

AIC International Crystallography School 2018
BARI, 29 AUG - 02 SEP

Powder Diffraction

Theory, Software and Applications

Tutorial Session: *EXPO* software

Main Functions

- Indexing
- Space group determination
- Estimation of integrated intensities
- Structure solution by direct methods
- Structure model optimization
- Real space techniques
- Rietveld refinement
- Crystal structure visualization

Registration and download



Free download of software and documentation from

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

Supporting materials

<http://www.ba.ic.cnr.it/softwareic/expo/tutorials-and-lectures/>

- Presentation slides
- Examples

Technical details

- Program is built using OpenGL and GTK libraries
- Program written in Fortran and C++
- Licensed under terms of a Licence Agreement



System requirements



Windows 32-bit or 64-bit: Windows 7/8/10, Vista, XP
File *.exe

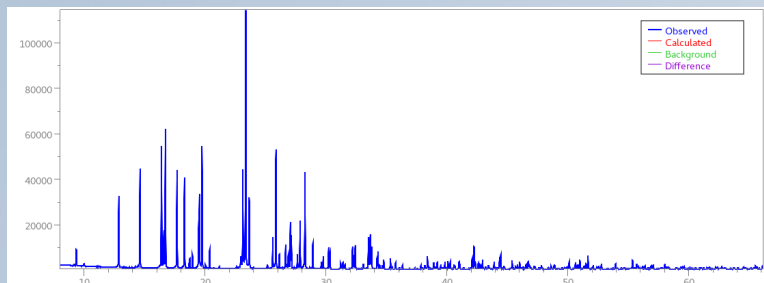


OS X: macOS 10.12 (Sierra), OS X 10.11 (El Capitan),
OS X 10.10 (Yosemite), Mac OS X 10.9 (Mavericks)
File *.dmg



GNU/Linux: Ubuntu 16.04 LTS, Ubuntu 14.04 LTS,
Fedora 24, Centos 7
File *.deb, *.rpm, *.tar.gz

Crystal Structure Determination Process



Experimental powder diffraction pattern

*Indexing &
space group
determination*



unit cell & space group

Structure solution
the biggest challenge



initial structural model

Rietveld method



final crystal structure

The indexing equation

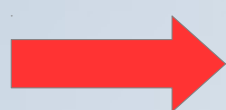
$$d_{hkl}^* = h a^* + k b^* + l c^*$$

$$d_{hkl}^* = \frac{1}{d_{hkl}} \quad a = \frac{b^* \wedge c^*}{V^*} \quad b = \frac{c^* \wedge a^*}{V^*} \quad c = \frac{a^* \wedge b^*}{V^*} \quad V^* = a^* \cdot b^* \wedge c^* = \frac{1}{V}$$



$$d_{hkl} = f(h, k, l, a, b, c, \alpha, \beta, \gamma)$$

Peak search



$$\theta_{hkl}^{obs}$$

Bragg's equation



$$d_{hkl}^{obs} = \frac{\lambda}{2} \sin \theta_{hkl}^{obs}$$

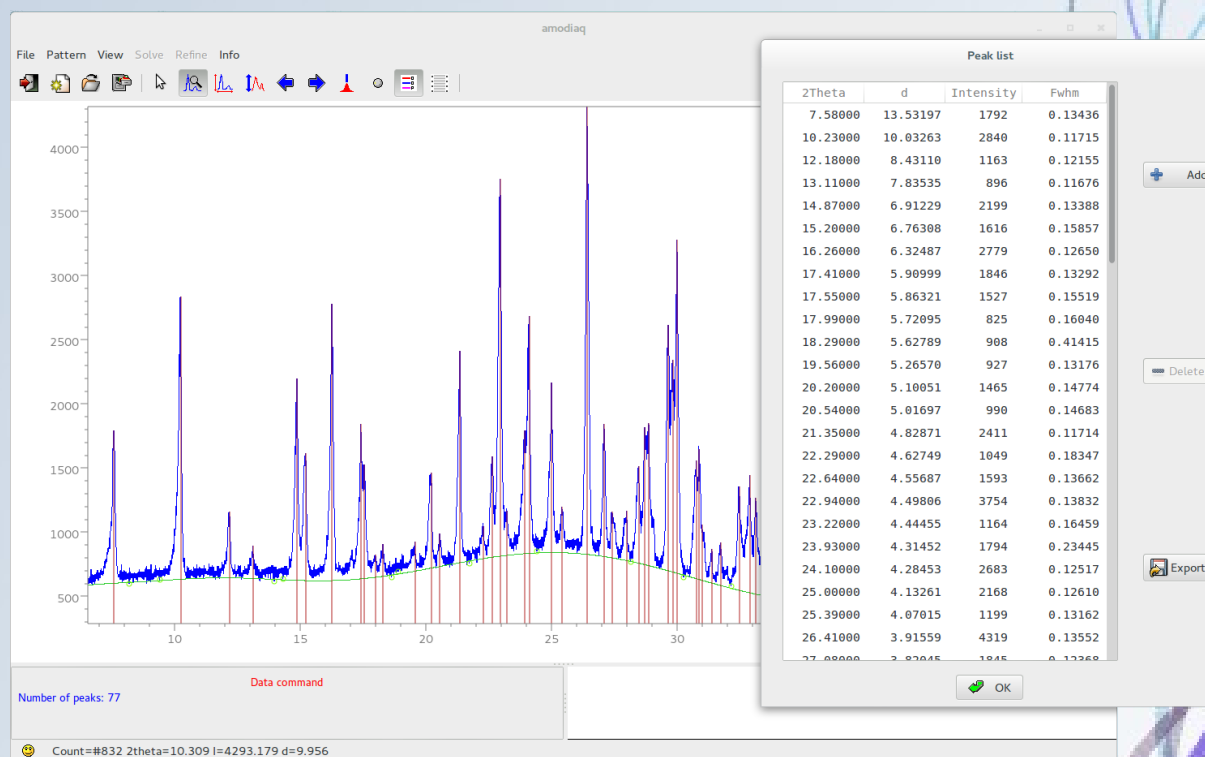
$$d_{hkl}^{obs} \simeq f(h, k, l, a, b, c, \alpha, \beta, \gamma)$$

Peak search

Peak search is needed in order to locate the peak positions

The method is a combinations of:

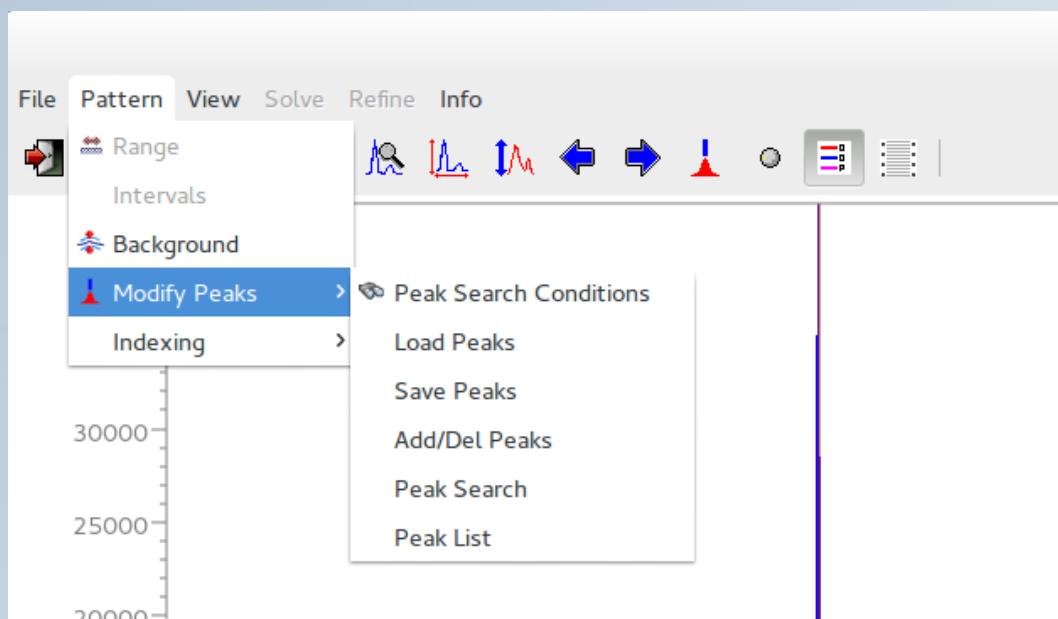
- background subtraction
- $K\alpha_2$ stripping
- smoothing
- calculation of the derivatives



25 first peaks can generally solve the indexing problem. Uncertain and not clearly resolved peaks should be not included. Try different selection of experimental peaks.

Peak search

Graphical tools are provided to improve the peak search



Set Search Conditions

Minimum 2-theta:

Maximum 2-theta:

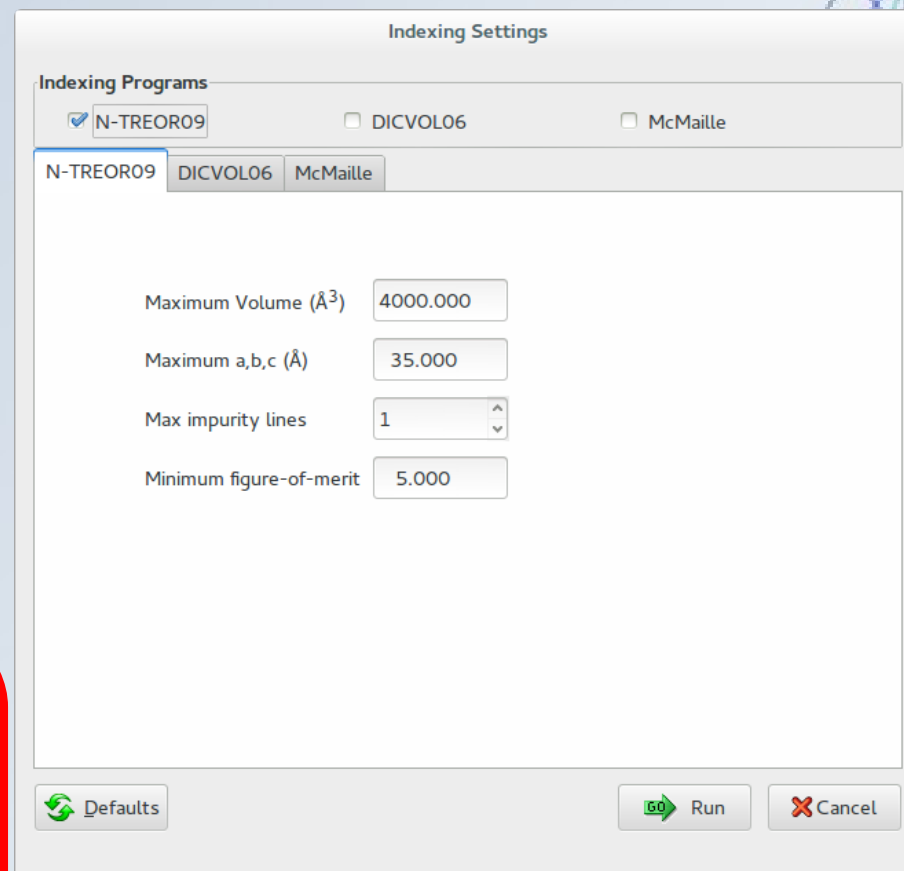
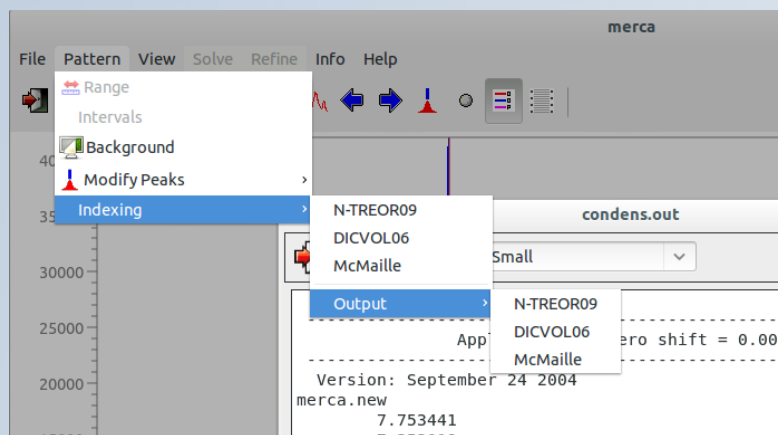
% Int. Threshold:

Number of Peaks:

Indexing by N-TREOR09

- Maximum volume and cell dimensions
- Maximum number of impurity lines
- Minimum figure of merit

Output files: condens.out, ntreor.out



DICVOL

DICVOL is an exhaustive trial-and-error indexing program with variation of parameters by successive dichotomy and partitioning of the unit cell volume.

The screenshot shows the 'Indexing Settings' dialog box for the DICVOL program. It features several sections: 'Indexing Programs' with checkboxes for N-TREOR09, DICVOL06 (selected), and McMaille; a tabbed interface with 'DICVOL06' selected; 'Number of lines used for searching solutions' set to 20; input fields for Volume (Å³), a,b,c (Å), B (°), Peak position error, Measured density, Molecular weight, Max impurity lines, and Minimum figure-of-merit; a 'Crystal System' section with checkboxes for Cubic, Hexagonal, Monoclinic, Tetragonal, Orthorhombic, and Triclinic; a 'Zero-point' section with checkboxes for Refine zero-point, Estimate zero-point, and Experimental zero-point; and an 'Exhaustive search' checkbox. At the bottom are buttons for 'Defaults', 'Run', and 'Cancel'.

- GUI may also be used to set maximum lengths of the unit cell edges, monoclinic angle and unit cell volume.
- Error in peak positions can be supplied and is assumed identical for every observed peak.
- It is possible to increase a minimum figure of merit, M_{20} (the default is 10)
- Any crystal system can be included or excluded from the indexing process
- Measured density and formula weight
- Maximum number of unindexed Bragg peaks
- Estimate and correct for the presence of zero-shift error

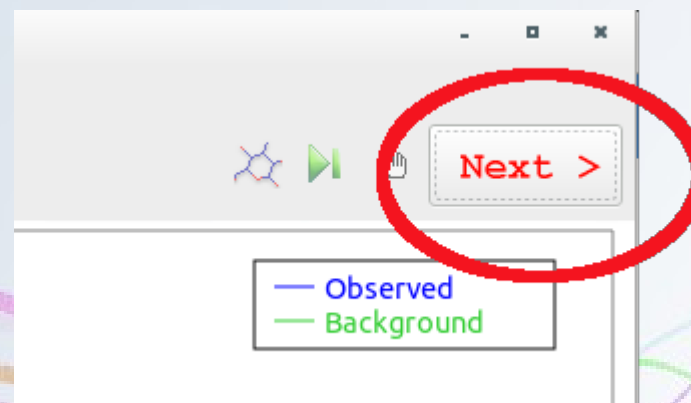
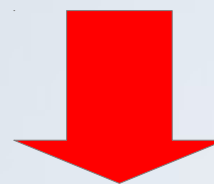
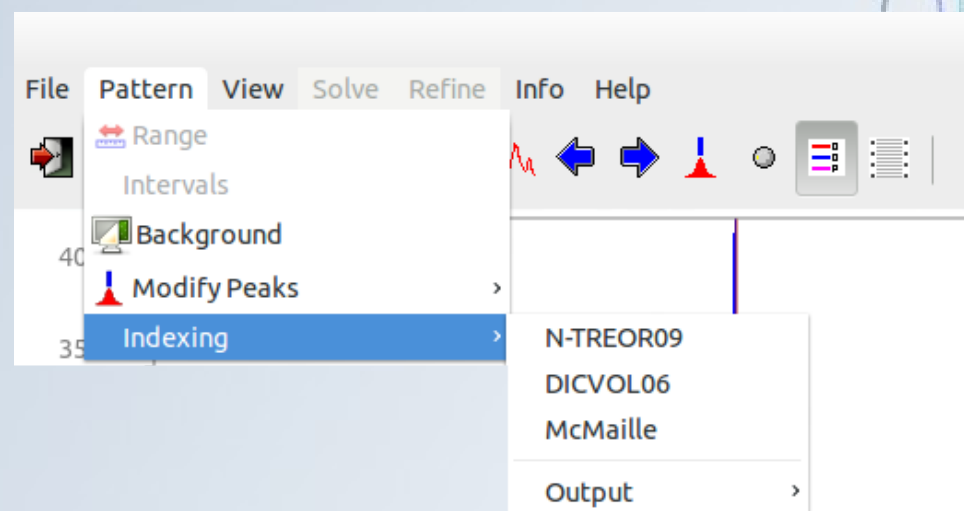
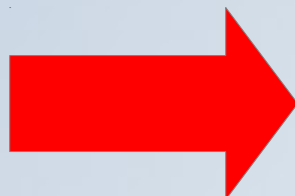
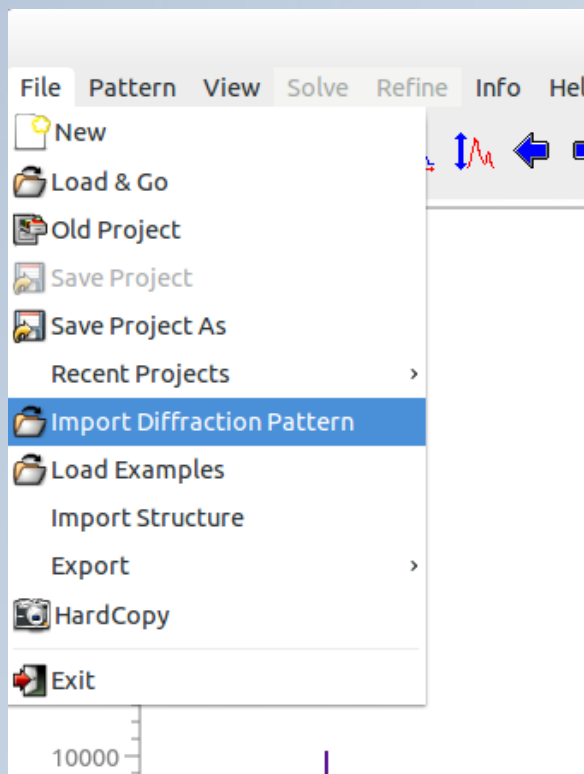
McMaille

The screenshot shows the 'Indexing Settings' dialog box for McMaille. At the top, under 'Indexing Programs', there are three checkboxes: 'N-TREOR09' (unchecked), 'DICVOL06' (unchecked), and 'McMaille' (checked). Below this is a tabbed interface with three tabs: 'N-TREOR09', 'DICVOL06', and 'McMaille' (selected). The 'McMaille' tab contains two sections. The first section, 'Option for black box mode', has six radio button options: 'Monte Carlo (NGrid=0)', 'Grid search (NGrid=1)', 'Monte Carlo + Grid search (NGrid=2)', 'Monte Carlo on all symmetries (NGrid=3)', 'Monte Carlo on all symmetries but without triclinic search (NGrid=3)' (selected), and 'Monte Carlo on all symmetries + grid search (NGrid=4)'. The second section, 'Crystal System', has six checkboxes: 'Cubic' (checked), 'Hexagonal' (checked), 'Monoclinic' (checked), 'Tetragonal' (checked), 'Orthorhombic' (checked), and 'Triclinic' (unchecked). Below these are two rows of input fields for volume and cell parameters. The first row is 'Volume (Å³)' with 'Min' set to 20.000 and 'Max' set to 4000.000. The second row is 'a,b,c (Å)' with 'Min' set to 0.000 and 'Max' set to 30.000. At the bottom, there are three buttons: 'Defaults' (with a circular arrow icon), 'Run' (with a green arrow icon), and 'Cancel' (with a red X icon).

- Treor and Dicvol run much faster than McMaille.
- The recommendation would be to start with TREOR or DICVOL, and finally McMaille if no convincing result is obtained.

Run Expo2014 for indexing

by **graphic interface**

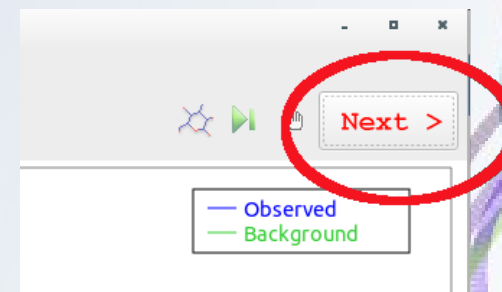
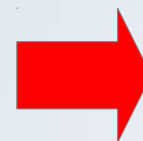
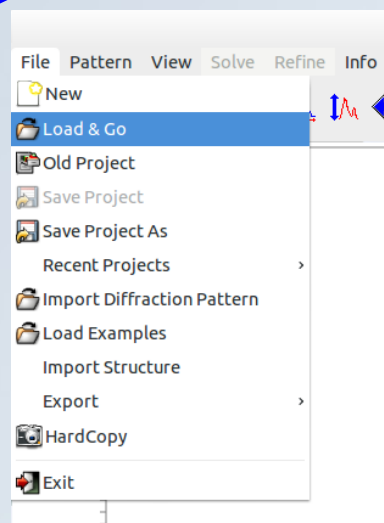


Run Expo2014 for indexing

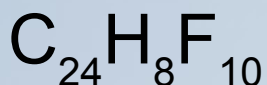
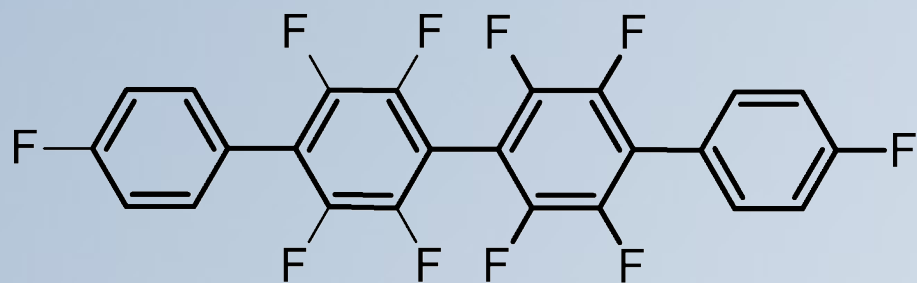
via an **ASCII input file (*.exp)** requiring minimal information and consisting of:

- **commands** (the first character in the line must be '%')
- **directives** (sub-commands following the related command)

```
%structure dfqp
%job decafluoroquarterphenyl (C24 H8 F10)
%data
  pattern dfqp.dat
  wave 1.790000
  range 7.000 70.000
%ntreor
%continue
```



Indexing of decafluoroquarterphenyl



decafluoroquarterphenyl

```
%structure dfqp
%job decafluoroquarterphenyl
%data
  pattern dfqp.dat
  wave 1.790000
```

```
%ntreor
%dicvol
```

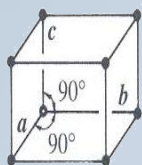
Published cell *

a: 24.0519

b: 6.1529

c: 12.4207


β : 102.755





Plausible cell parameters

Select cell

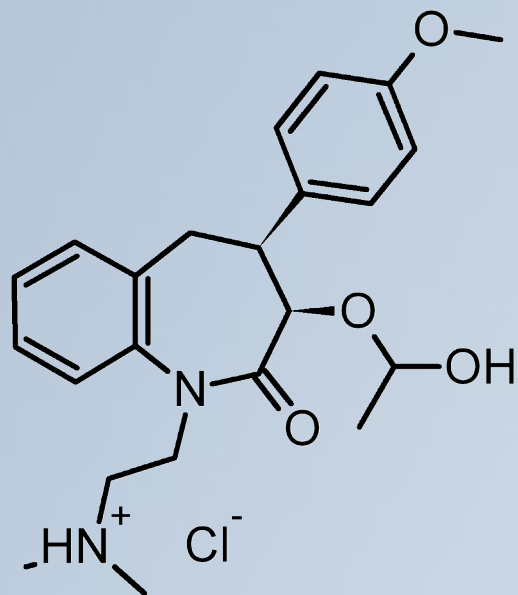
| Nr. | Prog. | a | b | c | alpha | beta | gamma | Vol. | M20 | FOMnew | Mc20 | shift | NIX | Symmetry Info |
|-----|-------|----------|---------|----------|--------|---------|--------|--------|-------|--------|------|--------|------------|---------------|
| 1 | N | 24.06143 | 6.14687 | 12.42651 | 90.000 | 102.739 | 90.000 | 1792.7 | 10.00 | 2.306 | - | -0.040 | 0 I 1 a 1 | |
| 2 | N | 24.10209 | 6.16368 | 12.45434 | 90.000 | 102.759 | 90.000 | 1804.5 | 10.00 | 2.298 | - | 0.040 | 0 I 1 a 1 | |
| 3 | N | 23.51948 | 6.15050 | 6.22038 | 90.000 | 92.246 | 90.000 | 899.1 | 22.00 | 1.931 | - | 0.000 | 5 P 1 21 1 | |
| 4 | D | 24.09430 | 6.14870 | 12.44520 | 90.000 | 102.720 | 90.000 | 1798.5 | 11.50 | - | - | -0.002 | 0 Mono | |

OK

Cancel

Export

Indexing of *diltiazem hydrochloride*



$C_{22}H_{27}N_2O_4S \cdot Cl$
Diltiazem hydrochloride

Change default settings by GUI
or adding the directive CEM in
the input file

Published cell

a: 42.190
b: 9.075
c: 6.037

$a \neq b \neq c$

Default maximum cell
axis value = **35 Å**

Indexing Settings

Indexing Programs

☒ N-TREOR09 ☐ DICVOL06 ☐ McMaille

N-TREOR09 DICVOL06 McMaille

Maximum Volume (Å³) 4000.000

Maximum a,b,c (Å) **50.0**

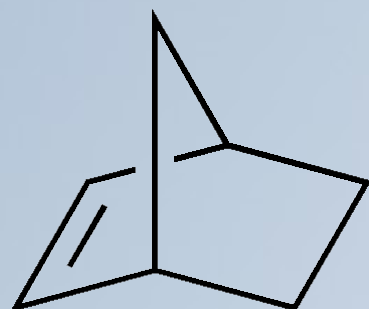
Max impurity lines 1

Minimum figure-of-merit 5.000

Run Cancel

```
%structure diltia
%job diltiazem
%data
pattern pd_0029.pow
wave 1.540560
%ntreor
CEM=50,
```

Indexing of norbornene



$C_{10}H_{10}$
norbornene

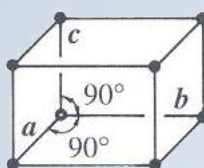
Published cell *

a : 7.6063

b : 8.6220

c : 8.749

β : 97.24



Indexing Settings

Indexing Programs

☒ N-TREOR09 ☐ DICVOL06 ☐ McMaille

N-TREOR09 DICVOL06 McMaille

Maximum Volume (\AA^3) 4000.000

Maximum a,b,c (\AA) 35.000

Max impurity lines 2

Minimum figure-of-merit 5.000

Defaults Run Cancel

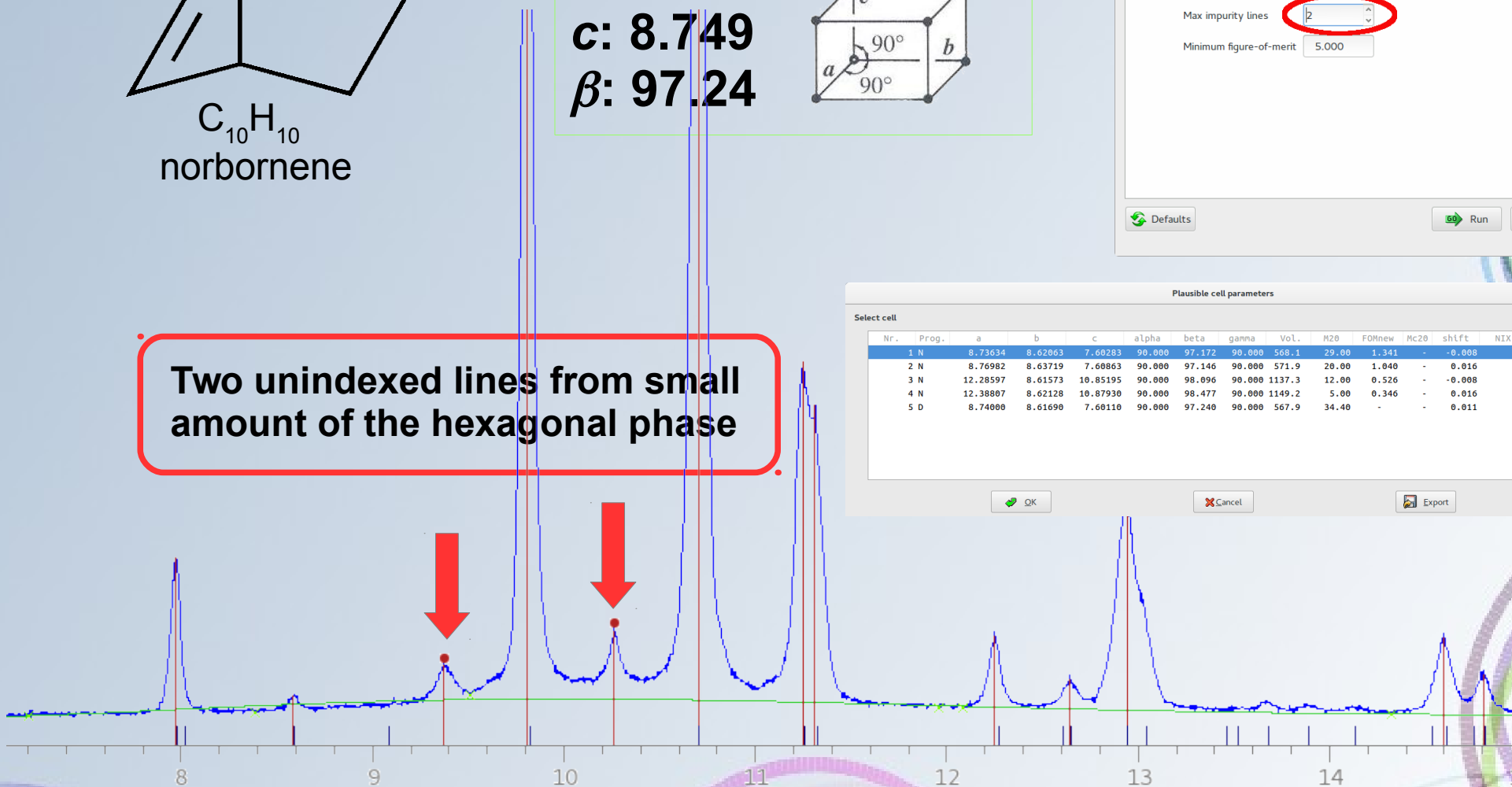
Two unindexed lines from small amount of the hexagonal phase

