







Chevreau *et al.*, J. Phys. Chem. C, 2015

Quantitative Phase Analysis - QPA

Principle:

The scale factor (s) refined during a Rietveld analysis is directly proportional to the volume fraction of that specific phase in the mixture.

- The weight fraction (W_α) of phase α in a multiphase mixture of N phases is given by:

$$W_\alpha = \frac{s_\alpha (ZMV)_\alpha}{\sum_{i=1}^{N} s_i (ZMV)_i}$$

s_i = Rietveld scale factor of phase i
Z_i = Number of formula units per unit cell
M_i = Mass of the formula unit
V_i = Volume of the unit cell

Practical Requirements for Accurate QPA:

- All phases in the sample must be identified and included in the model.

CheMin XRD/XRF instrument on Mars



Rover Curiosity 2015

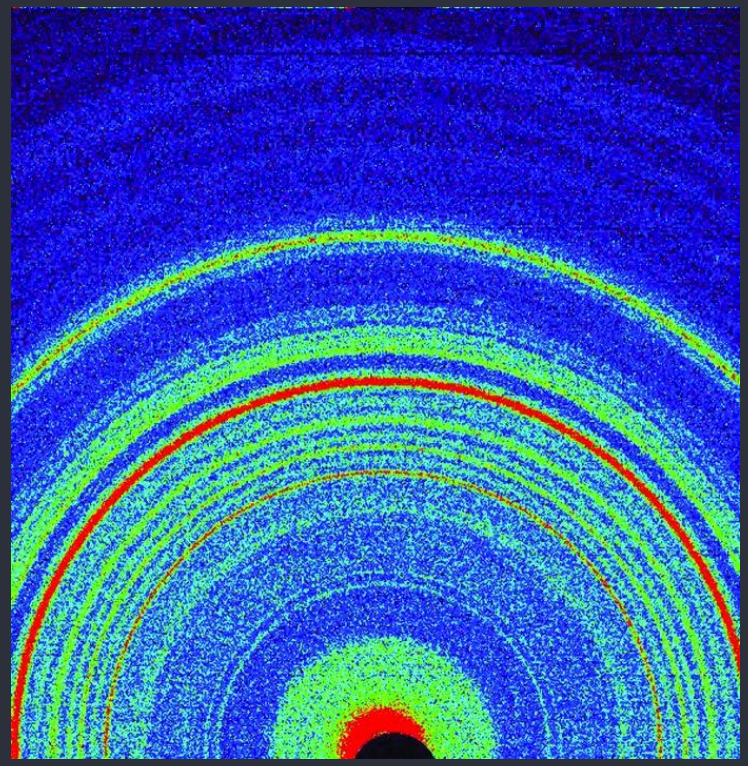


Image on the CCD detector

Mineral	**John Klein**
Plagioclase	22
Fe-forsterite	3
Augite	4
Pigeonite	6
Orthopyroxene	3
Magnetite	4
Anhydrite	3
Bassanite	1
Sanidine	1
Quartz	0.4*
Hematite	0.6*
Akaganeite	1
Pyrite	0.3*
Pyrrhotite	1
Phyllosilicate	22
Amorphous	28

Biskh *et al.*, Science, 2013.
Blake *et al.*, J. Minerals, 2024.

Microstructure Analysis: Size & Strain Broadening

Diffraction peak profiles broaden beyond instrumental effects due to physical crystal imperfections: crystallite size and microstrain:

$$B_{\text{tot}} = \text{Instrument } (B_{\text{inst}}) + \text{Size } (B_{\text{size}}) + \text{Strain } (B_{\text{strain}})$$

1. Crystallite Size Broadening : The Scherrer Equation:

$$D = \frac{K \cdot \lambda}{\beta_{\text{size}} \cdot \cos\theta}$$

2. Microstrain Broadening

- Cause: Local lattice distortions, dislocations, composition gradients, or mechanical stress causing variations in d-spacing ($\Delta d/d$).
- Stokes-Wilson Relation:

$$\epsilon = \frac{\beta_{\text{strain}}}{4\tan\theta}$$

Computational and Data-Driven Crystallography

« Density Functionnal Theory » calculations

A complementary tool to powder XRD

Principle : Solves the many-body **Schrödinger equation** to calculate the total ground-state energy of a crystal structure based on electron density:

$$\hat{H}\psi = E\psi$$

$\hat{H}$: Hamiltonien operator
E : Energy « Eigenvalue »
ψ: Wavefunction

Why use DFT with powder diffraction?

