

EXPO

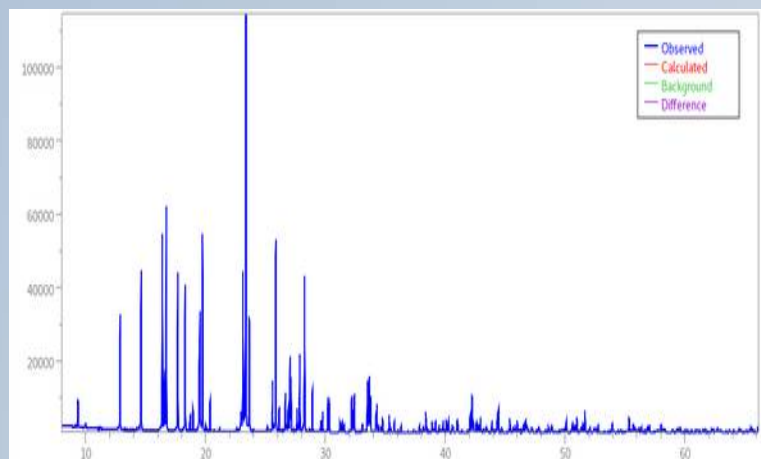
**software for solving crystal structures by powder
diffraction data: methods and application**

Corrado Cuocci

Institute of Crystallography — National Research Council (CNR), Bari, Italy

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

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

Methods of Structure Solution

Structure solution methods

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

Traditional approaches:

- direct methods
- Patterson methods

Direct space methods

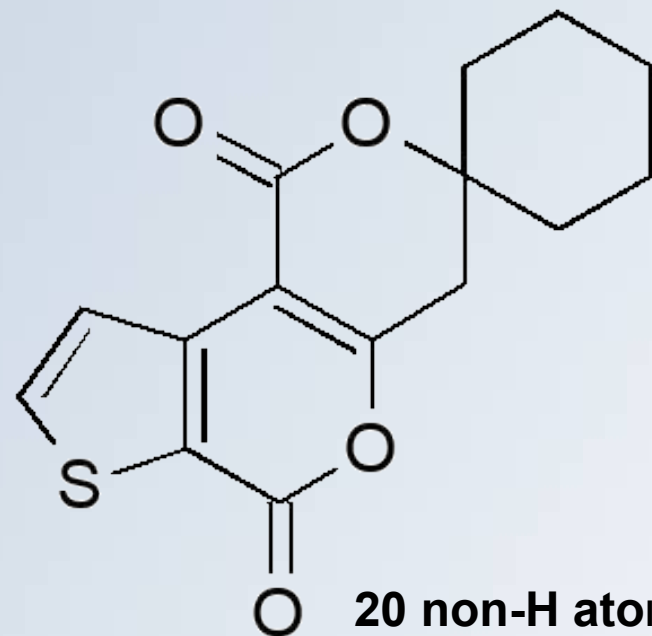
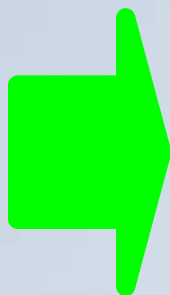
Alternative expressions: real space, global optimization, global search

Other methods:

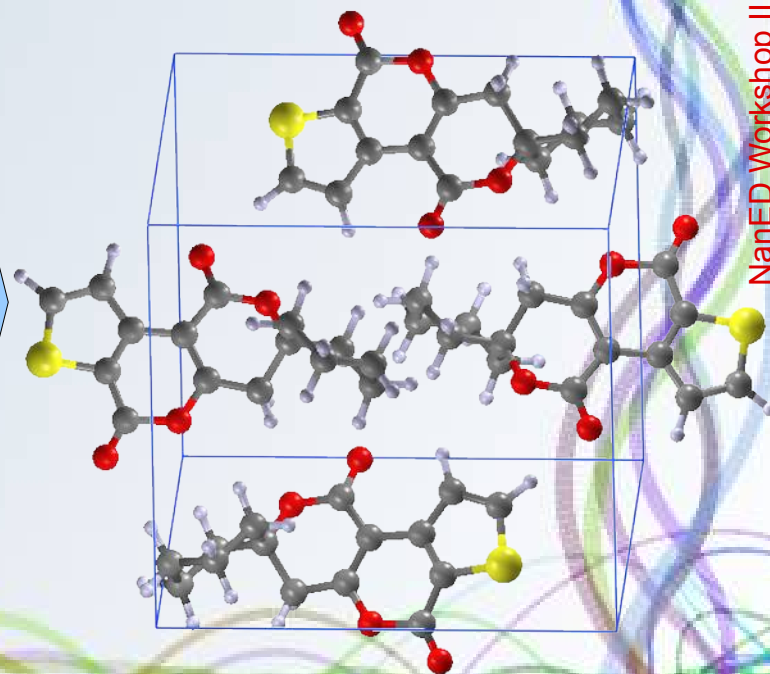
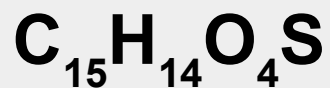
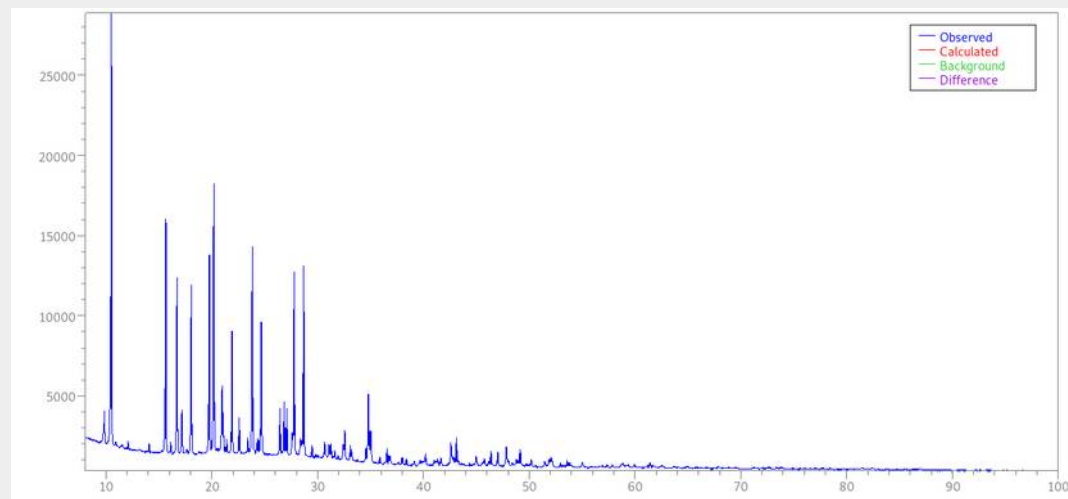
- charge flipping
- molecular replacement
-

Crystal structure determination by direct methods

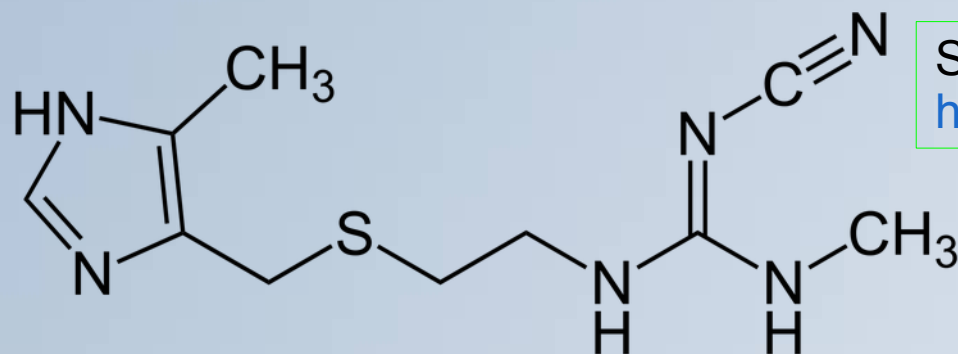
Laboratory of Industrial and
Synthetic Organic Chemistry
(LISOC)



Required information



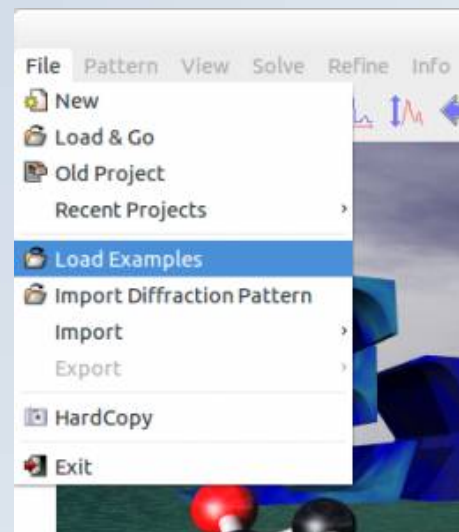
Crystal structure determination of cimetidine from powder synchrotron diffraction data



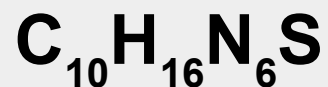
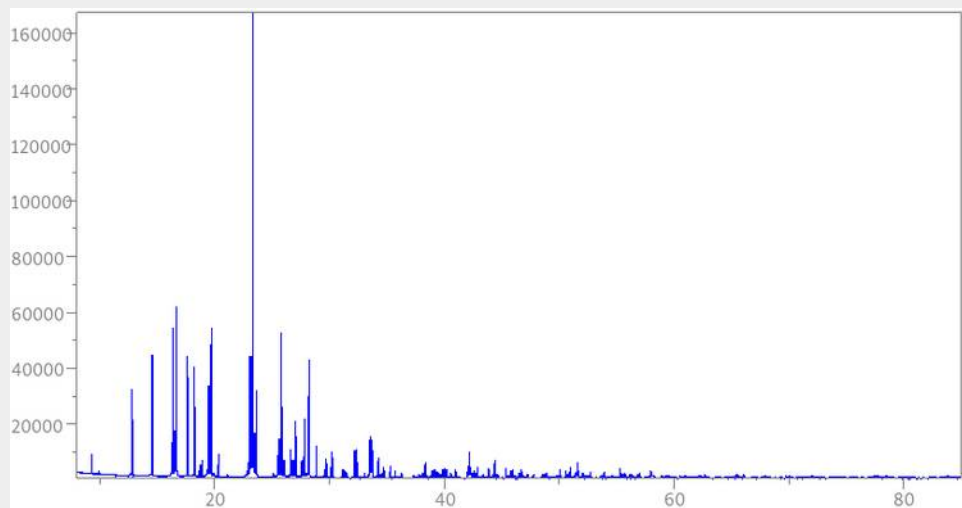
cimetidine

See tutorial 1:

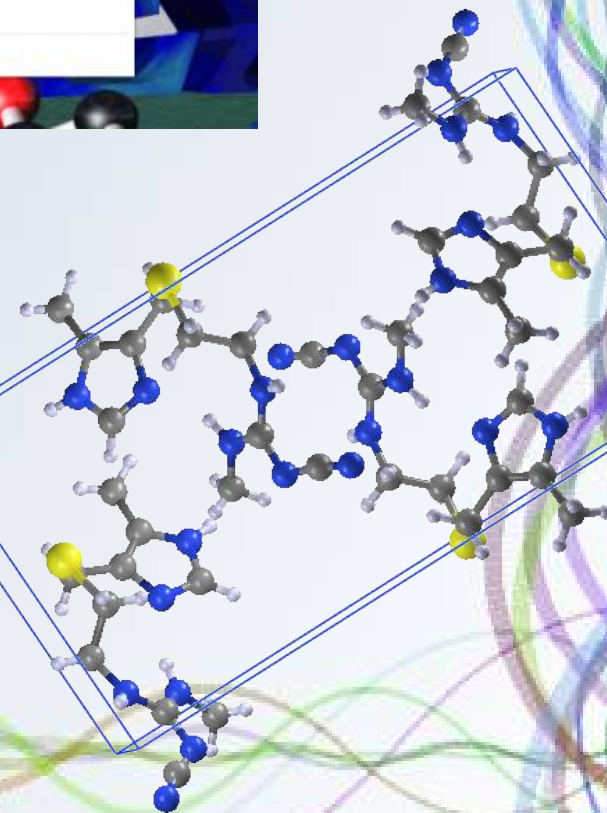
<https://www.ba.ic.cnr.it/softwareic/expo/tutorials/>



Required information



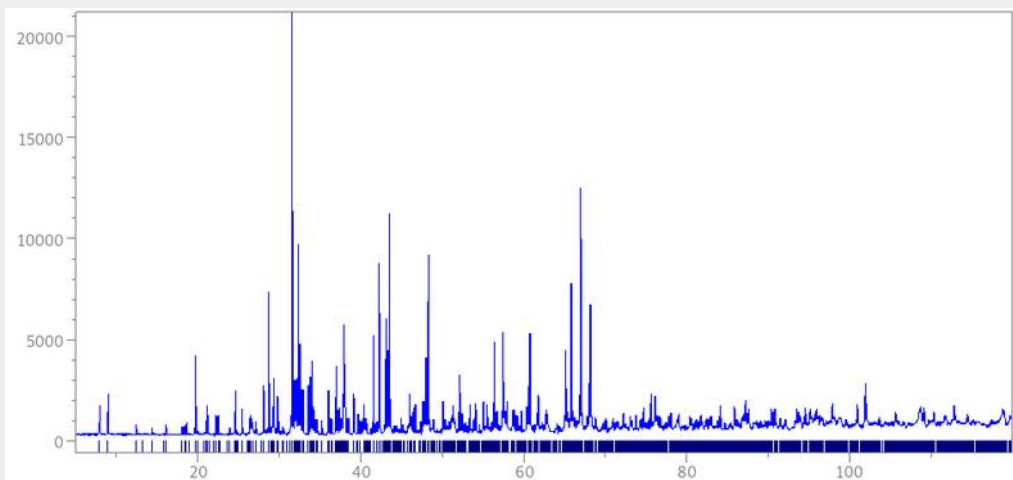
DM



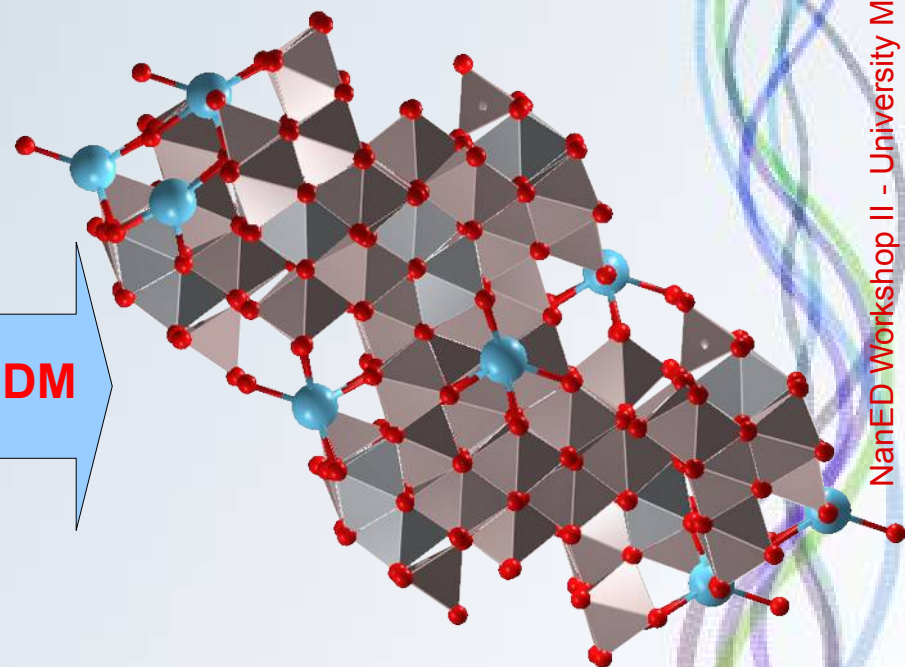
Crystal structure determination of $\text{La}_3\text{Ti}_5\text{Al}_{15}\text{O}_{37}$

```
%structure LaTi
%job LaTi from Acta Cryst. (2011).B67,455-460
%data
cell 22.59355 10.99919 9.72968 90 98.5634 90
spacegroup C2/c
content (Al9LaO19Ti2)8
pattern LaTi.rtv
%continue
```

Required information



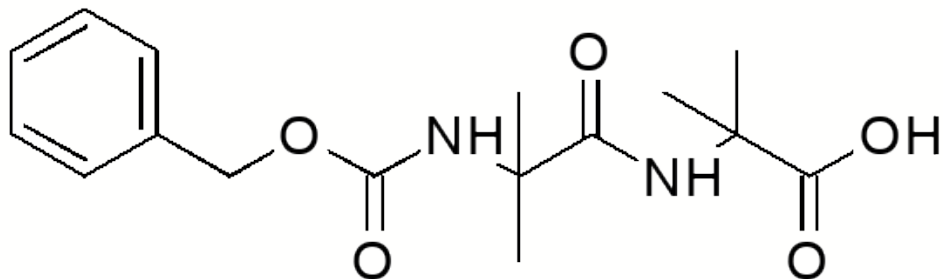
DM



Example included in the folder:

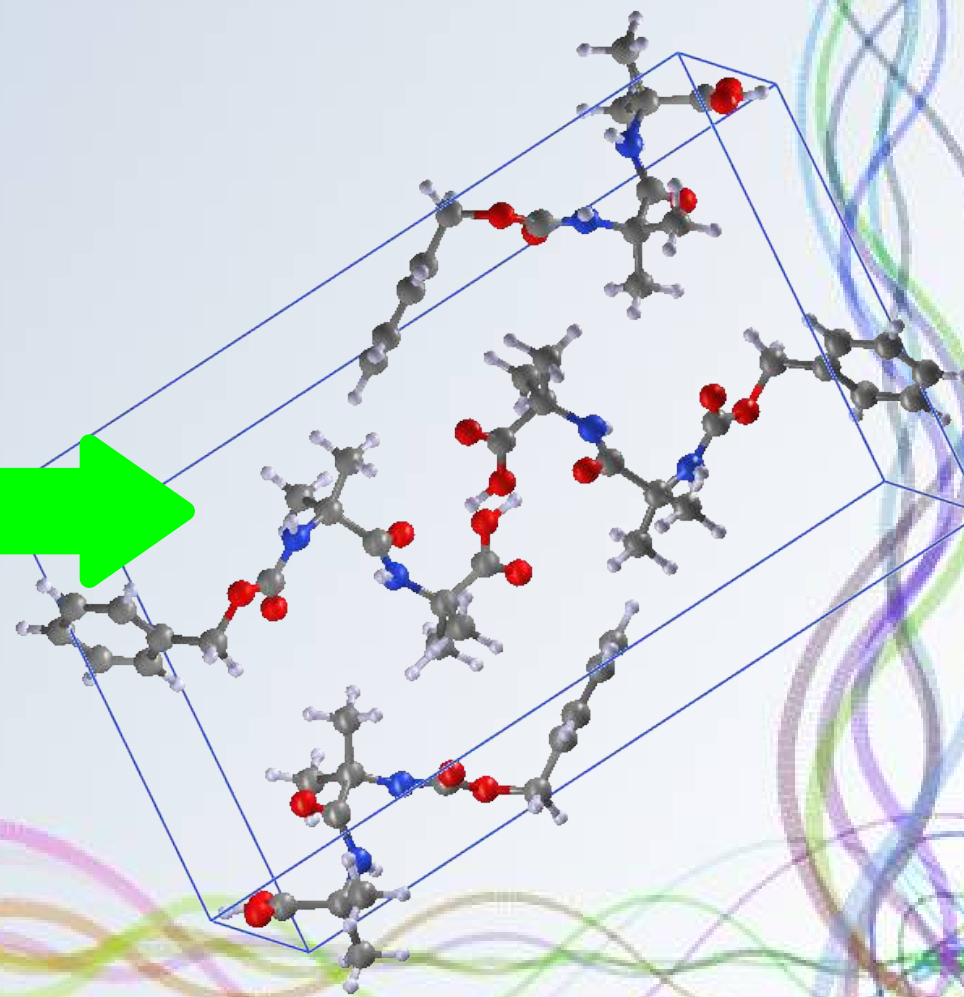
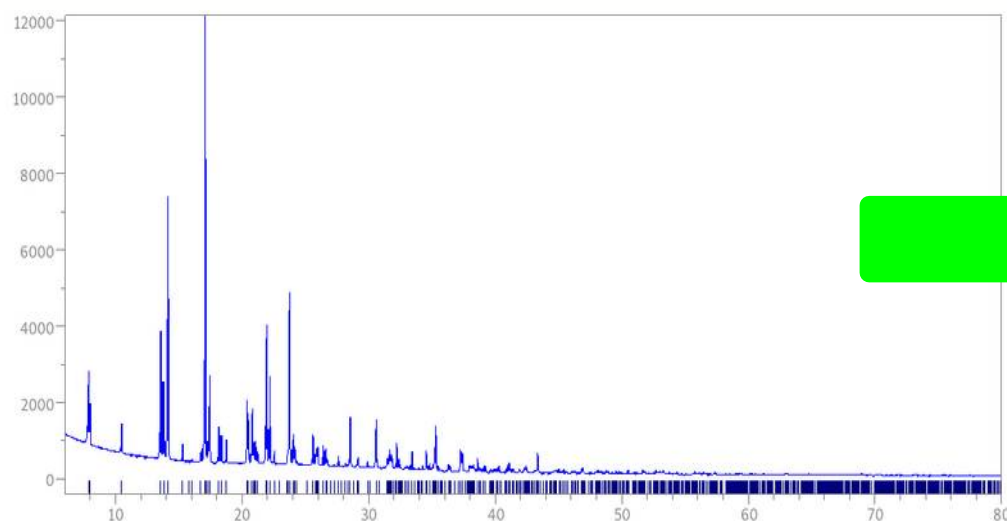
EXPO&more computer sessions EXPO/Solution by Direct Methods/Default

