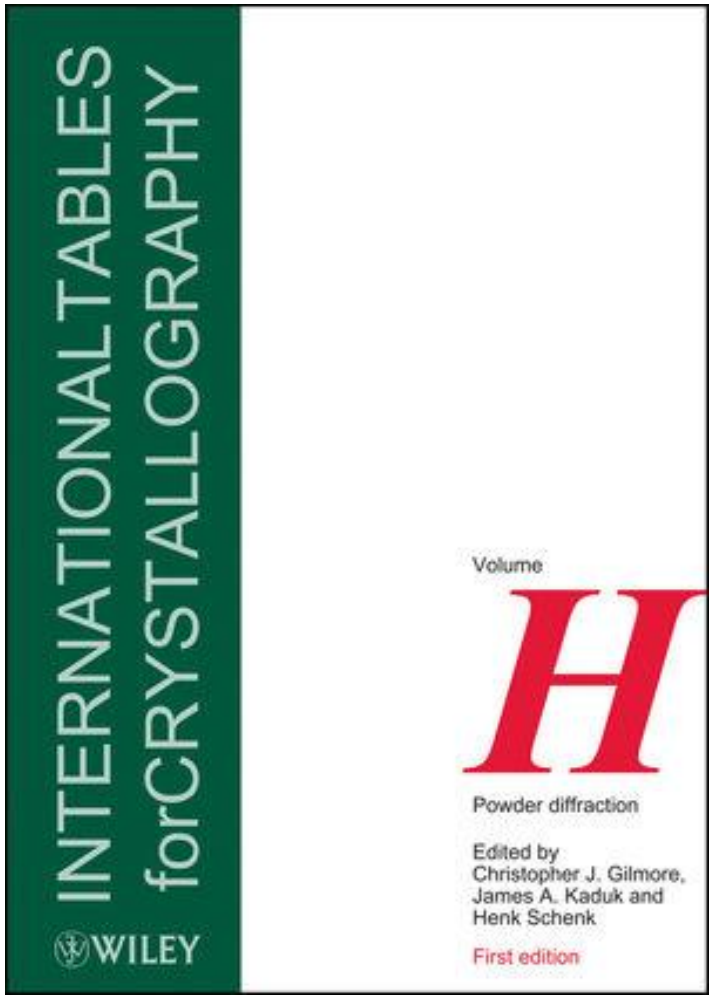

# Structure solution from powder diffraction: an overview. The software EXPO.

*Angela Altomare*

*angela.altomare@ic.cnr.it*

Institute of Cristallography - CNR, Bari, Italy


C. J. Gilmore, J. A. Kaduk, H. Schenk (Eds),
International Tables for Crystallography, Volume H, Powder Diffraction, Wiley, New York, 2019

A. Altomare, C. Cuocci, A. Moliterni and R. Rizzi, **Indexing a powder diffraction pattern, pp. 270-281.**

A. Altomare, C. Cuocci, A. Moliterni and R. Rizzi, **Solving crystal structures using reciprocal-space methods, pp. 395–413.**

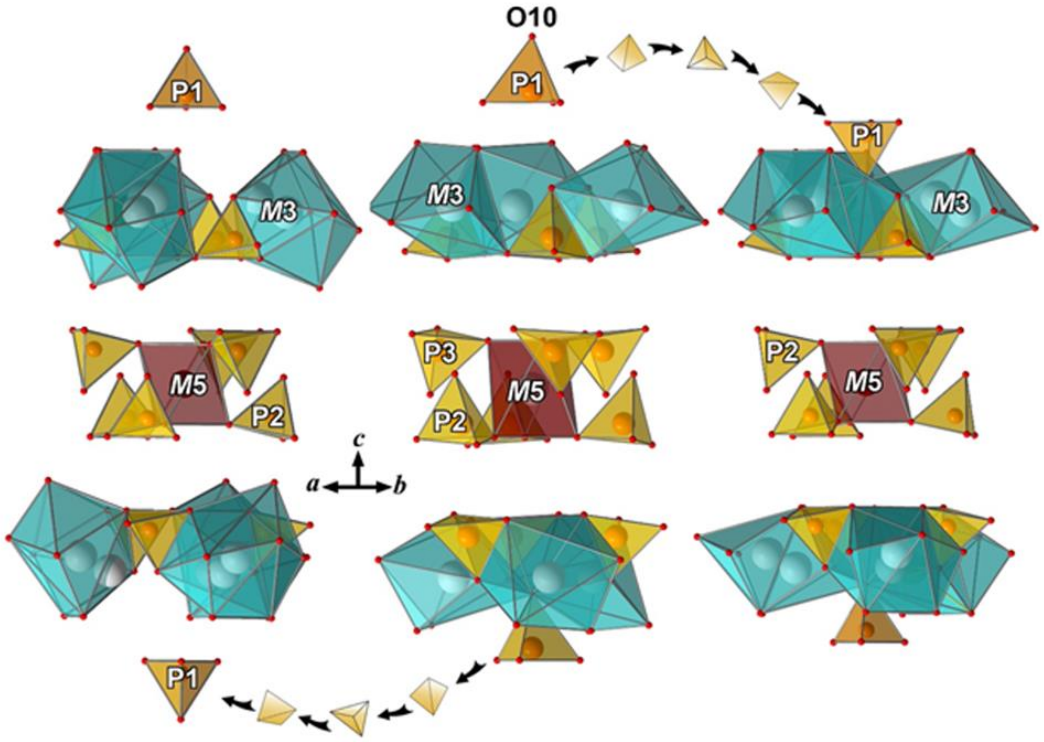
W. I. F. David, **Real-space methods for structure solution from powder-diffraction data: application to molecular structures, pp. 414–432.**

R. Černý and V. Favre-Nicolin, **Solving and refining inorganic structures, pp. 442–451.**

C. J. Gilmore, J. A. Kaduk and H. Schenk, **Survey of computer programs for powder diffraction, pp. 698-715.**

# Structure solution from powder diffraction

**Structure characterization at atomic level is required to discover Structure-Property relationships.**





CRYSTAL GROWTH & DESIGN

pubs.acs.org/crystal Article

**A Comprehensive Study of $Ca_9Tb(PO_4)_7$ and $Ca_9Ho(PO_4)_7$ Doped $\beta$-Tricalcium Phosphates: *Ab initio* Crystal Structure Solution, Rietveld Analysis, and Dielectric Properties**

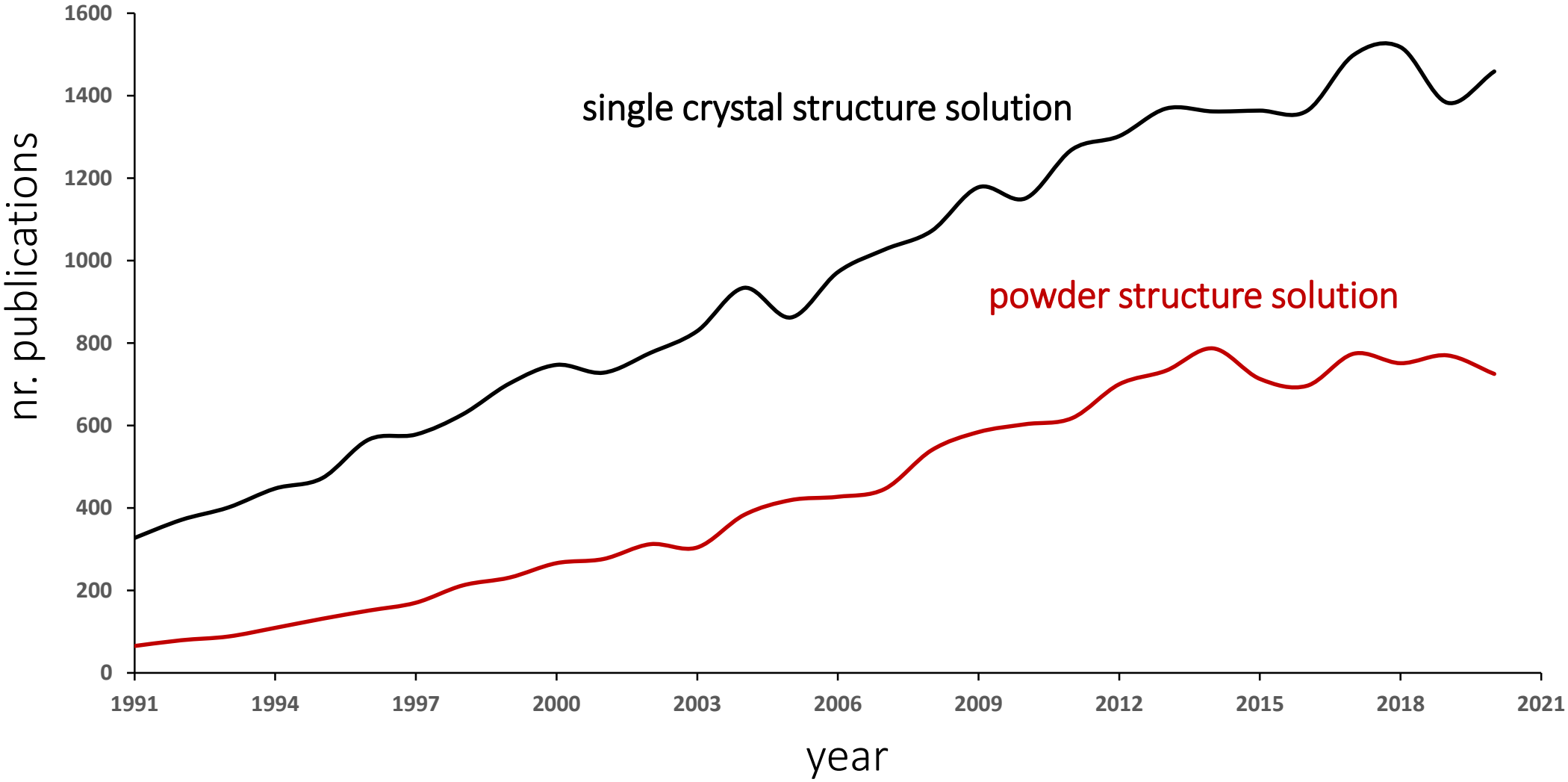
*Published as part of a Crystal Growth and Design virtual special issue on The Rietveld Refinement Method: Half of a Century Anniversary*

Rosanna Rizzi, Francesco Capitelli, Bogdan I. Lazoryak, Vladimir A. Morozov, Fabio Piccinelli, and Angela Altomare*

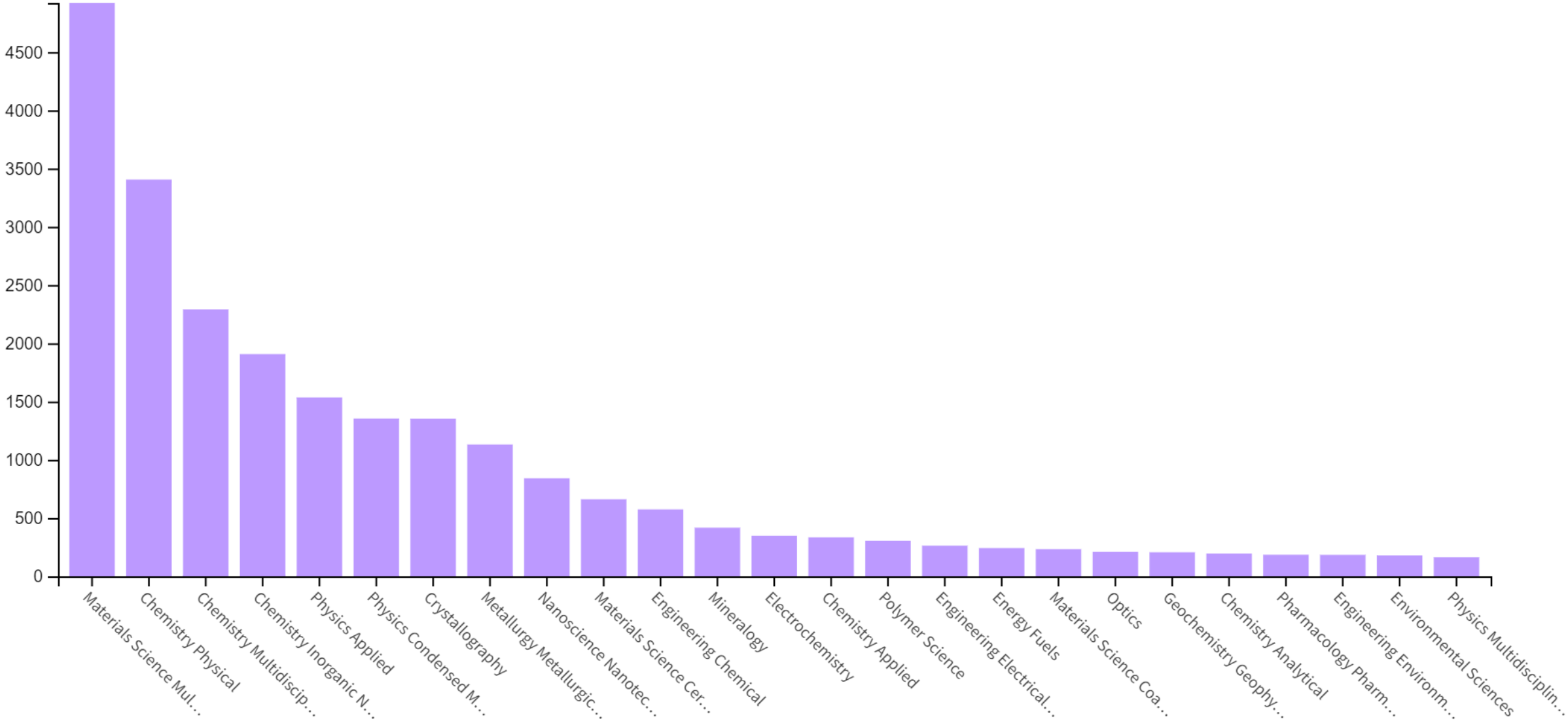
The change of orientation of the tethraedron $P_1O_4$ is responsible of the phase transition ferroelectric $\Longleftrightarrow$ paraelectric

X-ray powder diffraction (XRPD) is the most appropriate methodology for the determination of atom positions of an unknown microcrystalline material to clarify its chemical, physical and biological properties.

# Search on Web of Science

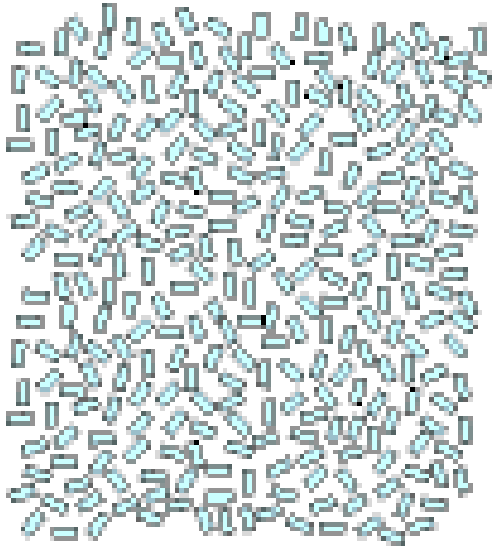


# Scientific areas (Web of Science)



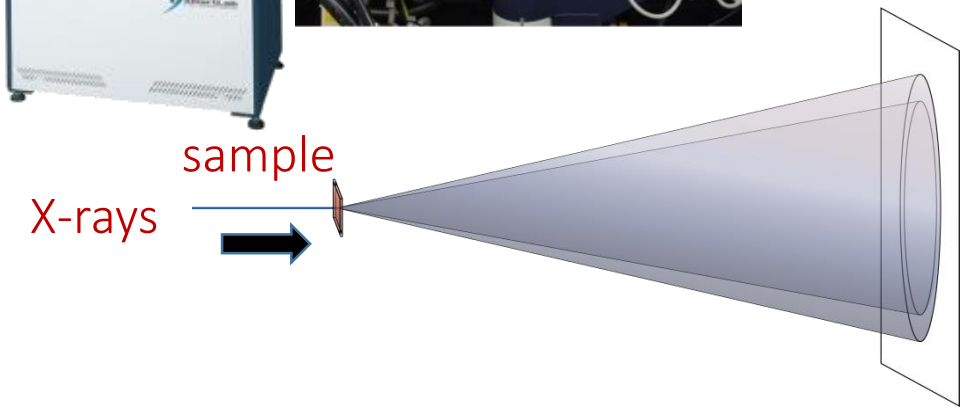
# Microcrystalline powder

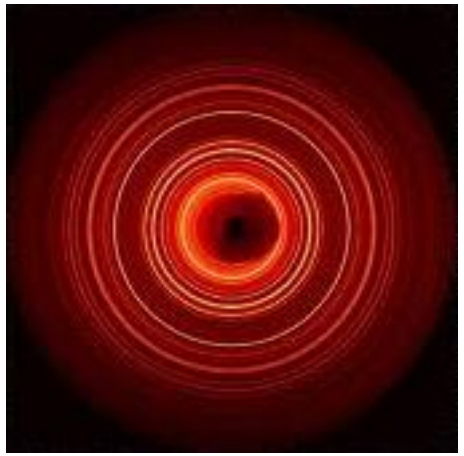
A microcrystalline powder consists of many microcrystallites (tens of micron size) randomly oriented in the space.

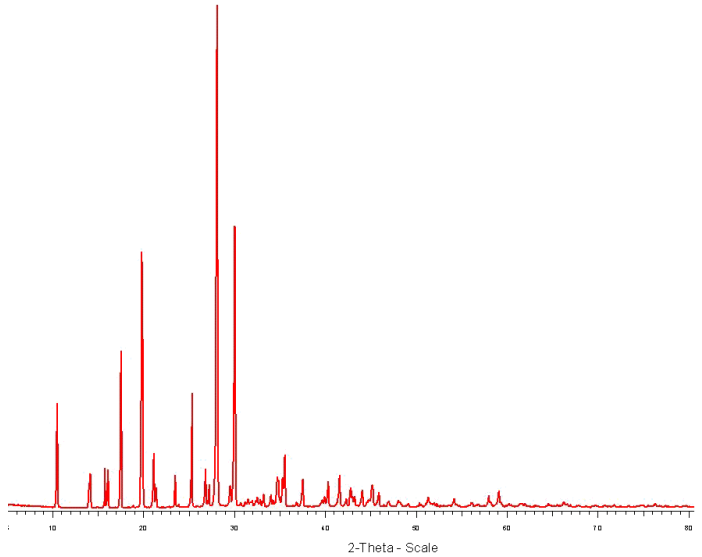


# Powder diffraction pattern





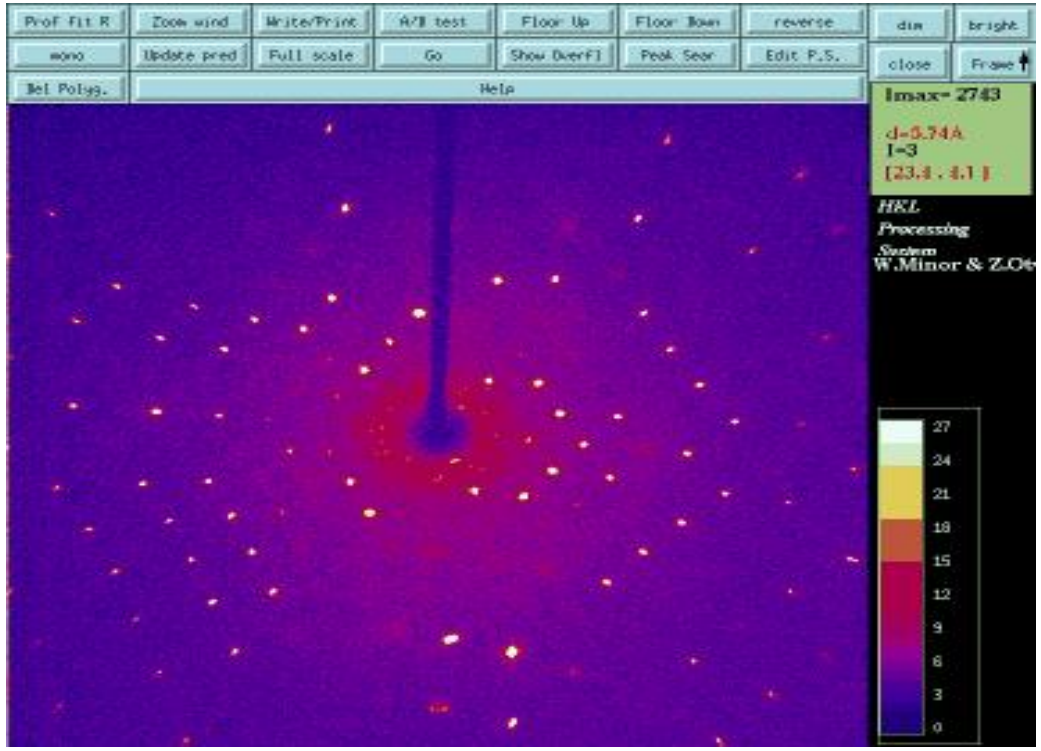



X-ray beams diffracted by a set of crystalline planes satisfying the Bragg law ($2d\sin\theta=\lambda$) lie on cones whose angular semi-aperture is $2\theta$.

When a detector intercepts diffraction cones, circles are observed, known as Debye rings.

