EXPO&more International Workshop

Direct space methods solution

Methods of Structure Solution

Structure solution methods

Traditional approaches:

- direct methods
- Patterson methods

Direct space methods

Alternative words:real space, global optimization, global search

Other methods:

- charge flipping
- molecular replacement

.....

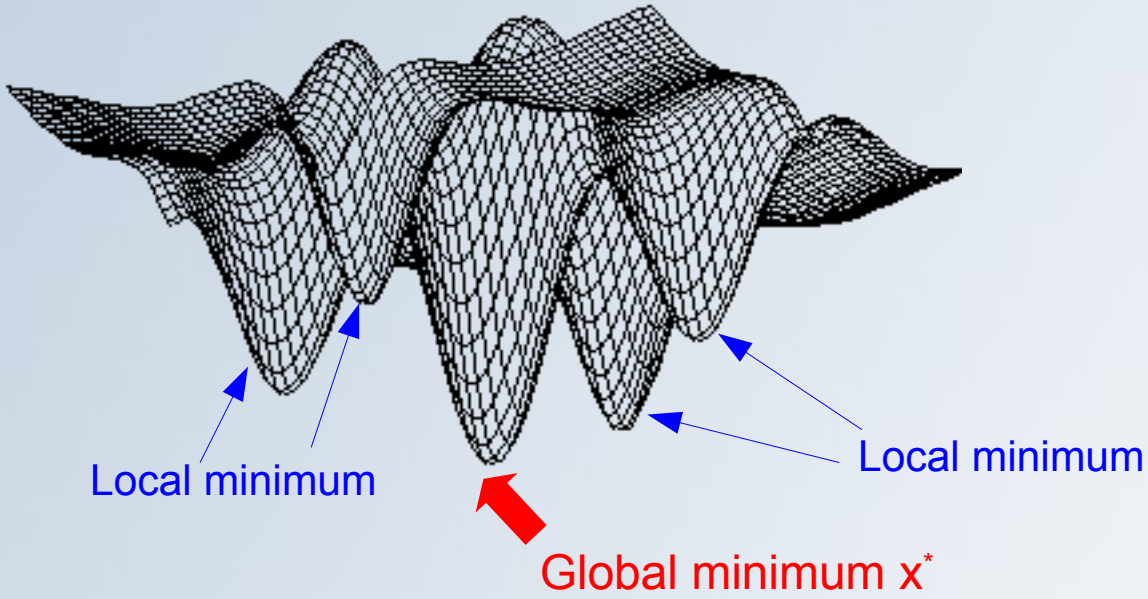
W. I. F. David and K. Shankland. Structure determination from powder diffraction data. Acta Cryst. A64, 52 (2008)
https://doi.org/10.1107/S0108767307064252

Global Optimization Methods

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

X = fractional coordinates of (*x,y,z*) *or*

X = postion (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*
  - *…….*

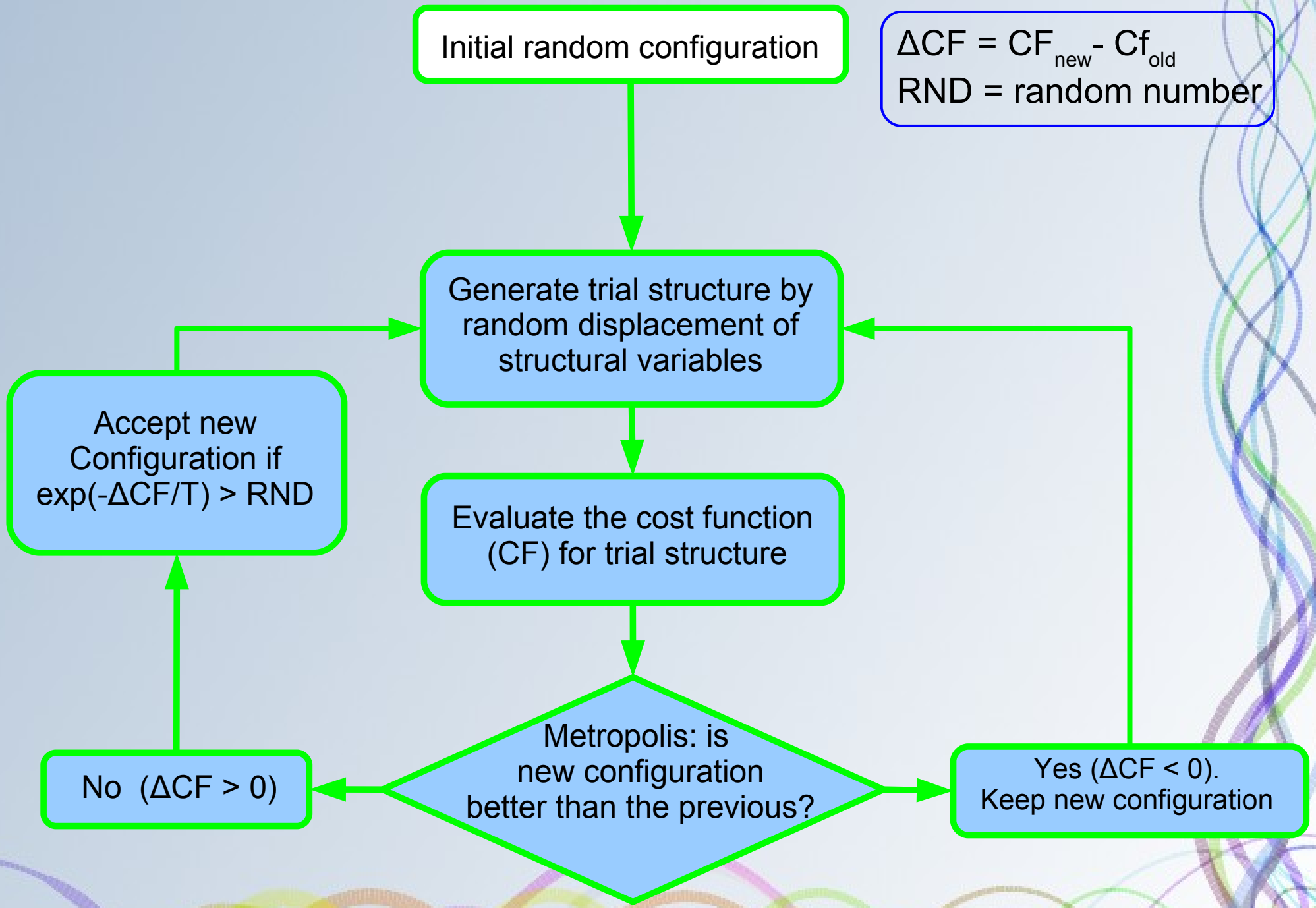
- Heuristic strategies
  - *Genetic Algorithms* (GA)*
  - ***Simulated Annealing (SA)****
  - *Tabu Search*
  - *Ant Colony Optimization*
  - *Particle Swarm Optimization* (PS)*
  - *Bee Algorithms*
  - *Firefly Algorithms*
  - *Harmony Search*
  - *Big Bang-Big Crunch*
  - *……*

Widely used and with the largest impact

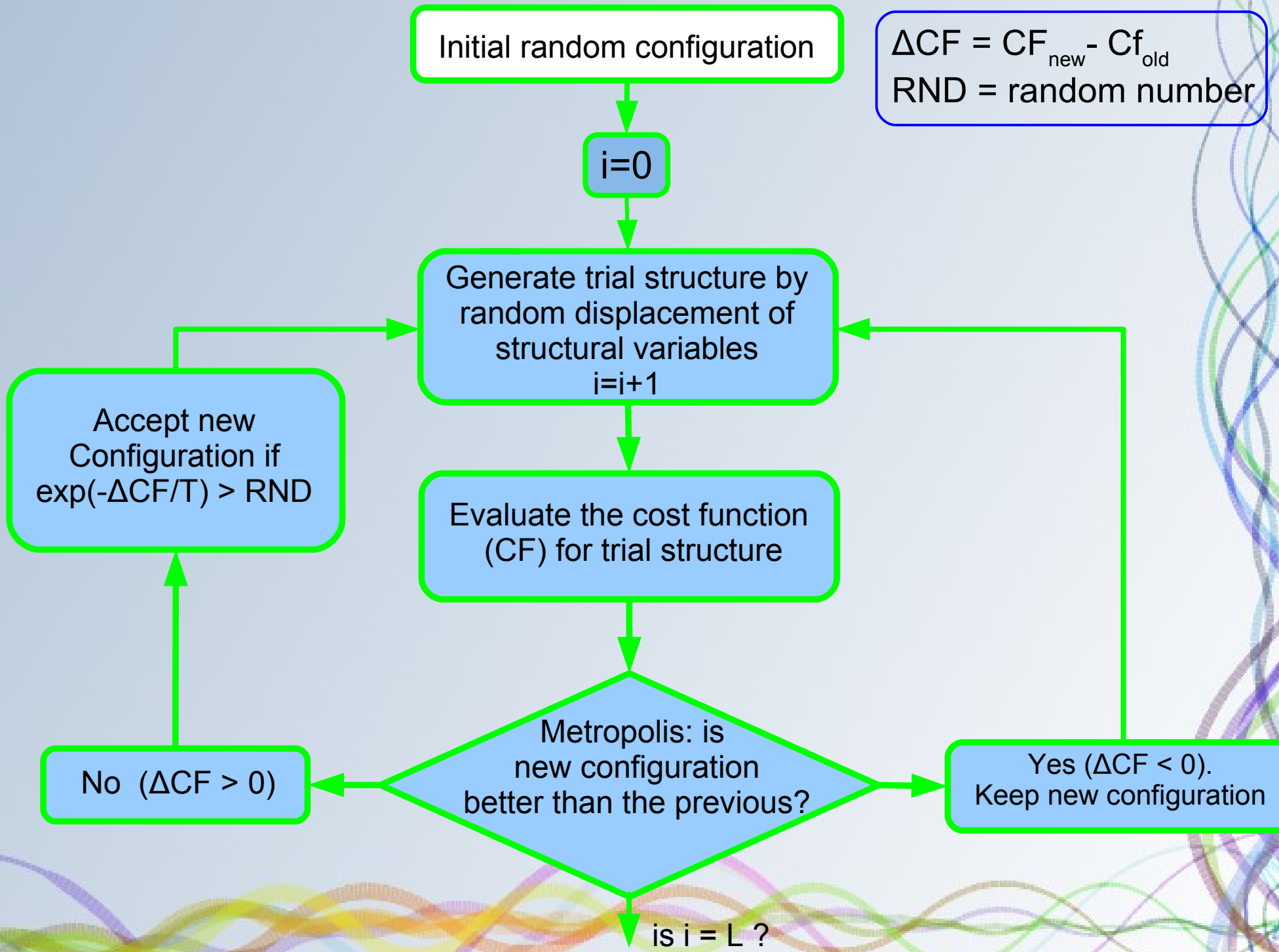
Various modifications:
- *parallel tempering* (PT)
- *adaptive simulated annealing*

(*) employed in solving crystal structure

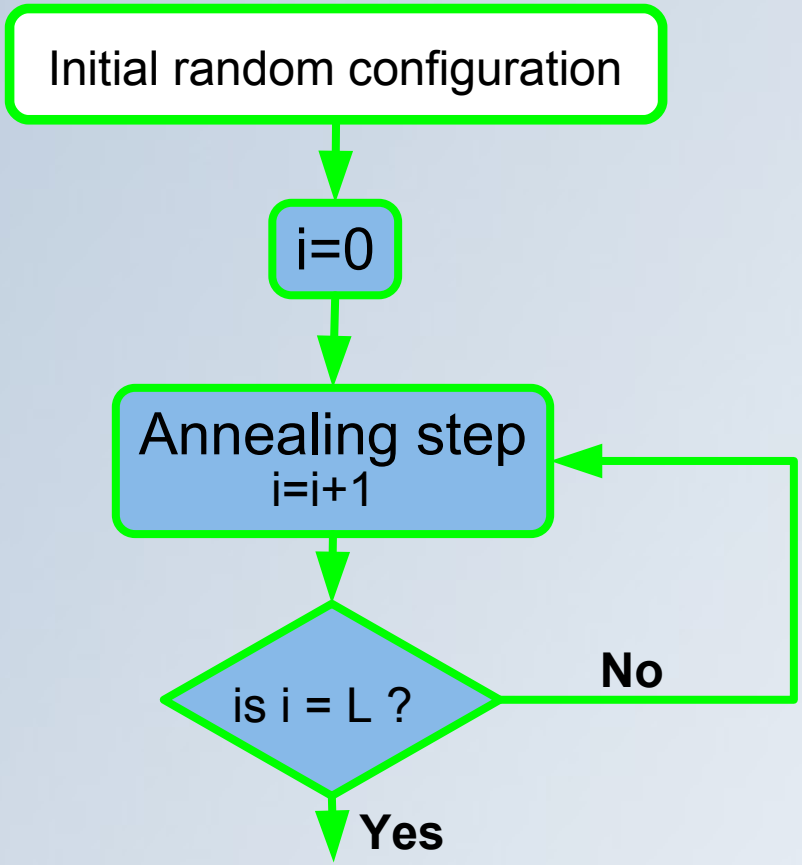
Simulated annealing algorithm: the annealing steps



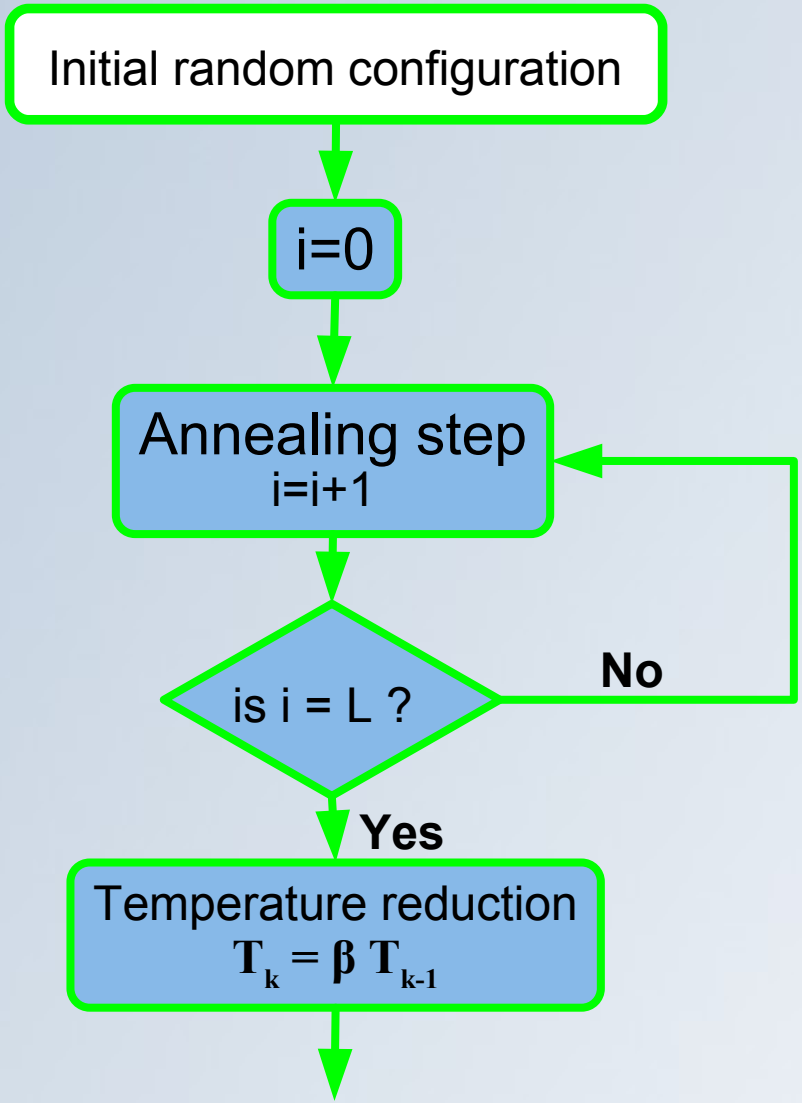
Simulated annealing algorithm: the annealing steps



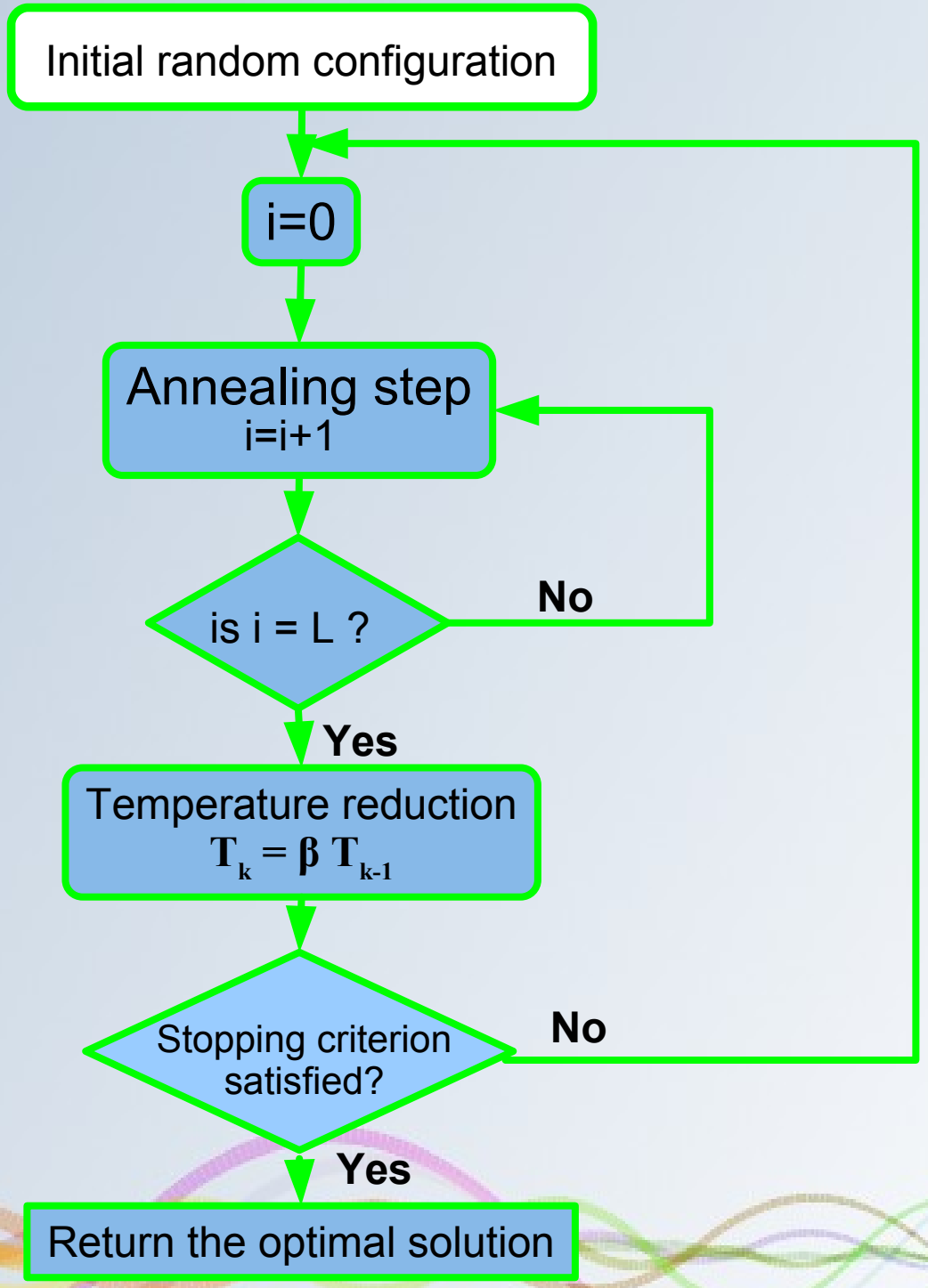
Simulated annealing algorithm: the annealing steps