- Structure Validation & Relaxation
- Locating Light Atoms
- Bonding & Interaction Analysis

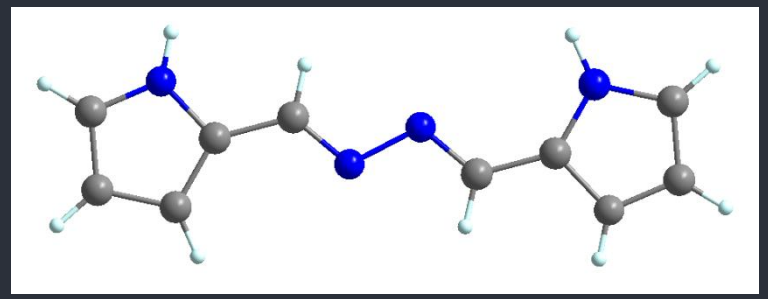
Software: Quantum Espresso, VASP, CASTEP, CRYSTAL, …

Example 1: Re-visiting Ag - π interactions in pyrrolic derivatives

Chimie de la MATIÈRE COMPLEXE
UMR 7140 - CMC

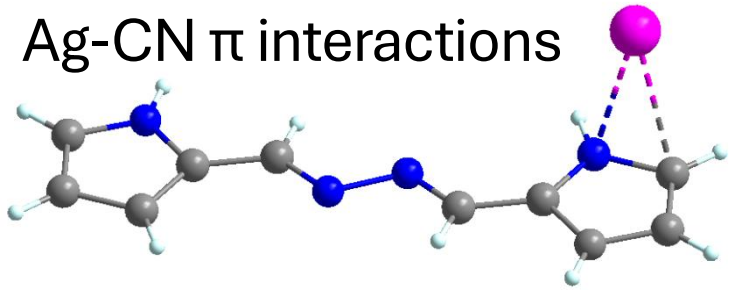
S. Baudron

bis((pyrrol-2-yl)methylene)hydrazine



+ $Ag(NO_3)$

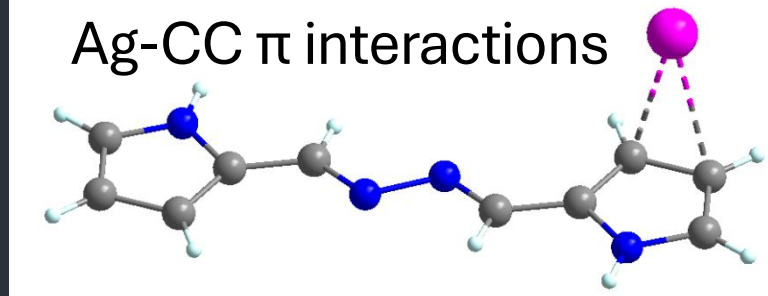
Ag-CN π interactions



