

The background features a complex molecular structure with grey carbon atoms, white hydrogen atoms, red oxygen atoms, and blue nitrogen atoms. Some atoms are connected by dashed lines, indicating hydrogen bonds. A central part of the molecule includes orange and green spheres, possibly representing a metal complex or a specific functional group. The entire scene is set against a light blue gradient background, with colorful wavy lines in shades of purple, yellow, and green at the bottom and right edges.

Global Optimisation Methods **simulated annealing & co.**

Corrado Cuocci

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

Methods of Structure Solution

Measured diffraction
intensities



Approximate positions
of atoms

- Direct methods
- Charge flipping
- Molecular replacement
- Direct space methods

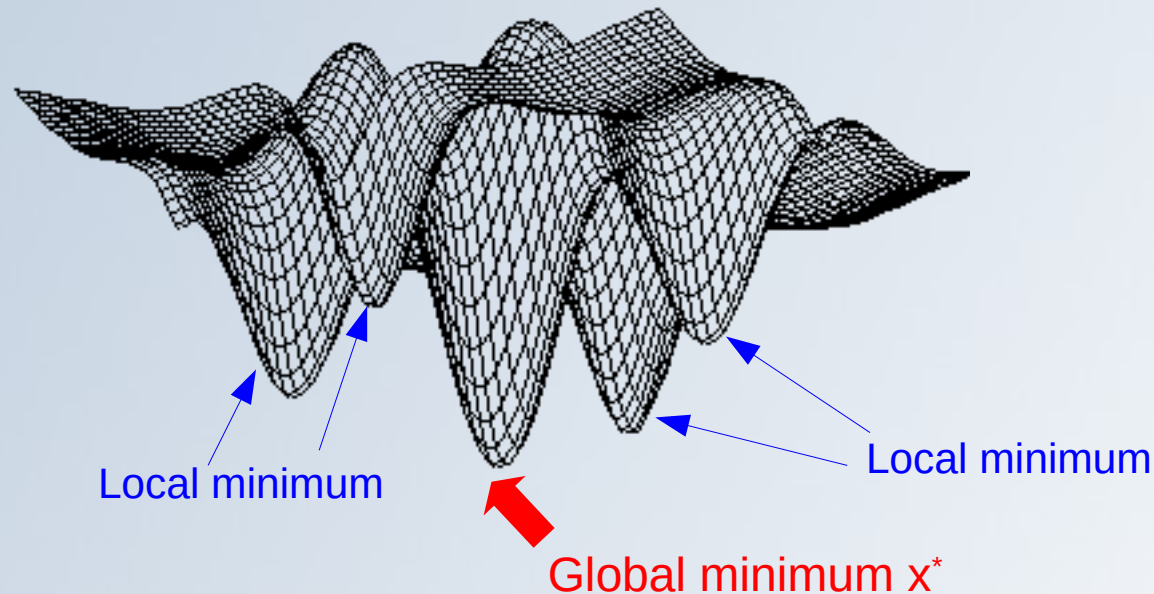
*Alternative expressions: real space,
global optimization, global search*

Global Optimization Methods

Find $x^* = \min\{F(x)\}$, 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 $(\tau_1, \tau_2, \dots, \tau_n)$



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

.....

✓ Advantages

- Theoretical guarantee of finding the global optimum
- Repeatability

✗ Disadvantages

- Computationally expensive

✓ Advantages

- Greater flexibility
- Practical efficiency
- Well-suited for problems with many local minima
- Generally easy to implement

✗ Disadvantages

- No guarantee of finding the global optimum.
- Non-deterministic results
- May require careful parameter tuning to perform well.

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)**

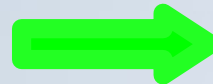
Bee Algorithms

Firefly Algorithms

Harmony Search

*Big Bang-Big Crunch**

.....



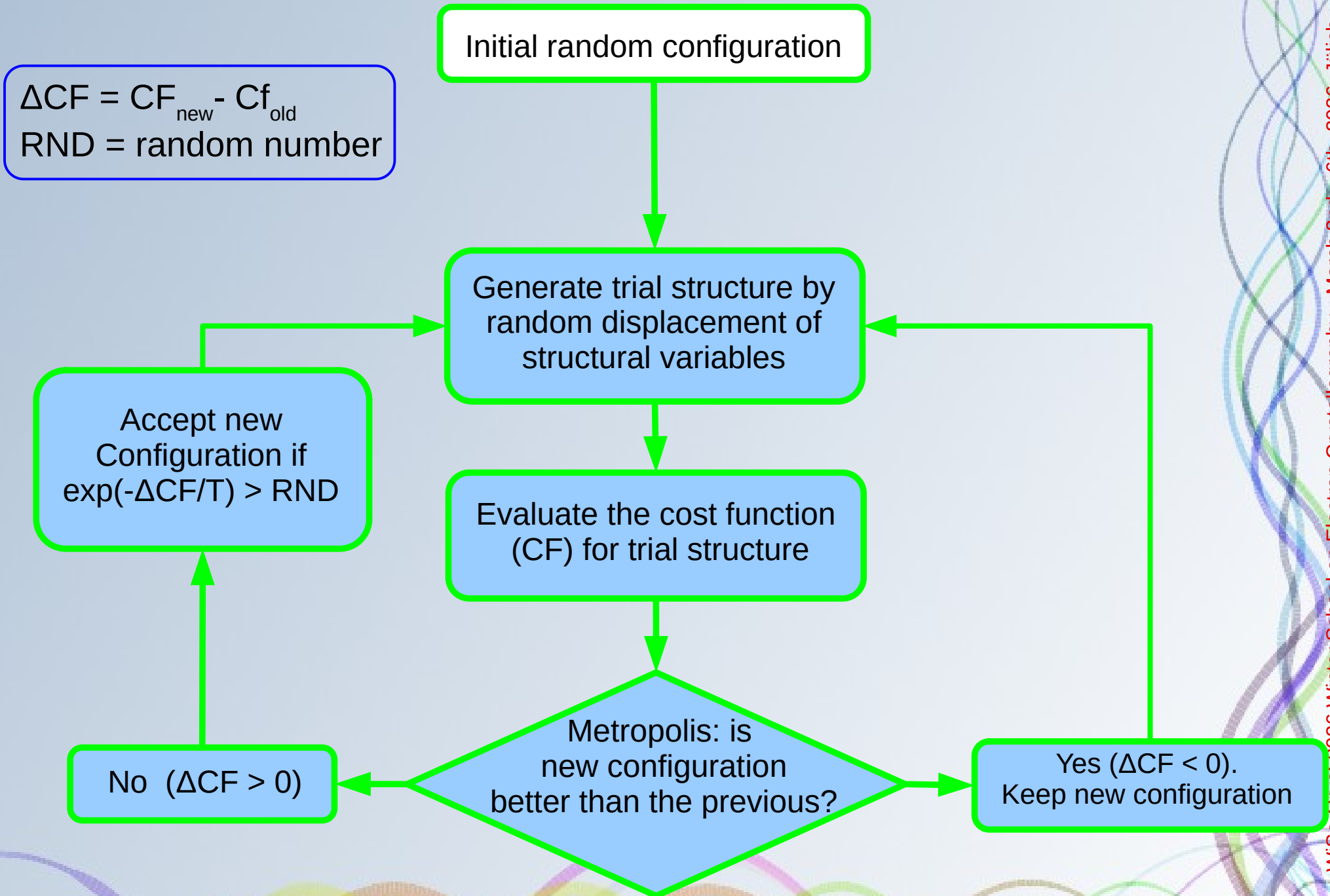
Widely used and with
the largest impact

Various modifications:

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

(*) *employed in solving crystal structure*

Simulated annealing algorithm: the annealing steps



Simulated annealing algorithm: the annealing steps

$$\Delta CF = CF_{\text{new}} - CF_{\text{old}}$$

RND = random number

Initial random configuration

i=0

Generate trial structure by
random displacement of
structural variables
i=i+1

Evaluate the cost function
(CF) for trial structure

Metropolis: is
new configuration
better than the previous?

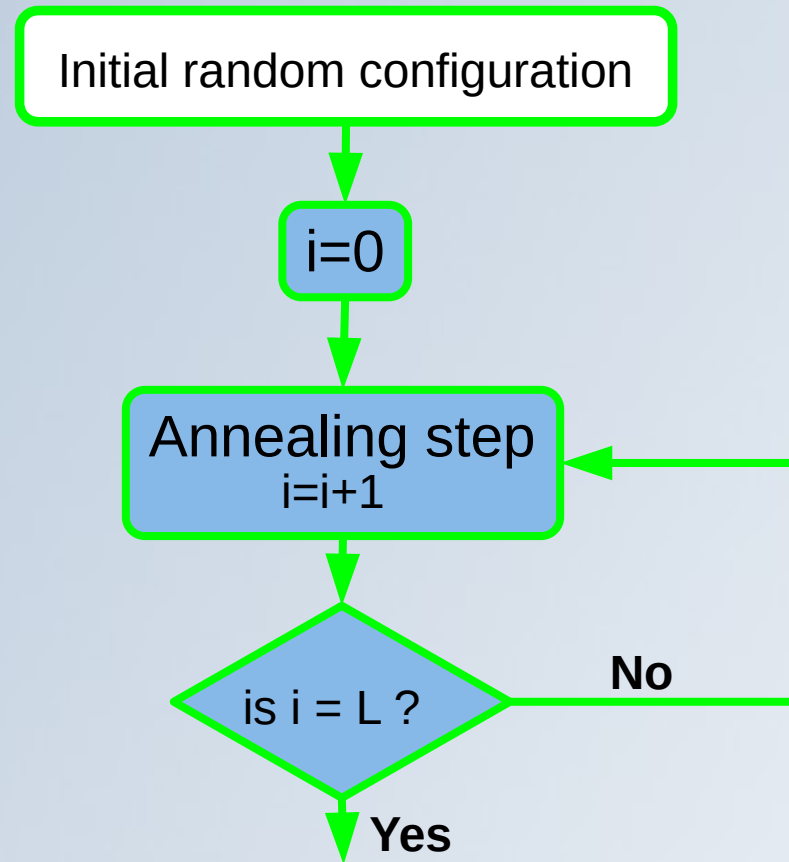
Accept new
Configuration if
 $\exp(-\Delta CF/T) > \text{RND}$

No ($\Delta CF > 0$)

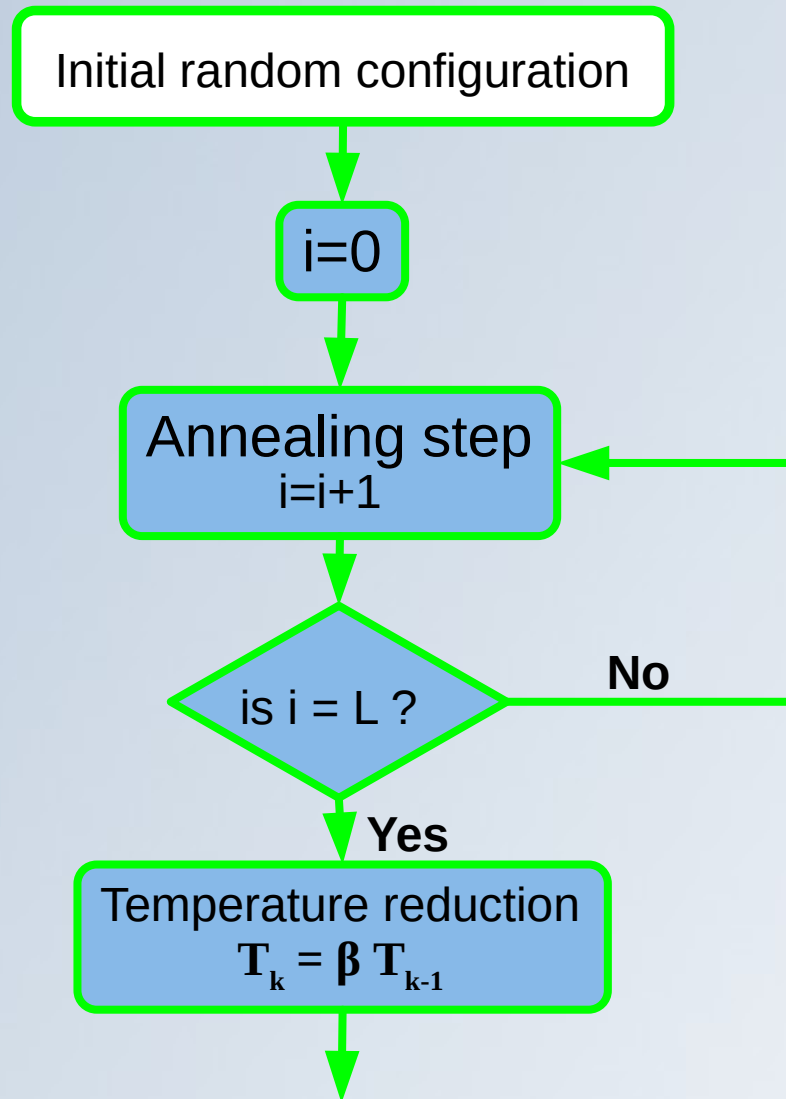
Yes ($\Delta CF < 0$).
Keep new configuration

is i = L ?