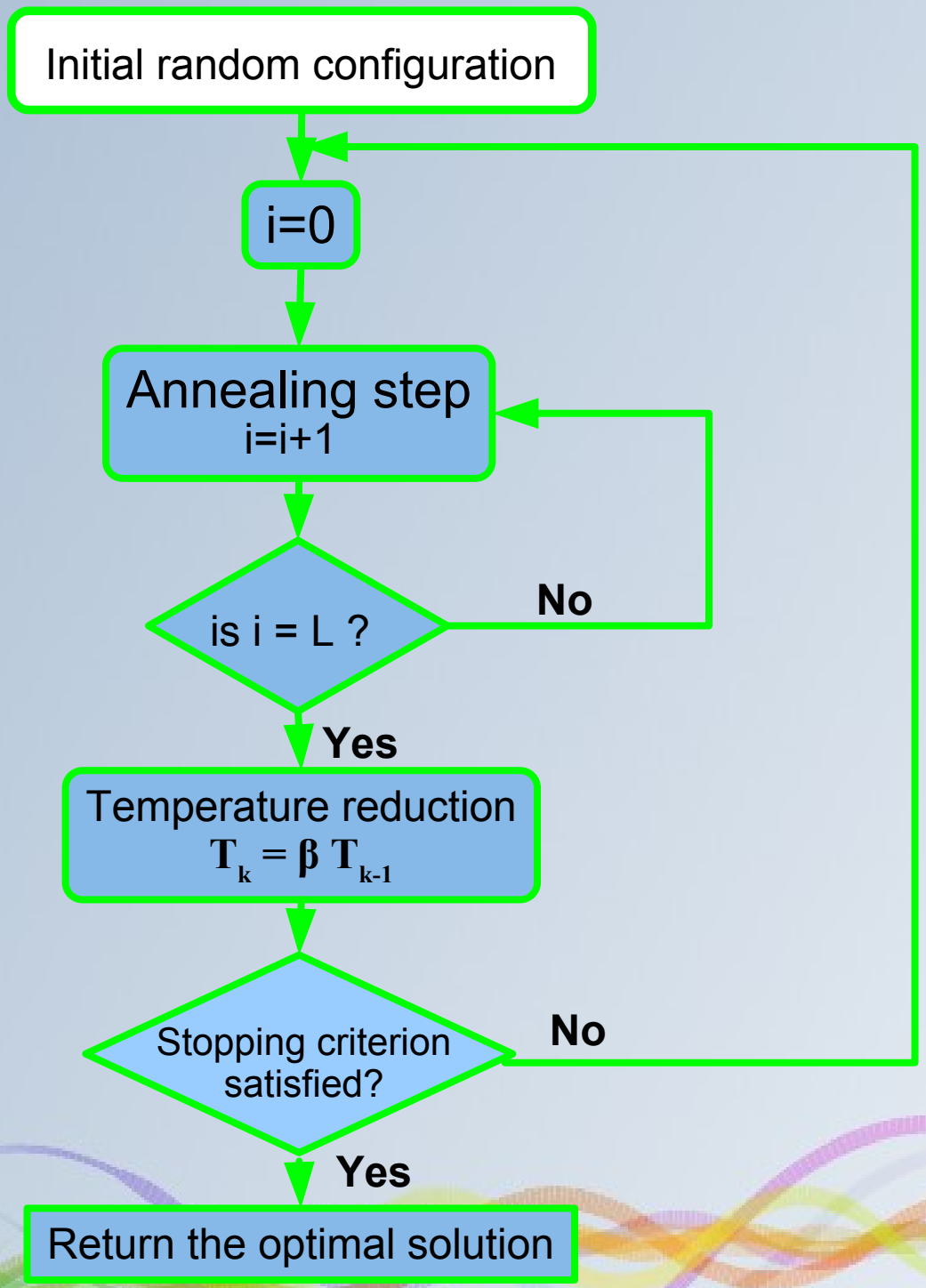
Simulated annealing algorithm

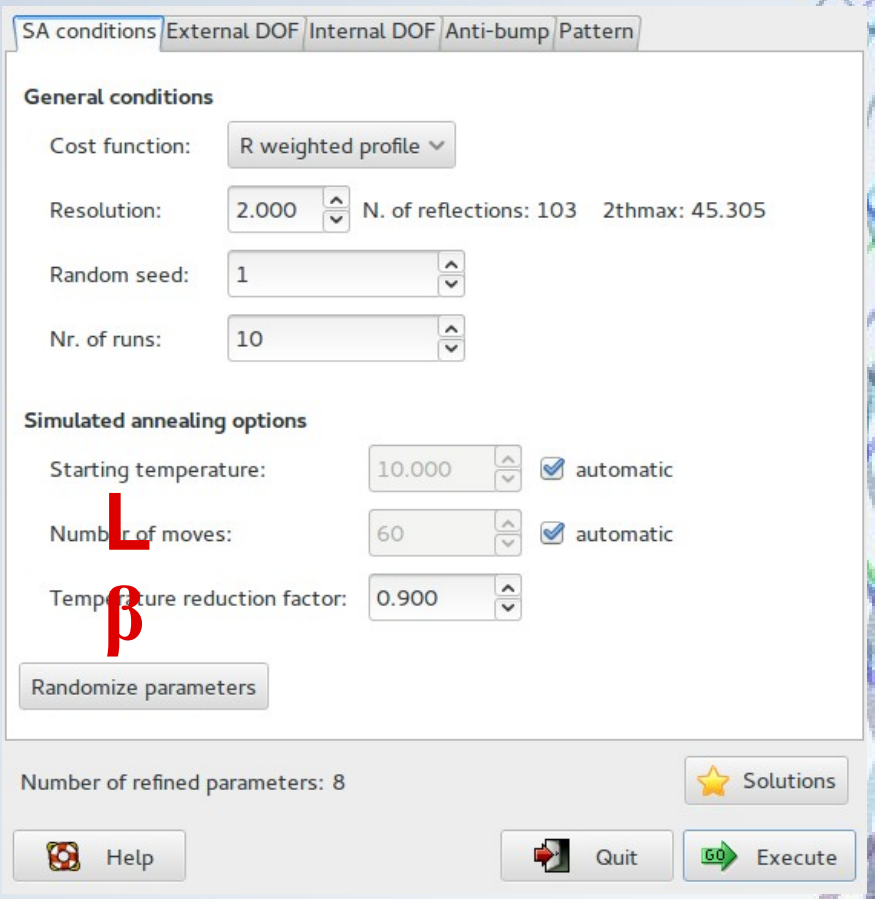


Simulated annealing algorithm



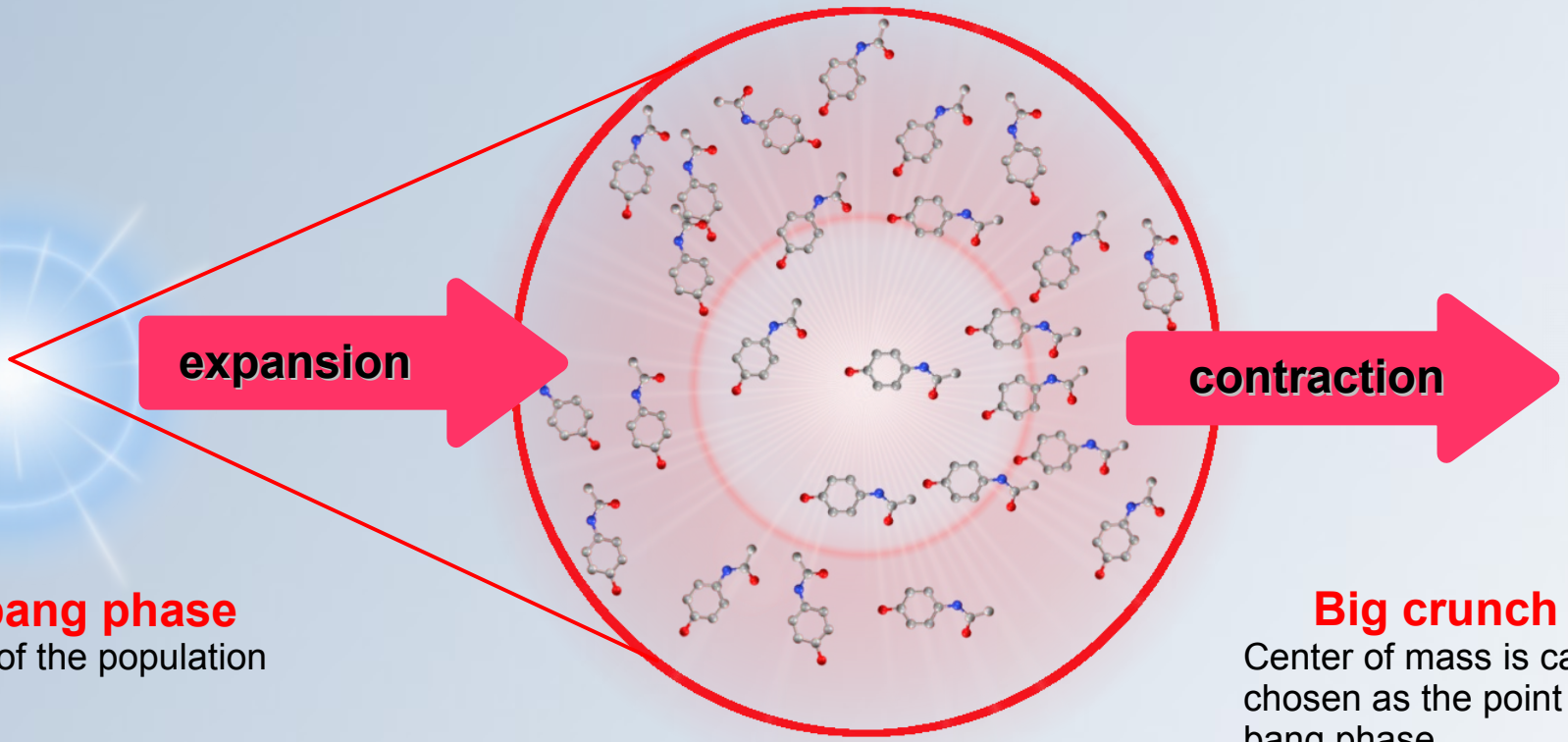
Simulated annealing options





- **Number of moves**
- **Number of SA runs**

Big bang-big crunch optimization



Big bang phase
Creation of the population

Candidate solutions are spread all over the search space in a uniform manner. Cost function values of the individuals are calculated

Big crunch phase
Center of mass is calculated and chosen as the point of new big bang phase

Hybrid version: SA is applied on selected individuals before the big crunch

Comparison

Traditional approaches

-Do not use chemical knowledge

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

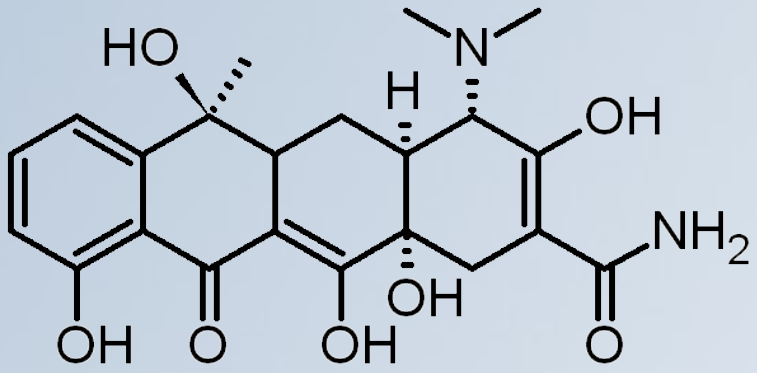
-Take advantage by using data of higher resolution

+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).



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

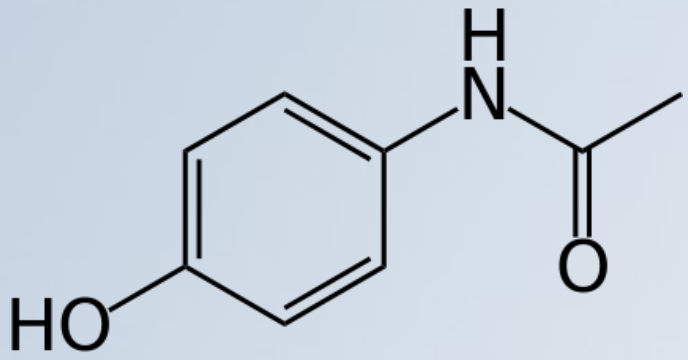
Clegg, W. & Teat, S. J., (2000). Acta Cryst. C56, 1343-1345.

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

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

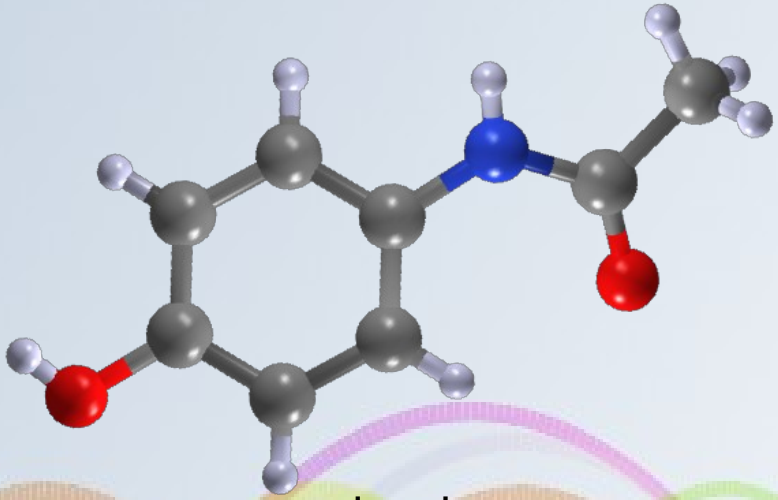
Building starting model

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

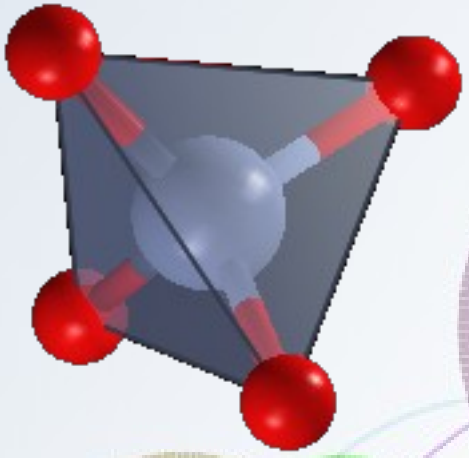


- Crystal structure can be described as a combination of building blocks

atom



molecule



polyhedron

Building starting model

- **Check for similar molecules** in databases or in the literature

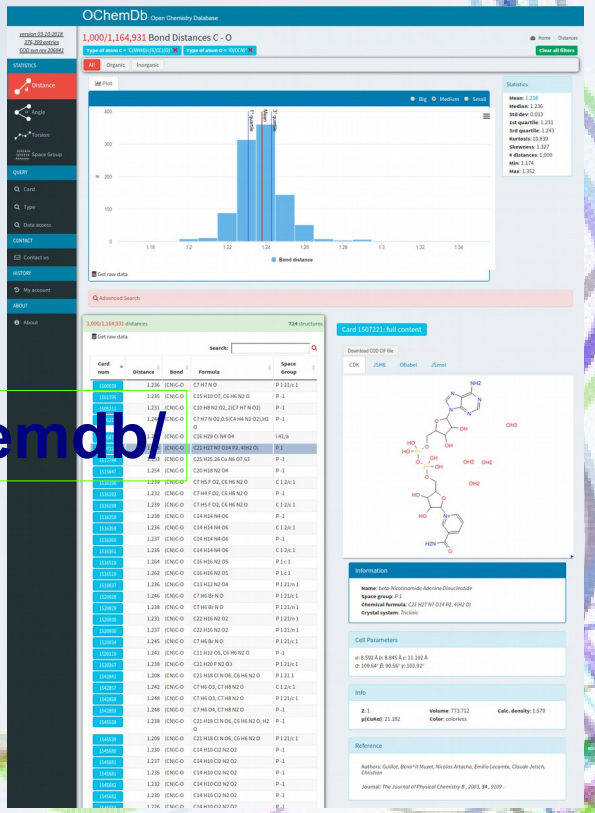
- **Optimize molecular geometry** by computational chemistry programs

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)

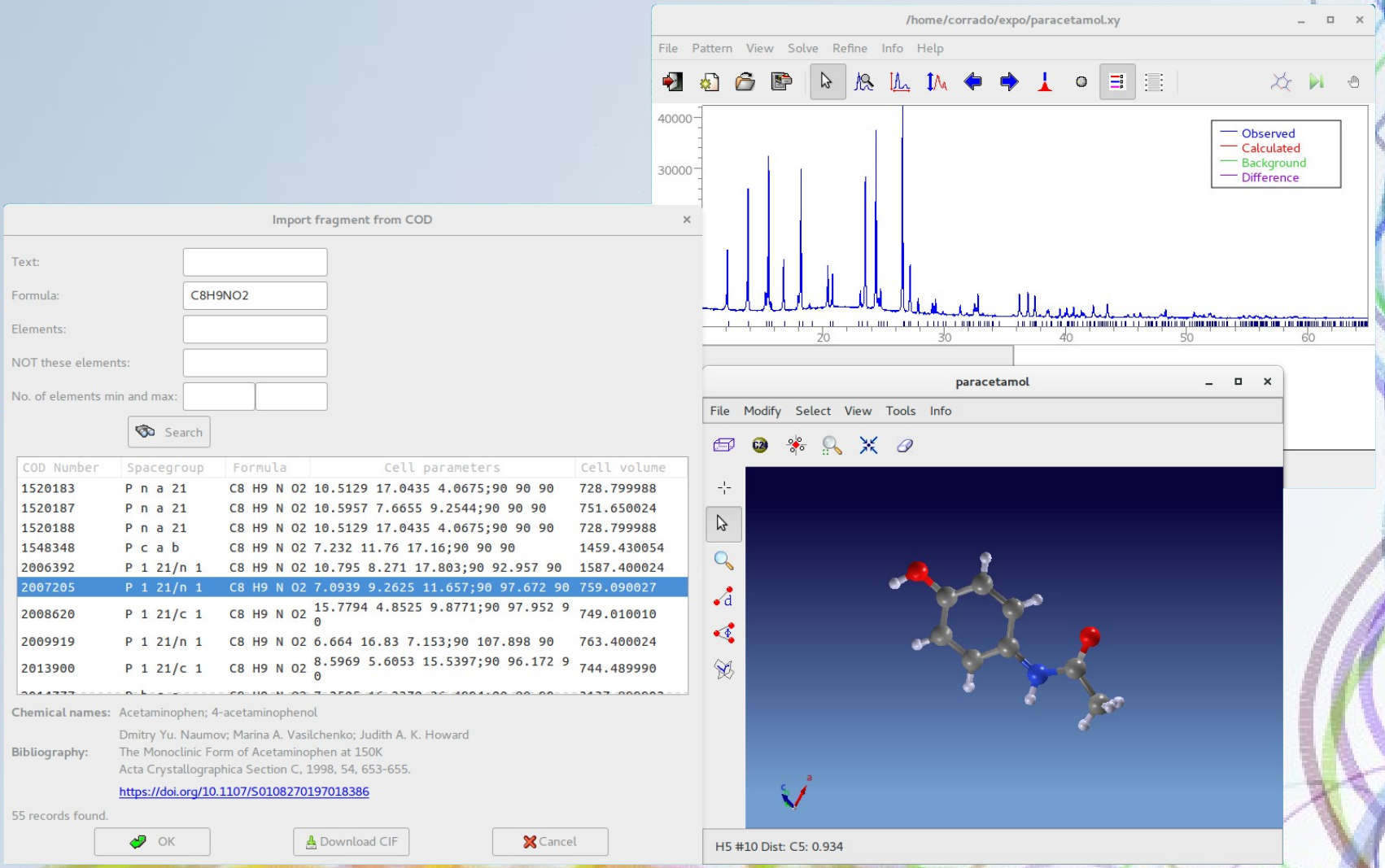
*Joint special issue: *Acta Cryst. B58, 317-422 (2002)*