Plausible cell parameters

Select cell

| Nr. | Prog. | a | b | c | alpha | beta | gamma | Vol. | M20 | FOHnew | Mc20 | shift | NIX | Symmetry | Id |
|-----|-------|----------|---------|----------|--------|--------|--------|--------|-------|--------|------|--------|-----|--------------|----|
| 1 | N | 8.73634 | 8.62063 | 7.60283 | 90.000 | 97.172 | 90.000 | 568.1 | 29.00 | 1.341 | - | -0.008 | | 2 P 1 21/a 1 | |
| 2 | N | 8.76982 | 8.63719 | 7.60863 | 90.000 | 97.146 | 90.000 | 571.9 | 20.00 | 1.040 | - | -0.016 | | 2 P 1 21/a 1 | |
| 3 | N | 12.28597 | 8.61573 | 10.85195 | 90.000 | 98.096 | 90.000 | 1137.3 | 12.00 | 0.526 | - | -0.008 | | 1 P 1 21 1 | |
| 4 | N | 12.38807 | 8.62128 | 10.87930 | 90.000 | 98.477 | 90.000 | 1149.2 | 5.00 | 0.346 | - | -0.016 | | 0 P 1 21 1 | |
| 5 | D | 8.74000 | 8.61690 | 7.60110 | 90.000 | 97.240 | 90.000 | 567.9 | 34.40 | - | - | 0.011 | | 3 Mono | |

OK Cancel Export



*M. Brunelli et al. "Crystal and Molecular Structures of Norbornene." Zeitschrift für Kristallographie, 2001 : 51–55.
<https://doi.org/10.1524/zkri.216.1.51.18996>

Errors in powder indexing

The solution is not always straightforward because of several reasons:

- Inaccuracy in peak positions due to:
 - 1.zero-point error
 - 2.sample misplacement
 - 3.low resolution
 - 4.bad crystallinity
- The presence of impurities providing spurious additional peaks

Data quality remains the most important factor in indexing

Some advice

- Do not accept unindexed peaks, unless you are able to explain them
- To obtain a proper solution, it is necessary to eliminate the impurity Bragg peaks from the indexing process
- Every observed peak must correspond to a calculated reflection
- Finally, in the case of a new material, the correctness of the *ab initio* indexing is generally ensured by solving and refining the crystal structure
- Do not waste computer time on bad data

Space group determination

**Analysis of the systematically
absent reflections**



Extinction symbol (ES)

| <i>Cryst. system</i> | <i>Ext. symb</i> | <i>Space groups</i> |
|---------------------------------|-----------------------------|---|
| <i>Mon.</i> | <i>P1 - 1</i> | <i>P2, Pm, P2/m</i> |
| <i>Mon.</i> | <i>P1 2₁ 1</i> | <i>P2₁, P2₁/m</i> |
| <i>Mon.</i> | <i>P1 2₁/c 1</i> | <i>P2₁/c</i> |
| <i>Orth.</i> | <i>P- - -</i> | <i>P222, Pm2m, P2mm, Pmm2, Pmmm</i> |
| <i>Tetr.</i> | <i>P- - -</i> | <i>P4, P$\bar{4}$, P4/m, P422, P4mm, P$\bar{4}$2m, P$\bar{4}$m2, P$\bar{4}$2₁m, P4/mmm</i> |

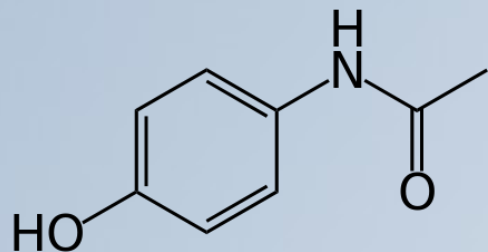
The ES does not unambiguously define the space group

The Expo procedure for space group determination

This may be synthesized into the following steps:

1. The experimental powder diffraction diagram is decomposed *via* the **Le Bail algorithm** into single diffraction intensities in the space group having the largest Laue symmetry and no extinction conditions (e.g., P2/m for monoclinic, P2/m2/m2/m for orthorhombic, P4/mmm for tetragonal, P6/mmm for trigonal-hexagonal systems, and Pm3m for the cubic system).
2. The normalized intensities $z_h = |E_h|^2$ are submitted to **statistical analysis** for the determination of the space group symmetry.
3. The algorithm provides a probability value for each **extinction symbol** compatible with the lattice symmetry established by the indexing procedure.

Indexing and space group determination of paracetamol



Paracetamol (form I polymorph)
 $C_8H_9NO_2$

Plausible cell parameters

Select cell

| Nr. | Prog. | a | b | c | alpha | beta | gamma | Vol. | M20 | FOMnew | Mc20 | shift | NIX | Symmetry Info |
|-----|-------|----------|---------|---------|--------|--------|--------|-------|-------|--------|------|-------|-----|---------------|
| 1 | N | 11.70872 | 9.38085 | 7.09982 | 90.000 | 97.412 | 90.000 | 773.3 | 53.00 | 4.875 | - | 0.000 | 0 | P 1 21/n 1 |
| 2 | D | 11.70610 | 9.37910 | 7.09960 | 90.000 | 97.413 | 90.000 | 773.0 | 62.00 | - | - | 0.006 | 0 | Mono |

OK Cancel Export

Missing Information

Cell Parameters

a: 11.706100 b: 9.379100 c: 7.099600 α : 90.0000 β : 97.41300 γ : 90.0000

Volume: 772.969

Space Group

☒ Find Space Group

Space Group Symbol:

P 2/m

Cell Content:

Content Volume: Volume per Atom: Density:

OK

Cell content

Cell content = $Z \cdot$ molecular formula

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

V_{mol} = volume of molecule

$$\mathbf{18 \text{ \AA}^3 \text{ rule: } V_{\text{mol}} = M \cdot 18}$$

M = number of non-hydrogen atoms

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

Accurate estimation of molecular volume

$$V_{\text{mol}} = \sum n_i v_i$$

n_i = number of atoms of the i^{th} type in the structure

v_i = volume contribution (in \AA^3) for the i^{th} atom type

Some approximate atomic volumes:

$$v_{\text{H}} = 5 \text{ \AA}^3$$

$$v_{\text{C}} = 14 \text{ \AA}^3$$

$$v_{\text{N}} = 12 \text{ \AA}^3$$

$$v_{\text{O}} = 11 \text{ \AA}^3$$

$$v_{\text{S}} = 25 \text{ \AA}^3$$

$$v_{\text{F}} = 11 \text{ \AA}^3$$

Complete list of average volume of the atoms in appendix A:

<http://www.ba.ic.cnr.it/softwareic/expo/average-volume-of-the-atoms/>

Cell content of paracetamol

Chemical formula: $C_8H_9NO_2$

Volume of unit cell = $V_{\text{cell}} = 773 \text{ \AA}^3$

- **18 \AA^3 rule**

$$Z = \frac{V_{\text{cell}}}{18 \cdot M} = \frac{773}{18 \cdot 11} = 3.90 \sim 4$$

- **Accurate method**

$$V_{\text{mol}} = \sum n_i v_i = 8 v_c + 9 v_H + v_N + 2 v_O = 8 \cdot 14 + 9 \cdot 5 + 12 + 2 \cdot 11 = 191 \text{ \AA}^3$$

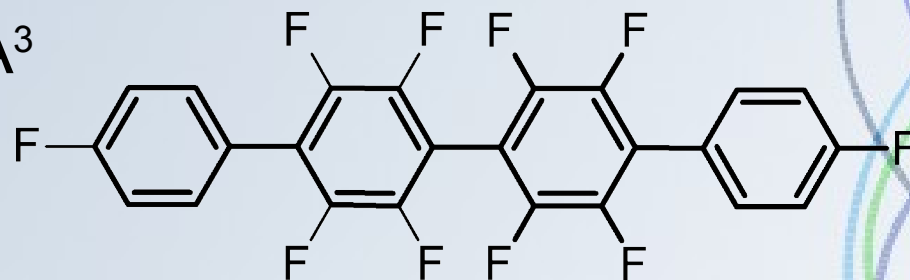
$$Z = \frac{V_{\text{cell}}}{V_{\text{mol}}} = \frac{773}{191} = 4.05$$

Cell content: $(C_8H_9NO_2)_4$ or $C_{32}H_{36}N_4O_8$

Cell content of decafluoroquarterphenyl

Chemical formula: $C_{24}H_8F_{10}$

Volume of unit cell = $V_{\text{cell}} = 1799 \text{ \AA}^3$



■ 18 \AA^3 rule

$$Z = \frac{V_{\text{cell}}}{18 \cdot M} = \frac{1799}{18 \cdot 34} = 2.90 \sim 2 \text{ or } 4$$

■ Accurate method

$$V_{\text{mol}} = \sum n_i v_i = 24 v_C + 8 v_H + 10 v_F = 24 \cdot 14 + 8 \cdot 5 + 10 \cdot 11 = 486 \text{ \AA}^3$$

$$Z = \frac{V_{\text{cell}}}{V_{\text{mol}}} = \frac{1799}{486} = 3.70 \sim 4$$

Cell content: $(C_{24}H_8F_{10})_4$ or $C_{96}H_{32}F_{40}$

Space group determination of paracetamol

Missing Information [X]

Cell Parameters

a: 11.706100 b: 9.379100 c: 7.099600 α : 90.0000 β : 97.41300 γ : 90.0000

Volume: 772.969

Space Group

☒ Find Space Group

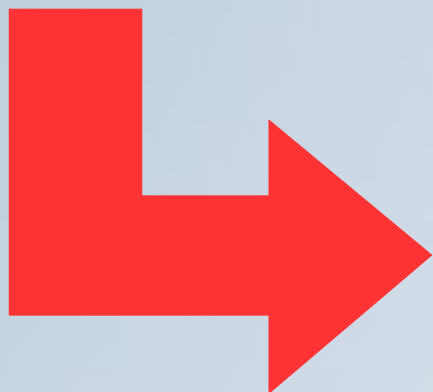
Cell Content: (C8 H9 N O2)4

Content Volume: 765.040 Volume per Atom: 17.567 Density: 1.299

Space Group Symbol:

P 2/m

OK



Find space group [X]

| Space Group | Extinction symbol | FoM | Nabs | Nasym | No. in CSD | % of CSD | Rank | Chiral |
|-------------|-------------------|-------|------|-------|------------|----------|------|--------|
| P 21/n | P 1 21/n 1 | 0.577 | 16 | 11 | 279041 | 34.57 | 1 | no |
| P 2/n | P 1 n 1 | 0.292 | 15 | 11 | 5232 | 0.65 | 14 | no |
| P n | P 1 n 1 | 0.292 | 15 | 22 | 3447 | 0.43 | 18 | no |
| P 21 | P 1 21 1 | 0.087 | 1 | 22 | 41791 | 5.18 | 5 | yes |
| P 21/m | P 1 21 1 | 0.087 | 1 | 11 | 4023 | 0.50 | 17 | no |
| P 2 | P 1 - 1 | 0.044 | 0 | 22 | 142 | 0.02 | 96 | yes |
| P 2/m | P 1 - 1 | 0.044 | 0 | 11 | 110 | 0.01 | 111 | no |
| P m | P 1 - 1 | 0.044 | 0 | 22 | 21 | 0.00 | 202 | no |
| P 21/c | P 1 21/c 1 | 0.024 | 15 | 11 | 279041 | 34.57 | 1 | no |
| P 2/c | P 1 c 1 | 0.012 | 14 | 11 | 5232 | 0.65 | 14 | no |
| P c | P 1 c 1 | 0.012 | 14 | 22 | 3447 | 0.43 | 18 | no |
| P 21/a | P 1 21/a 1 | 0.010 | 16 | 11 | 279041 | 34.57 | 1 | no |
| P 2/a | P 1 a 1 | 0.005 | 15 | 11 | 5232 | 0.65 | 14 | no |
| P a | P 1 a 1 | 0.005 | 15 | 22 | 3447 | 0.43 | 18 | no |
| I 2 | I 1 - 1 | 0.000 | 58 | 11 | 6826 | 0.85 | 12 | yes |
| I 2/m | I 1 - 1 | 0.000 | 58 | 6 | 4094 | 0.51 | 16 | no |
| I m | I 1 - 1 | 0.000 | 58 | 11 | 293 | 0.04 | 69 | no |
| I 2/a | I 1 a 1 | 0.000 | 65 | 6 | 67434 | 8.35 | 3 | no |

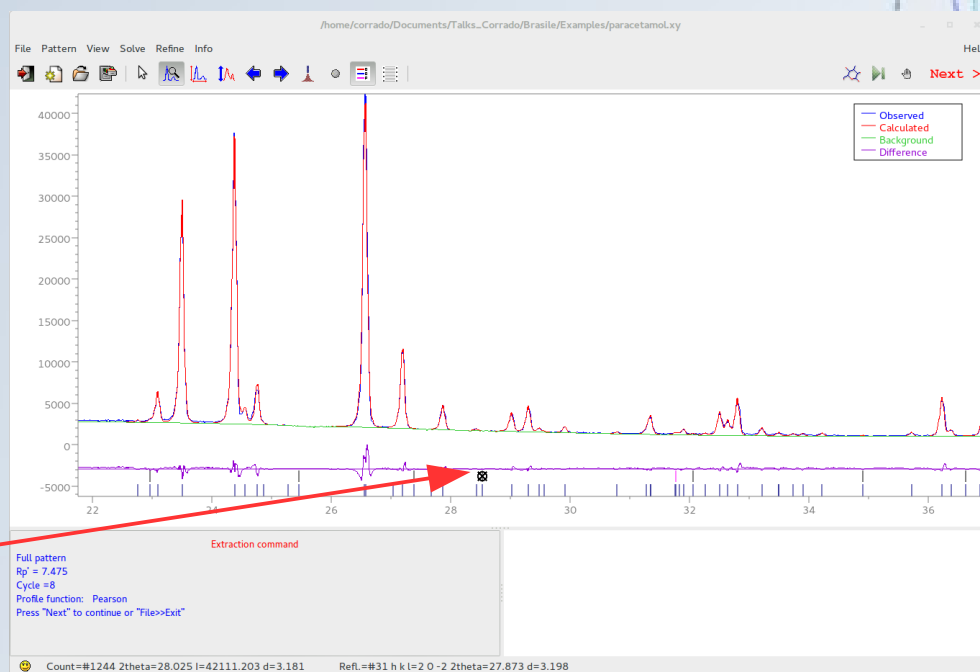
List OK Cancel

Space group determination of paracetamol

Find space group

| Space Group | Extinction symbol | FoM | Nabs | Nasym | No. in CSD | % of CSD | Rank | Chiral |
|-------------|-------------------|-------|------|-------|------------|----------|------|--------|
| P 21/n | P 1 21/n 1 | 0.577 | 16 | 11 | 279041 | 34.57 | 1 | no |
| P 2/n | P 1 n 1 | 0.292 | 15 | 11 | 5232 | 0.65 | 14 | no |
| P n | P 1 n 1 | 0.292 | 15 | 22 | 3447 | 0.43 | 18 | no |
| P 21 | P 1 21 1 | 0.087 | 1 | 22 | 41791 | 5.18 | 5 | yes |
| P 21/m | P 1 21 1 | 0.087 | 1 | 11 | 4023 | 0.50 | 17 | no |
| P 2 | P 1 - 1 | 0.044 | 0 | 22 | 142 | 0.02 | 96 | yes |
| P 2/m | P 1 - 1 | 0.044 | 0 | 11 | 110 | 0.01 | 111 | no |
| P m | P 1 - 1 | 0.044 | 0 | 22 | 21 | 0.00 | 202 | no |
| P 21/c | P 1 21/c 1 | 0.024 | 15 | 11 | 279041 | 34.57 | 1 | no |
| P 2/c | P 1 c 1 | 0.012 | 14 | 11 | 5232 | 0.65 | 14 | no |
| P c | P 1 c 1 | 0.012 | 14 | 22 | 3447 | 0.43 | 18 | no |
| P 21/a | P 1 21/a 1 | 0.010 | 16 | 11 | 279041 | 34.57 | 1 | no |
| P 2/a | P 1 a 1 | 0.005 | 15 | 11 | 5232 | 0.65 | 14 | no |
| P a | P 1 a 1 | 0.005 | 15 | 22 | 3447 | 0.43 | 18 | no |
| I 2 | I 1 - 1 | 0.000 | 58 | 11 | 6826 | 0.85 | 12 | yes |
| I 2/m | I 1 - 1 | 0.000 | 58 | 6 | 4094 | 0.51 | 16 | no |
| I 2/n | I 1 - 1 | 0.000 | 58 | 11 | 293 | 0.04 | 69 | no |
| I 2/a | I 1 a 1 | 0.000 | 65 | 6 | 67434 | 8.35 | 3 | no |

List OK Cancel



List of systematically absent reflections

| Num | h | k | l | Extinction condition | Probability for extinction | Type |
|-----|---|---|----|----------------------|----------------------------|------------|
| 33 | 0 | 3 | 0 | 21 (0 k 0 : k) | 1.000 | Single |
| 143 | 0 | 5 | 0 | 21 (0 k 0 : k) | 0.766 | Overlapped |
| 29 | 2 | 0 | 1 | n (h 0 l : h + l) | 1.000 | Single |
| 80 | 3 | 0 | -2 | n (h 0 l : h + l) | 0.998 | Single |
| 106 | 3 | 0 | 2 | n (h 0 l : h + l) | 0.996 | Single |
| 154 | 1 | 0 | 6 | n (h 0 l : h + l) | 0.995 | Single |
| 271 | 4 | 0 | -5 | n (h 0 l : h + l) | 0.992 | Overlapped |
| 14 | 1 | 0 | 2 | n (h 0 l : h + l) | 0.940 | Single |
| 10 | 1 | 0 | -2 | n (h 0 l : h + l) | 0.984 | Single |
| 44 | 1 | 0 | -4 | n (h 0 l : h + l) | 0.972 | Overlapped |
| 71 | 3 | 0 | 0 | n (h 0 l : h + l) | 0.965 | Overlapped |
| 75 | 0 | 0 | 5 | n (h 0 l : h + l) | 0.956 | Overlapped |
| 185 | 4 | 0 | 1 | n (h 0 l : h + l) | 0.948 | Overlapped |
| 122 | 3 | 0 | -4 | n (h 0 l : h + l) | 0.947 | Overlapped |
| 190 | 4 | 0 | -3 | n (h 0 l : h + l) | 0.944 | Overlapped |
| 2 | 1 | 0 | 0 | n (h 0 l : h + l) | 0.992 | Single |
| 201 | 0 | 0 | 7 | n (h 0 l : h + l) | 0.940 | Overlapped |
| 47 | 2 | 0 | -3 | n (h 0 l : h + l) | 0.928 | Single |
| 16 | 0 | 0 | 3 | n (h 0 l : h + l) | 0.926 | Single |

Close

Space group determination of decafluoroquarterphenyl

```
%structure dfqp
%job decafluoroquarterphenyl (C24 H8 F 10)
%data
pattern dfqp.dat
wave 1.790000
cont (C24 H8 F10)4
cell 24.085501 6.155077 12.435071 90.00000 102.68274 90.00000
%continue
```

| Space Group | Extinction symbol | FoM | Nabs | Nasym | No. in CSD | % of CSD | Rank | Chiral |
|-------------|-------------------|-------|------|-------|------------|----------|------|--------|
| I 2/a | I 1 a 1 | 0.415 | 49 | 17 | 67434 | 8.35 | 3 | no |
| I a | I 1 a 1 | 0.415 | 49 | 34 | 8450 | 1.05 | 9 | no |
| I 2 | I 1 - 1 | 0.307 | 41 | 34 | 6826 | 0.85 | 12 | yes |
| I 2/m | I 1 - 1 | 0.307 | 41 | 17 | 4094 | 0.51 | 16 | no |
| I m | I 1 - 1 | 0.307 | 41 | 34 | 293 | 0.04 | 69 | no |
| P 21/a | P 1 21/a 1 | 0.136 | 18 | 34 | 279041 | 34.57 | 1 | no |
| P 21/c | P 1 21/c 1 | 0.126 | 20 | 34 | 279041 | 34.57 | 1 | no |
| P 21/n | P 1 21/n 1 | 0.116 | 21 | 34 | 279041 | 34.57 | 1 | no |
| P 2/a | P 1 a 1 | 0.058 | 17 | 34 | 5232 | 0.65 | 14 | no |
| P a | P 1 a 1 | 0.058 | 17 | 68 | 3447 | 0.43 | 18 | no |
| P 21 | P 1 21 1 | 0.056 | 1 | 68 | 41791 | 5.18 | 5 | yes |
| P 21/m | P 1 21 1 | 0.056 | 1 | 34 | 4023 | 0.50 | 17 | no |
| P 2/c | P 1 c 1 | 0.054 | 19 | 34 | 5232 | 0.65 | 14 | no |
| P c | P 1 c 1 | 0.054 | 19 | 68 | 3447 | 0.43 | 18 | no |
| P 2/n | P 1 n 1 | 0.049 | 20 | 34 | 5232 | 0.65 | 14 | no |

List

OK Cancel

Methods of Structure Solution

**Structure
solution**

```
graph TD; A((Structure solution)) --> B[Traditional approaches:]; A --> C[Direct space methods]; A --> D[Other methods:];
```

Traditional approaches:

- direct methods
- Patterson methods

Direct space methods

*Alternative words: real space,
global optimization, global search*

Other methods:

- charge flipping
- molecular replacement
-

A Typical Direct Methods Procedure

Scaling and normalization of the structure factors $\mathbf{F}_h \rightarrow \mathbf{E}_h$

Triplet and negative quartet invariants are found among the reflections with largest \mathbf{E}_h

Random phases assignment

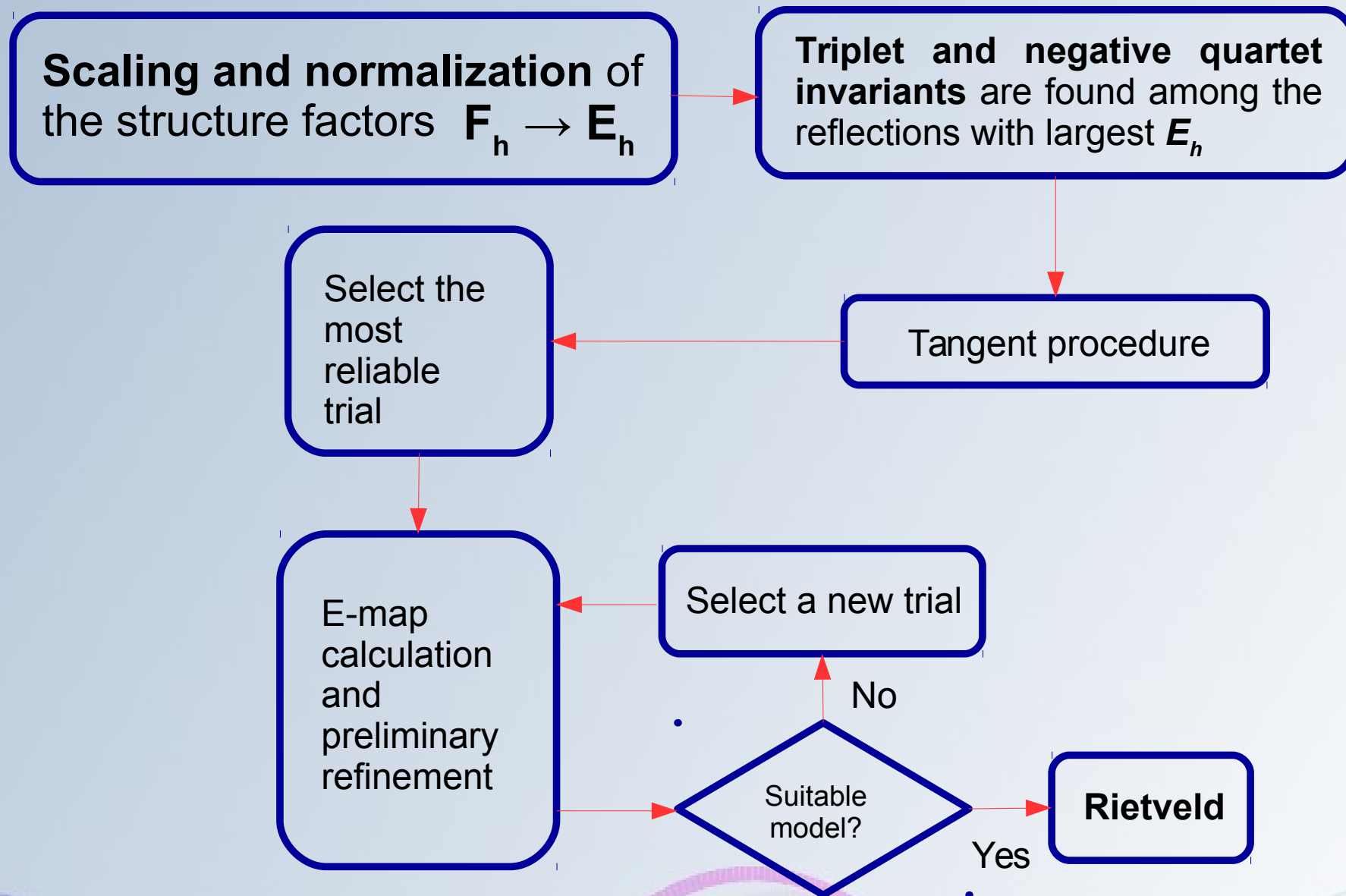
Cycles of tangent formula

New phase set (trial) and FOM

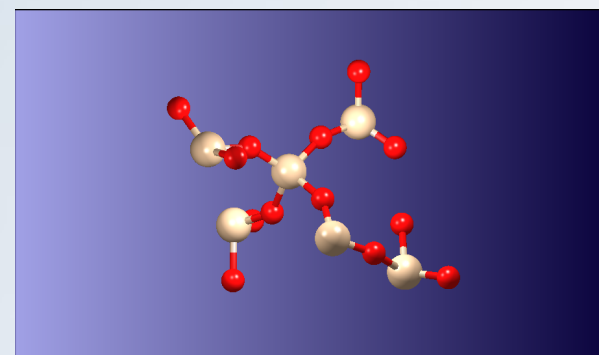
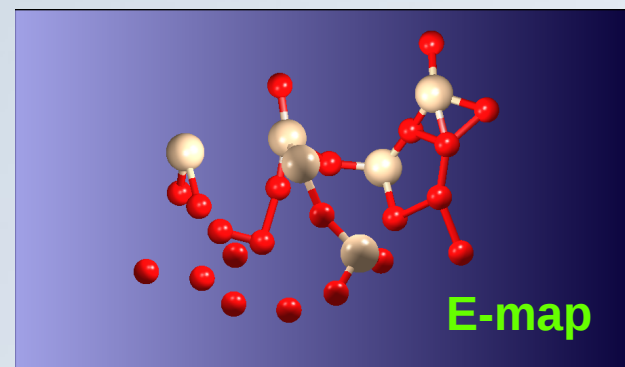
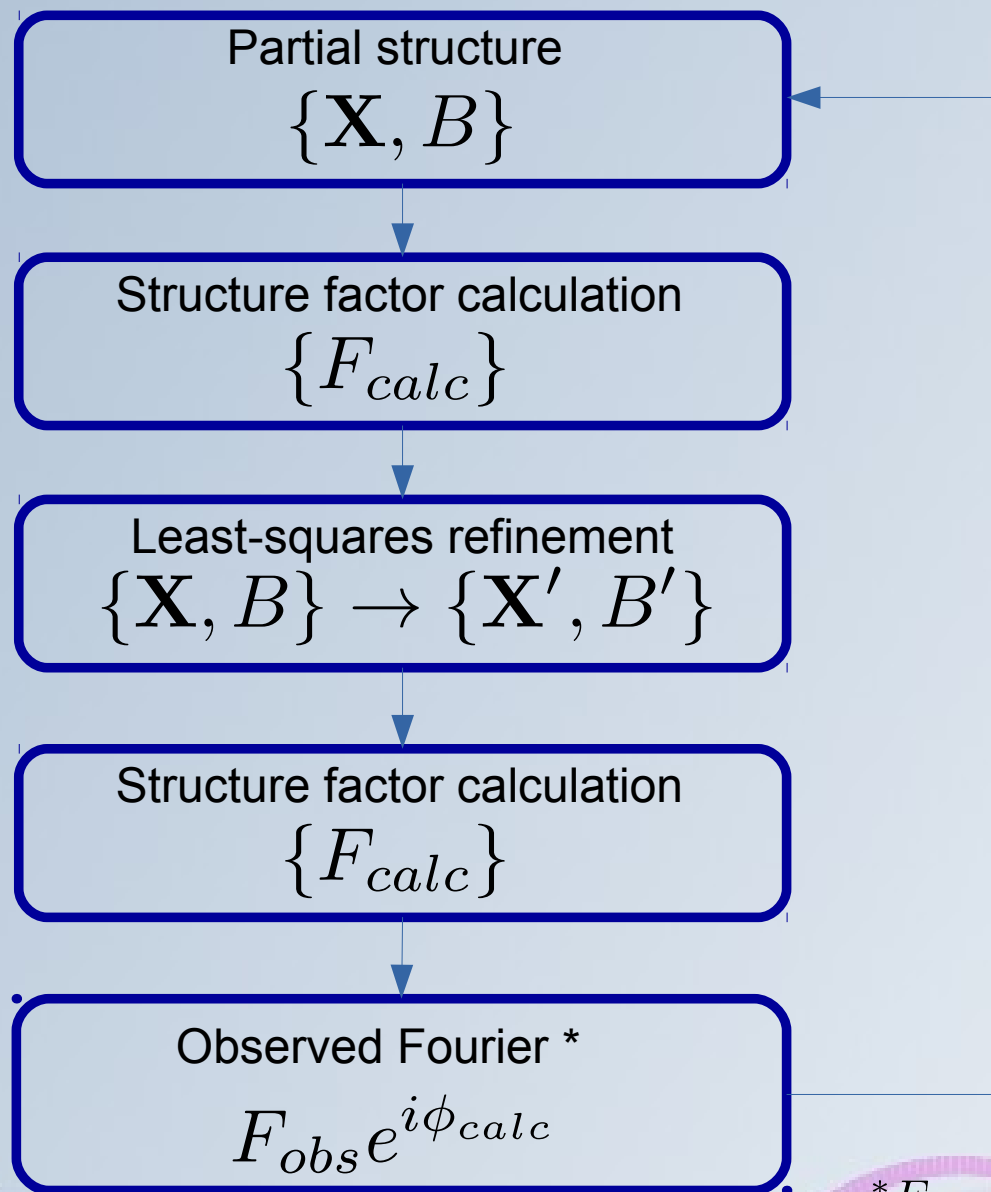
N_{\max} trial?