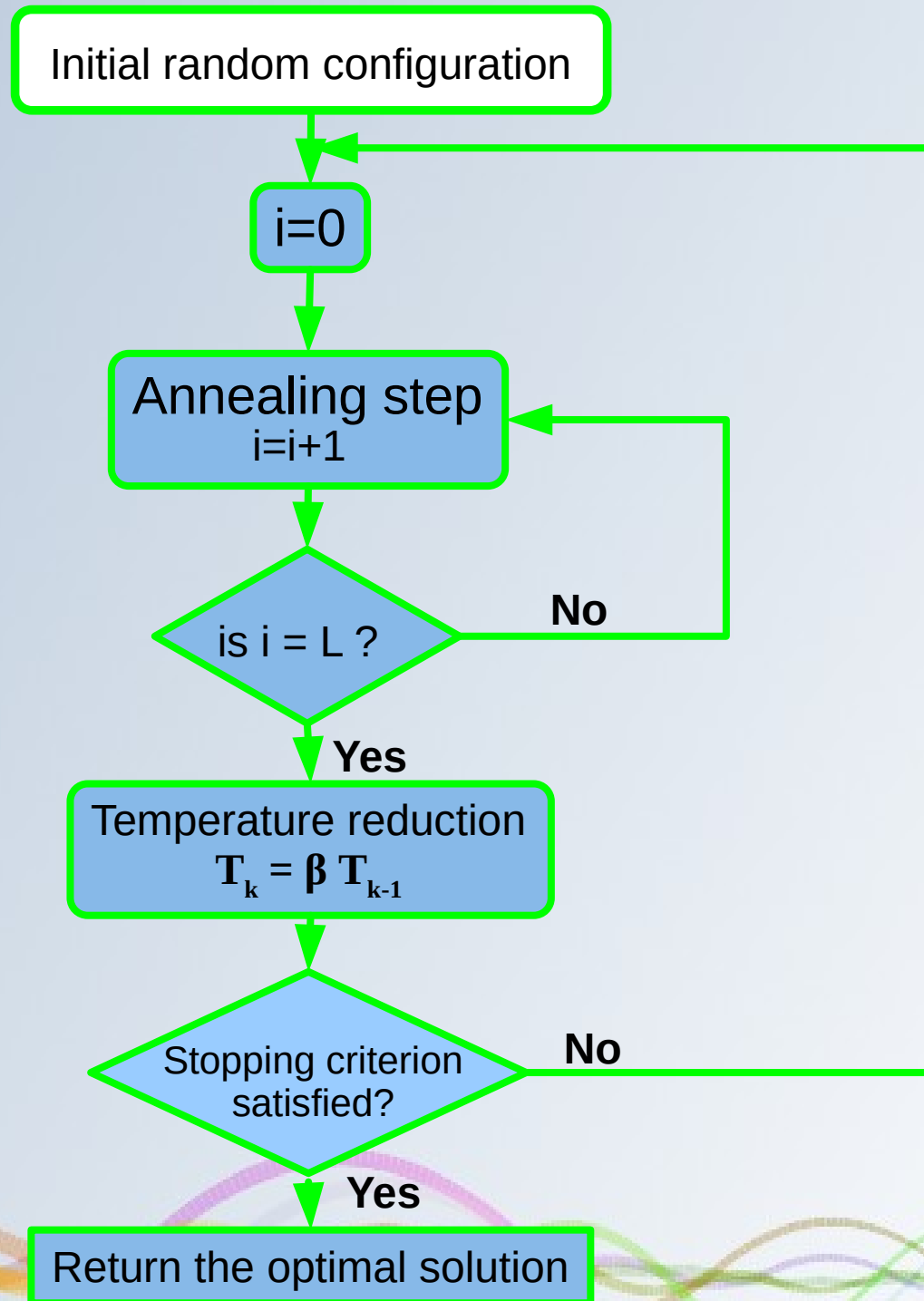
Simulated annealing algorithm: the annealing steps



Simulated annealing algorithm



Simulated annealing algorithm



Simulated annealing options

Initial random configuration

$i=0$

Annealing step
 $i=i+1$

is $i = L$?

No

Yes

Temperature reduction
 $T_k = \beta T_{k-1}$

Stopping criterion
satisfied?

No

Yes

Return the optimal solution

Global Optimization Dialog

Conditions Crystal

General Conditions

Cost function: R weighted profile

Resolution: 2.000 Nr. of reflections: 103 2 θ : 45.3049°

Random seed: 1

Nr. of runs: 10

Simulated Annealing Conditions

Number of moves: 60 ☒ automatic

Starting temperature: 10.00 ☒ automatic

Temperature reduction factor: 0.90

Randomize parameters

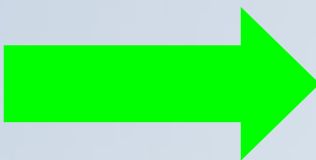
Solutions

Help Run Close

- Number of moves
- Number of SA runs

Simulated annealing options

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

A screenshot of the 'Global Optimization Dialog' window. The 'Conditions' tab is selected, showing 'Crystal' as the condition. Under 'General Conditions', the 'Cost function' is set to 'R weighted profile', 'Resolution' is '2.000', and 'Nr. of reflections' is '103 2θ: 45.3049°'. A green box highlights a list of factors: 'No. of molecular fragments', 'No. of external DoFs', 'No. of internal DoFs', and 'The flexibility of the molecule'. The 'Temperature reduction factor' is set to '0.90'. At the bottom, there are buttons for 'Randomize parameters', 'Solutions', 'Help', 'Run', and 'Close'.

Simulated Annealing Variations

Initial random configuration

$i=0$

Annealing step
 $i=i+1$

is $i = L$?

No

Yes

Temperature reduction

$$T_k = \beta T_{k-1}$$

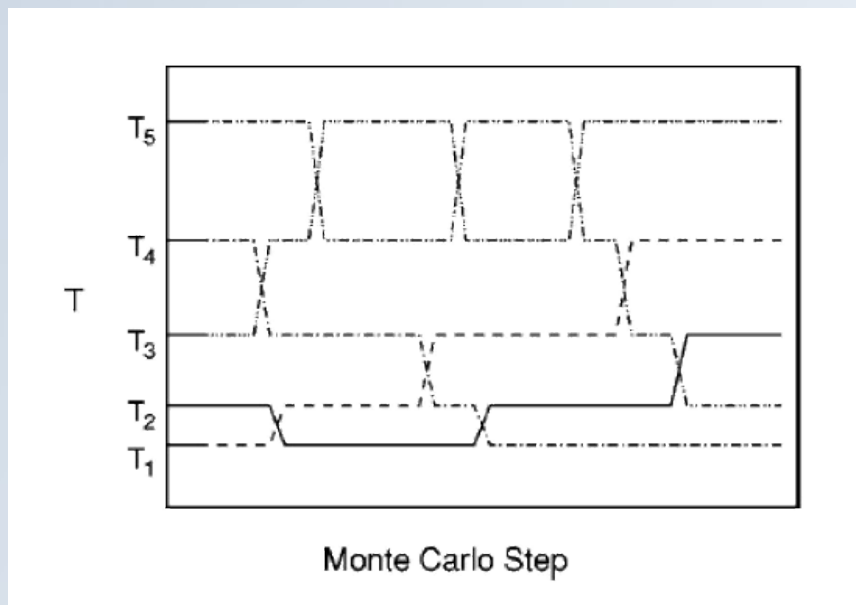
Stopping criterion
satisfied?

No

Yes

Return the optimal solution

- Adaptive Simulated Annealing (ASA)
- Combinations with other optimization techniques
- Parallel Tempering



<https://objcryst-fox.readthedocs.io>

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.

-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

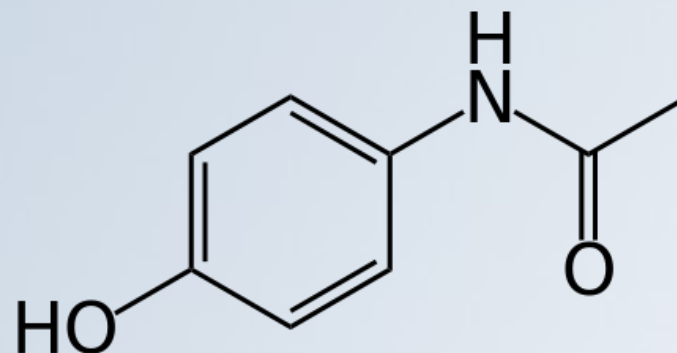
+Complexity of procedure depends on the number of degrees of freedom (DoF).

+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

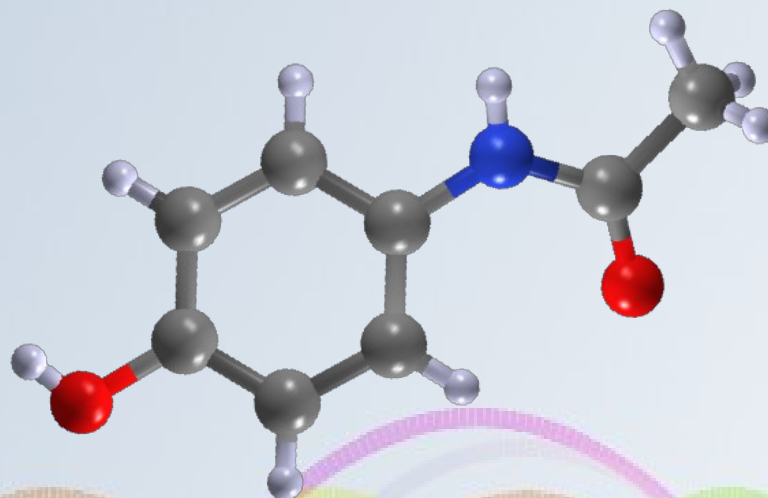
- It is necessary to know the molecular connectivity. Spectroscopic techniques (MS, NMR) can be useful



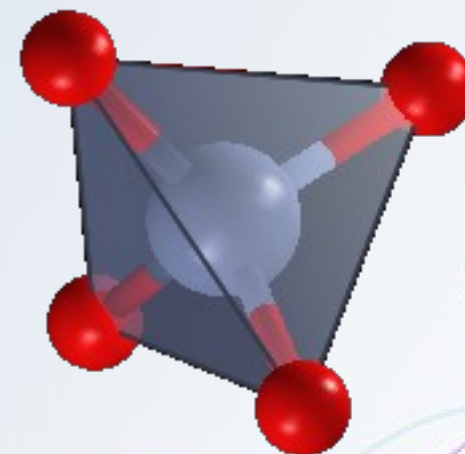
- Crystal structure can be described as a combination of building blocks