Load crystal structures from COD

“View” menu → **“Jav Molecular Viewer”**

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

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



Geometry optimization

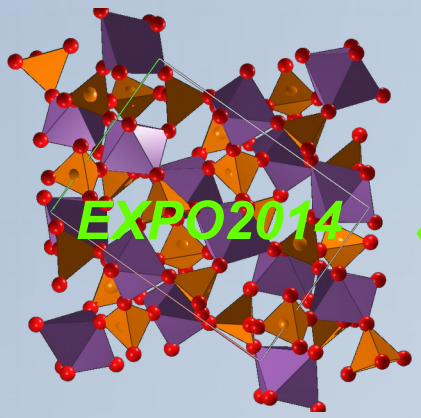
Three levels of theory

- 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



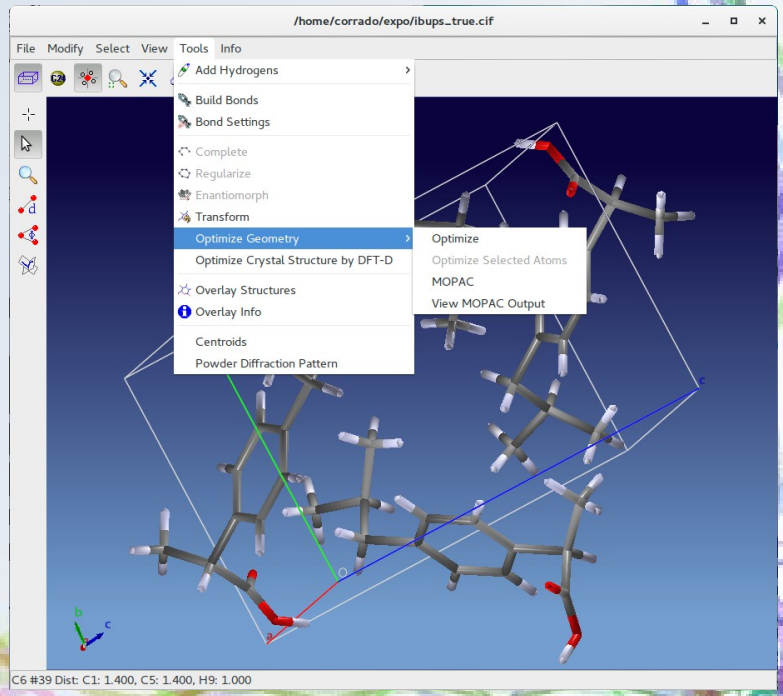
Fotran/C++

obmodule.f90

C++

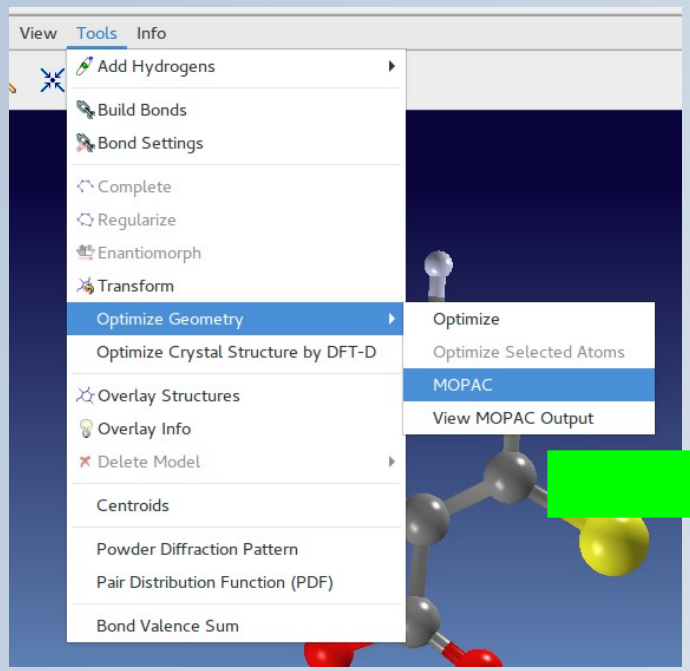
- Molecular-mechanics force fields (MMFF99 and UFF provided by **Open Babel library)**

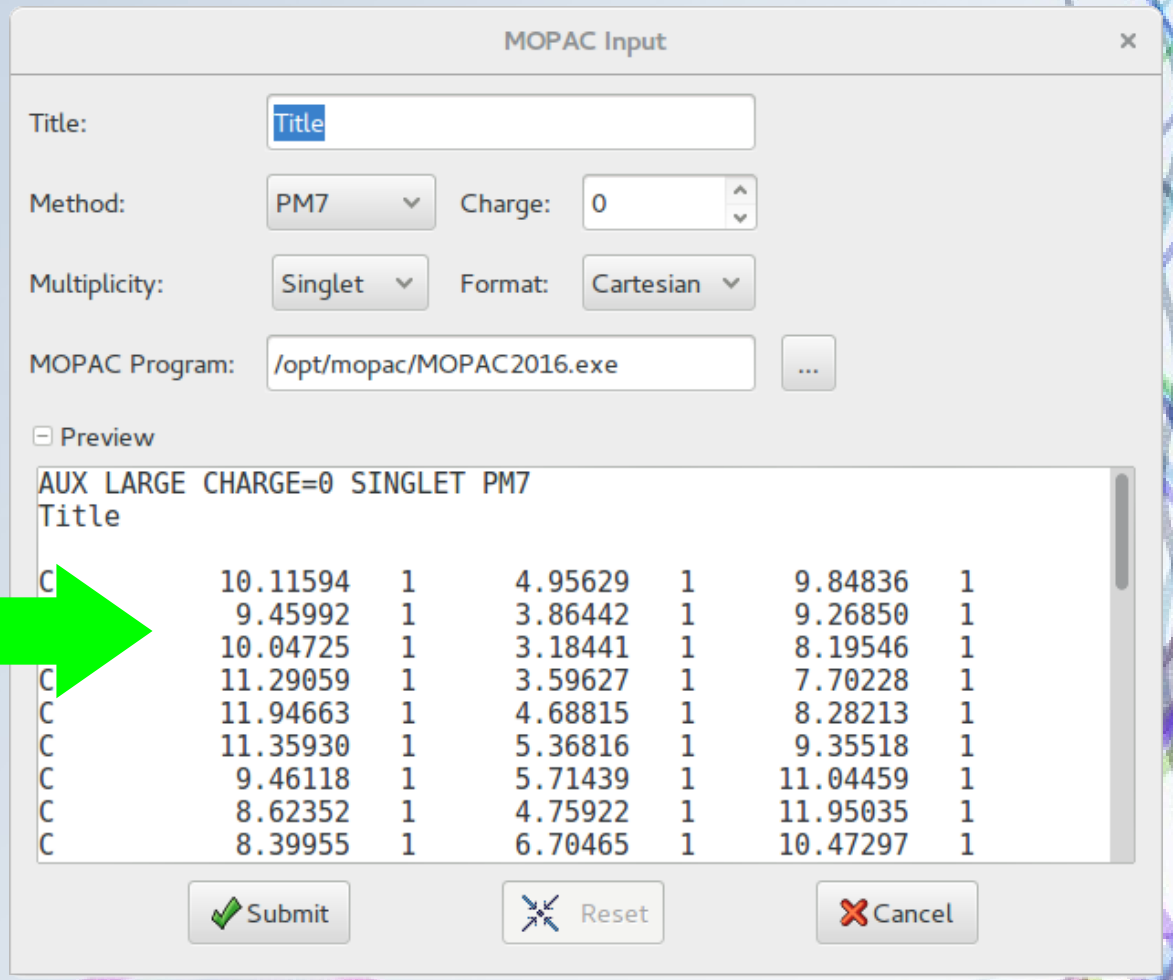
- Able to process input and output files of common quantum-chemistry packages: GAMESS-US, NWChem, Gaussian, CRYSTAL, ABINIT, Quantum ESPRESSO