Ma *et al.*, Eur. J. Inorg. Chem., 2004

OR

Ag-CC π interactions



Baudron *et al.*, Cryst. Eng. Comm., 2019

DFT calculations for stability comparison

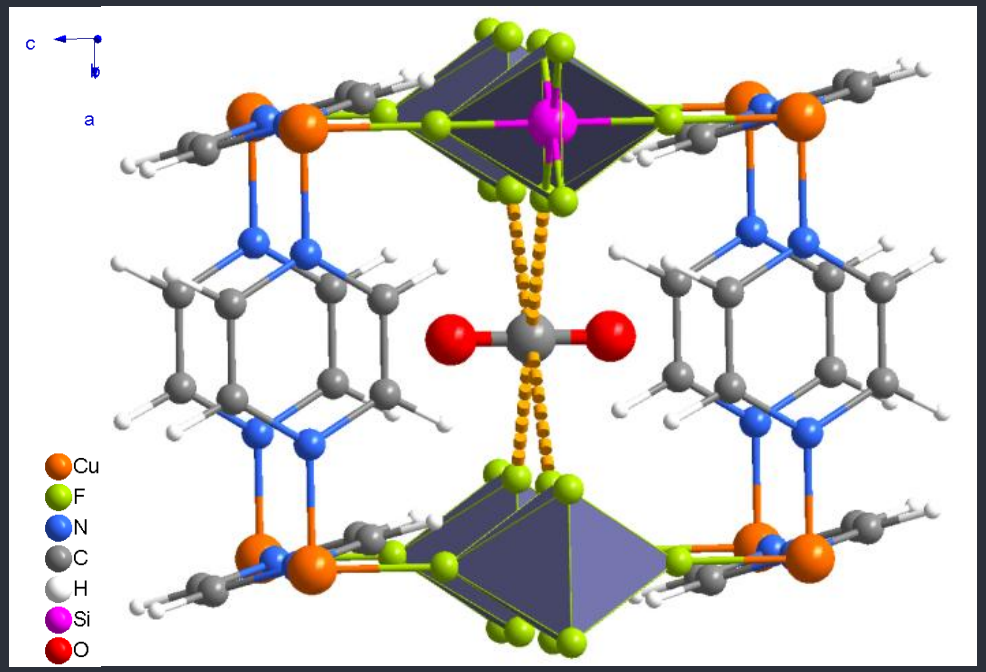
Ground-state total energy:

-2651.14 Ry

-2651.21 Ry ✓

Example 2: Complementarity Experiment / Theory

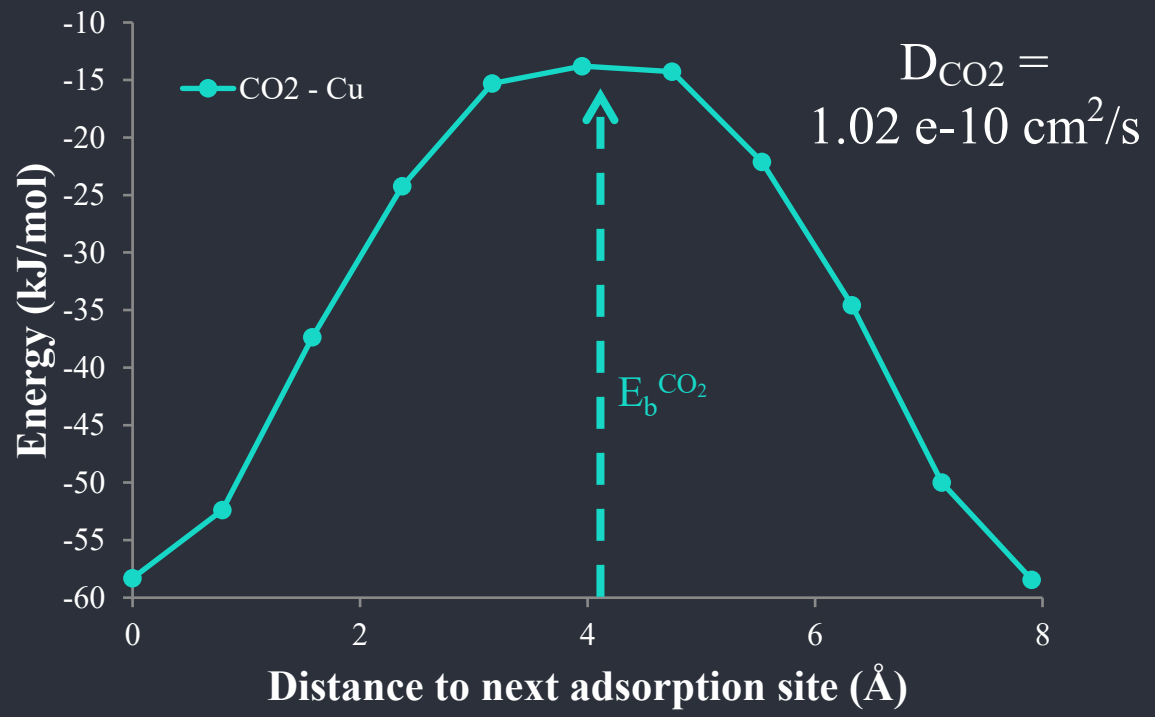
NPD to locate adsorption sites of CO_2 in SIFSIX-3-Cu



Method: Rietveld then Fourier Map

And diffusion ???