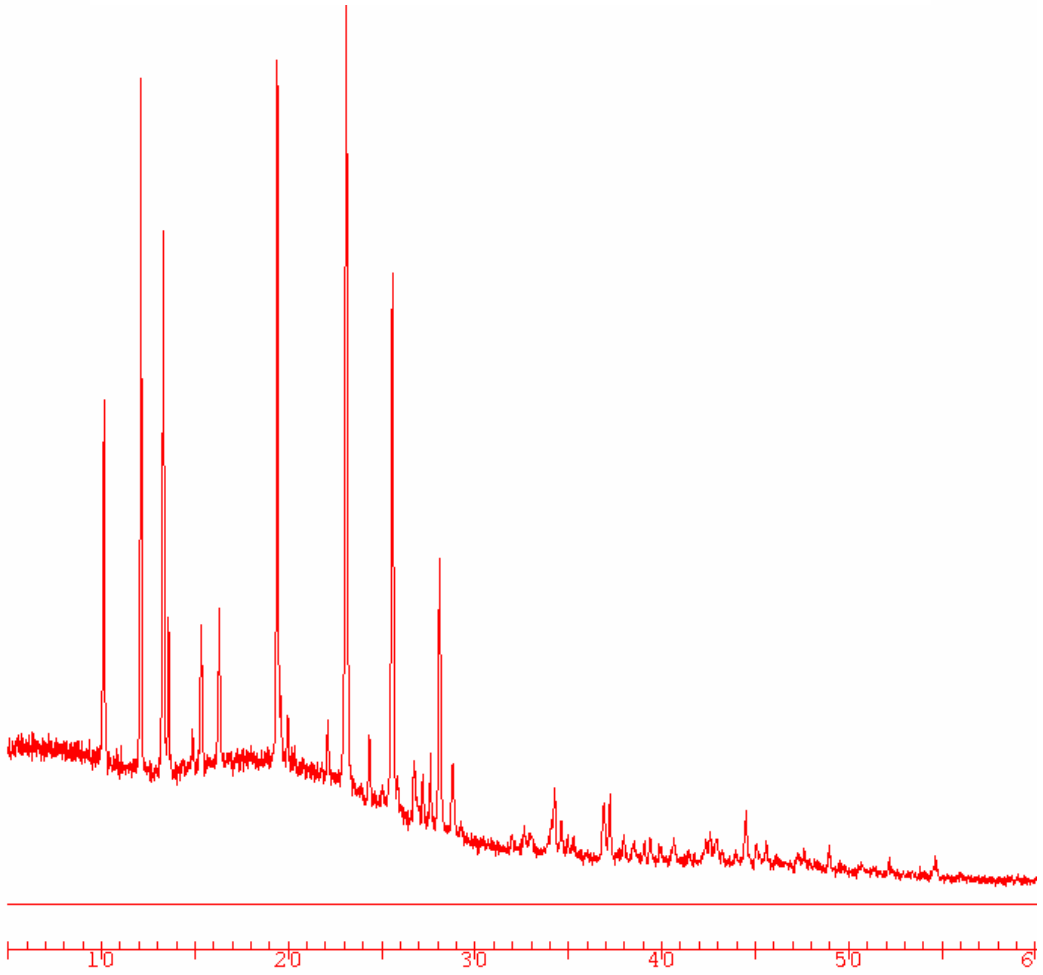
The diffraction pattern is one-dimensional, consisting of diffraction peaks as $2\theta$ angle increases.

# Single-crystal diffraction

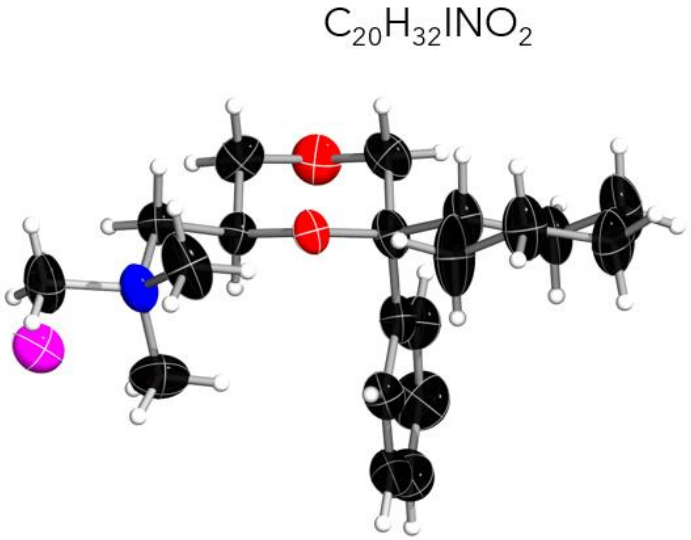



# Powder diffraction

$$f_j = f_j^0 \exp(-B_j \sin^2\theta/\lambda^2)$$



# Structure solution from single-crystal diffraction data




**Hydrogen atom** positions are located through the analysis of an **electron density map.**

**Atomic displacement parameters** are **anisotropically** refined.

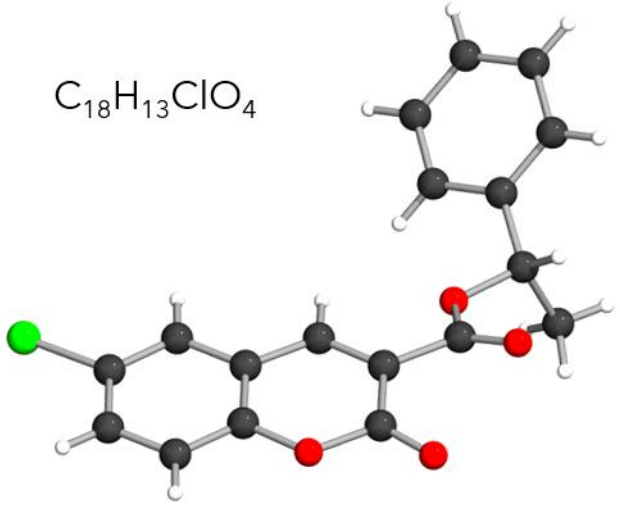
One single crystal sampled from a solid-state mass can be **not representative** of the material (polymorph, impurity?).

# Stucture solution from powder diffraction data

**Hydrogen atoms** are **geometrically** located.

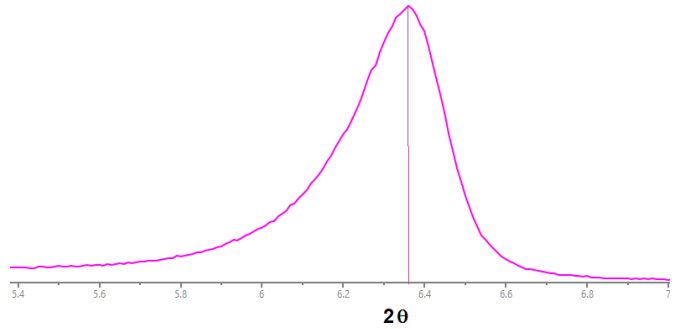
**Atomic displacement parameters** are **isotropically** refined.

**A full solid-state sample** is investigated.

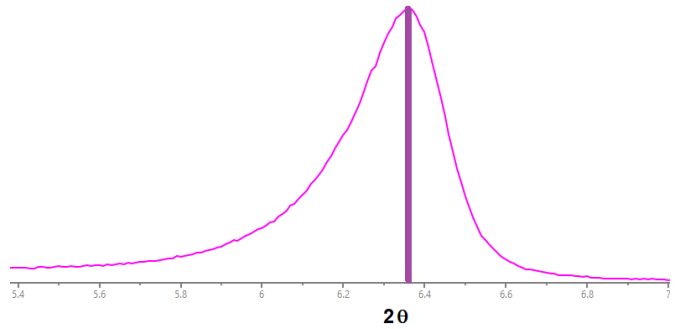


# Information from a powder diffraction pattern

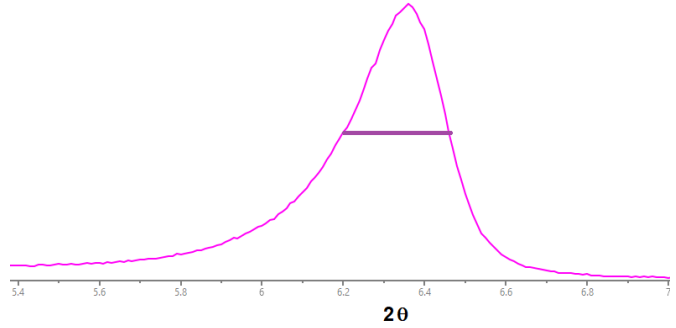
**2θ peak positions** depend on interplanar distances and the radiation wavelength, in particular, on unit cell parameters and the diffractometer alignment.




**Peak intensities** depend on the atomic content and atom positions in the unit cell.




**Peak widths** depend on the geometry adopted for data collection and crystallite size.



# Powder diffraction problems

## Peak overlap

- Systematic and/or casual (reduced with synchrotron radiation)
- Depends on symmetry and structure complexity
- Increases with the 2θ value

## Background

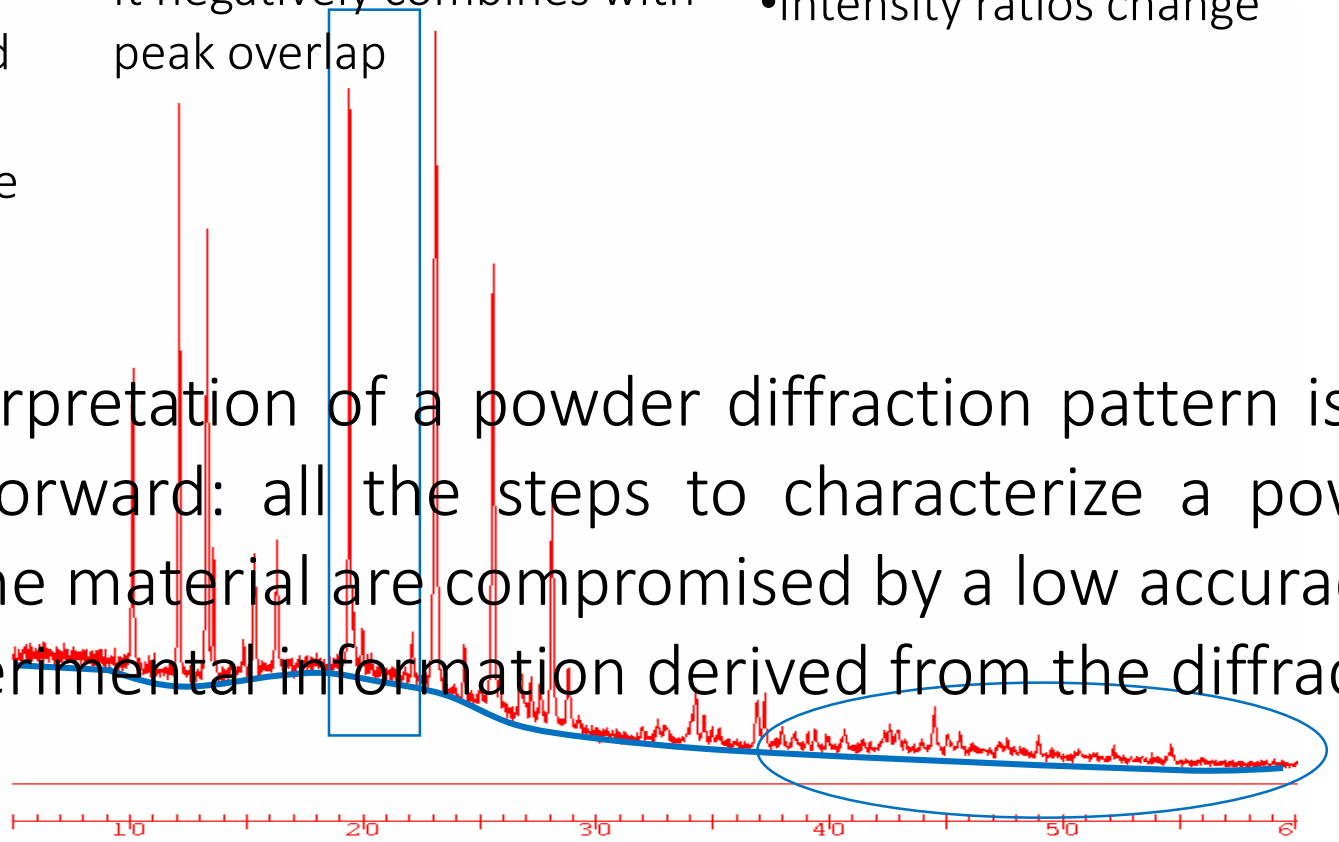
- Sometimes difficult to be evaluated
- It negatively combines with peak overlap

## Preferred orientation

- Crystallites can orient along preferred directions
- Intensity ratios change

## Experimental resolution

If the scattering power is low (light atoms) the experimental resolution is far from being atomic $d_{min}=\lambda/(2\sin\theta_{max})$




The interpretation of a powder diffraction pattern is not straightforward: all the steps to characterize a powder crystalline material are compromised by a low accuracy of the experimental information derived from the diffraction pattern.

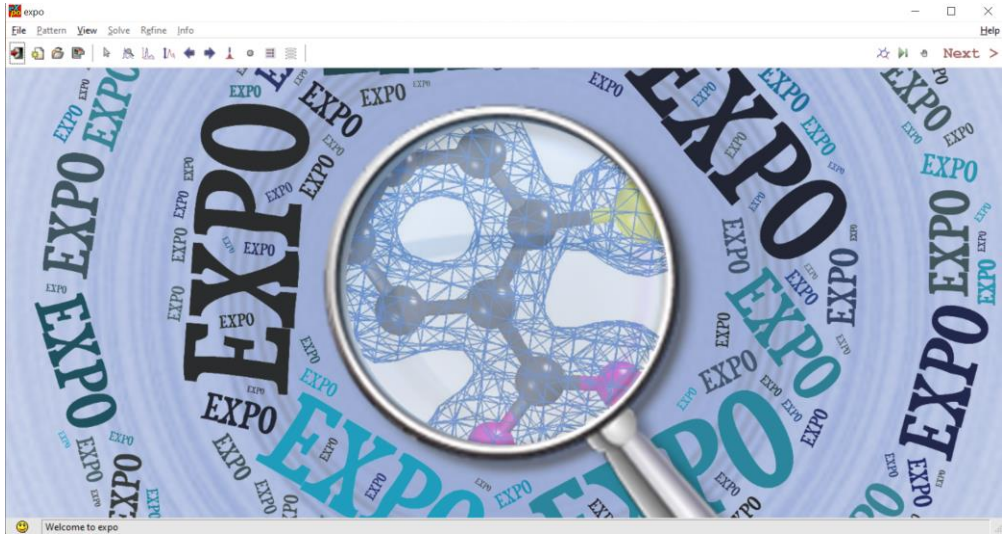
# Steps of the structure solution process

- **Determination of cell parameters**
- **Determination of space group**
  - Extraction of integrated intensities
  - Structure solution in the reciprocal space (Direct Methods, Patterson Methods, Maximum Entropy)

  Charge flipping, Molecular replacement
  - Structure solution in the direct space (Simulated Annealing, Genetic Algorithm, Monte Carlo, Grid Search)
    - Crystal structure refinement (Rietveld Method)

# EXPO software

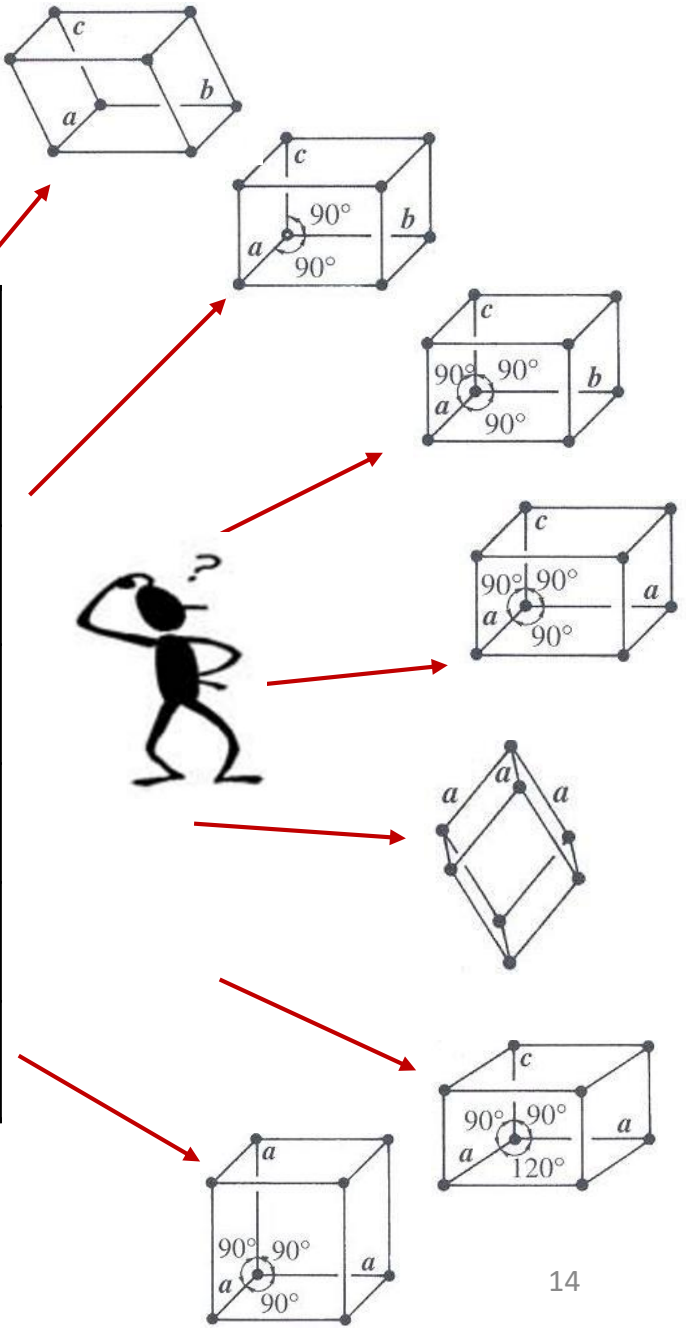
A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni, R. Rizzi, N. Corriero, A. Falcicchio, EXPO2013: a kit of tools for phasing crystal structures from powder data, J. Appl. Cryst. 46 (2013) 1231–1235.




- Determination of cell parameters
- Determination of space group
- Extraction of integrated intensities
- Structure solution in the reciprocal space (Direct Methods)
- Structure solution in the direct space (Simulated Annealing)
- Crystal structure refinement (Rietveld Method)

# Determination of cell parameters

| | | |
|---|---|---|
| Triclinic | $a \neq b \neq c$ | $\alpha \neq \beta \neq \gamma$ |
| Monoclinic | $a \neq b \neq c$ | $\alpha = \gamma = 90°, \beta \neq 90°$ |
| Orthorhombic | $a \neq b \neq c$ | $\alpha = \beta = \gamma = 90°$ |
| Tetragonal | $a = b \neq c$ | $\alpha = \beta = \gamma = 90°$ |
| Trigonal | $a = b = c$ | $\alpha = \beta = \gamma \neq 90°$ |
| Hexagonal | $a = b \neq c$ | $\alpha = \beta = 90°, \gamma = 120°$ |
| Cubic | $a = b = c$ | $\alpha = \beta = \gamma = 90°$ |



# Determination of cell parameters

$$d_h = \lambda/2\sin\theta_h$$

| Symmetry | $1/d_{hkl}$ |
|---|---|
| Cubic | $(h^2 + k^2 + l^2)^{1/2}\, a^*$ |
| Tetragonal | $[(h^2 + k^2)\, a^{*2} + l^2 c^{*2}]^{1/2}$ |
| Hexagonal | $[(h^2 + hk + k^2)\, a^{*2} + l^2 c^{*2}]^{1/2}$ |
| Orthorhombic | $[h^2 a^{*2} + k^2 b^{*2} + l^2 c^{*2}]^{1/2}$ |
| Monoclinic | $[h^2 a^{*2} + k^2 b^{*2} + l^2 c^{*2} + 2hl\, a^* c^* \cos\beta^*]^{1/2}$ |
| Triclinic | $[h^2 a^{*2} + k^2 b^{*2} + l^2 c^{*2} + 2hk a^* b^* \cos\gamma^* + 2hl\, a^* c^* \cos\beta^* + 2kl\, b^* c^* \cos\alpha^*]^{1/2}$ |

$d_h = f(a^*, b^*, c^*, \alpha^*, \beta^*, \gamma^*, \boldsymbol{h})$

$$\boldsymbol{a} = \frac{\boldsymbol{b}^* \wedge \boldsymbol{c}^*}{V^*}, \quad \boldsymbol{b} = \frac{\boldsymbol{c}^* \wedge \boldsymbol{a}^*}{V^*},$$

$$\boldsymbol{c} = \frac{\boldsymbol{a}^* \wedge \boldsymbol{b}^*}{V^*}$$

$$V^* = \boldsymbol{a}^* \cdot \boldsymbol{b}^* \wedge \boldsymbol{c}^* = 1/V$$

$d_h = f(a, b, c, \alpha, \beta, \gamma, \boldsymbol{h})$

# Determination of cell parameters

Determination in the direct space: Cell parameters ($a$, $b$, $c$, $\alpha$, $\beta$, $\gamma$) are randomly varied for reaching the best agreement between the experimental pattern and the pattern calculated from the current cell.

Methods: Grid Search, Monte Carlo, Genetic Algorithm, Simulated Annealing, …
Software: **AUTOX, EFLECH, Hmap, McMaille , SVD-Index TOPAS, FOX 34, GAIN, FIDEL-GO**

Determination in the reciprocal space: $h$, $k$, $l$ associated to $d_h$ corresponding to the smallest Bragg angle values are varied.