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





Graphical User Interface for MOPAC2016

Molecule editor

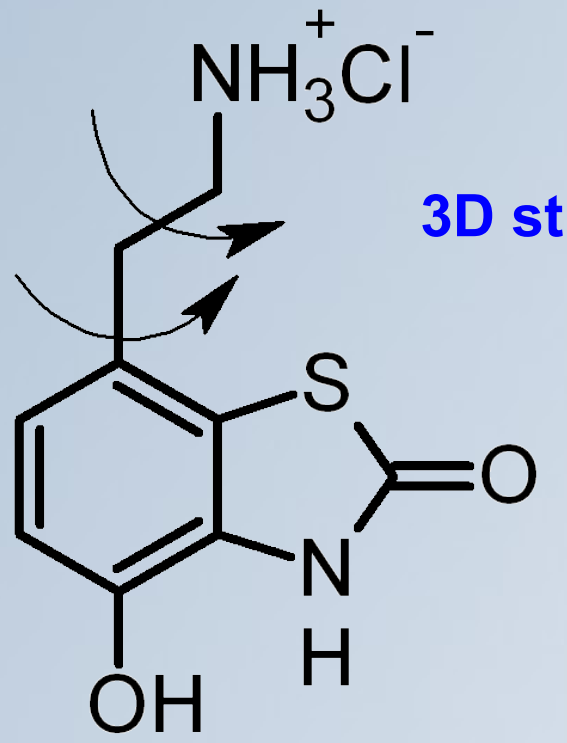
A molecule editor allows

- 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/

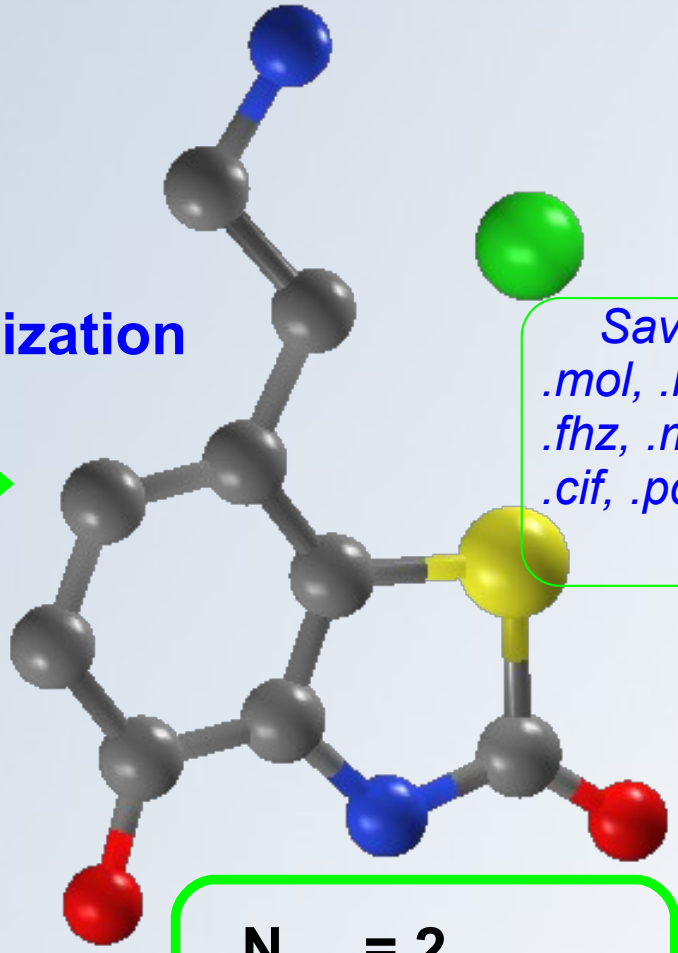
Building starting model



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

P 21/a

3D structure optimization

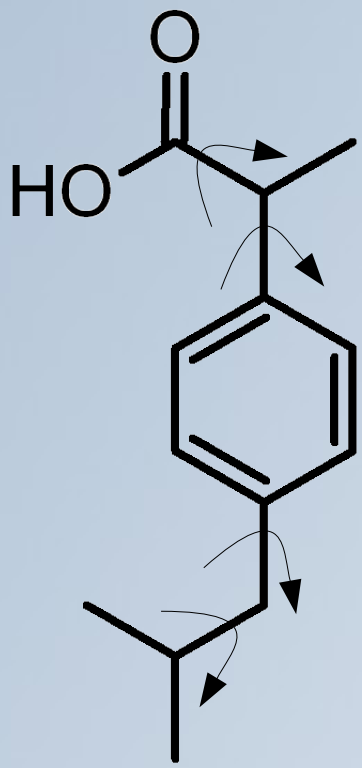


Save as
.mol, .mol2,
.fhz, .mop,
.cif, .pdb, .frac

$N_{frag} = 2$

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

Building starting model: example 2

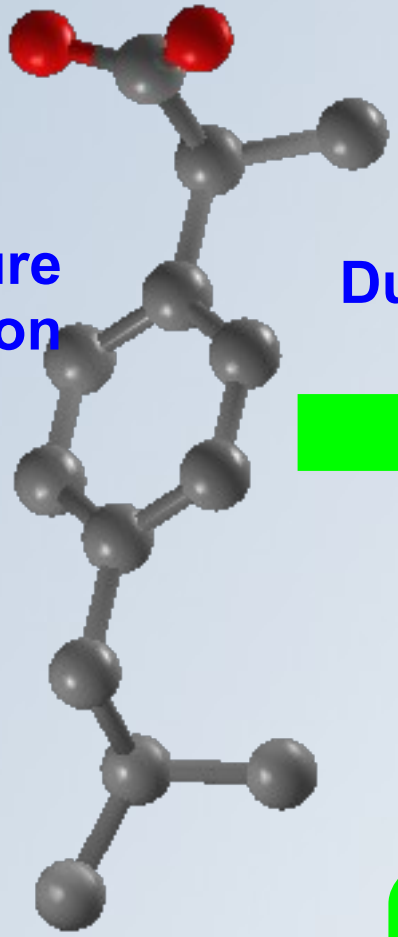


S-Ibuprofen

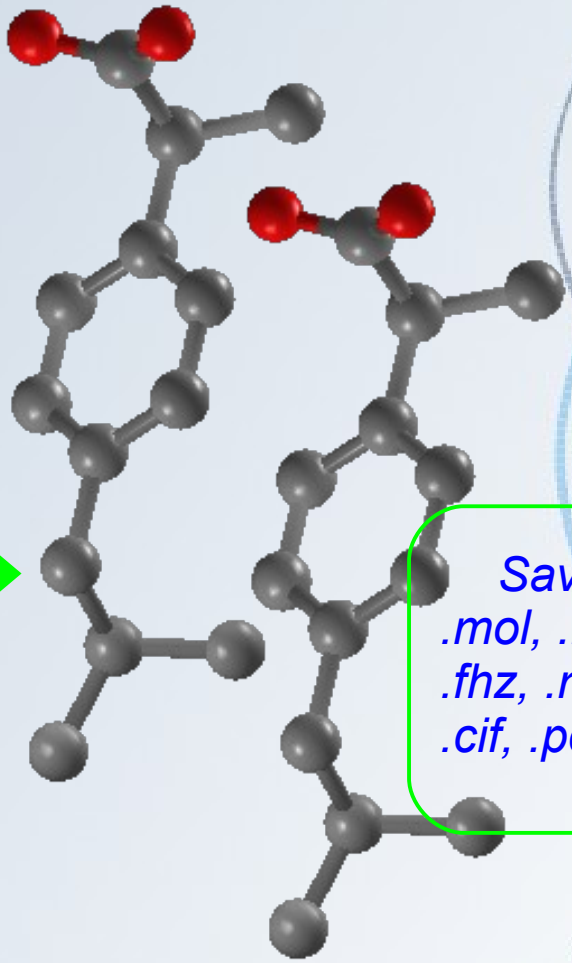
$P2_1$

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

3D structure optimization



Duplicate

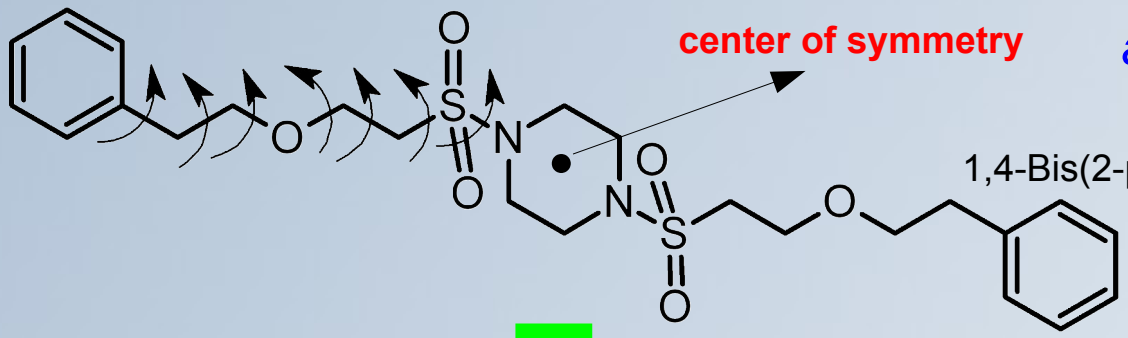


Save as .mol, .mol2, .fhz, .mop, .cif, .pdb, .frac

$N_{frag} = 2$

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

Building starting model: example 3

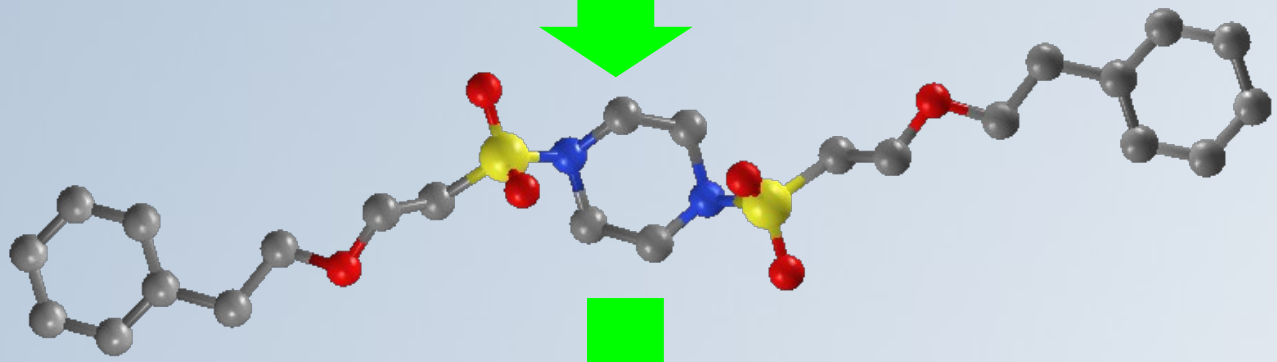


a.u. is a half molecule
Z' = 1/2

1,4-Bis(2-phenethyloxyethanesulfonyl)piperazine

P 21/a

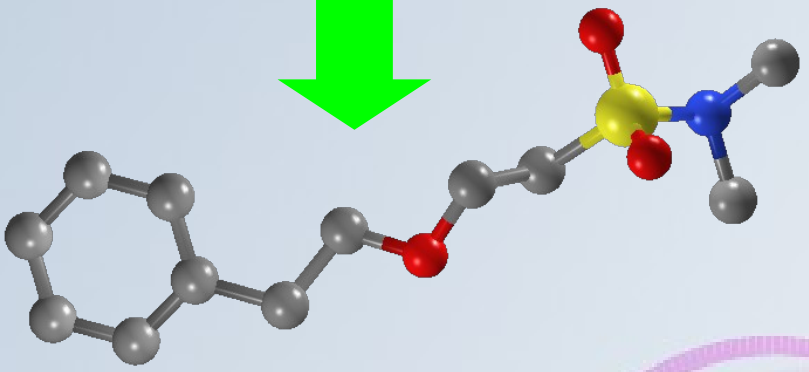
3D structure optimization



Structure solution by DSM + DOC

N_{dof} = 6 + 14

15 times slower



Structure solution by DSM

N_{dof} = 6 + 7

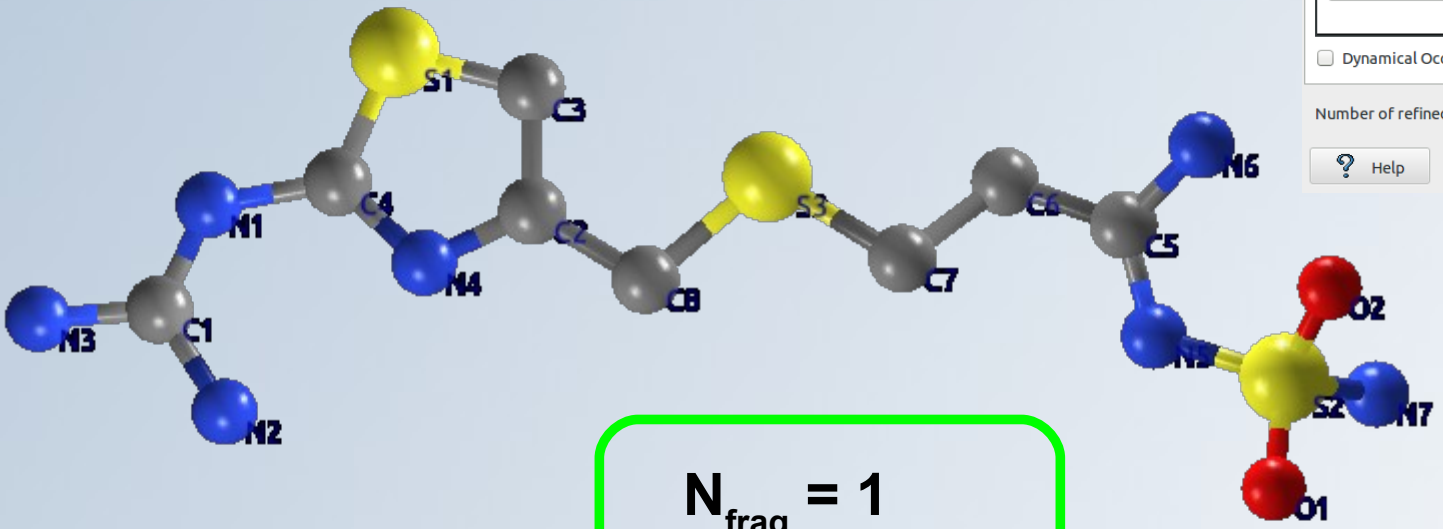
Recommended strategy

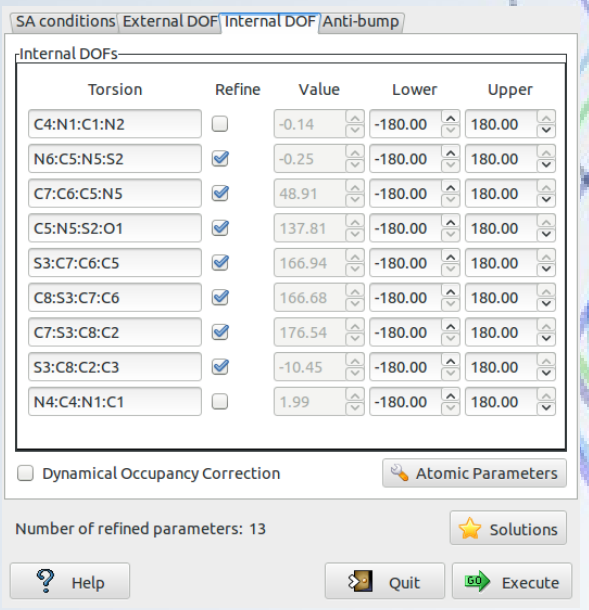
Building starting model: example 4

famotidine form B

P 21/c

3D structure optimization



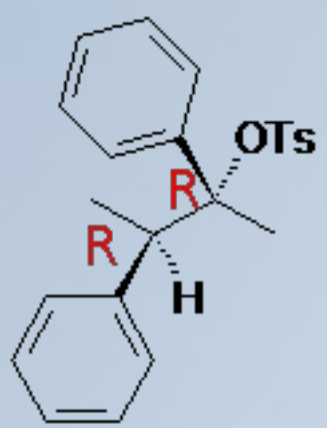


$N_{frag} = 1$

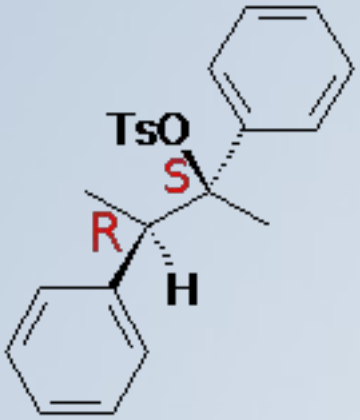
$N_{dof} = 6 + 6$

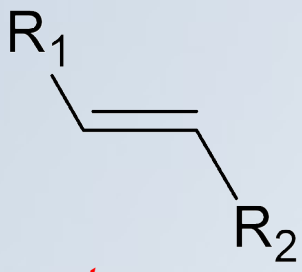
Building starting model

- Stereochemistries will not be altered during simulated annealing.
  Attention to compounds with more than one chiral center and cis/trans isomerism



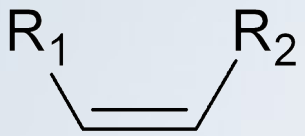
or





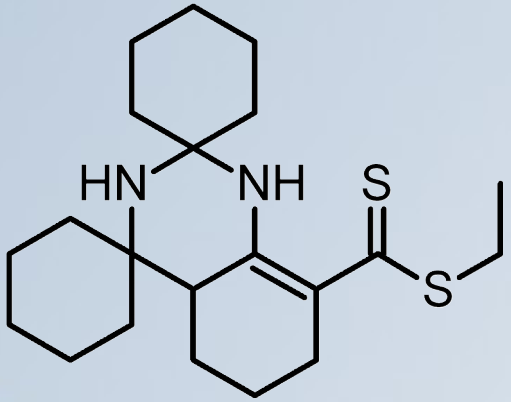
trans

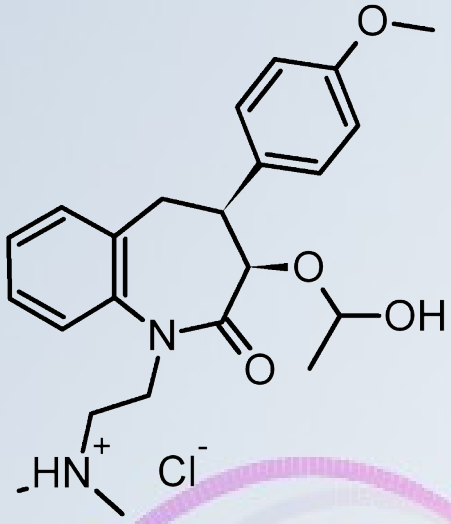
or

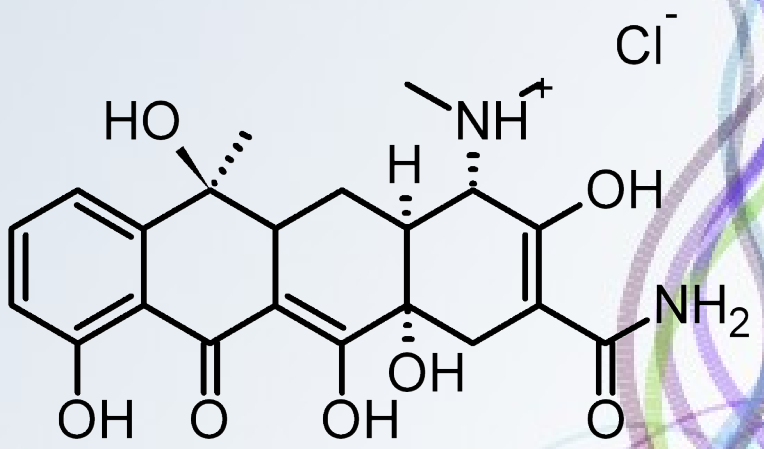


cis

- 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







Building starting model

- Check V_{cell}/V_{mol}

V_{mol} = number of non-hydrogen atoms · 18 (*18 Å³ rule*)

or

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

n_i = number of atoms of the i^{th} type in the structure
v_i = volume contribution (in Å³) for the i^{th} atom type

The volume of the unit cell can be used to determine the number of independent building blocks in the asymmetric unit from the known crystal density

Average volume occupied by atom ≈ 15 – 20 Å³

The success of the structure determination depends crucially on the accuracy of the input molecular model.

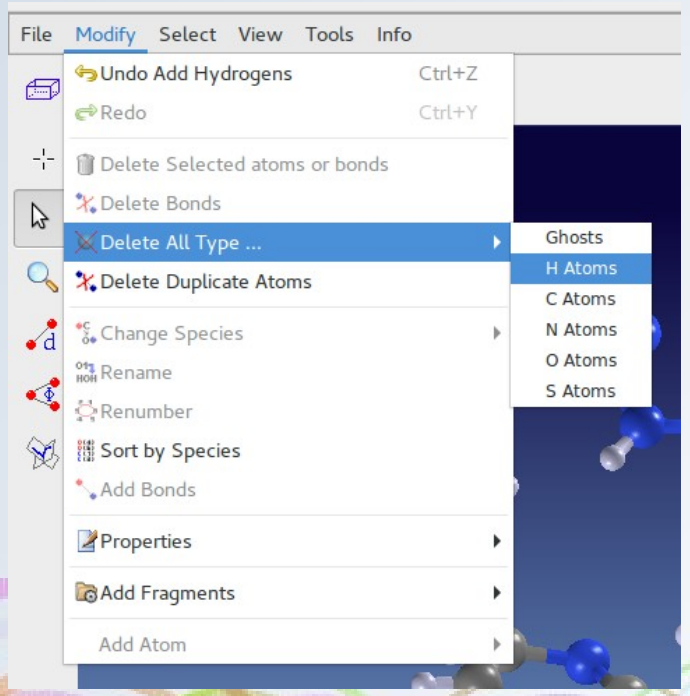
Experience and chemical intuition are required to build the correct model

H atoms

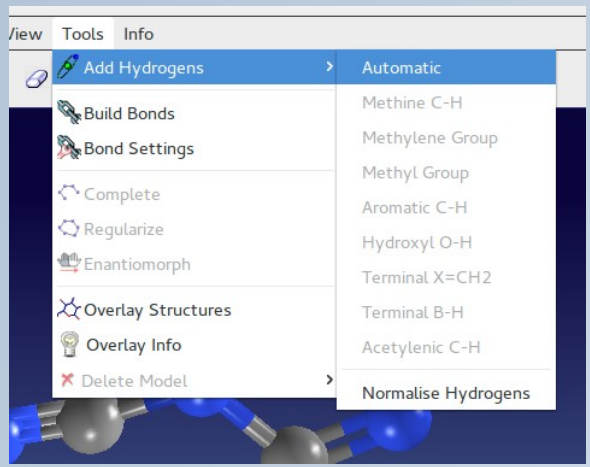
H atoms do not contribute significantly to X-ray diffraction, they can be ignored during the structure solution

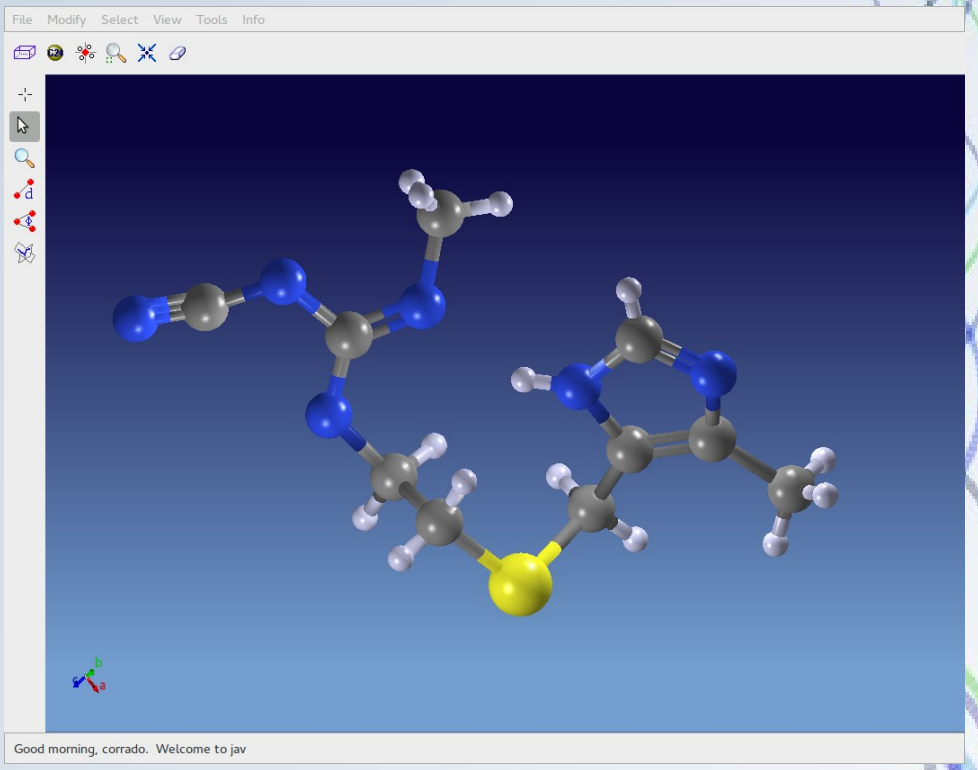
Eliminating the H atoms reduces the number of atoms and DoFs, decreasing the time to evaluate CF for each trial structure

Delete H atoms using the GUI or using the `deletehydro` directive



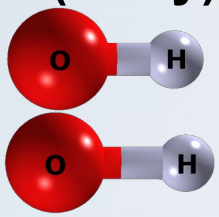
Hydrogen calculation





Hydrogen atoms are positioned geometrically at X-ray distances

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

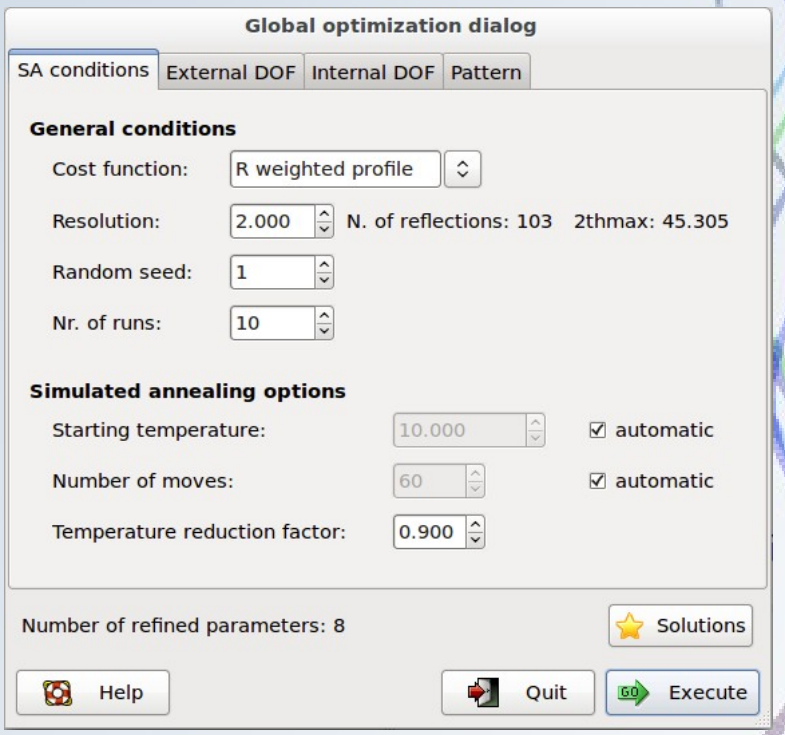


0.82 Å (X-ray)

0.95 Å (neutrons)