Crystal structure solution from powder diffraction data exploring all trials

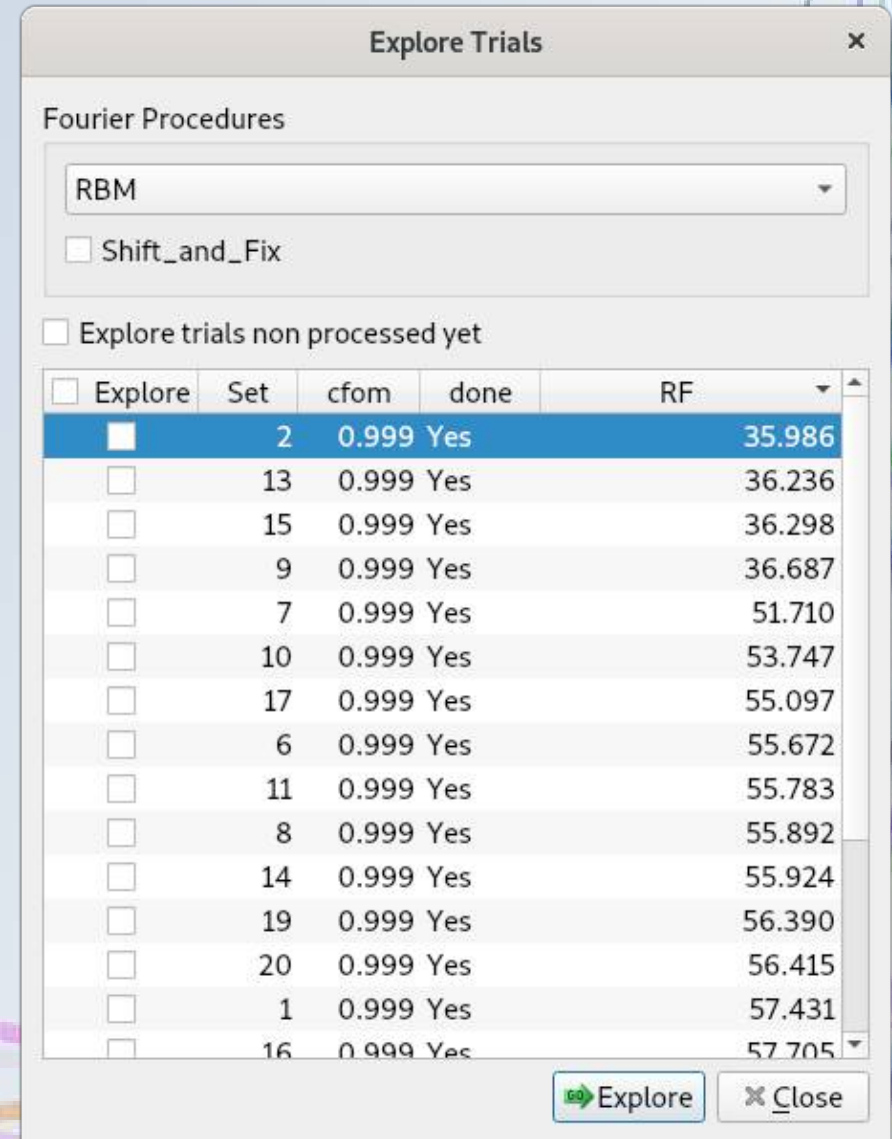
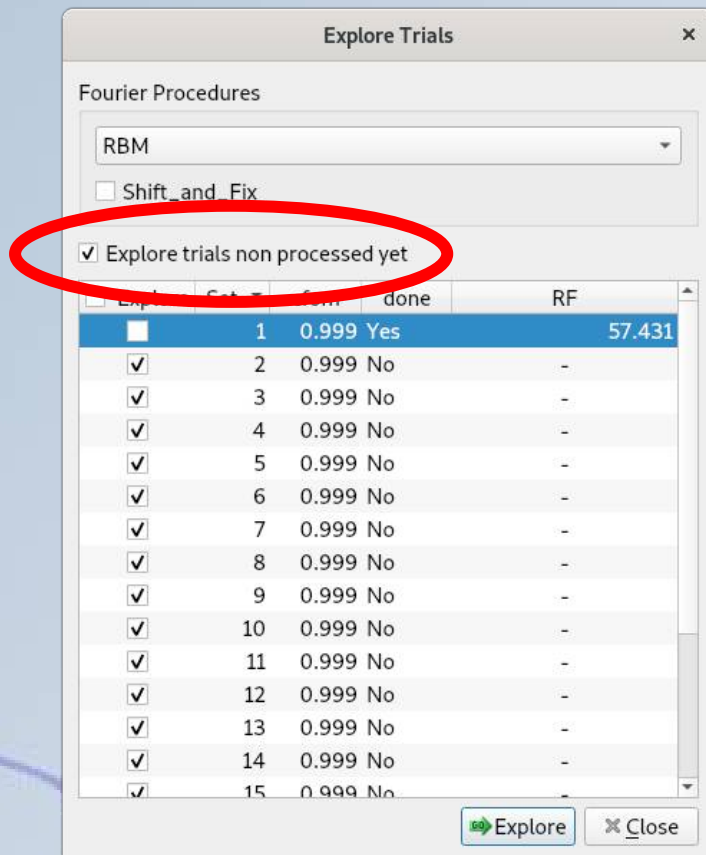
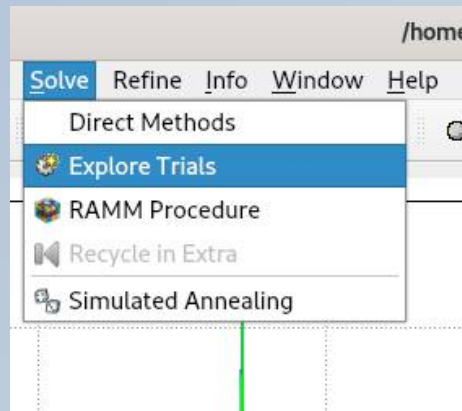


All trials were explored during the Direct Methods procedure

Required information



Exploring trials

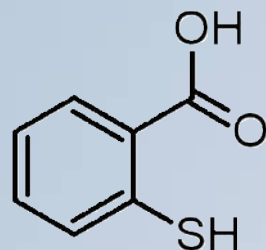


Crystal structure solution of 2-Mercaptobenzoic acid compound by direct methods

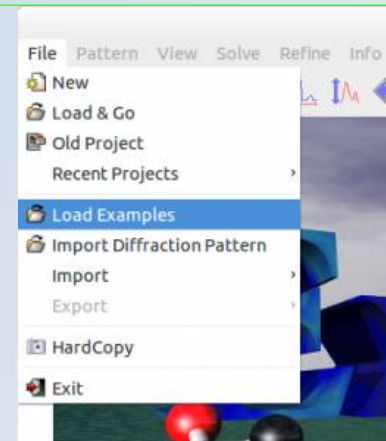
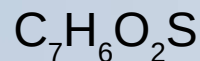
of 2-Mercaptobenzoic acid

See tutorial 2:

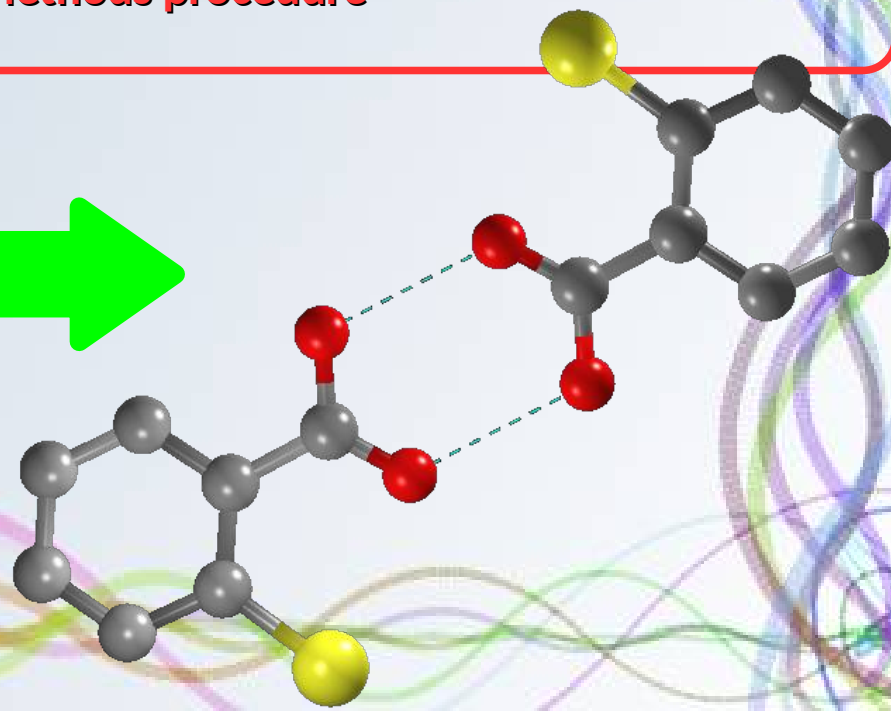
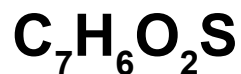
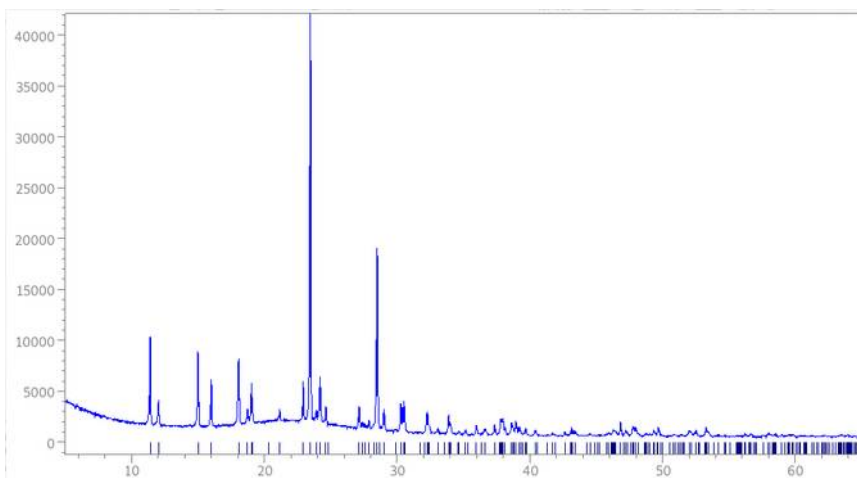
<https://www.ba.ic.cnr.it/softwareic/expo/tutorials/>



2-mercaptobenzoic acid



All trials were explored during the Direct Methods procedure



Crystal structure determination by direct methods

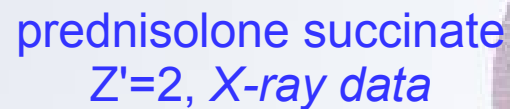
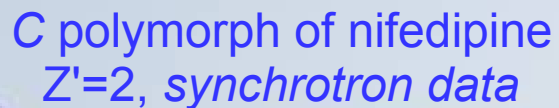
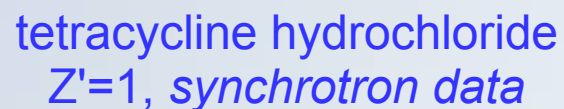
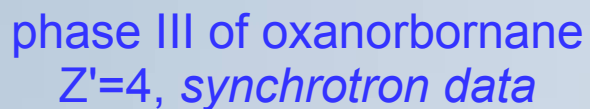
```
%job cri14 - C15H14O4S
%structure cri14
%data
pattern cri14_data.xy
wave 1.54056
%ntreor
%dicvol
%continue
```

Input file starting from indexing

```
%job cri14 - C15H14O4S
%structure cri14
%data
pattern cri14_data.xy
wave 1.54056
cell 9.1659 12.5632 11.5473 90.000 101.8688 90.000
space P 21/c
content (C15 H14 O4 S)4
%continue
```

*Input file when cell and space group
are already known*

NanED Workshop II - University Mainz, December 6.-8

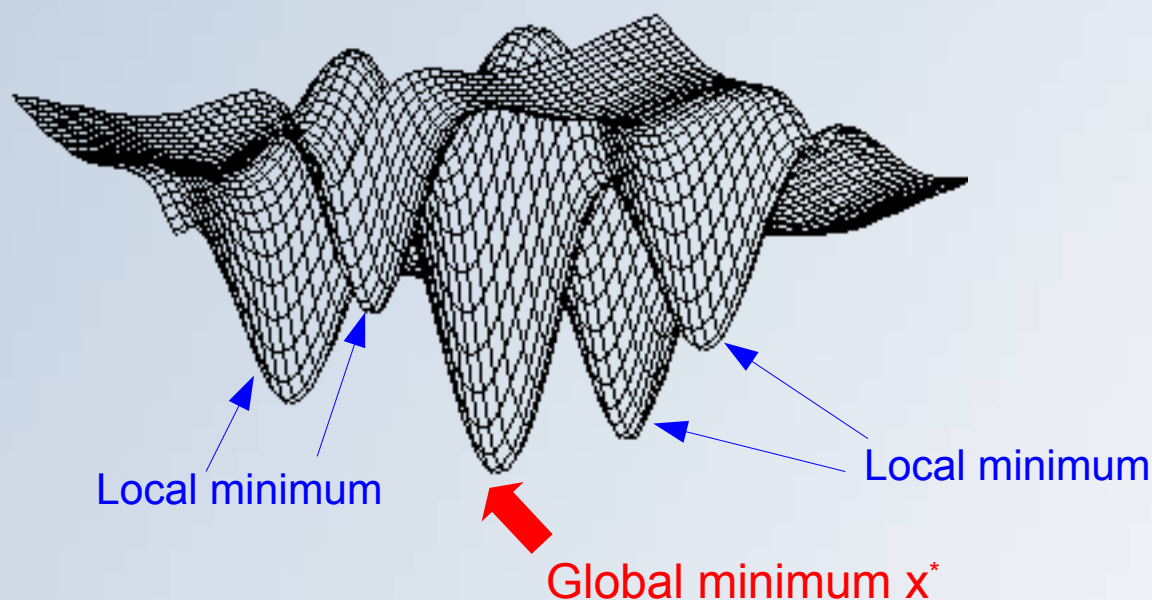


Global Optimization Methods

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

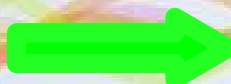


Local optimization methods



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)**

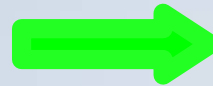
Bee Algorithms

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)**

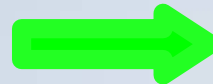
Bee Algorithms

Firefly Algorithms

Harmony Search

*Big Bang-Big Crunch**

.....



Widely used and with
the largest impact

Various modifications:

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

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

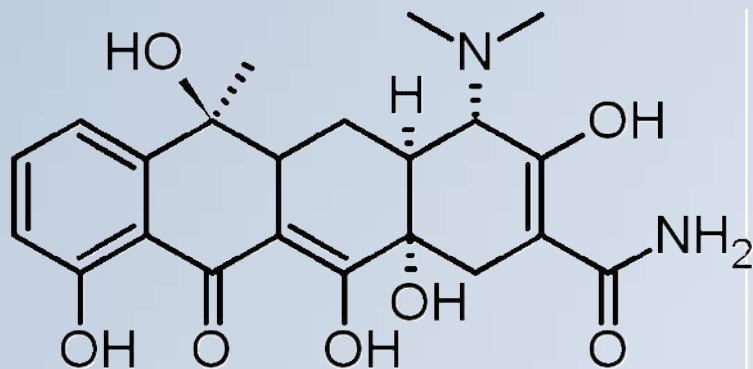
commercial software are in red

() employed in solving crystal structure*

Comparison between the two methods

Traditional approaches

- Chemical knowledge is not necessary
- 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).

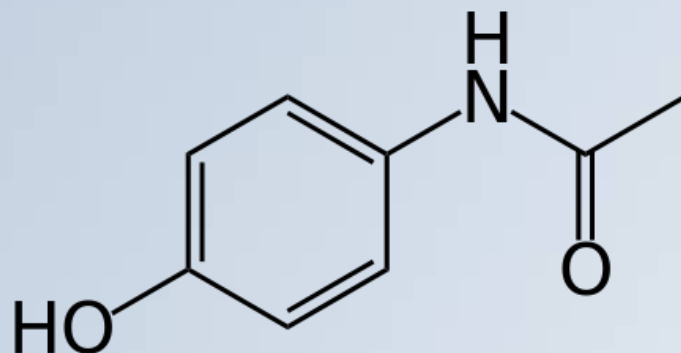
tetracycline (32 non-H atoms and 8 DoF) can be solved using global optimization

- +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

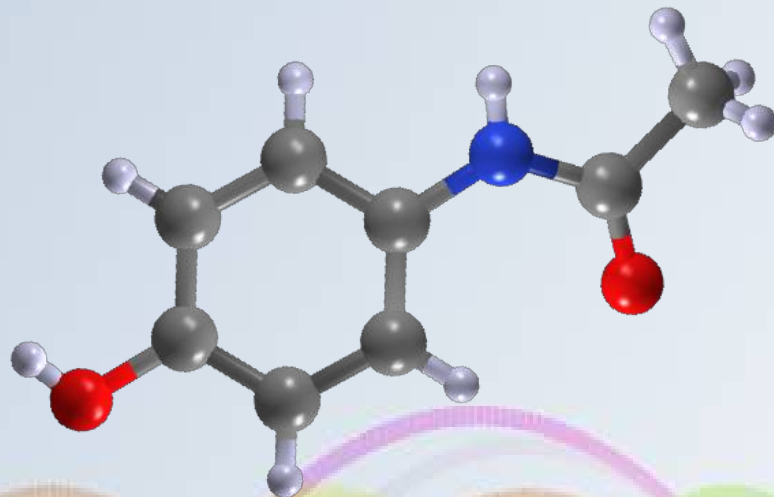
- It is necessary to know the molecular connectivity. Spectroscopic techniques (MS, NMR) can be useful



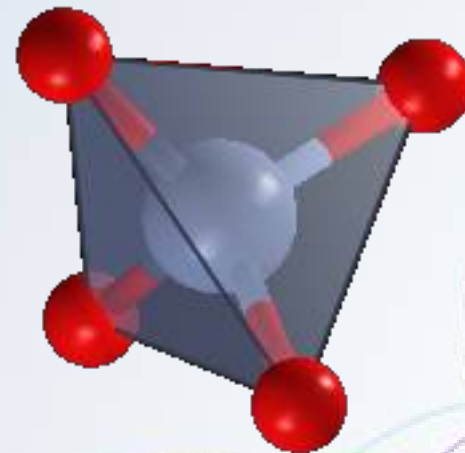
- Crystal structure can be described as a combination of building blocks



atom



molecule



polyhedron

Building starting model

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

Crystal Structure Databases*

Non-commercial databases 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)

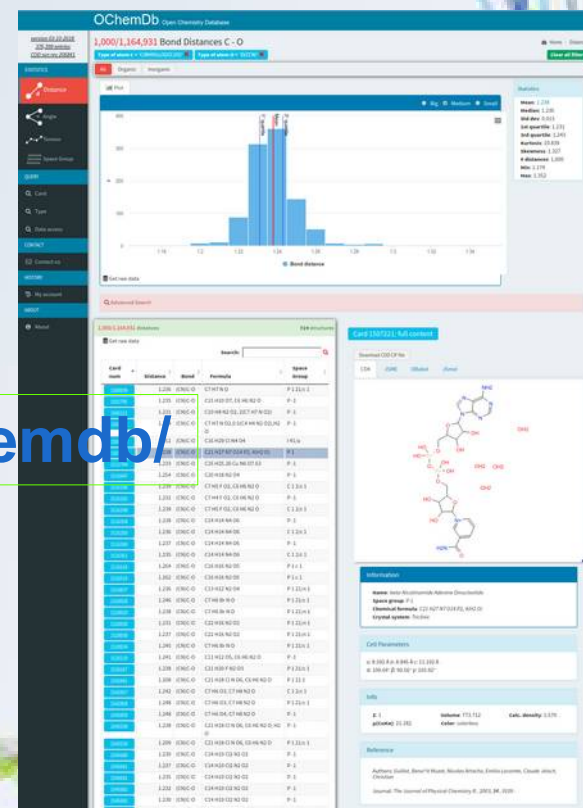
Free Chemistry Databases

- 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/>



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
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.

/home/corrado/expo/paracetamol.xy

File Pattern View Solve Refine Info Help

Observed
Calculated
Background
Difference

paracetamol

File Modify Select View Tools Info

H5 #10 Dist: C5: 0.934

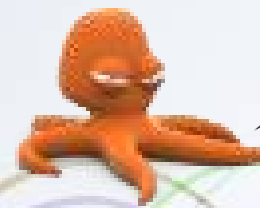
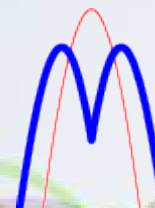
Geometry optimization

Three levels of theory

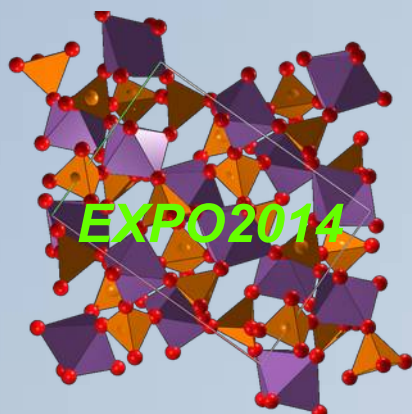
- Molecular-mechanics force fields (**MM**)
- Semi-empirical methods (**SE**)
- *Ab initio* methods: Hartree–Fock methods, density functional theory (**DFT**)