No

A Typical Direct Methods Procedure



Completion of the Crystal Structure and Preliminary Refinement



* F_{obs} may be replaced by $(F_{obs} - F_{calc})$ or by $(2F_{obs} - F_{calc})$

Resolution Bias Correction Algorithm (RBM)

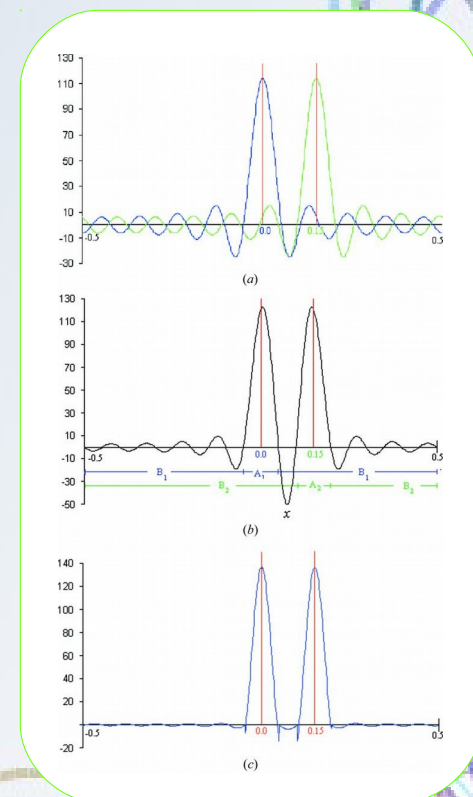
Electron density of a crystal structure $\rho(\mathbf{r}) = \sum_{j=1}^N \rho_j(\mathbf{r} - \mathbf{r}_j)$

Calculated electron-density map $\rho'(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{h}} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{r})$

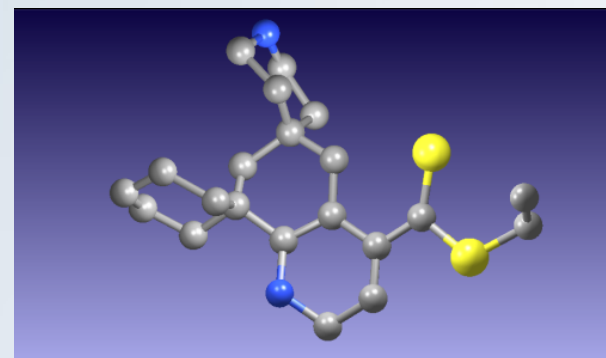
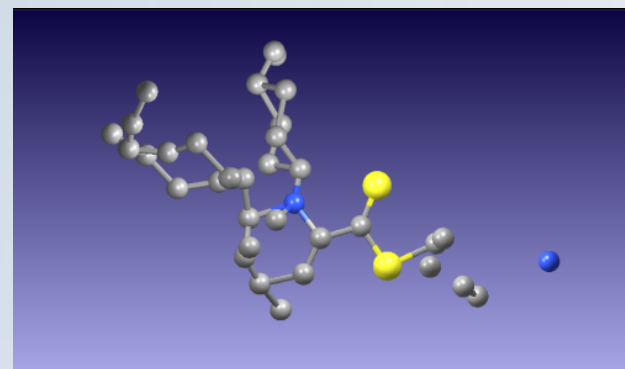
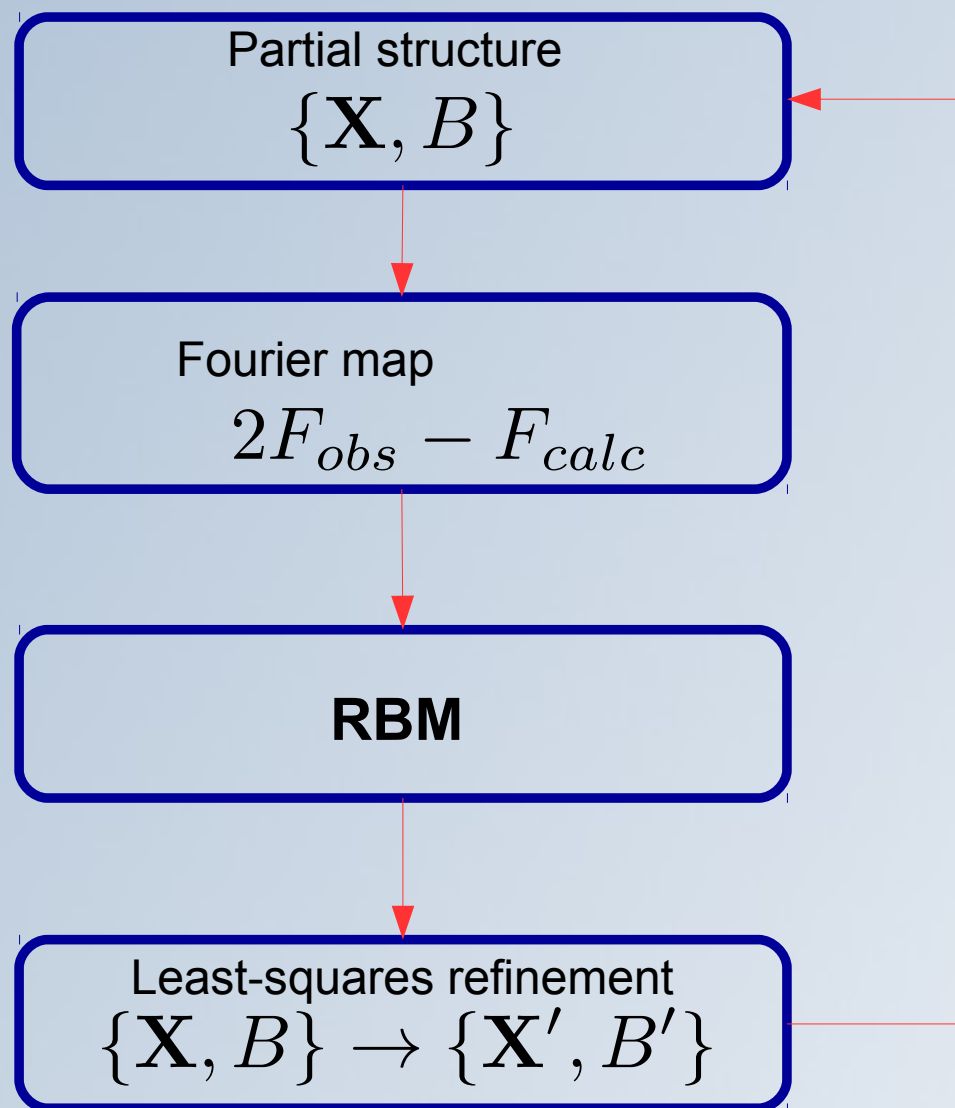
Features of $\rho'(\mathbf{r})$

- Negative in more or less extended regions
- Atomic peaks show deformed profile and are surrounded by series of negative and positive ripples
- Atomic peak are shifted from the correct position ($\mathbf{r}_j \rightarrow \mathbf{r}'_j$)

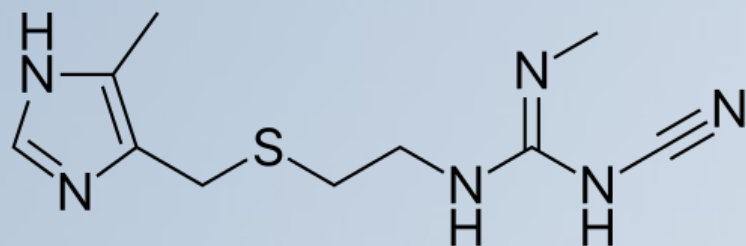
$$\rho'(\mathbf{r}) \xrightarrow{\text{RBM}} \rho'_{mod}(\mathbf{r}) \approx \rho(\mathbf{r})$$



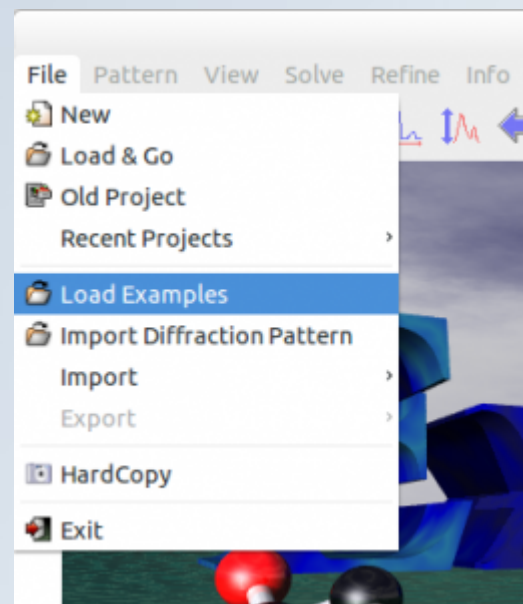
Default Strategy of Model Optimization in the EXPO Program



Crystal structure determination of Cimetidine compound*



cimetidine
 $C_{10}H_{16}N_6S$

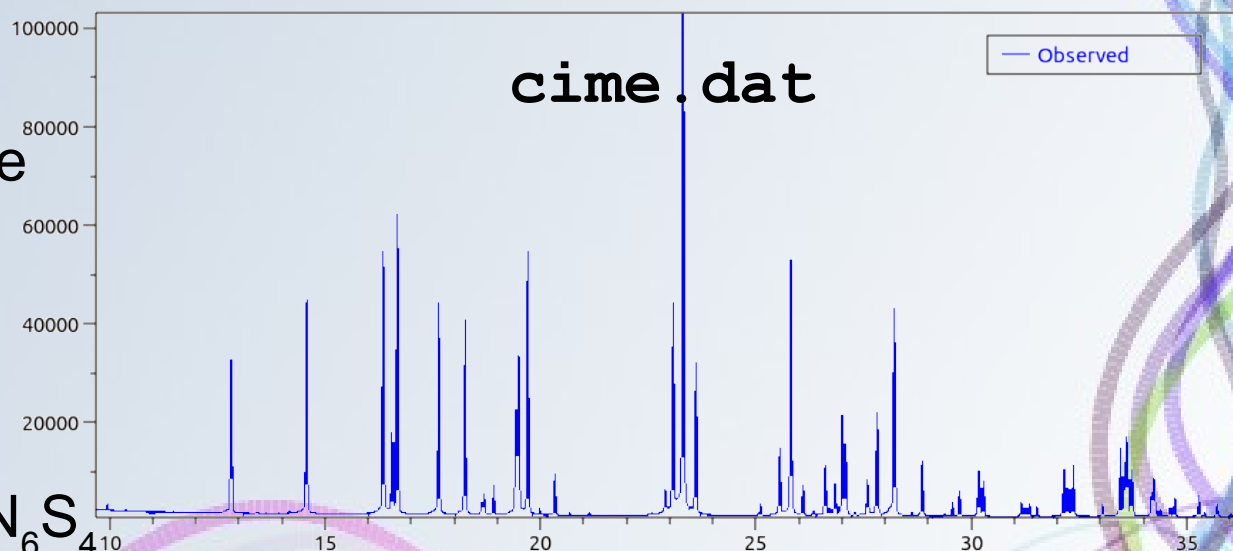


Required information:

- Powder diffraction data file

wavelength 1.52904
synchrotron

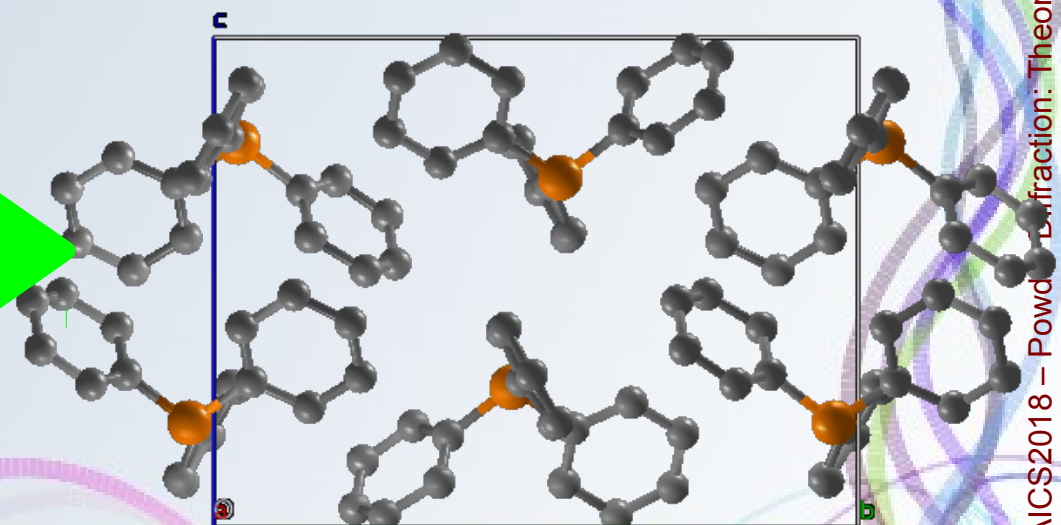
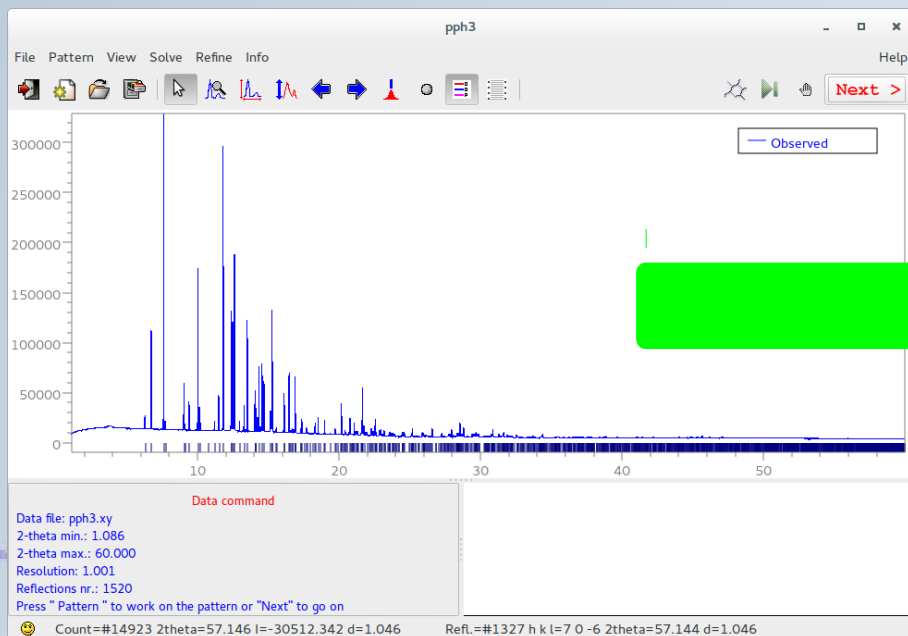
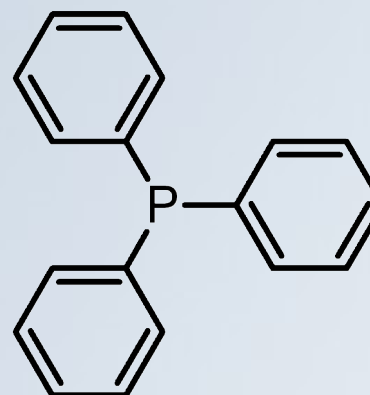
- Chemical formula: $C_{10}H_{16}N_6S$



*R. J. Cernik, A. K. Cheetham, C. K. Prout, D. J. Watkin, A. P. Wilkinson and B. T. M. Willis. *J. Appl. Cryst.* (1991). **24**, 222-226
<https://doi.org/10.1107/S0021889890013486>

Crystal structure solution of $\text{P}(\text{C}_6\text{H}_5)_3$

```
%structure pph3
%job triphenylphosphine
%data
  pattern pph3.xy
  wave 1.000972
  synchrotron
  range -1 60
  space P 21/c
  cell 8.495202 15.028483 11.430653 90.0 92.89145 90.0
  content (P (C6H5) 3) 4
%continue
```



Crystal structure determination of famotidine*

```
%job famotidine (C8 H15 N7 O2 S3)
```

```
%structure famo
```

```
%init
```

```
%data
```

```
pattern famo.xy
```

```
wave 0.8507473
```

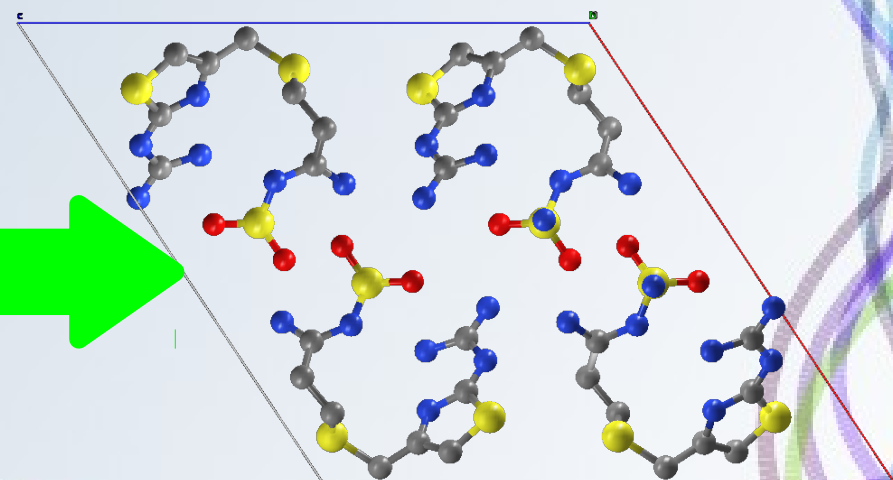
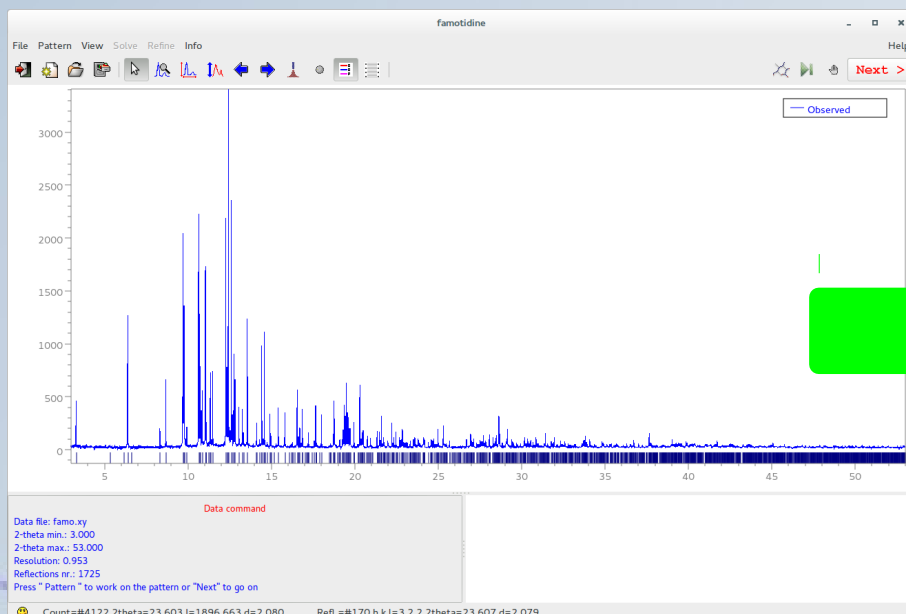
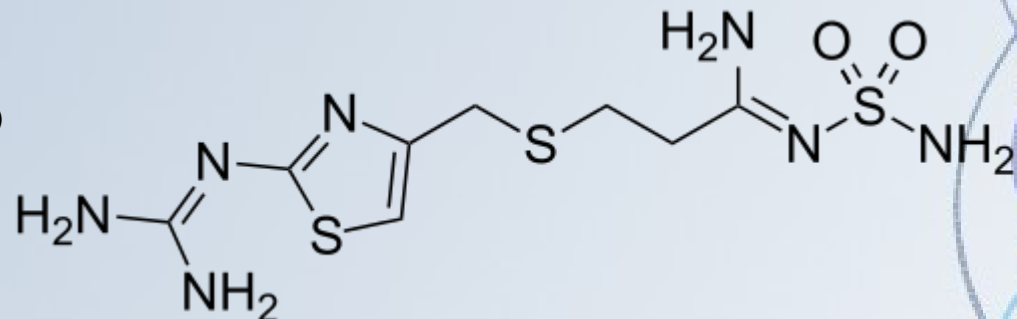
```
cell 17.6547 5.29320 18.2590 90.0 123.5580 90.0
```

```
space p 21/c
```

```
cont (C8 H15 N7 O2 S3) 4
```

```
synchrotron
```

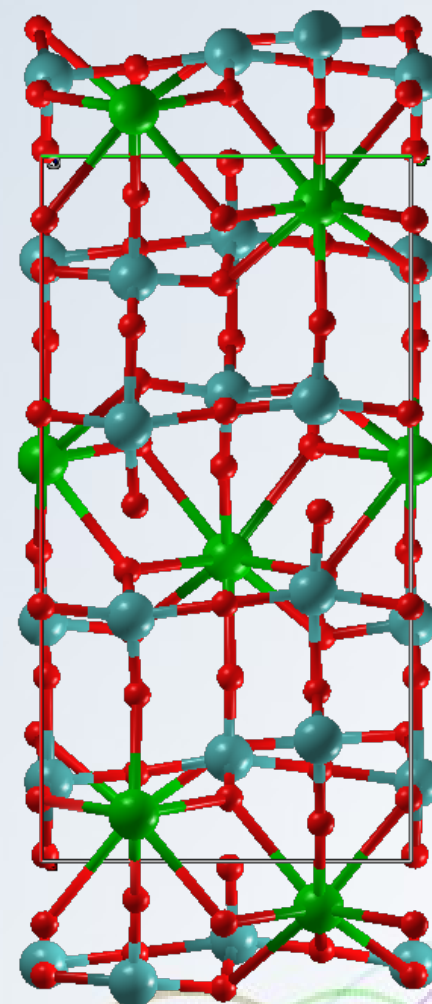
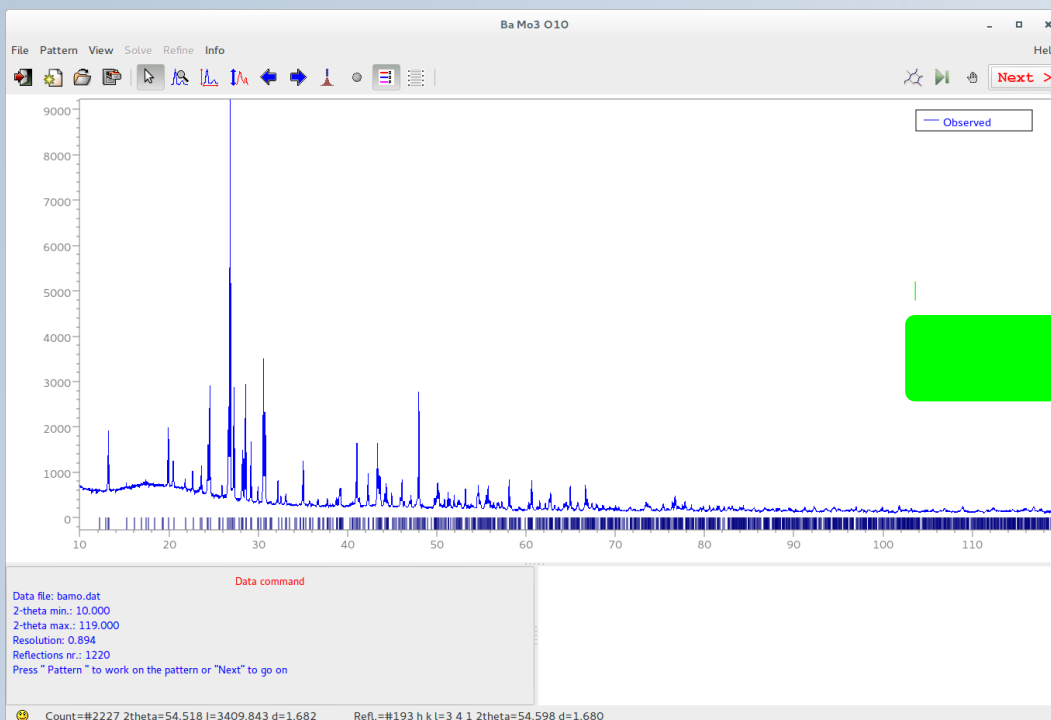
```
%continue
```



Crystal structure determination of BaMo₃O₁₀*

Space group: $P2_1$; $Z = 4$

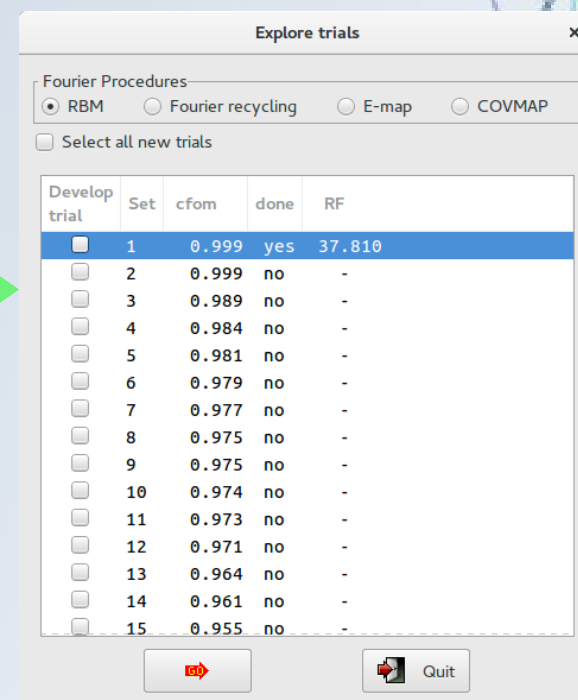
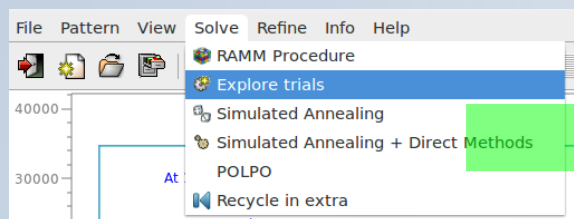
$a = 14.695(2) \text{ \AA}$, $b = 7.5704(7) \text{ \AA}$, $c = 6.9618(6) \text{ \AA}$, and $\beta = 100.381(8)^\circ$



*Werner, P.-E., Moustiakimov, M., Marinder, B.-O. & Knight, K. S. (1997). Z. Kristallogr. 212, 665–670. <https://doi.org/10.1524/zkri.1997.212.9.665>