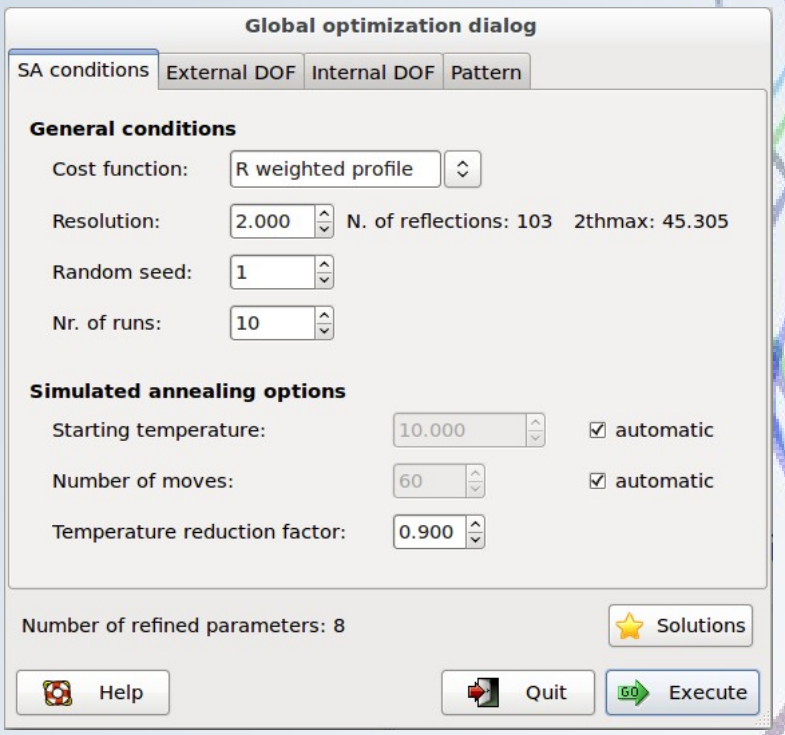
Simulated annealing options

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



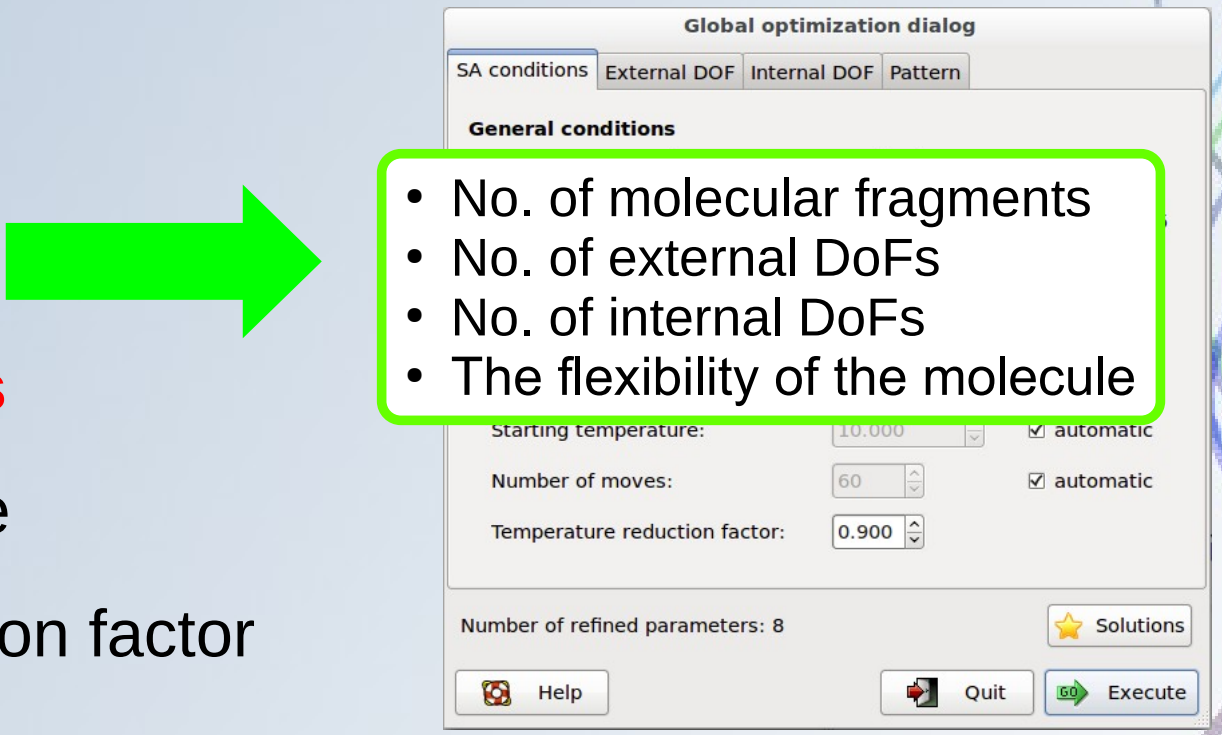
Simulated annealing options

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



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

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**

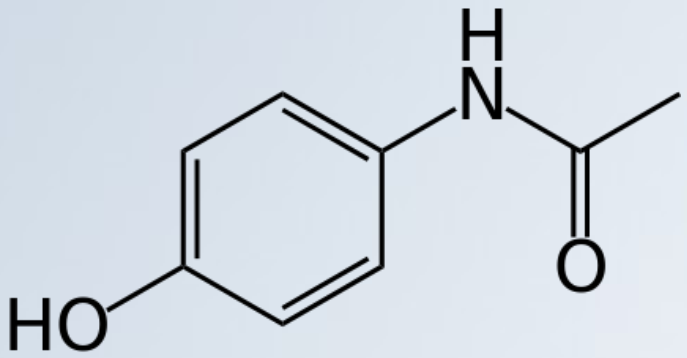
$$R_B = \frac{\sum_h |I_h^{exp} - I_h^{calc}|}{\sum_h I_h^{exp}}$$

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

Using expo2014 for direct-space solution

- **By input file (*.exp)**

  File → Load and 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**

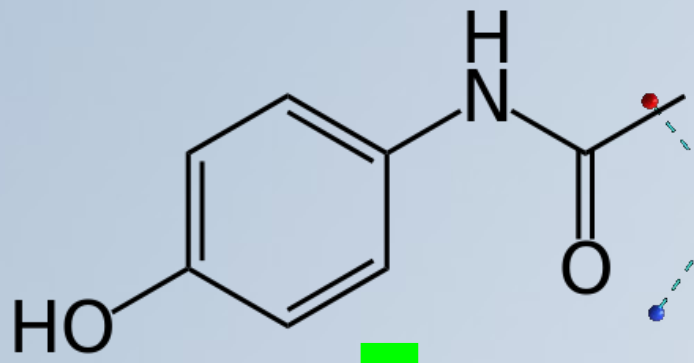
```
expo paracetamol.exp
```
or
```
expo paracetamol.exp -nogui
```
or
```
expo paracetamol.exp -auto
```

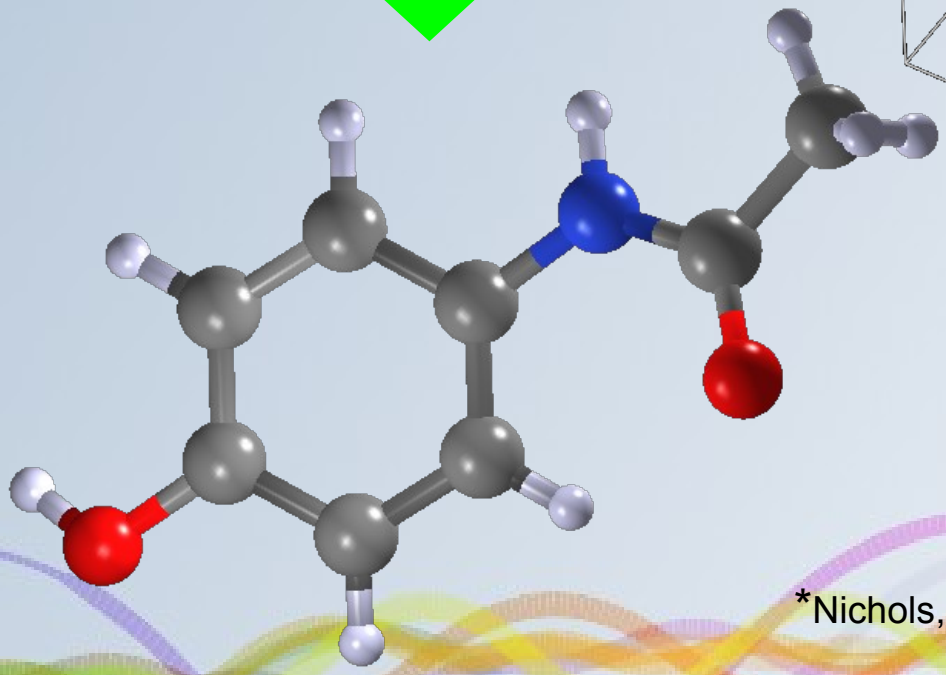
Usage of the command %fragment

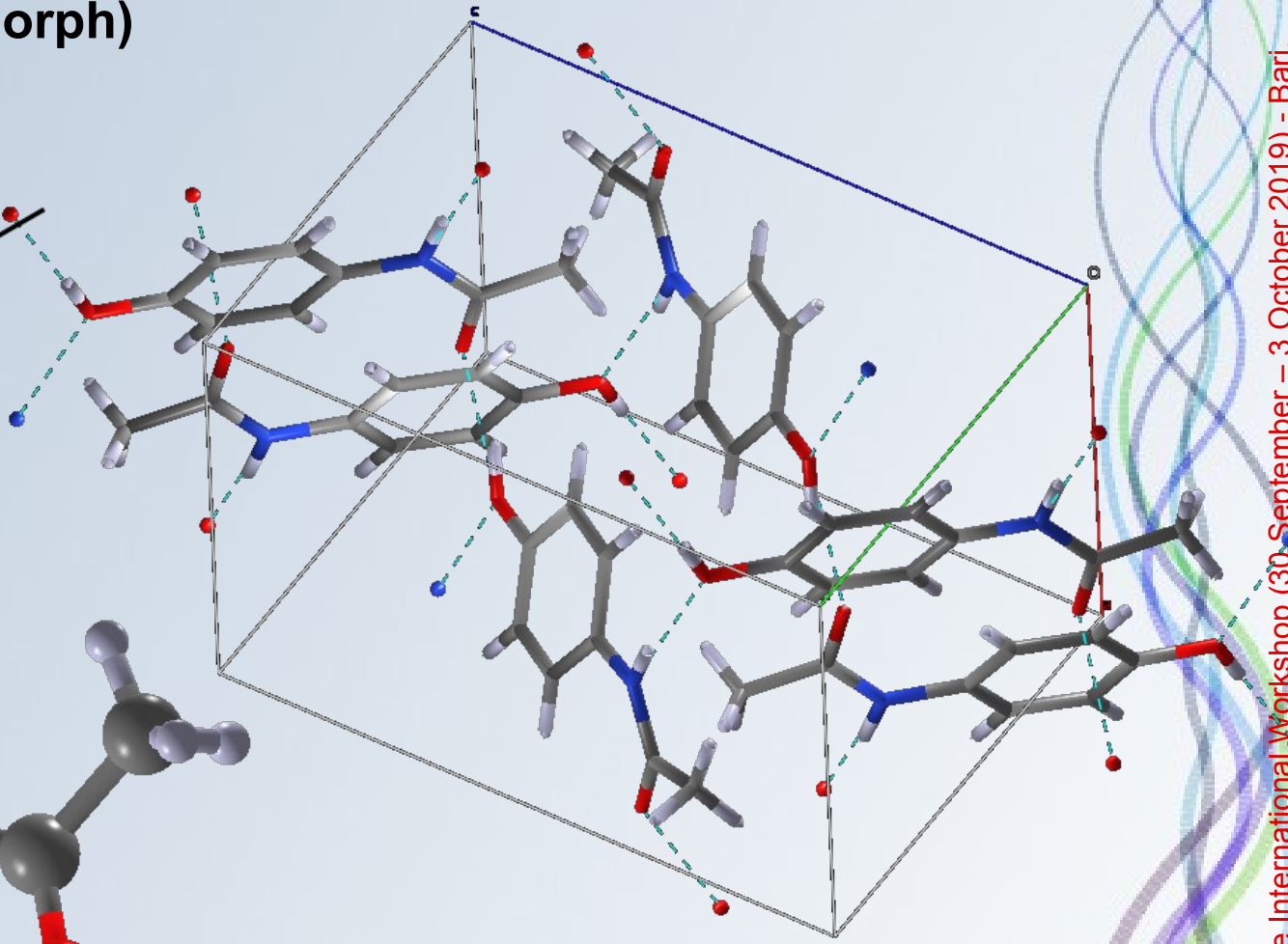
```
%fragment tetra AtC AtV [dist]
%fragment octa AtC AtV [dist]
%fragment square AtC AtV [dist]
%fragment cube AtC AtV [dist]
%fragment trigonal AtC AtV [dist]
%fragment prism_tetra AtC AtV [dist]
%fragment prism_trig AtC AtV [dist]
%fragment icosa AtC AtV [dist]
%fragment atoms chem_formula