Strategy: **MM** → **SE** → **DFT**



The Use of Open Babel



Fotran/C++

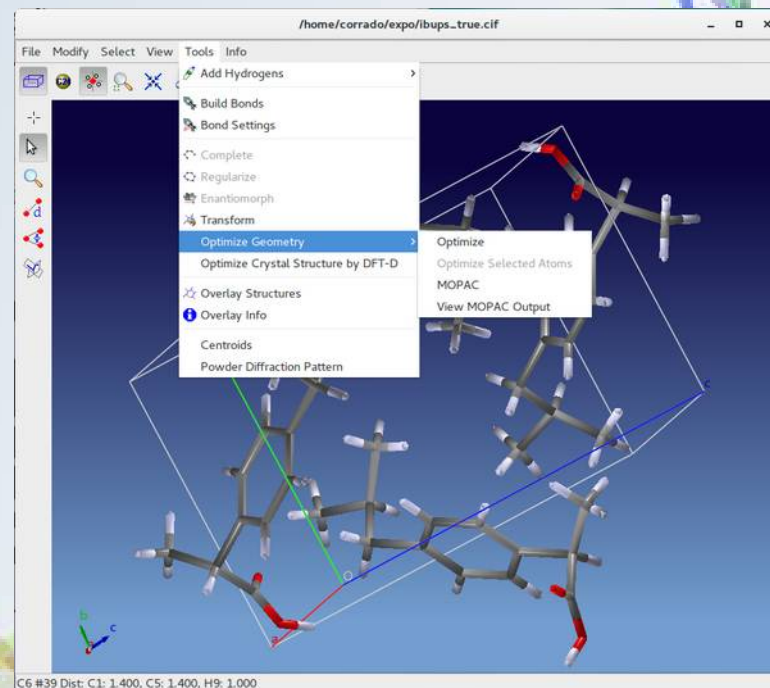


obmodule.f90



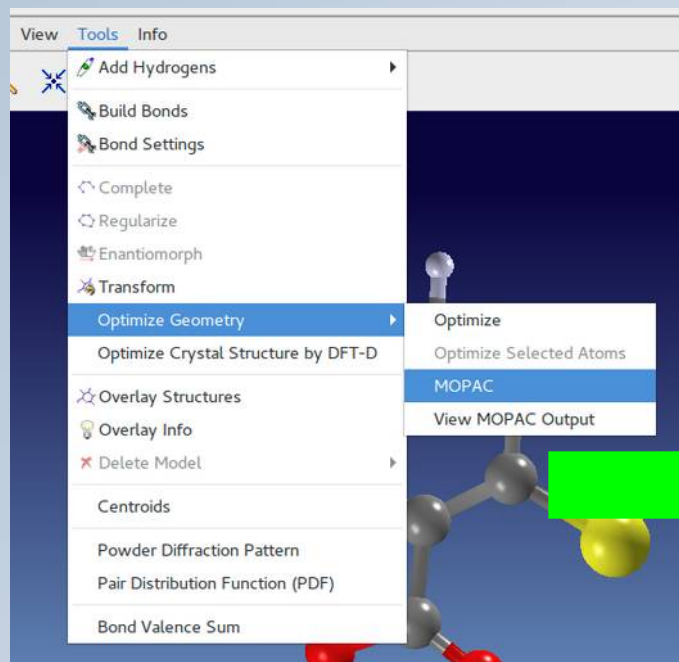
C++

- 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>



MOPAC Input

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

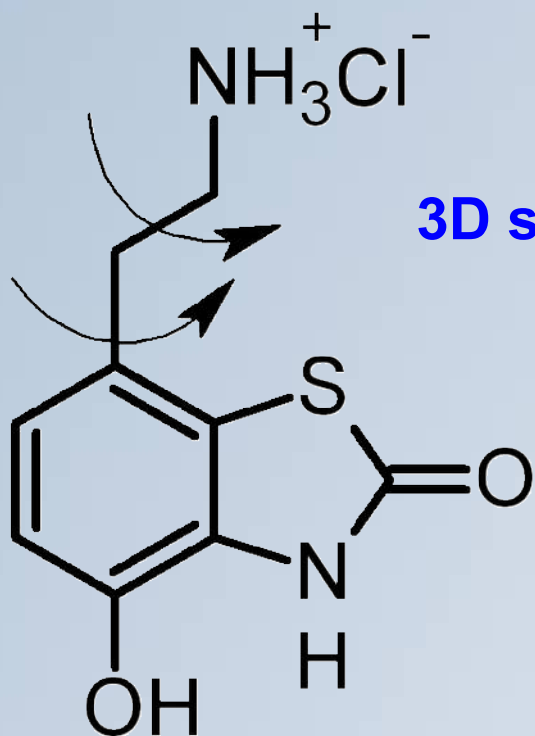
A molecule editor allows

- 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

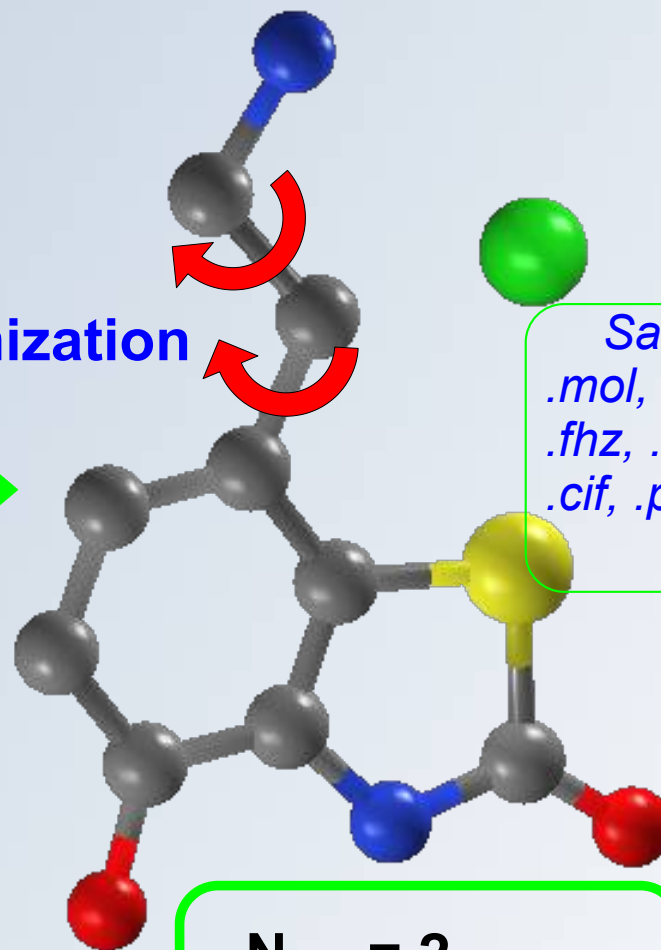
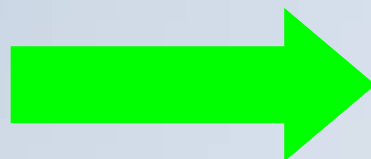
Some free available software

- 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



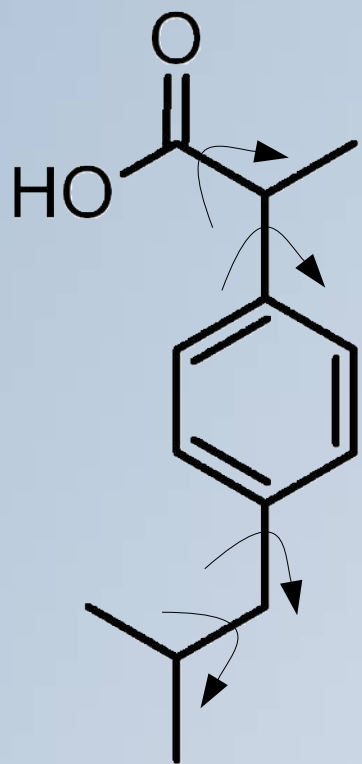
Save as
.mol, .mol2,
.fhz, .mop,
.cif, .pdb, .frac

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

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

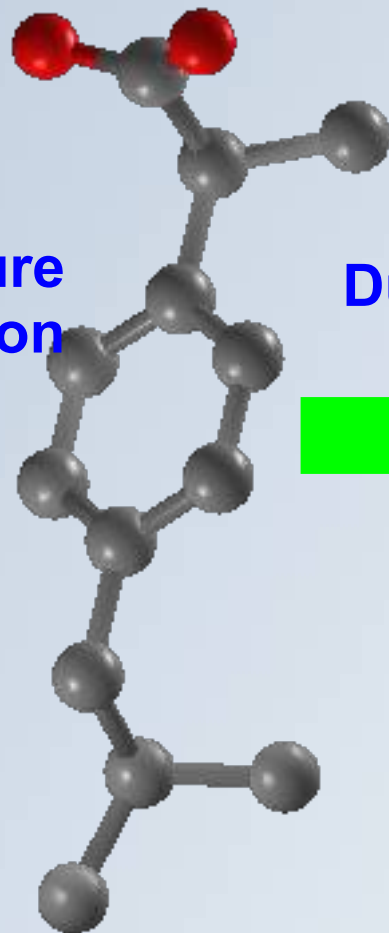
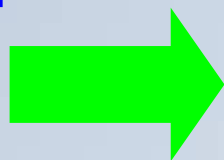
P 21/a

Building starting model: example 2

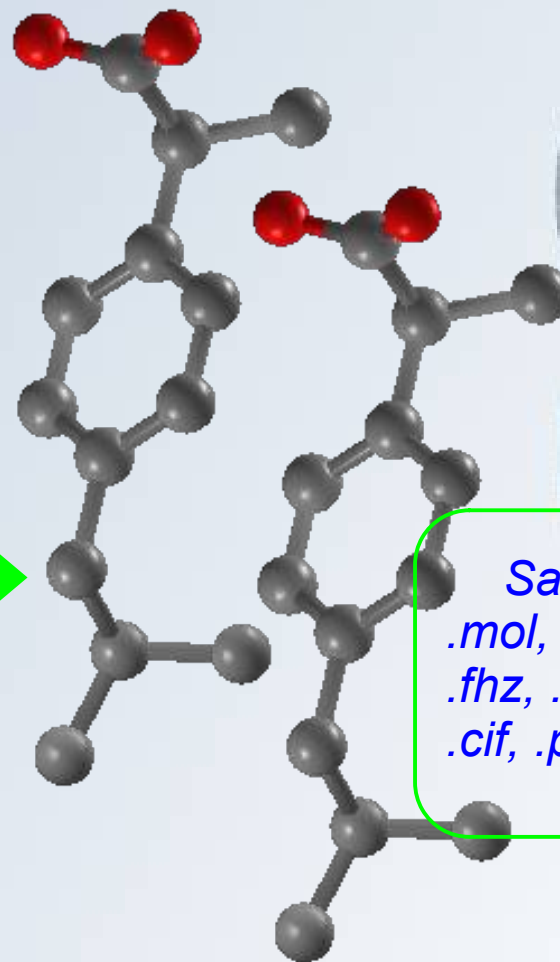


S-Ibuprofen
 $P2_1$

3D structure
optimization



Duplicate



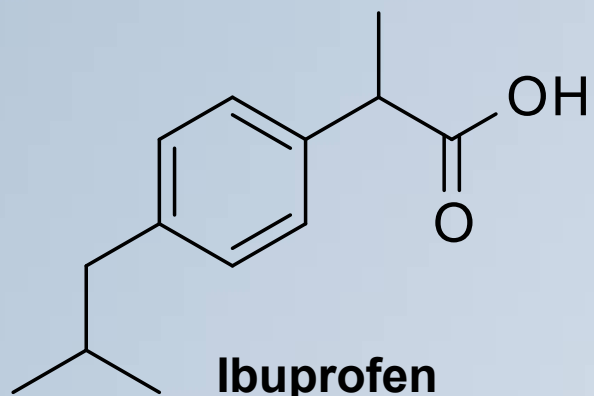
Save as
.mol, .mol2,
.fhz, .mop,
.cif, .pdb, .frac

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

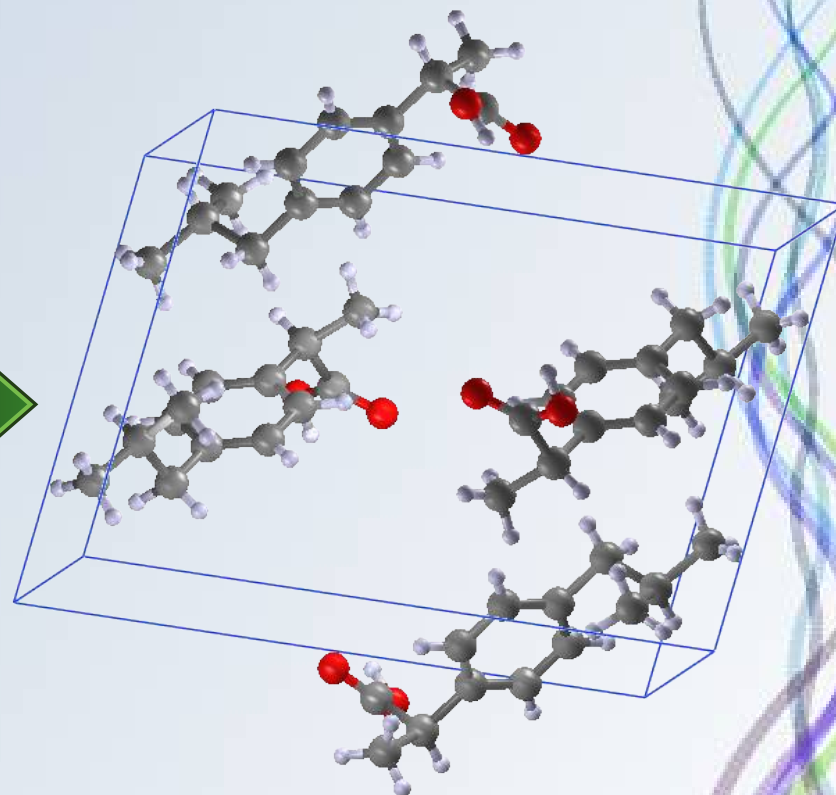
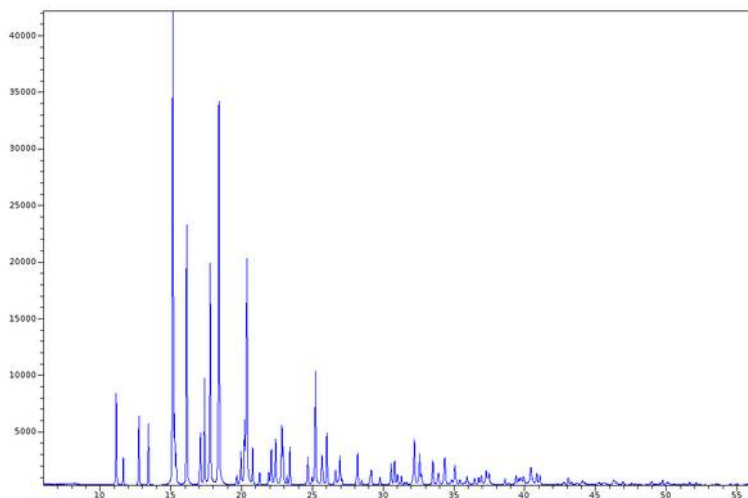
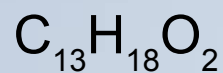
$$N_{\text{dof}} = 6 + 6 + 4 + 4$$

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

Crystal structure solution of ibuprofen by direct space method



Ibuprofen

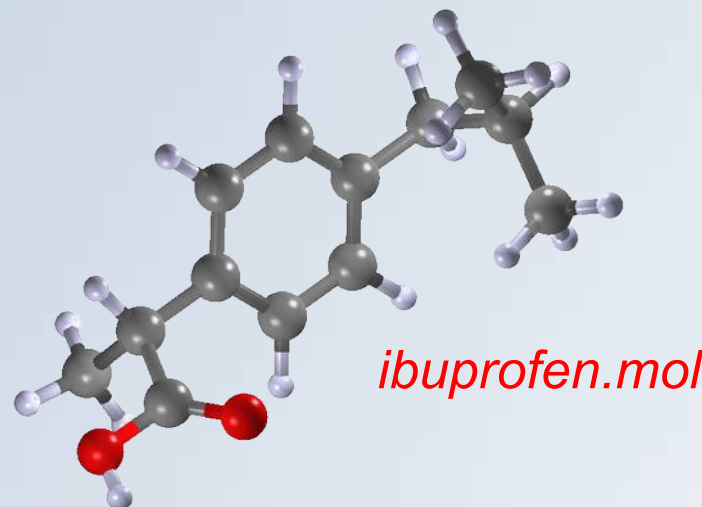


Using expo2014 for direct-space structure solution

- By input file (*.exp)

File → Load and Go

```
%Structure ibuprofen
%Job ibuprofen - C13H18O2
%Data
  Cell 14.6682 7.8933 10.7323 90 99.4092 90
  SpaceGroup p 21/c
  Pattern ibuprofen.dat
  Wavelength 1.54056
%fragment ibuprofen.mol
%sannel
```