Methods: trial-and-error, zone-search
Software: **TREOR, DICVOL, ITO, X-CELL**

# Determination of cell parameters in the reciprocal space

**Peak search in the experimental powder diffraction pattern → interplanar distance values $d_h$ ($2\, d_h \sin\theta_h = \lambda$ ).**

## Fundamental Equation of Indexation

$$Q_{hkl} = h^2A_{11} + k^2A_{22} + l^2A_{33} + h\cdot kA_{12} + h\cdot lA_{13} + k\cdot lA_{23}$$

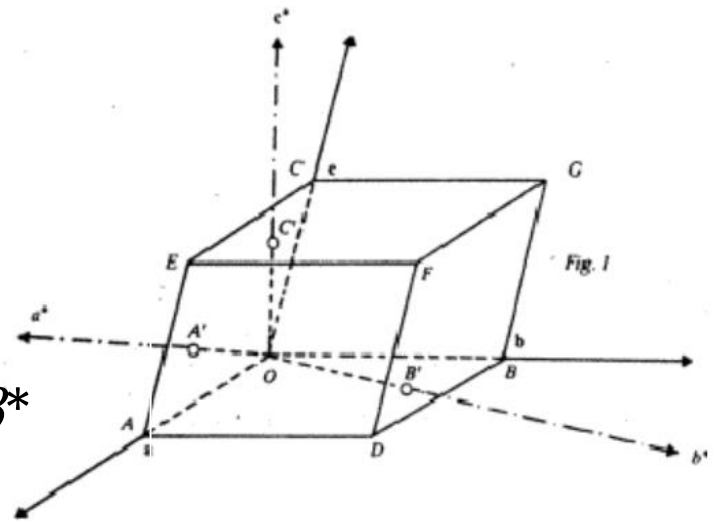
$$Q_{hkl} = 10^4/d^2_{hkl}$$

$$A_{11} = 10^4\cdot a^{*2} \quad A_{22} = 10^4\cdot b^{*2} \quad A_{33} = 10^4\cdot c^{*2}$$

$$A_{12} = 10^4\cdot 2\cdot a^*\cdot b^*\cdot\cos\gamma^* \quad A_{13} = 10^4\cdot 2\cdot a^*\cdot c^*\cdot\cos\beta^*$$

$$A_{23} = 10^4\cdot 2\cdot b^*\cdot c^*\cdot\cos\alpha^*$$

$\{A_{ij}\}$

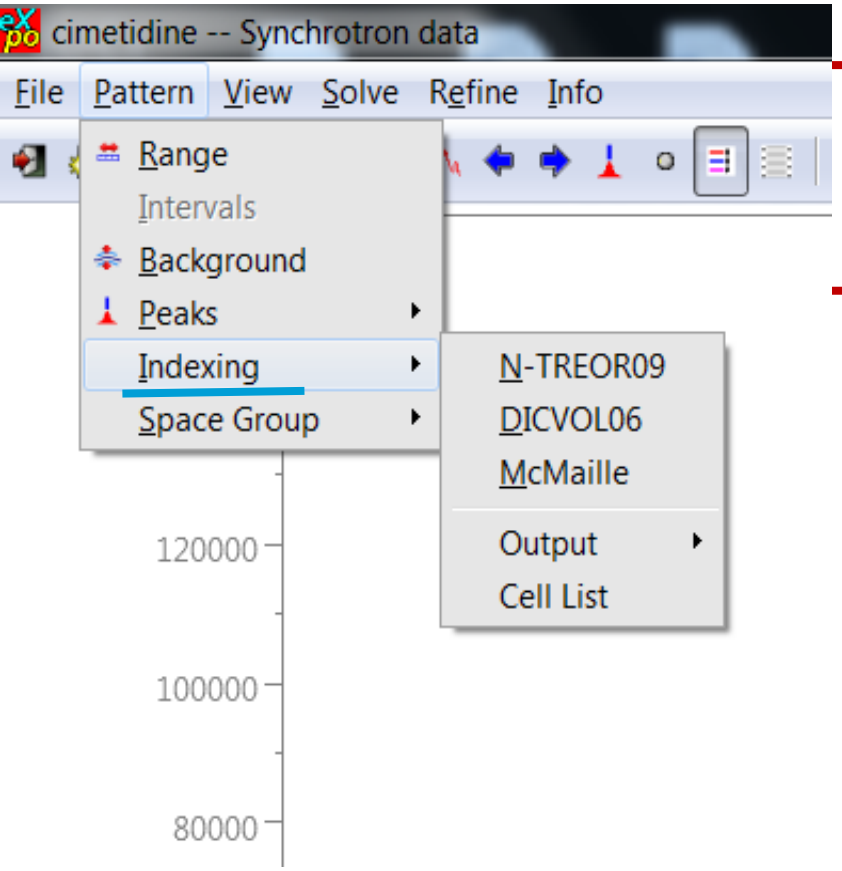


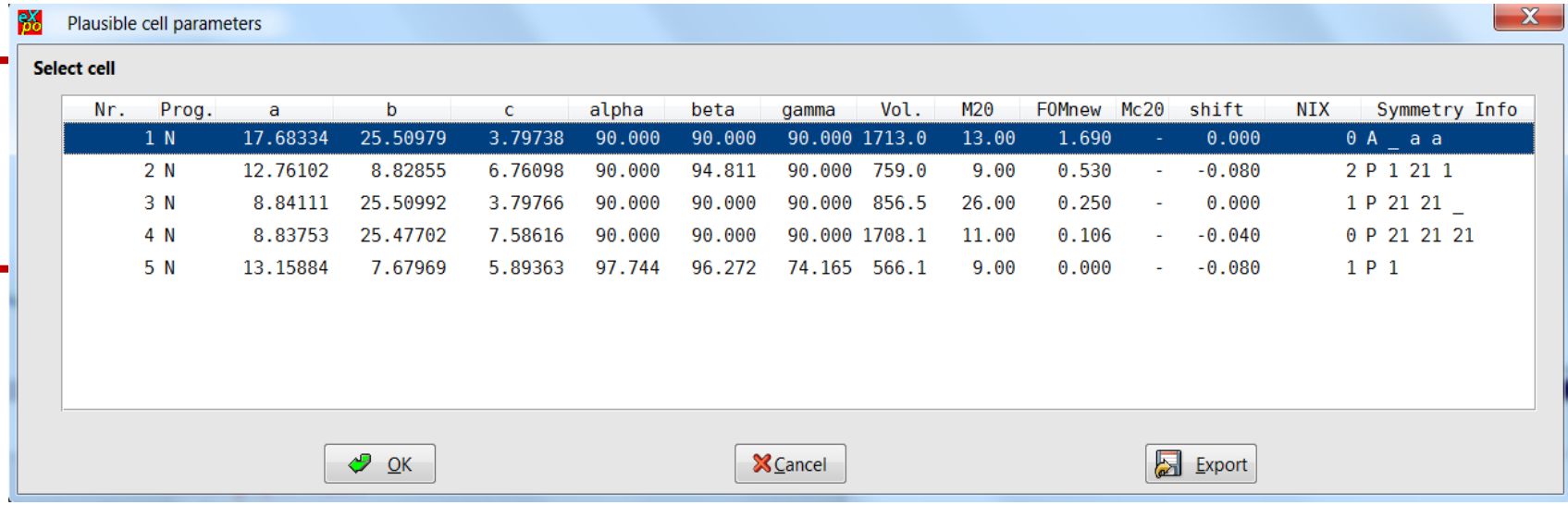

Reciprocal Lattice ($a^*$, $b^*$, $c^*$, $\alpha^*$, $\beta^*$, $\gamma^*$) → Direct Lattice ($a$, $b$, $c$, $\alpha$, $\beta$, $\gamma$)

Software provides lists of feasible cells compatible with experimental peak positions.

Figures of merit rank feasible cells.

# Determination of cell parameters: EXPO




| Nr. | Prog. | a | b | c | alpha | beta | gamma | Vol. | M20 | FOMnew | Mc20 | shift | NIX | Symmetry Info |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | N | 17.68334 | 25.50979 | 3.79738 | 90.000 | 90.000 | 90.000 | 1713.0 | 13.00 | 1.690 | - | 0.000 | 0 | A _ a a |
| 2 | N | 12.76102 | 8.82855 | 6.76098 | 90.000 | 94.811 | 90.000 | 759.0 | 9.00 | 0.530 | - | -0.080 | 2 | P 1 21 1 |
| 3 | N | 8.84111 | 25.50992 | 3.79766 | 90.000 | 90.000 | 90.000 | 856.5 | 26.00 | 0.250 | - | 0.000 | 1 | P 21 21 _ |
| 4 | N | 8.83753 | 25.47702 | 7.58616 | 90.000 | 90.000 | 90.000 | 1708.1 | 11.00 | 0.106 | - | -0.040 | 0 | P 21 21 21 |
| 5 | N | 13.15884 | 7.67969 | 5.89363 | 97.744 | 96.272 | 74.165 | 566.1 | 9.00 | 0.000 | - | -0.080 | 1 | P 1 |

2θ-zero shift
Impurity peaks
Low crystallinity
Poor peak resolution

# Determination of cell parameters: EXPO

## Remarks

Do not accept unindexed peaks, unless you can explain them. Each observed peak must correspond to a calculated reflection.

The correctness of the cell is generally confirmed by the solution process.

In case of failure, check the peak search result and, if necessary, modify the selection of peaks.

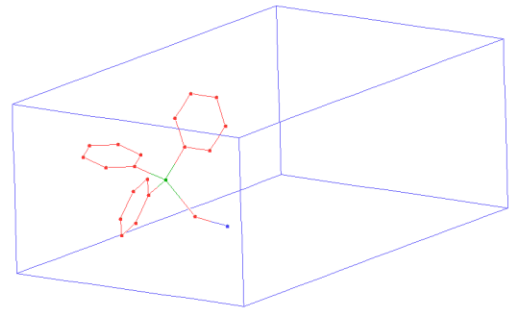
In case of failure, try non-default indexing strategies.

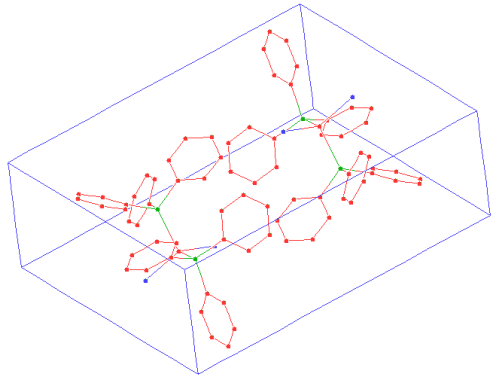
It can be useful to use more than one indexing software.

Data quality is a very important condition.

# Determination of space group

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | P 1 | 26 | P m c 21 | 51 | P m m a | 76 | P 41 | 101 | P 42 c m | 151 | P 31 1 2 | 176 | P 63/m | 201 | P n -3 | 226 | F m -3 c |
| 2 | P -1 | 27 | P c c 2 | 52 | P n n a | 77 | P 42 | 102 | P 42 n m | 152 | P 31 2 1 | 177 | P 6 2 2 | 202 | F m -3 | 227 | F d -3 m |
| 3 | P 1 2 1 | 28 | P m a 2 | 53 | P m n a | 78 | P 43 | 103 | P 4 c c | 153 | P 32 1 2 | 178 | P 61 2 2 | 203 | F d -3 | 228 | F d -3 c |
| 4 | P 1 21 1 | 29 | P c a 21 | 54 | P c c a | 79 | I 4 | 104 | P 4 n c | 154 | P 32 2 1 | 179 | P 65 2 2 | 204 | I m -3 | 229 | I m -3 m |
| 5 | C 1 2 1 | 30 | P n c 2 | 55 | P b a m | 80 | I 41 | 105 | P 42 m c | 155 | R 3 2 | 180 | P 62 2 2 | 205 | P a -3 | 230 | I a -3 d |
| 6 | P 1 m 1 | 31 | P m n 21 | 56 | P c c n | 81 | P -4 | 106 | P 42 b c | 156 | P 3 m 1 | 181 | P 64 2 2 | 206 | I a -3 | | |
| 7 | P 1 c 1 | 32 | P b a 2 | 57 | P b c m | 82 | I -4 | 107 | I 4 m m | 157 | P 3 1 m | 182 | P 63 2 2 | 207 | P 4 3 2 | | |
| 8 | C 1 m 1 | 33 | P n a 21 | 58 | P n n m | 83 | P 4/m | 108 | I 4 c m | 158 | P 3 c 1 | 183 | P 6 m m | 208 | P 42 3 2 | | |
| 9 | C 1 c 1 | 34 | P n n 2 | 59 | P m m n | 84 | P 42/m | 109 | I 41 m d | 159 | P 3 1 c | 184 | P 6 c c | 209 | F 4 3 2 | | |
| 10 | P 1 2/m 1 | 35 | C m m 2 | 60 | P b c n | 85 | P 4/n | 110 | I 41 c d | 160 | R 3 m | 185 | P 63 c m | 210 | F 41 3 2 | | |
| 11 | P 1 21/m 1 | 36 | C m c 21 | 61 | P b c a | 86 | P 42/n | 111 | P -4 2 m | 161 | R 3 c | 186 | P 63 m c | 211 | I 4 3 2 | | |
| 12 | C 1 2/m 1 | 37 | C c c 2 | 62 | P n m a | 87 | I 4/m | 112 | P -4 2 c | 162 | P -3 1 m | 187 | P -6 m 2 | 212 | P 43 3 2 | | |
| 13 | P 1 2/c 1 | 38 | A m m 2 | 63 | C m c m | 88 | I 41/a | 113 | P -4 21 m | 163 | P -3 1 c | 188 | P -6 c 2 | 213 | P 41 3 2 | | |
| 14 | P 1 21/c 1 | 39 | A b m 2 | 64 | C m c a | 89 | P 4 2 2 | 114 | P -4 21 c | 164 | P -3 m 1 | 189 | P -6 2 m | 214 | I 41 3 2 | | |
| 15 | C 1 2/c 1 | 40 | A m a 2 | 65 | C m m m | 90 | P 4 21 2 | 115 | P -4 m 2 | 165 | P -3 c 1 | 190 | P -6 2 c | 215 | P -4 3 m | | |
| 16 | P 2 2 2 | 41 | A b a 2 | 66 | C c c m | 91 | P 41 2 2 | 116 | P -4 c 2 | 166 | R -3 m | 191 | P 6/m m m | 216 | F -4 3 m | | |
| 17 | P 2 2 21 | 42 | F m m 2 | 67 | C m m a | 92 | P 41 21 2 | 117 | P -4 b 2 | 167 | R -3 c | 192 | P 6/m c c | 217 | I -4 3 m | | |
| 18 | P 21 21 2 | 43 | F d d 2 | 68 | C c c a | 93 | P 42 2 2 | 118 | P -4 n 2 | 168 | P 6 | 193 | P 63/m c m | 218 | P -4 3 n | | |
| 19 | P 21 21 21 | 44 | I m m 2 | 69 | F m m m | 94 | P 42 21 2 | 119 | I -4 m 2 | 169 | P 61 | 194 | P 63/m m c | 219 | F -4 3 c | | |
| 20 | C 2 2 21 | 45 | I b a 2 | 70 | F d d d | 95 | P 43 2 2 | 120 | I -4 c 2 | 170 | P 65 | 195 | P 2 3 | 220 | I -4 3 d | | |
| 21 | C 2 2 2 | 46 | I m a 2 | 71 | I m m m | 96 | P 43 21 2 | 121 | I -4 2 m | 171 | P 62 | 196 | F 2 3 | 221 | P m -3 m | | |
| 22 | F 2 2 2 | 47 | P m m m | 72 | I b a m | 97 | I 4 2 2 | 122 | I -4 2 d | 172 | P 64 | 197 | I 2 3 | 222 | P n -3 n | | |
| 23 | I 2 2 2 | 48 | P n n n | 73 | I b c a | 98 | I 41 2 2 | 123 | P 4/m m m | 173 | P 63 | 198 | P 21 3 | 223 | P m -3 n | | |
| 24 | I 21 21 21 | 49 | P c c m | 74 | I m m a | 99 | P 4 m m | 124 | P 4/m c c | 174 | P -6 | 199 | I 21 3 | 224 | P n -3 m | | |
| 25 | P m m 2 | 50 | P b a n | 75 | P 4 | 100 | P 4 b m | 125 | P 4/n b m | 175 | P 6/m | 200 | P m -3 | 225 | F m -3 m | | |





# Determination of space group

**The presence of symmetry operators has consequences in the reciprocal space.**

$P2_1/c$

Class of reflections (0 *k* 0), *k* = 2n +1

and

Class of reflections (*h* 0 *l*), *l* = 2n +1

are systematically absent (reflection intensities are zero) due to the presence of

$2_1$ axis || **b**

*c* glide ⊥ **b**

The statistical-probabilistical analysis of experimental diffraction intensities provides information on systematically absent reflections (extinctions).

# Determination of space group

Integrated intensities are extracted from the experimental pattern assuming the maximum symmetry of the Laue class compatible with the identified crystal system.

| Crystal System | Laue Classes | |
|---|---|---|
| Triclinic | $\bar{1}$ | |
| Monoclinic | $2/m$ | |
| Orthorhombic | $mmm$ | |
| Tetragonal | $4/m$ | $4/mmm$ |
| Trigonal | $\bar{3}$ | $\bar{3}m$ |
| Hexagonal | $6/m$ | $6/mmm$ |
| Cubic | $m\bar{3}$ | $m\bar{3}m$ |

**Statistical-probabilistical analysis of intensities → probability corresponding to each extinction group compatible with the crystal system.**

The probability of each extinction group is calculated by the product of $p$ probabilities corresponding to each symmetry element of the extinction group.

$p(2_1)=1-\langle z_{0k0}\rangle_{k=2n+1}$, ($\langle z\rangle$ small, probability $\cong 1$)

$\langle z_{0k0}\rangle_{k=2n+1}$ is the average of reflection intensities of type $(0\ k\ 0)$, $k=2n+1$.

Extinction groups are ranked according to calculated probability values.
The space group is derived.

# Determination of space group

The smaller the <z> value for a class of reflections, the closer to 1 the probability of the symmetry element corresponding to the extinction of that class.

**Monoclinic**

| Space group | Extinction group |
|---|---|
| P 21 | P 1 21 1 |
| P 21/m | P 1 21 1 |
| P 2 | P 1 - 1 |
| P 2/m | P 1 - 1 |
| P m | P 1 - 1 |
| P 21/c | P 1 21/c 1 |
| P 2/c | P 1 c 1 |
| P c | P 1 c 1 |
| P 21/n | P 1 21/n 1 |
| P 2/n | P 1 n 1 |
| P n | P 1 n 1 |

**Orthorhombic**

| Space group | Extinction group |
|---|---|
| P 21 c a | P - c a |
| P m c a | P - c a |
| P 21 m a | P - - a |
| P m m a | P - - a |
| P m 2 a | P - - a |
| P 2 n a | P - n a |
| P m n a | P - n a |
| P c c a | P c c a |
| P m c 21 | P - c - |
| P m c m | P - c - |
| P 2 c m | P - c - |

Nr. of extinction groups for each crystal system:

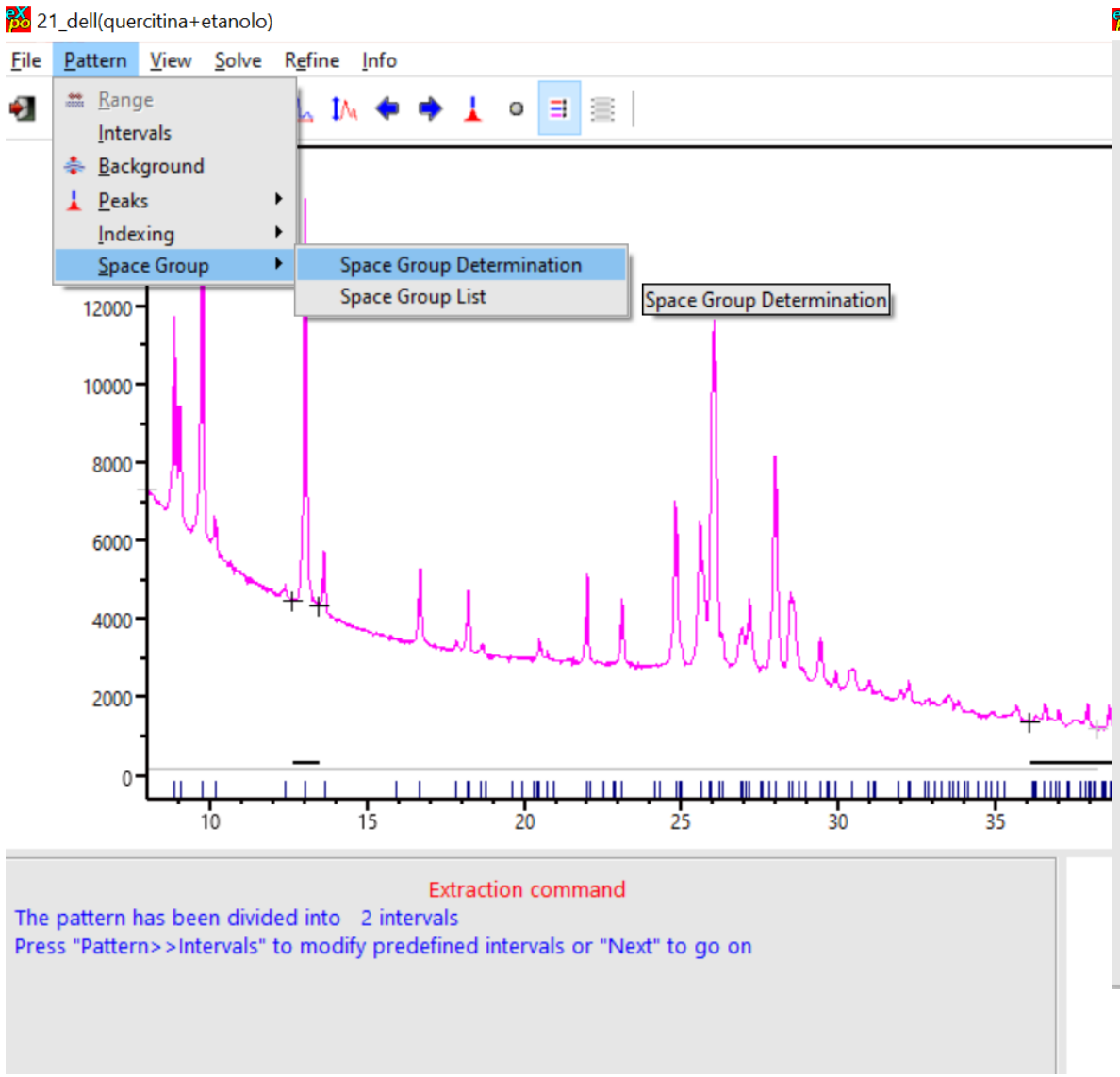
14 Monoclinic
111 Orthorhombic
31 Tetragonal
12 Trigonal-Hexagonal (only hexagonal axes),
18 Cubic

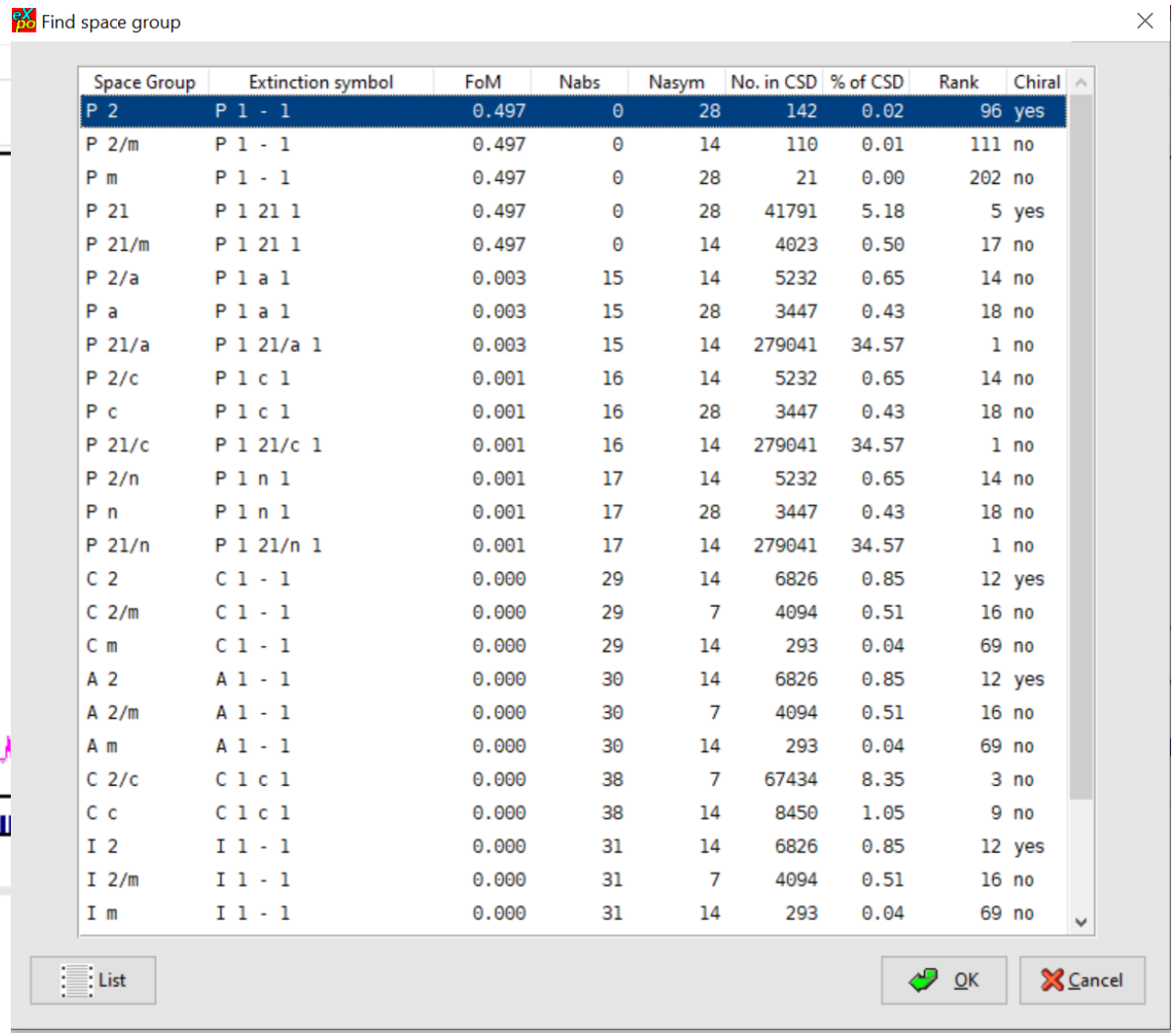
Extinction groups do not unambiguously define space groups.