%fragment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

Molecular compounds

Paracetamol (form I polymorph)
($C_8H_9NO_2$) *

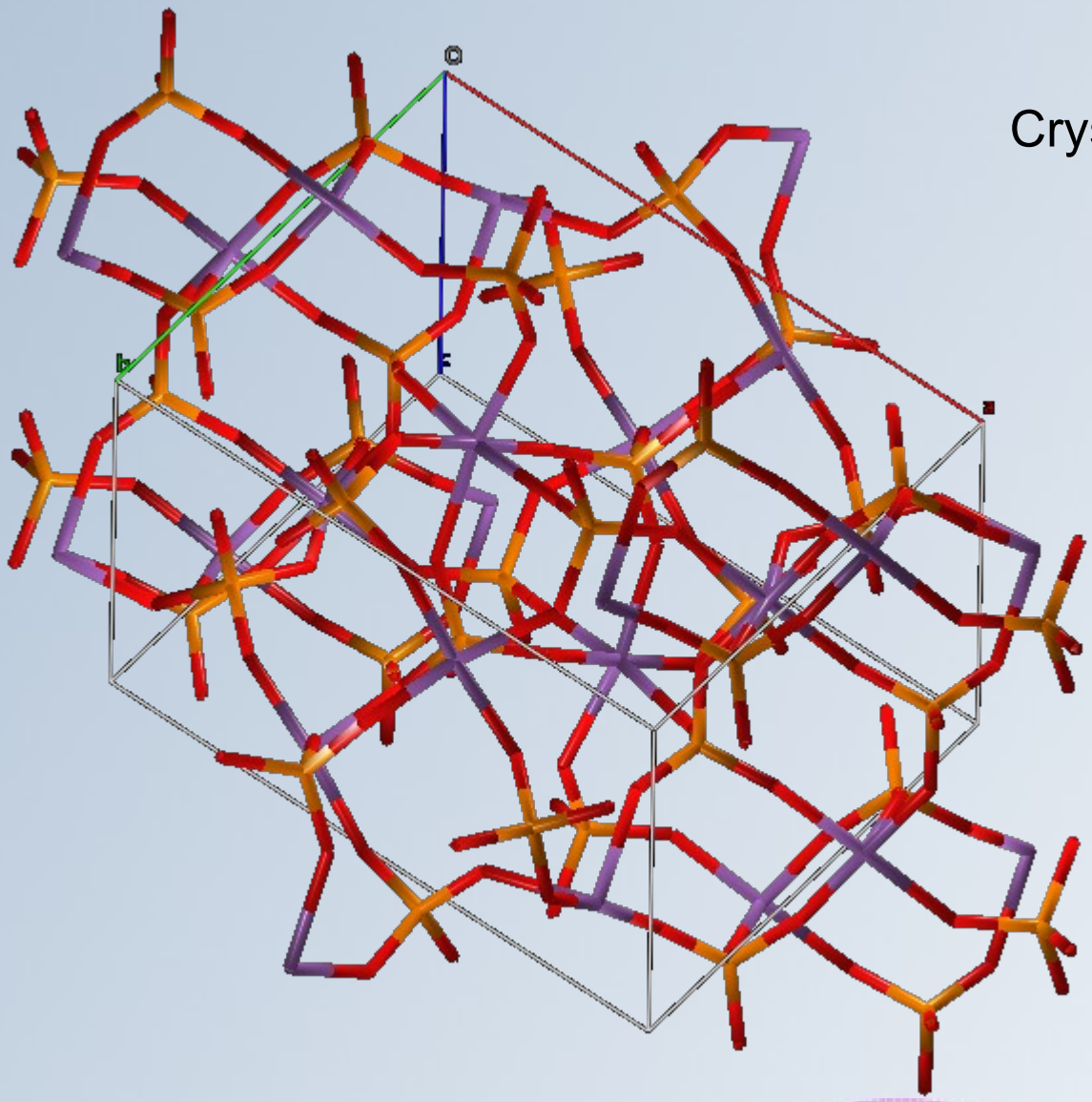






*Nichols, C. & Frampton, C. S. (1998). J. Pharm. Sci. 87, 684–693.

Non-molecular compounds

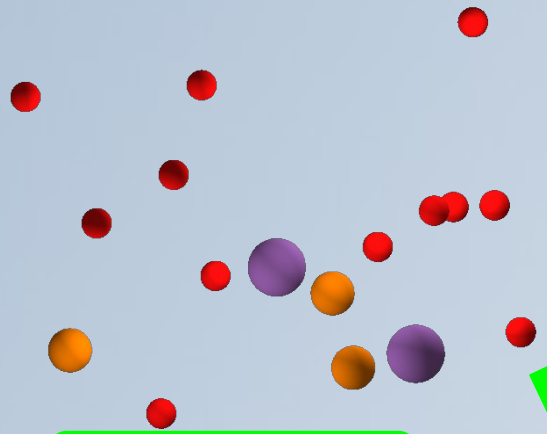


Crystal structure of $Sb_2(PO4)_3$ *

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

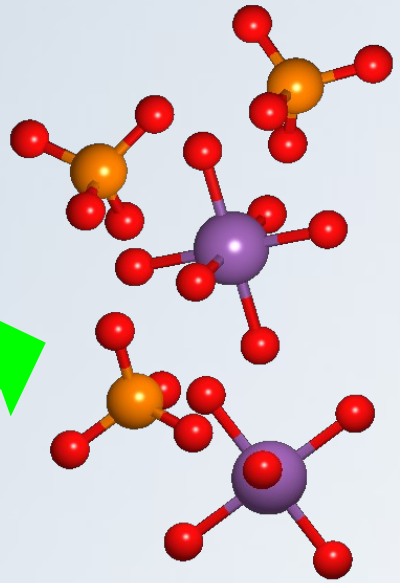
Non-molecular compounds

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

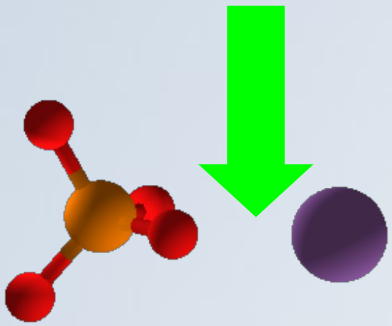


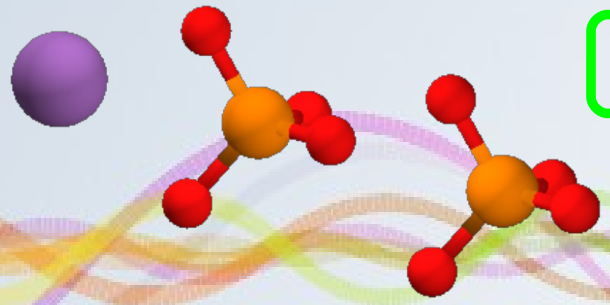
$N_{dof} = 51$

$Sb_2(PO_4)_3$



$N_{dof} = 30$

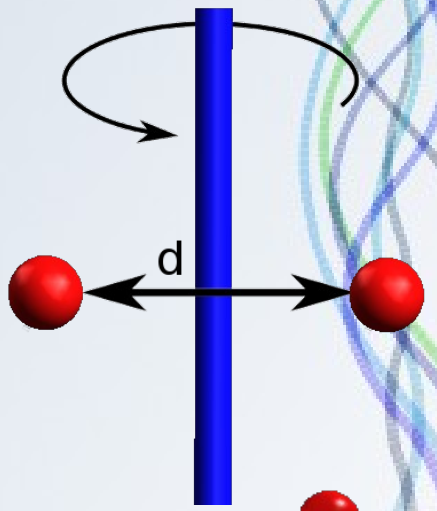


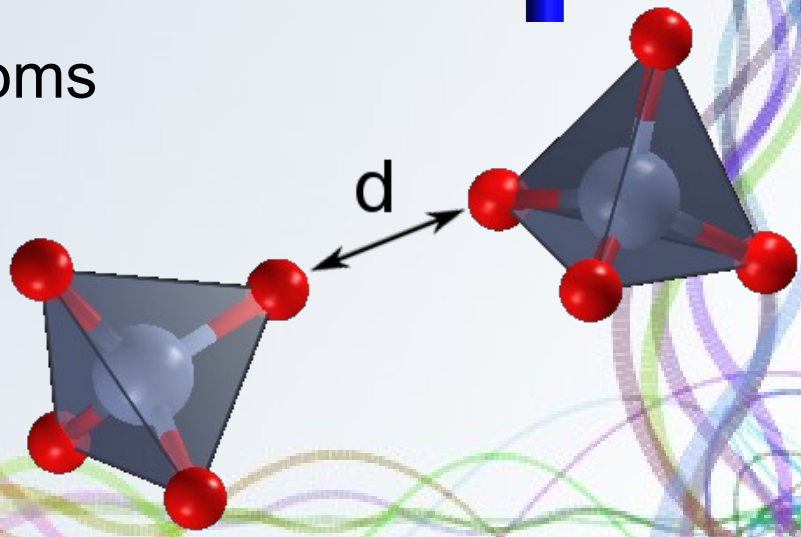


$N_{dof} = 24$

Non-molecular compounds

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

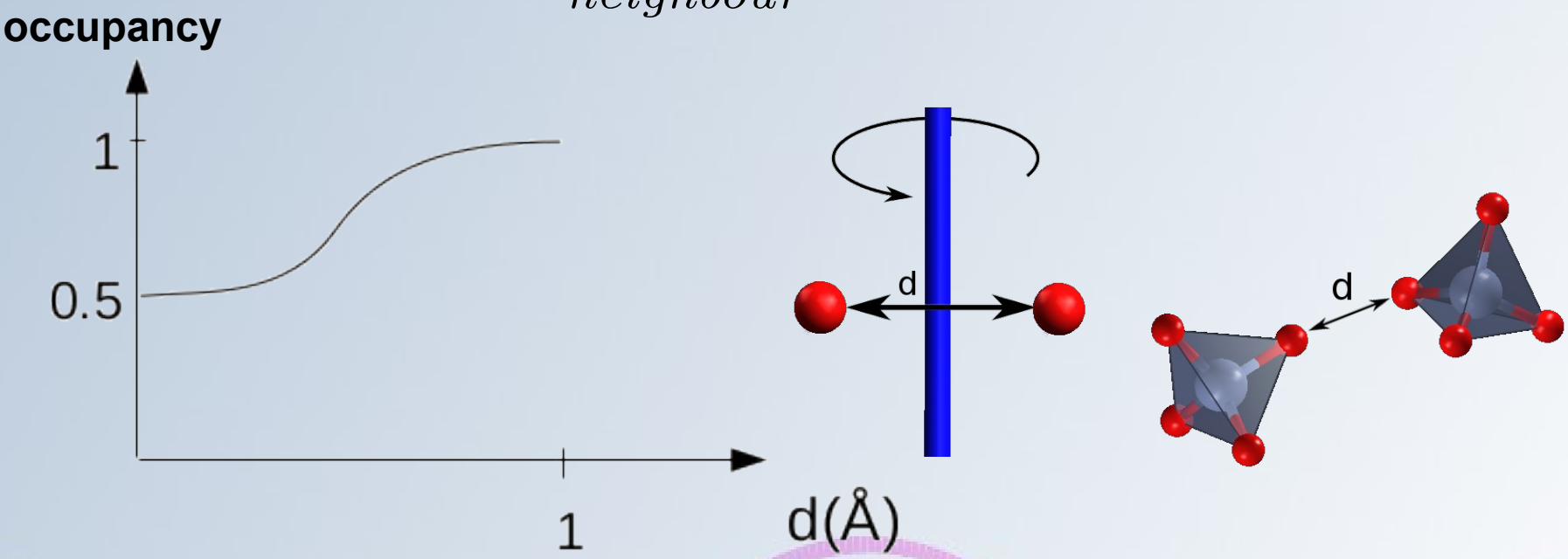




Dynamical occupancy correction (DOC)

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

$$occupancy = \frac{1}{1 + \sum_{neighbour} |d_{min} - d_i|} \qquad d_{min} = 1\mathring{A}$$

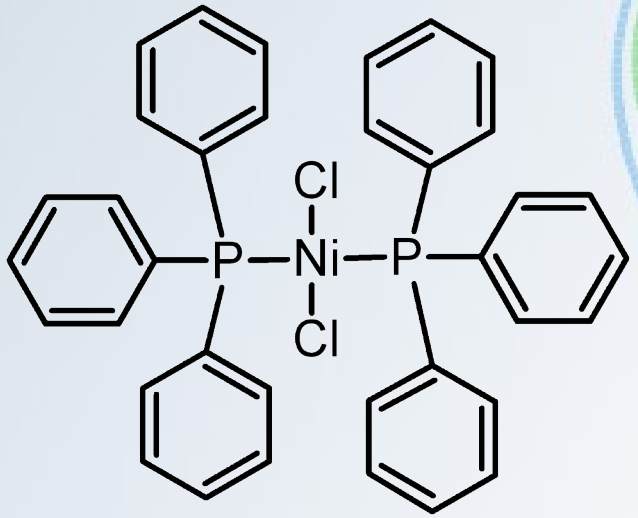


DOC is able to merge the excess atoms automatically

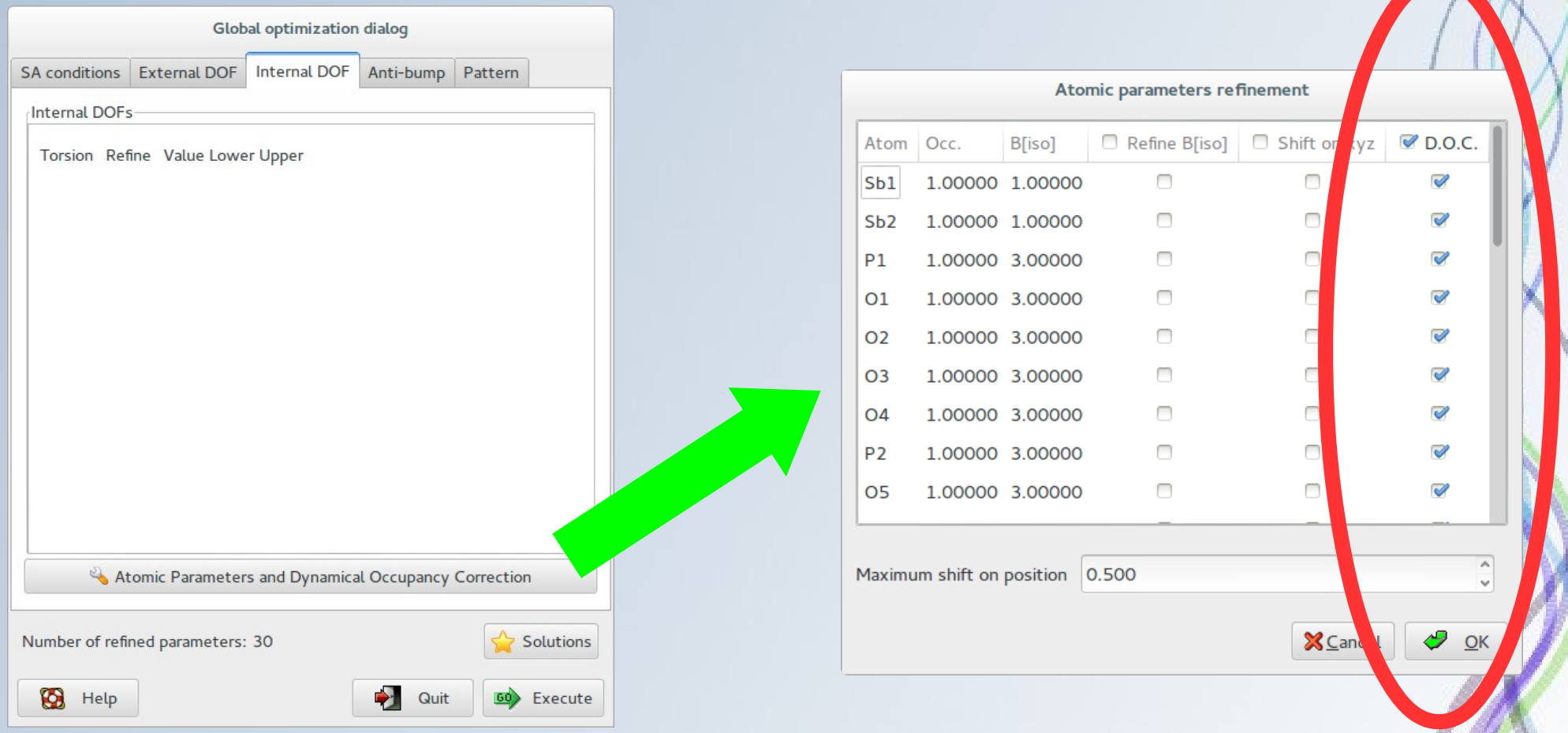
Dynamical occupancy correction (DOC)