- Command-line usage

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

Usage of the command %fragment

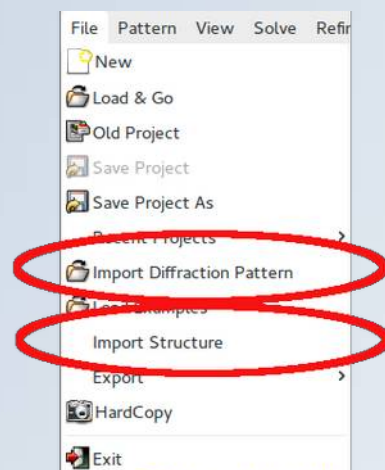
```
%fragment tetra AtC AtV [dist]
%fragment octa AtC AtV [dist]
%fragment square AtC AtV [dist]
%fragment cube AtC AtV [dist]
%fragment trigonal AtC AtV [dist]
%fragment prism_tetra AtC AtV [dist]
%fragment prism_trig AtC AtV [dist]
%fragment icosah AtC AtV [dist]
%fragment atoms chem_formula
%fragment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

Using expo2014 for direct-space solution

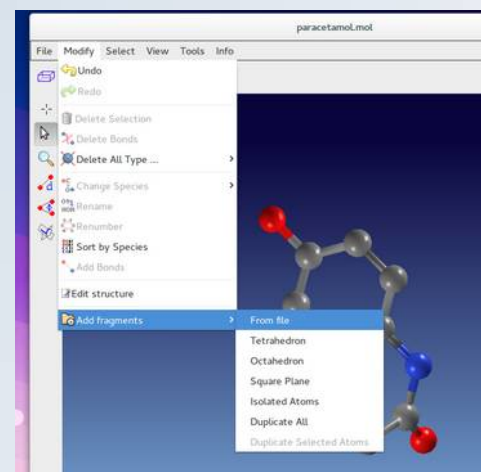
- By graphical interface

- File → Import Diffraction Pattern

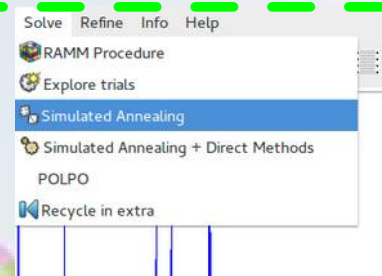
- File → Import Structure



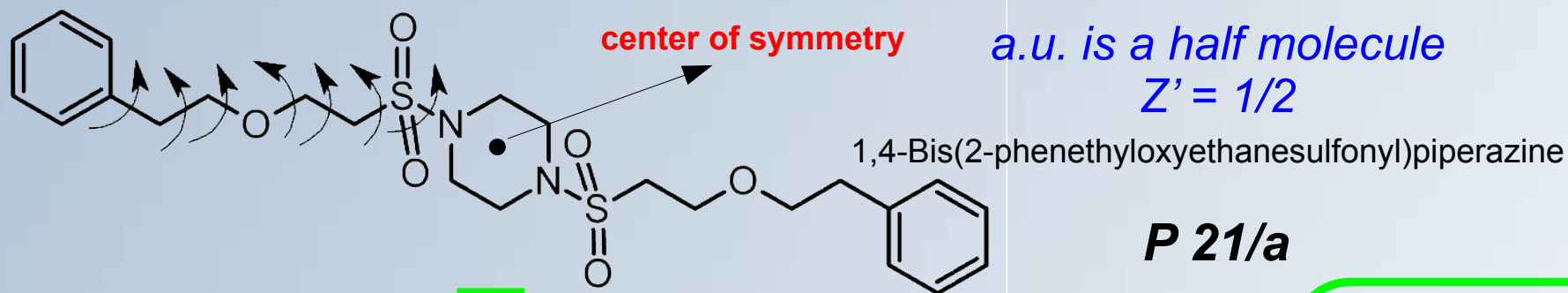
- Modify → Add Fragments



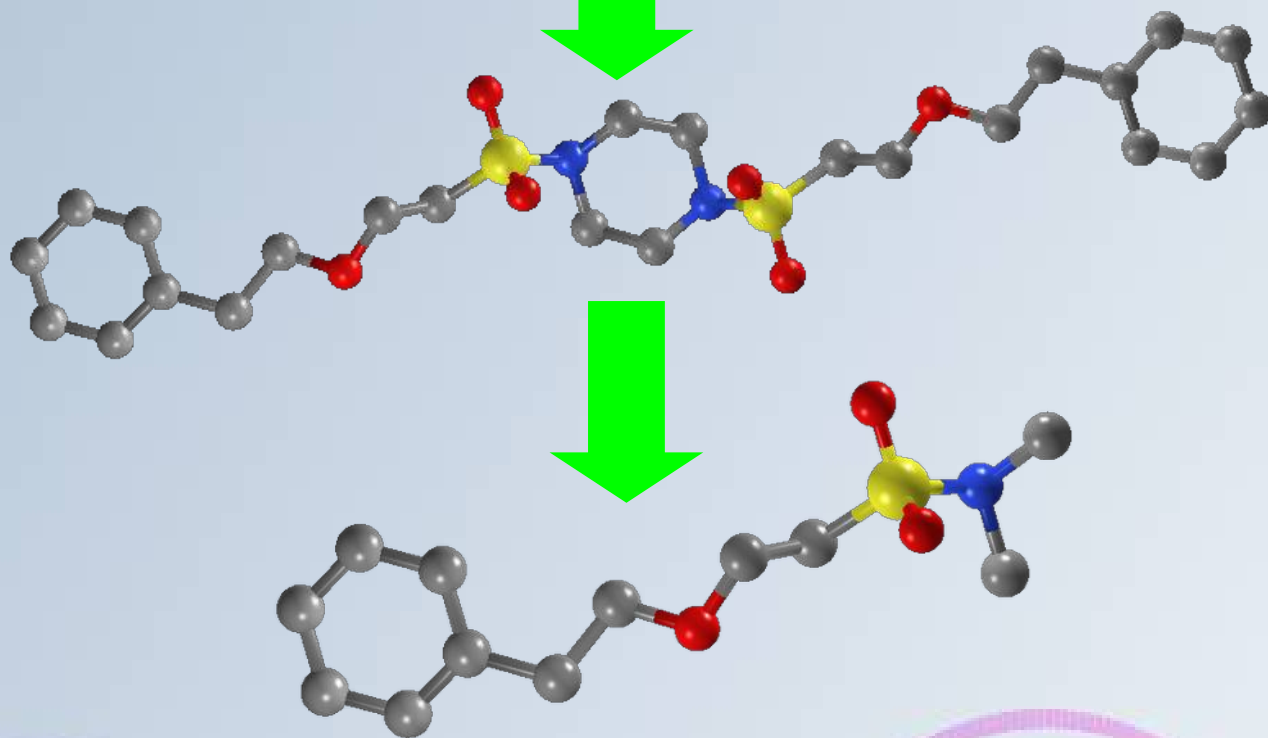
- Solve → Simulated Annealing



Building starting model: example 3



3D structure optimization



Structure solution
by DSM + DOC

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

15 times slower

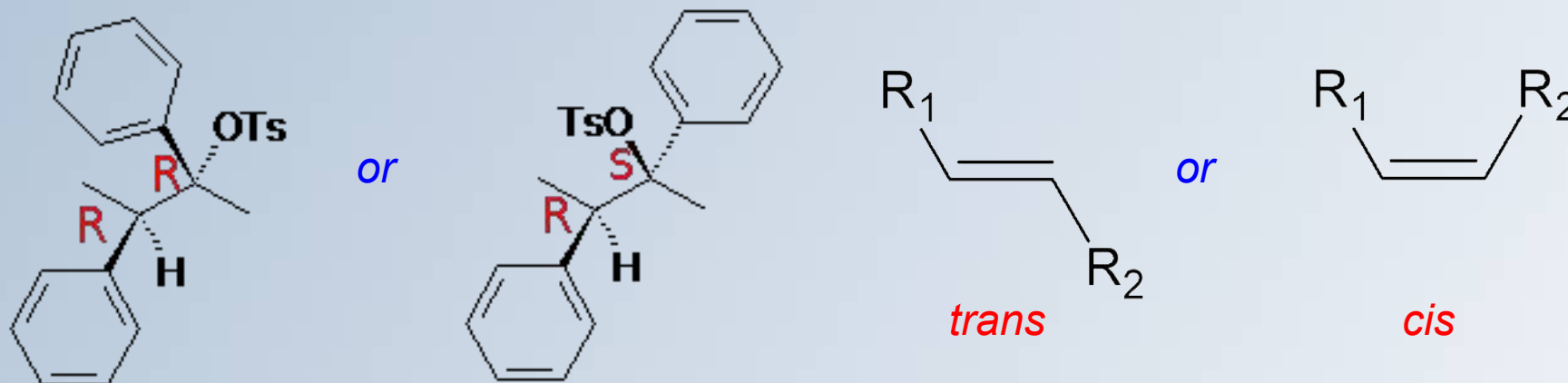
Structure solution
by DSM

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

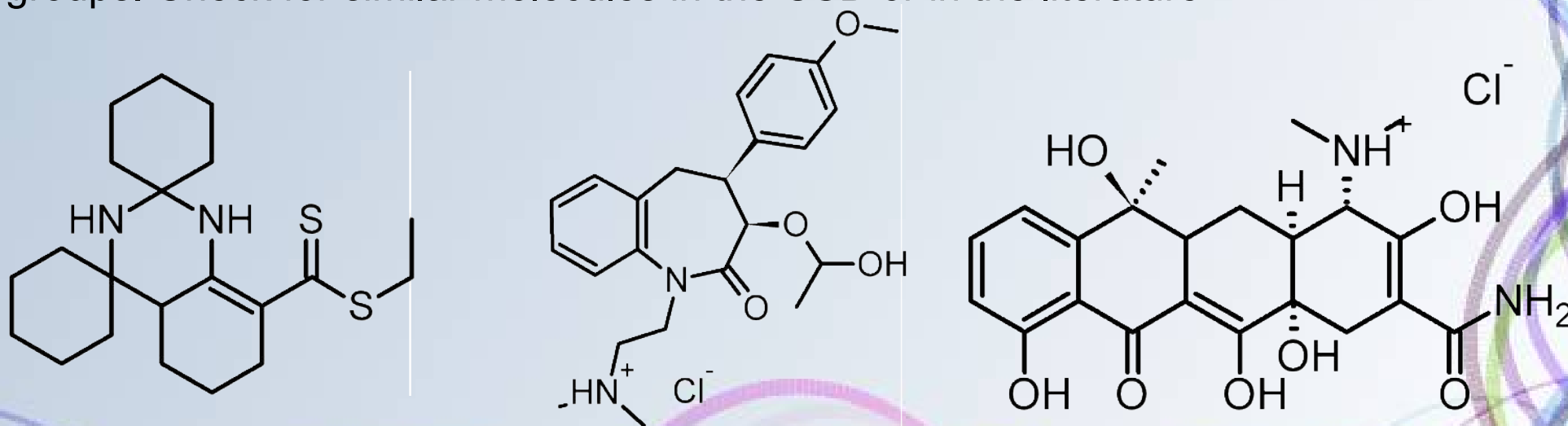
**Recommended
strategy**

Building starting model

- Stereochemistries will not be altered during simulated annealing. Pay particular attention to compounds with more than one chiral center and cis/trans isomers.



- Pay attention to non planar ring systems or unusual combinations of elements in functional groups. Check for similar molecules in the CSD or in the literature

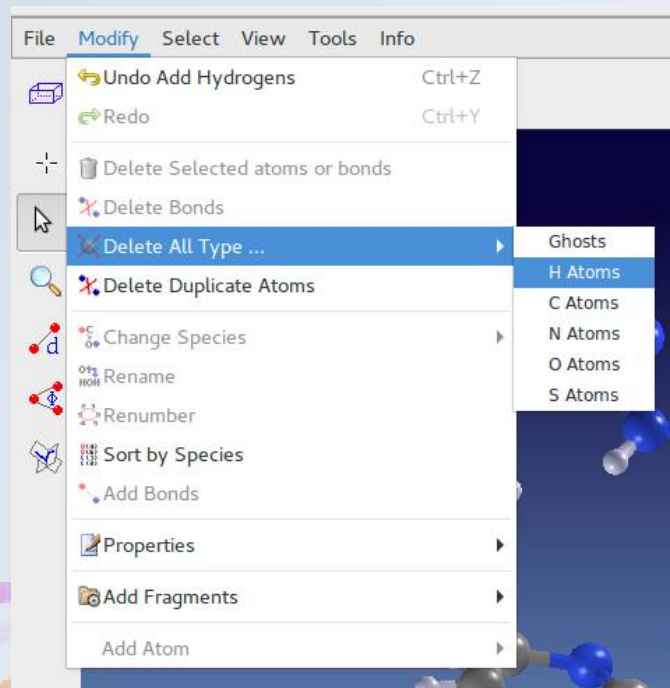


H atoms

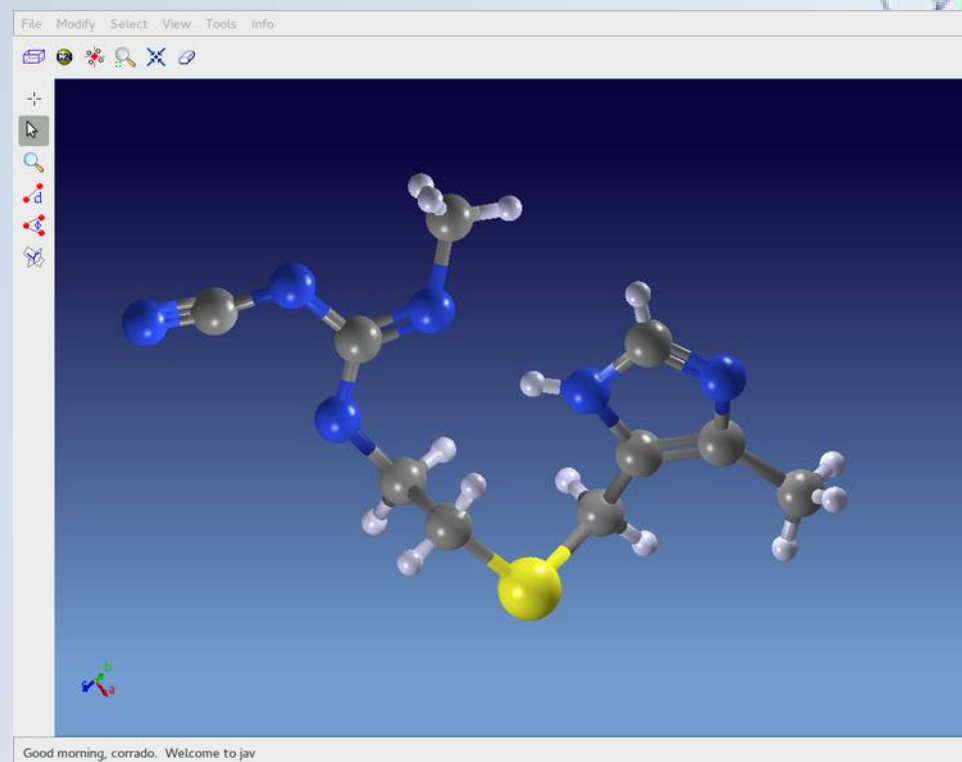
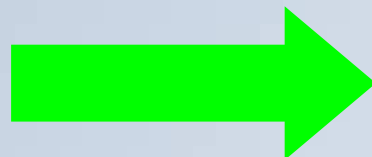
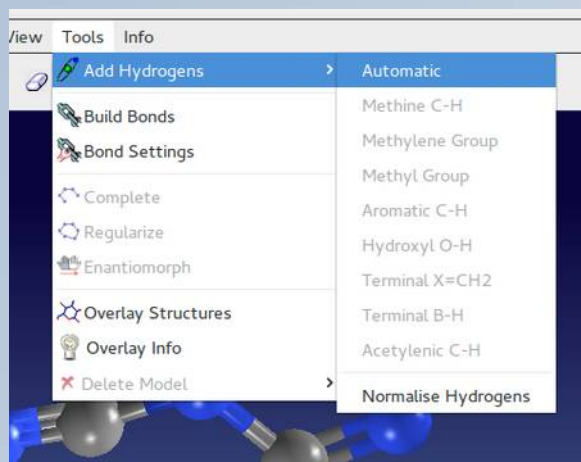
H atoms do not contribute significantly to X-ray diffraction, they can be ignored during the structure solution

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

Delete H atoms using the GUI or using the deletehydro directive

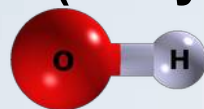


Hydrogen calculation



Hydrogen atoms are positioned geometrically at X-ray distances

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



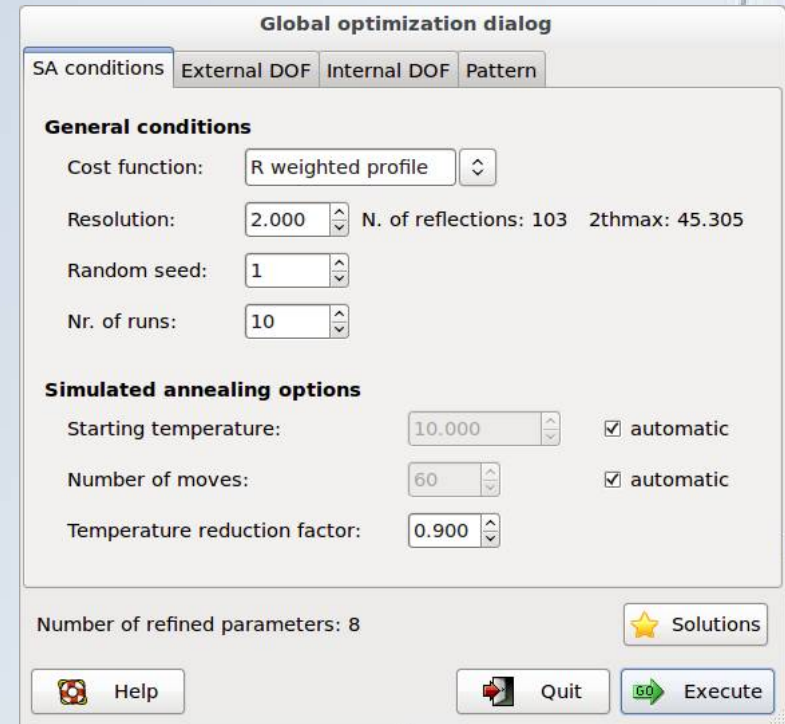
0.82 Å (X-ray)



0.95 Å (neutrons)

Simulated annealing options

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



The image shows a software window titled "Global optimization dialog" with four tabs: "SA conditions", "External DOF", "Internal DOF", and "Pattern". The "SA conditions" tab is active. It contains two sections: "General conditions" and "Simulated annealing options".

General conditions

- Cost function: R weighted profile (dropdown)
- Resolution: 2.000 (spinbox) N. of reflections: 103 2thmax: 45.305
- Random seed: 1 (spinbox)
- Nr. of runs: 10 (spinbox)

Simulated annealing options

- Starting temperature: 10.000 (spinbox) ☒ automatic
- Number of moves: 60 (spinbox) ☒ automatic
- Temperature reduction factor: 0.900 (spinbox)

Number of refined parameters: 8

Buttons: Help, Quit, Execute, Solutions (star icon)

Simulated annealing options

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



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

Temperature reduction factor: 0.900

Number of refined parameters: 8

Solutions

Help Quit Execute

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

Cost Functions

- **Whole profile R factor**

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

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

- **Integrated intensities R factor**

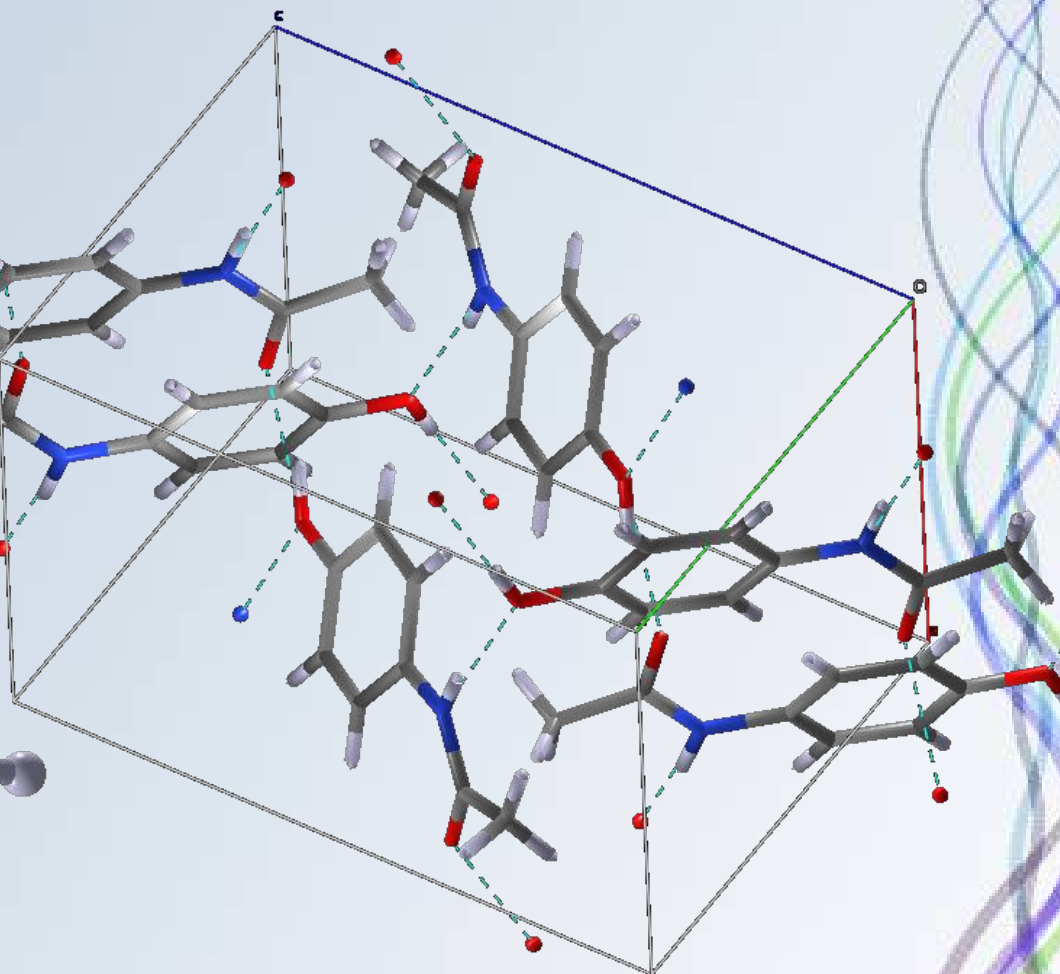
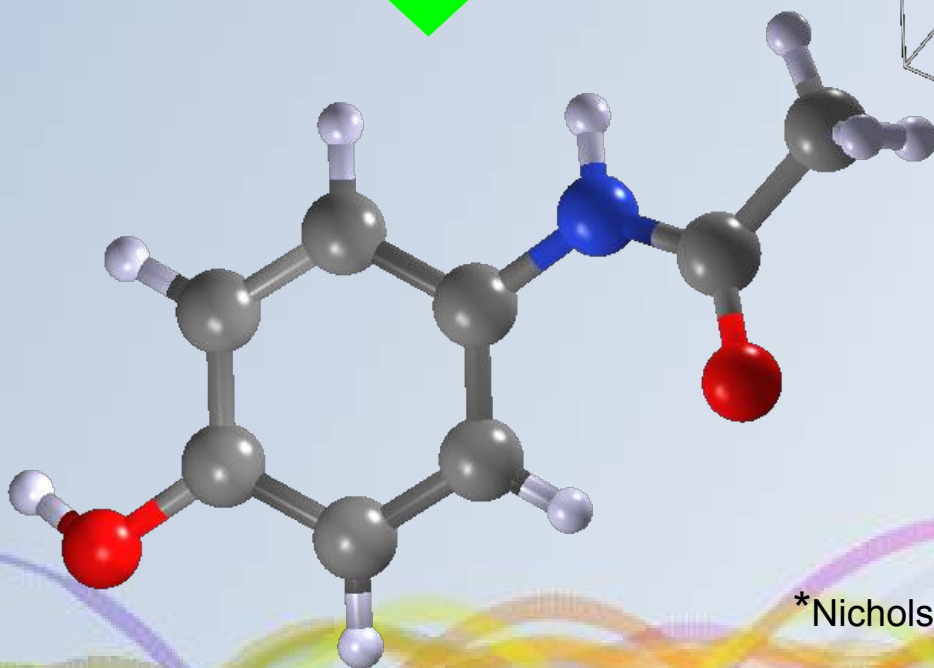
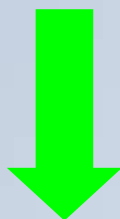
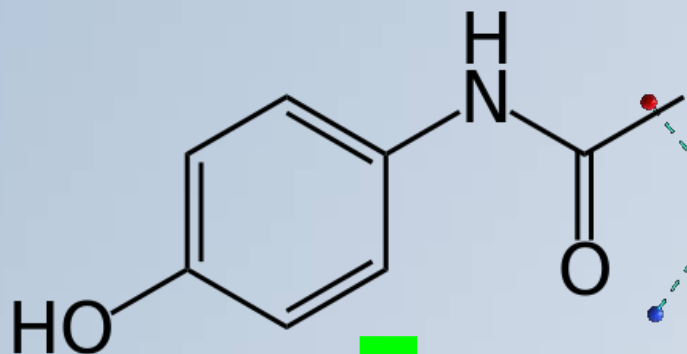
$$R_B = \frac{\sum_h |I_h^{exp} - I_h^{calc}|}{\sum_h I_h^{exp}}$$

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

Molecular compounds

Paracetamol (form I polymorph)

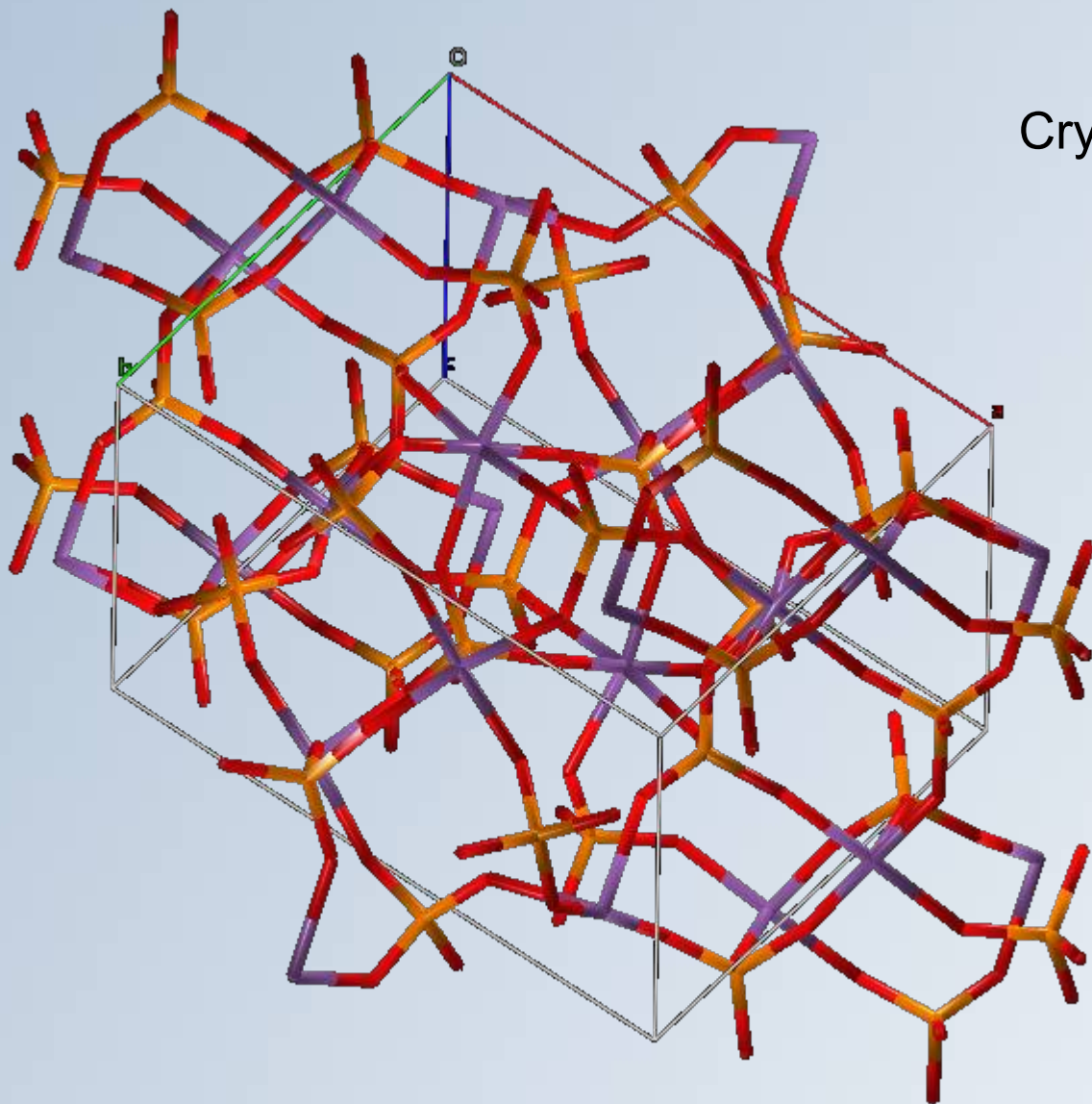
(C₈H₉NO₂) *



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

Non-molecular compounds

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

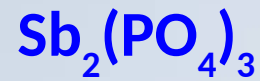


*Jouanneaux, A., Verbaere, A., Guyomard, D., Piffard, Y., Oyetola, S. & Fitch, A. N. (1991). *Eur. J. Solid State Inorg. Chem.* **28**, 755-765.

Non-molecular compounds

$$\frac{N_{obs}}{N_{dof}} > 8$$

$$N_{dof} = 51$$

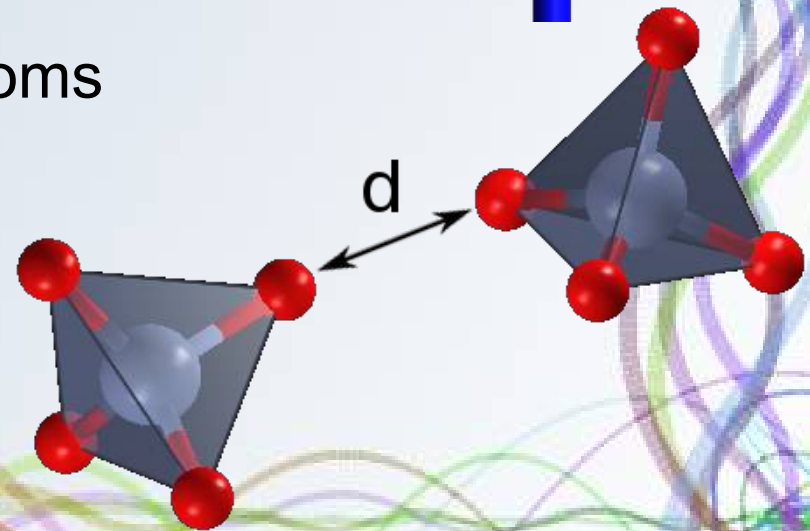
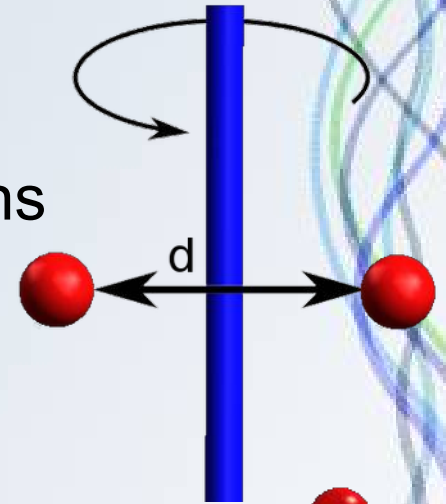


$$N_{dof} = 30$$

$$N_{dof} = 24$$

Non-molecular compounds

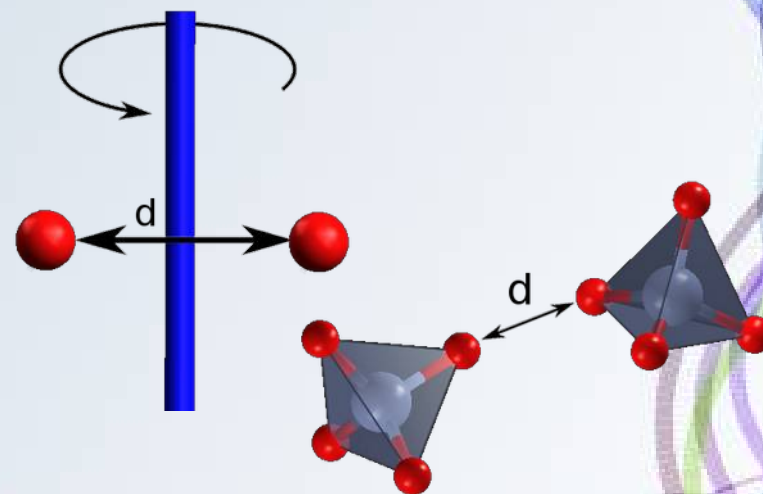
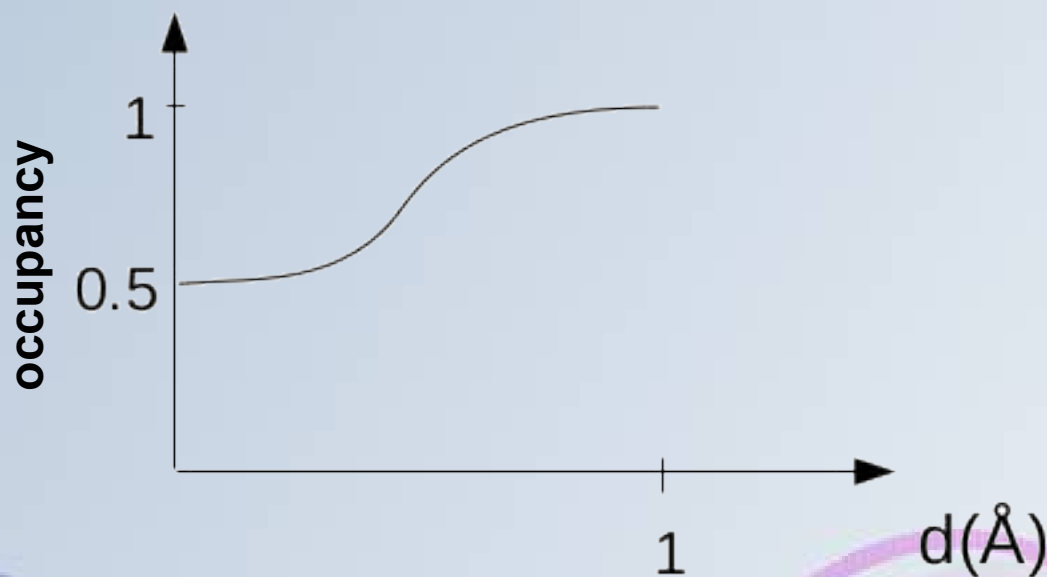
- You cannot know the number and the type of the polyhedra
- Some atoms are expected to fall on special positions
- Different building blocks share some atoms



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

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



DOC is able to merge the excess atoms automatically

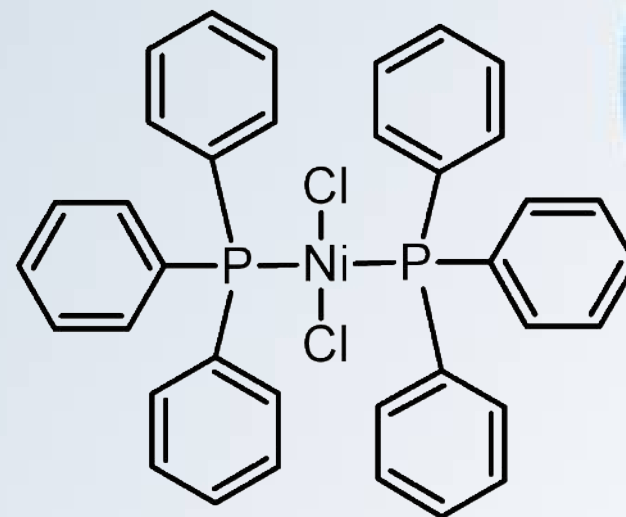
Dynamical occupancy correction (DOC)

`doc`

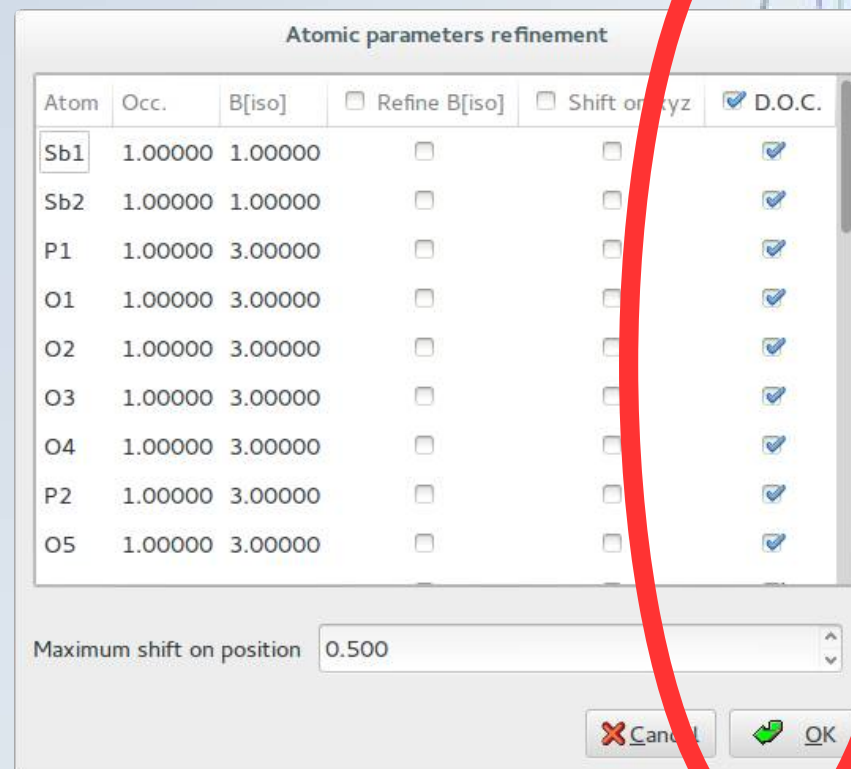
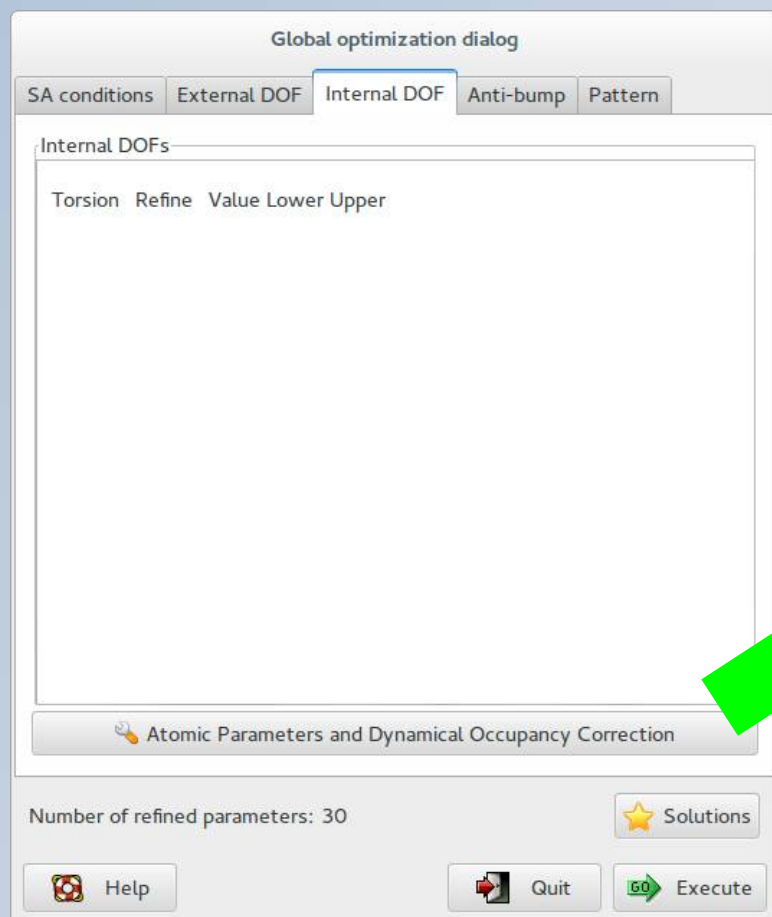
or

`doc atom1 atom2 ...`

`doc Ni1`

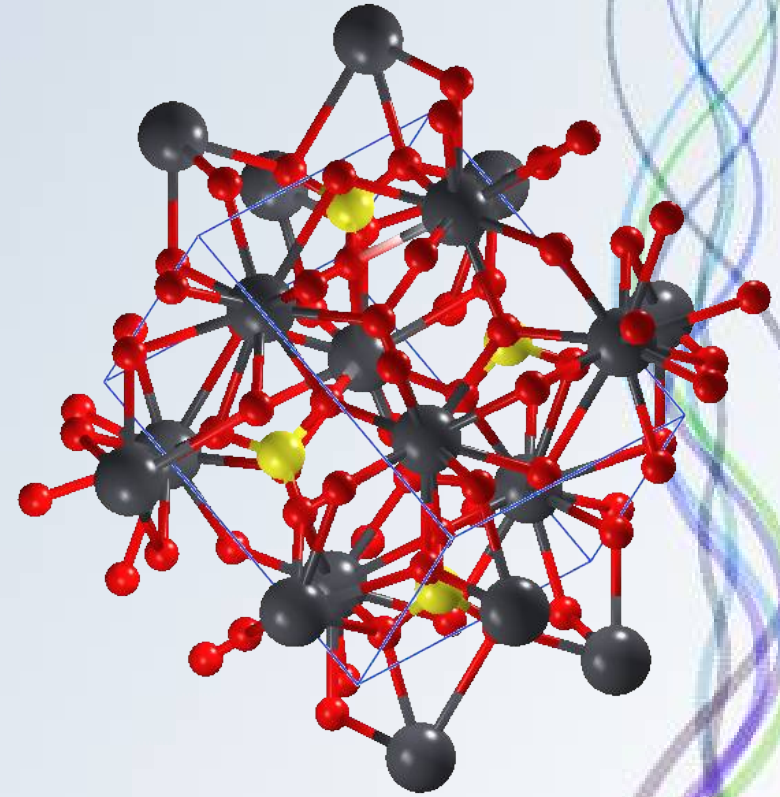
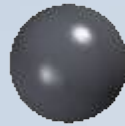
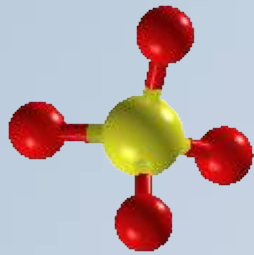


Dynamical occupancy correction (DOC)

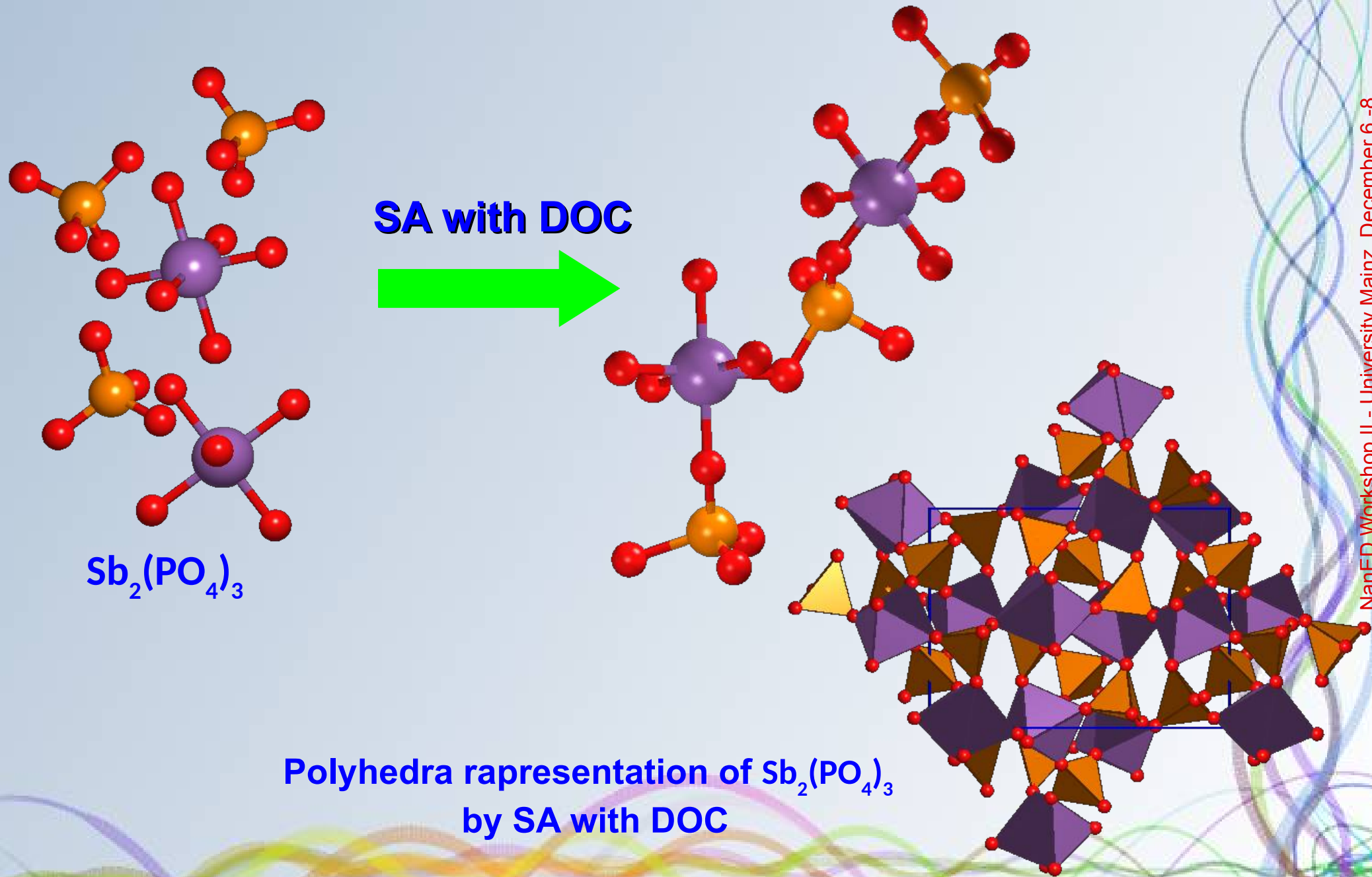


DOC slows down the computation time so it should be avoided if no special positions or shared atoms are expected.

SA applied to non-molecular compounds



SA applied to non-molecular compounds




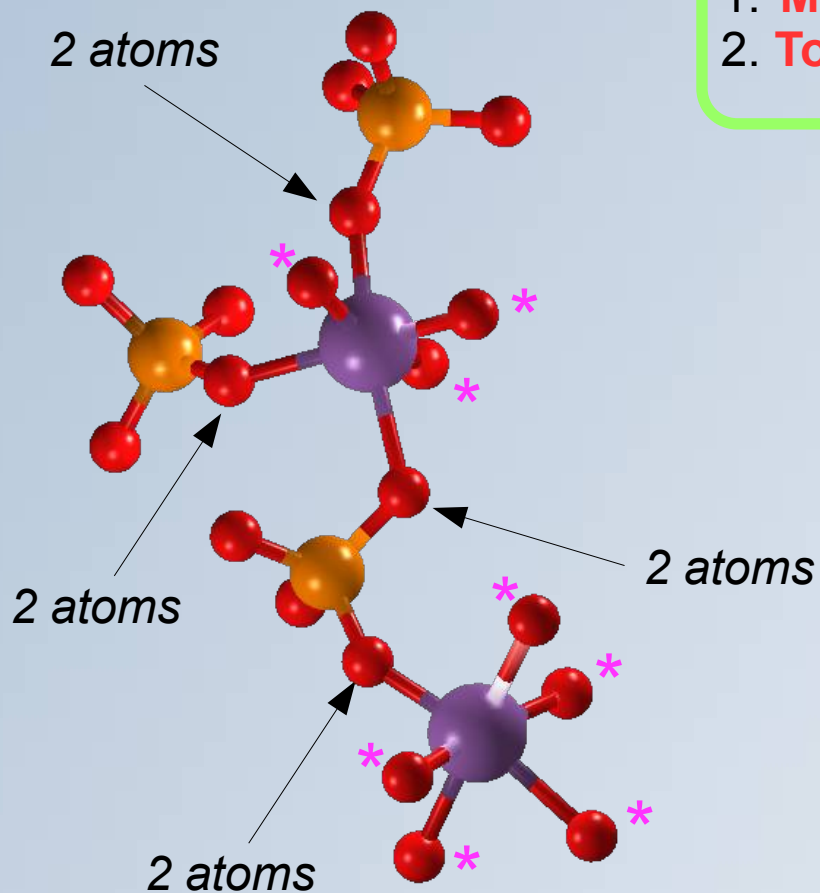
Delete duplicate atoms

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

Set threshold value

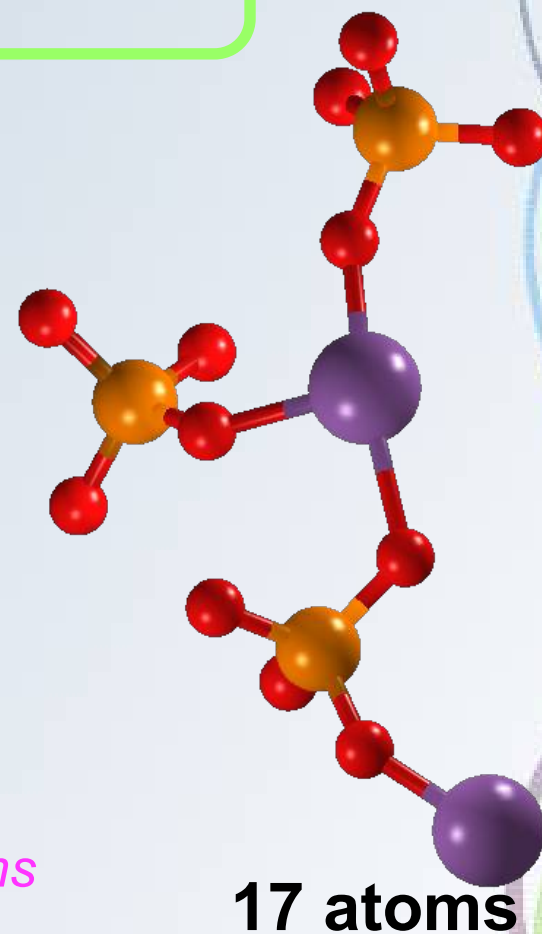
Threshold:

 OK



29 atoms

** symmetry equivalent atoms*



17 atoms

Direct Space with Low Quality Diffraction Pattern

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

Imposing anti-bumping restraints

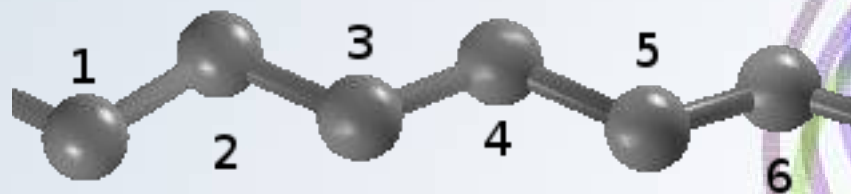
$$CF_{bump} = \sum_{ij}^n w_{ij} (d_{ij}^{min} - d_{ij}^{model})^{2k}$$

$$k = 2$$

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

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

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



Only 1-5 interactions are considered

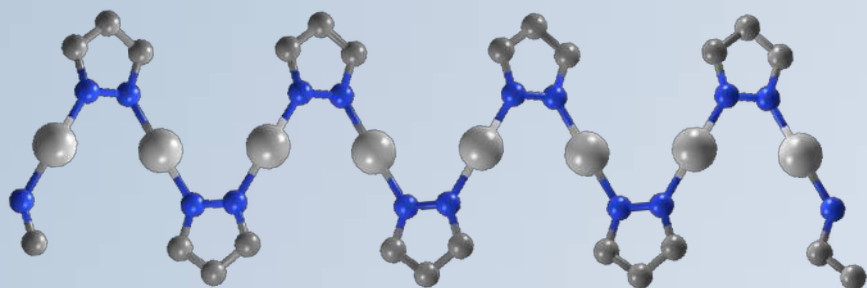
Imposing anti-bumping restraints

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

$$k = 2$$

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

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



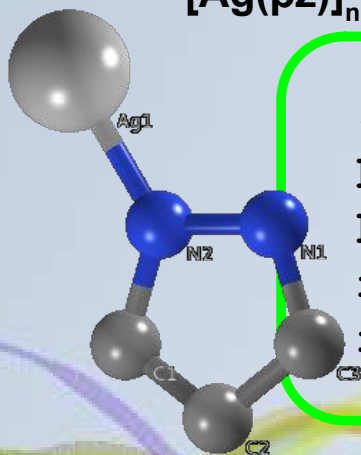
[Ag(pz)]_n Hpz = pyrazole

Directive:

bump

or

bump atoms1 atoms2 [dist]