atom



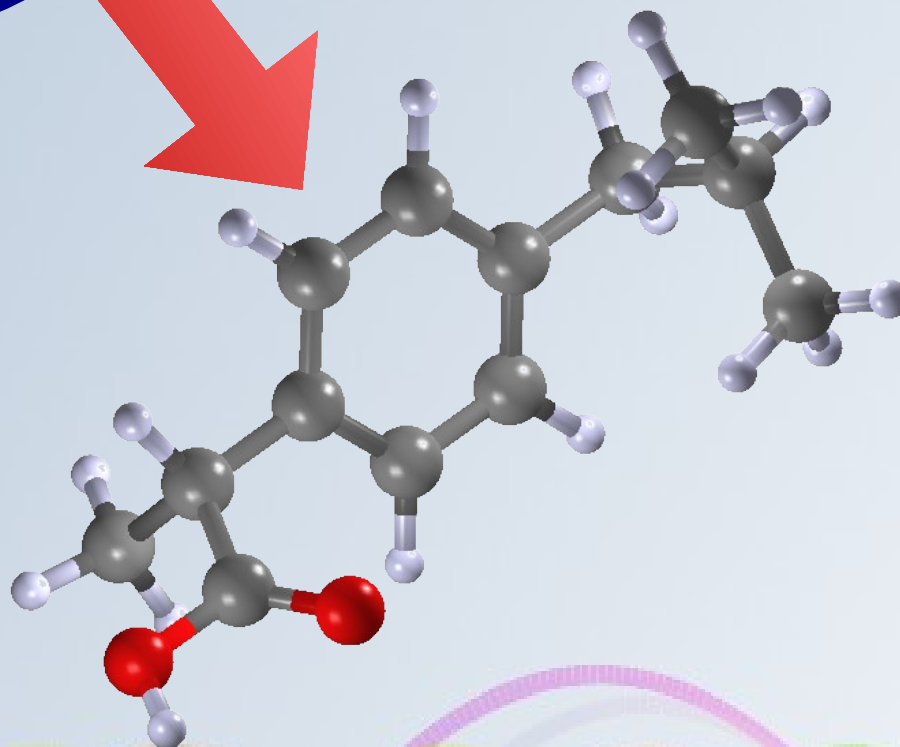
molecule



polyhedron

Building starting model

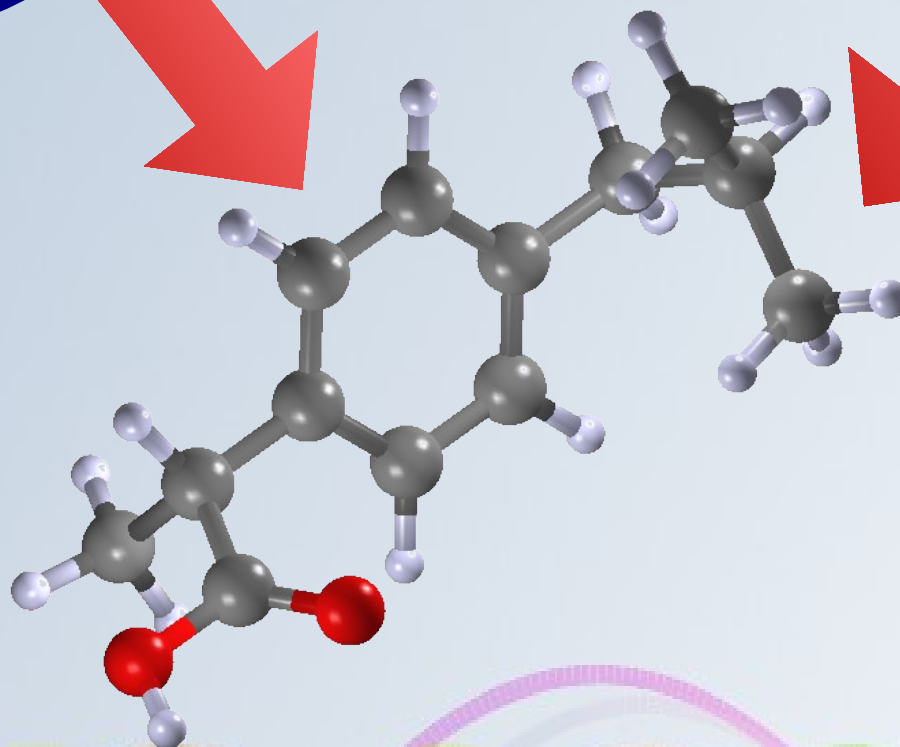
Check for
similar
molecules in
databases



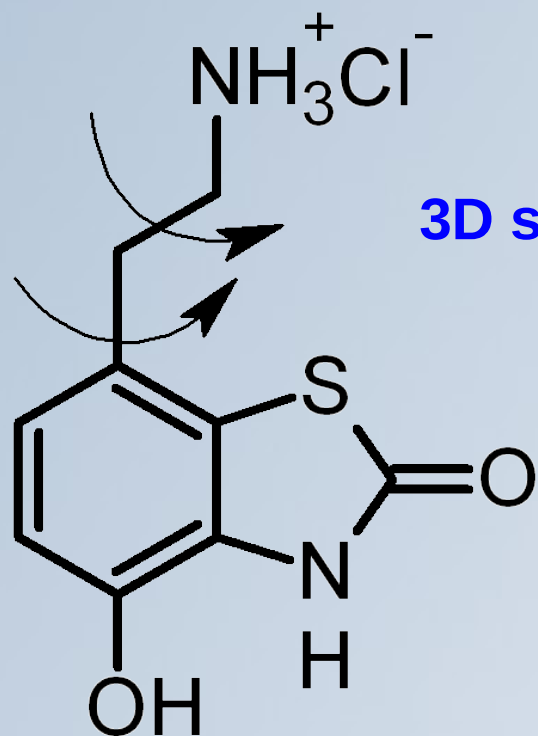
Building starting model

**Check for
similar
molecules in
databases**

**Optimize
Molecular
geometry by
computational
chemistry**



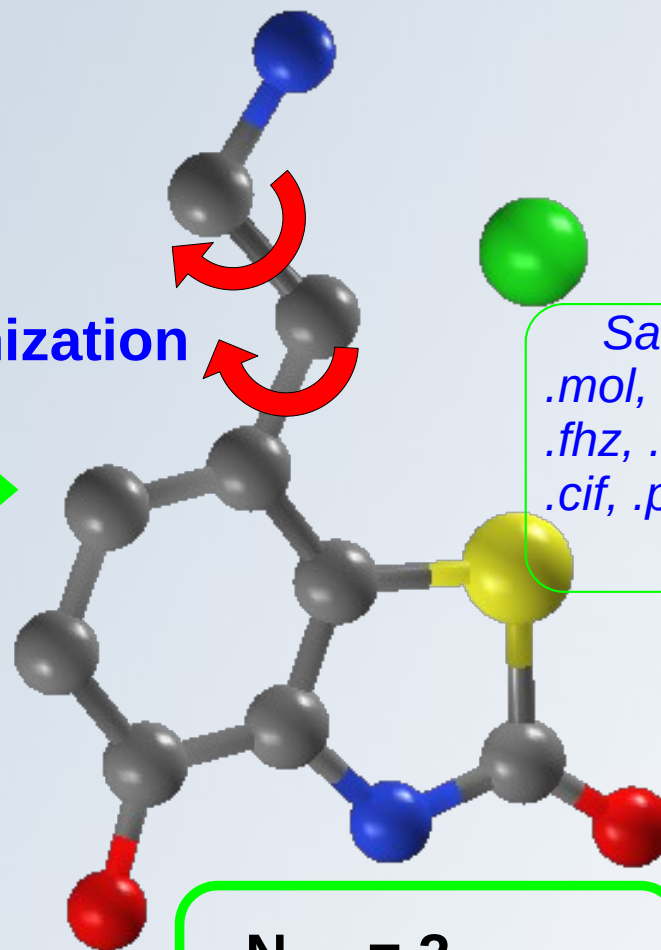
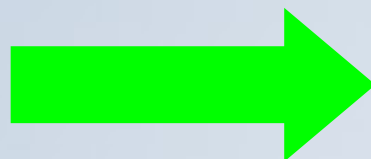
Building starting model



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

P 21/a

3D structure optimization

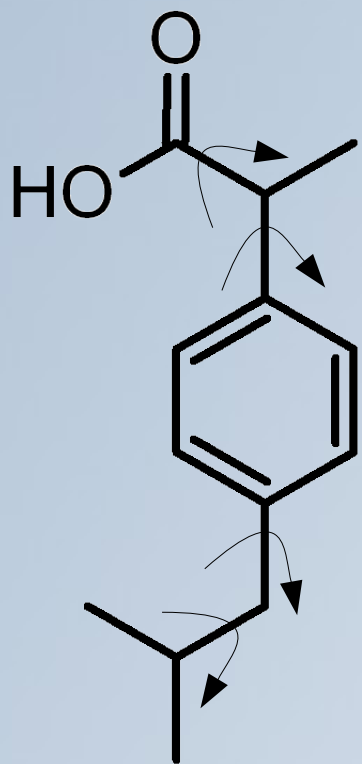


Save as
.mol, .mol2,
.fhz, .mop,
.cif, .pdb, .frac

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

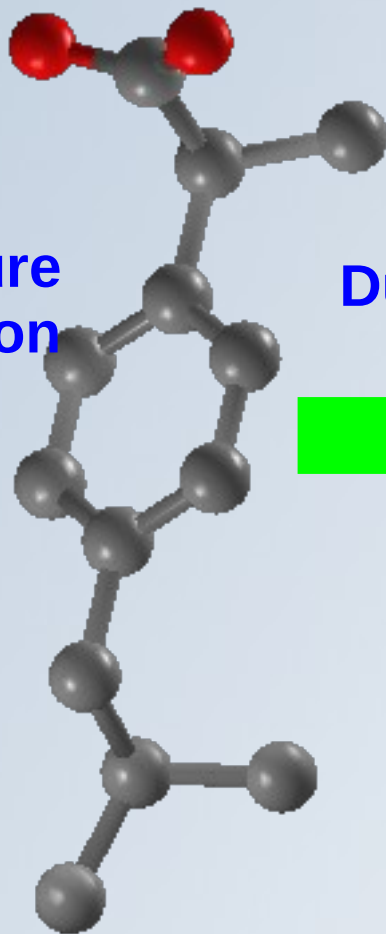
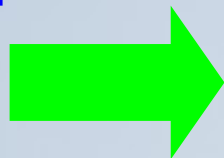
$$N_{\text{dof}} = 6 + 3 + 2$$

Building starting model: example 2

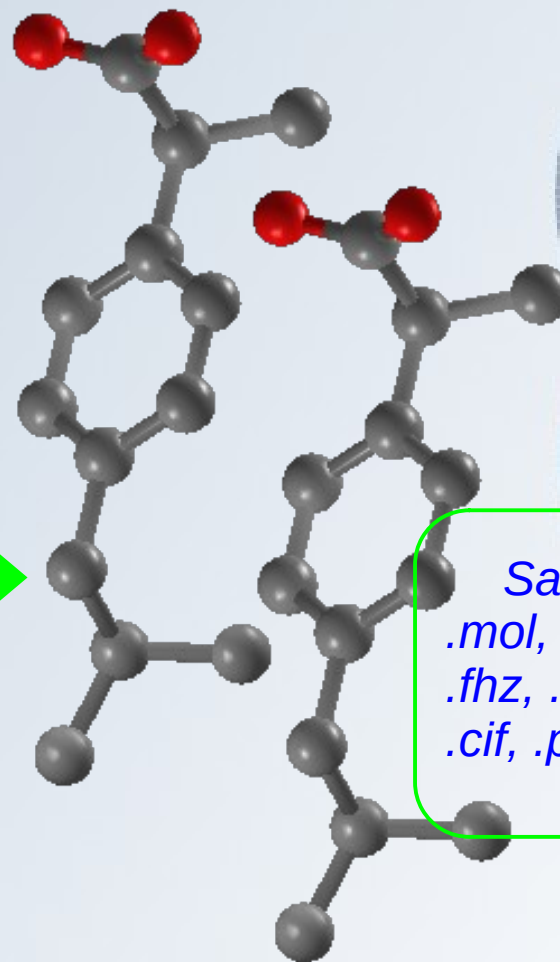


S-Ibuprofen
 $P2_1$

3D structure
optimization



Duplicate



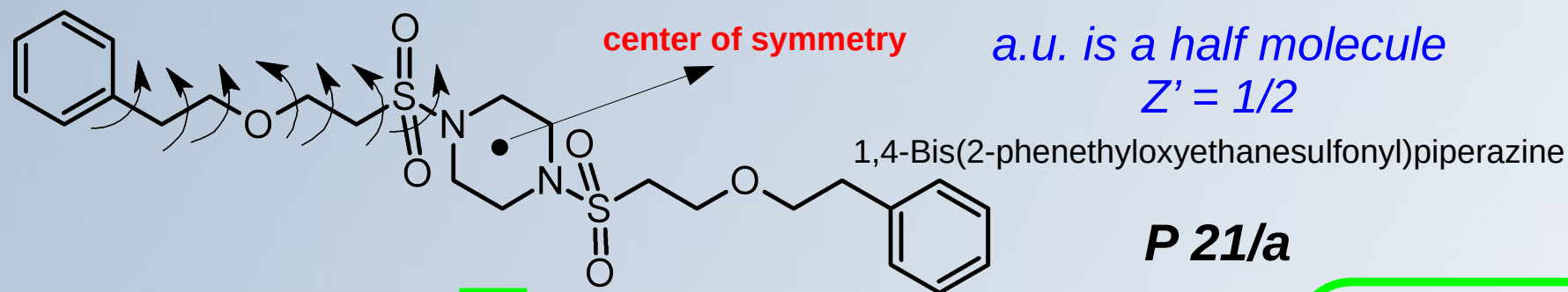
Save as
.mol, .mol2,
.fhz, .mop,
.cif, .pdb, .frag