```
doc

or

doc atom1 atom2 ...
```

```
doc Ni1
```

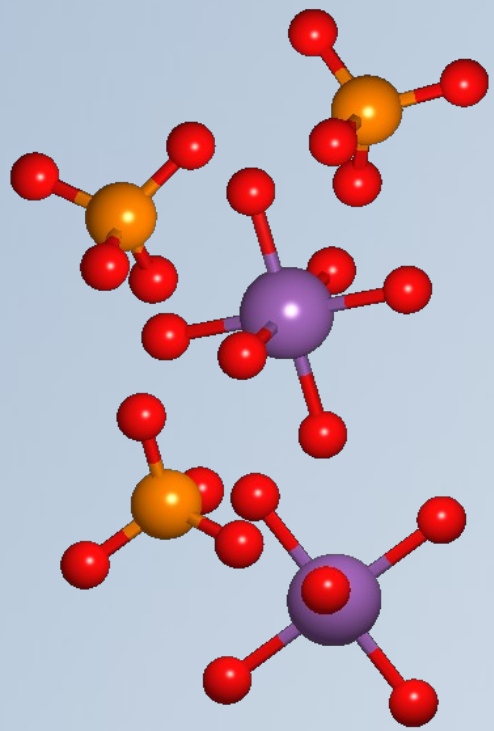


Dynamical occupancy correction (DOC)



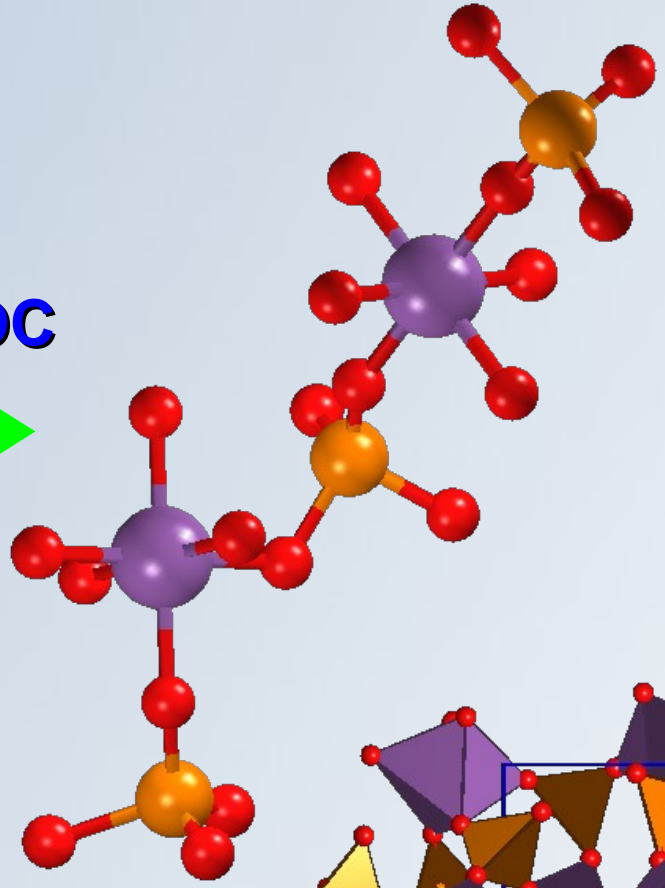
DOC slows down the computation time so it should be avoid if no special positions or shared atoms are expected.

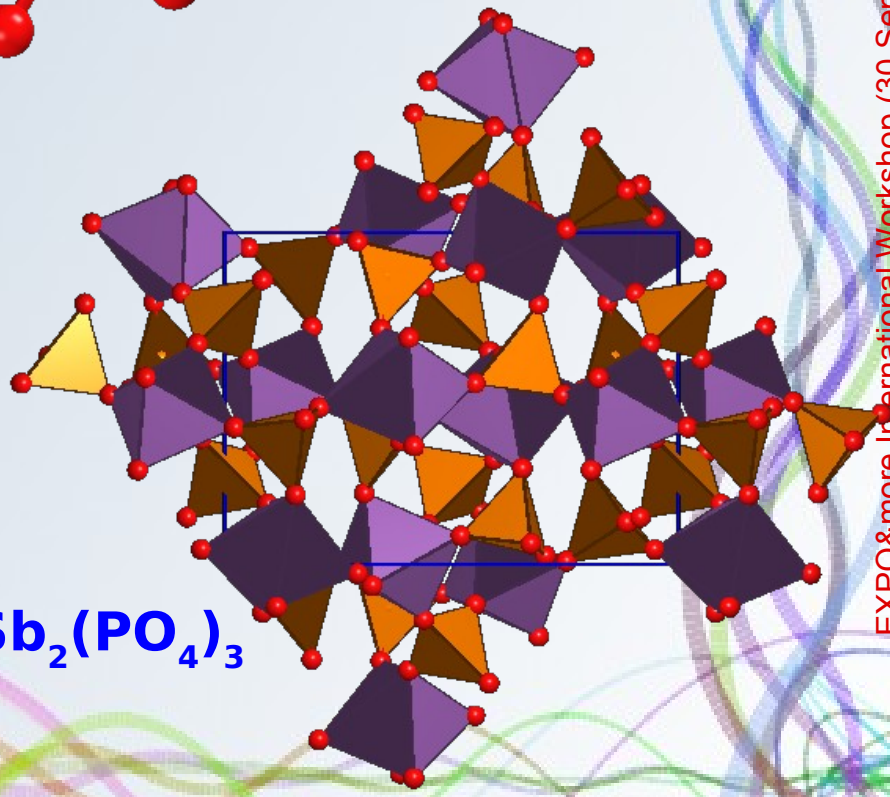
SA applied to non-molecular compounds



$Sb_2(PO_4)_3$

SA with DOC

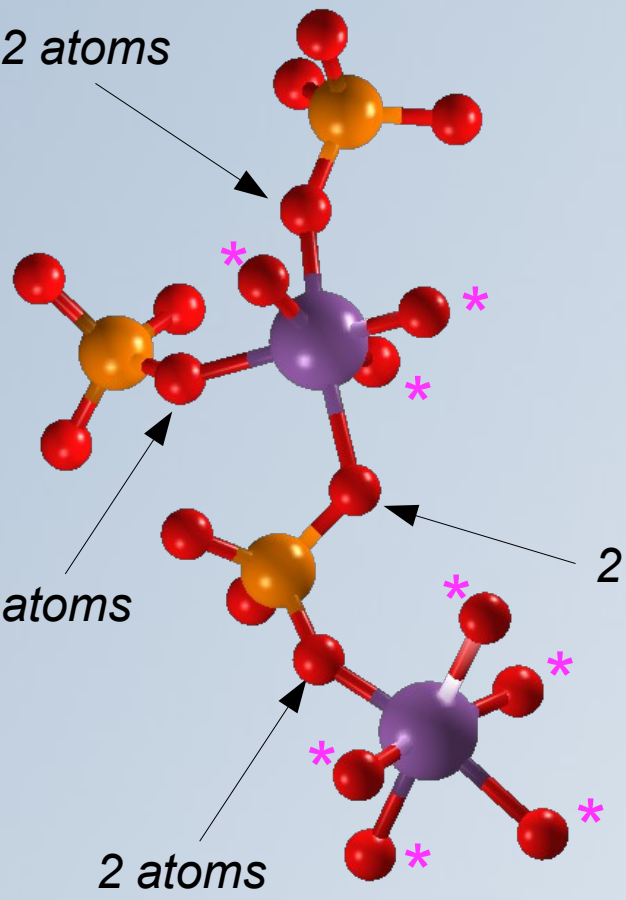


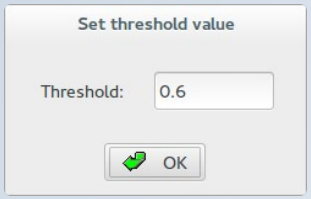


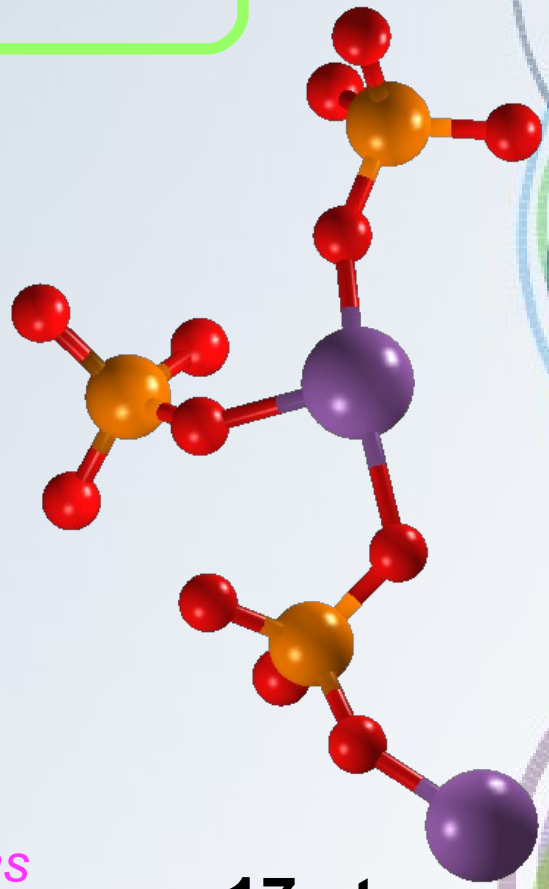
Polyhedra rapresentation of $Sb_2(PO_4)_3$
by SA with DOC

Delete duplicate atoms

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







* *symmetry equivalent atoms*

29 atoms

17 atoms

Direct Space 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}

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

R_{ij} distance between atoms *i* and *j*

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

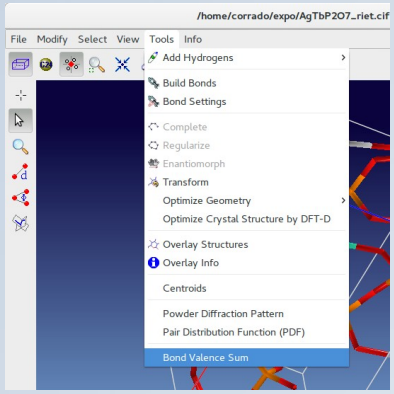
$$G_{ii} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(V_i - V_i^0)^2}$$ **global-instability index**

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



/home/corrado/expo/AgTbP2O7_riet.cif

File Modify Select View Tools Info

Tb1 #12 Dist: O2: 2.259, O6: 2.532, O3: 2.380, O7: 2.309

Bond Valence Sum

GII = 0.057188

Edit bond valence parameters

Export

Number	Label	Type	BVS	CN
1	Tb1	Tb	3.07897	8
2	Ag1	Ag	0.88664	7
3	P1	P	4.67892	4
4	P2	P	5.10217	4
5	O1	O	1.93926	3
6	O2	O	2.01691	3
7	O5	O	1.89685	4
8	O6	O	1.97989	4
9	O3	O	1.99975	4
10	O7	O	2.00836	3
11	O4	O	1.90568	2

Bonds of atom Tb1

No.	Label	Type	x	y	z	Symm.Op.	Distance	Bond Valence
5	O1	O	0.7449	-0.0131	0.5730	(x, y, z)	2.510	0.274
5	O1	O	0.7551	0.4869	0.9270	(-x+1/2, y+1/2, -z+1/2)+(1,0,1)	2.282	0.508
6	O2	O	0.7257	0.1770	1.2368	(x, y, z)+(0,0,1)	2.259	0.541
7	O5	O	0.5711	0.1740	0.8702	(x+1/2, -y+1/2, z+1/2)	2.363	0.409
8	O6	O	0.9406	0.2950	1.0802	(x, y, z)+(0,0,1)	2.532	0.259
9	O3	O	0.6998	-0.1650	0.9334	(-x+1/2, y+1/2, -z+1/2)+(1,-1,1)	2.380	0.390
9	O3	O	0.8002	0.3350	0.5666	(x, y, z)	2.585	0.224
10	O7	O	0.9298	-0.0191	0.8490	(-x, -y, -z)+(2,-1,1)	2.309	0.473

Close

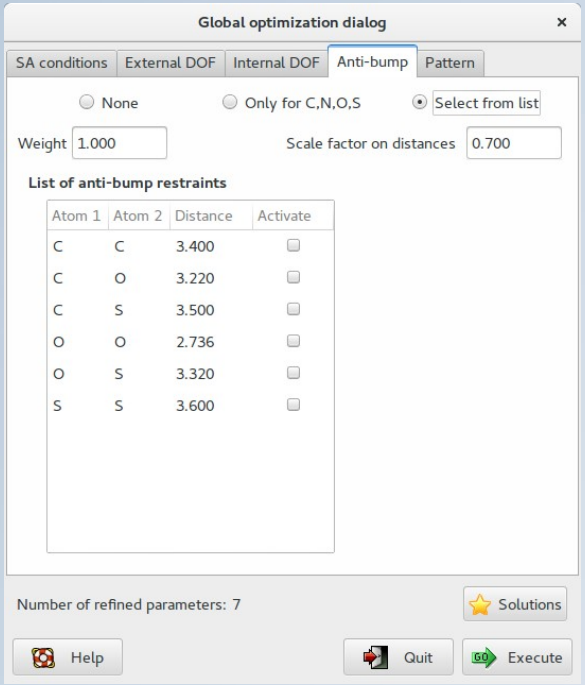
Imposing anti-bumping restraints

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

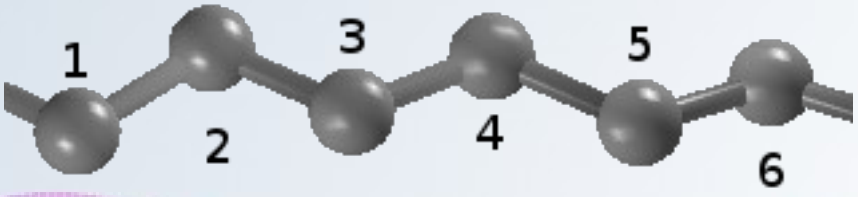
$$k = 2$$

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

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



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



Only 1-5 interactions are considered

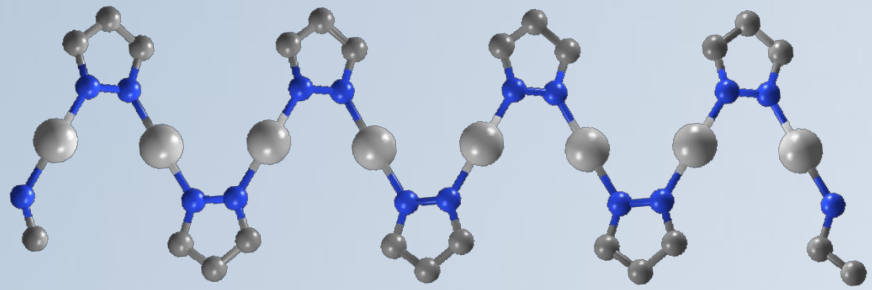
Imposing anti-bumping restraints

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

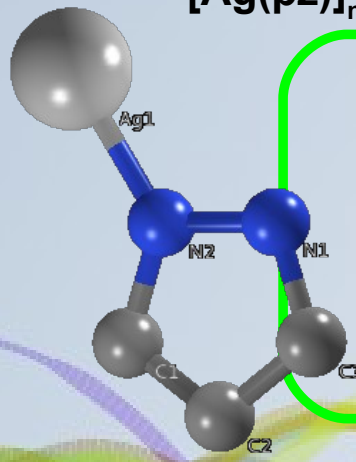
$$k = 2$$

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

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



$[Ag(pz)]_n$ Hpz = pyrazole




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

Directive:

bump

or

bump atoms1 atoms2 [dist]

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

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

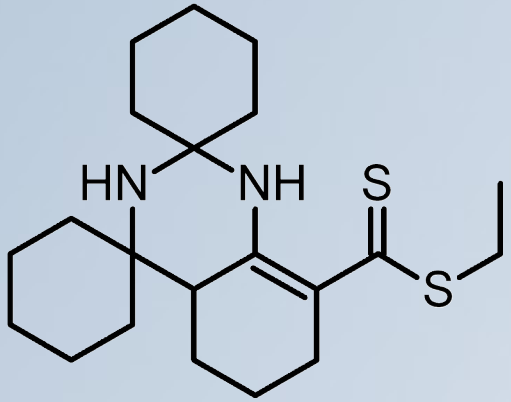
d_{target} and tol are optional:

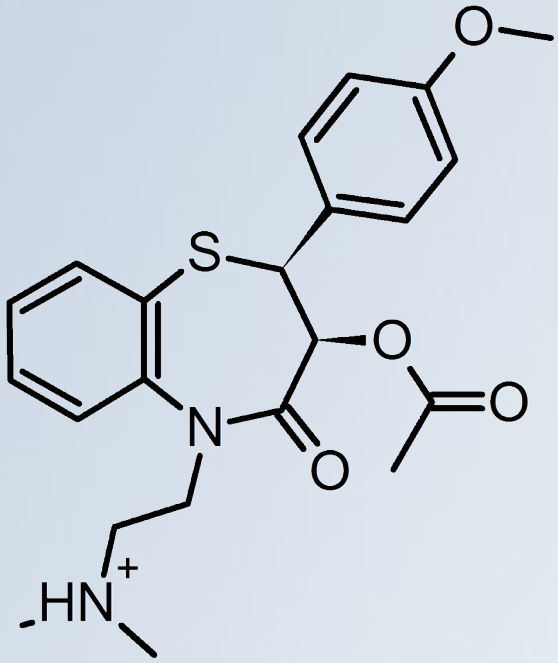
rest A B

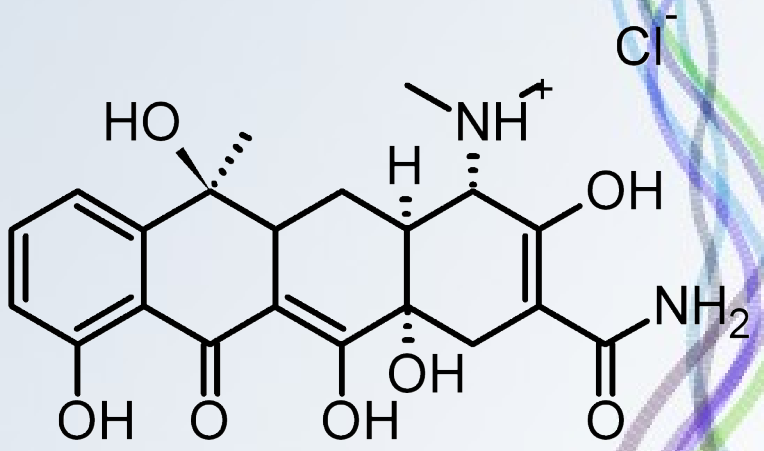
- Rarely improve the success rate of the solution search for good quality data
- Restraints can slow or prevent a structure solution

Non planar ring systems

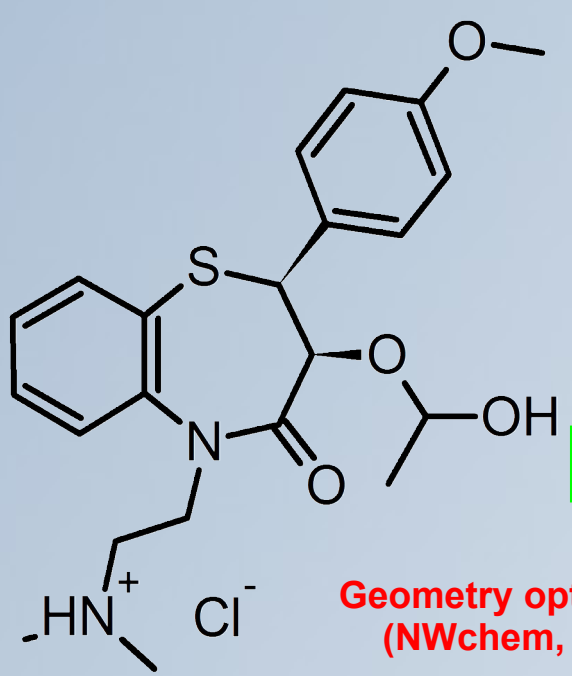
Attention to non planar ring systems or unusual combinations of elements in functional groups

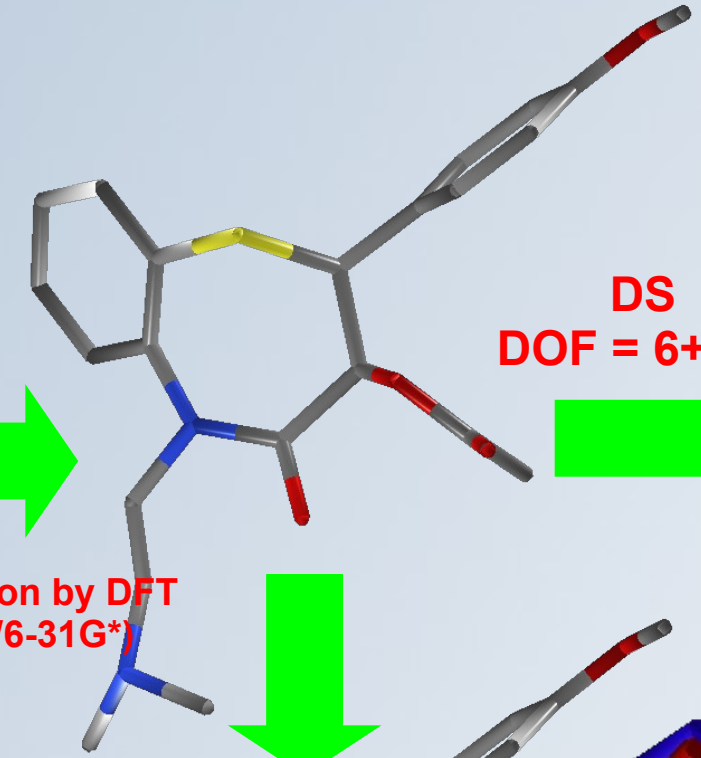






Structure Solution of Diltiazem Hydrochloride*

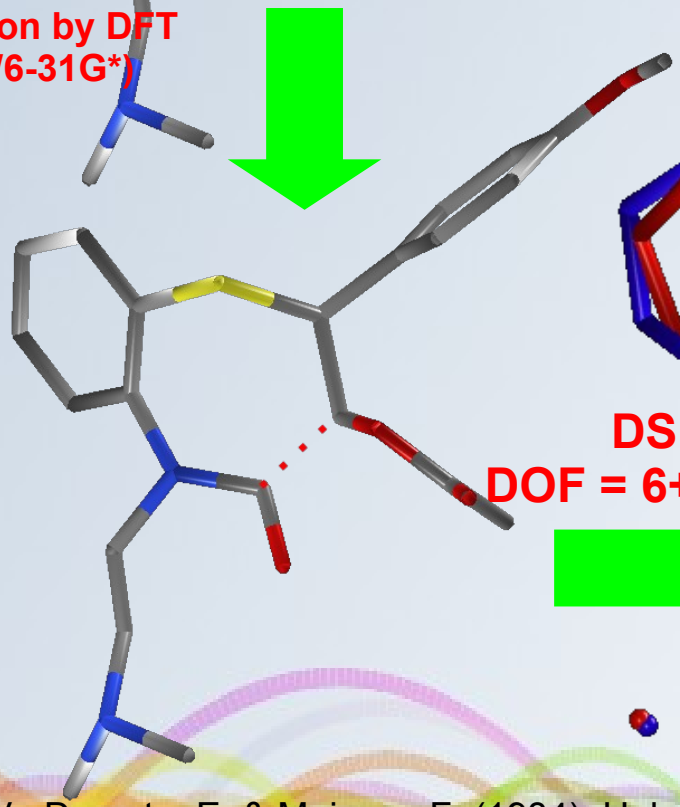




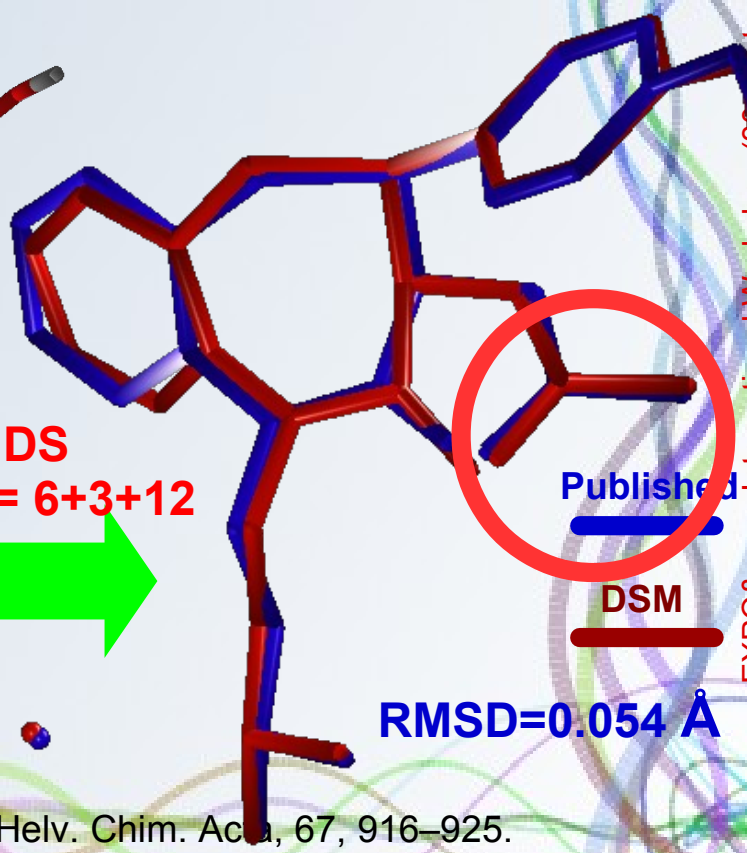
DS
DOF = 6+3+7

NO SOLUTION!

Geometry optimization by DFT
(NWchem, B3LYP/6-31G*)



DS
DOF = 6+3+12

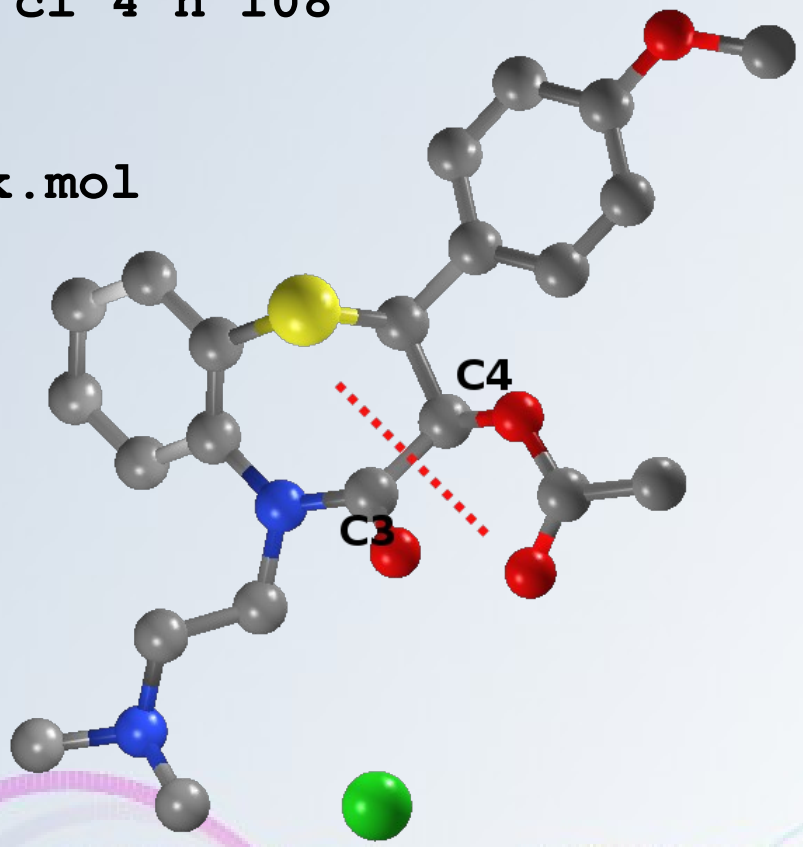


EXPO&more International Workshop (2 September – 3 October 2019) - Bari,

*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