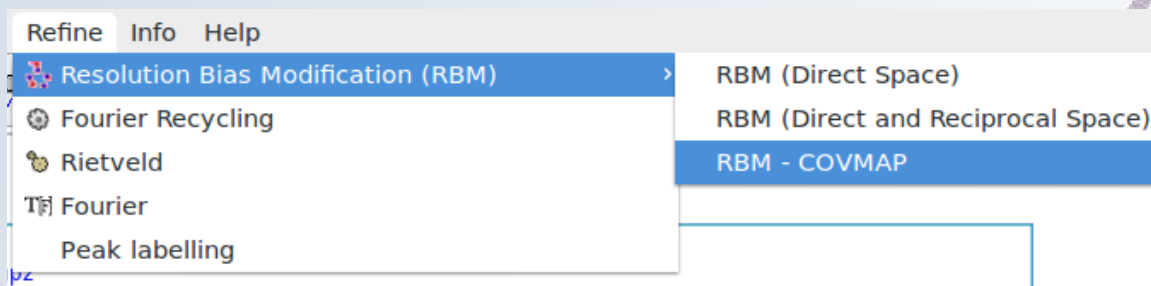
When default Expo2014 fails: strategies

- Explore trials



- Apply the RAMM (Random based Method)

- Structure model optimization



Crystal structure solution of the 2-mercaptobenzoic acid*

%Structure merca

%Job 2-Mercaptobenzoic acid (C7 H6 O2 S)

%Data

Pattern merca.xy

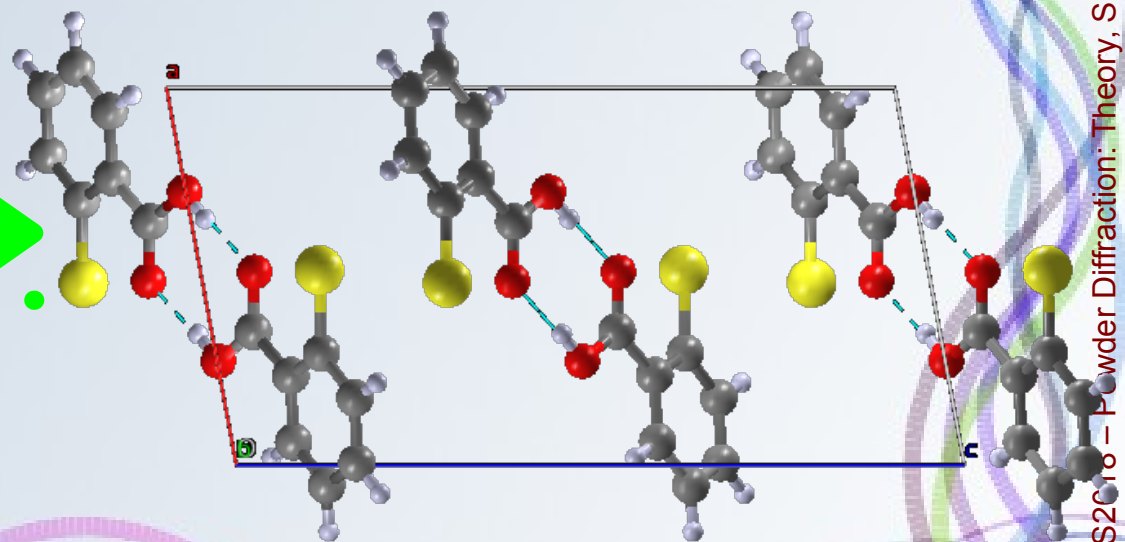
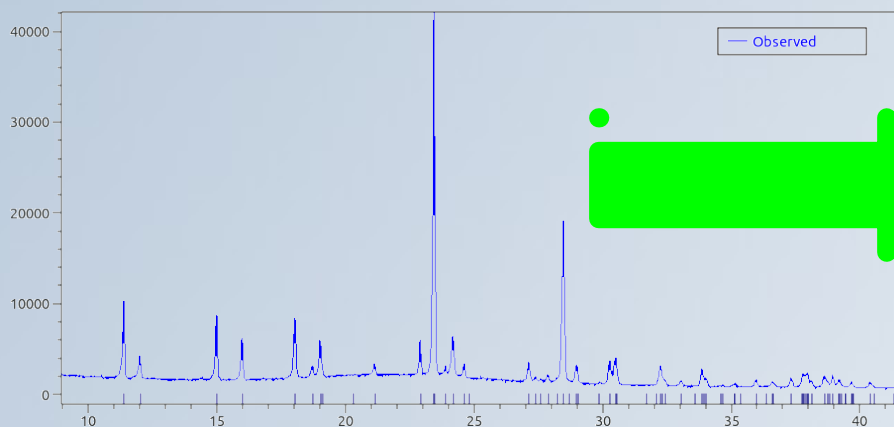
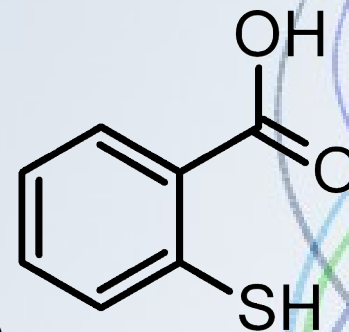
Wavelength 1.54056

Cell 7.885 5.976 14.949 90.0 100.48 90

SpaceGroup p 21/c

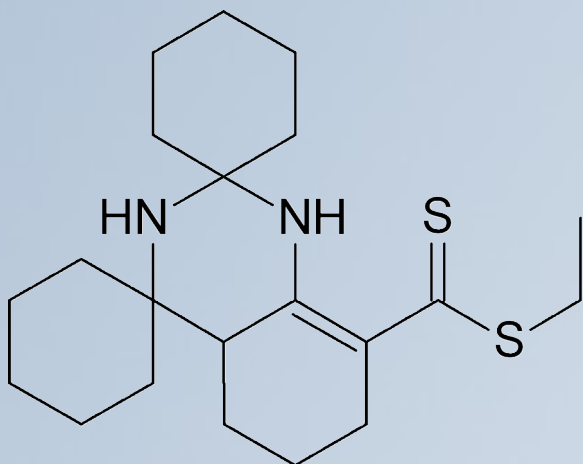
Content (C7 H6 O2 S) 4

%continue

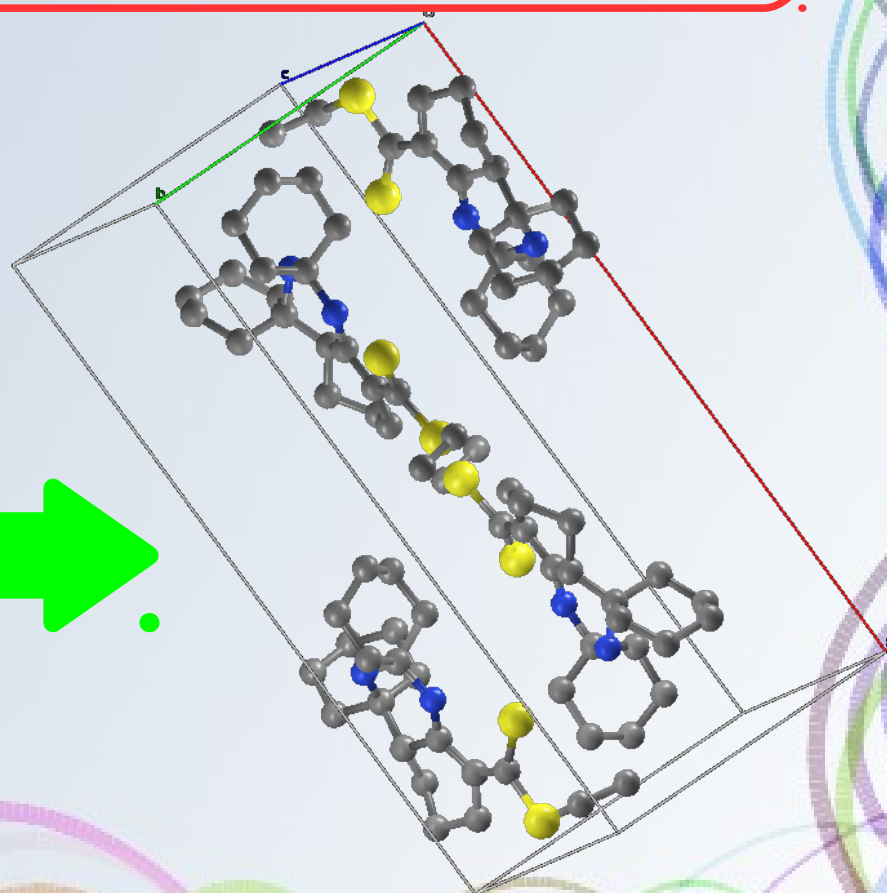
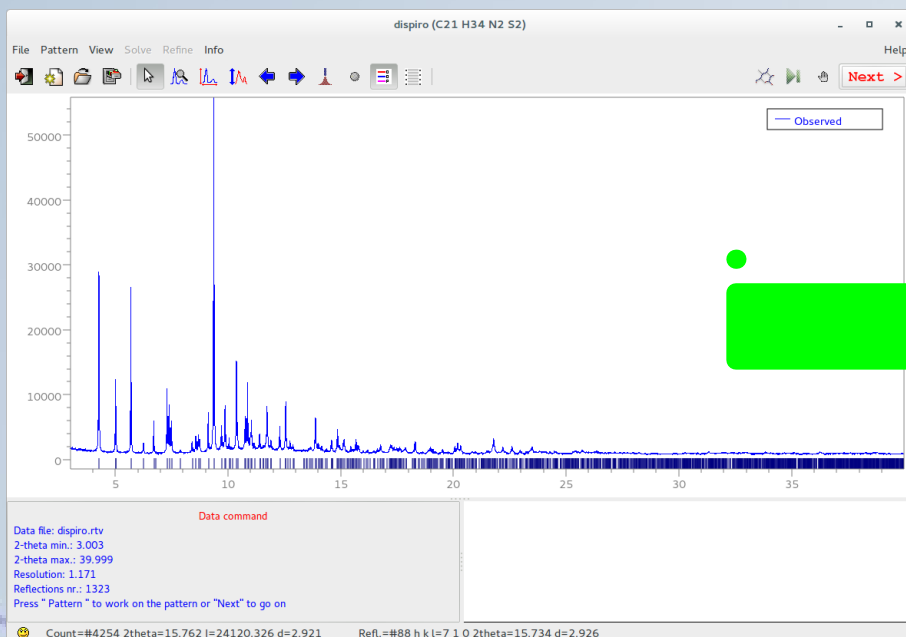


*Steiner, T. Acta Cryst. (2000). C56, 876±877. <https://doi.org/10.1107/S0108270100005898>

Crystal structure solution of a bidentate pro-ligand, $C_{21}H_{34}N_2S_2$ from powder synchrotron diffraction data*

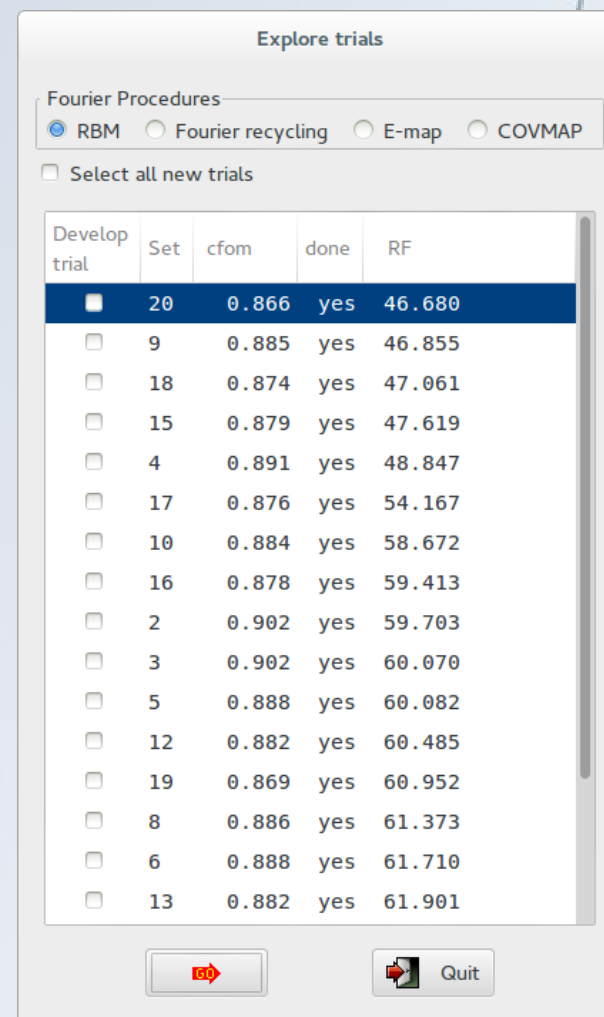
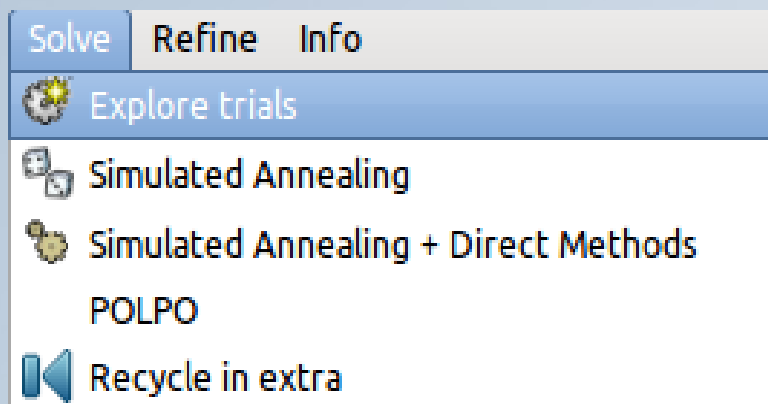


All trials were explored during the Direct Methods procedure



*Ávila, E. E. (2009). Acta Cryst. B65, 639–646

Exploring trials



Command %alltrials

The use of the command **%alltrials** in the input file automatically activate the 'Explore trials' approach

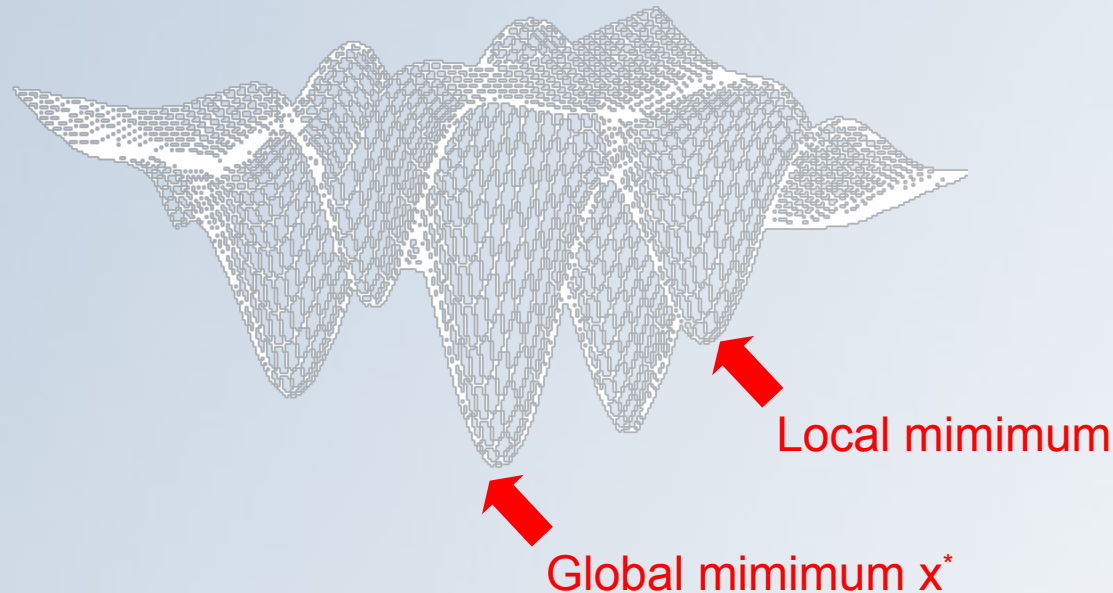
```
%structure dispiro
%job bidentate pro-ligand C21 H34 N2 S2
%data
  pattern dispiro.rtv
  cont (C21 H34 N2 S2) 4
  wave 0.80098
  synchrotron
  range 3.003 40
  cell      21.7356 10.0565 9.4510  90.000  99.9602  90.000
  space p 21/n
%alltrials
```

Global Optimization Methods

Find $\mathbf{x}^* = \min\{F(\mathbf{x})\}$, where $F: \mathbb{R}^n \rightarrow \mathbb{R}$

\mathbf{X} = fractional coordinates of (x,y,z) *or*

\mathbf{X} = position (x,y,z), orientation (θ , ϕ , ψ), torsion angles (τ_1 , τ_2 , ..., τ_n) of molecular fragments



Local optimization methods  Structure refinement

Global optimization methods  Structure solution

Global optimization methods

- Deterministic methods

 - Branch and Bound methods*

 - Cutting Plane methods*

 - Interval methods*

 -

- Heuristic strategies

 - Genetic Algorithms (GA)*

 - Simulated Annealing (SA)*

 - Tabu Search*

 - Ant Colony Optimization*

 - Particle Swarm Optimization (PS)*

 - Bee Algorithms*

 - Firefly Algorithms*

 - Harmony Search*

 - Big Bang-Big Crunch*

 -

Global optimization methods

- Deterministic methods

 - Branch and Bound methods*

 - Cutting Plane methods*

 - Interval methods*

 -

- Heuristic strategies

 - Genetic Algorithms (GA) **

 - Simulated Annealing (SA)**

 - Tabu Search*

 - Ant Colony Optimization*

 - Particle Swarm Optimization (PS)**

 - Bee Algorithms*

 - Firefly Algorithms*

 - Harmony Search*

 - Big Bang-Big Crunch **

 -

(*) *employed in solving crystal structure*

Global optimization methods

- Deterministic methods

Branch and Bound methods

Cutting Plane methods

Interval methods

.....

- Heuristic strategies

*Genetic Algorithms (GA) **

Simulated Annealing (SA) *

Tabu Search

Ant Colony Optimization

*Particle Swarm Optimization (PS) **

Bee Algorithms

Firefly Algorithms

Harmony Search

*Big Bang-Big Crunch **

.....

Widely used and with
the largest impact

Various modifications:

- *parallel tempering (PT)*

- *adaptive simulated annealing*

(*) *employed in solving crystal structure*

Global optimization methods

- Deterministic methods

Branch and Bound methods

Cutting Plane methods

Interval methods

.....

- Heuristic strategies

*Genetic Algorithms (GA) **

Simulated Annealing (SA) *

Tabu Search

Ant Colony Optimization

*Particle Swarm Optimization (PS) **

Bee Algorithms

Firefly Algorithms

Harmony Search

*Big Bang-Big Crunch **

.....

Widely used and with
the largest impact

Various modifications:

- *parallel tempering (PT)*

- *adaptive simulated annealing*

(*) *employed in solving crystal structure*

Global optimization methods

- Deterministic methods

Branch and Bound methods

Cutting Plane methods

Interval methods

.....

- Heuristic strategies

*Genetic Algorithms (GA) **

Simulated Annealing (SA) *

Tabu Search

Ant Colony Optimization

*Particle Swarm Optimization (PS) **

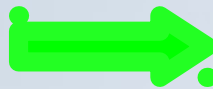
Bee Algorithms

Firefly Algorithms

Harmony Search

*Big Bang-Big Crunch **

.....



Widely used and with
the largest impact

Various modifications:

- *parallel tempering (PT)*
- *adaptive simulated annealing*

Software **: DASH (SA), EXPO (SA),
ENDEAVOUR (SA), FOX (PT), GEST (GA),
PeckCryst (PS), PowderSolve (SA), PSSP (SA),
TOPAS (SA), ...

commercial software are in red

(**) Cerny, R. & Favre-Nicolin, V. (2007). Z. Kristallogr. 222, 105–113.

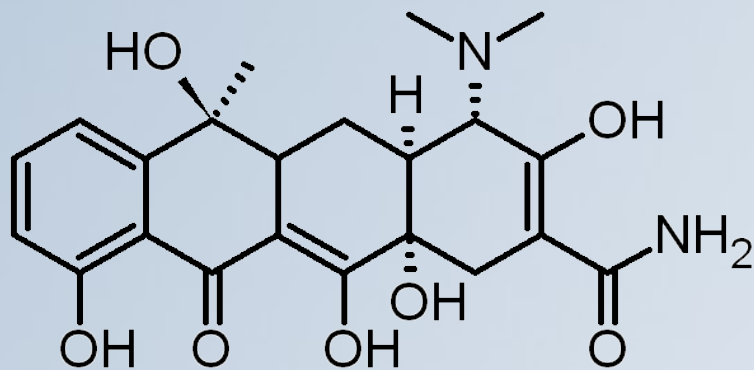
(*) employed in solving crystal structure

Comparison

Traditional approaches

Do not use chemical knowledge

Complexity of the problem depends on the number of non H-atoms in the a.u.



Clegg, W. & Teat, S. J., (2000). *Acta Cryst.* C56, 1343-1345.

Take advantage by using data of higher resolution

Generally require less time to run

Direct space methods

Can incorporate a massive amount of prior chemical information

Complexity of procedure depends on the number of degrees of freedom (DoF).

tetracycline (32 non-H atoms and 8 DoF) can be solved using global optimization

High resolution is not needed.
Default resolution: 2-2.5 Å.

Take time and patience. For large molecules: faster computer, run overnight, parallel program

Building starting model

- **Check for similar molecules** in databases or in the literature
- **Optimize molecular geometry** by computational chemistry programs

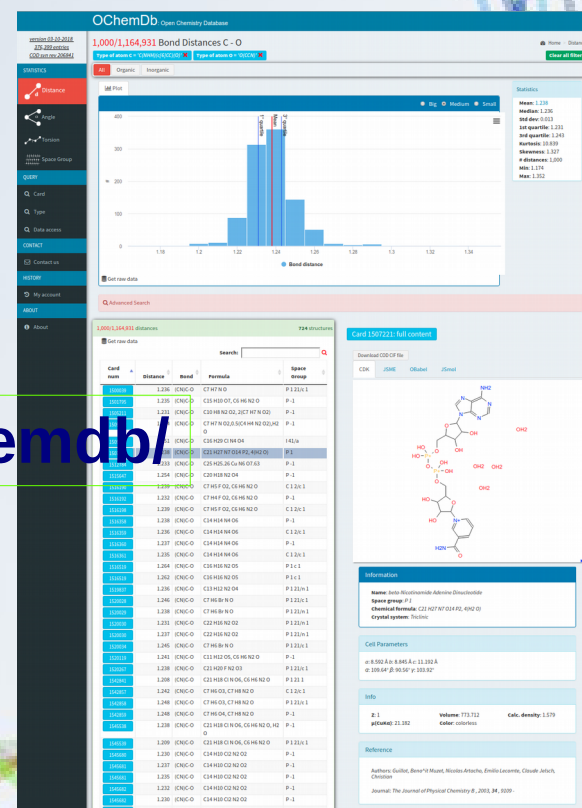
Free Chemistry Databases

- PubChem: <https://pubchem.ncbi.nlm.nih.gov/>
- NIST Chemistry WebBook: <http://webbook.nist.gov/chemistry/>
- Drugbank: <http://www.drugbank.ca/>

Other databases: ZINC, eMolecules, ChEBI, NMRShiftDB, ...

Chemical file formats: *sdf*, *mol*, *mol2*, *cml*, *SMILES*, ...

<http://www.ba.ic.cnr.it/ochemdb/>



Crystal Structure Databases*

Non-commercial database are in red

- **CSD** (Cambridge Structural Database) (organics & organometallics):
<http://www.ccdc.cam.ac.uk/>
- **ICSD** (Inorganic Crystal Structure Database)
(inorganics, elements, minerals & intermetallics): <http://icsd.ill.fr/>
- **COD** (Crystallography Open Database) (general database):
<http://www.crystallography.net/>

Other databases: ICDD PDF-4+, American Mineralogist Crystal Structure Database, MINCRYST, Zeolite Structures Database, ...

File format: CIF (Crystallographic Information File)

* Joint special issue: *Acta Cryst. B58*, 317-422 (2002)

Load crystal structures from COD

“View” menu → “Jav Molecular Viewer”

“File” menu → “Import Fragment” menu → “From COD”

“Modify” menu → “Add Fragments” menu → “From COD”

Import fragment from COD

Text:

Formula:

Elements:

NOT these elements:

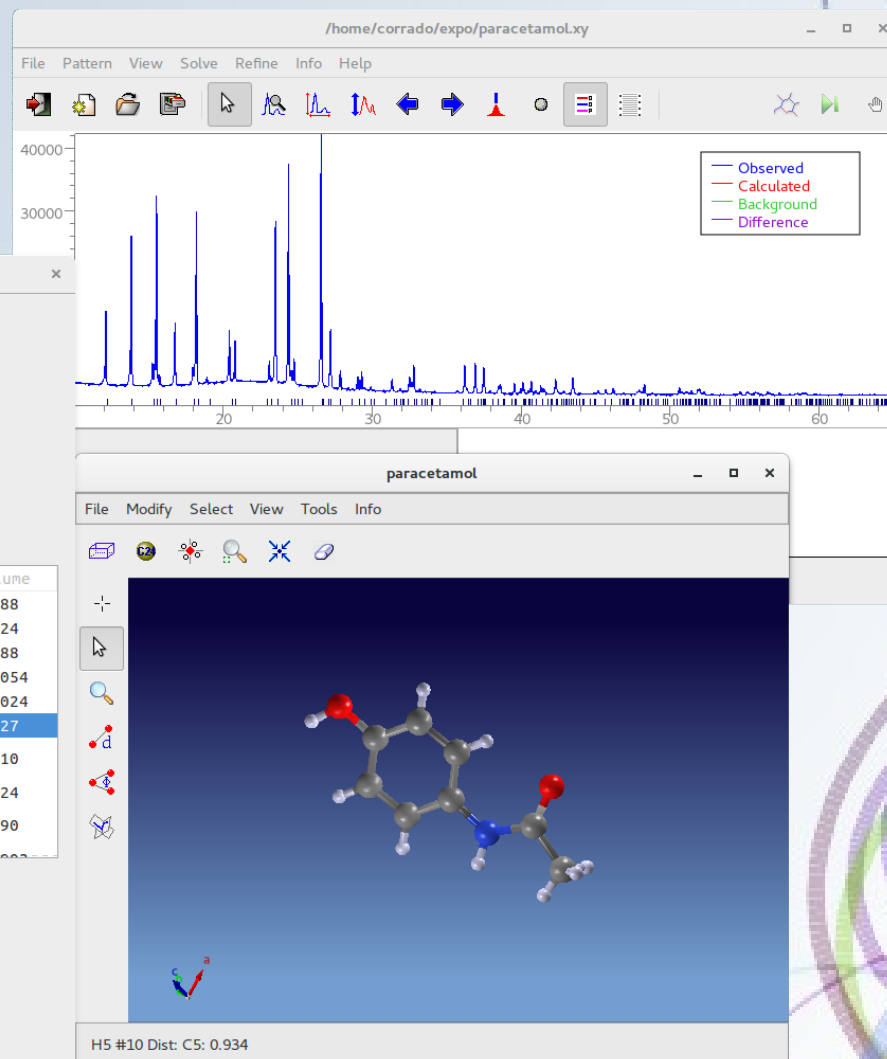
No. of elements min and max:

| COD Number | Spacegroup | Formula | Cell parameters | Cell volume |
|------------|------------|------------|------------------------------------|-------------|
| 1520183 | P n a 21 | C8 H9 N O2 | 10.5129 17.0435 4.0675;90 90 90 | 728.799988 |
| 1520187 | P n a 21 | C8 H9 N O2 | 10.5957 7.6655 9.2544;90 90 90 | 751.650024 |
| 1520188 | P n a 21 | C8 H9 N O2 | 10.5129 17.0435 4.0675;90 90 90 | 728.799988 |
| 1548348 | P c a b | C8 H9 N O2 | 7.232 11.76 17.16;90 90 90 | 1459.430054 |
| 2006392 | P 1 21/n 1 | C8 H9 N O2 | 10.795 8.271 17.803;90 92.957 90 | 1587.400024 |
| 2007205 | P 1 21/n 1 | C8 H9 N O2 | 7.0939 9.2625 11.657;90 97.672 90 | 759.090027 |
| 2008620 | P 1 21/c 1 | C8 H9 N O2 | 15.7794 4.8525 9.8771;90 97.952 90 | 749.010010 |
| 2009919 | P 1 21/n 1 | C8 H9 N O2 | 6.664 16.83 7.153;90 107.898 90 | 763.400024 |
| 2013900 | P 1 21/c 1 | C8 H9 N O2 | 8.5969 5.6053 15.5397;90 96.172 90 | 744.489990 |

Chemical names: Acetaminophen; 4-acetaminophenol

Bibliography: Dmitry Yu. Naumov; Marina A. Vasilchenko; Judith A. K. Howard
The Monoclinic Form of Acetaminophen at 150K
Acta Crystallographica Section C, 1998, 54, 653-655.
<https://doi.org/10.1107/S0108270197018386>

55 records found.