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

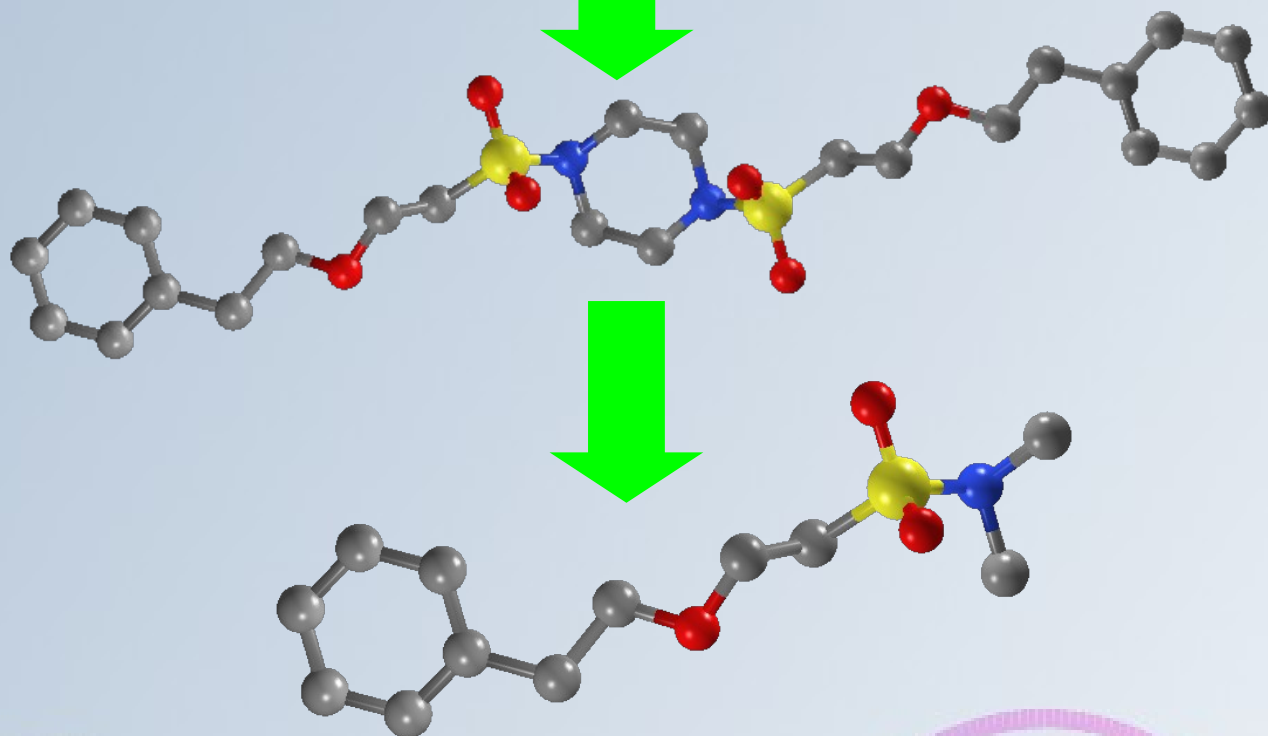
$$N_{\text{dof}} = 6 + 6 + 4 + 4$$

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

Building starting model: example 3



3D structure optimization



Structure solution
by DSM + DOC

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

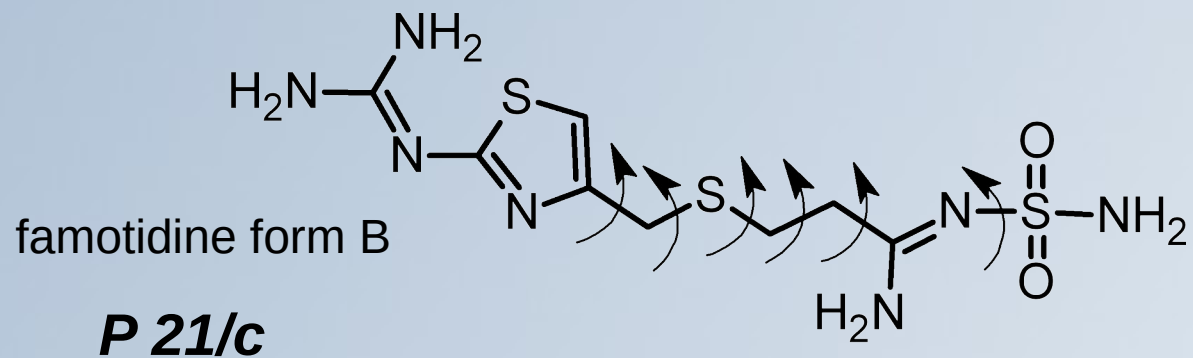
15 times slower

Structure solution
by DSM

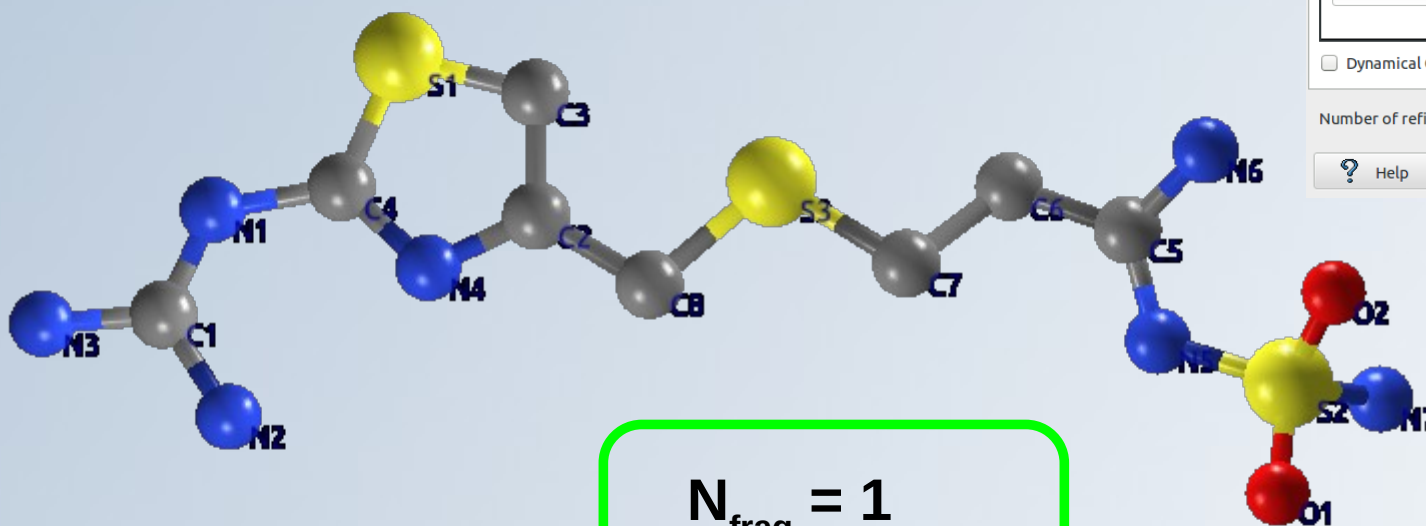
$$N_{\text{dof}} = 6 + 7$$

*Recommended
strategy*

Building starting model: 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

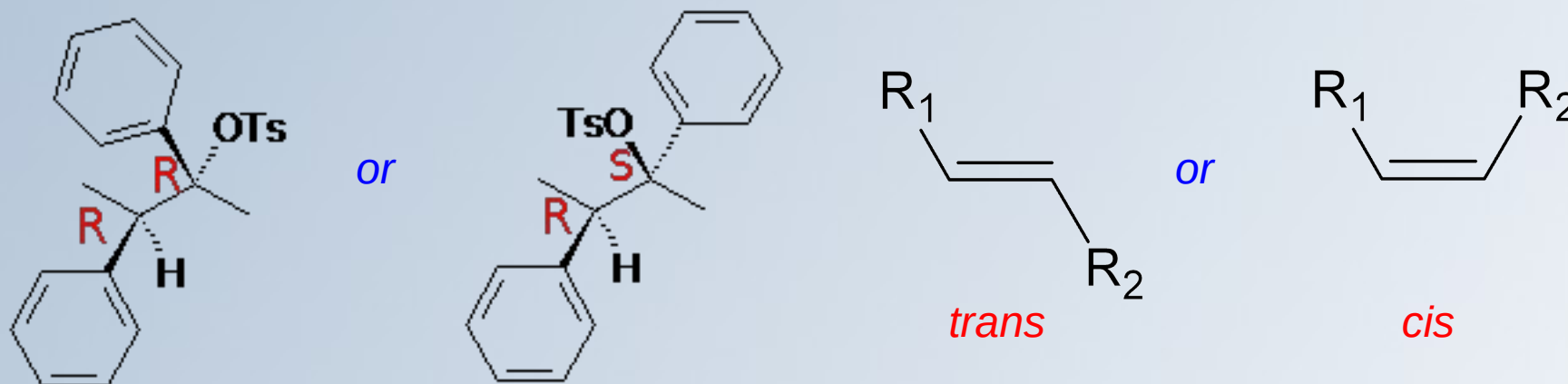
☐ Dynamical Occupancy Correction ☒ Atomic Parameters