```
%sannel
bscale 0.9
bump * *
nobump Ag1 N1
nobump Ag1 N2
```

Warning: time-consuming procedure, use only if the diffraction data are not of sufficient quality

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>)

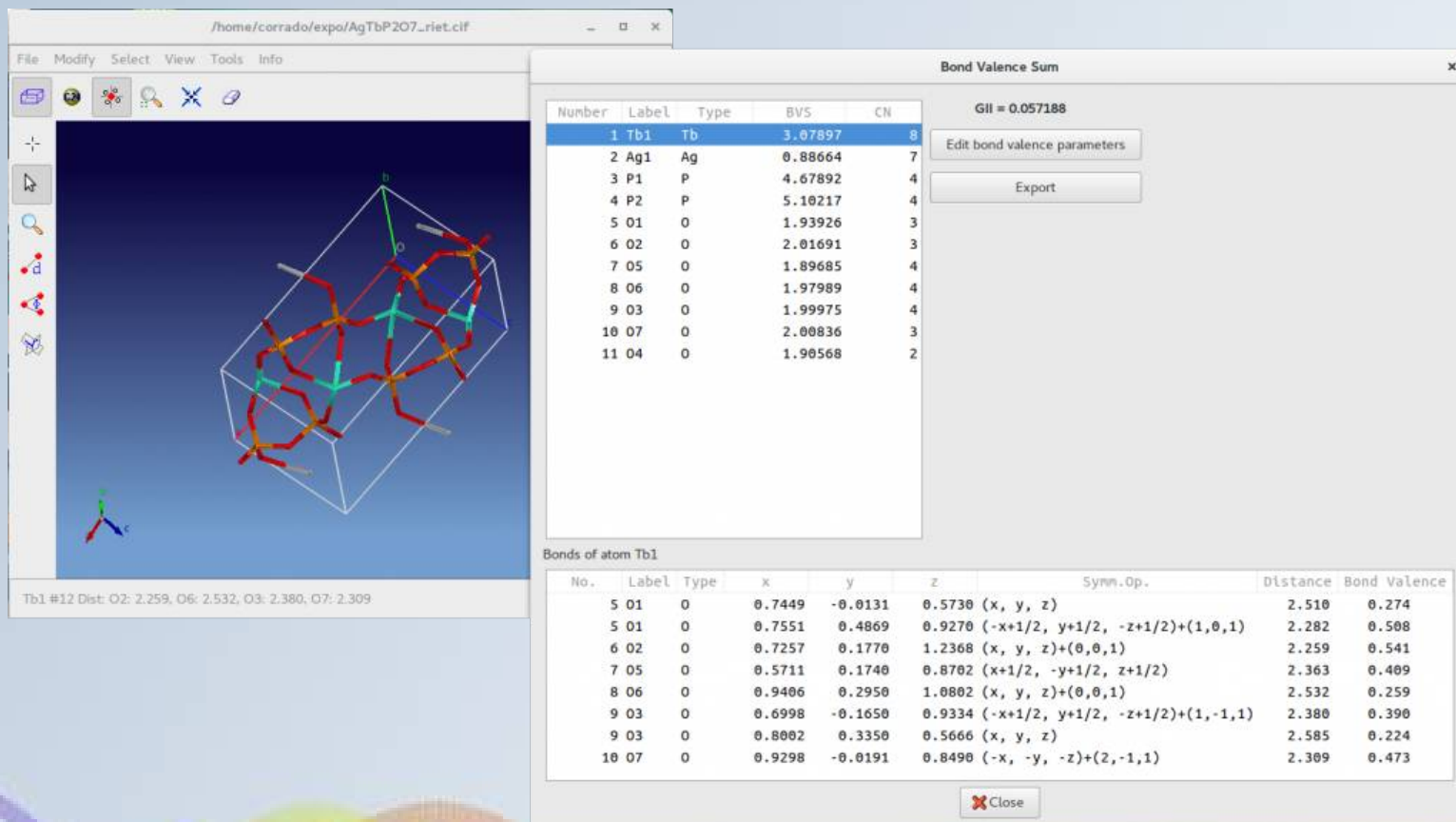
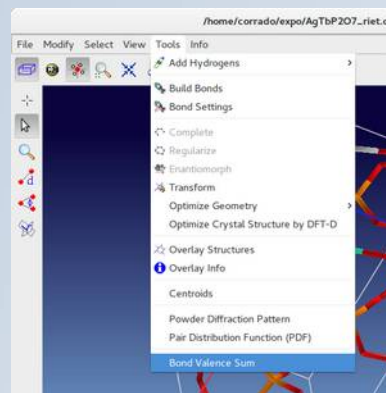
$$G_{ii} = \sqrt{\frac{1}{N} \sum_{i=1}^N (V_i - V_i^0)^2} \quad \text{global-instability index}$$

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



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

Directive:

rest A B d_{target} tol

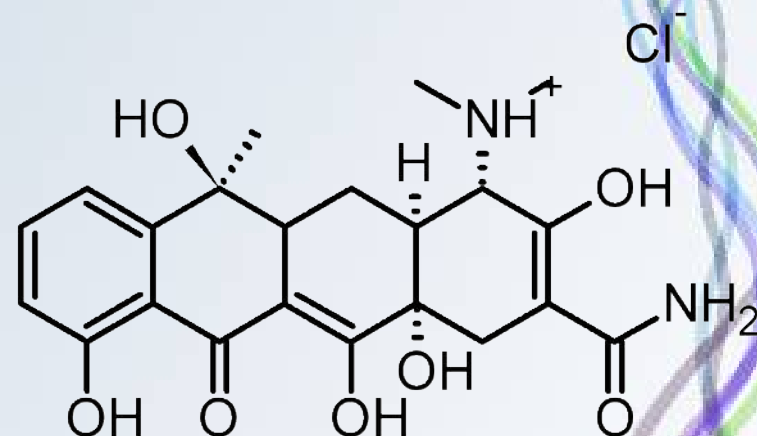
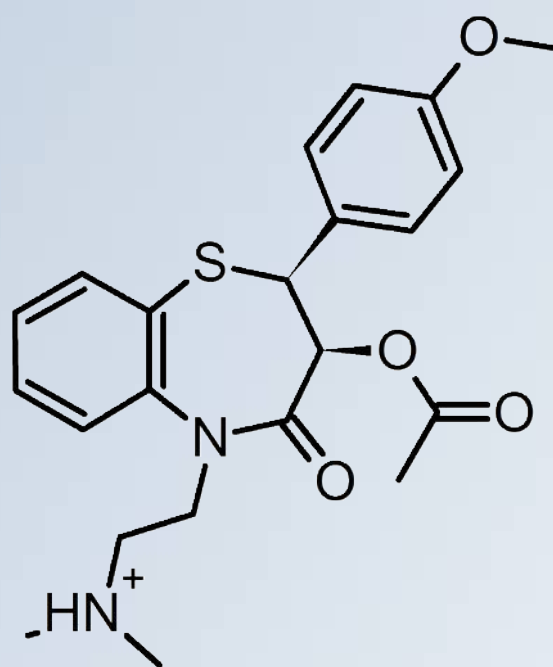
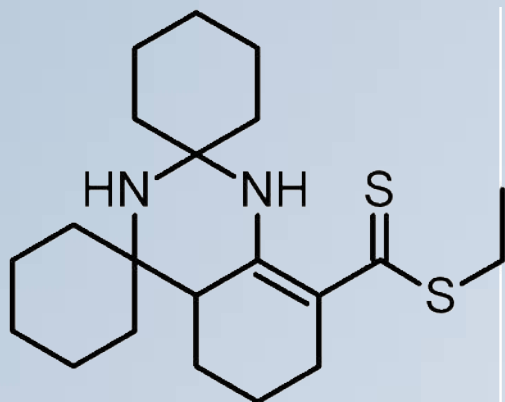
d_{target} and tol are optional:

rest A B

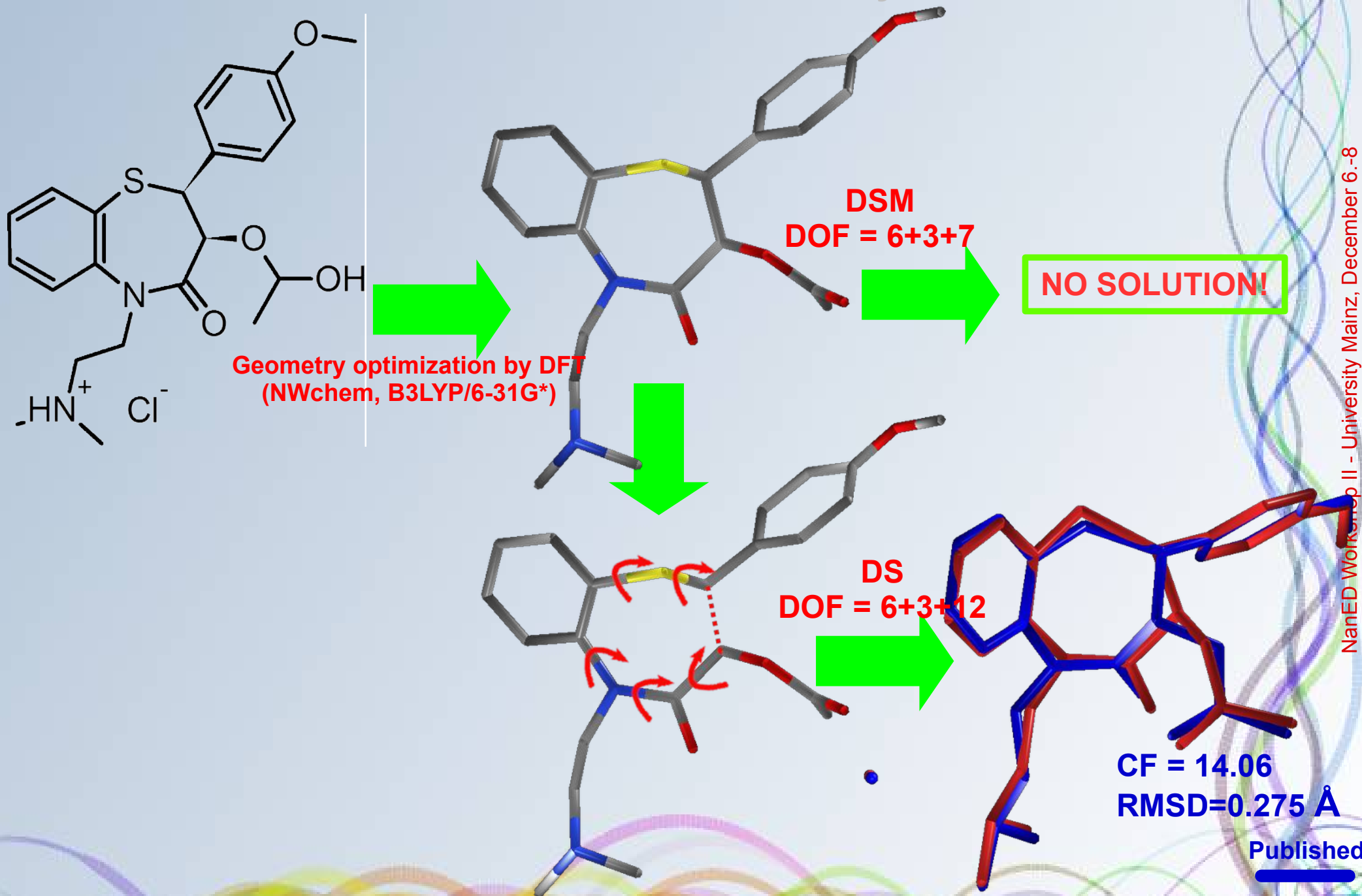
- Rarely improves the success rate of the solution search for good quality data
- Restraints can slow or prevent a structure solution process

Non planar ring systems

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



Structure Solution of Diltiazem Hydrochloride*



*Kojic Prodic, B., Ruzic Toros, Z., Sunjic, V., Decorte, E. & Moimas, F. (1984). *Helv. Chim. Acta*, 67, 916–925.
<https://doi.org/10.1002/hlca.19840670333>

Structure Solution of Diltiazem Hydrochloride

File: diltia.exp