Solution processes are executed considering each possible space group.

Already solved structures similar to the structure under study can help to selecting the correct space group.

# Determination of space group: EXPO




| Space Group | Extinction symbol | FoM | Nabs | Nasym | No. in CSD | % of CSD | Rank | Chiral |
|---|---|---|---|---|---|---|---|---|
| P 2 | P 1 - 1 | 0.497 | 0 | 28 | 142 | 0.02 | 96 | yes |
| P 2/m | P 1 - 1 | 0.497 | 0 | 14 | 110 | 0.01 | 111 | no |
| P m | P 1 - 1 | 0.497 | 0 | 28 | 21 | 0.00 | 202 | no |
| P 21 | P 1 21 1 | 0.497 | 0 | 28 | 41791 | 5.18 | 5 | yes |
| P 21/m | P 1 21 1 | 0.497 | 0 | 14 | 4023 | 0.50 | 17 | no |
| P 2/a | P 1 a 1 | 0.003 | 15 | 14 | 5232 | 0.65 | 14 | no |
| P a | P 1 a 1 | 0.003 | 15 | 28 | 3447 | 0.43 | 18 | no |
| P 21/a | P 1 21/a 1 | 0.003 | 15 | 14 | 279041 | 34.57 | 1 | no |
| P 2/c | P 1 c 1 | 0.001 | 16 | 14 | 5232 | 0.65 | 14 | no |
| P c | P 1 c 1 | 0.001 | 16 | 28 | 3447 | 0.43 | 18 | no |
| P 21/c | P 1 21/c 1 | 0.001 | 16 | 14 | 279041 | 34.57 | 1 | no |
| P 2/n | P 1 n 1 | 0.001 | 17 | 14 | 5232 | 0.65 | 14 | no |
| P n | P 1 n 1 | 0.001 | 17 | 28 | 3447 | 0.43 | 18 | no |
| P 21/n | P 1 21/n 1 | 0.001 | 17 | 14 | 279041 | 34.57 | 1 | no |
| C 2 | C 1 - 1 | 0.000 | 29 | 14 | 6826 | 0.85 | 12 | yes |
| C 2/m | C 1 - 1 | 0.000 | 29 | 7 | 4094 | 0.51 | 16 | no |
| C m | C 1 - 1 | 0.000 | 29 | 14 | 293 | 0.04 | 69 | no |
| A 2 | A 1 - 1 | 0.000 | 30 | 14 | 6826 | 0.85 | 12 | yes |
| A 2/m | A 1 - 1 | 0.000 | 30 | 7 | 4094 | 0.51 | 16 | no |
| A m | A 1 - 1 | 0.000 | 30 | 14 | 293 | 0.04 | 69 | no |
| C 2/c | C 1 c 1 | 0.000 | 38 | 7 | 67434 | 8.35 | 3 | no |
| C c | C 1 c 1 | 0.000 | 38 | 14 | 8450 | 1.05 | 9 | no |
| I 2 | I 1 - 1 | 0.000 | 31 | 14 | 6826 | 0.85 | 12 | yes |
| I 2/m | I 1 - 1 | 0.000 | 31 | 7 | 4094 | 0.51 | 16 | no |
| I m | I 1 - 1 | 0.000 | 31 | 14 | 293 | 0.04 | 69 | no |

# Unit cell content

Unit cell content= Z · molecular formula

$$Z = \frac{V_{cell}}{V_{mol}}$$

$V_{mol}$ = Volume of molecule

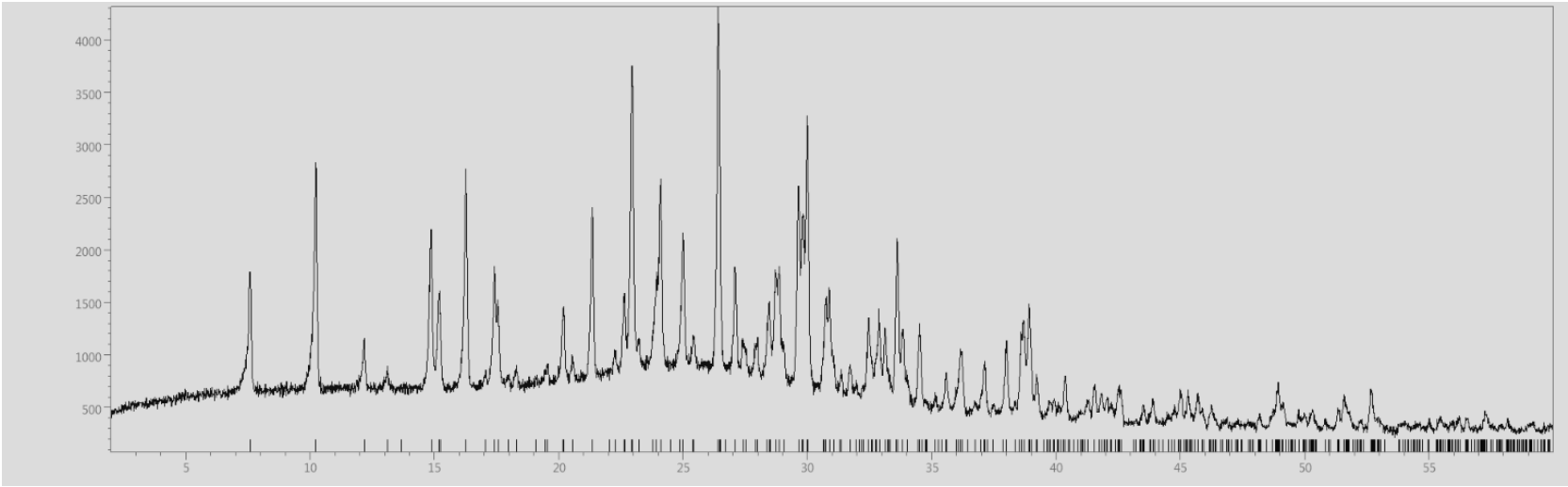
Rule 18 Å$^3$ : $V_{mol} = M \cdot 18$

M = number of non-H atoms

$$Z = \frac{V_{cell}}{18 \cdot M}$$

# Peak overlap

## Number of independent observations << Number of observations




$C_{20}H_{24}Cl\,N_3\,O^{2+}\cdot 2Cl^{-}\cdot 2H_2O$

cell 7.84 26.99 10.81 90 92.96 90
space group $P\,2_1/c$

```
**************************************************

  Independent observations calculation routine

  Number of independent observations =     217

  Independent observations percentage =   51.89%

**************************************************
```

217/29 <8

```
*****************************************************
*                Pattern information                *
*                                                   *
*          2-Theta min     =   2.0000               *
*          2-Theta max     =  59.9800               *
*          Step            =   0.0100               *
*          Ncounts         =     5799               *
*                                                   *
*       Generated reflections information           *
*                                                   *
*            rho min       =  0.00010               *
*            rho max       =  0.07807               *
*            d (resolution)=  1.78946               *
*            reflections nr.=     418               *
*            h k l (max)   =    4  14   6           *
*****************************************************

Number of atoms in asymmetric unit = 29.00
```

RESEARCH PAPERS

*J. Appl. Cryst.* (1995). **28**, 738-744
https://doi.org/10.1107/S0021889895009757

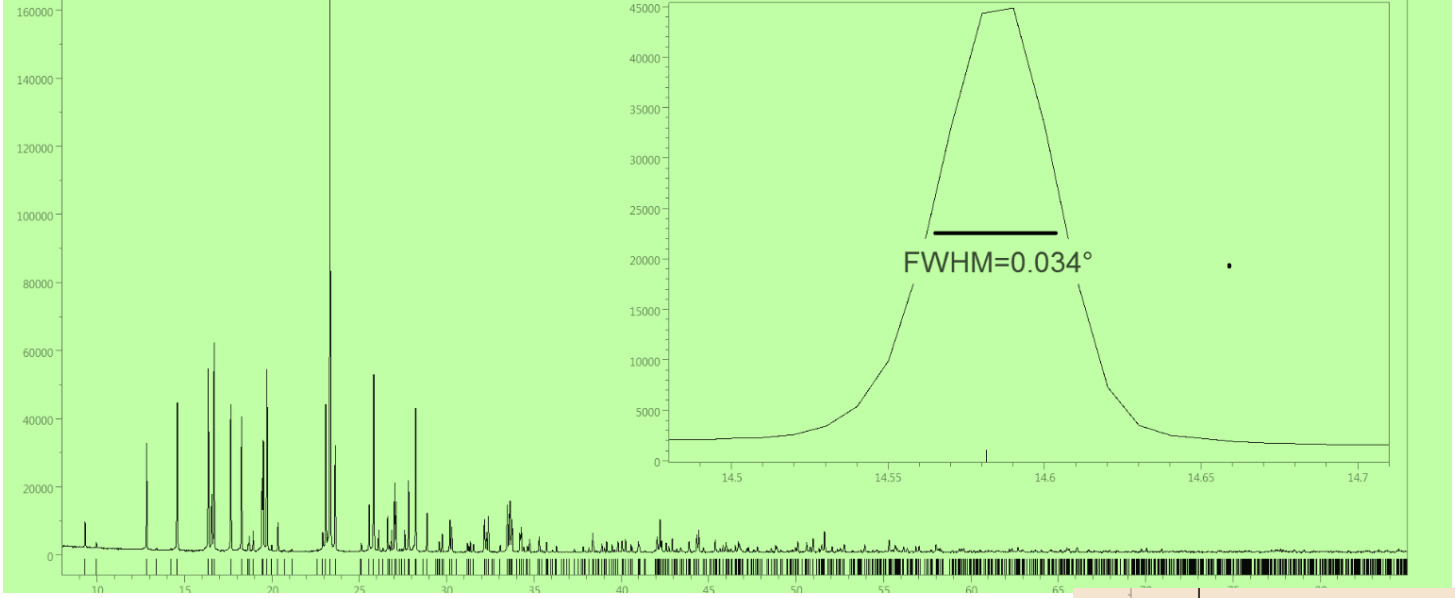
**On the Number of Statistically Independent Observations in a Powder Diffraction Pattern**

**A. Altomare, G. Cascarano, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, M. C. Burla and G. Polidori**

A method for deriving the number of statistically independent observations in a powder diffraction pattern is described.

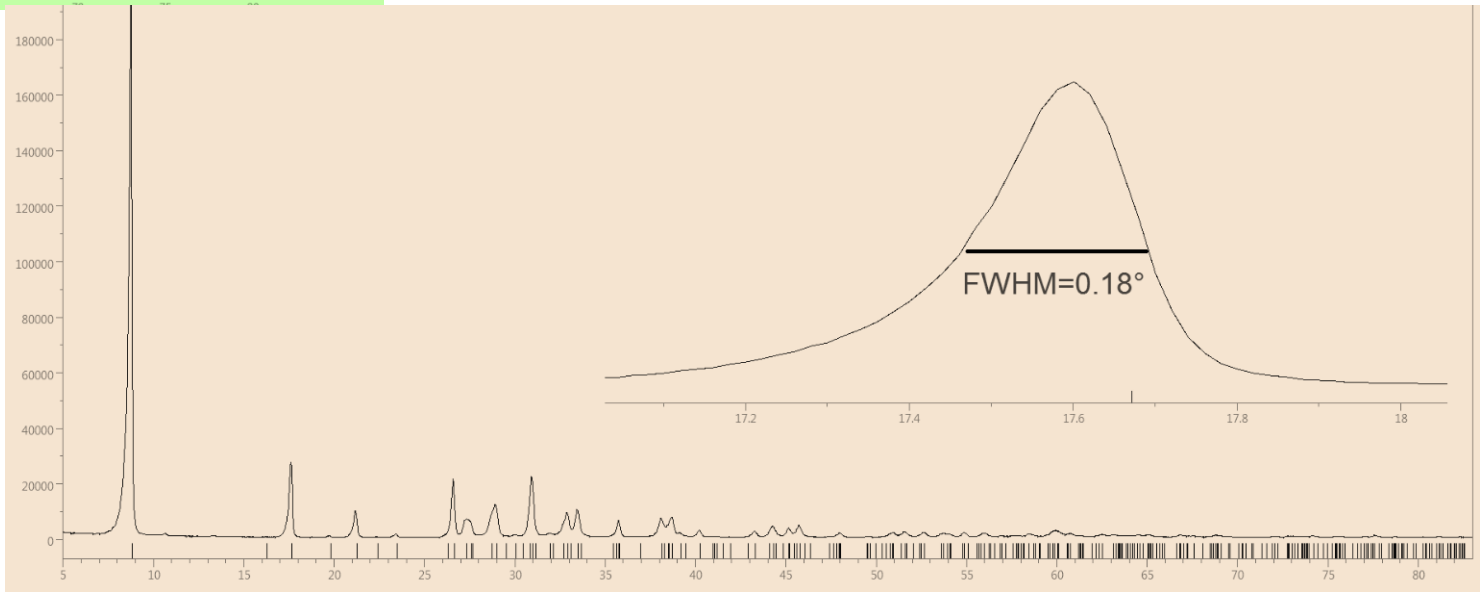
# Peak overlap

When peak overlap is severe (independent observation percentage < 50%) the success of the *ab-initio* solution process can be critical.




Synchrotron X-ray data

Independent observation percentage=65%



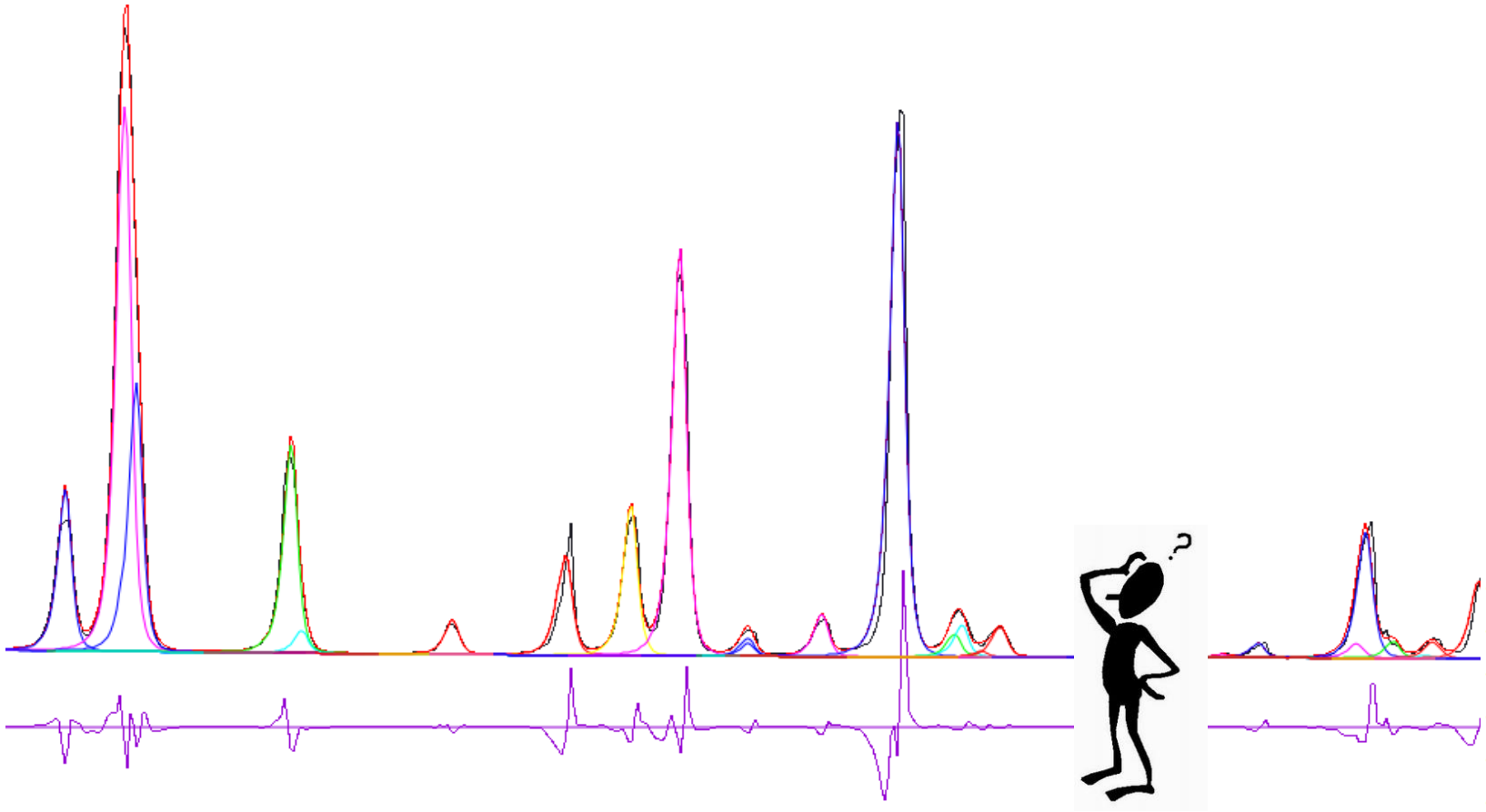

Laboratory X-ray data

Independent observation percentage=49%

# Steps of the structure solution process

- Determination of cell parameters
- Determination of space group
    - **Extraction of integrated intensities**
    - **Structure solution in the reciprocal space (Direct Methods, Patterson Methods, Maximum Entropy)**
    - Structure solution in the direct space (Simulated Annealing, Genetic Algorithm, Monte Carlo, Grid Search)
        - Crystal structure refinement (Rietveld Method)

# Extraction of integrated intensities







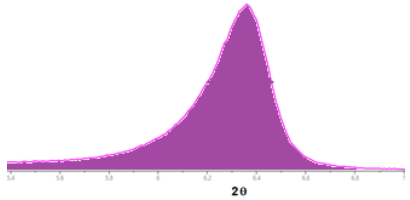
| n. | h | k | l | 2 theta | d | \|Fo\| |
|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 7.55810 | 11.68702 | 8.90578 |
| 2 | 2 | 0 | 0 | 15.14933 | 5.84351 | 31.32463 |
| 3 | 1 | 1 | 0 | 17.33083 | 5.11256 | 1.47938 |
| 4 | 0 | 1 | 1 | 19.50874 | 4.54644 | 10.53517 |
| 5 | 1 | 1 | -1 | 20.53053 | 4.32242 | 4.44129 |
| 6 | 1 | 1 | 1 | 21.35852 | 4.15669 | 4.16626 |
| 7 | 2 | 1 | 0 | 21.79216 | 4.07495 | 4.43352 |
| 8 | 3 | 0 | 0 | 22.80815 | 3.89567 | 1.04215 |
| 9 | 0 | 0 | 2 | 23.47877 | 3.78590 | 7.03053 |
| 10 | 1 | 0 | -2 | 23.98028 | 3.70784 | 13.24663 |
| 11 | 2 | 1 | -1 | 24.07613 | 3.69330 | 6.58350 |
| 12 | 1 | 0 | 2 | 25.39749 | 3.50406 | 8.73999 |
| 13 | 2 | 1 | 1 | 25.48833 | 3.49178 | 2.04203 |
| 14 | 2 | 0 | -2 | 26.77442 | 3.32690 | 3.78907 |
| 15 | 3 | 1 | 0 | 27.73659 | 3.21364 | 4.73911 |
| 16 | 0 | 1 | 2 | 28.29770 | 3.15118 | 5.94230 |
| 17 | 1 | 1 | -2 | 28.72053 | 3.10574 | 9.97870 |
| 18 | 3 | 1 | -1 | 29.29089 | 3.04655 | 2.12633 |
| 19 | 2 | 0 | 2 | 29.29354 | 3.04628 | 2.53219 |
| 20 | 1 | 1 | 2 | 29.92918 | 2.98301 | 3.38198 |
| 21 | 4 | 0 | 0 | 30.57185 | 2.92175 | 17.45384 |
| 22 | 3 | 1 | 1 | 31.05749 | 2.87717 | 2.60420 |
| 23 | 2 | 1 | -2 | 31.12123 | 2.87142 | 3.06918 |
| 24 | 3 | 0 | -2 | 31.29406 | 2.85595 | 1.87647 |
| 25 | 0 | 2 | 0 | 31.44354 | 2.84272 | 4.28453 |
| 26 | 1 | 2 | 0 | 32.38517 | 2.76218 | 1.79019 |
| 27 | 2 | 1 | 2 | 33.34112 | 2.68513 | 1.02032 |

# Extraction of integrated intensities

Area under each peak → $I_h$ diffraction integrated intensity

$I_h \propto |F_h|^2$

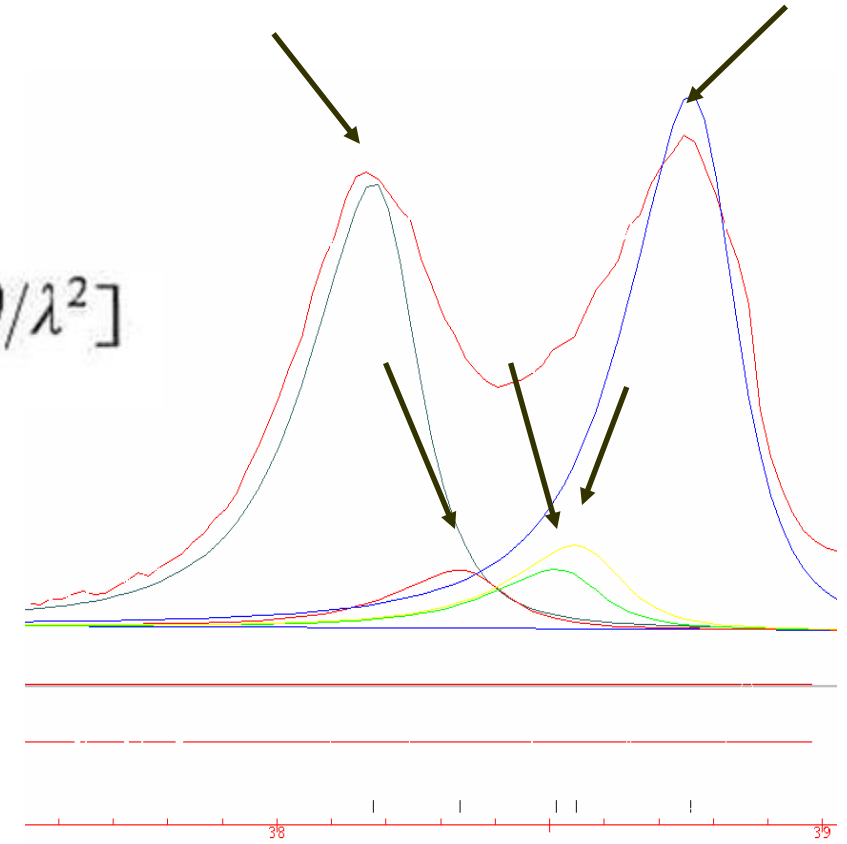
$|F_h|$ structure factor modulus




$$\mathbf{F_h} = \sum_j f_j \exp[2\pi i(hx_j + ky_j + lz_j)] \exp[-B_j \sin^2\theta/\lambda^2]$$

Two experimental peaks

**5 diffraction effects (reflections)**




Decomposition of the full diffraction pattern into single peaks and extraction of integrated intensities corresponding to each ***h*** reflection.