Number of refined parameters: 13

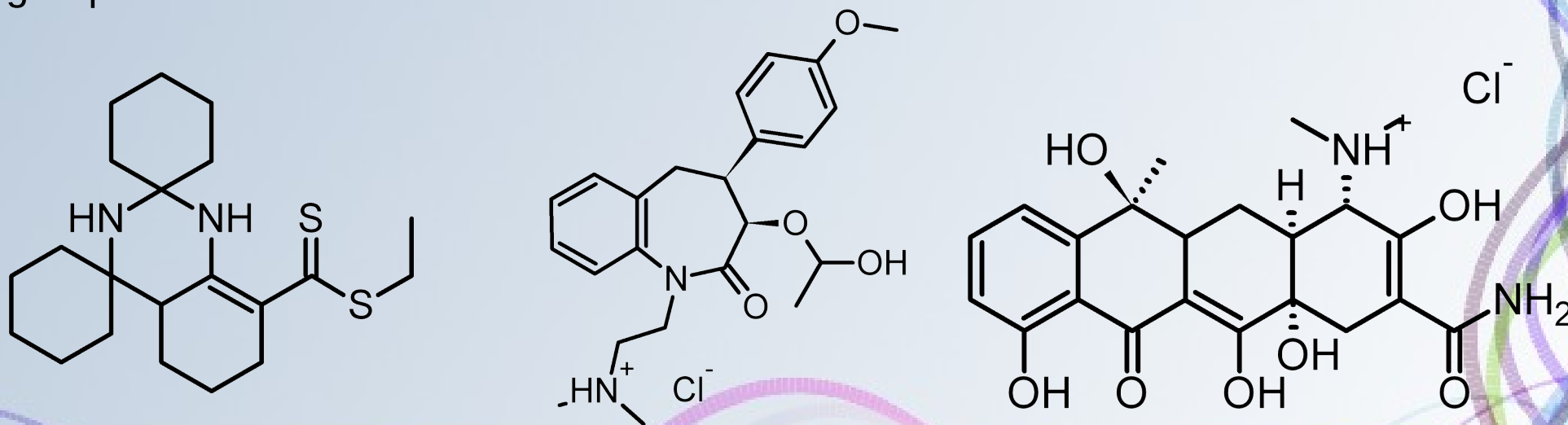
☒ Solutions

Building starting model

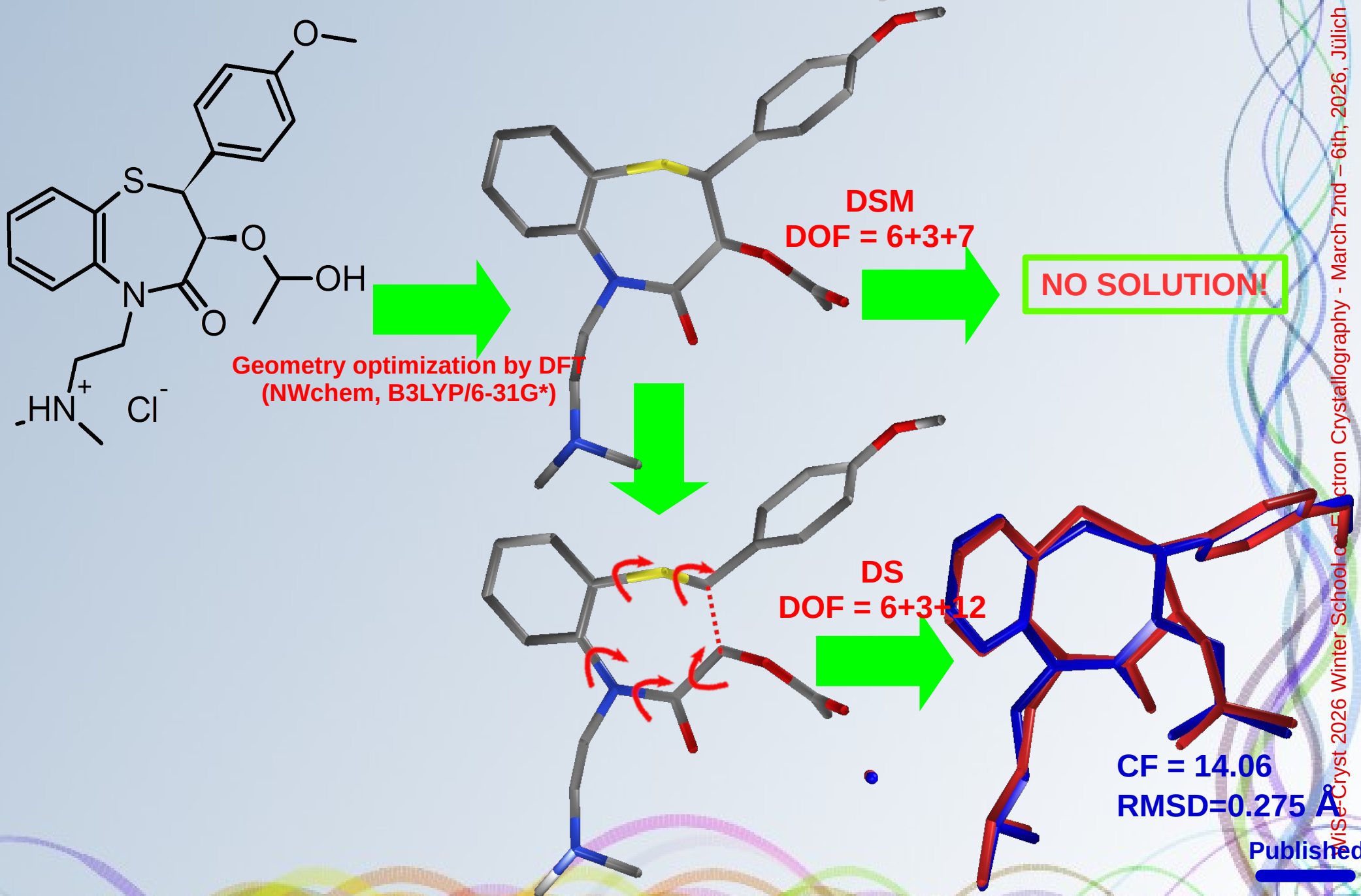
- Stereochemistries will not be altered during simulated annealing. Pay particular attention to compounds with more than one chiral center and cis/trans isomers.



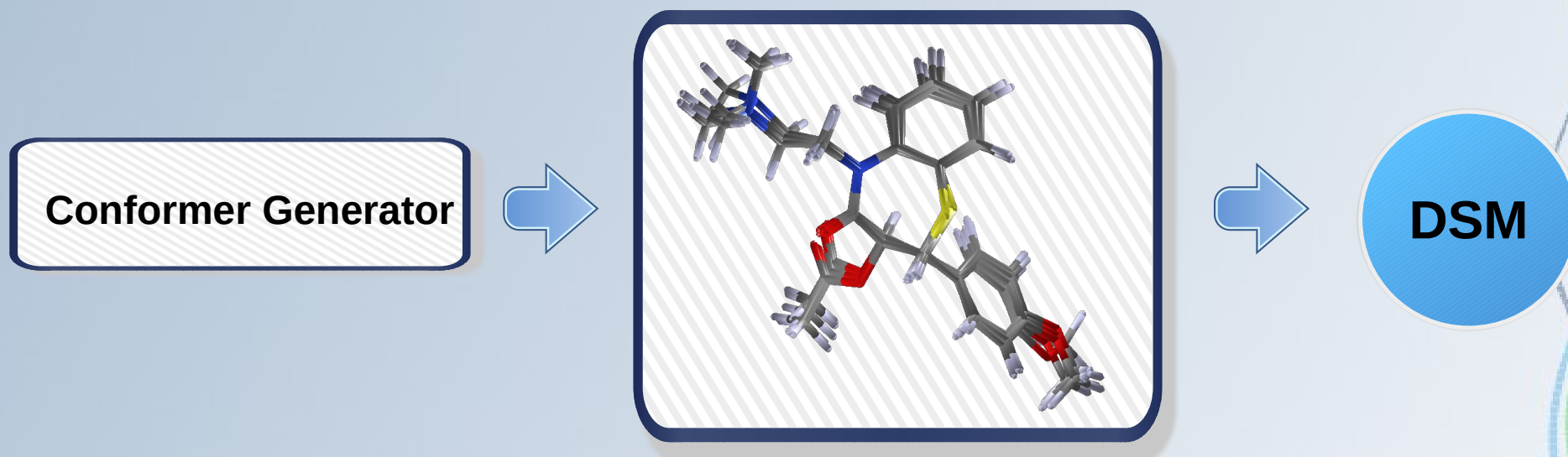
- Pay attention to non planar ring systems or unusual combinations of elements in functional groups. Check for similar molecules in the CSD or in the literature



Structure Solution of Diltiazem Hydrochloride*



Combining conformational analysis with DSM



- BALLOON
- ~~CONFAB~~
- FROG2
- RDKit



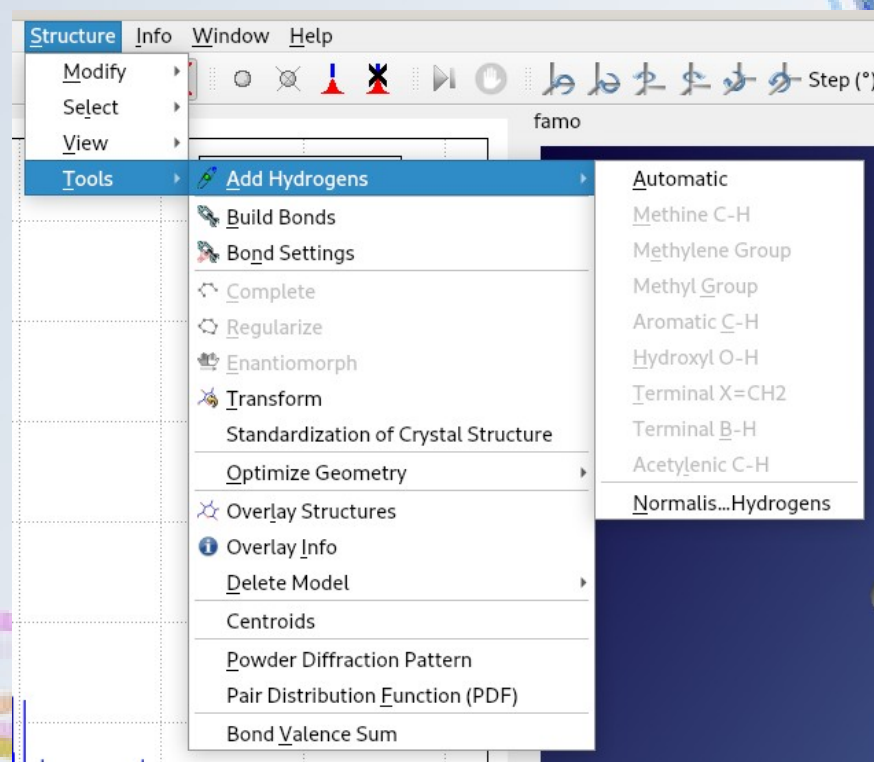
- CORINA
- MOE
- OMEGA

H atoms

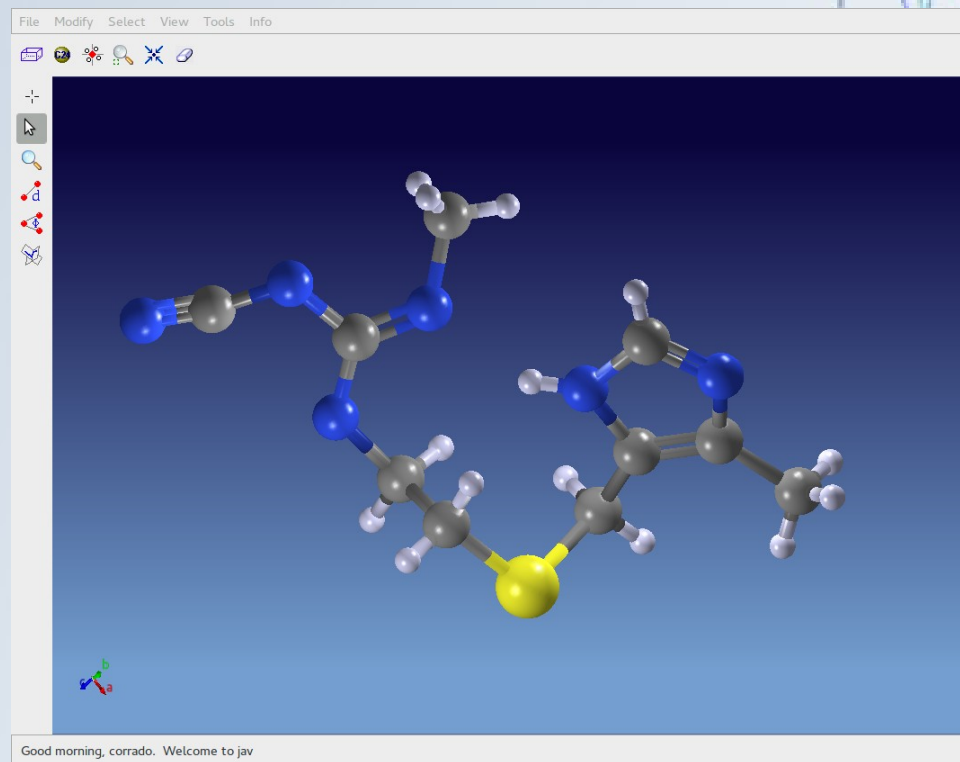
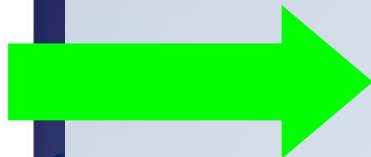
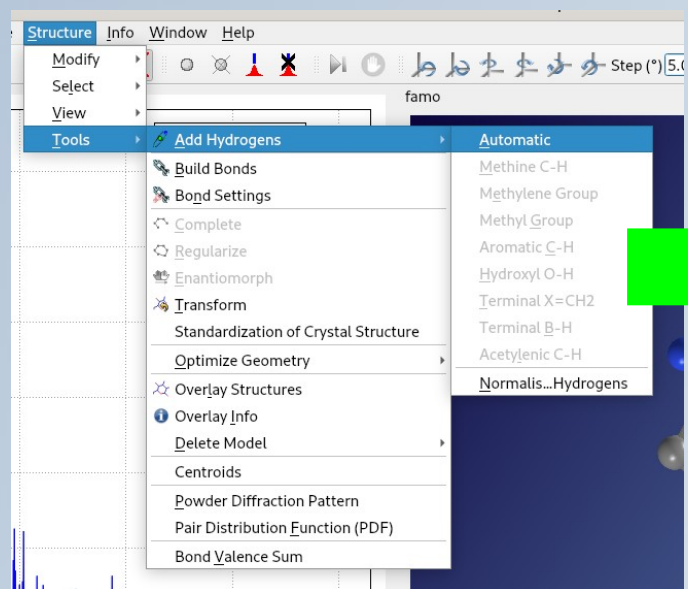
H atoms do not contribute significantly to X-ray diffraction, they can be ignored during the structure solution