Geometry optimization

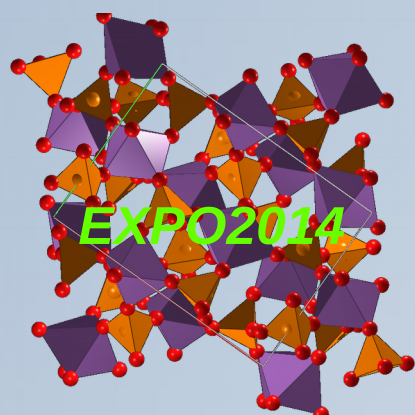
Three levels of theory

- Molecular-mechanics force fields (**MM**)
- Semi-empirical methods (**SE**)
- *Ab initio* methods: Hartree–Fock methods, density functional theory (**DFT**)

Strategy: **MM** → **SI** → **DFT**

Programs: MOPAC, Gamess, NWChem, Gaussian, ABINIT, Quantum ESPRESSO, ORCA, Molpro, Q-Chem, octopus, etc.

The Use of Open Babel



Fotran/C++

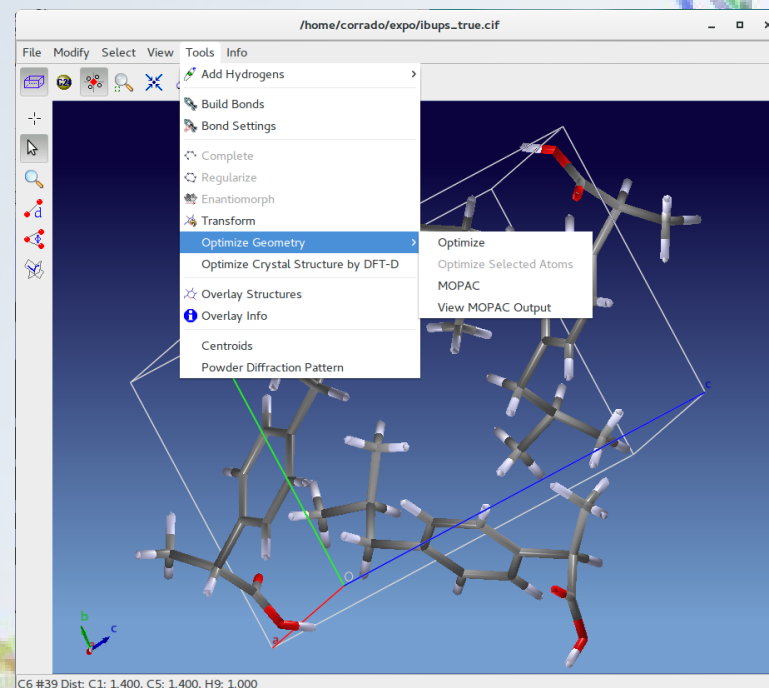


obmodule.f90



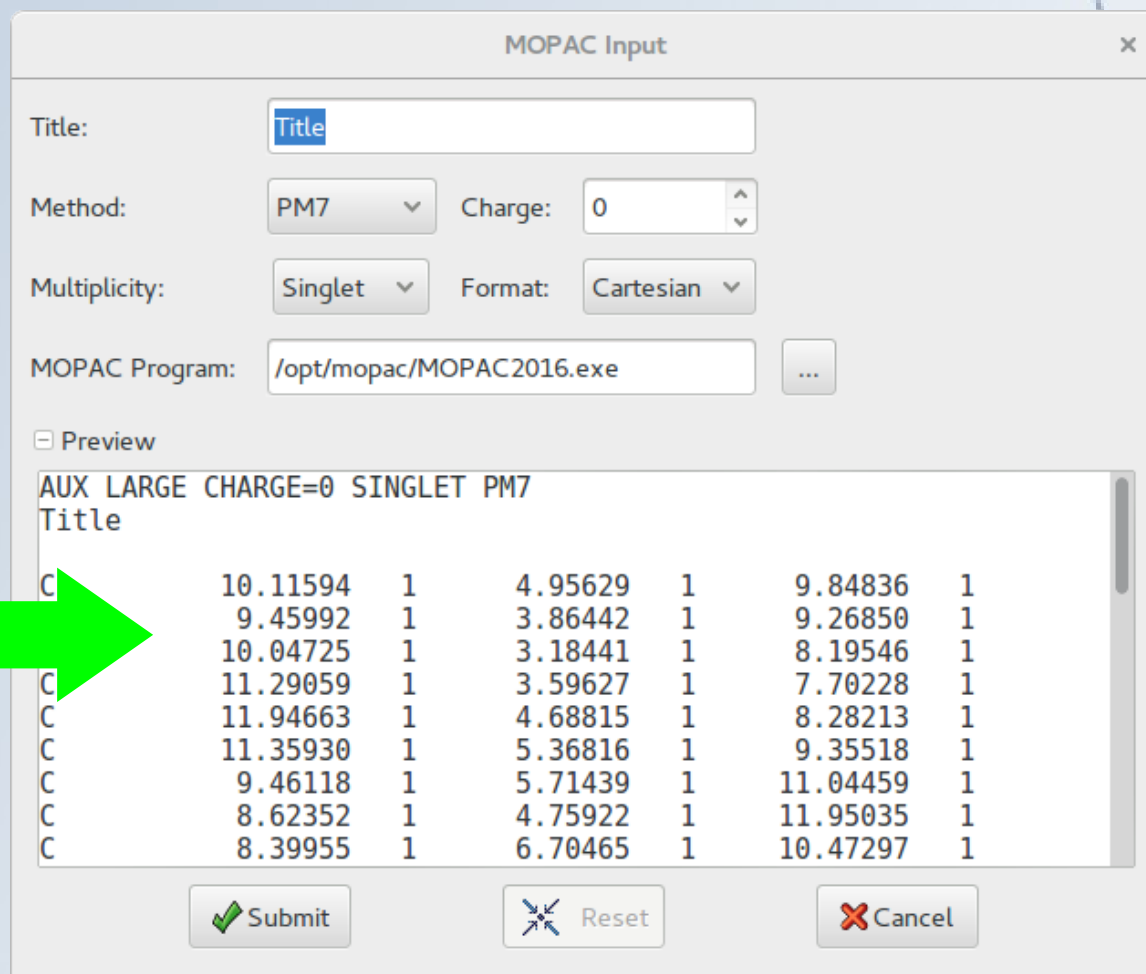
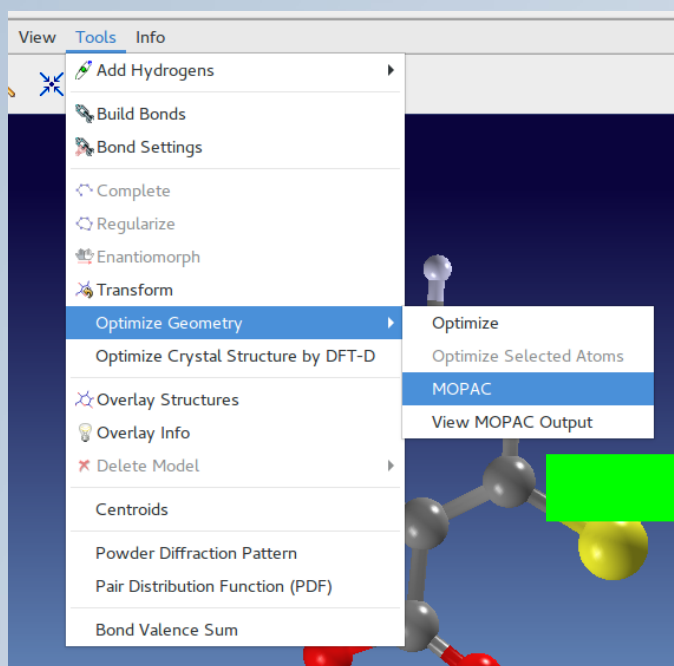
C++

- Molecular-mechanics force fields (MMFF99 and UFF provided by **Open Babel library**)
- Able to process input and output files of common quantum-chemistry packages: GAMESS-US, NWChem, Gaussian, CRYSTAL, ABINIT, Quantum ESPRESSO



Geometry optimization by MOPAC

MOPAC2016™ is a semiempirical quantum chemistry software package available FREE for academic, not-for-profit use. Download link: <http://openmopac.net/downloads.html>



The 'MOPAC Input' dialog box is shown. It contains fields for Title, Method, Charge, Multiplicity, Format, and MOPAC Program. A 'Preview' section displays the input file content. At the bottom are 'Submit', 'Reset', and 'Cancel' buttons.

Title:

Method: Charge:

Multiplicity: Format:

MOPAC Program:

☐ Preview

```
AUX LARGE CHARGE=0 SINGLET PM7
Title
C      10.11594  1      4.95629  1      9.84836  1
      9.45992  1      3.86442  1      9.26850  1
      10.04725  1      3.18441  1      8.19546  1
C      11.29059  1      3.59627  1      7.70228  1
C      11.94663  1      4.68815  1      8.28213  1
C      11.35930  1      5.36816  1      9.35518  1
C      9.46118  1      5.71439  1     11.04459  1
C      8.62352  1      4.75922  1     11.95035  1
C      8.39955  1      6.70465  1     10.47297  1
```

Graphical User Interface for MOPAC2016

Molecule editor

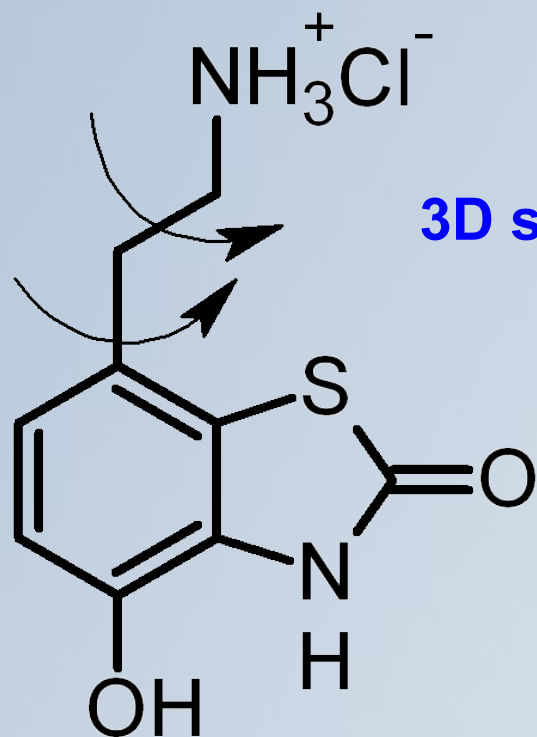
A molecule editor allows

- Sketch molecules in 2D or 3D format
- Optimize the geometry by force field method
- Create input file for the quantum-chemistry calculations
- Read output files of the most common computational packages

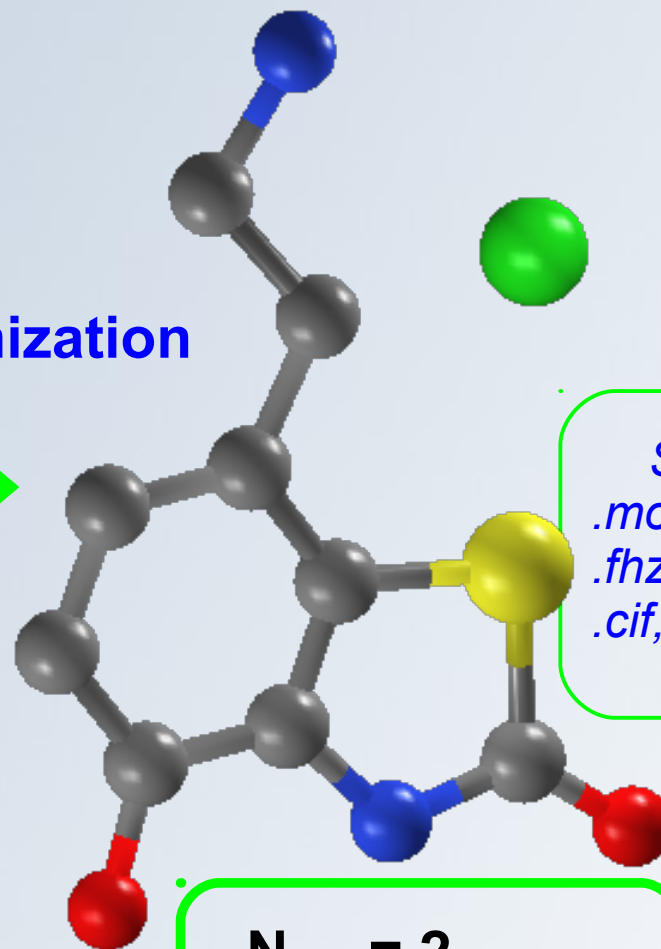
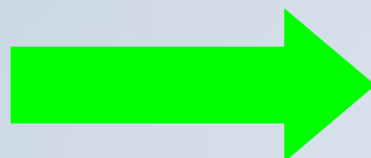
Some free available software

- ACD/ChemSketch - <http://www.acdlabs.com>
- Avogadro - http://avogadro.openmolecules.net/wiki/Main_Page
- MarvinSketch - <http://www.chemaxon.com/products/marvin/>
- Gabedit: <http://gabedit.sourceforge.net/>

Building starting model



3D structure optimization



Save as
.mol, .mol2,
.fhz, .mop,
.cif, .pdb, .frac

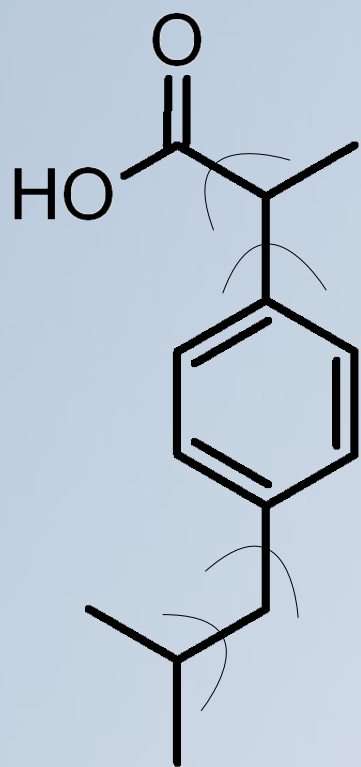
$$N_{\text{frag}} = 2$$

$$N_{\text{dof}} = 6 + 3 + 2$$

2-(4-Hydroxy-2-oxo-2,3-dihydro-1,3-benzothiazol-7-yl)
ethylammonium chloride

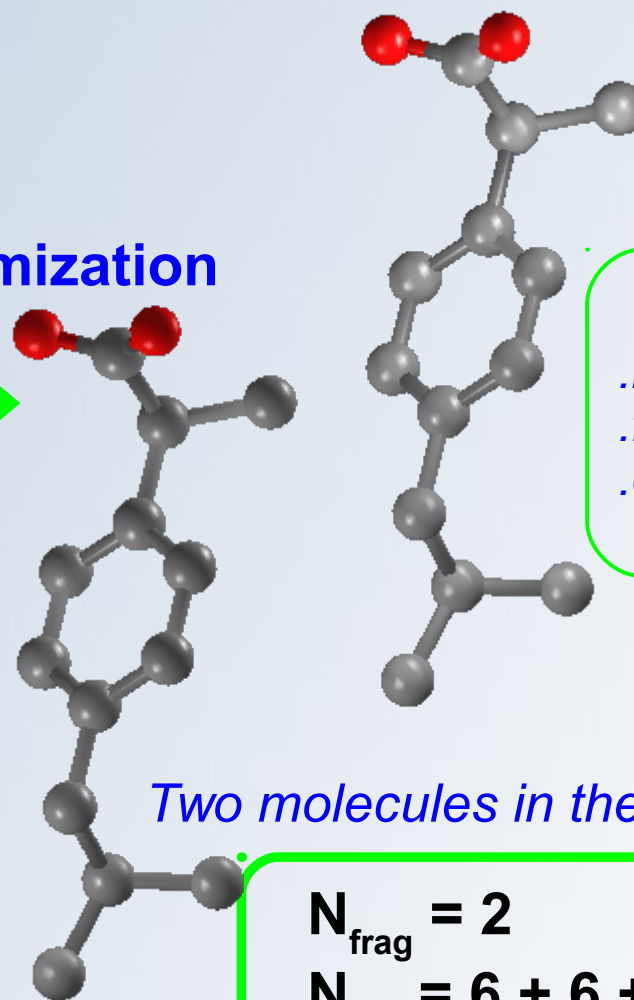
P 21/a

Building starting model: example 2



S-Ibuprofen
 $P2_1$

3D structure optimization

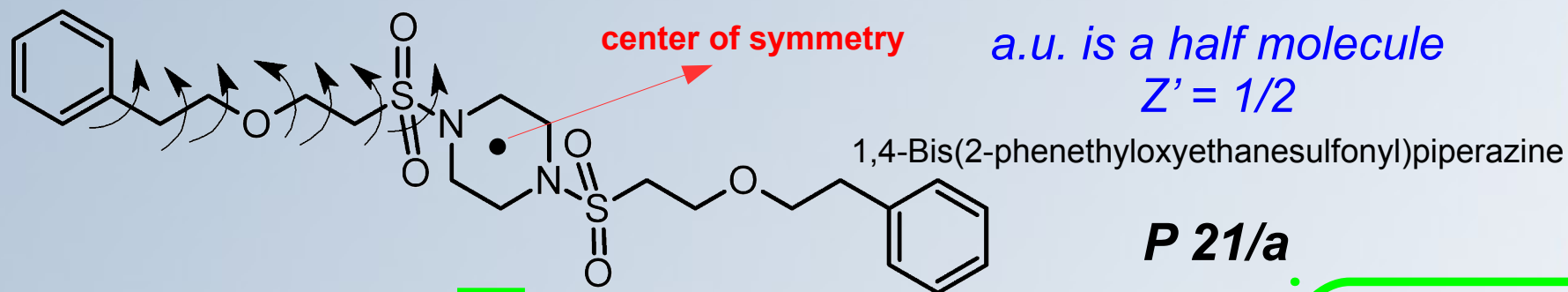


Save as
.mol, .mol2,
.fhz, .mop,
.cif, .pdb, .frac

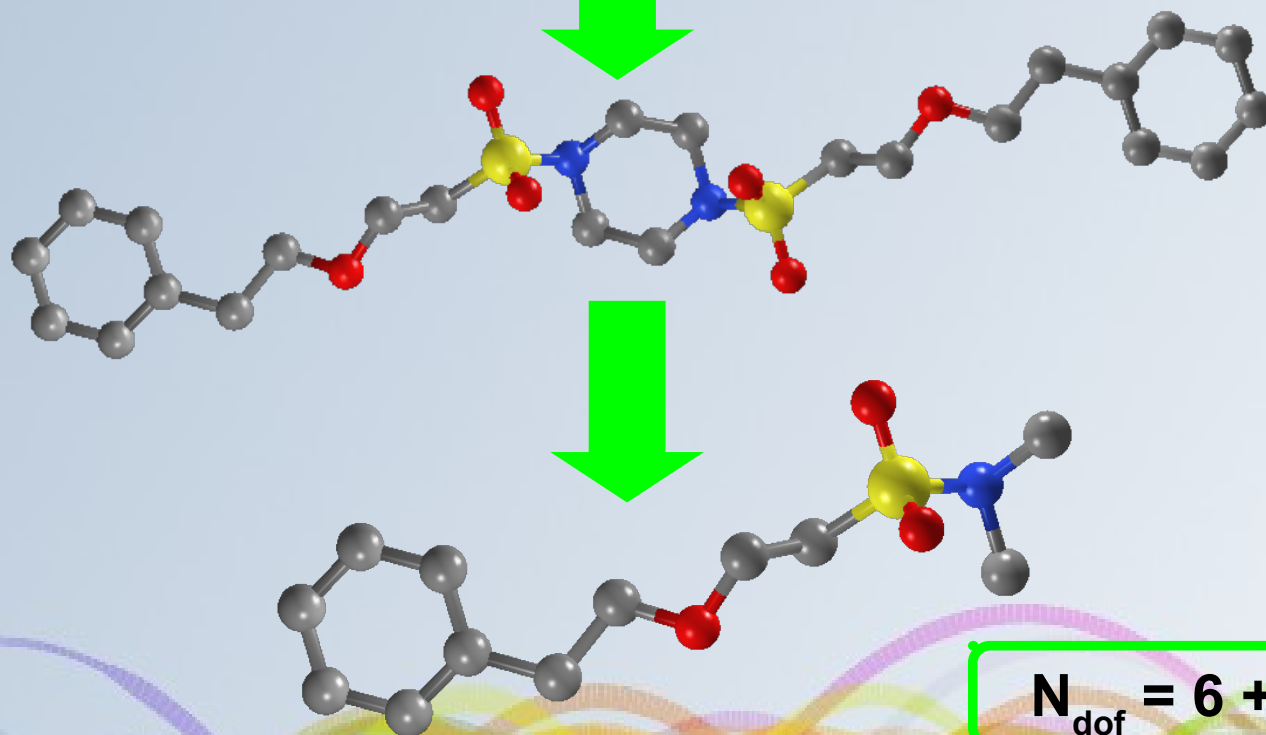
Two molecules in the a.u. ($Z'=2$)

$$\begin{aligned} N_{\text{frag}} &= 2 \\ N_{\text{dof}} &= 6 + 6 + 4 + 4 \end{aligned}$$

Building starting model: example 3



3D structure optimization



**Structure solution
by DSM + DOC**

$$N_{\text{dof}} = 6 + 14$$

15 times slower

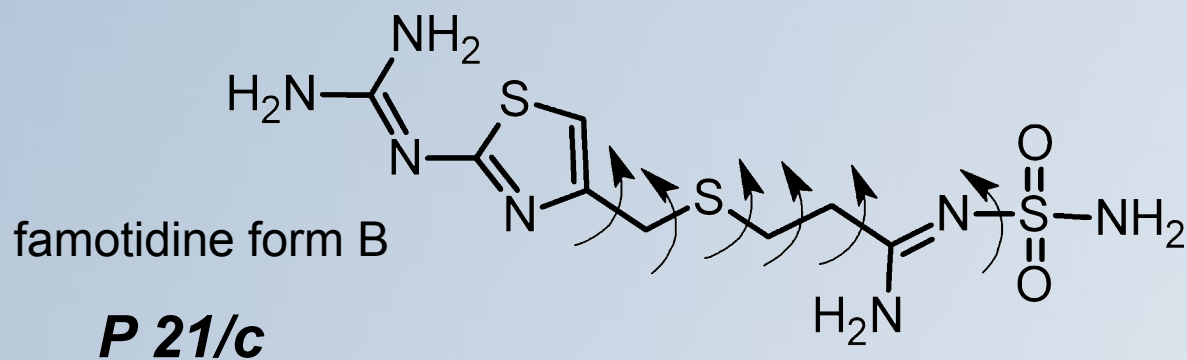
**Structure solution
by DSM**

$$N_{\text{dof}} = 6 + 7$$

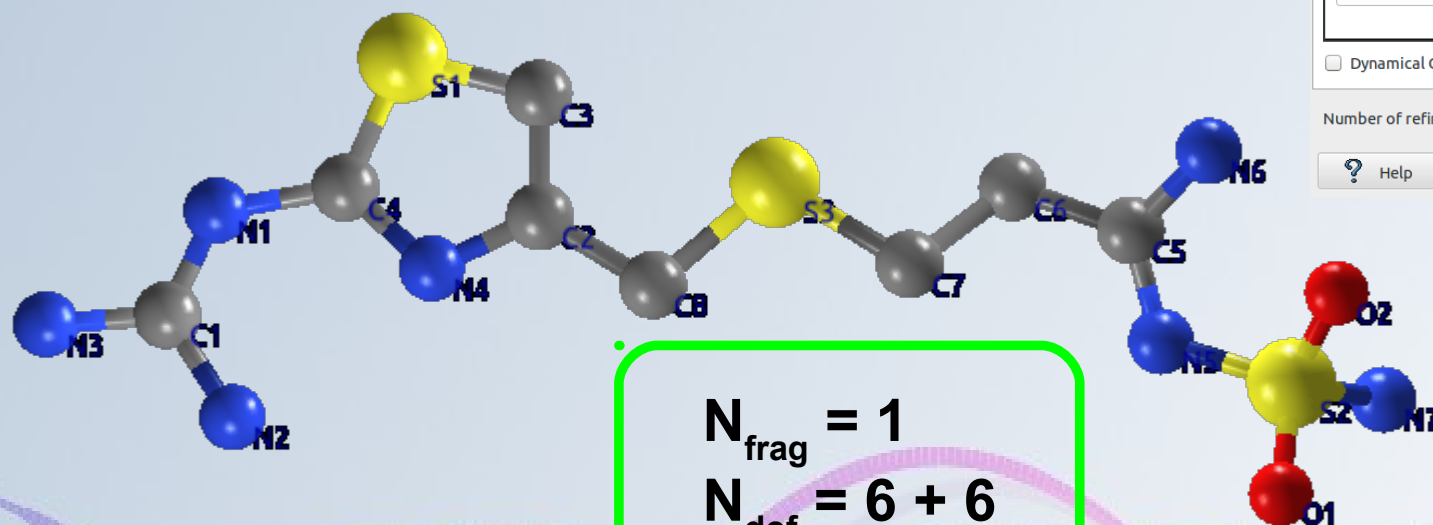
**Recommended
strategy**

$$N_{\text{dof}} = 6 + 7$$

Building starting model: example 4



3D structure optimization



$$N_{\text{frag}} = 1$$

$$N_{\text{dof}} = 6 + 6$$

SA conditions External DOF Internal DOF Anti-bump

Internal DOFs

| Torsion | Refine | Value | Lower | Upper |
|-------------|-------------------------------------|--------|---------|--------|
| C4:N1:C1:N2 | <input type="checkbox"/> | -0.14 | -180.00 | 180.00 |
| N6:C5:N5:S2 | <input checked="" type="checkbox"/> | -0.25 | -180.00 | 180.00 |
| C7:C6:C5:N5 | <input checked="" type="checkbox"/> | 48.91 | -180.00 | 180.00 |
| C5:N5:S2:O1 | <input checked="" type="checkbox"/> | 137.81 | -180.00 | 180.00 |
| S3:C7:C6:C5 | <input checked="" type="checkbox"/> | 166.94 | -180.00 | 180.00 |
| C8:S3:C7:C6 | <input checked="" type="checkbox"/> | 166.68 | -180.00 | 180.00 |
| C7:S3:C8:C2 | <input checked="" type="checkbox"/> | 176.54 | -180.00 | 180.00 |
| S3:C8:C2:C3 | <input checked="" type="checkbox"/> | -10.45 | -180.00 | 180.00 |
| N4:C4:N1:C1 | <input type="checkbox"/> | 1.99 | -180.00 | 180.00 |

☐ Dynamical Occupancy Correction ☒ Atomic Parameters

Number of refined parameters: 13

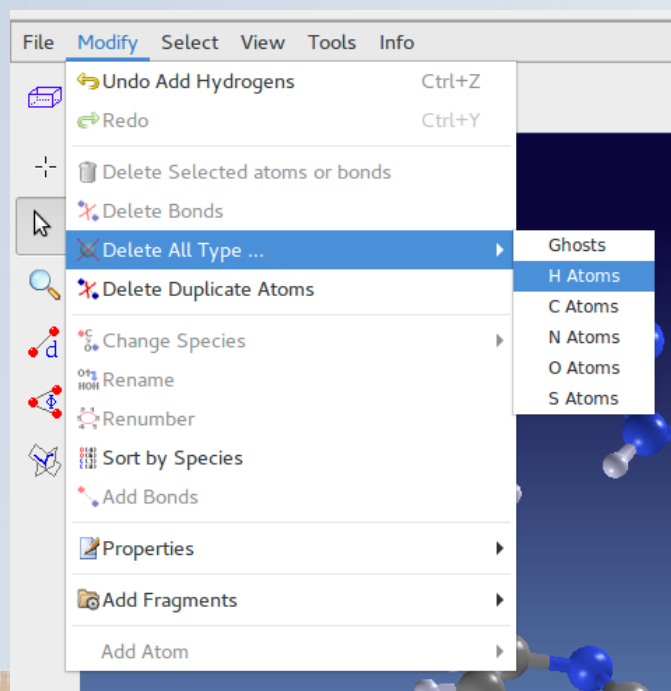
☒ Solutions

H atoms

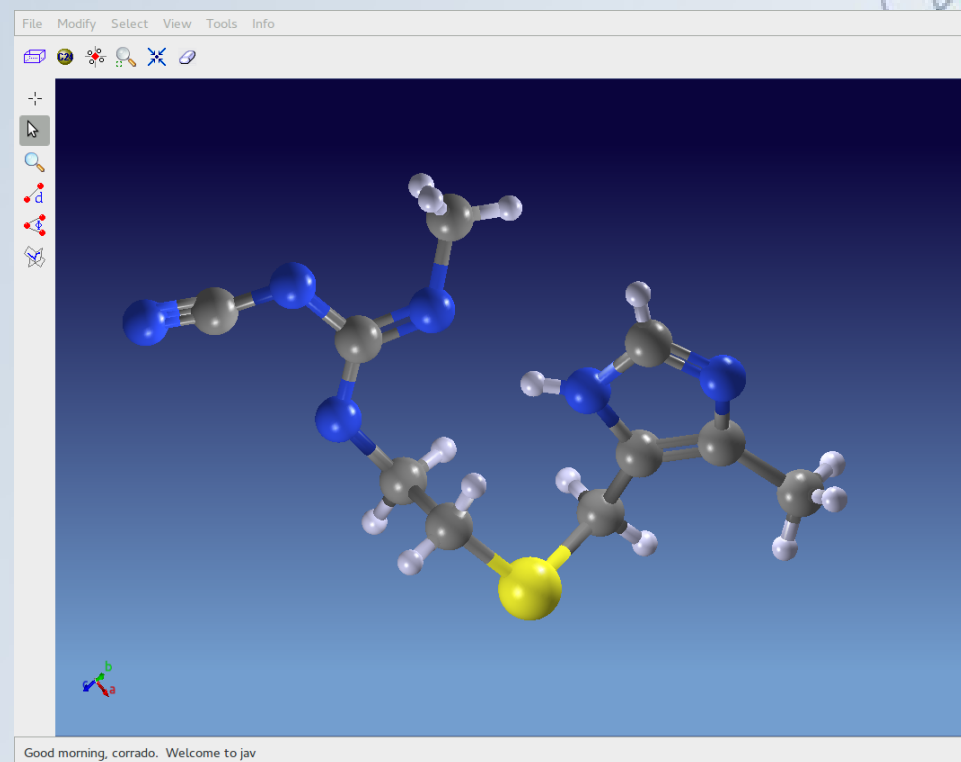
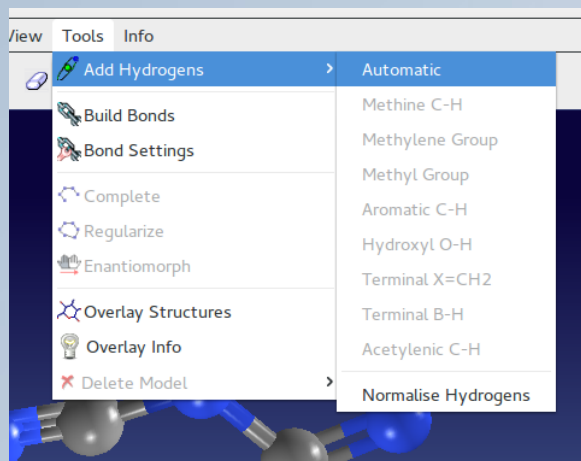
H atoms do not contribute significantly to X-ray diffraction, they can be ignored during the structure solution