```
%Structure diltia
%Job CSD reftype: CEYHUJ01

%Data
Cell      42.190   9.075   6.037   90   90   90
SpaceGroup p 21 21 21
Pattern    pd_0029.xy
Wavelength 1.54056

%fragment diltiazem_nw_break.mol
%fragment atoms Cl
deletehydro

%anneal
nrun 100
niter 5000
```

Structure Solution of Diltiazem Hydrochloride (approach with restraint)

File: diltia_res.exp

```
%Structure diltia
```

```
%Job CSD refcode: CEYHUJ01
```

```
%Data
```

```
Cell      42.190   9.075   6.037   90   90   90
```

```
SpaceGroup p 21 21 21
```

```
Pattern    pd_0029.xy
```

```
Wavelength 1.54056
```

```
%fragment diltiazem_nw_break.mol
```

```
%fragment atoms C1
```

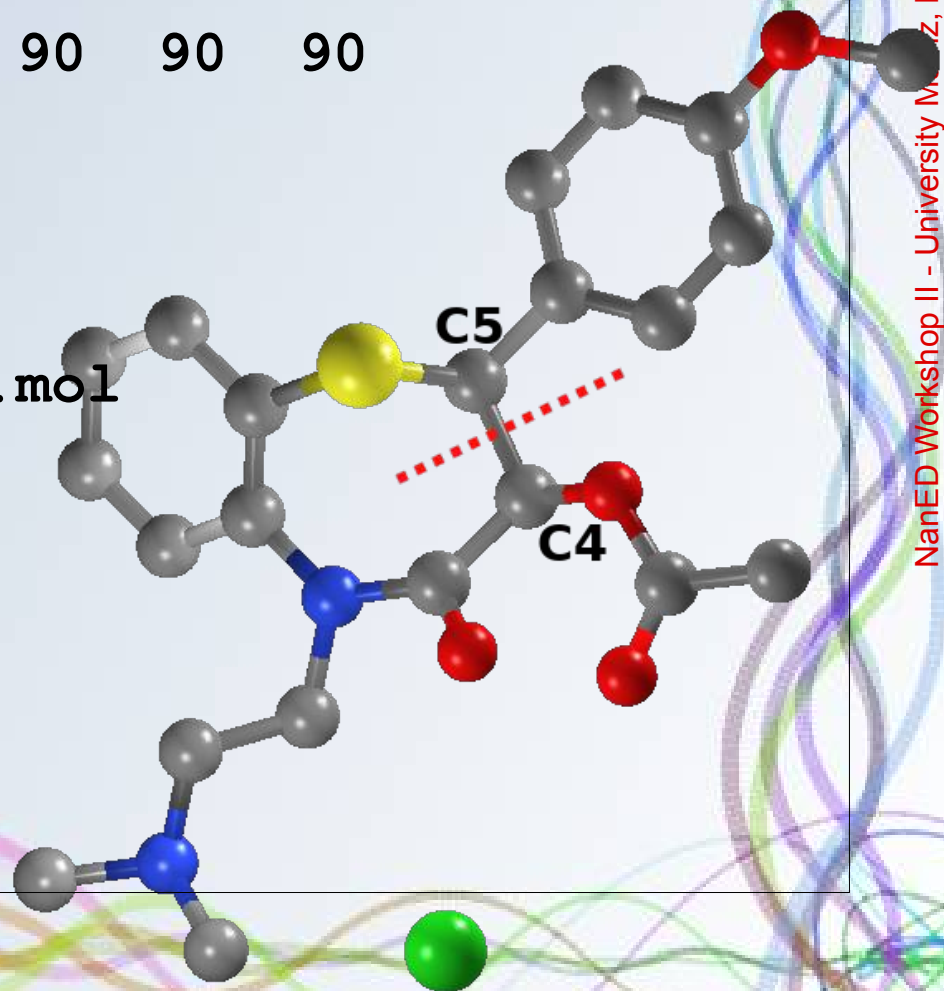
```
deletehydro
```

```
%sannel
```

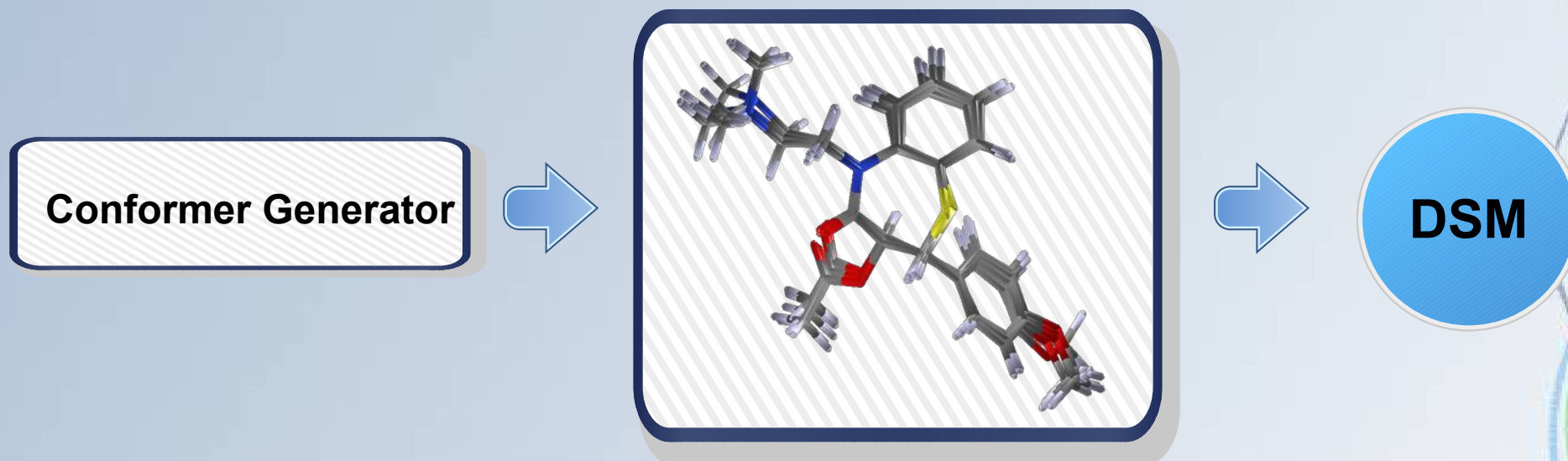
```
nrun 100
```

```
niter 1000
```

```
rest C4 C5
```



Combining conformational analysis with DSM

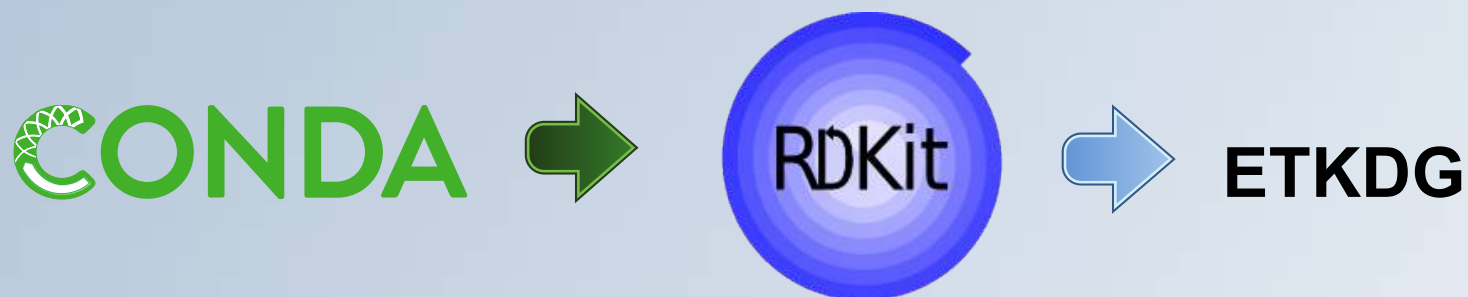


- BALLOON
- CONFAB
- FROG2
- RDKit



- CORINA
- MOE
- OMEGA

Conformer Generation with RDKit



- RDKit environment must be activated

```
cd ~/anaconda3/bin/  
source activate my-rdkit-env
```

- Simple script for generating conformers from single molfile

https://github.com/iwatobipen/rdk_confgen

Change line `param = rdDistGeom.ETKDGv2()` (line 16)

in `param = rdDistGeom.ETKDGv3()`

- Then run confgen command

```
confgen --input diltia_nw.mol --numconf 100
```

- Split SDF file into single MOL files

```
obabel gen_confs.sdf -O diltia1.mol -m
```

Running a structure solution process on each estimated conformer

```
#!/bin/bash
```

```
filein=molecules.txt  
template=diltia_temp1
```

```
rm results.dat summary.dat *.expo *.bin *_best*.cif  
c=`wc -l $filein | awk '{print $1}'`  
  
for i in `seq 1 $c`;  
do  
a=`cat $filein | head -$i | tail -1`  
file=${a%.*}  
echo $file $template  
sed "s%STRNAME%$file%" $template > "$file.exp"  
#expo "$file.exp" "$file.out" --nogui  
mpirun -np 20 $HOME/expo/expo_mpi "$file.exp" "$file.out"  
echo "======" >> results.dat  
echo "$file.out" >> results.dat  
grep @@ "$file.out" | cut -c3- >> results.dat  
grep --with-filename @@best "$file.out" | sed 's/:@@best//g'  
>> summary.dat  
done
```

Running a structure solution process on each estimated conformer

```
#!/bin/bash
```

```
filein=molecules.txt  
template=diltia_temp1
```

```
rm results.dat summary.dat *.expo *.bir  
c=`wc -l $filein | awk '{print $1}'`  
  
for i in `seq 1 $c`;  
do  
a=`cat $filein | head -$i | tail -1`  
file=${a%.*}  
echo $file $template  
sed "s%STRNAME%$file%" $template > "$file"  
#expo "$file.exp" "$file.out" --nogui  
mpirun -np 20 $HOME/expo/expo_mpi "$file.exp" "$file.out"  
echo "===== " >> results.dat  
echo "$file.out" >> results.dat  
grep @@ "$file.out" | cut -c3- >> results.dat  
grep --with-filename @@best "$file.out" | sed 's/:@@best//g'  
>> summary.dat  
done
```

```
diltia1.mol  
diltia2.mol  
diltia3.mol  
diltia5.mol  
diltia4.mol  
diltia8.mol  
diltia7.mol  
diltia6.mol  
diltia9.mol  
diltia11.mol  
diltia10.mol  
diltia15.mol  
diltia14.mol  
diltia13.mol  
diltia12.mol  
diltia20.mol  
...
```

Running a structure solution

```
#!/bin/bash
```

```
filein=molecules.txt  
template=diltia_tmpl
```

```
rm results.dat summary.dat  
c=`wc -l $filein | awk '{p=$1}'`  
  
for i in `seq 1 $c`;  
do  
a=`cat $filein | head -$i`  
file=${a%.*}  
echo $file $template  
sed "s%STRNAME%$file%" $template > "$file.exp"  
#expo "$file.exp" "$file.out" --nogui  
mpirun -np 20 $HOME/expo/expo_mpi "$file.exp" "$file.out"  
echo "===== " >> results.dat  
echo "$file.out" >> results.dat  
grep @@ "$file.out" | cut -c3- >> results.dat  
grep @@best "$file.out" | cut -c3- >> summary.dat  
done
```

```
%Structure STRNAME
```

```
%Job CSD refcode: CEYHUJ01
```

```
%Data
```

```
Cell 42.190 9.075 6.037 90 90 90
```

```
SpaceGroup p 21 21 21
```

```
Pattern pd_0029.pow
```

```
Wavelength 1.54056
```

```
%fragment STRNAME.mol
```

```
%fragment atoms Cl
```

```
deletehydro
```

```
%sanneal
```

```
nrun 20
```

```
%save STRNAME.expo
```

Running a structure solution process on each estimated conformer