- DFT-based Molecular Dynamics
- Nudget Elastic Band



Artificial Intelligence

Hype vs. Reality

Key Application Areas

- **Indexing & Cell Parameters:** Neural networks predicting unit cell metrics ($a, b, c, \alpha, \beta, \gamma$) from raw peak positions.
- **Space Group & Phase Classification:** Automated classifiers parsing systematic absences or pattern features to propose symmetry and phase identity (e.g., public web-inference classifiers like *Galaxi*).
- **Generative Structure Solution:** Deep learning and diffusion models (e.g., *PXRDGen*, *Crystalyze*, *PXRDnet*) mapping 1D patterns directly to 3D atomic coordinates or candidate structures.

Robustness & Capabilities : Test by yourself !!!

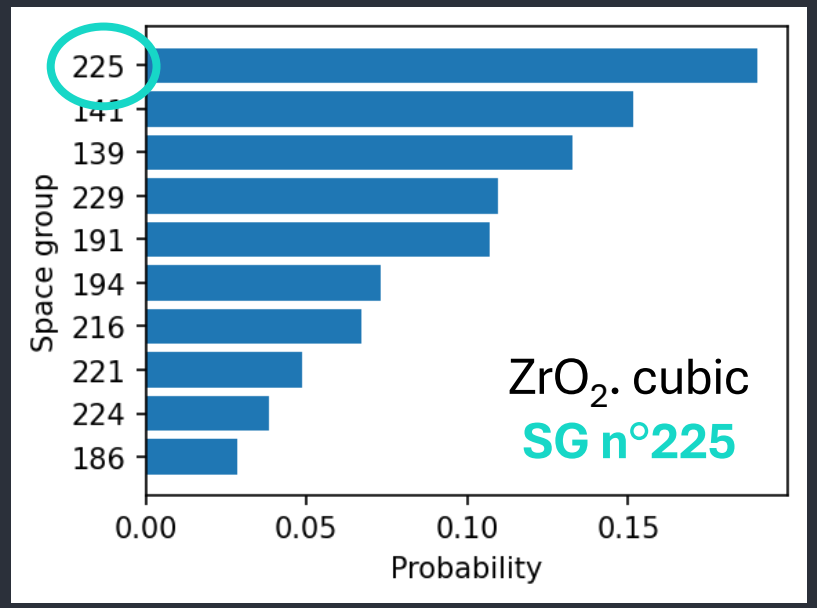
Limitations :

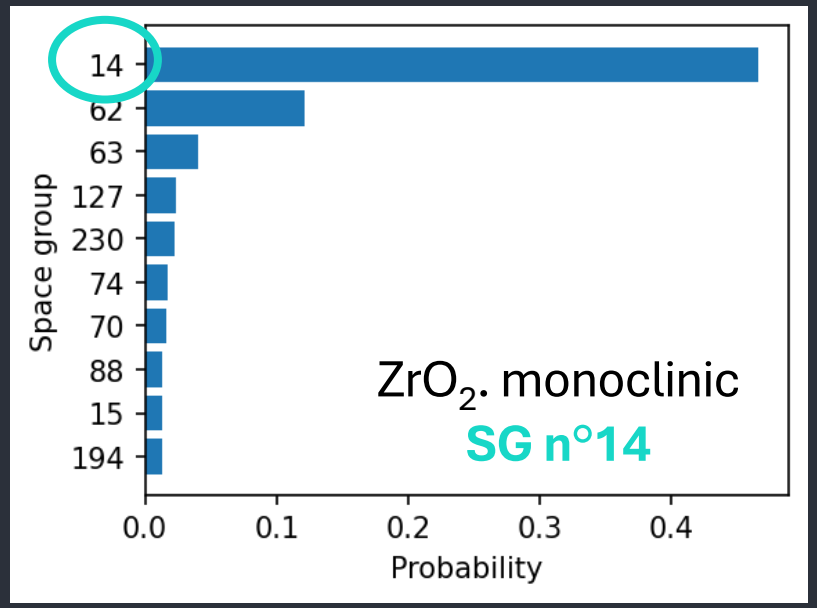
- **Information Collapse:** 1D powder diffraction loses 3D directional orientation, meaning AI models cannot universally solve arbitrarily large or disordered structures without chemical constraints.
- **Artifacts & Overlapping Peaks:** Severe preferred orientation, peak broadening, or unindexed impurities can heavily degrade deep learning predictions.