Eliminating the H atoms reduces the number of atoms and DoFs, decreasing the time to evaluate CF for each trial structure

Delete H atoms

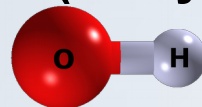


Hydrogen calculation

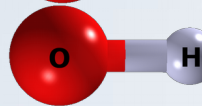


Hydrogen atoms are positioned geometrically at X-ray distances

X-H (X-ray) < X-H (neutrons)



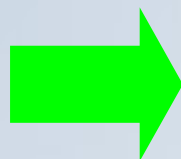
0.82 Å (X-ray)



0.95 Å (neutrons)

Simulated annealing options

- Cost function
- Resolution
- Random seed
- **Number of moves**
- **Number of SA runs**
- Starting temperature
- Temperature reduction factor



- No. of molecular fragments
- No. of external DoFs
- No. of internal DoFs
- The flexibility of the molecule

Global optimization dialog

SA conditions External DOF Internal DOF Pattern

General conditions

Cost function: R weighted profile

Resolution: 2.000 N. of reflections: 103 2thmax: 45.305

Random seed: 1

Nr. of runs: 10

Simulated annealing options

Starting temperature: 10.000 ☒ automatic

Number of moves: 60 ☒ automatic

Temperature reduction factor: 0.900

Number of refined parameters: 8

Solutions

Help Quit Execute

Cost Functions

Whole profile R factors: R_p, R_{wp}, χ^2

$$R_{wp} = \sqrt{\frac{\sum_i w_i (y_{exp}(\theta_i) - y_{calc}(\theta_i))^2}{\sum_i w_i y_{exp}(\theta_i)^2}}$$

$$2\theta_0 - f * FWHM < y(\theta_i) < 2\theta_0 + f * FWHM \quad f = 1$$

Integrated intensities R factors: R_B, R_{Bg}

$$R_B = \frac{\sum_h |I_h^{exp} - I_h^{calc}|}{\sum_h I_h^{exp}}$$

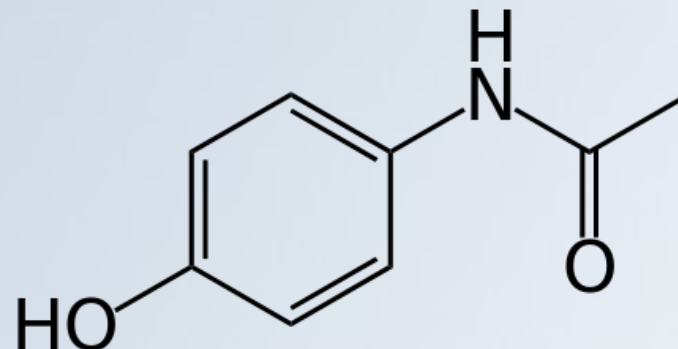
Other cost functions: $CF_{\text{geometry restraints}}, CF_{\text{bond valence}}, CF_{\text{antibumping}}$

Using expo2014 for direct-space solution

- By input file (*.exp)

File > Load & Go

```
%Structure paracetamol
%Job Paracetamol (C8H9NO2)
%Data
  Cell 7.100 9.380 11.708 90.0 97.42 90.0
  SpaceGroup p 21/n
  Pattern paracetamol.xy
  Wavelength 1.54056
%fragment paracetamol.mol
%sannel
```



- Command-line usage

```
expo paracetamol.exp
or
expo paracetamol.exp -nogui
or
expo paracetamol.exp -auto
```

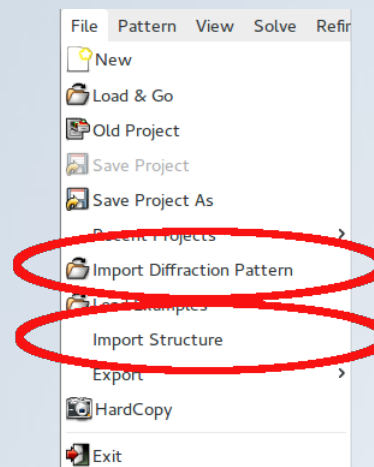
Usage of common %fragment

```
%fragment tetra AtC AtV [dist]
%fragment octa AtC AtV [dist]
%fragment square AtC AtV [dist]
%fragment cube AtC AtV [dist]
%fragment trigonal AtC AtV [dist]
%fragment prism_tetra AtC AtV [dist]
%fragment prism_trig AtC AtV [dist]
%fragment icosah AtC AtV [dist]
%fragment atoms chem_formula
%fragment smiles SMILES_string
```

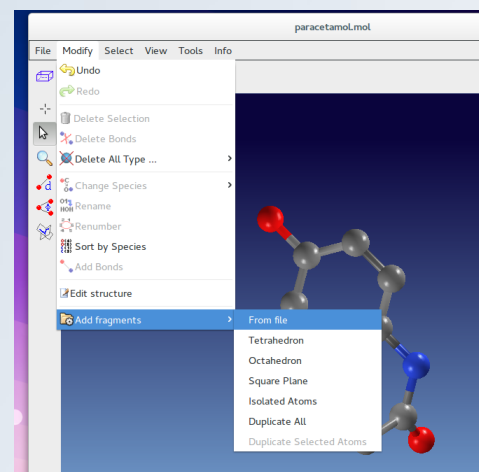
Using expo2014 for direct-space solution

- By graphical interface

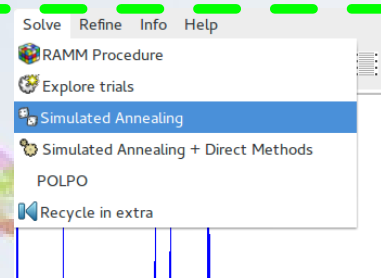
- *File > Import Diffraction Pattern*
- *File > Import Structure*



- *Modify > Add Fragments*



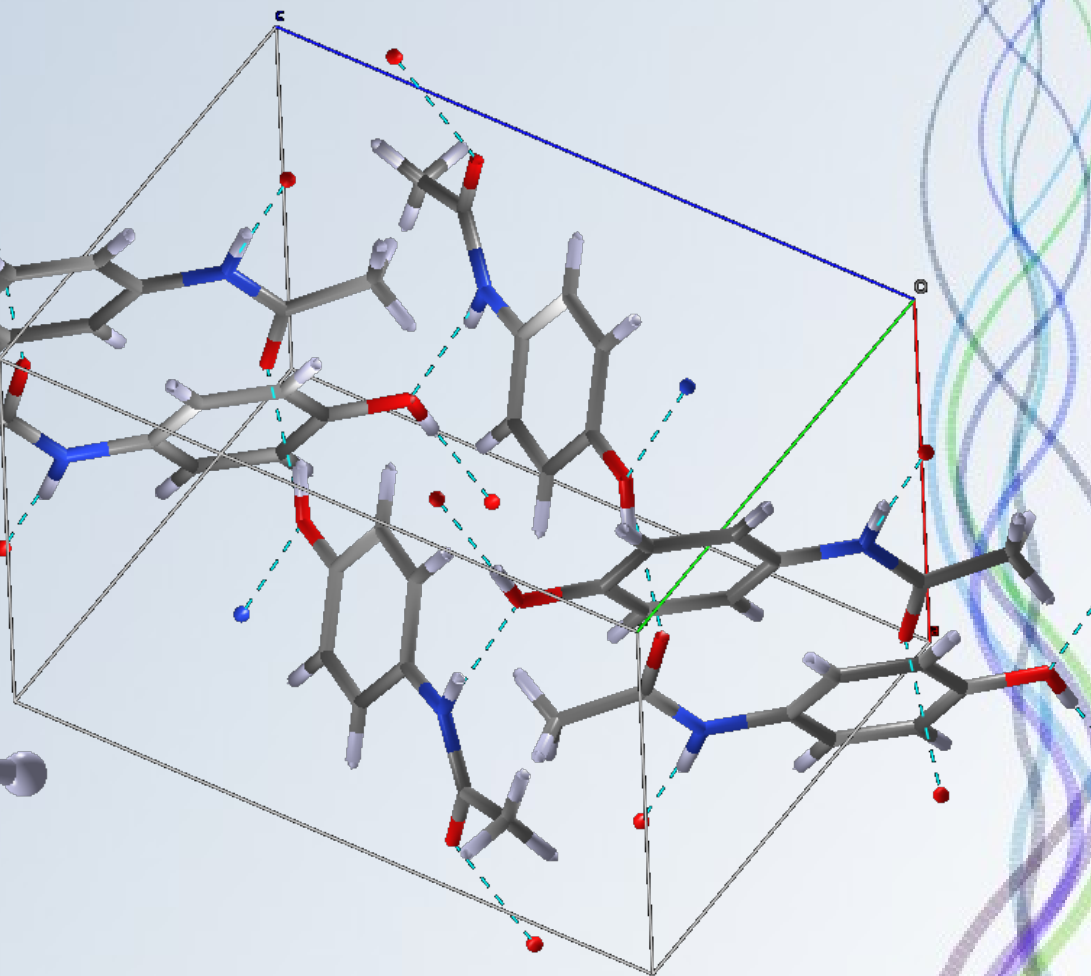
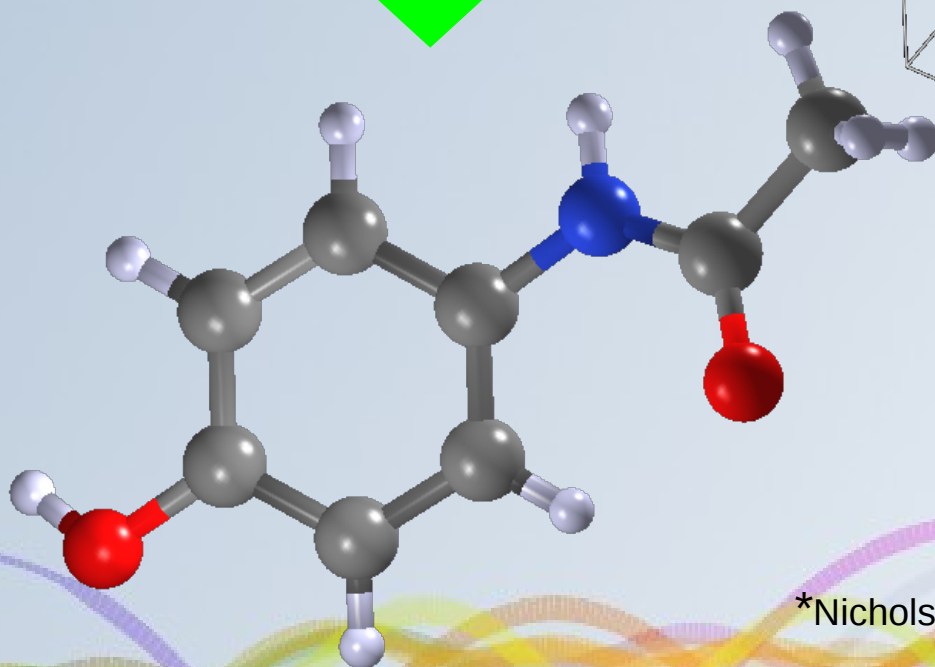
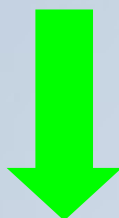
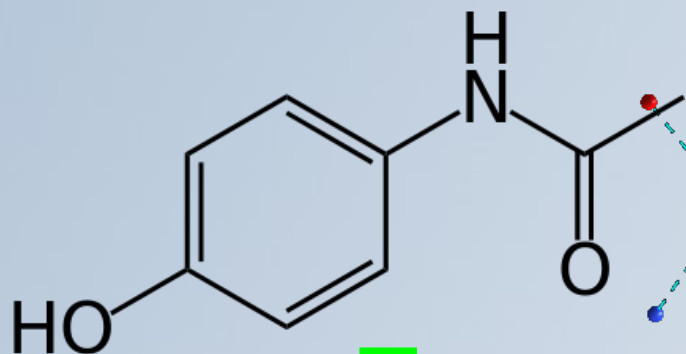
- *Solve > Simulated Annealing*



Molecular compounds

Paracetamol (form I polymorph)

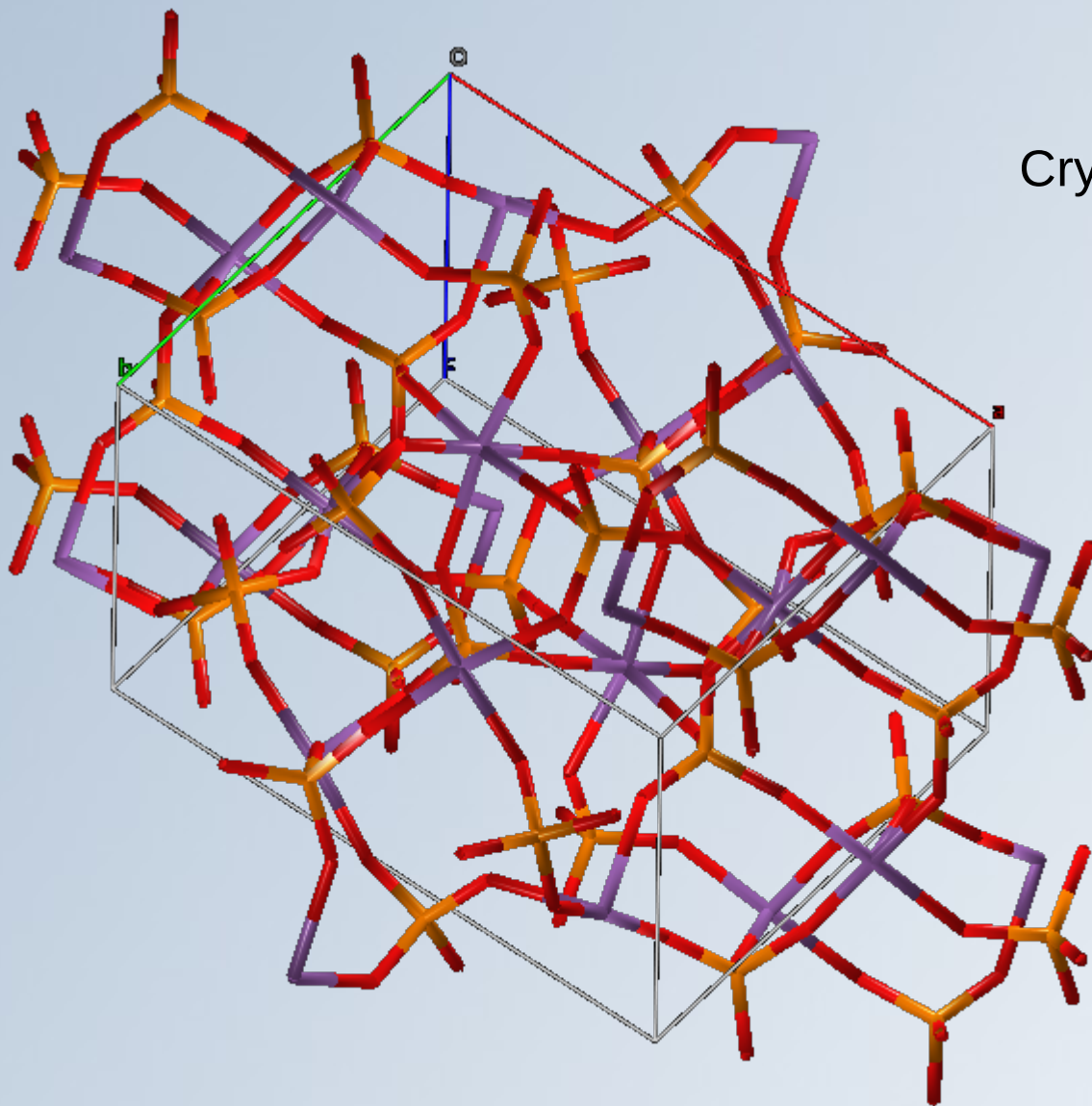
$(C_8H_9NO_2)^*$



*Nichols, C. & Frampton, C. S. (1998). J. Pharm. Sci. 87, 684–693.

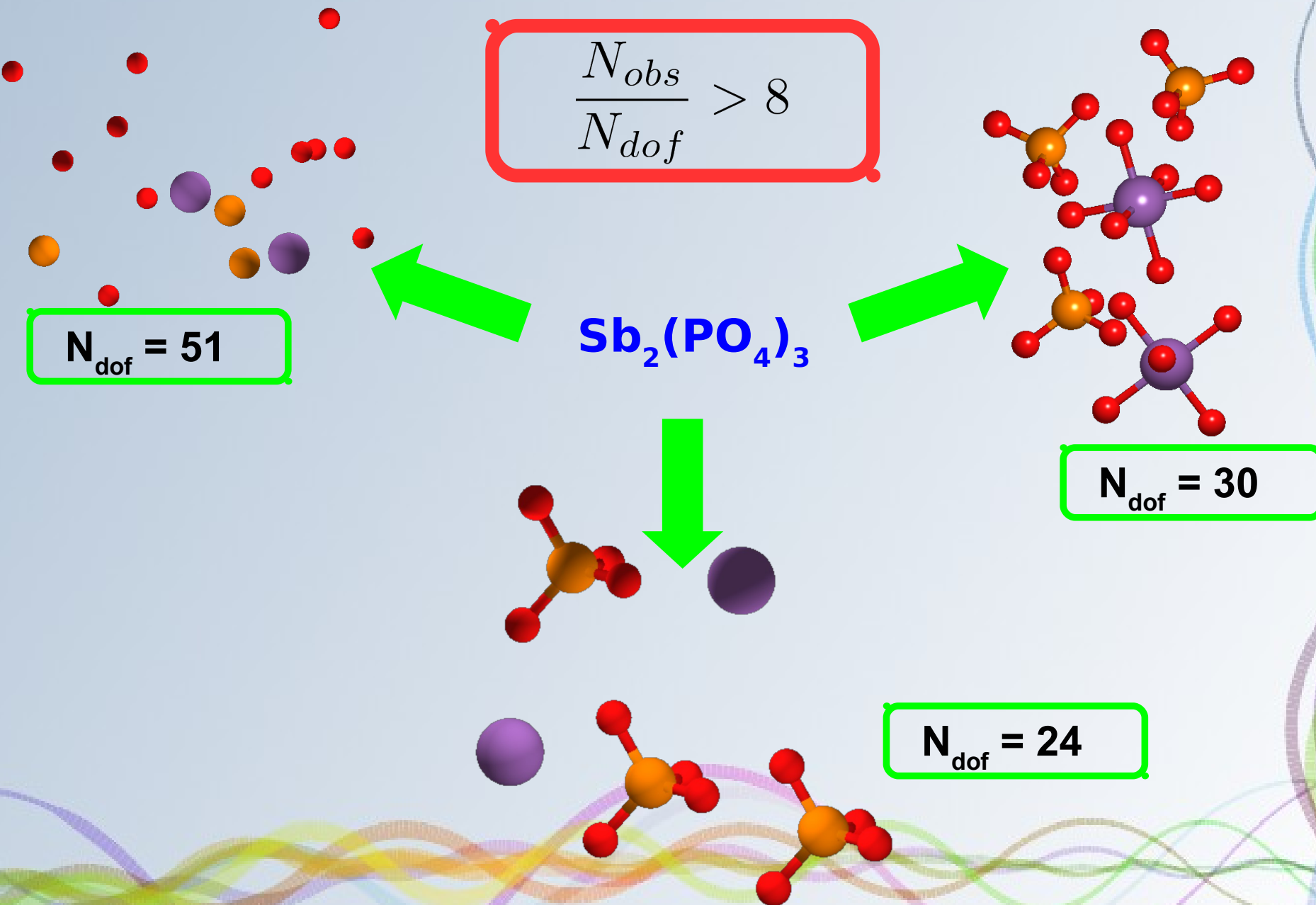
Non-molecular compounds

Crystal structure of $\text{Sb}_2(\text{PO}_4)_3$ *



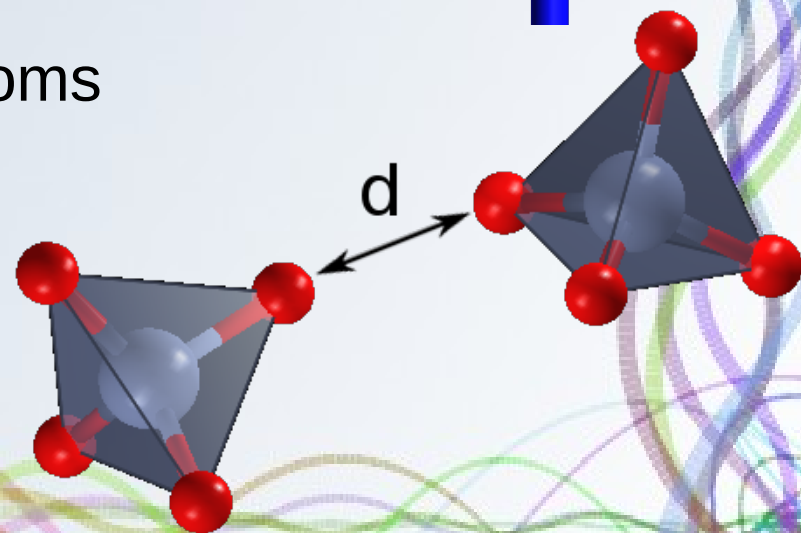
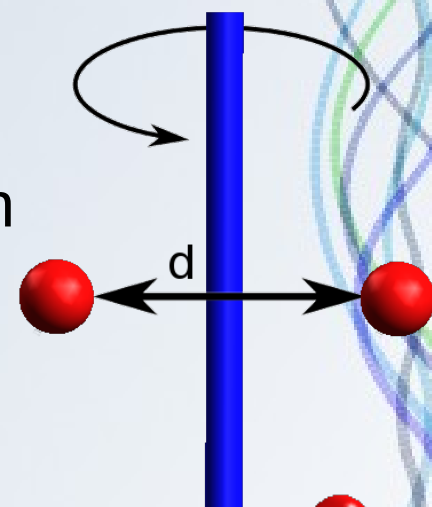
*Jouanneaux, A., Verbaere, A., Guyomard, D., Piffard, Y., Oyetola, S. & Fitch, A. N. (1991). *Eur. J. Solid State Inorg. Chem.* **28**, 755-765.

Non-molecular compounds



Non-molecular compounds

- You cannot know the number and the type of the polyhedra
- Some atoms are expected to fall on special position
- Different building blocks share some atoms



Dynamical occupancy correction (DOC)

- Falcioni, M. & Newsam, J. M. (1989). *Nature* **342**, 260-262.
- Favre-Nicolin, V. & Černý, R. (2002). *J. Appl. Cryst.* **35**, 734-743

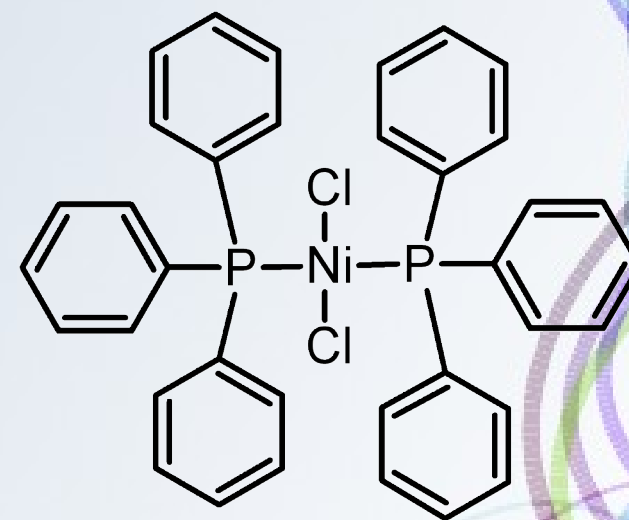
$$occupancy = \frac{1}{1 + \sum_{neighbour} |d_{min} - d_i|} \quad d_{min} = 1\text{\AA}$$

```
doc
```

```
or
```

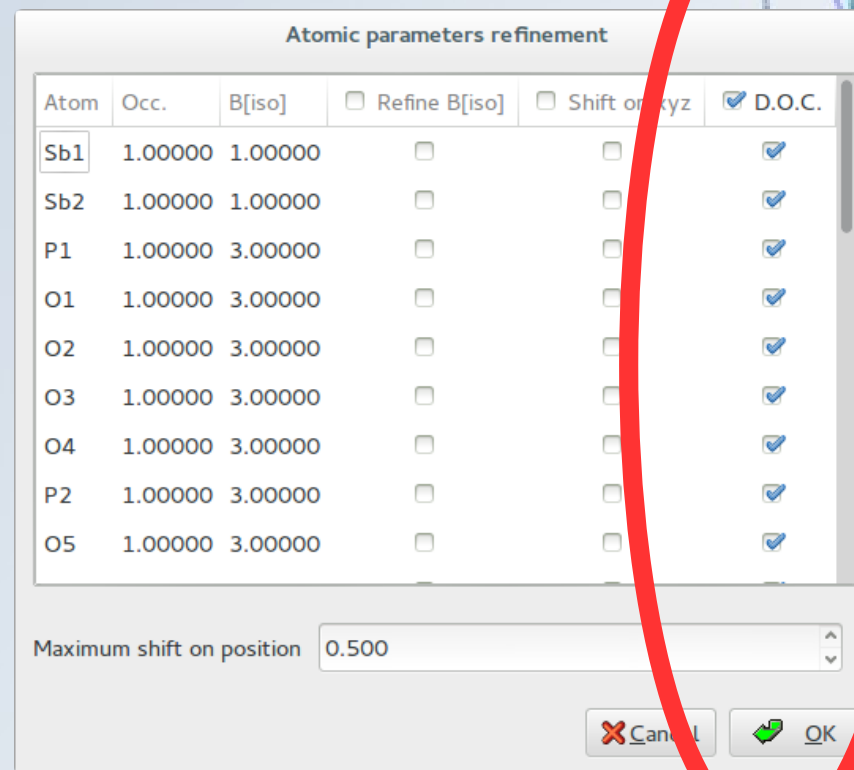
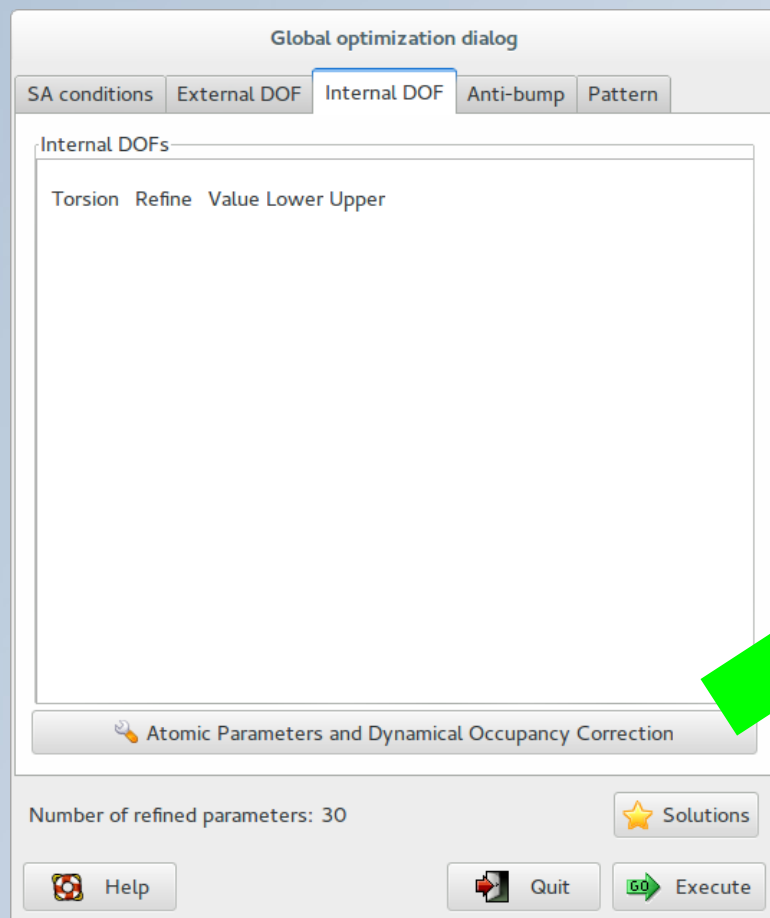
```
doc atom1 atom2 ...
```

```
doc Ni1
```