The larger the peak overlap degree, the less accurate the estimate of integrated intensities and experimental structure factor moduli.

# Extraction of integrated intensities

## LE BAIL *et al*., 1988

Iterative decomposition algorithm following the Rietveld formula.

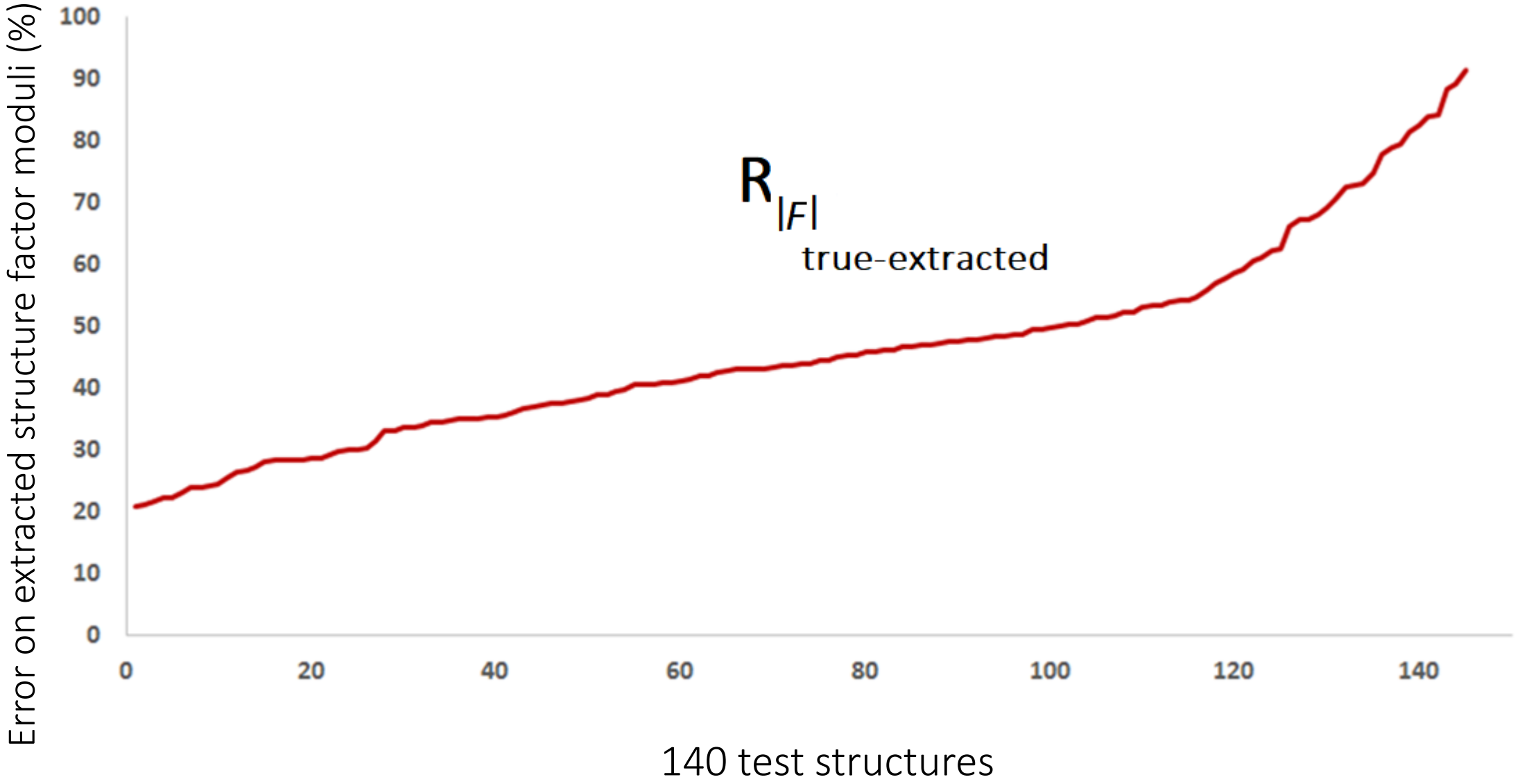
## PAWLEY, 1981

Non-linear least squares procedure. The integrated intensities are refinable variables in addition to the profile parameters.

Both methods are combined with *full profile fitting* techniques based on least squares minimizing the difference between observed and calculated profiles.

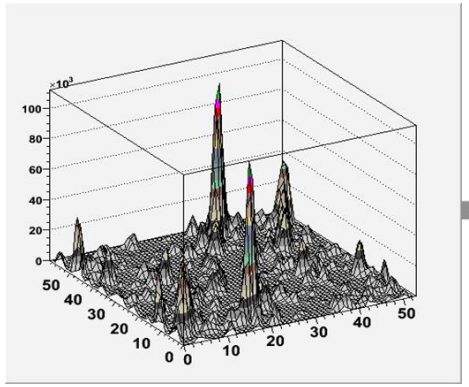
Over the years, the procedures for extracting integrated intensities have been considerably improved, but the two methods present a maximum level of reachable accuracy.

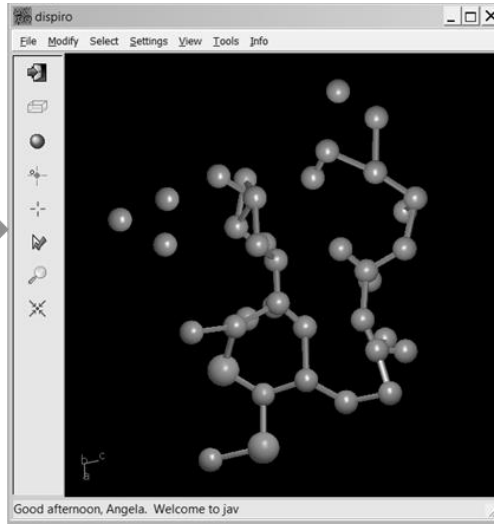
# Extraction of integrated intensities



# Structure solution by Direct Methods

$$\rho(\mathbf{r}) = V^{-1} \sum_{\mathbf{h}} |F_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) \exp(-2\pi i \bar{\mathbf{h}} \cdot \mathbf{r})$$






$$F_{hkl} = |F_{hkl}| \exp(i\varphi_{hkl})$$

What do we know? $|F_{\mathbf{h}}|$ ($I_{\mathbf{h}} \propto |F_{\mathbf{h}}|^2$)

What do we need? $\varphi_{\mathbf{h}}$

**The phase problem**

# The phase problem

Phases dominate the result of the electron density map calculation.
Incorrect phases generate wrong maps.

Positivity. The electron density map is everywhere positive: $\rho(\mathbf{r}) > 0$

Atomicity. Electrons are concentrated around nuclei and not dispersed in the unit cell.

Uniform distribution of atoms in the cell.

Giacovazzo, C. (2013). Phasing in Crystallography: A Modern Perspective. Oxford: IUCr/Oxford University Press.

# Structure solution by Direct Methods

**Direct Methods are statistical-probabilistical methods able to estimate phases of structure factors.**

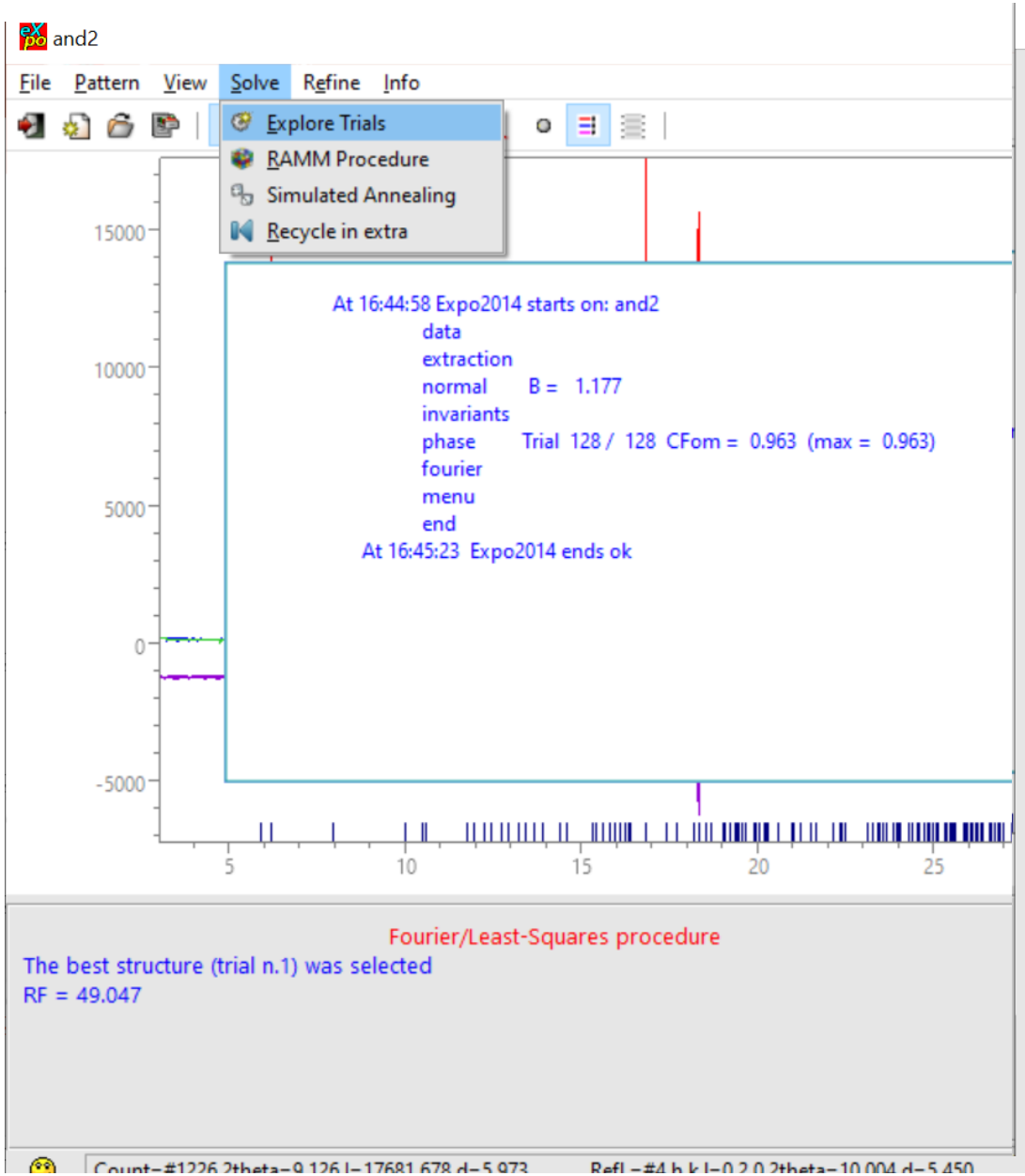
Direct Methods usually generate several phasing trials.

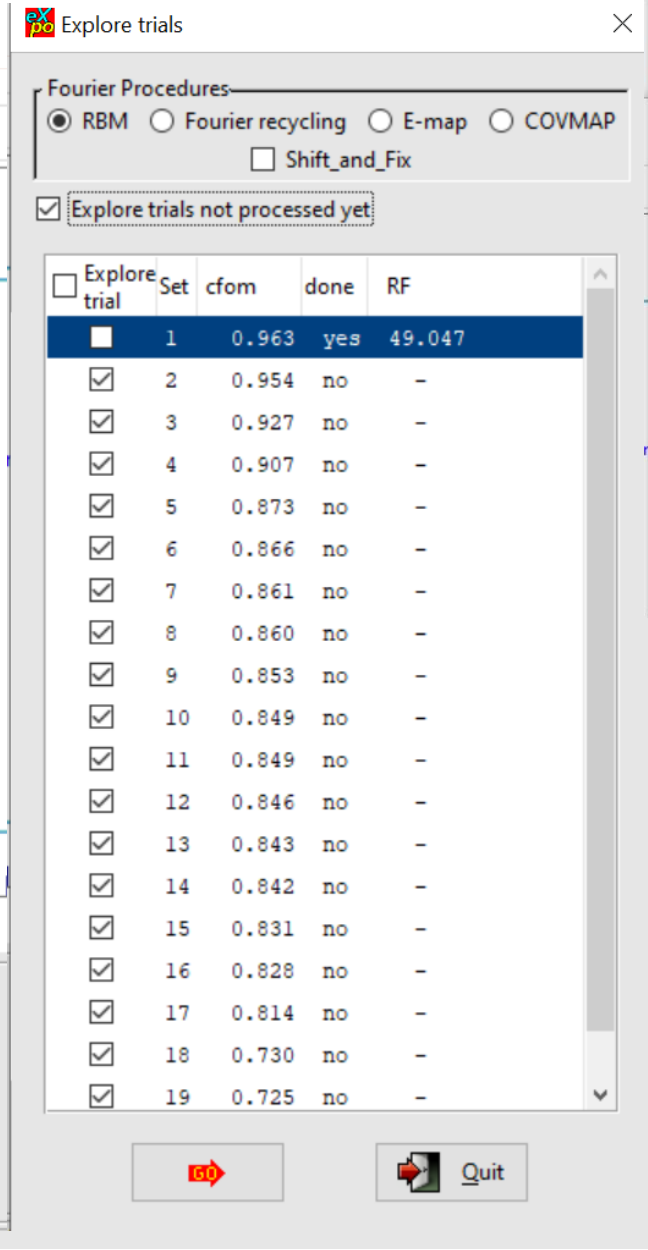
Phasing sets are ranked according to a Figure of Merit (CFOM).

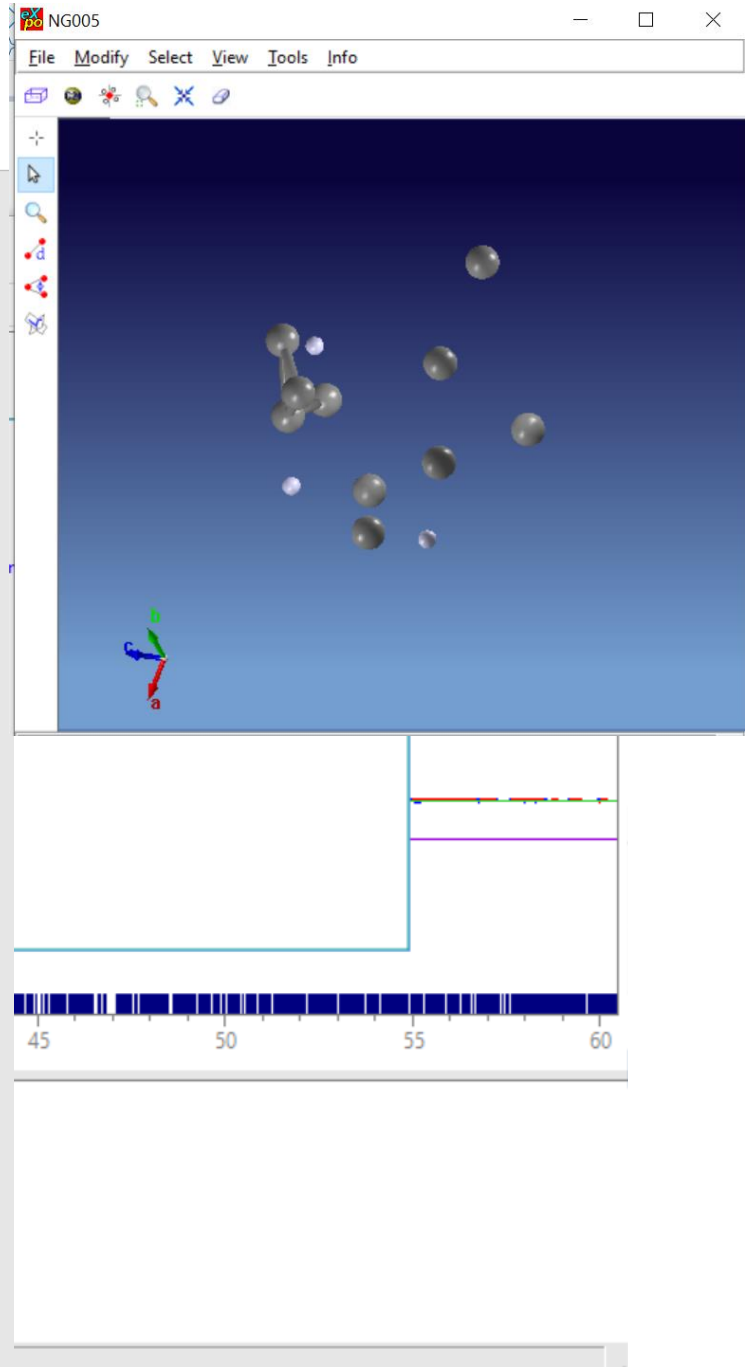
The set corresponding to the largest CFOM value is probabilistically estimated as the most reliable: its corresponding phases are automatically used to calculate the electron density map.

When the first ranked phasing set does not provide the correct solution, it can be necessary to explore all the sets generated by Direct Methods.

# Structure solution by Direct Methods: EXPO



The quality of the electron density map calculated after the phasing process by Direct Methods mainly depends on:

experimental resolution $d_{min}=\lambda/(2\sin\theta_{max})$

atomic resolution is the best condition

reliability of integrated intensities extracted from the experimental pattern

peak overlap should be low (percentage of independent observations larger than 50%)

structure complexity

Depending on the number of non-H atoms in the asymmetric unit (35 is a current reasonable limit)

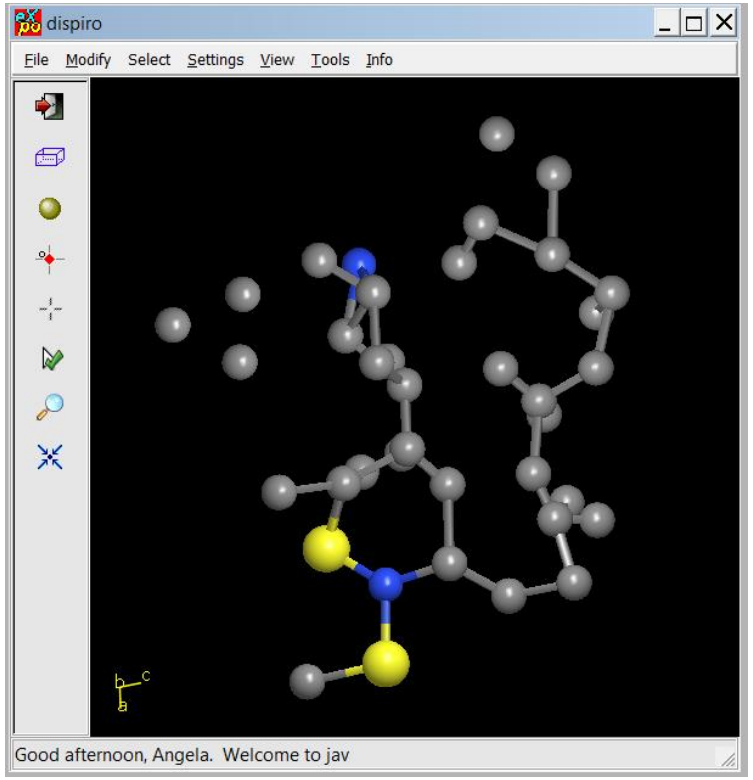
For organic compounds, the experimental resolution is usually far away from 1 Å because of the rapid scattering factor decay of light atoms.