Phase identification with ML

- Cristalline phase identification

Symmetry and SG determination with scikit-learn using 10 diffraction reflections





- Machine Learning: fast (<3min on CPU)
- Deep Learning with Convolutional Neural Network
- Multivariate Curve Resolution with Alternative Least Square

Suzuki *et al.*, Nature, 2020

Detection of capillaries for automated alignment on CRISTAL beamline

Erik Elkaïm
Pierre Fertey
Arafat Noureddine

Positioning: Robot

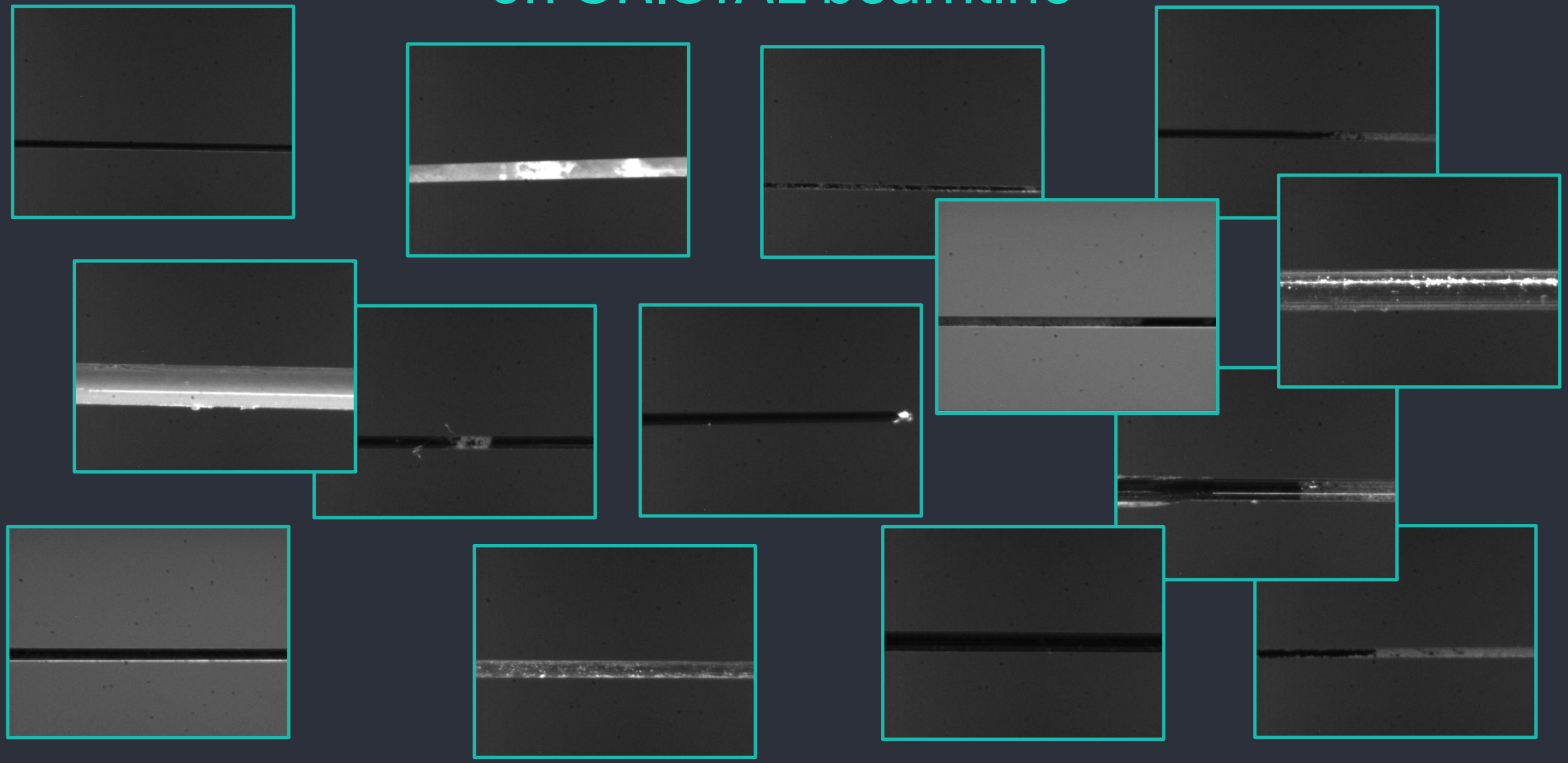
Alignment with incident beam: Deep Learning model