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



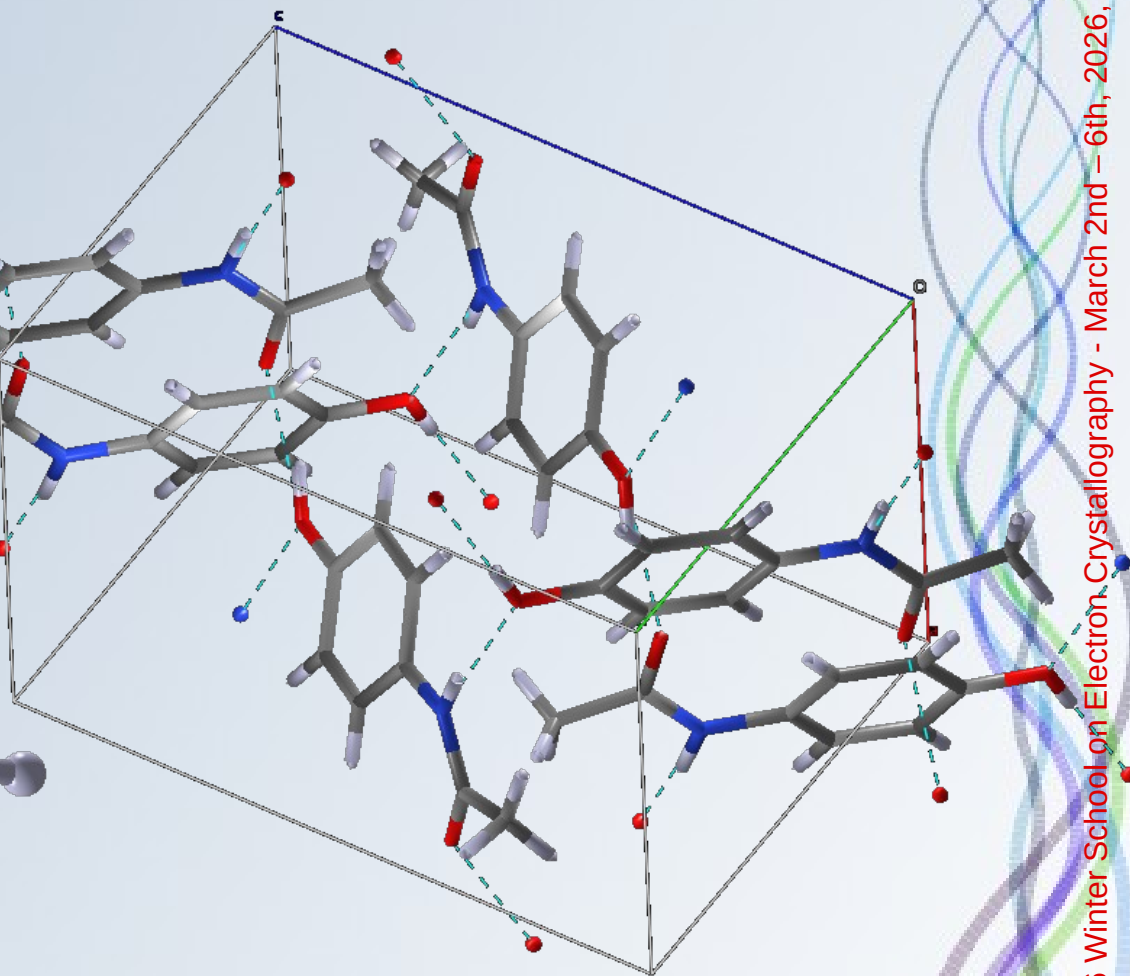
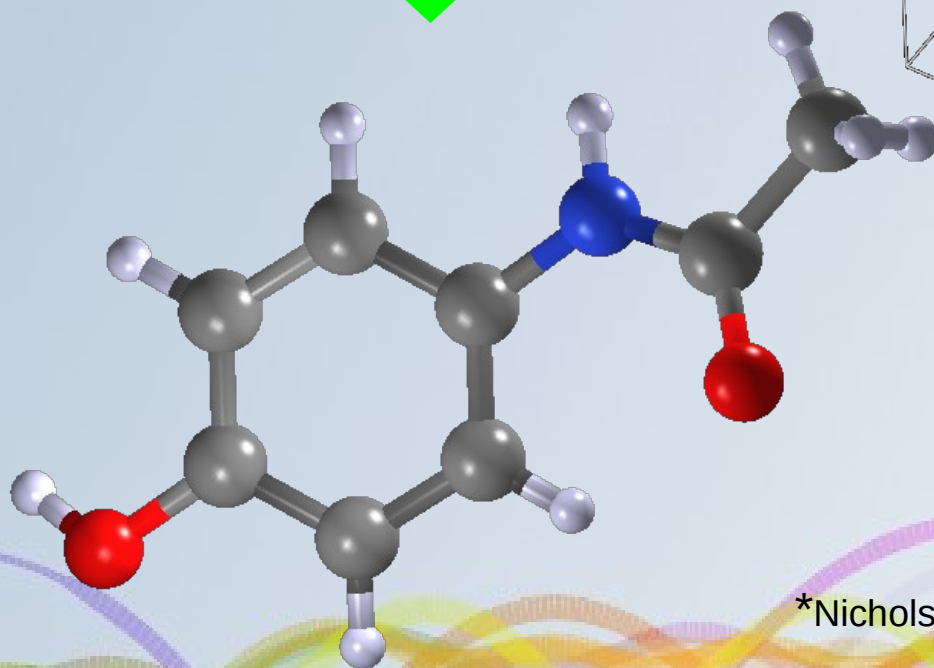
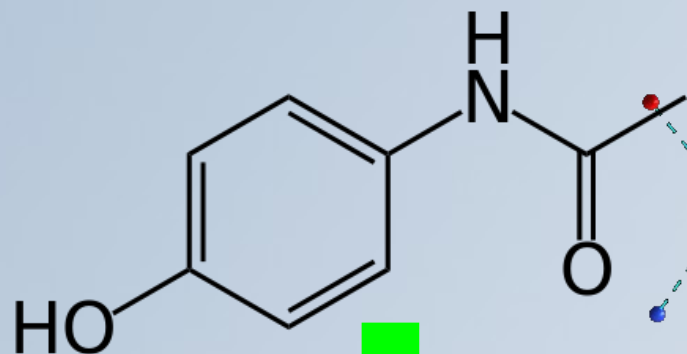
Hydrogen calculation



Molecular compounds

Paracetamol (form I polymorph)

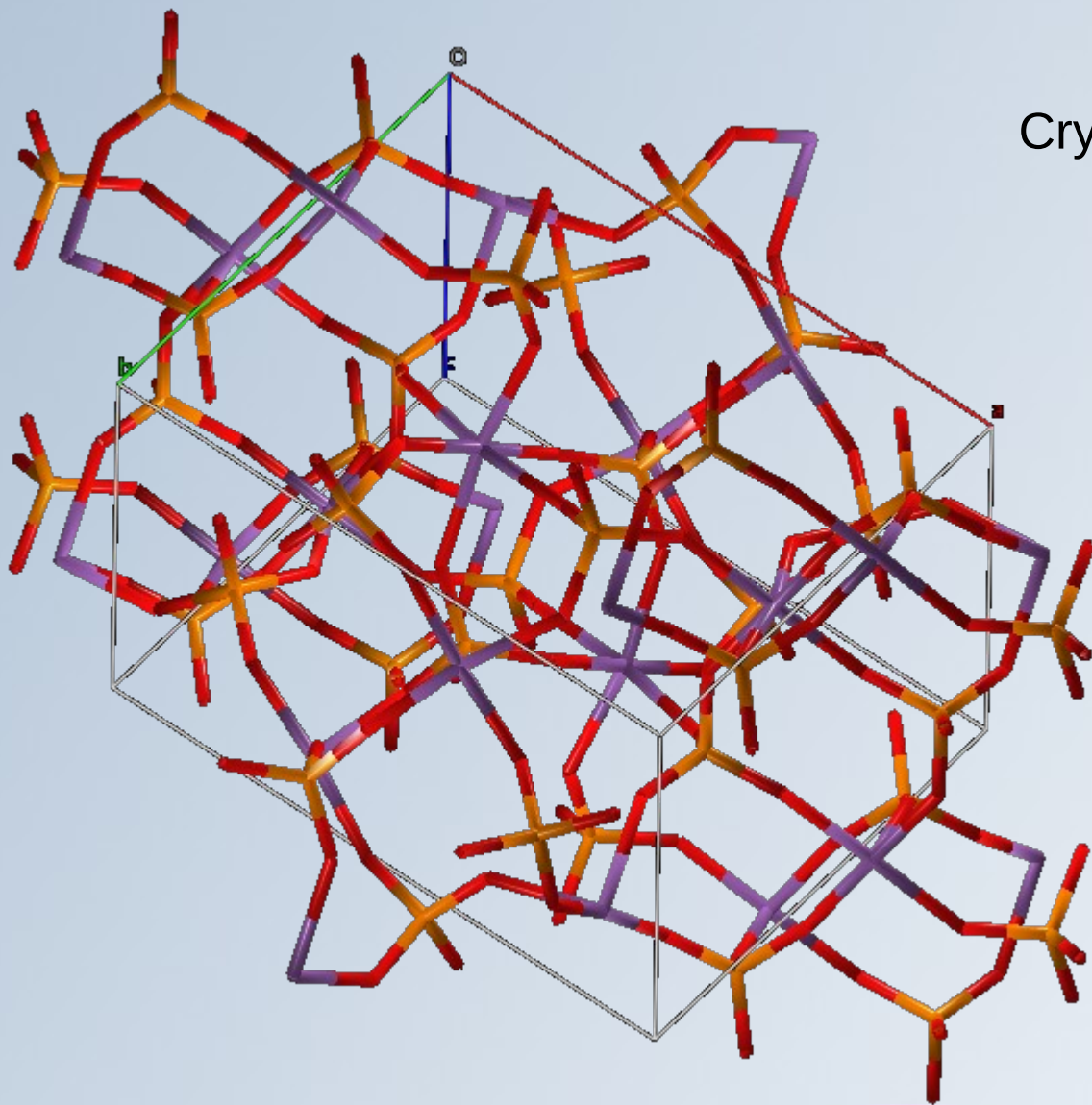
(C₈H₉NO₂) *



*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

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

$$N_{dof} = 51$$

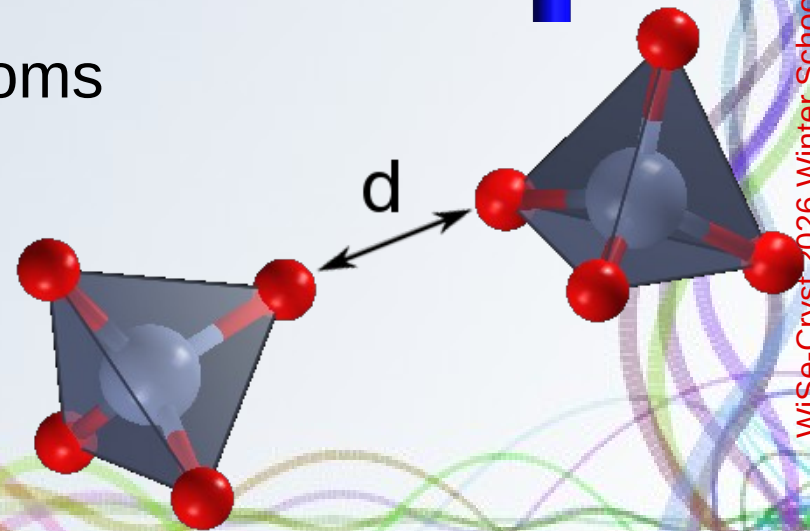
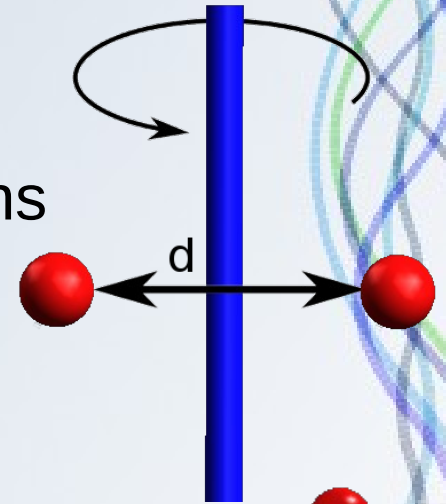