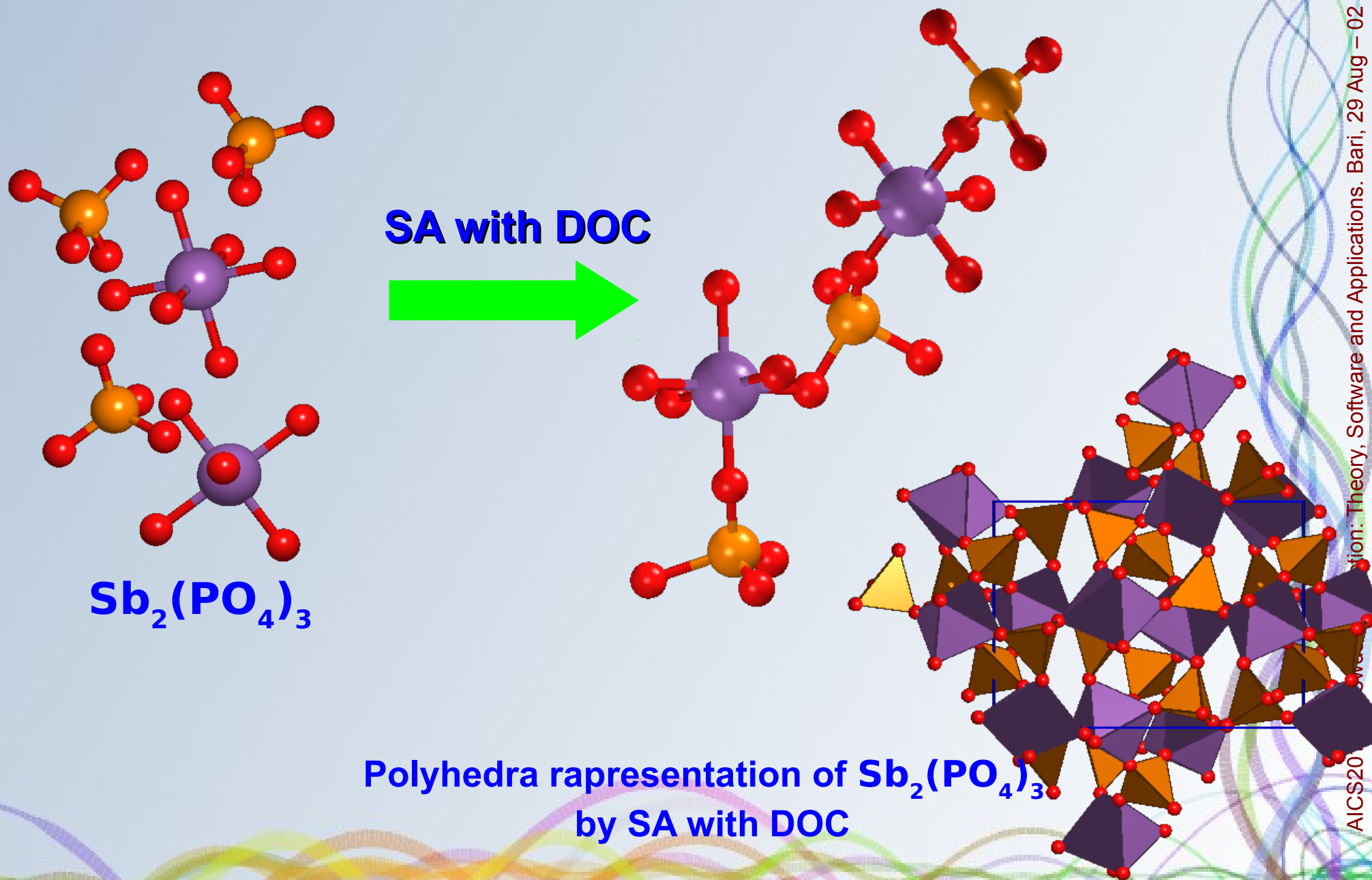
DOC is able to merge the excess atoms automatically

Dynamical occupancy correction (DOC)



DOC slows down the computation time so it should be avoided if no special positions or shared atoms are expected.

SA applied to non-molecular compounds



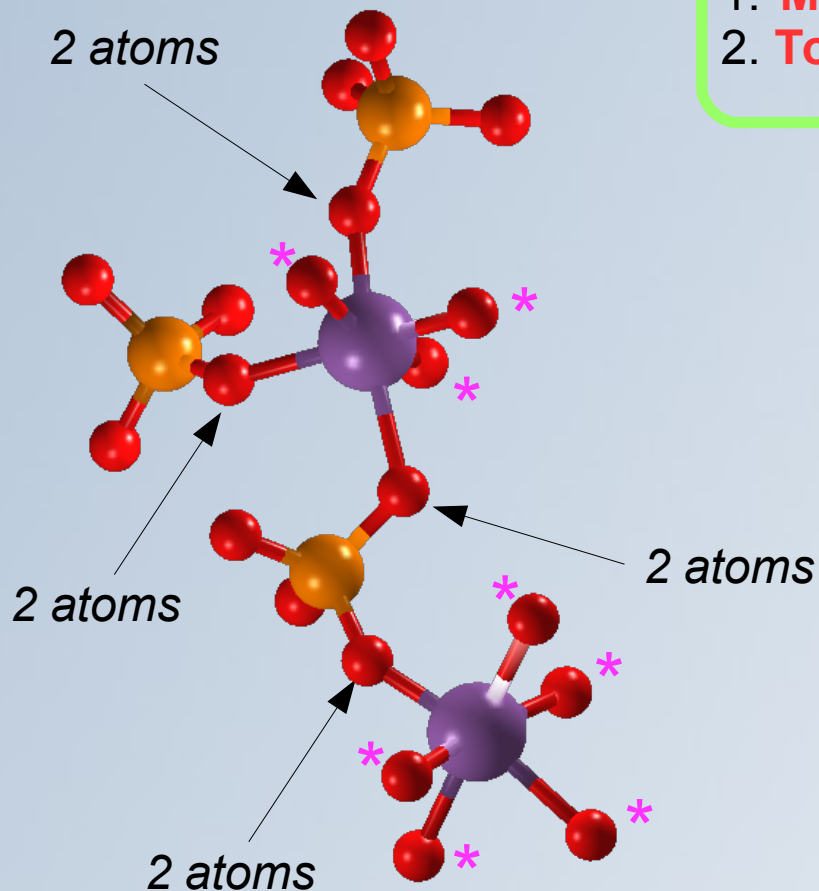
Delete duplicate atoms

1. **Modify > Delete Duplicate Atoms**
2. **Tools > Build Bonds**

Set threshold value

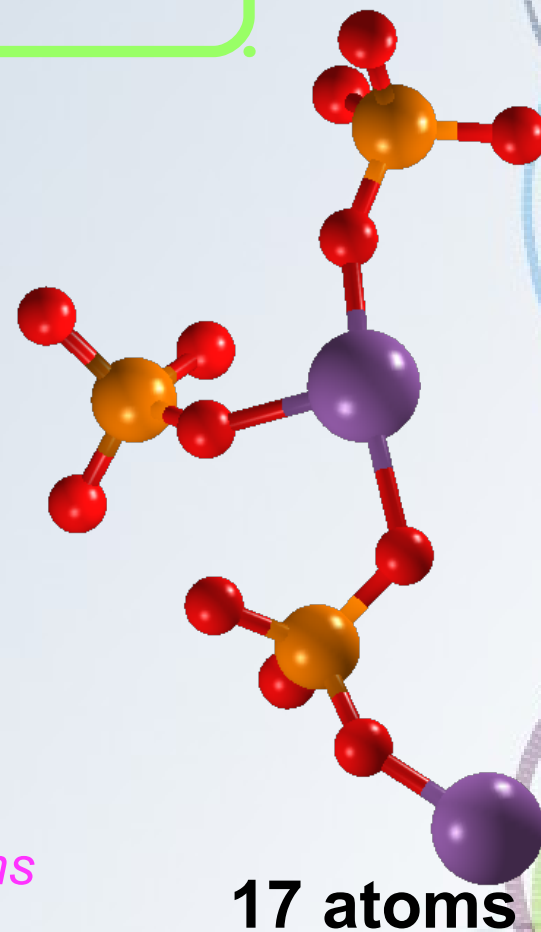
Threshold:

☒ OK



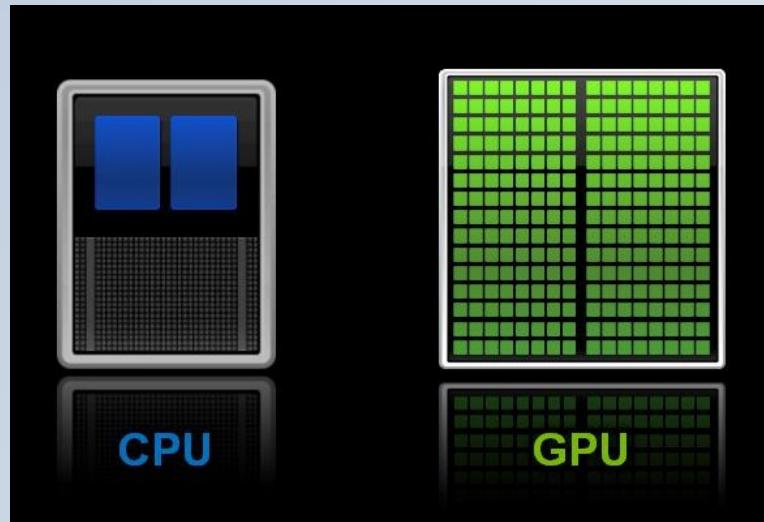
* symmetry equivalent atoms

29 atoms

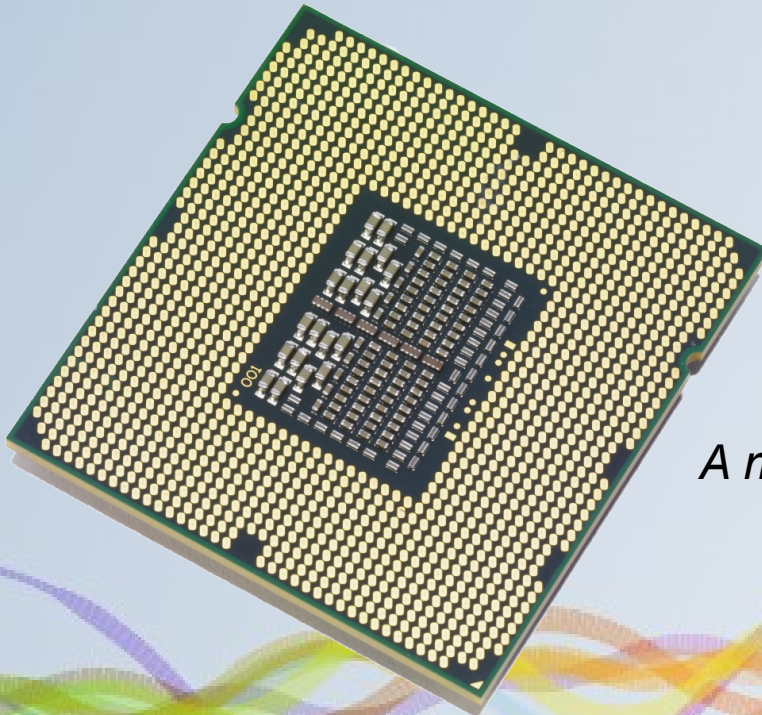


17 atoms

Understanding processors



The difference between a CPU and GPU



A modern quad-core CPU-Intel's Core i7 © Intel

Parallel Machines

Notebooks



Typically 2-6 cores

Smartphones



Typically 2-10 cores

Workstations



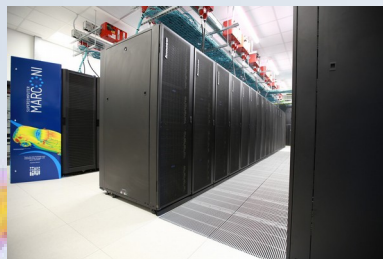
Typically 4-56 cores

Graphical Processing Units



Up to 3000 cores

Supercomputers



Marconi by CINECA (Italy)
244.800 cores in total

Three Programming models

- **Message Passing Interface (MPI)**
Distributed-memory architecture
- **Open MultiProcessing (OpenMP)**
Shared-memory architecture
- **Compute Unified Device Architecture (CUDA)
Open Computing Language (OpenCL)**
Coprocesor architecture

Three Programming models

- **Message Passing Interface (MPI)**

Distributed-memory architecture

- **Open MultiProcessing (OpenMP)**

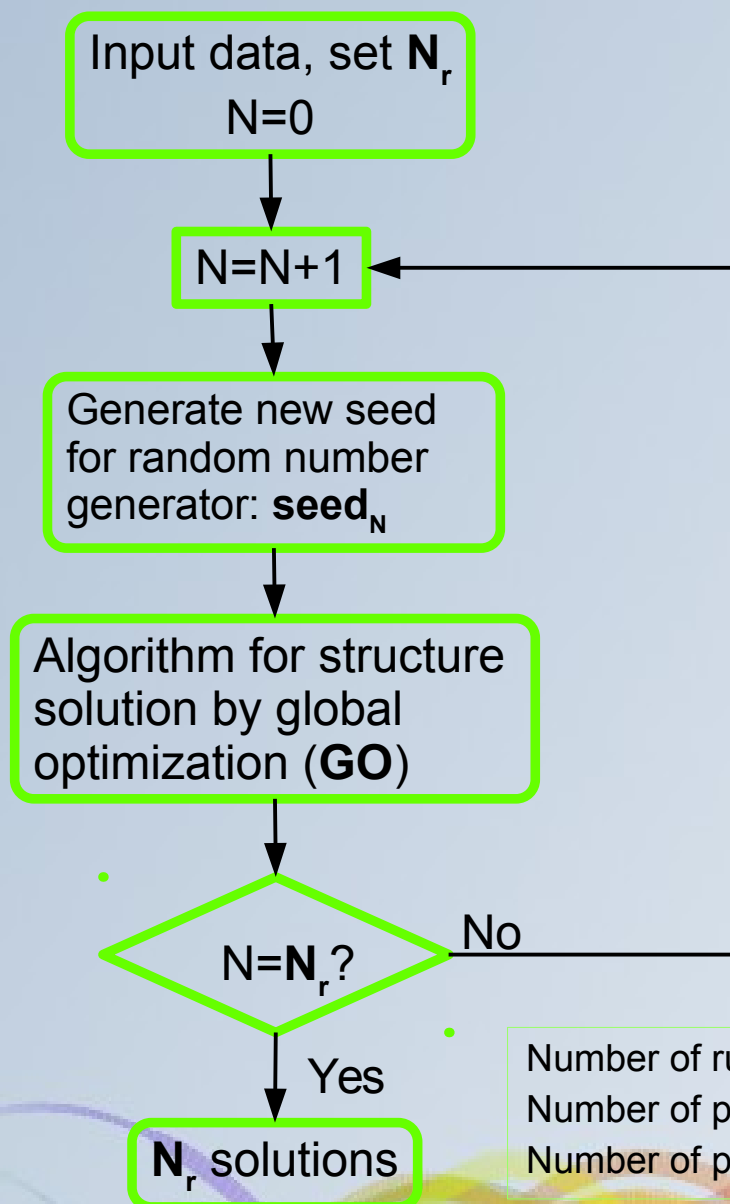
Shared-memory architecture

- **Compute Unified Device Architecture (CUDA)
Open Computing Language (OpenCL)**

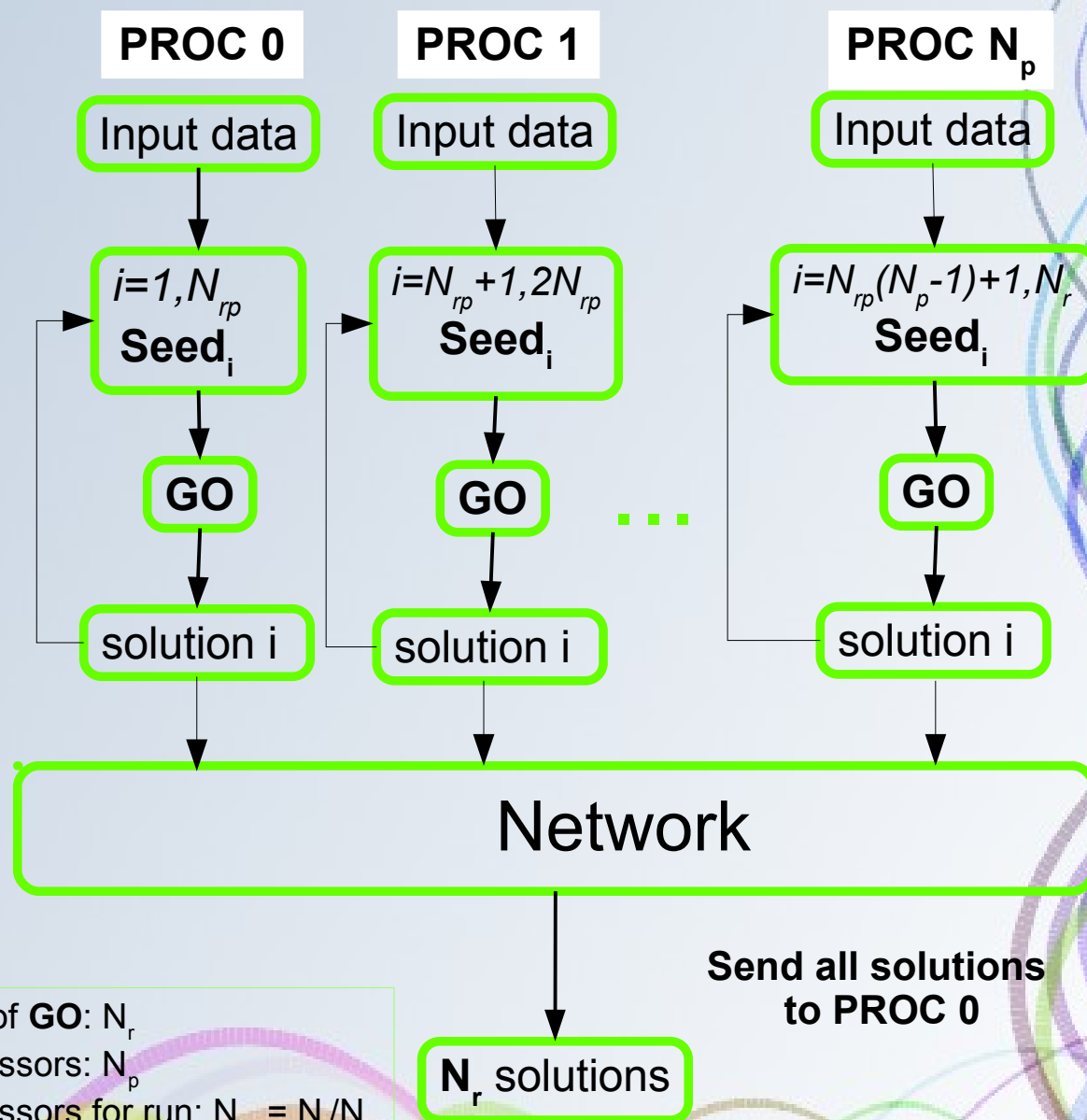
Coprocessor architecture

Parallelization of structure solution by global optimization

Serial program



Parallel program



Number of runs of GO: N_r
Number of processors: N_p
Number of processors for run: $N_{rp} = N_r/N_p$

Send all solutions
to PROC 0

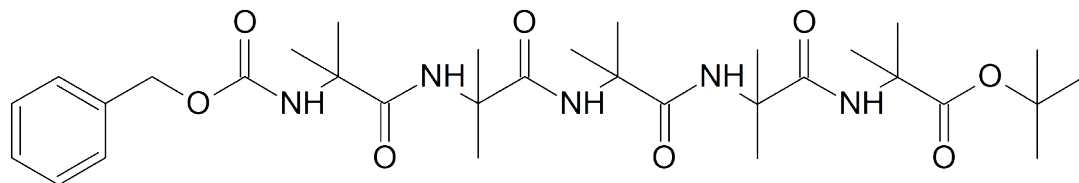
N_r solutions

Running the Parallel Version of Expo2014

- Computer with multi-core CPUs and Linux environment.
- Open MPI installed.
- Compiling Expo2014 from source and linking with MPI libraries
- Run Expo2014 by using the launcher **mpirun** with the appropriate options.

```
mpirun -np 10 expo input_file.exp
```

Structure solution of small peptide



Z-(Aib)₅-O-t-Bu

DOF

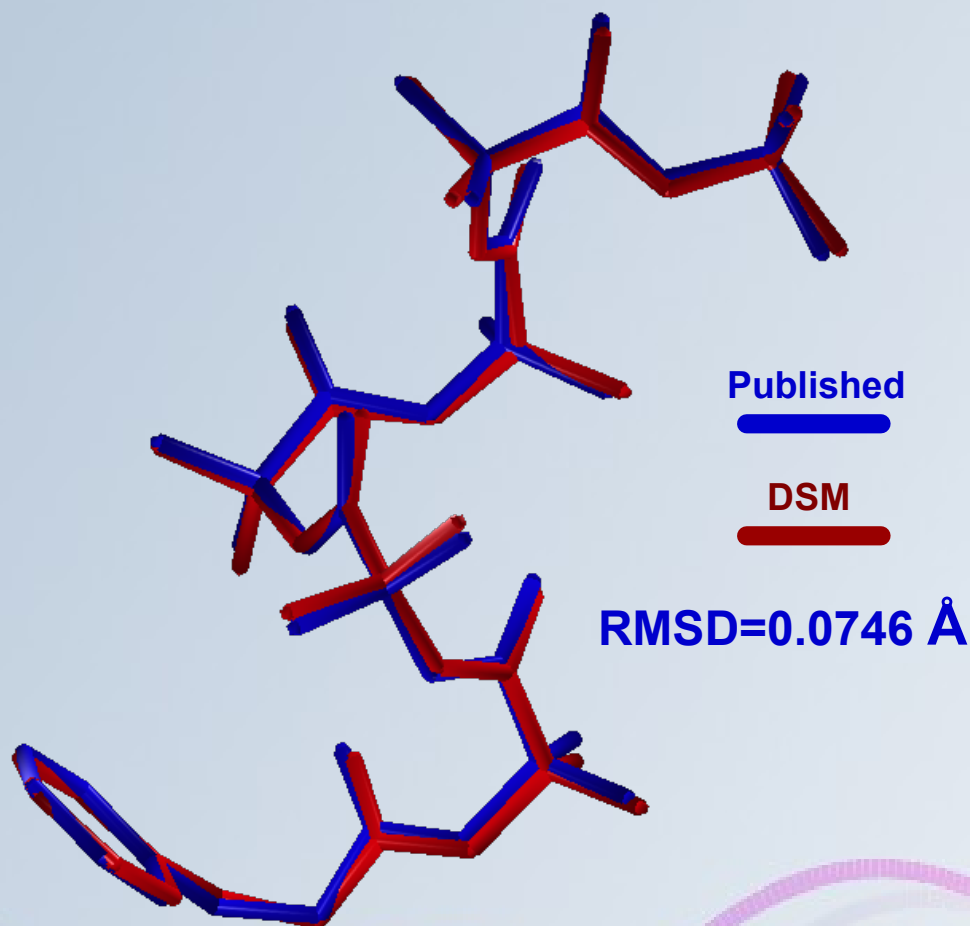
Internal: 20
External: 6

Algorithm settings

NRUN: 100
NITER: 1000

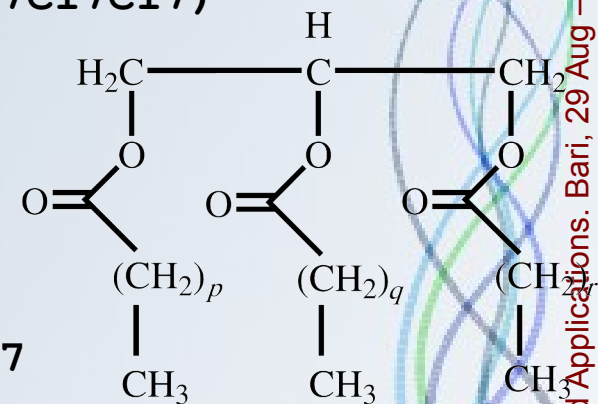
Time

Single CPU-core: 500 h
20 CPU-cores: 25 h
Parallel speedup: 20



Structure of mono-acid β -triacylglycerols

```
%job beta-1,2,3-tris(heptadecanoyl)glycerol (beta-C17C17C17)
%structure glycerol17
%data
pattern av0044C17C17C17sup3.rtv
wave 0.850047
synch
space P-1
cell 11.86642 51.4495 5.43208 72.765 100.0950 120.577
%frag glycerol17_rand.mol
%sannel
nrun 100
bump
```



RMSD=0.0746 Å

DOF = 6+53

Published

DSM

DS with Low Quality Diffraction Pattern

- **Bond valence restraints**
- **Anti-bumping restraints**
- **Molecular geometry restraints**

Bond Valence Restraints

Atomic valence V_i of atom i in crystal structure is the sum of individual bond valences S_{ij}

$$V_i = \sum_j S_{ij} \quad S_{ij} = \exp\left(\frac{R_0 - R_{ij}}{B}\right)$$

R_{ij} distance between atoms i and j

R_0, B bond valence parameters (**bvparmyyyy.cif** maintained by I.D. Brown and available from <http://www.iucr.org/resources/data/datasets/bond-valence-parameters>)

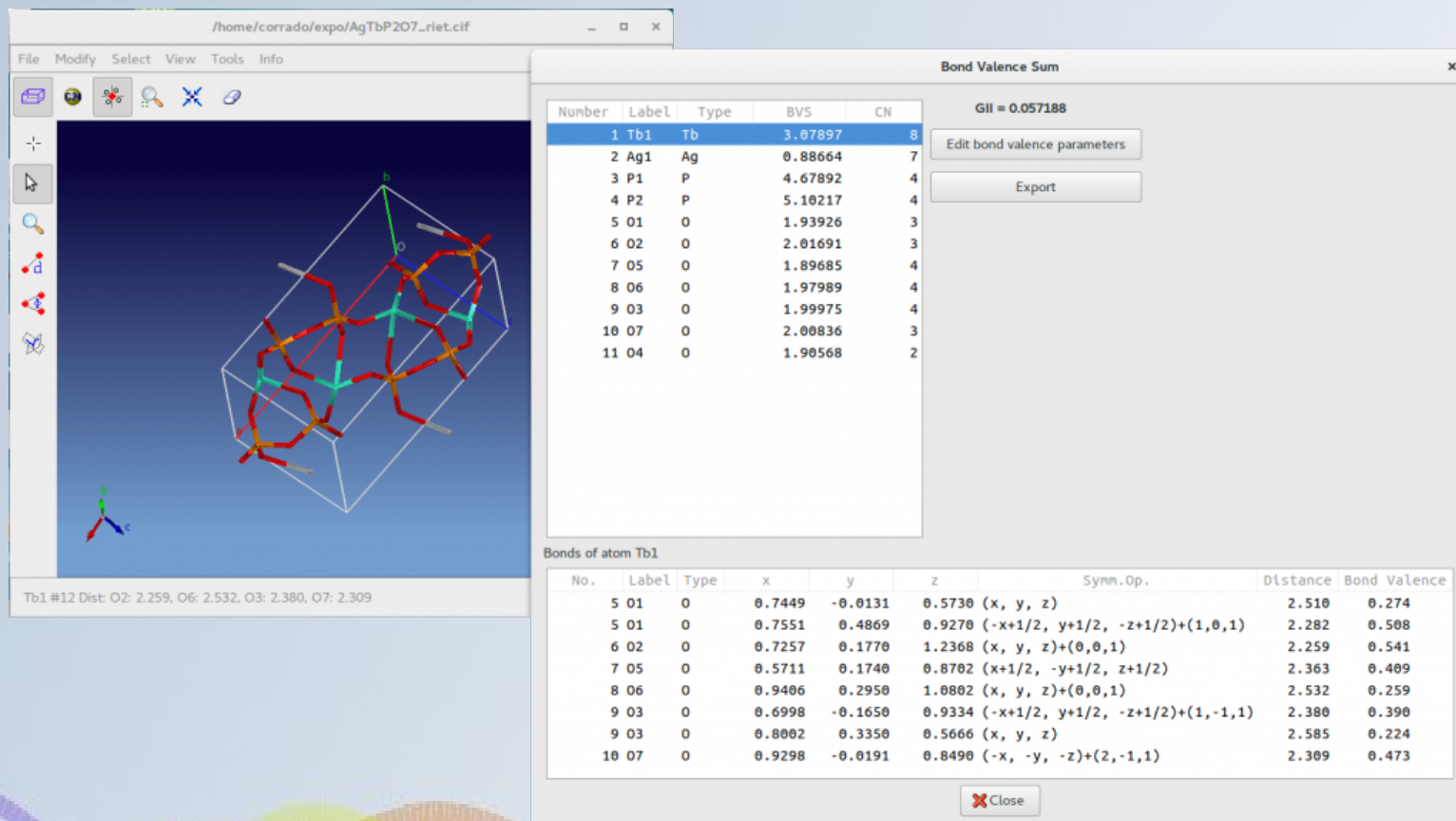
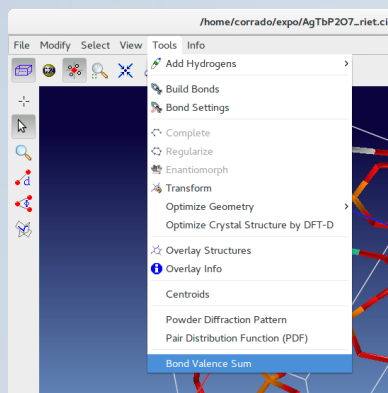
$$G_{ii} = \sqrt{\frac{1}{N} \sum_{i=1}^N (V_i - V_i^0)^2} \quad \text{global-instability index}$$