Capillaries on magnetic support

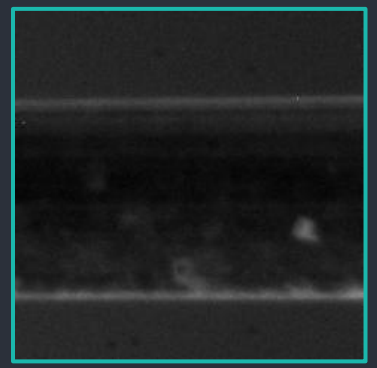
2-circles Diffractometer

Detection of capillaries for automated alignment on CRISTAL beamline

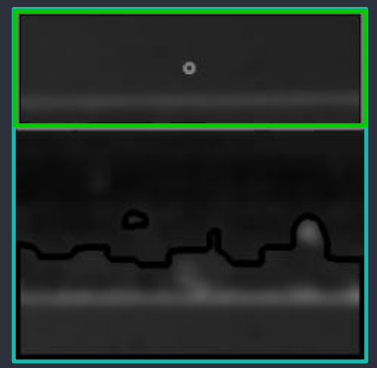


Detection of capillaries for automated alignment on CRISTAL beamline

Raw

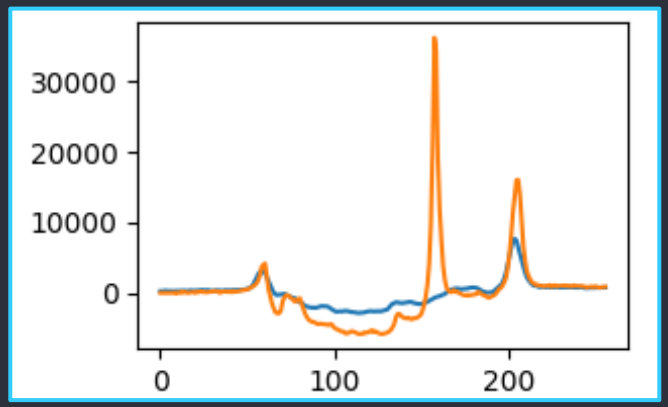


Classical Methods



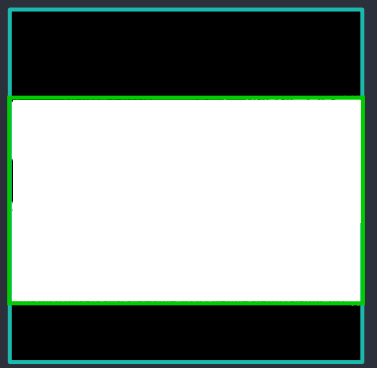
75%

Projection & Maximum Correlation



85%

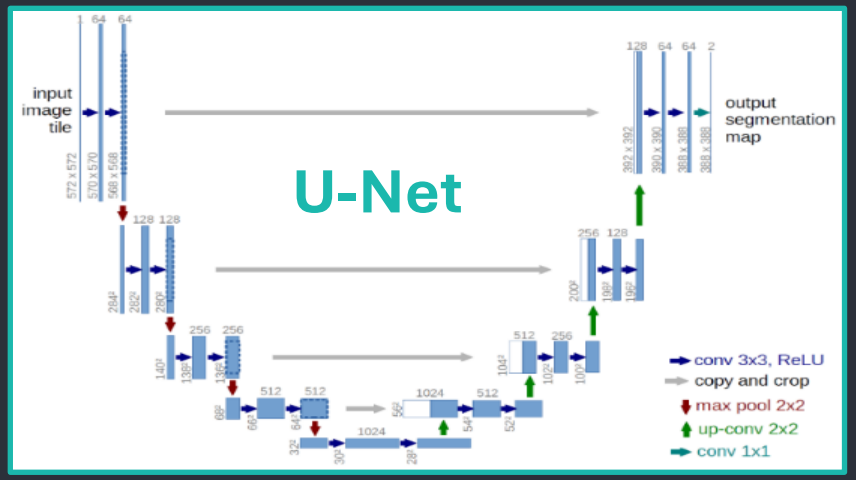
Classical Methods & Deep Learning



95%

Methods

1. Creation of a data-base: 50…150 images
2. Training and Validation of Neural Network
3. Deployment of the AI model
   - Device « ImageProcessor » in python
4. No issues reported over the past 2 years



Ronneberger *et al., 2015, arXiv.*

DiffractGPT

Solving a 3D structure from a 1D powder pattern... in a few seconds...

Accelerated "inverse design"

To be tested:

- *Pattern quality (overlapping, background noise, peak broadening)*
- *Inorganic vs. Organic vs. Hybrid*
- *Large cell parameter*
- *Partial occupancy & Disorder*

AI agent
for structure determination ?