Organic structures are the most resistant to attempts of solution in the reciprocal space.

In the case of inorganic structures, the heavy atoms are quite easily located. It is sometimes difficult to recover positions of light atoms and complete the structure.

# Optimization of Direct Methods structure model

**The structure model provided by Direct Methods is often approximate: atoms are missed to the complete solution, additional false positions occur, atoms are not perfectly located.**




25 non-H atoms in the asymmetric unit

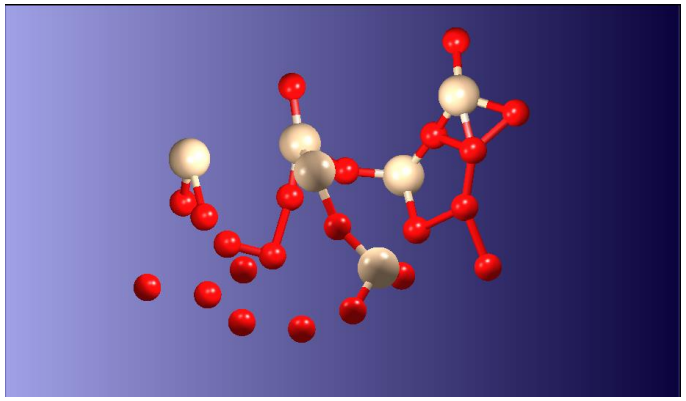
It is useful to optimize the model before the final Rietveld refinement, owing to the modest convergence rate.

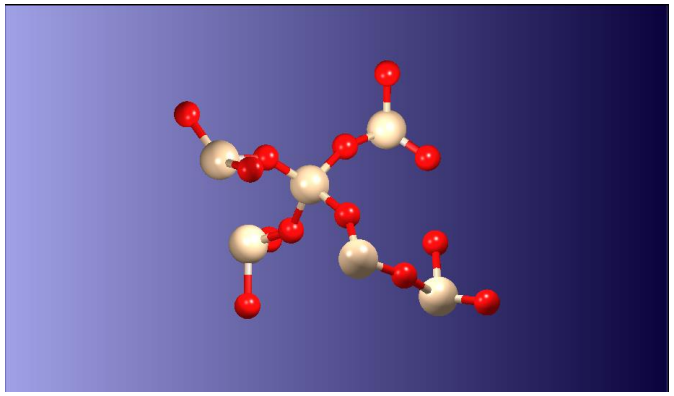
Structure model optimization strategies are based on suitable combinations of weighted least-squares and Fourier map calculations.


- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Moliterni, A.G.G.; Rizzi, R. *Powder diffraction: the new automatic least-squares Fourier recycling procedure in EXPO2005*. J. Appl. Cryst. **2006**, *39*, 558-562.
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Kamel, G.S.; Moliterni, A.; Rizzi, R. *Minimally resolution biased electron-density maps*. Acta Cryst. **2008**, *A64*, 326–336.
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Moliterni, A.; Rizzi, R. *Correcting resolution bias in electron density maps of organic molecules derived by direct methods from powder data*. J. Appl. Cryst. **2008**, *41*, 592–599.
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Maggi, S.; Moliterni, A.; Rizzi, R. *Correcting electron-density resolution bias in reciprocal space*. Acta Cryst. **2009**, *A65*, 183–189.
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Moliterni, A.; Rizzi, R. *The dual-space resolution bias correction algorithm: application to powder data*. J. Appl. Cryst. **2010**, *43*, 798–804.
- Altomare, A.; Cuocci, C.; Giacovazzo, C.; Moliterni, A.; Rizzi, R. *COVMAP: a new algorithm for structure model optimization in the EXPO package*. J. Appl. Cryst. **2012**, *45*, 789–797.
- Altomare, A.; Cuocci, C.; C.; Moliterni, A.; Rizzi, R. *RAMM: a new random-model-based method for solving ab initio crystal structure using the EXPO package*. J. Appl. Cryst. **2012**, *45*, 789–797.
- Altomare, A.; Cuocci, C.; Moliterni, A.; Rizzi, R., Corriero, N., Falcicchio, A. *The Shift_and_Fix procedure in EXPO: advances for solving ab initio crystal structures by powder diffraction data*. J. Appl. Cryst. **2017**, *50*, 1812–1820.

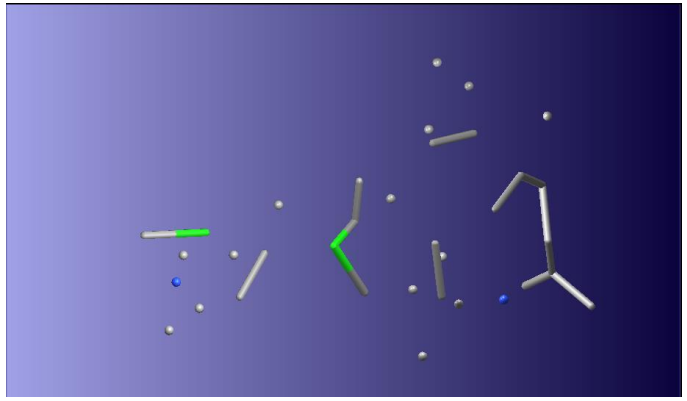
# Optimization of Direct Methods structure model

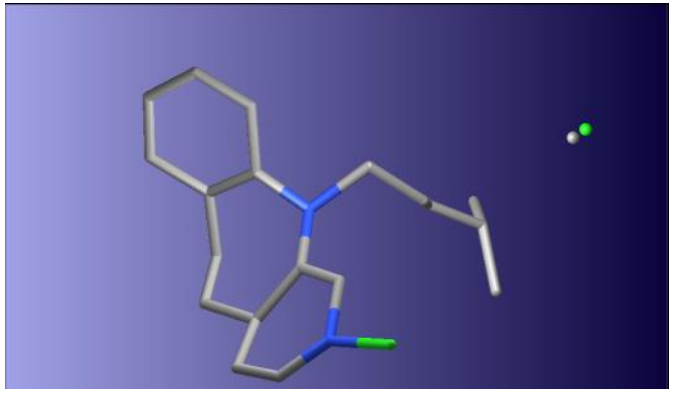
UTM-1 zeolite



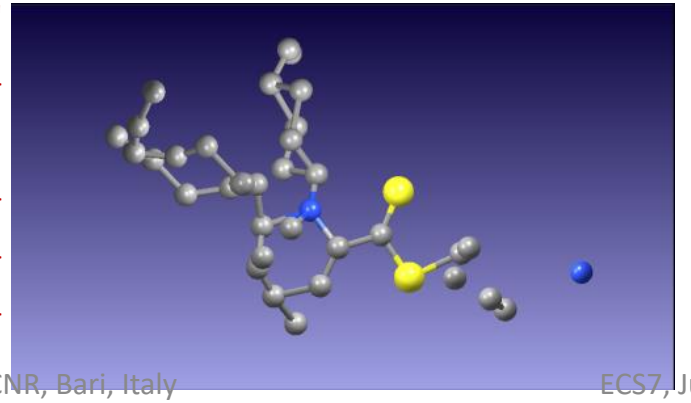


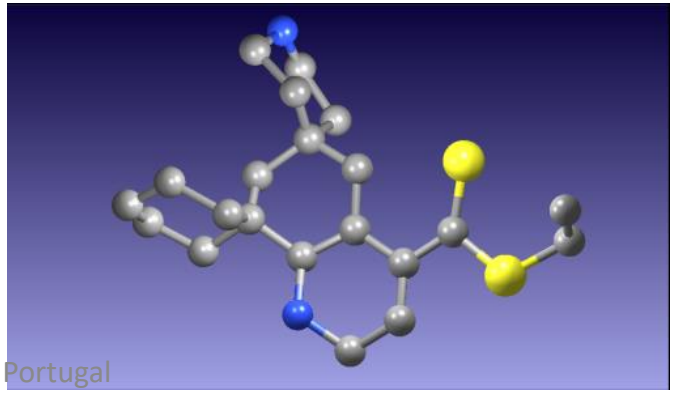
Clomipramine hydrochloride $C_{19}H_{24}ClN_2 \cdot Cl$



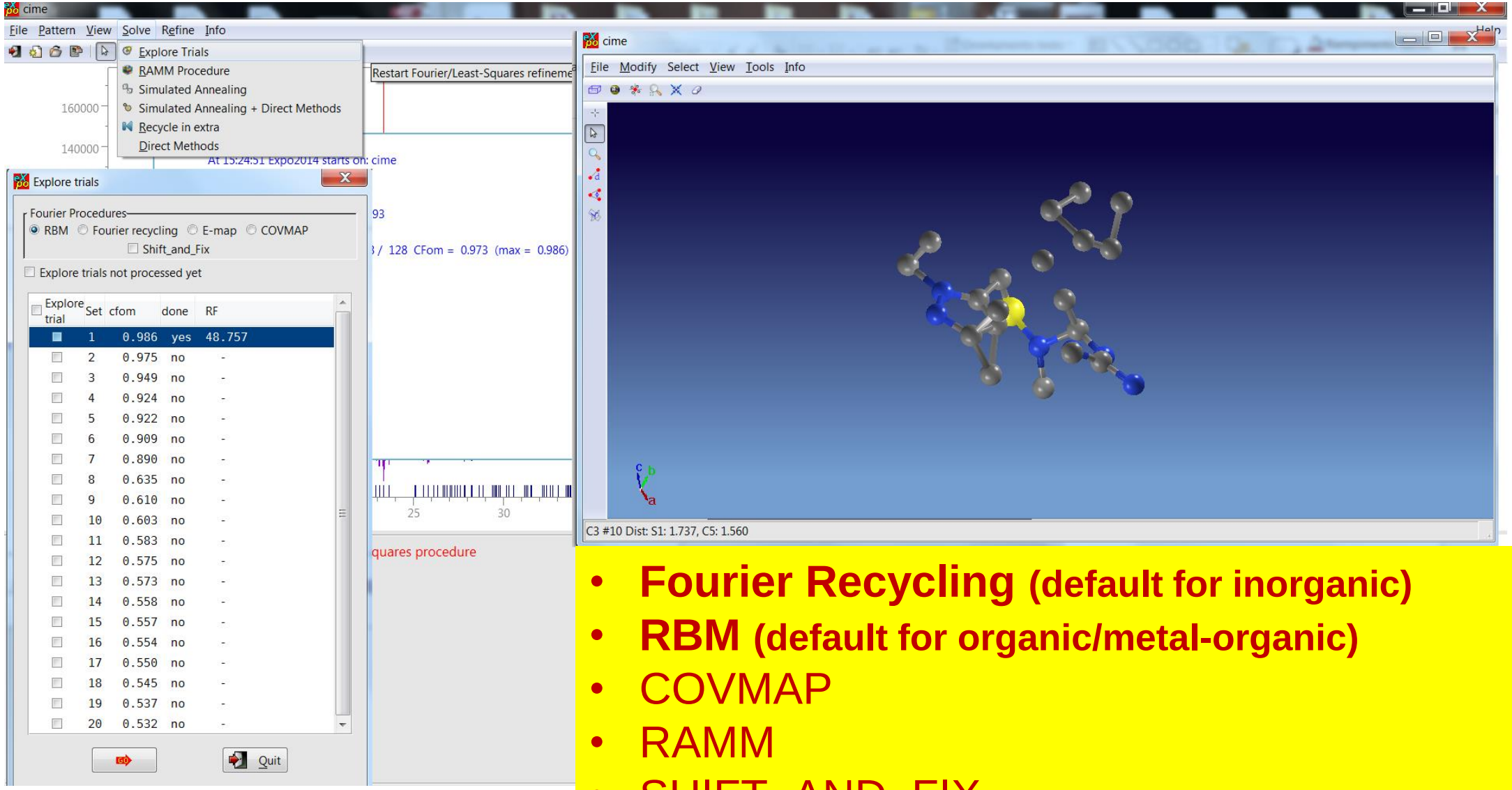


Ethyl 1' ,2' ,3' ,4' ,4a' ,5' ,6' ,7' -octahydrodispiro[cyclohexane-1,2' -quinazoline-4' ,1'' -cyclohexane]-8' -carbodithioate





# Optimization of Direct Methods structure model: EXPO

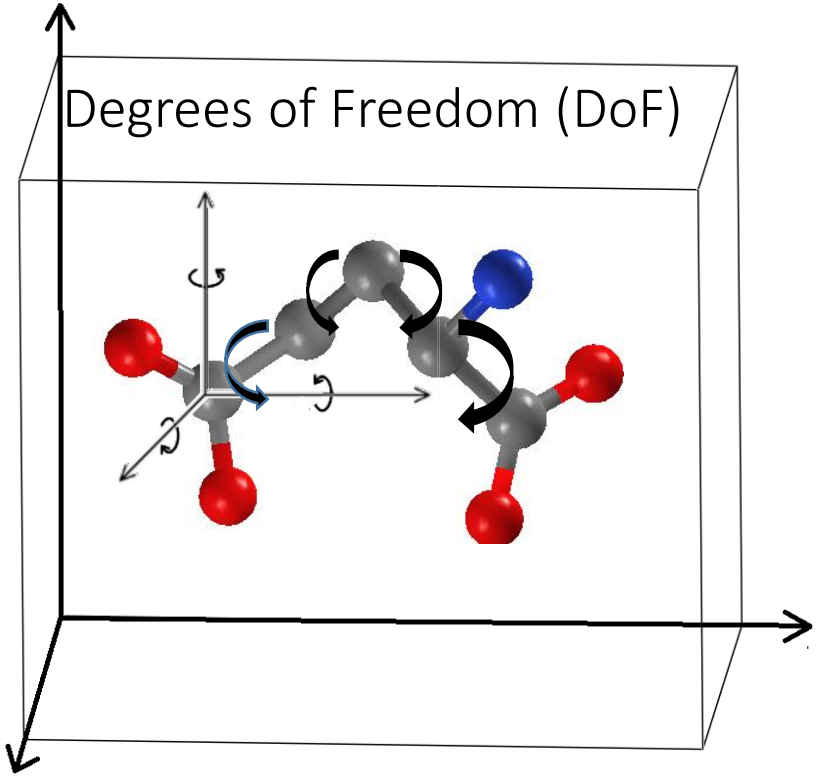



- **Fourier Recycling (default for inorganic)**
- **RBM (default for organic/metal-organic)**
- COVMAP
- RAMM
- SHIFT_AND_FIX

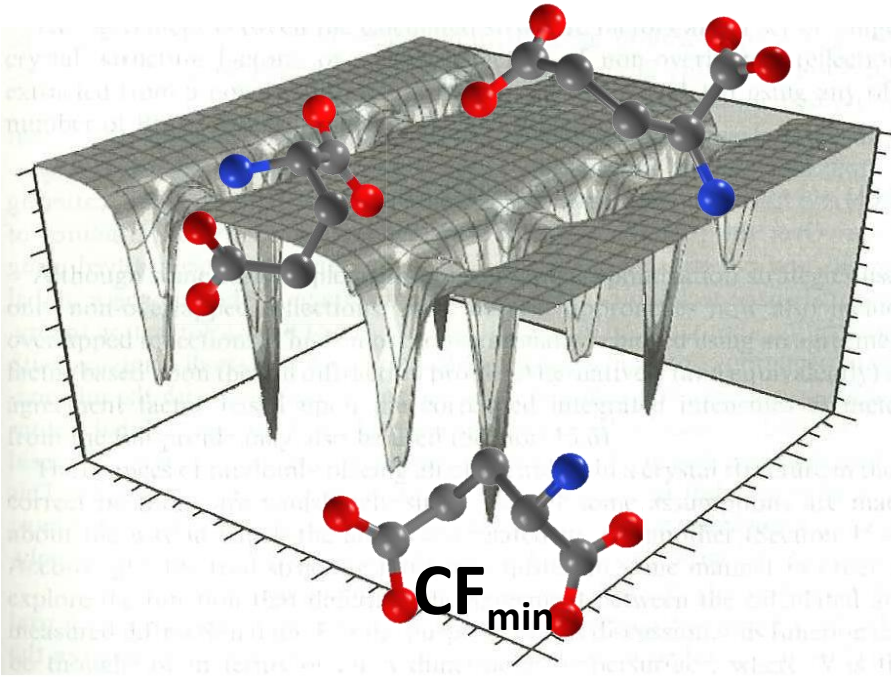
# Steps of the structure solution process

- Determination of cell parameters
- Determination of space group
    - Extraction of integrated intensities
    - Structure solution in the reciprocal space (Direct Methods, Patterson Methods, Maximum Entropy)
    - **Structure solution in the direct space (Simulated Annealing, Genetic Algorithm, Monte Carlo, Grid Search)**
        - Crystal structure refinement (Rietveld Method)

# Structure solution in the direct space




$$\mathbf{CF} = \sqrt{\frac{\sum_i w_i \left(y_{oss}(\theta_i) - y_{calc}(\theta_i)\right)^2}{\sum_i w_i \, y_{oss}(\theta_i)^2}}$$




**Heuristic methods of global optimization**: Genetic algorithm
Grid search
Simulated Annealing, Parallel Tempering
Particle Swarm Optimization
Big Bang-Big Crunch …

# Structure solution in the direct space

**The extraction of structure factor moduli from the experimental pattern is not required.**

Molecular connectivity must be known (good knowledge of bond distances and angles).

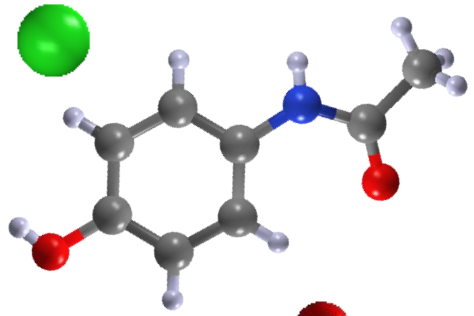
The starting model (built according to structures already solved similar to the structure under study and/or by using *molecule editors*) is a combination of building blocks: atoms, molecules, polyhedrons.

The global minimum of the cost function (CF) in the hypersurface of DoF is searched.

The success depends on how far the starting model is from the true model and the number of DoF (DoF<15 is a good condition).

# Simulated Annealing

$\Delta CF = CF_{current} - CF_{previous}$
RND = random number



Starting random model

Trial generation of structure models by variation of DoF

CF computation corresponding to the trial structure

Metropolis: is the new configuration better than the previous one?

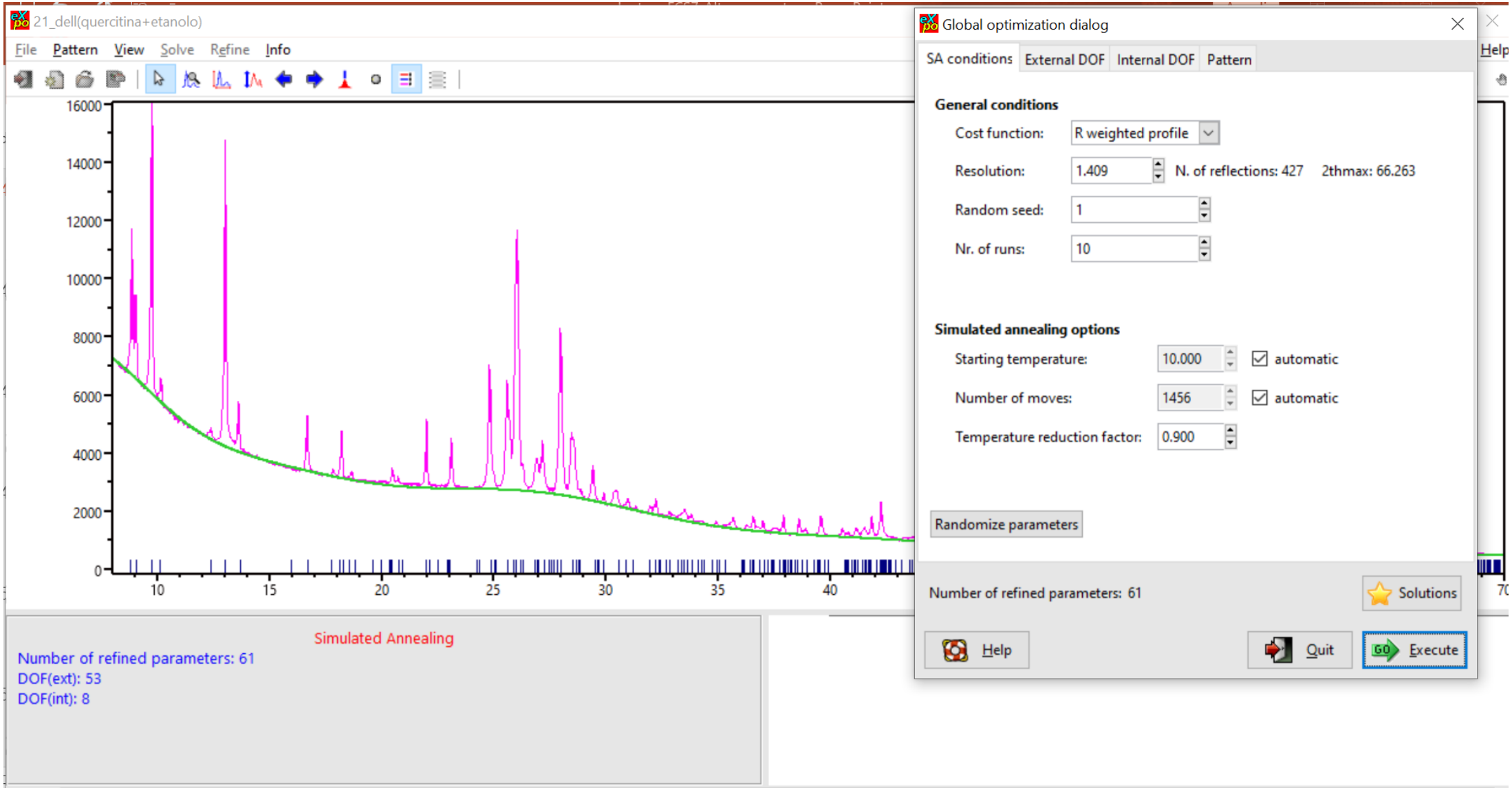
Yes (**ΔCF** < 0) The new configuration is kept

No (**ΔCF** > 0)