```
#!/bin/bash
```

```
filein=molecules.txt  
template=diltia_temp1
```

```
rm results.dat summary.dat *.expo *.bin *_best*.cif  
c=`wc -l $filein | awk '{print $1}'`
```

```
for i in `seq 1 $c`;  
do
```

```
a=`cat $filein | head -$i | tail -1`  
file=${a%.*}
```

```
echo $file $template
```

```
sed "s%STRNAME%$file%" $template > "$file.exp"
```

```
#expo "$file.exp" "$file.out" --nogui
```

```
mpirun -np 20 expo_ompi "$file.exp" "$file.out"
```

```
echo "===== " >> results.dat
```

```
echo "$file.out" >> results.dat
```

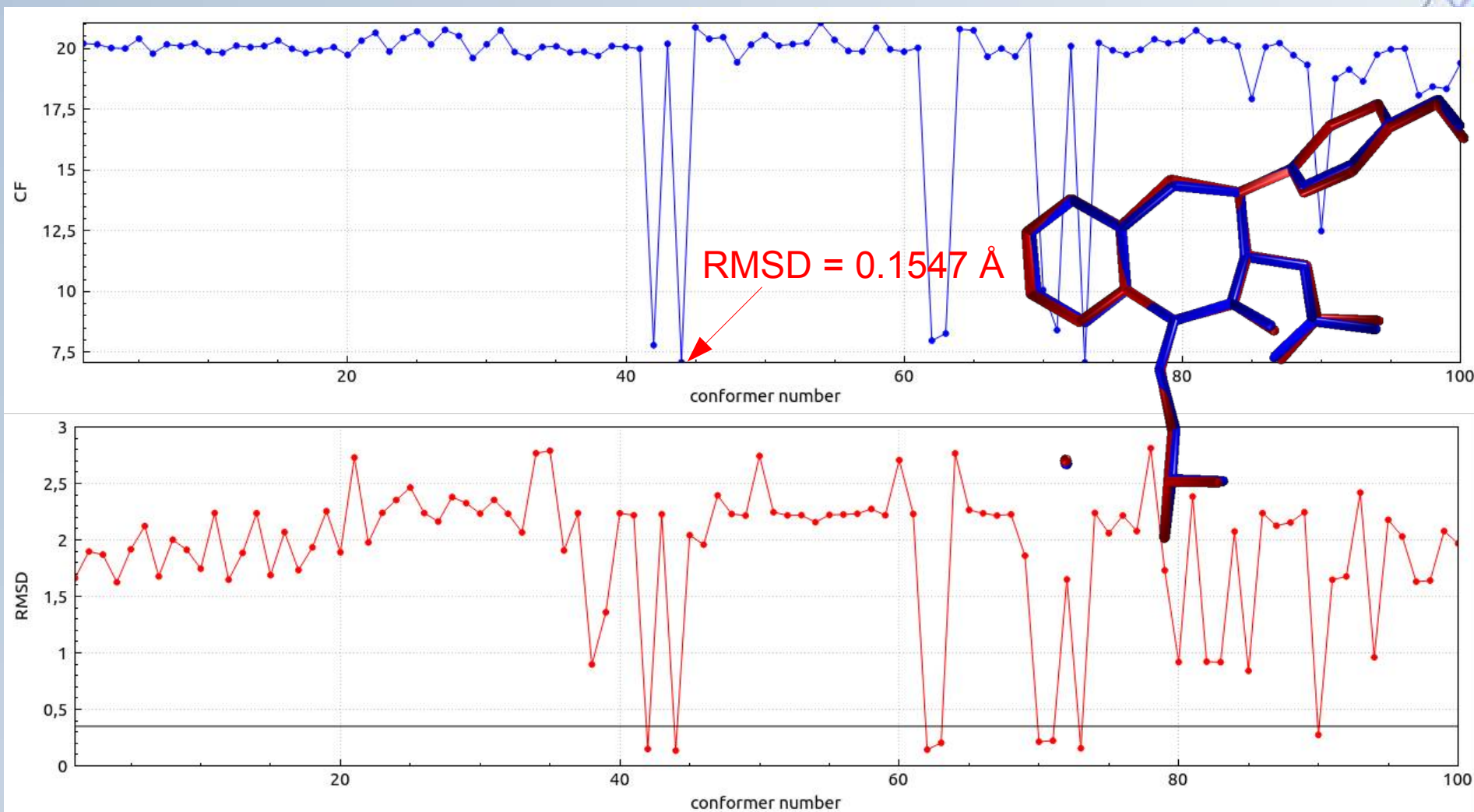
```
grep @@ "$file.out" | cut -c3- >> results.dat
```

```
grep --with-filename @@best "$file.out" | sed 's/:@@best//g'
```

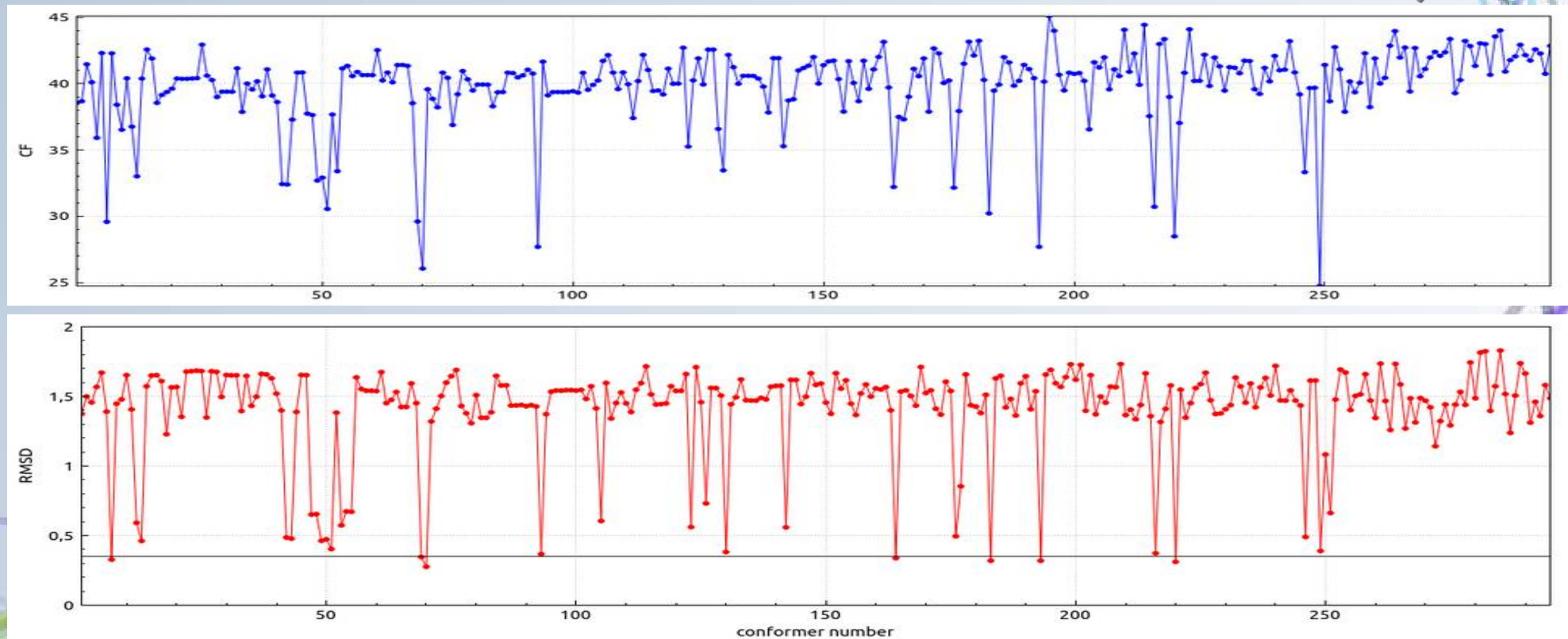
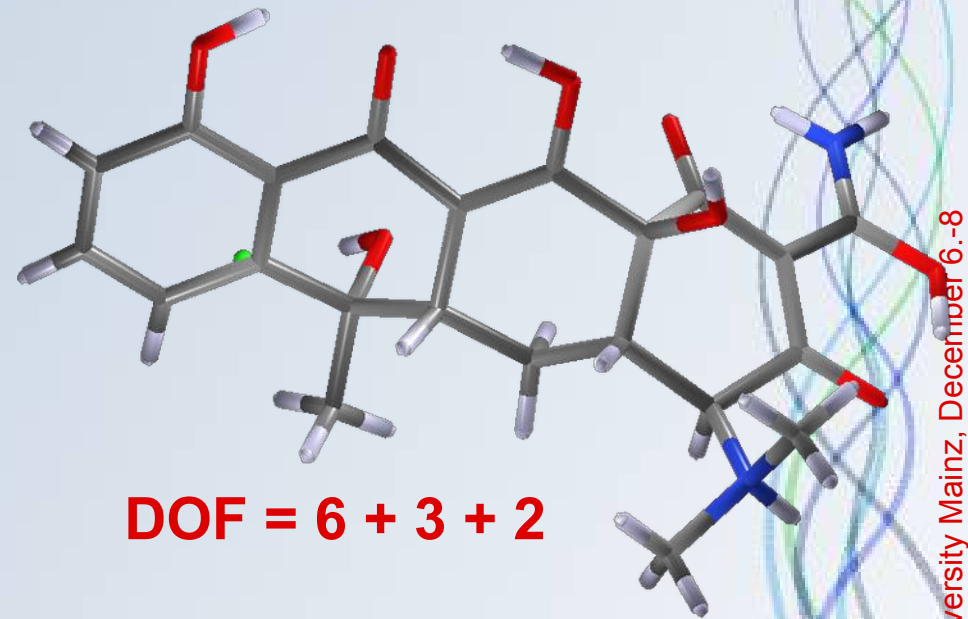
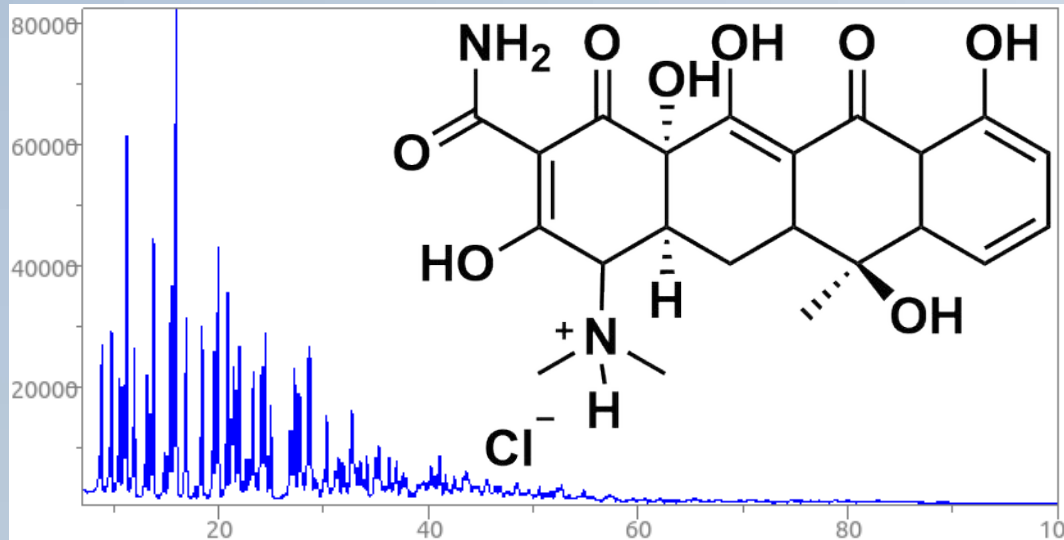
```
>> summary.dat
```

```
done
```


Analysis of the results



The crystal structure of tetracycline hydrochloride



Parallel Machines

Notebooks



Typically 2-8 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

Parallelization Strategies

❖ CPU-Based Parallelization Strategies



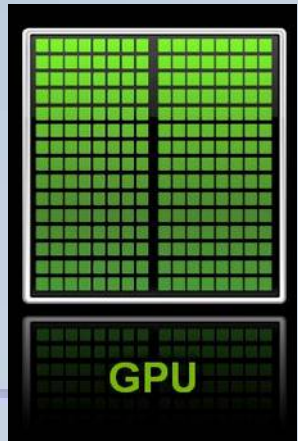
- **Message Passing Interface (MPI)**

Distributed-memory architecture

- **Open MultiProcessing (OpenMP)**

Shared-memory architecture

❖ GPU-Based Parallelization Strategies



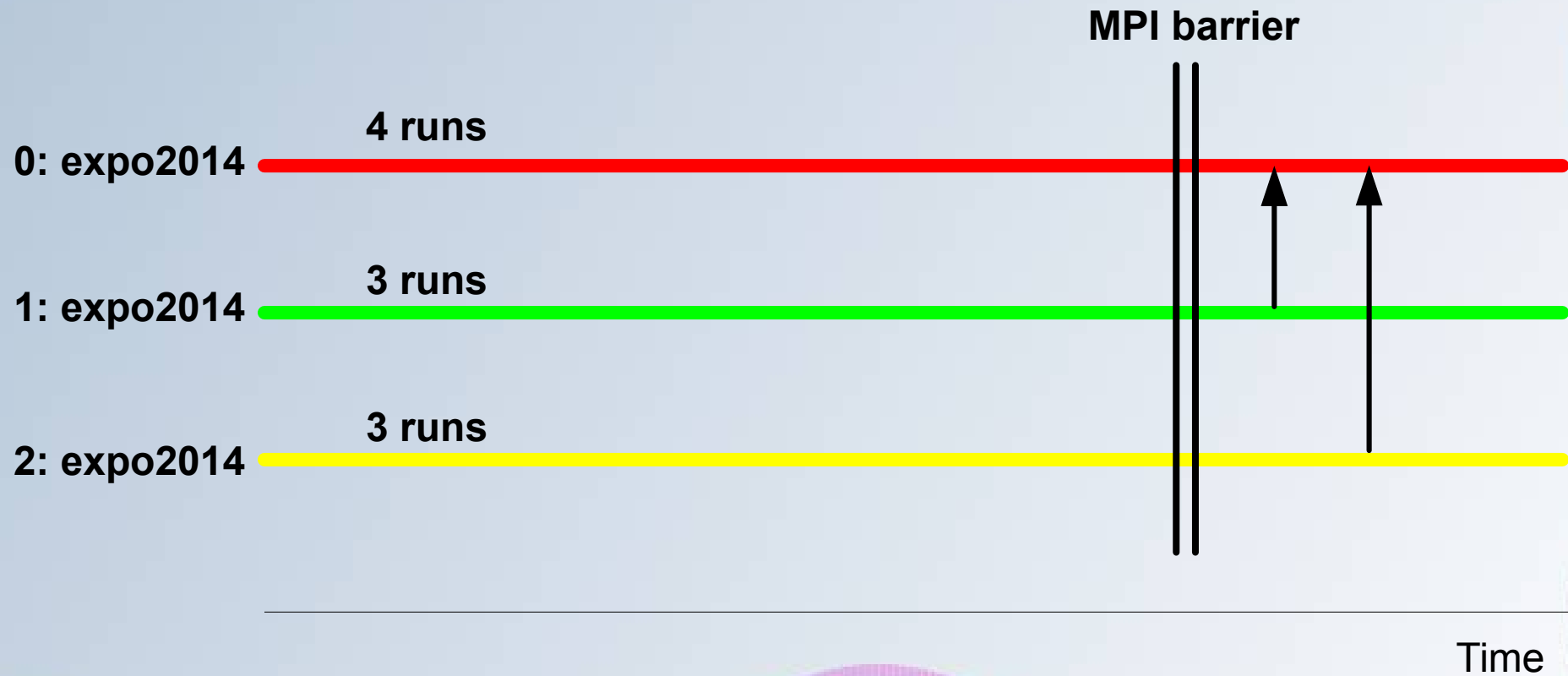
- **Compute Unified Device Architecture (CUDA)**

- **Open Computing Language (OpenCL)**

- **OpenACC**

Message Passing Interface (MPI)

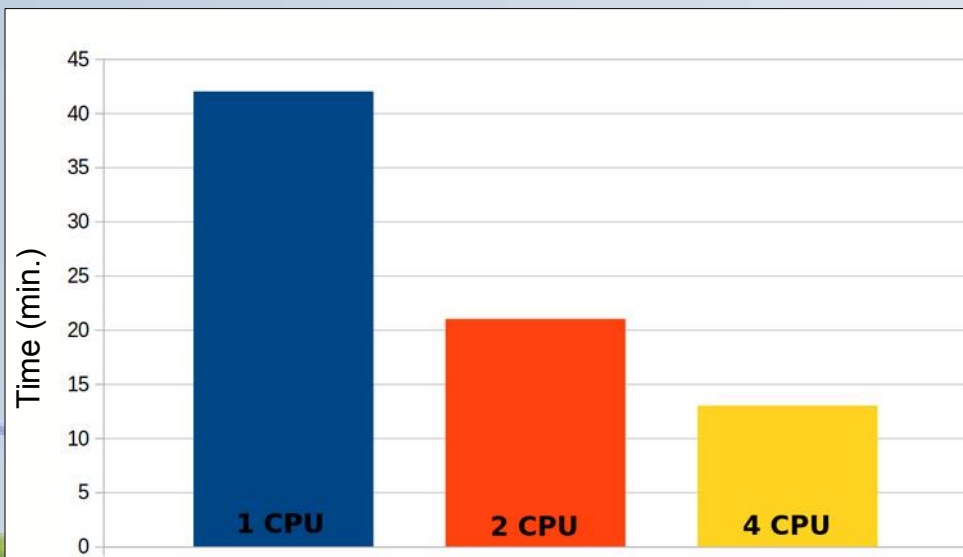
```
mpirun -np 3 expo input_file.exp
```



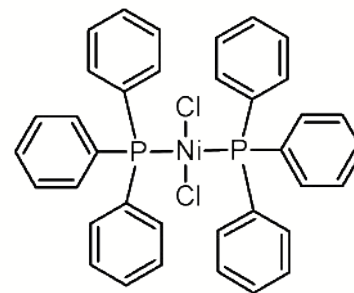
Running the Parallel Version of EXPO2014

- Computer with multi-core CPUs and Linux environment.
- MPI installed.
- Compiling EXPO2014 from source and linking with MPI libraries
- Run Expo2014 by using the launcher `mpirun` with the appropriate options.

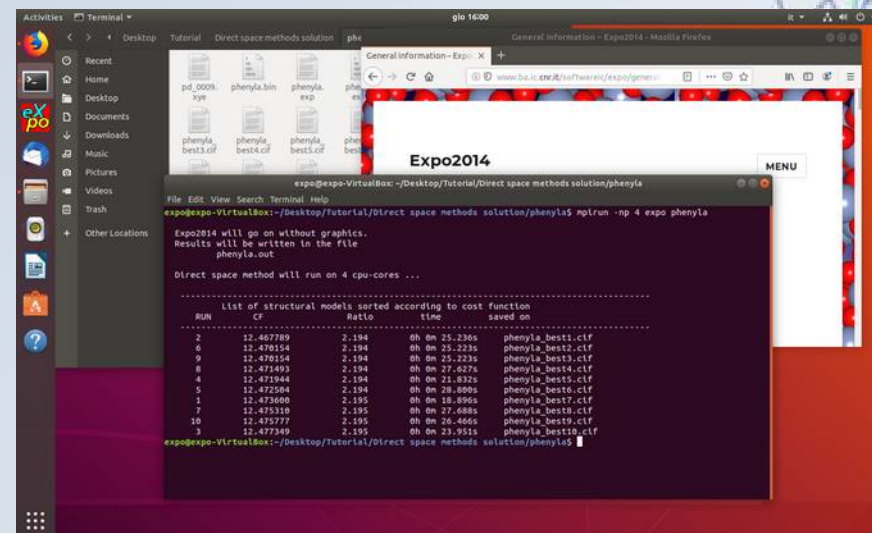
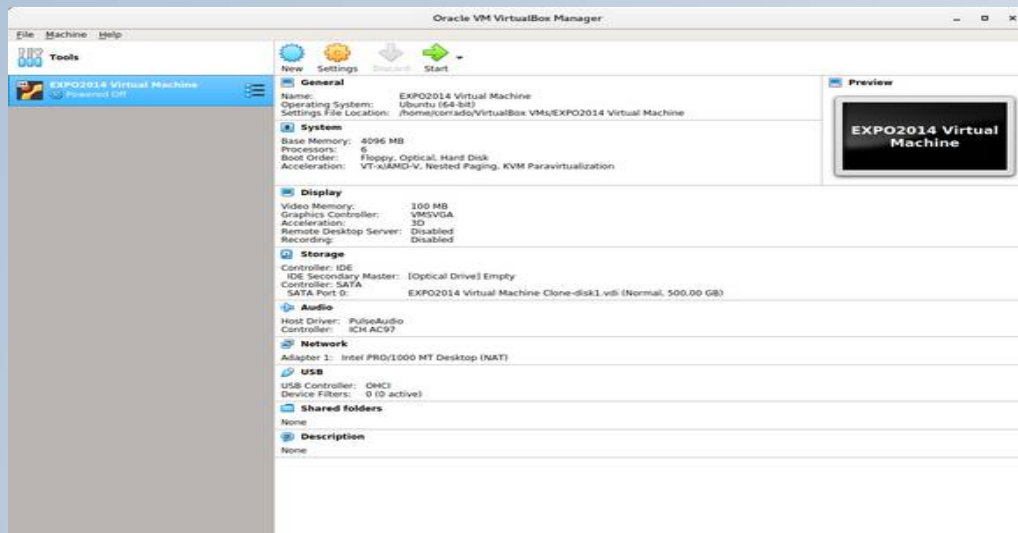
```
mpirun -np 10 expo input_file.exp
```



Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz



Running parallel version of EXPO2014 on Windows

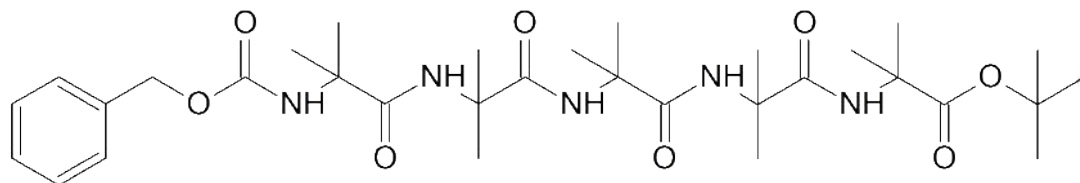


Download and installation

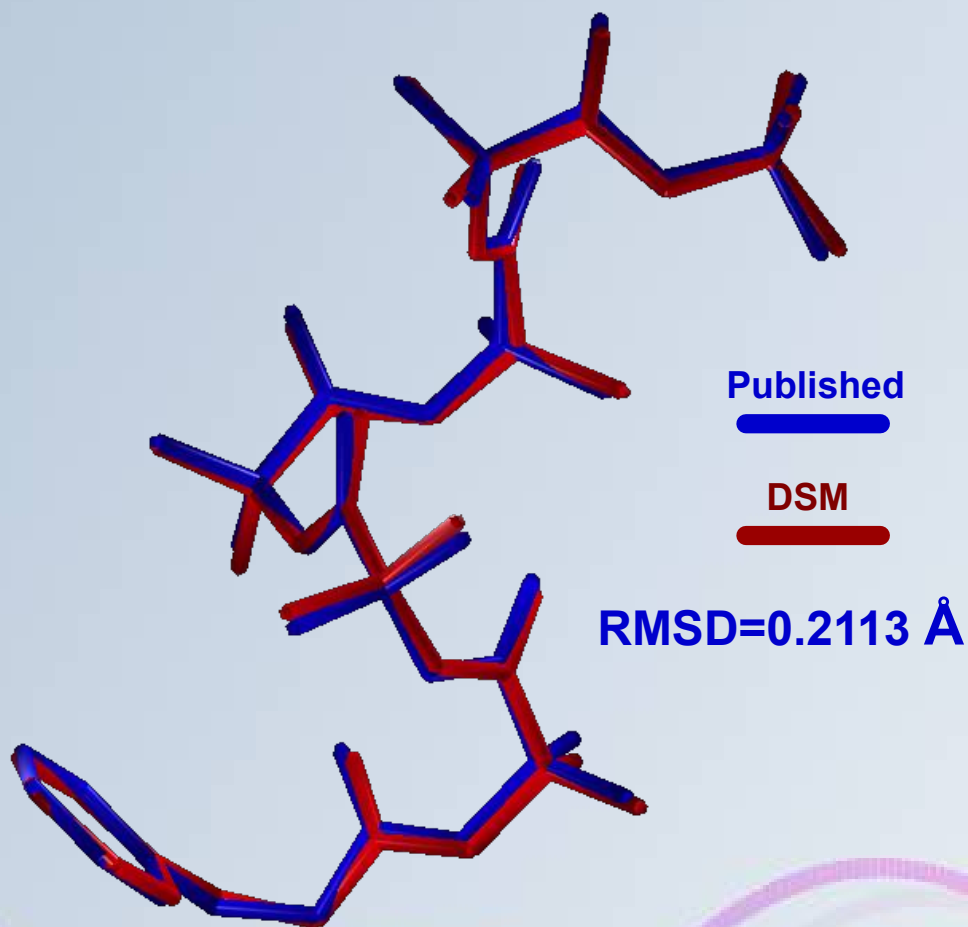
<http://www.ba.ic.cnr.it/softwareic/expo/parallelism-in-expo2014-for-structure-solution-by-direct-space-method/>

Windows Subsystem for Linux (WSL)

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 solution of selexipag form I