The estimated values V_i can be incorporated as restraints in the cost function (*):

$$CF_{VB} = \sum_i w_i (V_i - V_i^0)^2$$

* J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal & V. Caignaert, (1990). *Nature* 346, 343 - 345

Check Bond Valence Sum



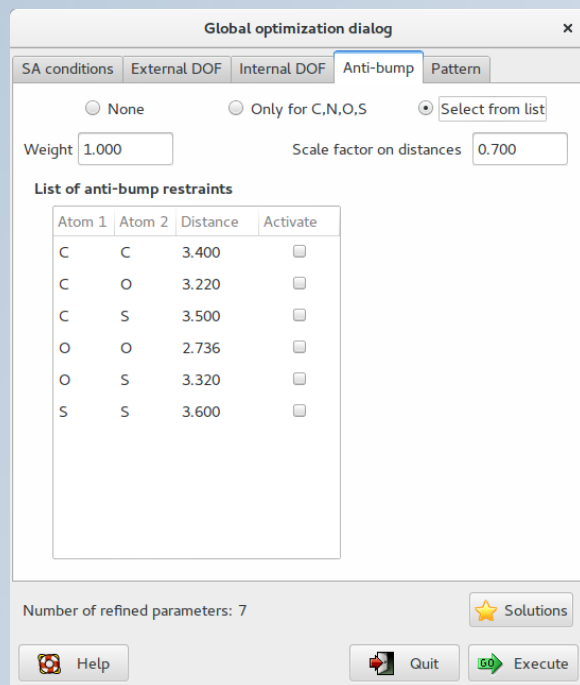
Imposing anti-bumping restraints

$$CF_{bump} = \sum_{ij}^n w_{ij} (d_{ij}^{min} - d_{ij}^{model})^{2k}$$

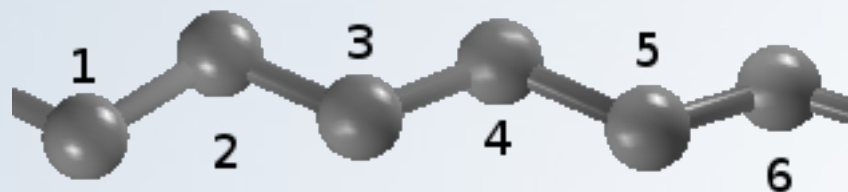
$$k = 2$$

$$d_{ij}^{model} < d_{ij}^{min}$$

$$d_{ij}^{min} = \epsilon(R_i^{vdW} + R_j^{vdW})$$



All nonbonded interactions between atoms that are separated by a path of bonds containing 4 rotatable bonds or less are excluded



Only interaction between atom 1 and atom 5 is considered

Molecular Geometry Restraints

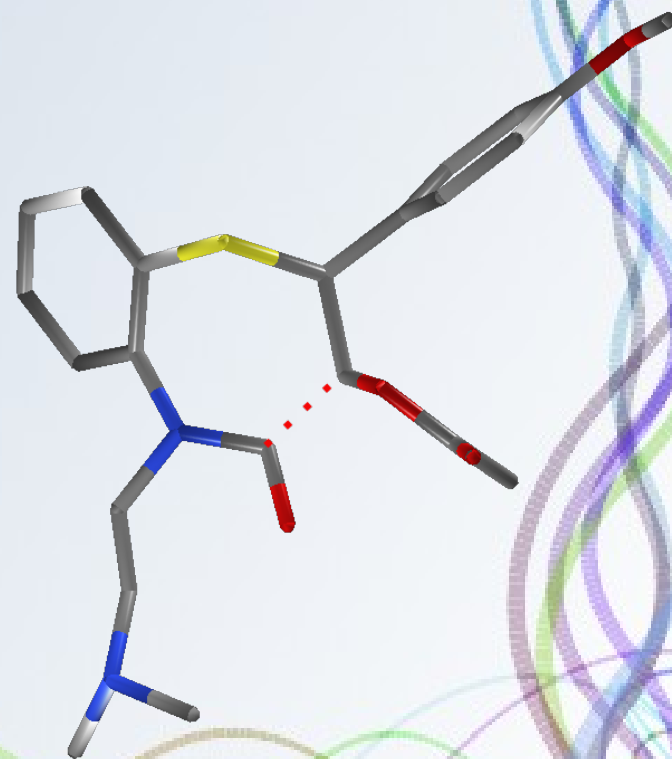
$$CF_{restraints} = \sum_i w_i MAX(0.0, |d_{target_i} - d_{AB_i}| - tol_i)^2$$

d_{AB_i} = distance between two atoms A and B

d_{target_i} = ideal distance

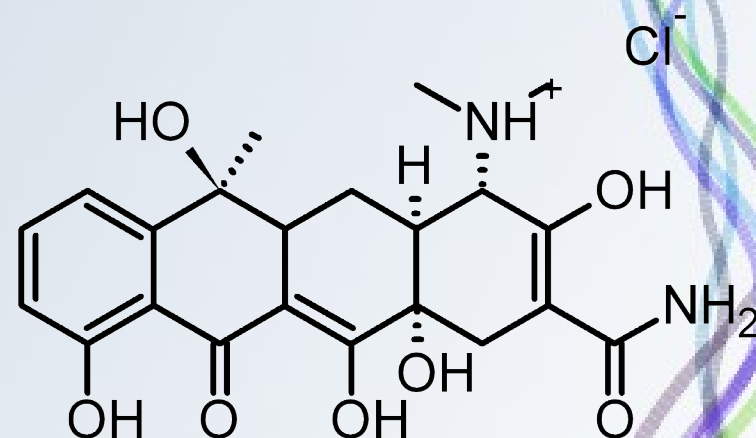
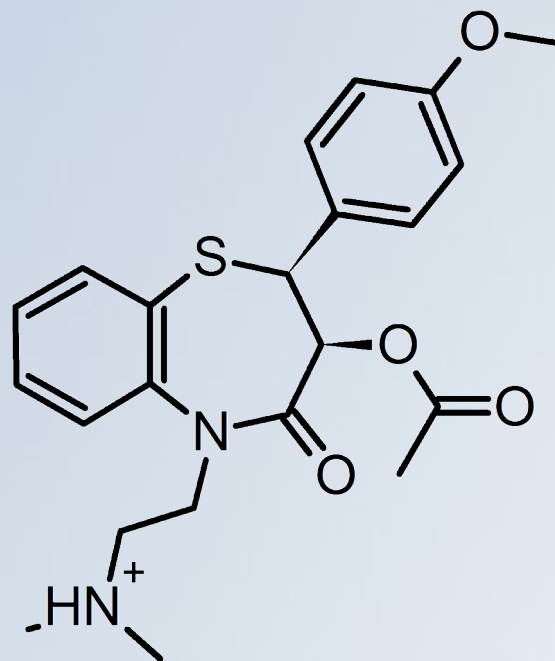
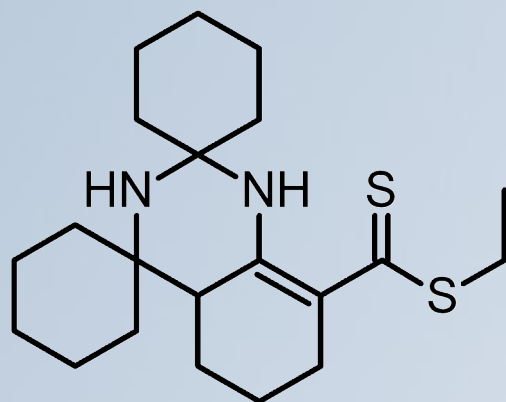
tol_i = permitted tolerance

w_i = user supplied weight

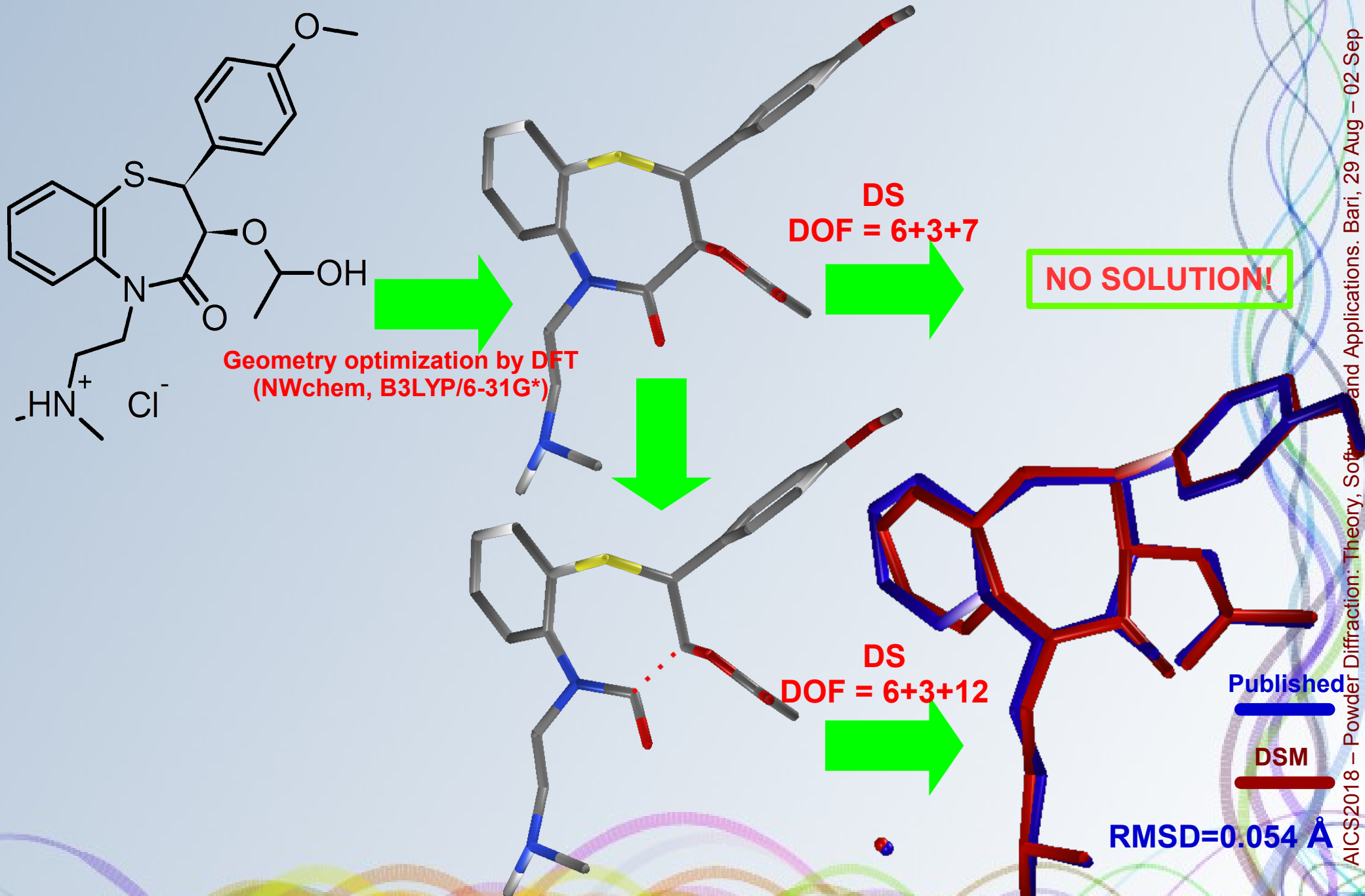


Non planar ring systems

Attention to non planar ring systems or unusual combinations of elements in functional groups



Structure Solution of Diltiazem Hydrochloride



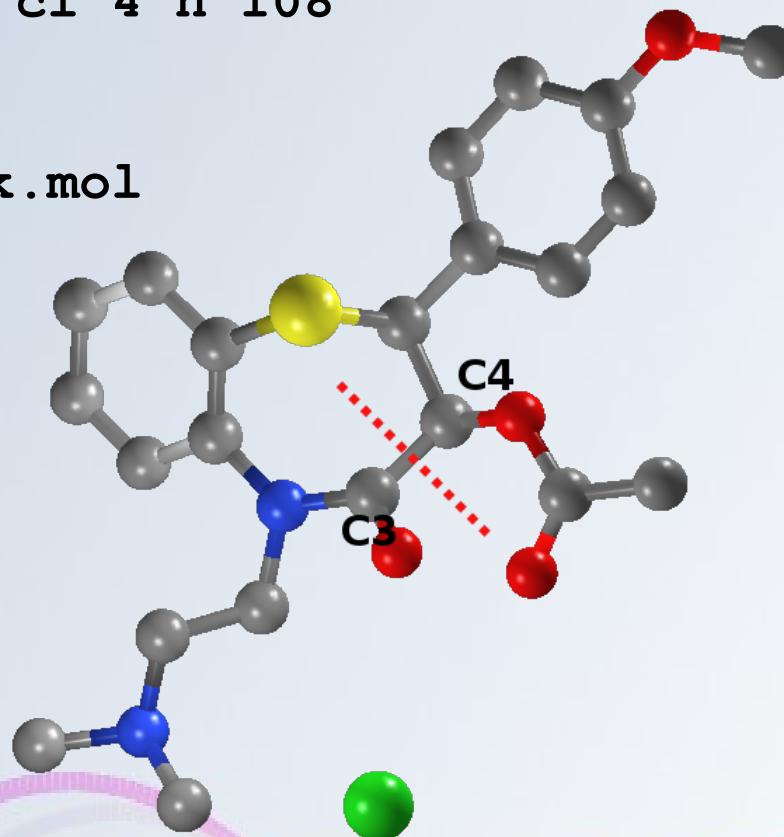
Published

DSM

RMSD=0.054 Å

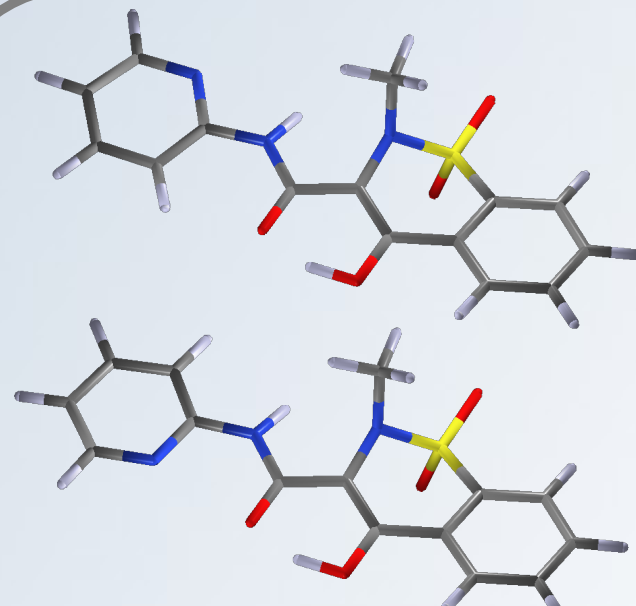
Structure Solution of Diltiazem Hydrochloride

```
%Structure diltia
%Job diltiazem Hydrochloride
%Data
Cell      42.190   9.075   6.037   90   90   90
SpaceGroup p 21 21 21
Content    c 88 n 8 o 16 s 4 cl 4 h 108
Pattern    pd_0029.pow
Wavelength 1.54056
%fragment diltia_nw_noH_break.mol
%frag atoms C1
%sannel
nrun 100
niter 5000
rest C3 C4
%save diltia.expo
```



Combined powder X-ray diffraction data and quantum-chemical calculations

- **Optimization of the molecular geometry** to obtain accurate starting models
- **Restraints** in the Rietveld refinement
- **H atoms**
- **Solve ambiguities**
(e.g., space groups, torsion angles)
- **Refinement of crystal structure**
- **Validation** of experimental crystal structures





Two possible orientations of the pyridyl ring in the piroxicam molecule (Naelapää, K., van de Streek, J., Rantanen, J., and Bond, A. D. (2012), *J.Pharm. Sci.* 101, 4214–4219)

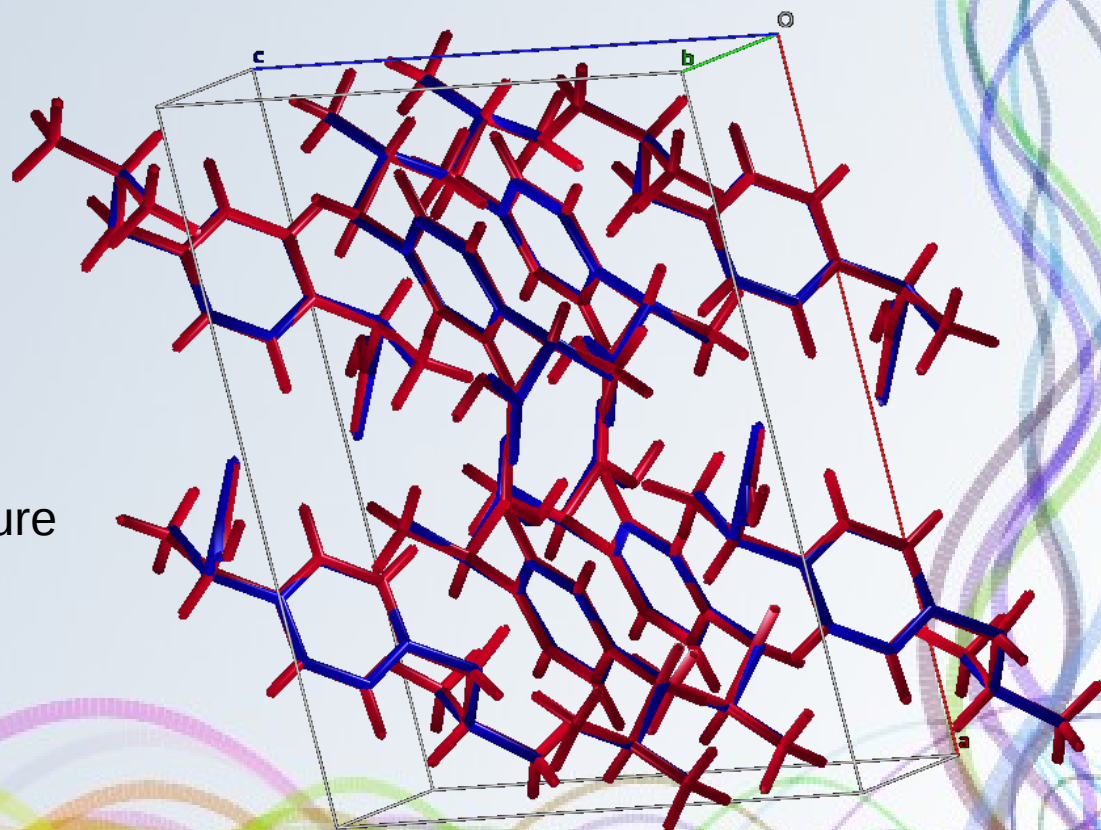
Assessing the solution with DFT-D

Theoretical approach: plane wave (PW) density functional theory with dispersion correction (DFT-D)

RMSD for non H-atoms above 0.25 Å could indicate incorrect experimental crystal structure *

Ibuprofen
RMSD=0.023 Å

-  Experimental crystal structure
-  DFT-D3 with NWChem



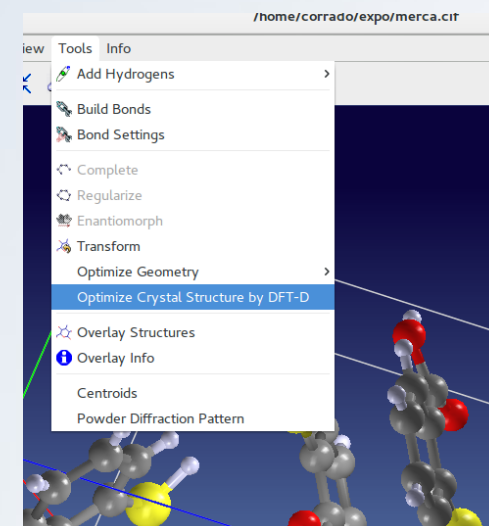
*Jacco ven de Streek *et al.* Validation of molecular crystal structures *Acta Cryst.* (2010). B66, 544–558

DFT-D: Howto

| Software | Academic price (€) | Link |
|------------------|--------------------|--|
| VASP | 4,000 | www.vasp.at |
| CASTEP | 1,800 | www.castep.org |
| CRYSTAL | 1,000 | www.crystal.unito.it |
| Quantum ESPRESSO | free | www.quantum-espresso.org |
| NWChem | free | www.nwchem-sw.org |
| Abinit | free | www.abinit.org |

Hardware: multi-core Linux Workstation

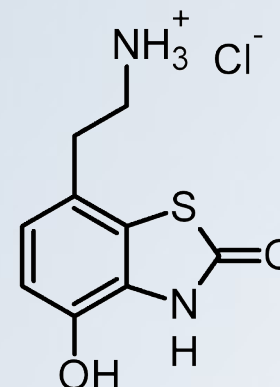
Time: approx. 100 hrs for small molecules on single CPU



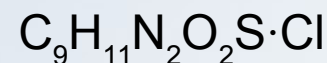
Structure refinement of $\text{C}_9\text{H}_{11}\text{N}_2\text{O}_2\text{S}\cdot\text{Cl}$

Input file for Rietveld refinement:

```
%Structure ammonium
%Initialize
%Job ethylammonium chloride (C9H11N2O2SCl)
%Data
  Cell 7.555 14.640 10.246 90 109.30 90
  SpaceGroup p 21/a
  Pattern ammonium.xy
  Wavelength 1.54056
%fragment ammonium_riet.cif
%rietveld
```



2-(4-Hydroxy-2-oxo-2,3-dihydro-1,3-benzothiazol-7-yl)
ethylammonium chloride



From graphical interface:

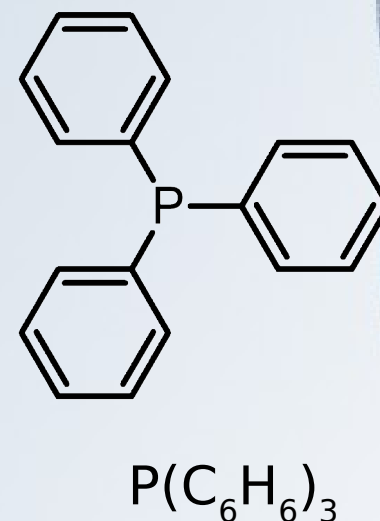
- File > Import Diffraction Pattern
- File > Import Structure
- Refine > Rietveld

Exercise 1 – structure determination of $\text{P}(\text{C}_6\text{H}_6)_3$

- Perform all the steps of the crystal structure determination of the Triphenylphosphine
Wavelength: 1.000972 (synchrotron radiation).
Data file: `Examples/pph3/pph3.xy`

Tips

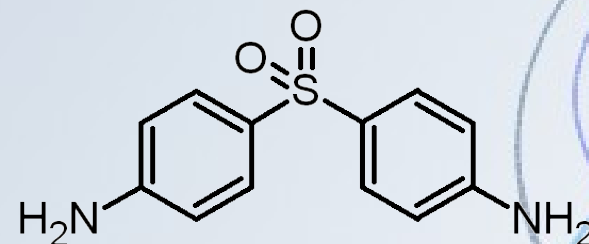
File > Import Diffraction Pattern
Pattern > Range (2-theta max = 60°)
Pattern > Indexing
 $Z = 4$



- Compare the final model with the published model `pph3_pub.cif` in the folder `Examples/pph3` (**Tools → Overlay structures**)

Exercise 2 – structure determination of dapsonsone ($\text{C}_{12}\text{H}_{12}\text{N}_2\text{O}_2\text{S}$)*

- Perform all the steps of the crystal structure determination of the dapsonsone loading the input file `Examples/dapsonsone/dapsonsone.exp`.



- The structure model obtained at the end of Direct Methods procedure, executed on the first set of phases (default choice), is not interpretable. You have to try different phasing trials. The following steps automatically carry out the structure solution process for each stored phasing trial and finally the best structure model is selected:
 - “Solve” → “Explore trials”
 - “Select all new trial”
 - “Go”

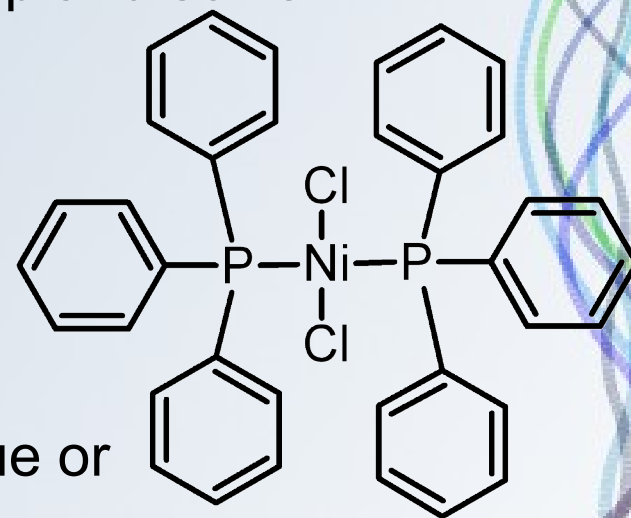
Exercise 2 – structure determination of $\text{C}_{36}\text{H}_{30}\text{Cl}_2\text{NiP}_2$

Part 1

- Load the input file `Examples/nickel/nickel_dm.exp` (**File** → **Load & go**), determine the space group and solve the structure by using the Direct Methods

Tip

correct space group is $P 2_1/c$

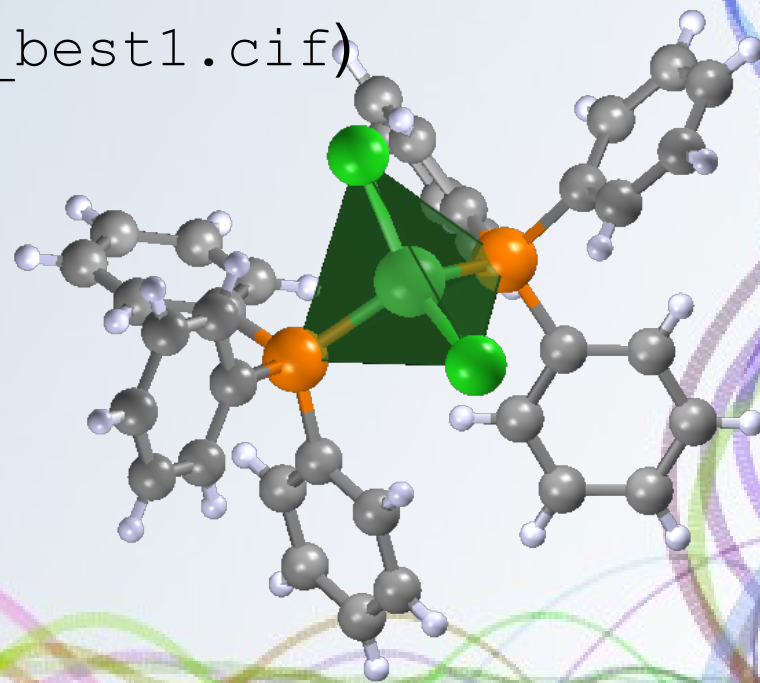


- If direct methods fail apply the RAMM technique or the `alltrials` procedure
- Save your best model in a `nickel_part1.cif` format

Exercise 2 – structure determination of $\text{C}_{36}\text{H}_{30}\text{Cl}_2\text{NiP}_2$

Part 2

- Solve the structure using the real space technique:
 1. Load the input file `Examples/nickel/nickel_ds.exp` (**File** → **Import**)
 2. Run Simulated Annealing (**Execute**)
- Compare the best solution of SA (`nickel_best1.cif`) with the `nickel_part1.cif`
- Use the polyhedra tool to visualize the tetrahedral coordination of Ni



Download this lecture

<http://www.ba.ic.cnr.it/softwareic/expo/tutorials-and-lectures/>

Contact, software download and info

<http://www.ba.ic.cnr.it/softwareic/expo/>

Acknowledgements

Colleagues of the research team

A. Altomare, A. Moliterni, R. Rizzi, N. Corriero and A. Falcicchio