Let’s see that together!!!

From Natural Language to Scientific Results





LLM – the semantic translator

- Role: understand natural langage, handle typos.
- Tool: convert to a structured object (JSON, pydantic schema)

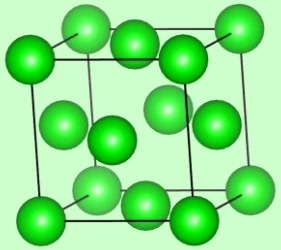
What it does NOT do:
It does not compute or make scientific decisions



The Decision Maker

- Role: deterministic « brain » and safety guardrails.
- Check: are the prerequisites met? (ex.: indexed before solving)

What it does NOT do:
No probabilistic decision, every rule is explicite

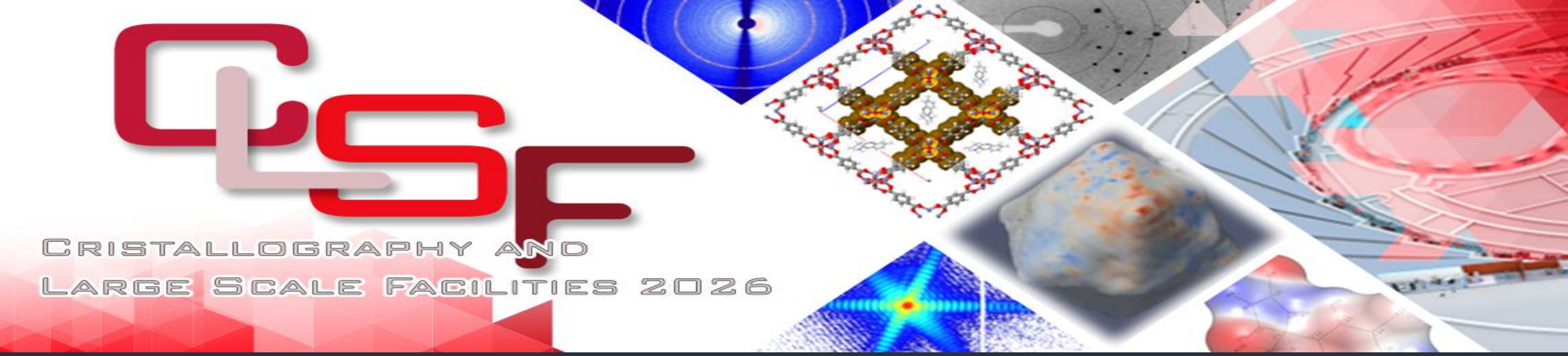


The Physics Engine - PyObjCryst

- Role: the « muscle ». Performs the real scientific computations.
- Actions: Peaks search, indexing, SG, Le Bail, Simulated Annealing

True Crystallography
Mathematics, geometry, physical law

Takeaways	The LLM is for confort. It lets you speak naturally	The Decision maker is for safety It enforces the right workflow	The physics engine delivers rigourous scientific process



Thank you !!! Questions ???

Hubert Chevreau - Synchrotron SOLEIL

hubert.chevreau@synchrotron-soleil.fr