$$N_{dof} = 30$$

$$N_{dof} = 24$$

Non-molecular compounds

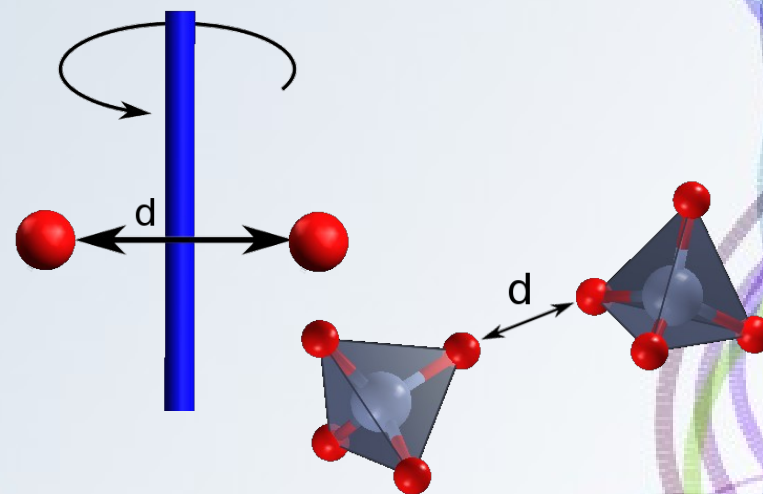
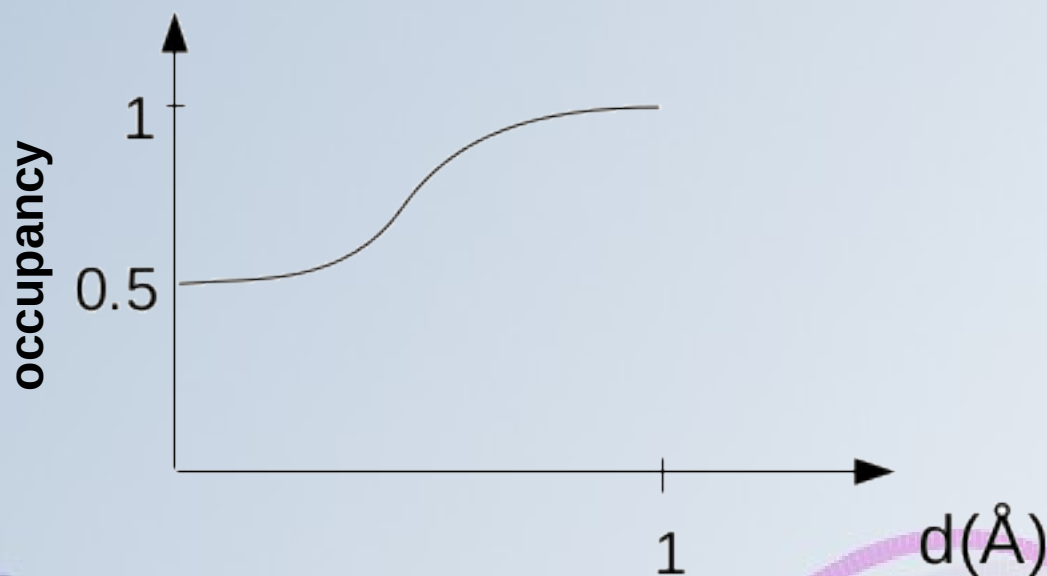
- You cannot know the number and the type of the polyhedra
- Some atoms are expected to fall on special positions
- Different building blocks share some atoms



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

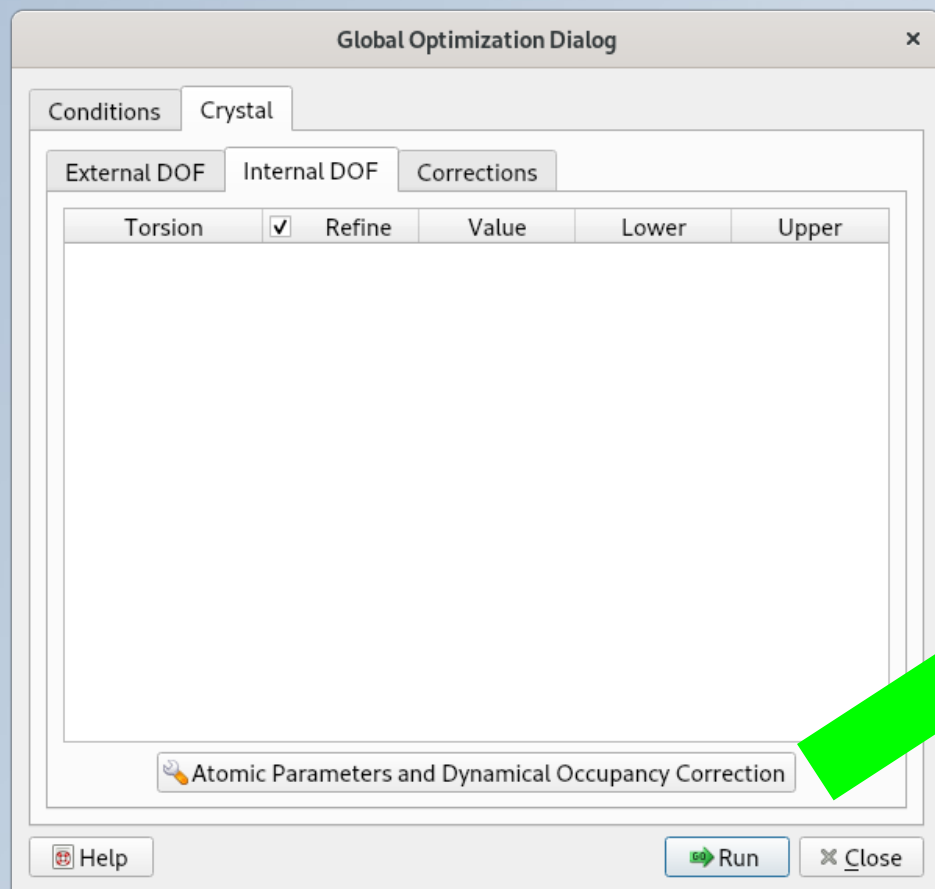
or

doc atom1 atom2 ...

doc Ni1



Dynamical occupancy correction (DOC)



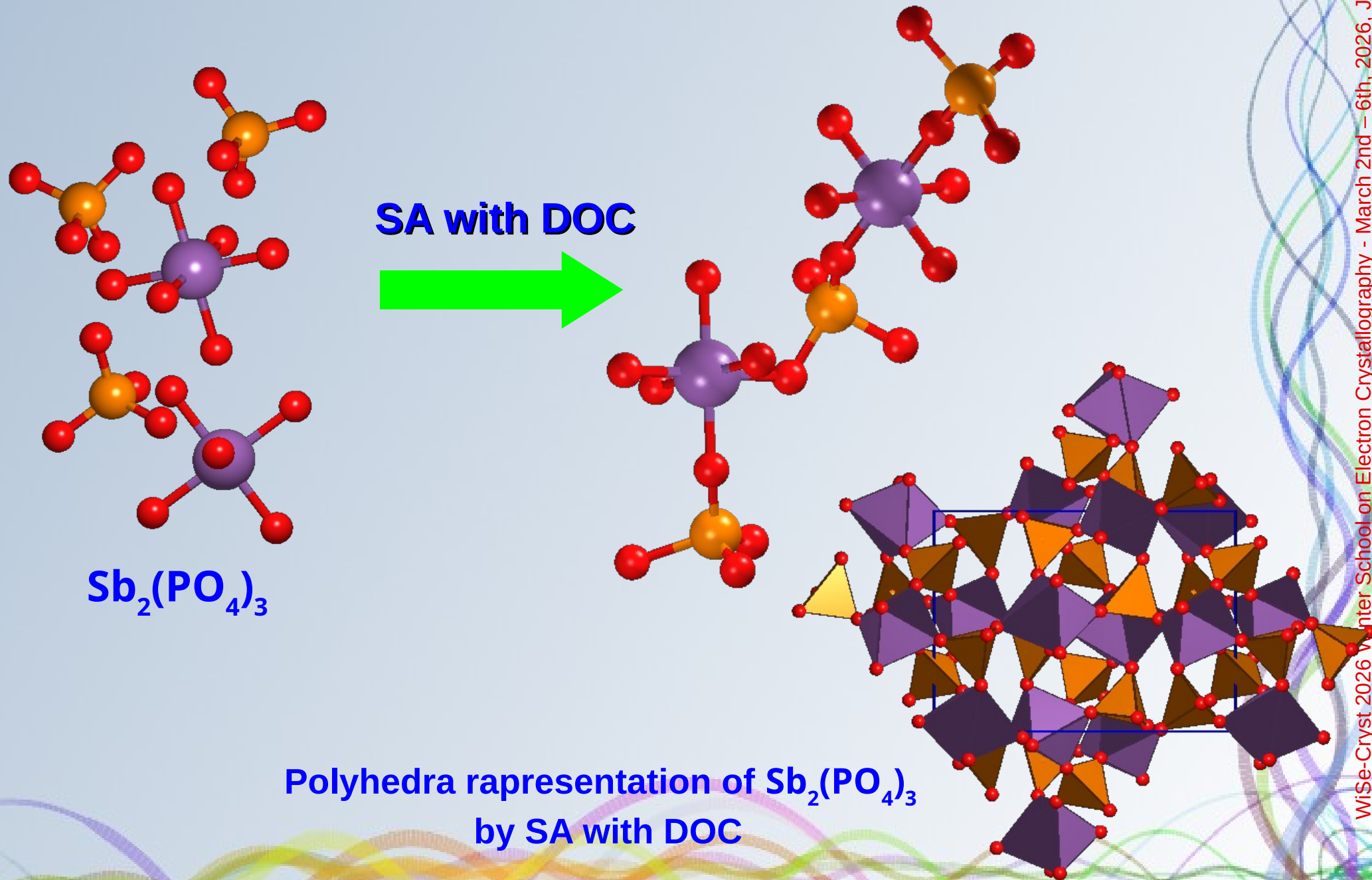
Atomic Parameters and Refinement

Atom	Occ.	D.O.C.	B[iso]	<input type="checkbox"/> Refine B[iso]
P1	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O1	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O2	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O3	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O4	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
P2	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O5	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O6	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O7	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O8	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
P3	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O9	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O10	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O11	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O12	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
Sb1	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O13	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O14	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O15	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O16	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O17	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O18	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
Sb2	1.00000	<input checked="" type="checkbox"/>	1.00000	<input type="checkbox"/>
O19	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O20	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O21	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O22	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O23	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>
O24	1.00000	<input checked="" type="checkbox"/>	3.00000	<input type="checkbox"/>

Cancel OK

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



Cost Functions

$$CF = \sum_{\mathbf{h}} w_{\mathbf{h}} (Y_{\text{obs}}(\mathbf{h}) - kY_{\text{calc}}(\mathbf{h}))^2$$

$$Y_{\text{obs}} = |F_{\text{obs}}| \quad \text{or} \quad Y_{\text{obs}} = F_{\text{obs}}^2$$

- **Other cost functions:**

$CF_{\text{geometry restraints}}$, $CF_{\text{bond valence}}$, $CF_{\text{antibumping}}$

Structure solution of carpyridine derivative

%Structure uzh

%Job Sir2021_test 14/02/2023

%Data

Cell 19.67210 17.33140 44.59000 90.0000 90.2613

space P21/c

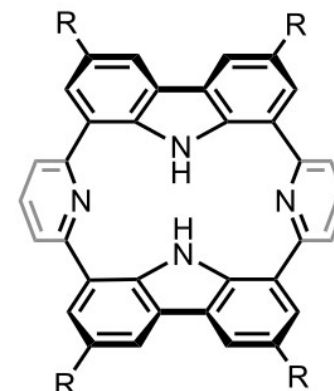
Formula C68H90N4

Electrons

Reflections ../../data/UZH_C5_0m.hkl

Format (3i4,2f8.2)

Fosq



R = -C5

%fragment ../../model/model.mol

%fragment ../../model/model.mol

%fragment ../../model/model.mol

delethydro

%sanneal

nrun 100

niter 2000

%end

DOF: 18 + 48 = 66

Time: 7.3 days on 50 CPU-cores (2.90 GHz)