```
%structure vohvia
```

```
%data
```

```
pattern VOHVIA.cif
```

```
cell 37.96347 6.110426 22.47454 90 98.3273 90
```

```
space P21/c
```

```
%fragment vohvia_mopac.mol
```

```
%fragment vohvia_mopac.mol
```

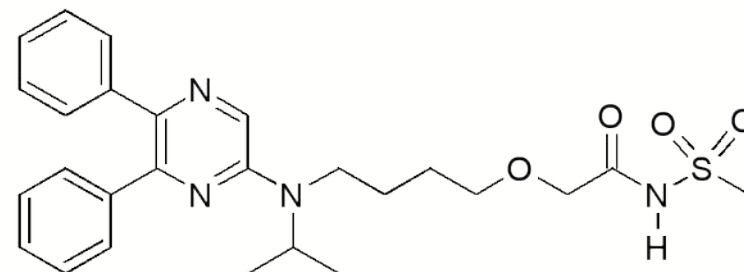
```
deletehydro
```

```
%sannel
```

```
nrun 100
```

```
niter 2000
```

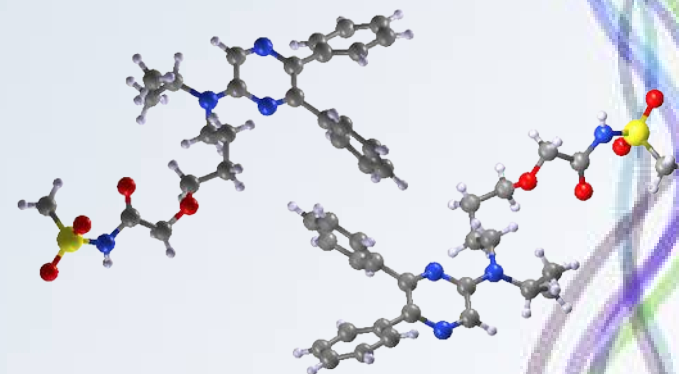
```
resm 2.5
```



DOF: 6+6+13+13 = 38

Time: ~45h on 25 CPU-cores (2.90 GHz)

RMSD: 0.3234 Å



Reference

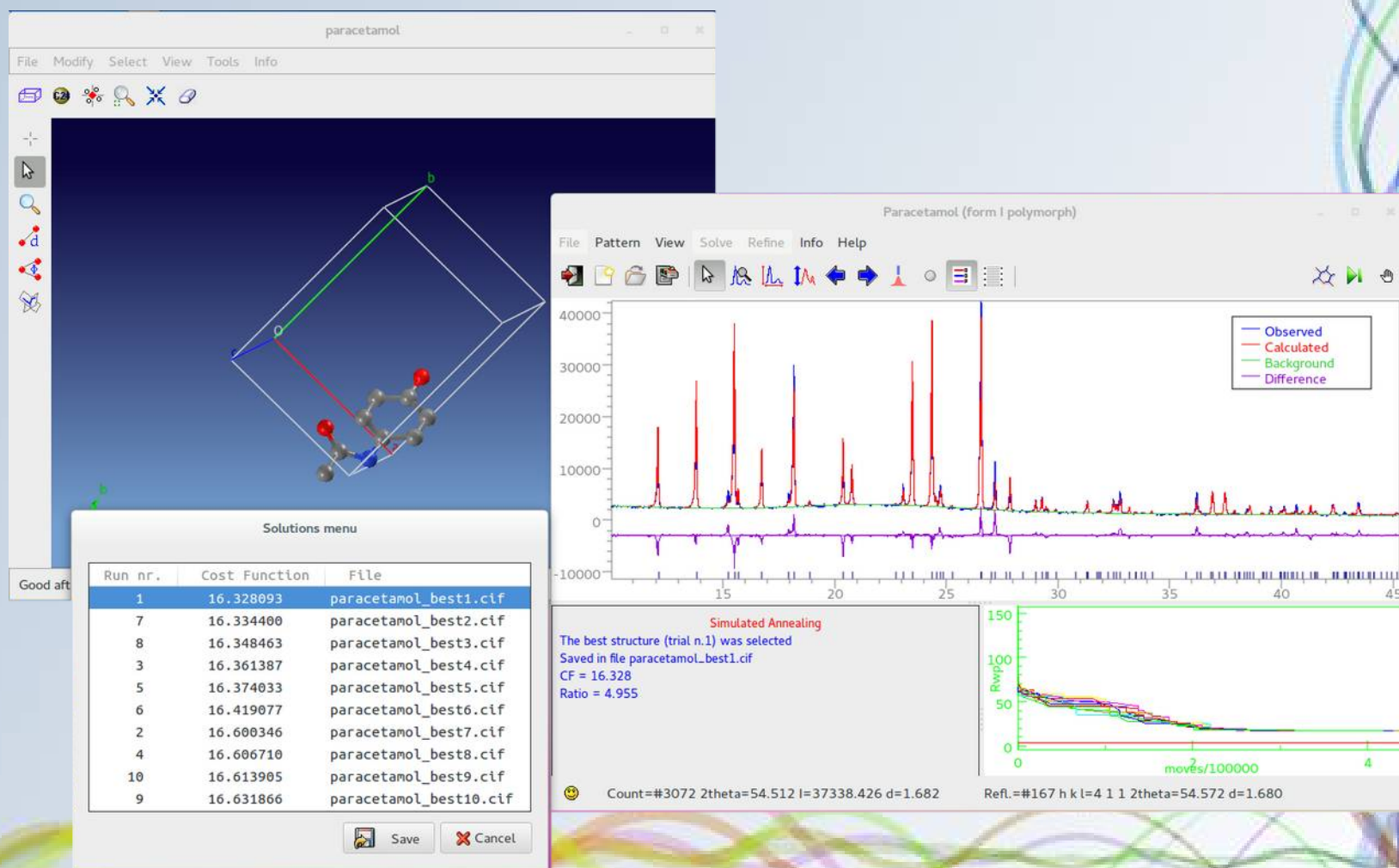
Successful Strategy for High Degree of Freedom Crystal Structure Determination from Powder X-Ray Diffraction Data: A Case Study for Selexipag Form I with 38 DOF

Michal Hušák, Alexandr Jegorov, Jiri Czernek, Jan Rohlíček, Simona Žižková, Pavel Vraspír, Pavel Kolesa, Andrew Fitch, and Jiri Brus

Cryst. Growth Des. 2019, 19, 8

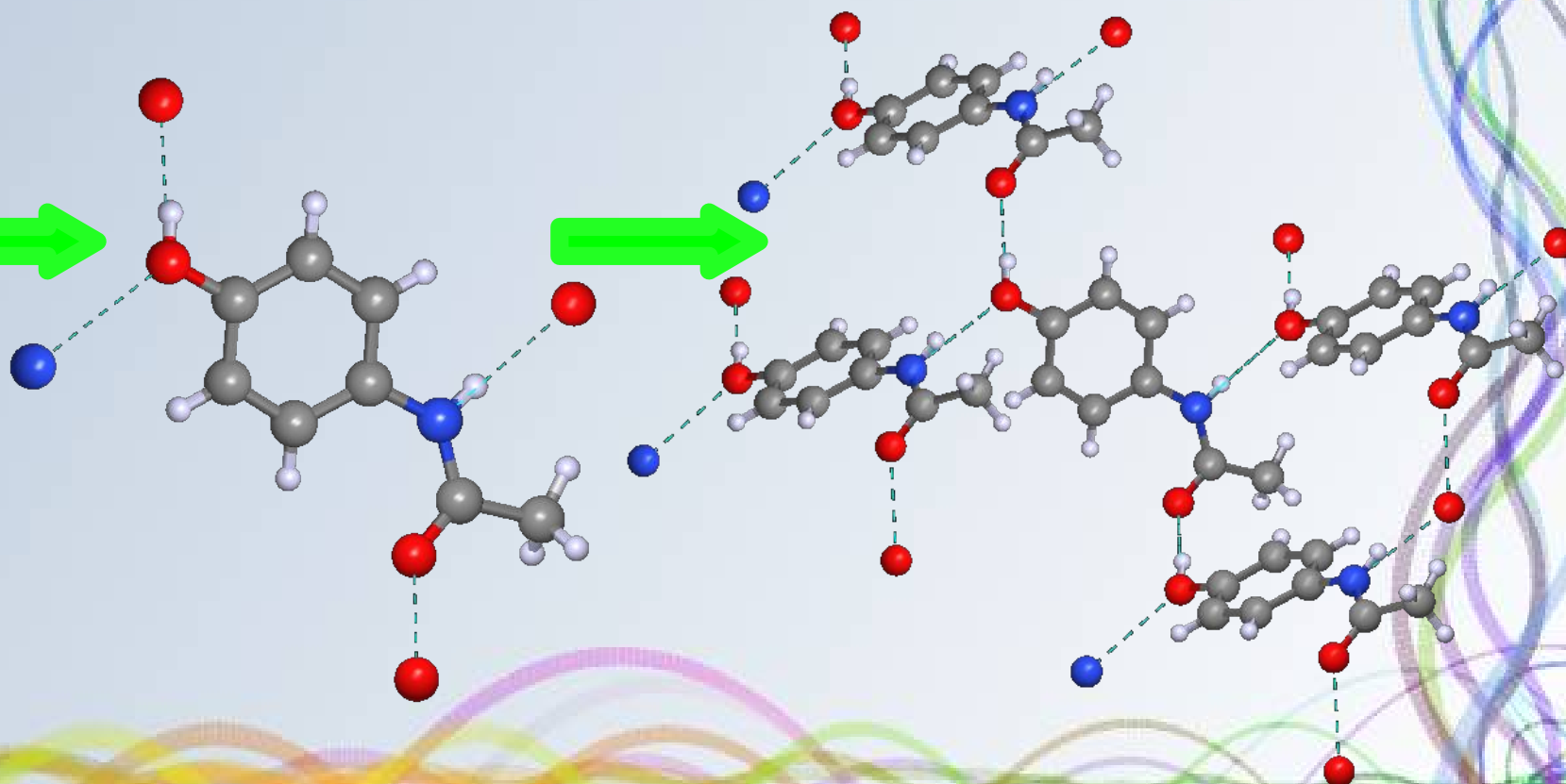
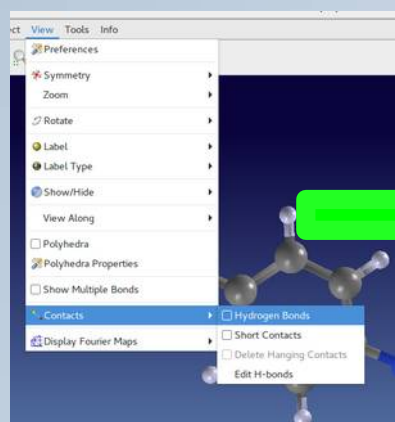
Assessing the solution

- Agreements factors
- Visual match between the calculated and observed profile
- Reproducibility of solution



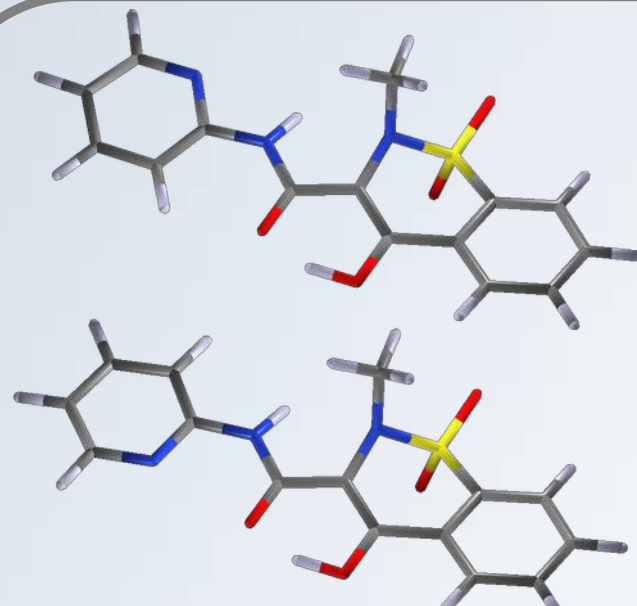
Assessing the solution

- Crystal packing
- Check close contacts, void spaces, likely interactions
- Network of interactions: hydrogen bonds and short contacts



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)

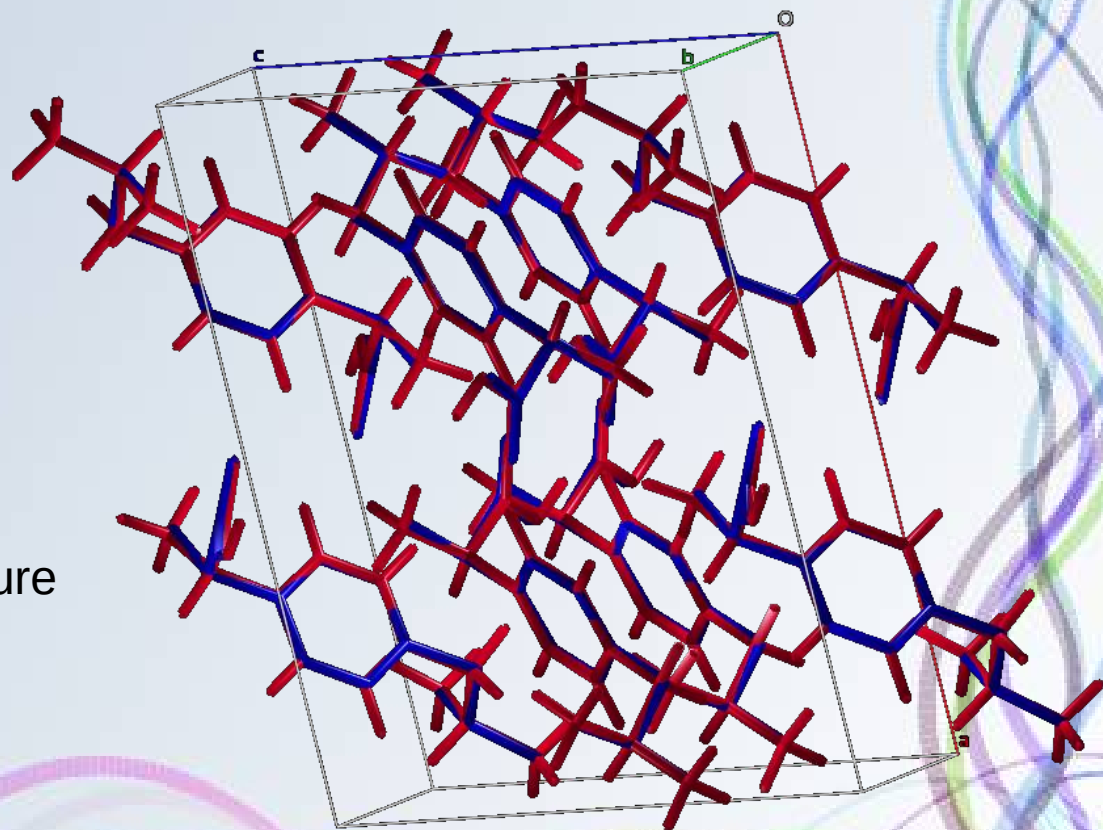
Assessing the solution with DFT-D

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

RMSD for non H-atoms above 0.35 Å could indicate incorrect experimental crystal structure *

Ibuprofen
RMSD=0.091 Å

-  Experimental crystal structure
-  DFT-D3 with NWChem



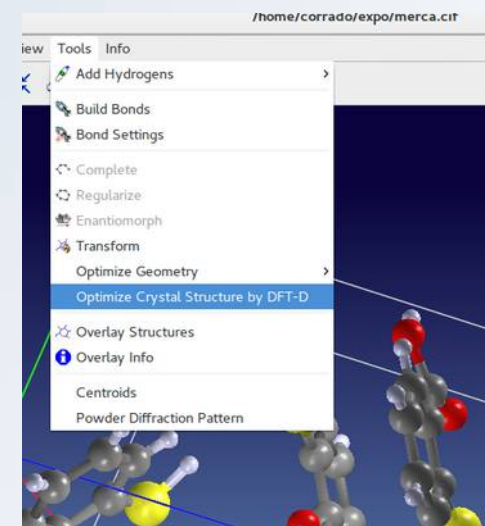
*Jacco ven de Streek *et al.* Validation of molecular crystal structures from powder diffraction data with dispersion-corrected density functional theory (DFT-D) *Acta Cryst.* (2010). B66, 544–558

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



When Structure Solution Fails

■ Starting model is incorrect:

- ❖ *chemical formula is wrong*
- ❖ *bond distance and angle are not entirely accurate*
- ❖ *number of building blocks is wrong*
- ❖ *missing solvent*
- ❖ *.....*



Solution:

- ❖ *Check the compositional information (MS, SEM/EDS, XRF, ICP, NMR)*
- ❖ *Try different combination of building blocks*
- ❖ *Check the molecular stereochemistry, or ring conformation*
- ❖ *Improve your model with CSD or building packages*
- ❖ *Use the SS-NMR to deduce the number of molecules in the asymmetric unit (Z')*

When Structure Solution Fails

- Poor quality diffraction pattern



Solution:

- *Collect new data*
- *Add restraints or anti-bump restraints*

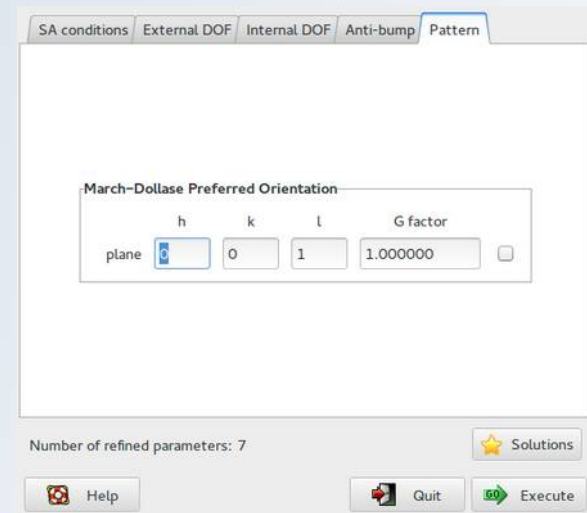
- Systematic problem in powder diffraction data

- *Preferred orientation*
- *Ka2 contributions*



Solution:

- *Collect new data*
- *Refine preferred orientation parameters*



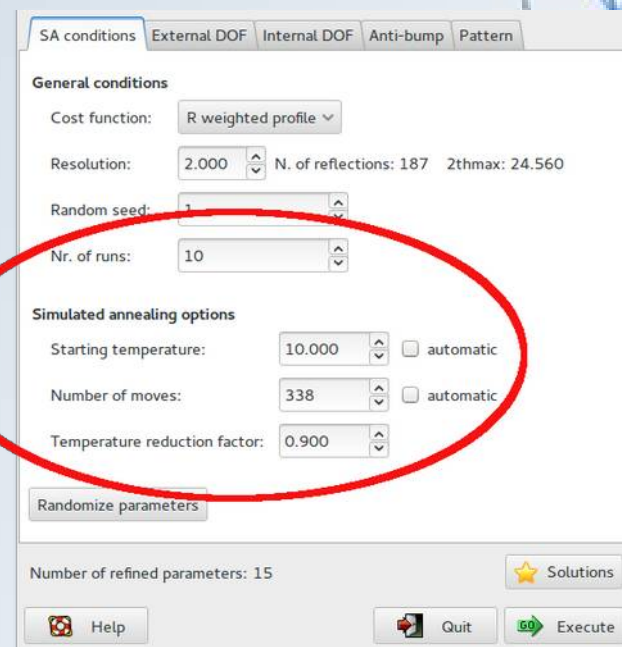
When Structure Solution Fails

- For complex structure (internal DOF > 10) the default SA conditions could be insufficient



Solution:

- *Increase the number of moves (`niter` directive) and/or runs (`nrun` directive)*



- The assumptions about thermal factors are invalid



Solution:

- *Try altering the non-hydrogen atom temperature factors*
- *Check temperature factors for similar structures*

When Structure Solution Fails

- Space group and cell are not correct



Solution:

- *It may be necessary to carry out a series of independent calculations to test different potential space groups and/or unit-cell choices*

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, A. Falcicchio and
F. Baldassarre

Computer resources

CINECA under ISCRA initiative, project HP10CQK3FF

Thank you for your kind attention