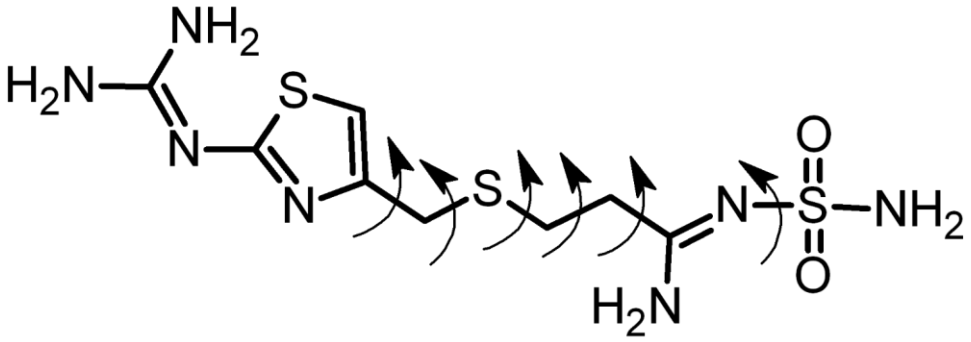
The trial model is accepted if exp(-ΔCF/T) > RND

Annealing

# Simulated Annealing: EXPO



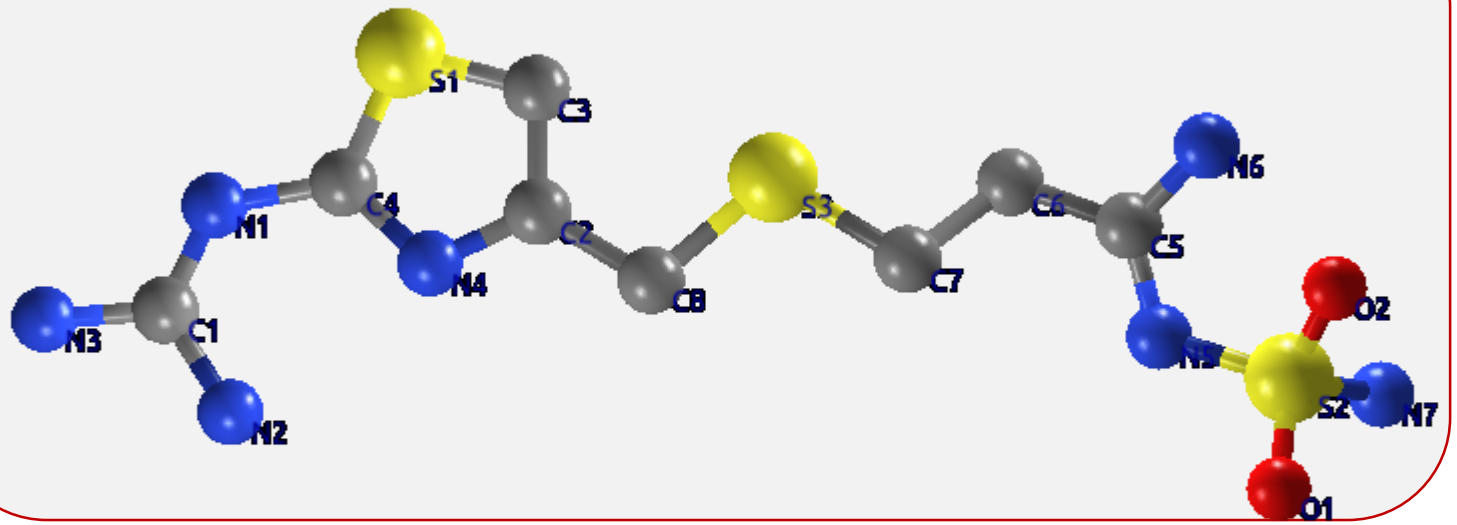
# Structure solution in the direct space



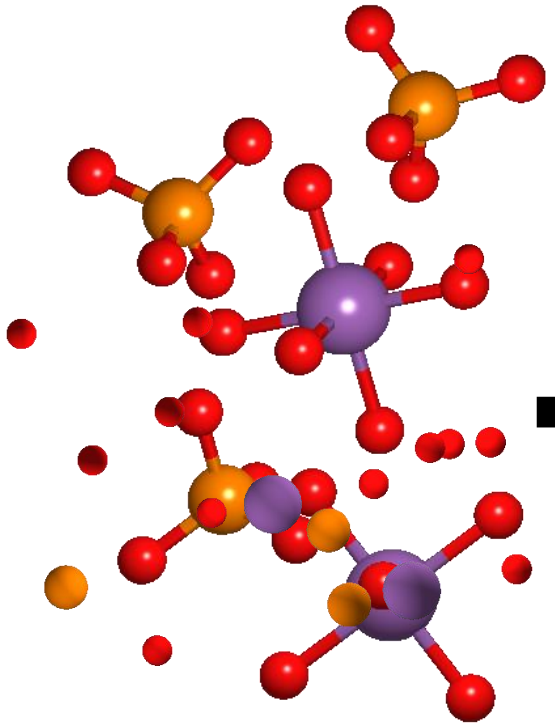

Starting model
Famotidine form B

$N_{frag} = 1$
$N_{DoF} = 6 + 6$

Simulated Annealing



# Structure solution in the direct space

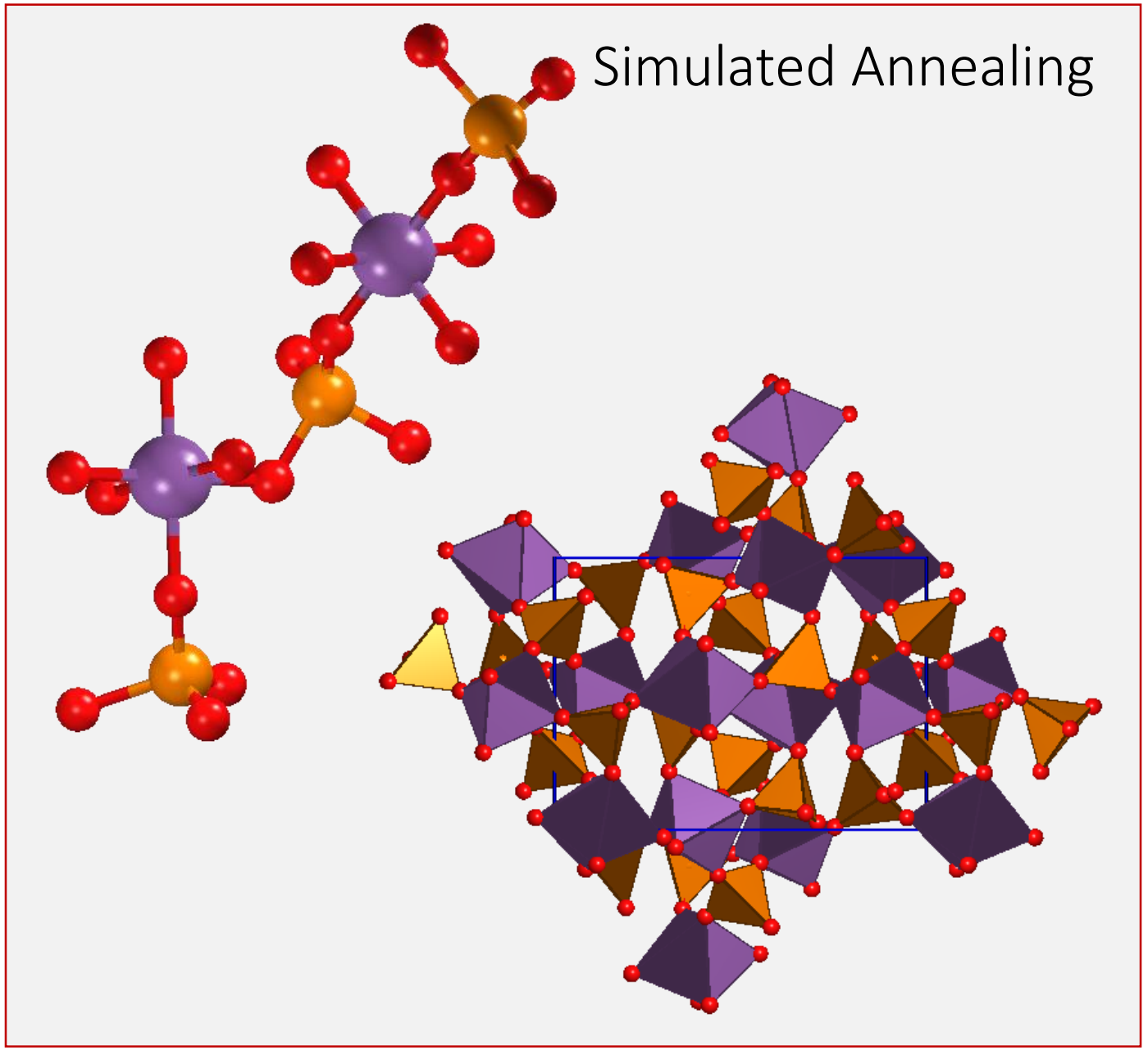


Starting model

$Sb_2(PO_4)_3$

$N_{DoF} = 51$

$N_{DoF} = 6 + 6 + 6 + 6 + 6$



# Structure solution in the direct space

## Validation

CF value

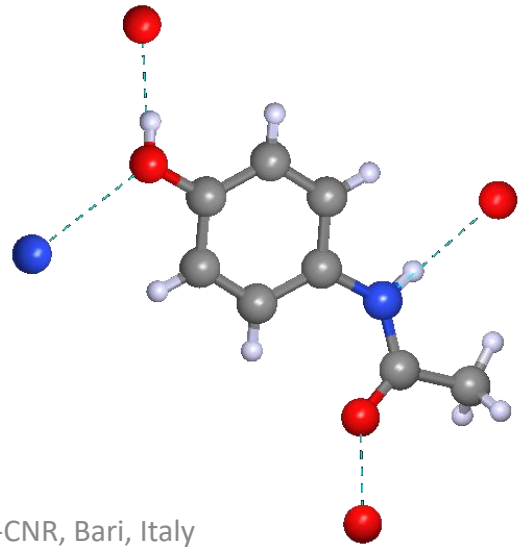
Reproducibility of solution

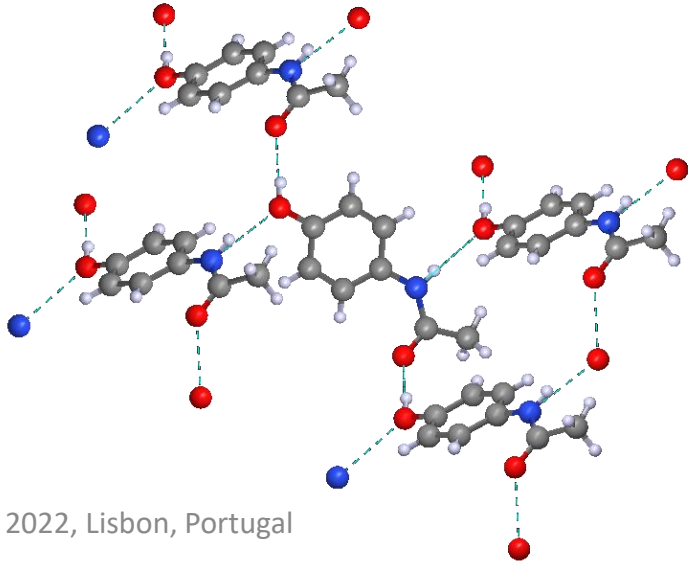
Difference between observed and calculated profiles

Crystal packing

Close contacts and voids

Network of interactions: hydrogen bonds and short contacts





## Structure solution in the reciprocal space (*ab-initio, single-crystal like*)

Only chemical formula and experimental diffraction profile (**easy**)

Possibly atomic resolution (**hard**)
Short execution time (**easy**)
Extraction of integrated intensities from the powder pattern to derive the experimental structure factor moduli (**hard**)

## Structure solution in the direct space

In addition to the chemical formula and experimental profile, also the knowledge of the expected geometry (**easy for molecular compounds, less easy for non-molecular**)
About 2-2.5 Å resolution (**easy**)
Usually long execution time (**hard**)
No extraction of the experimental structure factor moduli (**easy**)

COMBINATION: EXPO, SUPERFLIP, TOPAS, FOCUS

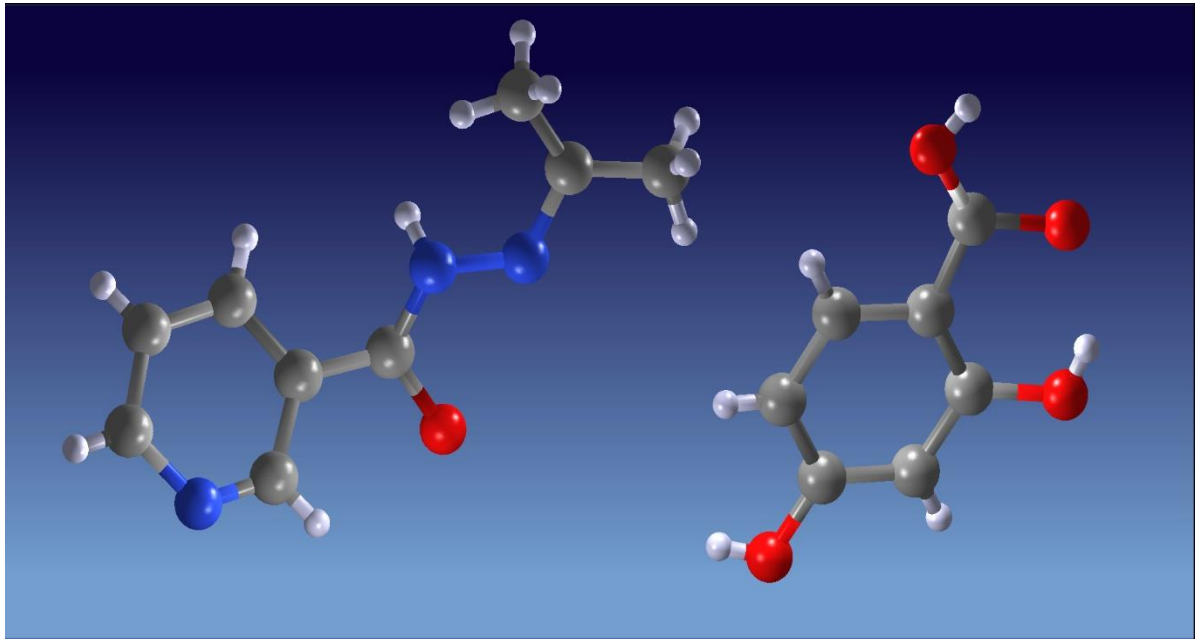
*overlapping, experimental resolution* *starting model, DoF*

SOFTWARE

EXPO, MICE, XLENS, POWSIM

DASH (SA), EXPO (SA), ENDEAVOUR (SA), FOX (PT), GEST (GA), PeckCryst (PS), PowderSolve (SA), PSSP (SA), TOPAS (SA), FIDEL-GO (MC)…

# Structure solution



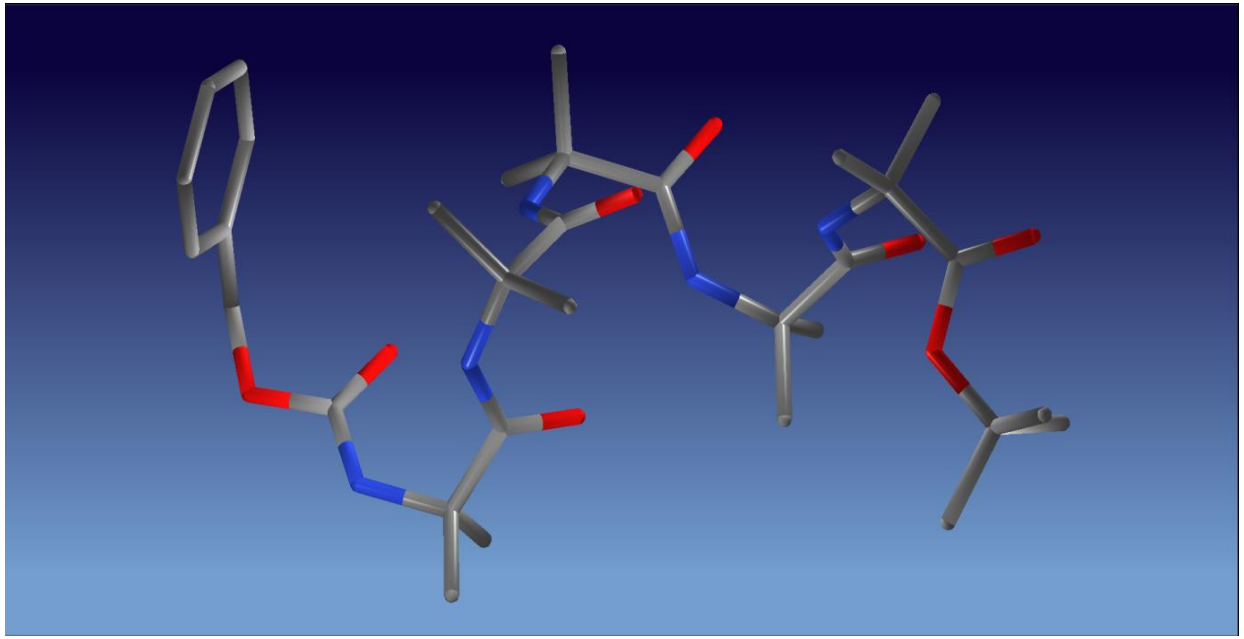
Co-crystal $C_9H_{11}N_3O \cdot C_7H_6O_4$*

***ab-initio*** **solution by EXPO**

COMPUTATIONAL TIME: A FEW MINUTES

**24 non-H atoms in the asymmetric unit**

*S.H. Lapidus *et al.*, *Acta Crystallogr.,* 2012, C68, o335.



Pentapeptide Z-$(Aib)_5$-O-*t*Bu*

**solution by Simulated Annealing in EXPO**

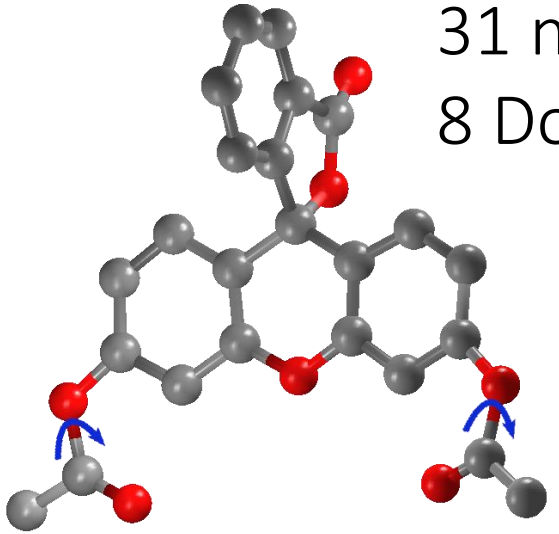
COMPUTATIONAL TIME: 500h

PARALLEL COMPUTING 20 CPU-cores: 25 h

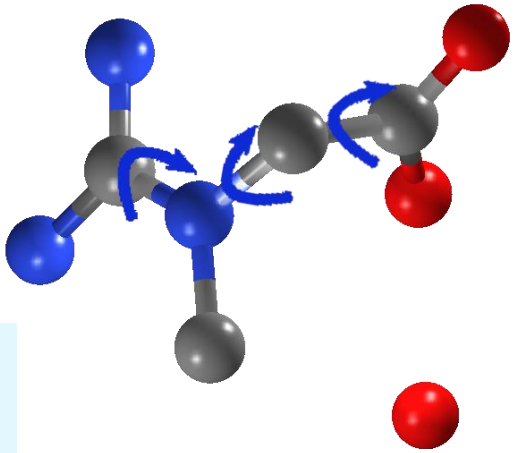
**26 DoF (45 non-H atoms in the asymmetric unit)**

*E. Benedetti, et al., J Am. Chem. Soc., 1982, 104, 2437.