Reference

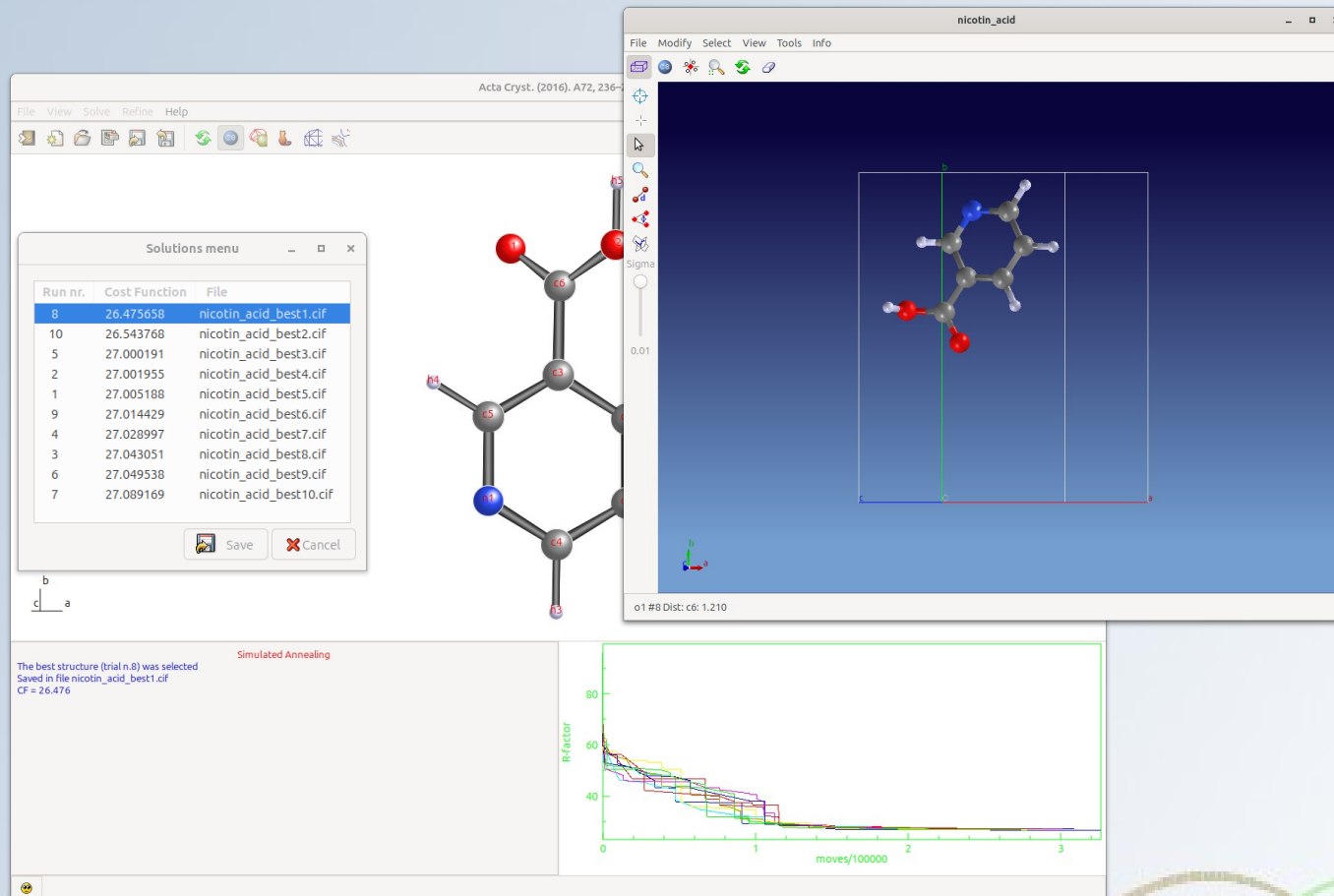
Saddles as rotational locks within shape-assisted self-assembled nanosheets

Joseph F. Woods, Lucía Gallego, Amira Maisch, Dominik Renggli, Corrado Cuocci, Olivier Blacque, Gunther Steinfeld, Andres Kaech, Bernhard Spingler, Andreas Vargas Jentzsch & Michel Rickhaus

Nature Communications | (2023)14:4725

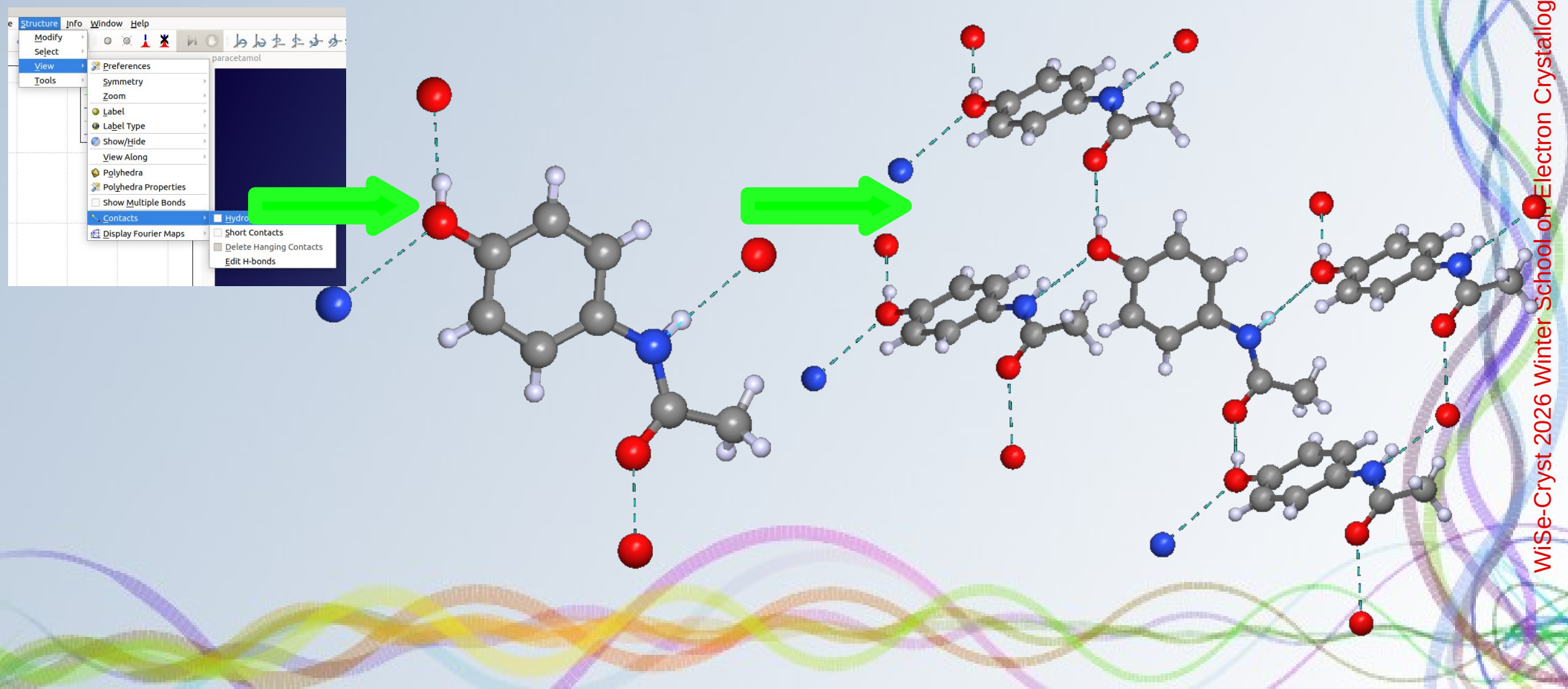
Assessing the solution

- Agreements factors
- Reproducibility of solution



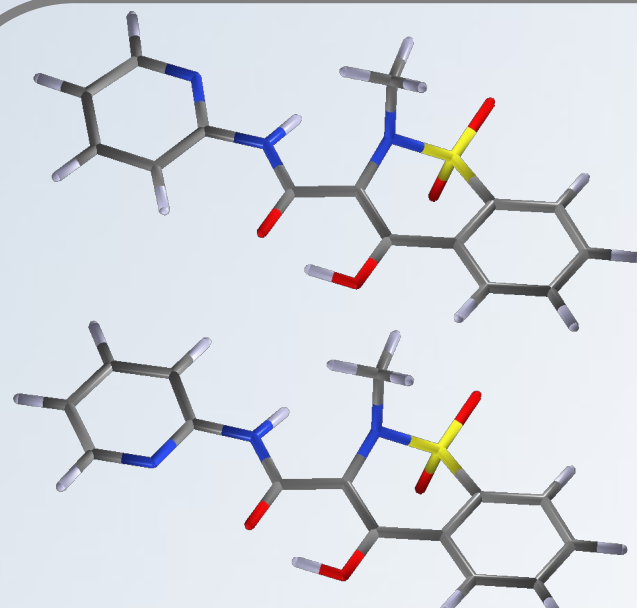
Assessing the solution

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



Combined powder diffraction data and quantum-chemical calculations

- **Optimization of the molecular geometry** to obtain accurate starting models
- **Restraints** in the 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)

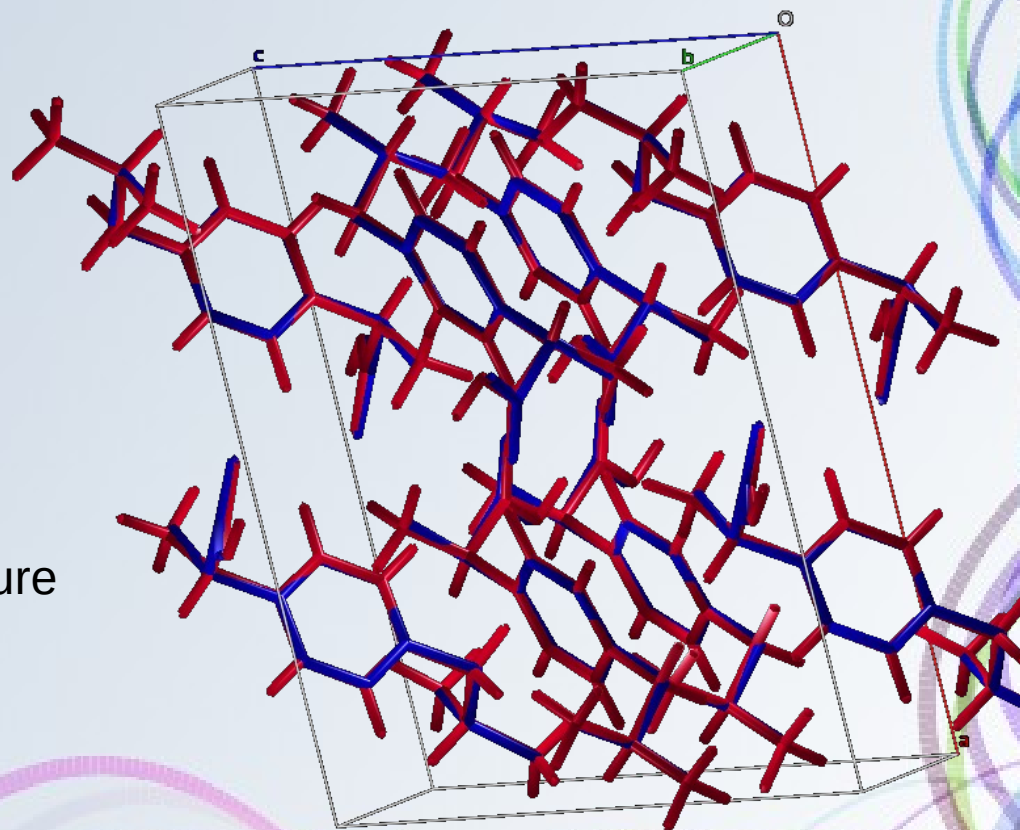
Assessing the solution with DFT-D

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 van 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

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')*
- *Use void analysis software to detect the presence of solvent molecules (e.g., PLATON, Mercury).*

When Structure Solution Fails

- Poor quality diffraction pattern



Solution:

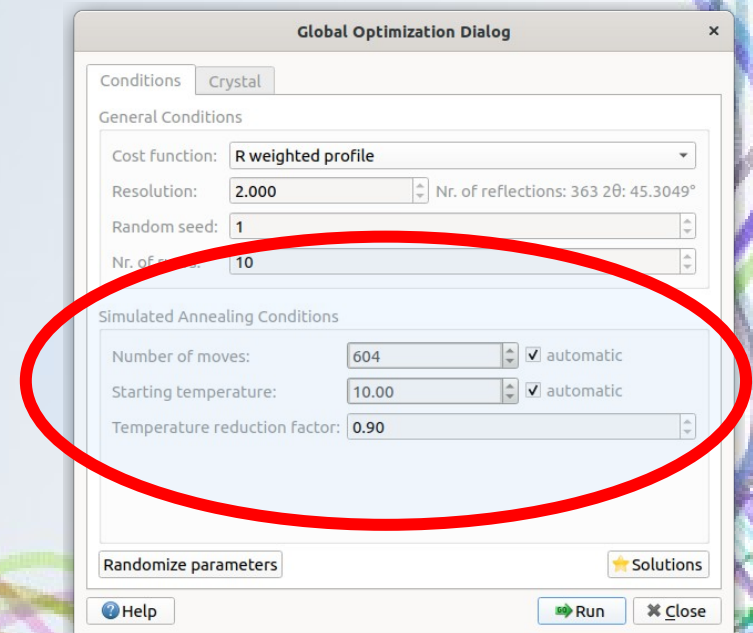
- *Collect new data*
- *Add restraints or anti-bump restraints*

- 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)*



When Structure Solution Fails

- The assumptions about thermal factors are invalid



Solution:

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

- 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*

Thank you for your kind attention