# Solution in the reciprocal space/ Solution in the direct space

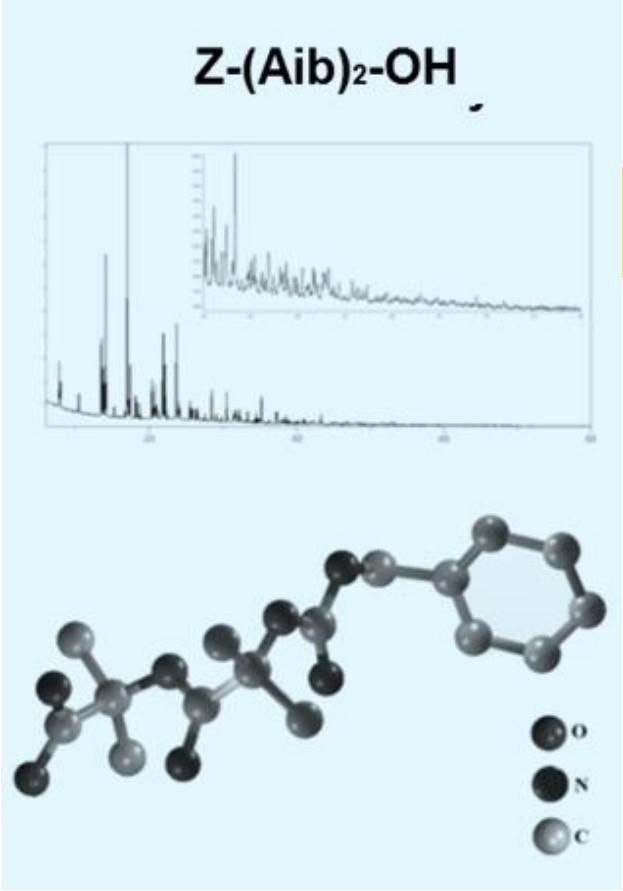


31 non-H atoms
8 DoF



10 non-H atoms
2 fragments
12 DoF

**23** non-H atoms in the asymmetric unit
**15** DoF




RSS → 14m 57s

DSS → 164m 40s

# Steps of the structure solution process

- Determination of cell parameters
- Determination of space group
  - Extraction of integrated intensities
  - Structure solution in the reciprocal space (Direct Methods, Patterson Methods, Maximum Entropy)
  - Structure solution in the direct space (Simulated Annealing, Genetic Algorithm, Monte Carlo, Grid Search)
    - **Crystal structure refinement (Rietveld Method) and depositing the CIF file into a structure database**

# Rietveld refinement

$$\Phi = \sum_i w_i \cdot (y_{oi} - y_{ci})^2$$

$y_{oi}$ the observed intensity at the *i*-th data point,

$y_{ci}$ the calculated intensity at the *i*-th data point, $y_c(i) = s \sum_{\mathbf{h}} \mathrm{I}_{\mathbf{h}} P(i,\mathbf{h}) A(i,\mathbf{h}) + y_b(i)$

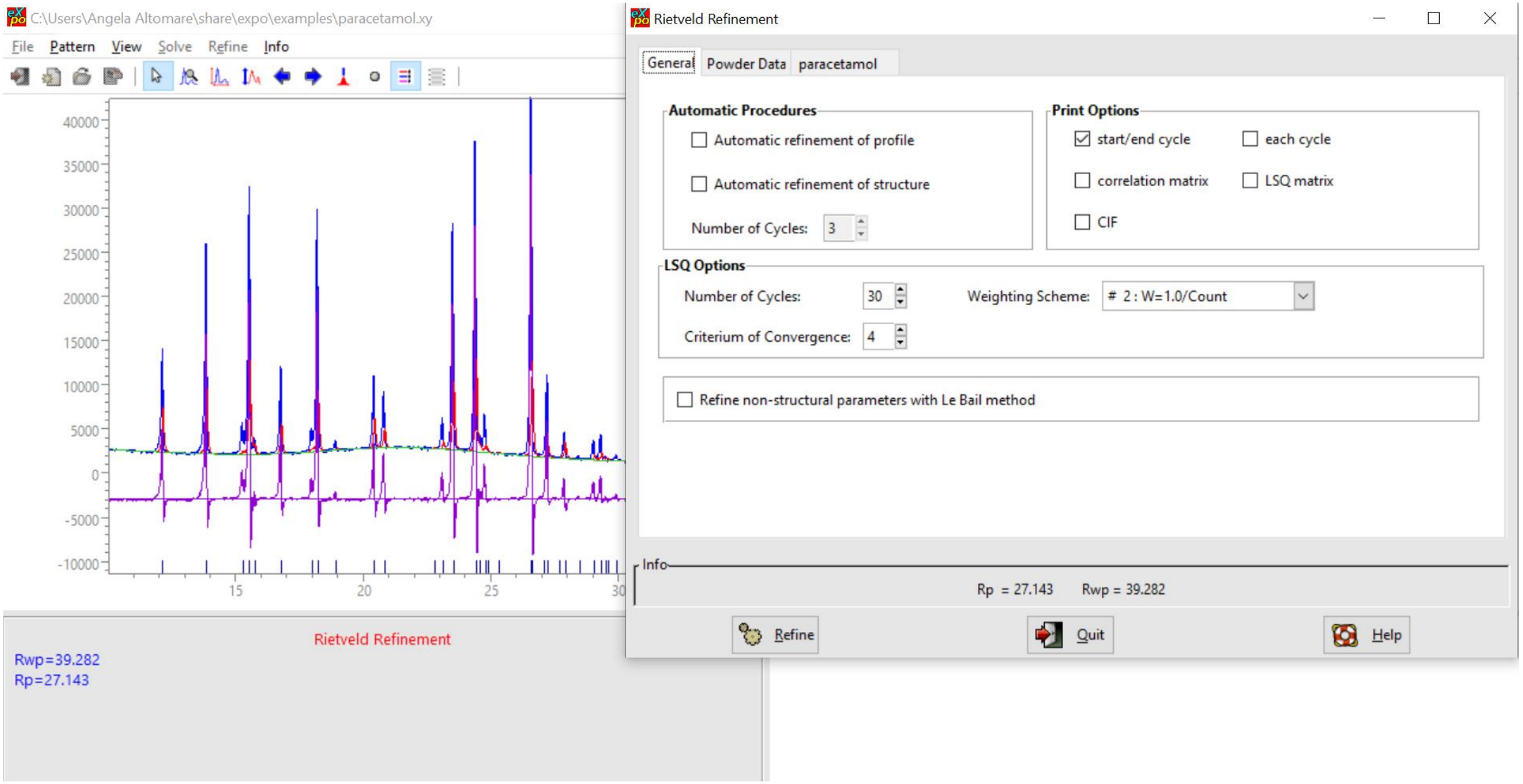
$w_i$ a weight $1/\sigma^2(y_{oi})$

Reliability parameters of refinement

- Weighted profile R-factor $R_{Wp} = \frac{\sum_i w_i \cdot (y_{oi} - y_{ci})^2}{\sum_i w_i y_{oi}^2}$
- Unweighted profile R-factor $R_p = \frac{\sum_i |(y_{oi} - y_{ci})|}{\sum_i y_{oi}}$
- Expected R value $R_{\exp} = \sqrt{\frac{(N-P)}{\sum_i w_i y_{oi}^2}}$
- Goodness-of-fit $\chi^2 = \frac{R_{wp}}{R_{\exp}}$
- Other residual on *F* or *F*$^2$: $R_F = \frac{\sum_{hkl} |(F_{hkl}^{obs} - F_{hkl}^{calc})|}{\sum_{hkl} F_{hkl}^{obs}}$ $R_B = \frac{\sum_{hkl} |(I_{hkl}^{obs} - I_{hkl}^{calc})|}{\sum_{hkl} I_{hkl}^{obs}}$

- Scale
- 2θ correction
- Background coefficients
- Profile parameters (FWHM, asymmetry, peak form coefficient)
- Atom coordinates
- Isotropic displacements
- Preferred orientation
- Cell parameters

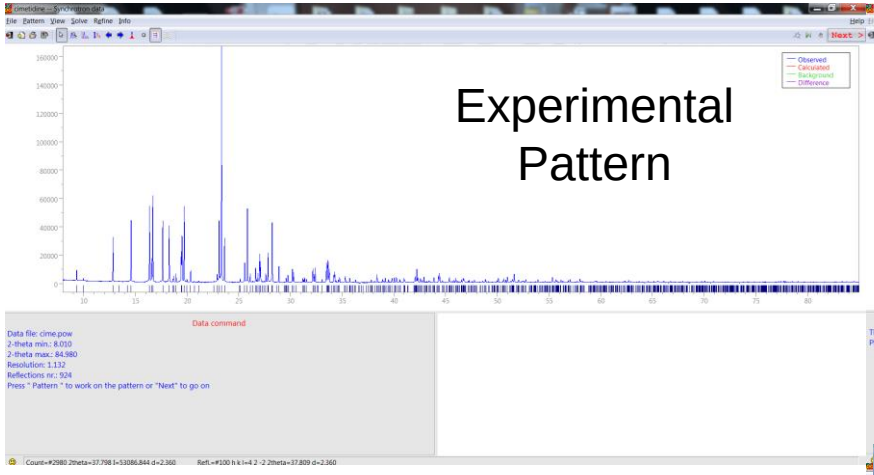
# Rietveld refinement: EXPO

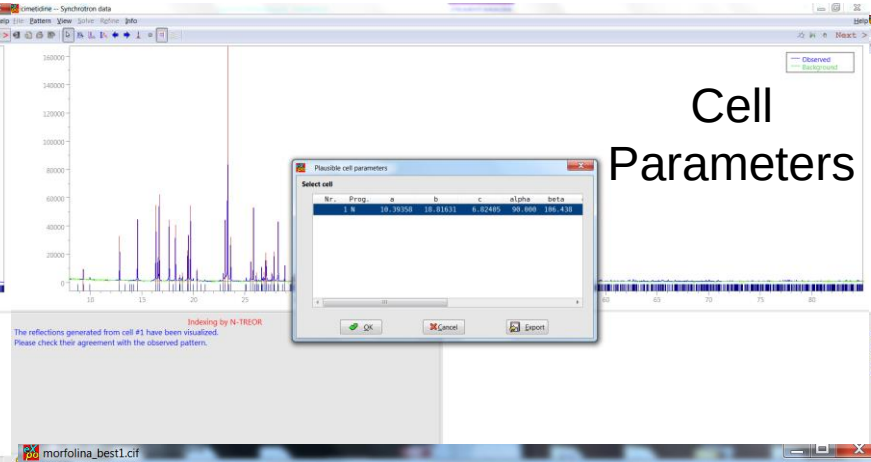


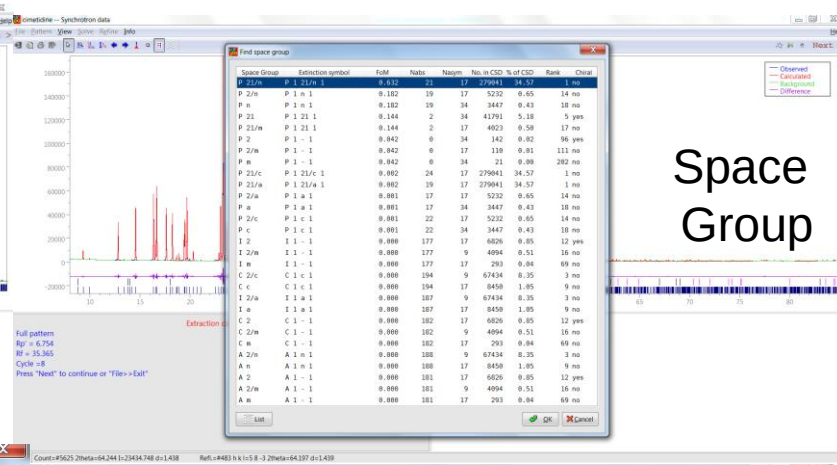
# The CIF File

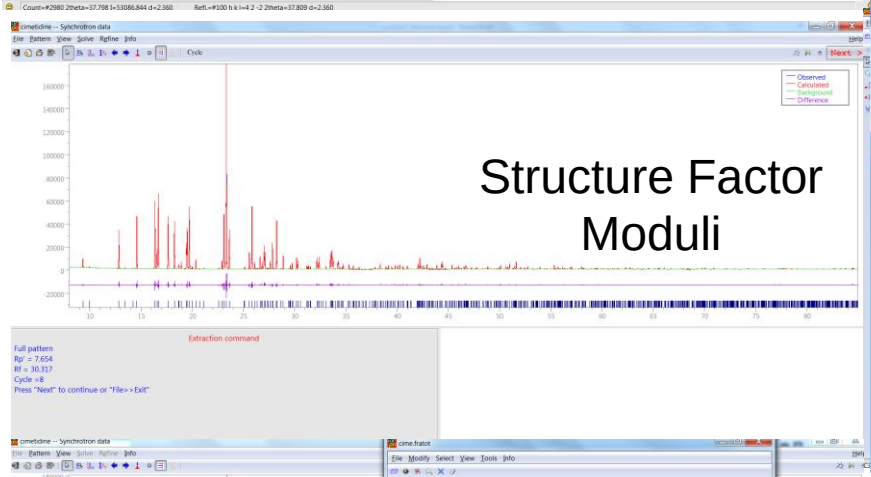
```
#======================================================================
data_global
#======================================================================

_audit_creation_method           Expo2014

_chemical_name_systematic    ?
_chemical_formula_moiety     'C11 H7 O2 S'
_chemical_formula_sum        'C11 H7 O2 S'
_chemical_formula_weight     203.240

loop_
    _atom_type_symbol
    _atom_type_description
    _atom_type_scat_source
   'H'   'Hydrogen'   'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
   'C'   'Carbon'   'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
   'O'   'Oxygen'   'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
   'S'   'Sulphur'   'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'

_cell_length_a                10.1317(3)
_cell_length_b                19.5045(7)
_cell_length_c                9.6954(3)
_cell_angle_alpha             90.000
_cell_angle_beta              90.000
_cell_angle_gamma             90.000
_cell_volume                  1915.95(11)
_cell_formula_units_Z         8
_cell_measurement_temperature    ?

_exptl_crystal_density_diffrn       1.409
_exptl_crystal_density_meas     ?
_exptl_crystal_density_method 'not measured'
_exptl_absorpt_coefficient_mu       2.738
_symmetry_Int_Tables_number       61
_symmetry_cell_setting            orthorhombic
_symmetry_space_group_name_H-M     'P c a b'
_symmetry_space_group_name_hall    '-P 2bc 2ac'

loop_
   _symmetry_equiv_pos_site_id
   _symmetry_equiv_pos_as_xyz
   1   'x, y, z'
   2   'x+1/2, -y, -z+1/2'
   3   '-x, -y+1/2, z+1/2'
   4   '-x+1/2, y+1/2, -z'
   5   '-x, -y, -z'
   6   '-x+1/2, y, z+1/2'
   7   'x, y+1/2, -z+1/2'
   .....
```

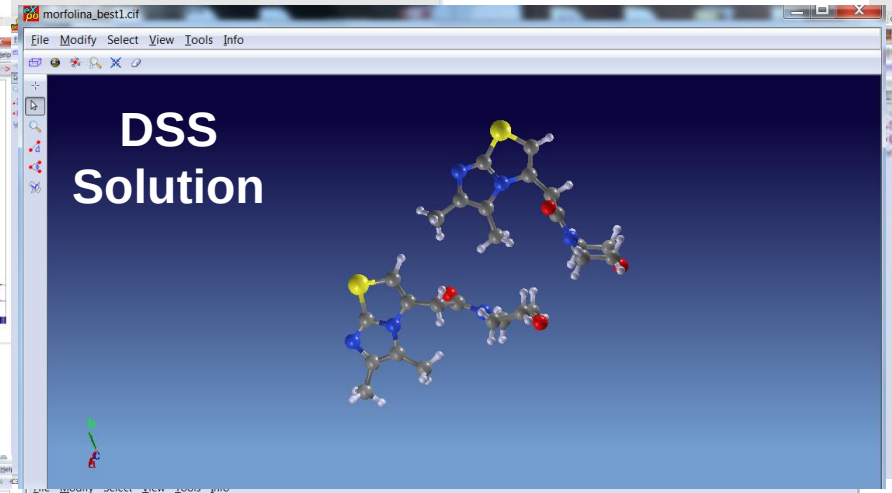
# The solution process: EXPO

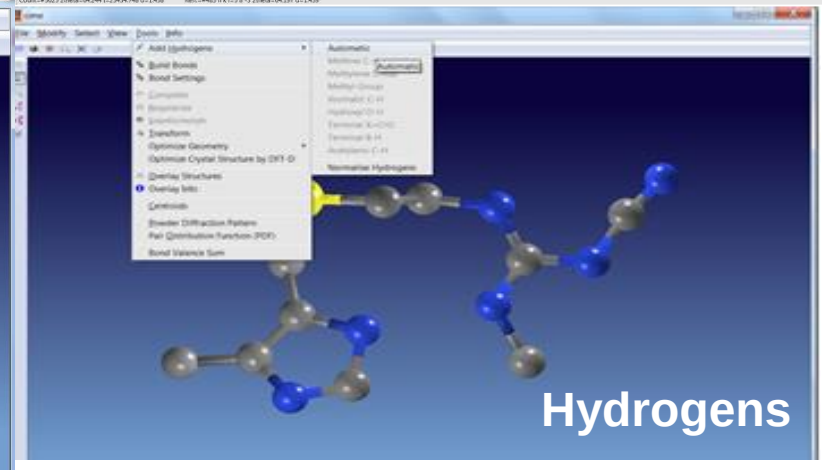


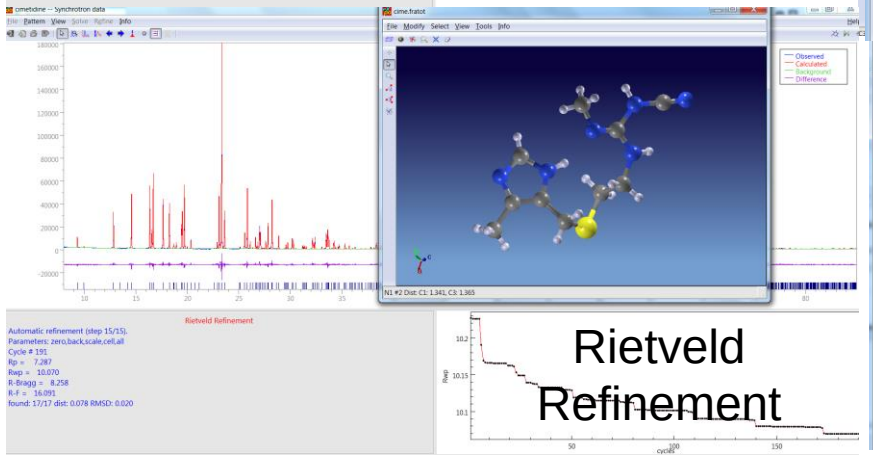



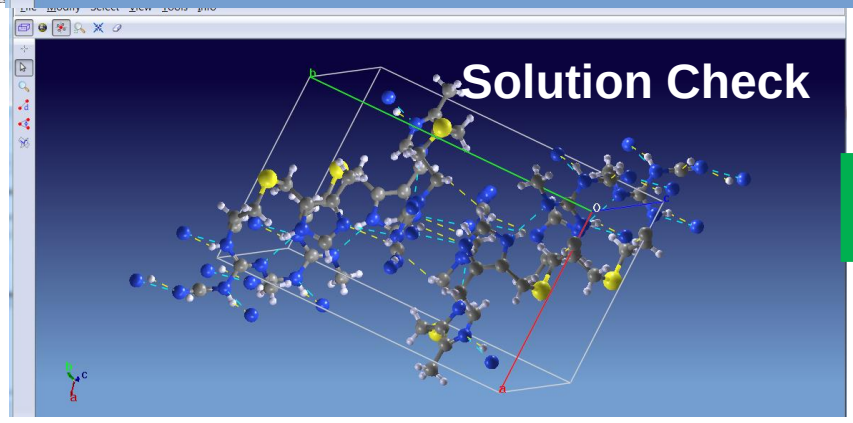




CIF FILE

# Conclusions

Powder diffraction is a powerful tool for solving structures of microcrystalline materials, assuming that cell parameters and space group have been determined.

If the quality of diffraction data (almost atomic resolution, limited peak overlap) and crystallinity are good, Direct Methods can correctly and automatically solve structures, organic, inorganic, metallorganic.

Direct space methods are widely used for solving both molecular and non-molecular structures. A crystallochemical inspection of the solution is mandatory. DFT (Density Functional Theory) calculations effectively support the solution process.

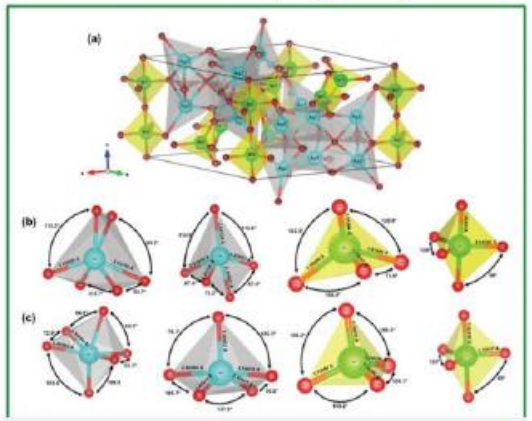
# EXPO software*

Free; Advanced methodologies; Simple use; Unique software; Automatism; User-friendly graphical interface; Authors' support

**> 1000 citations**

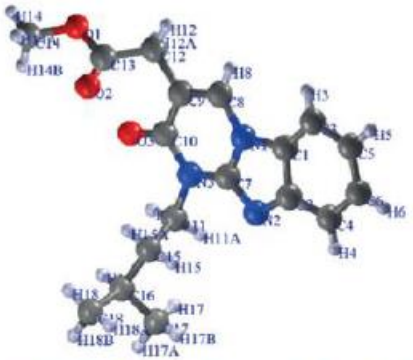
**> 9000 downloads**

metastable β-$Ag_2WO_4$*



Lemos, PS ; Altomare, A ; Gouveia, AF; Nogueira, IC; Gracia, L; Llusar, R; Andres, J; Longo, E; Cavalcante, (2016) LS DALTON TRANSACTIONS Volume: 45 Issue: 3 Pages: 1185-1191.

benzo[4,5]imidazo[1,2-*a*]pyrimidin-2(1*H*)-one



*R. Mancuso, L. Veltri, P. Russo, G. Grasso, C. Cuocci, R. Romeo, B. Gabriele*

Palladium-Catalyzed Carbonylative Synthesis of Functionalized Benzimidazopyrimidinones

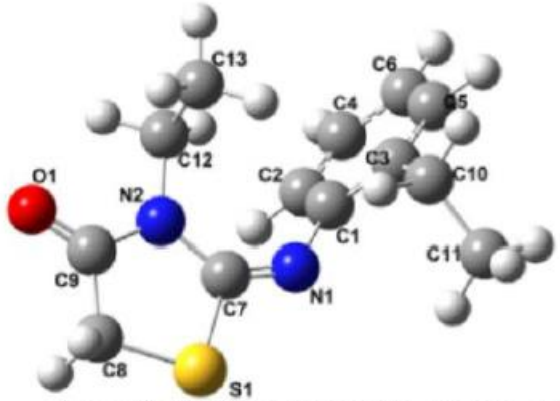
SYNTHESIS-STUTTGART

**Volume:** 50 **Issue:** 2 **Pages:** 267-277

**DOI:** 10.1055/s-0036-1591835

**Published:** JAN 2018

3N-ethyl-2N'-(2-ethylphenylimino) thiazolidin-4-one



Synthesis, PXRD structural determination, Hirshfeld surface analysis and DFT/TD-DFT investigation of 3N-ethyl-2N'-(2-ethylphenylimino) thiazolidin-4-one

Nour El Houda Belkafouf [a], Fayssal Triki Baara [b], Angela Altomare [c], Rosanna Rizzi [c], Abdelkader Chouaih [a,*], Ayada Djafri [b], Fodil Hamzaoui [d]

JOURNAL OF MOLECULAR STRUCTURE

**Volume:** 1189 **Pages:** 8-20

**DOI:** 10.1016/j.molstruc.2019.04.028

**Published:** AUG 5 2019

*Altomare, A.; Cuocci, C.; Giacovazzo, C.; Moliterni, A.; Rizzi, R.; Corriero, N; Falcicchio, A. EXPO2013: a kit of tools for phasing crystal structures from powder data. J. Appl. Cryst. 2013, 46, 1231-1235.

# EXPO: Registration and Download

Free software and documentation download for no-profit institution:

http://www.ba.ic.cnr.it/softwareic/expo/expo2014-download/

**Click on 'Register' and enter your personal data**

**After registration you will receive a confirmation e-mail and you will be allowed to the download**

**Feedback** **http://www.ba.ic.cnr.it/softwareic/expo/contact-us/**

Suggestions
Bug Reports
Future requests
Help

# Thank you for your kind attention