```
%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
%fragment atoms Cl
%sannel
nrun 100
niter 5000
rest C3 C4
%save diltia.expo
```



Parallel Machines

Notebooks

Typically 2-6 cores

Smartphones

Typically 2-10 cores

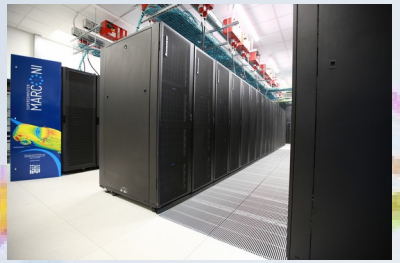
Workstations

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)

  Coprocessor architecture

Three Programming models

- **Message Passing Interface (MPI)**

  Distributed-memory architecture

- **Open MultiProcessing (OpenMP)**

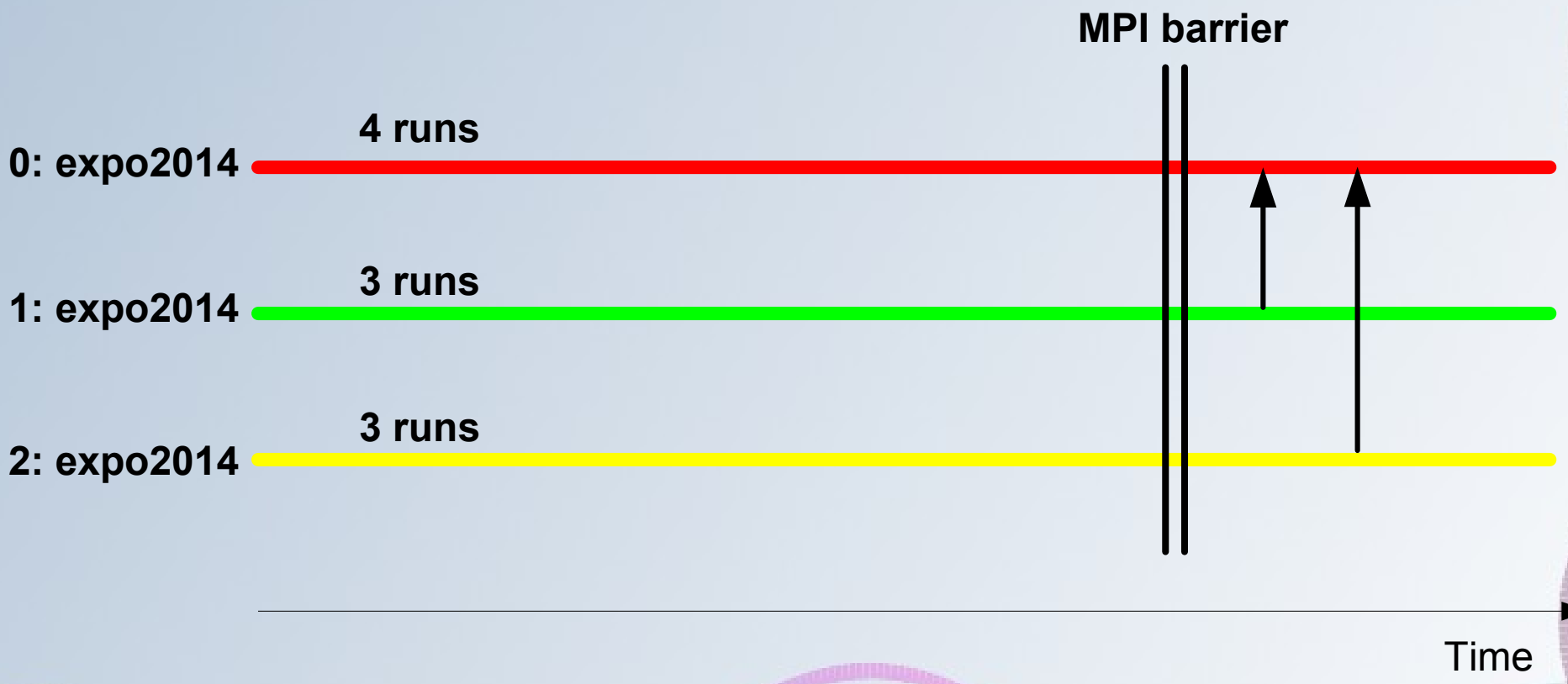
  Shared-memory architecture

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

  Coprocessor architecture

Message Passing Interface (MPI)

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

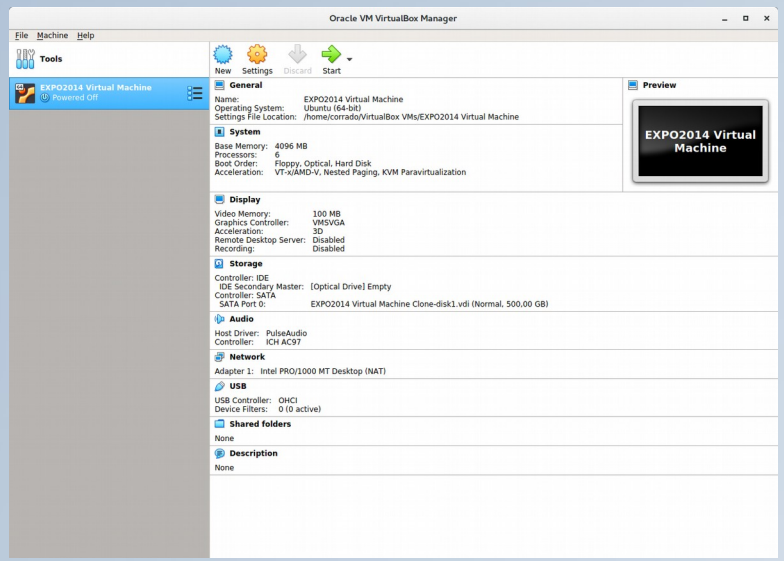


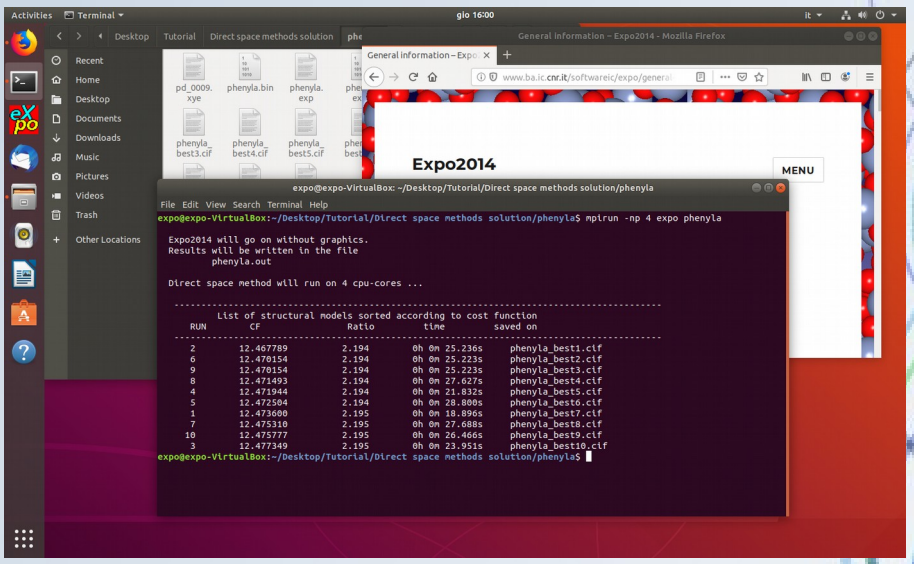
Running the Parallel Version of EXPO2014

- Computer with multi-core CPUs and Linux environment.
- 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
```

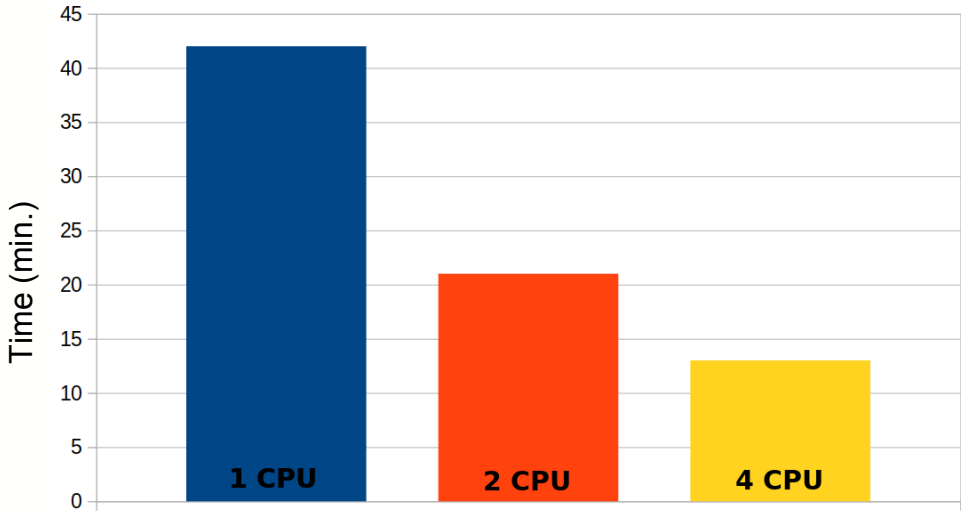
EXPO2014 Virtual Machine



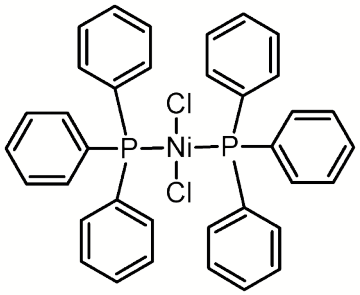


Download and installation

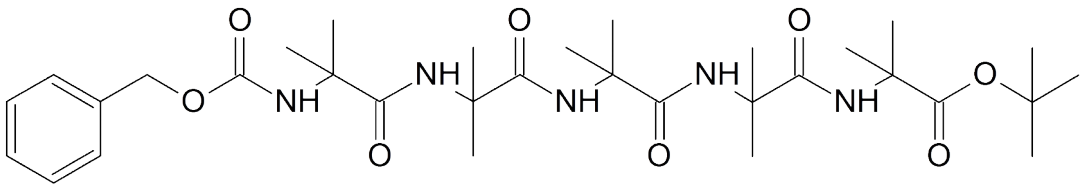
http://www.ba.ic.cnr.it/softwareic/expo/parallelism-in-expo2014-for-structure-solution-by-direct-space-method/



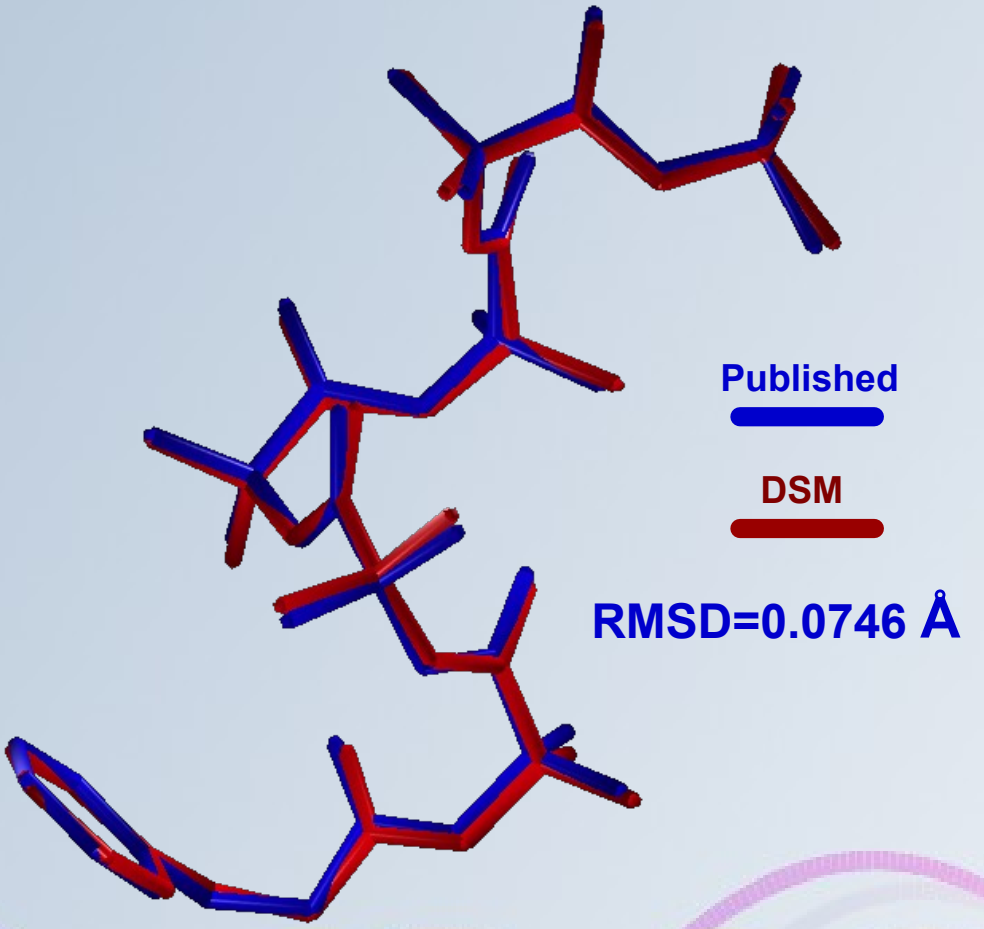
Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz



Structure solution of small peptide



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



DOF

Internal: 20
External: 6

Algorithm settings

NRUN: 100
NITER: 1000

Time

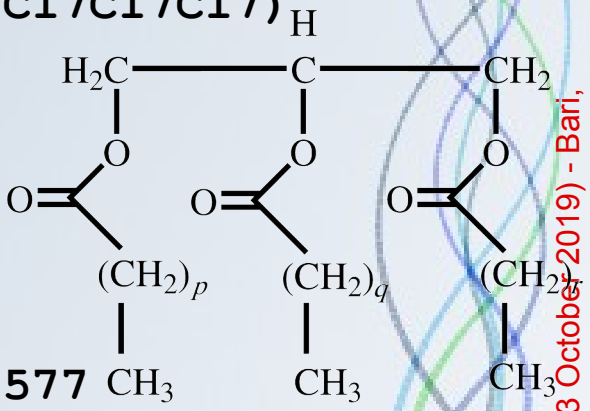
Single CPU-core: 500 h
20 CPU-cores: 25 h
Parallel speedup: 20

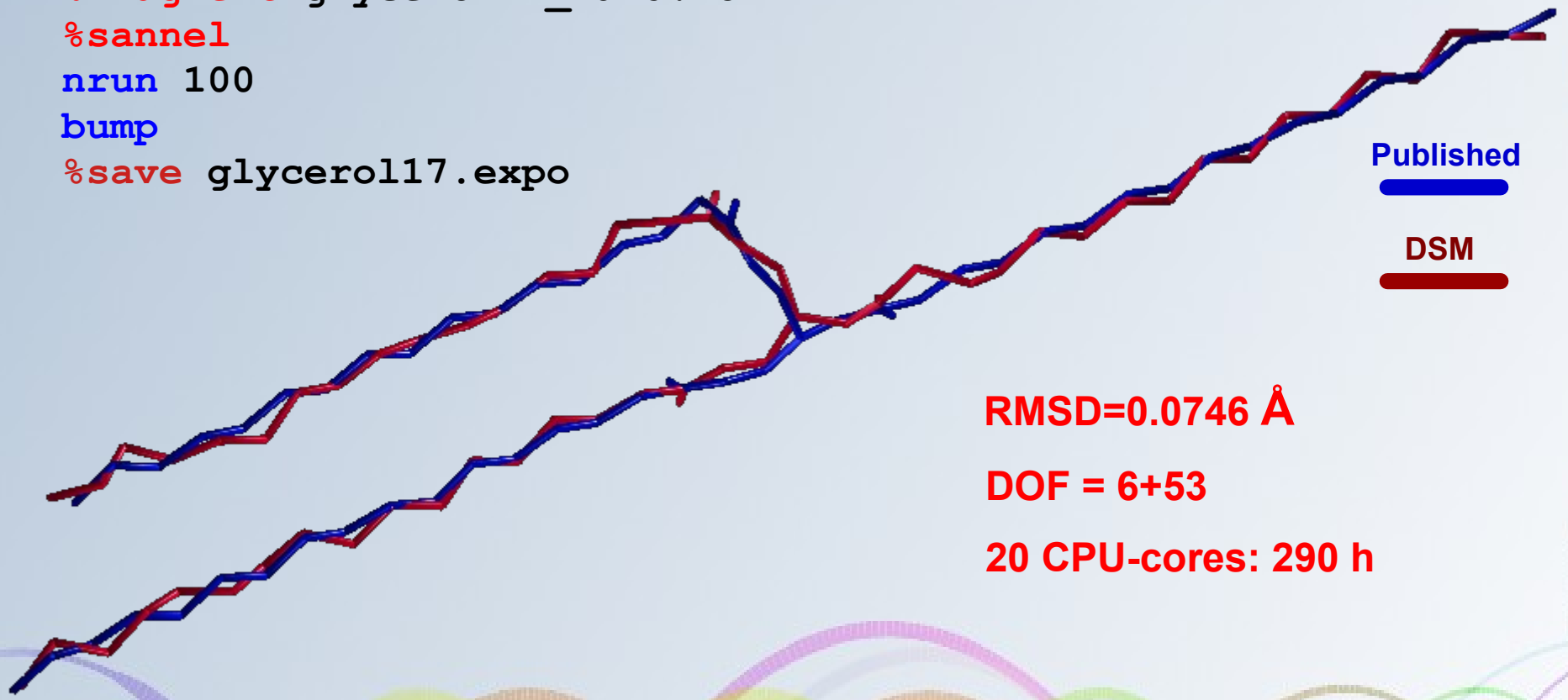
EXPO&more International Workshop (30 September – 3 October 2019) - Bari,

E. Benedetti, A. Bavoso, B. Di Blasio, V. Pavone, C. Pedone, M. Crisma, G. M. Bonora, and C. Toniolo, *J. Am.Chem. Soc.* **104**, 2437 (1982)

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
%fragment glycerol17_rand.mol
%sannel
nrun 100
bump
%save glycerol17.expo
```





RMSD=0.0746 Å

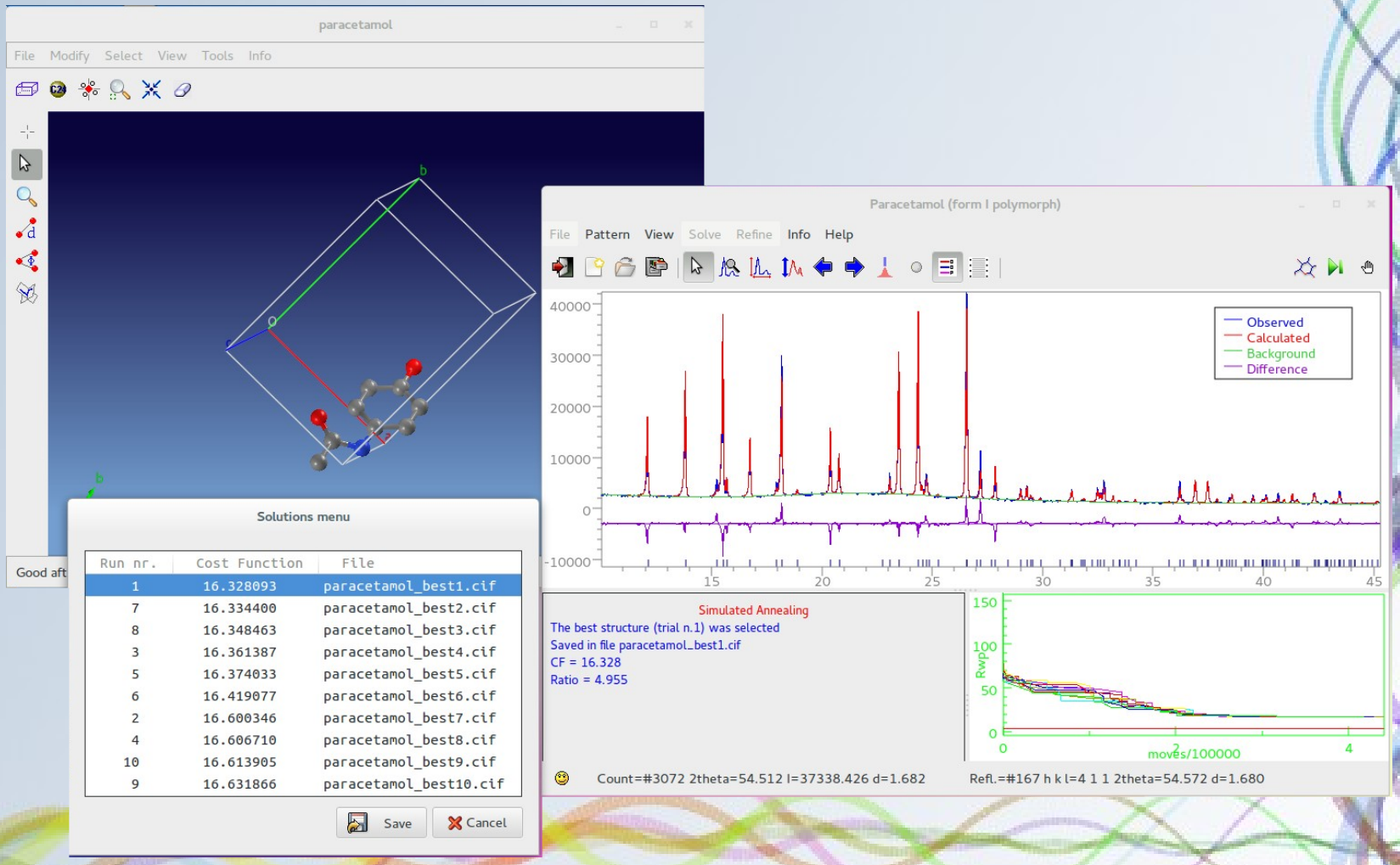
DOF = 6+53

20 CPU-cores: 290 h

Robert B. Helmholdt, René Peschar and Henk Schenk. Acta Cryst. (2002). B**58**,134±139
https://doi.org/10.1107/S0108768101016330

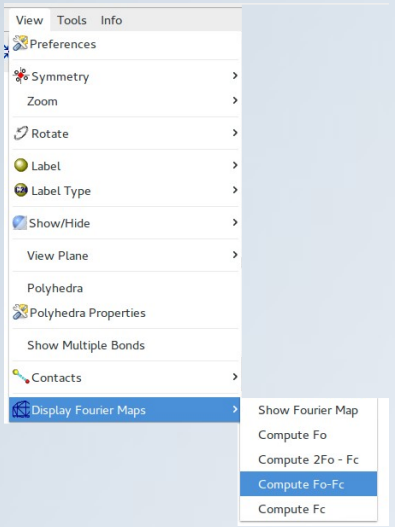
Assessing the solution

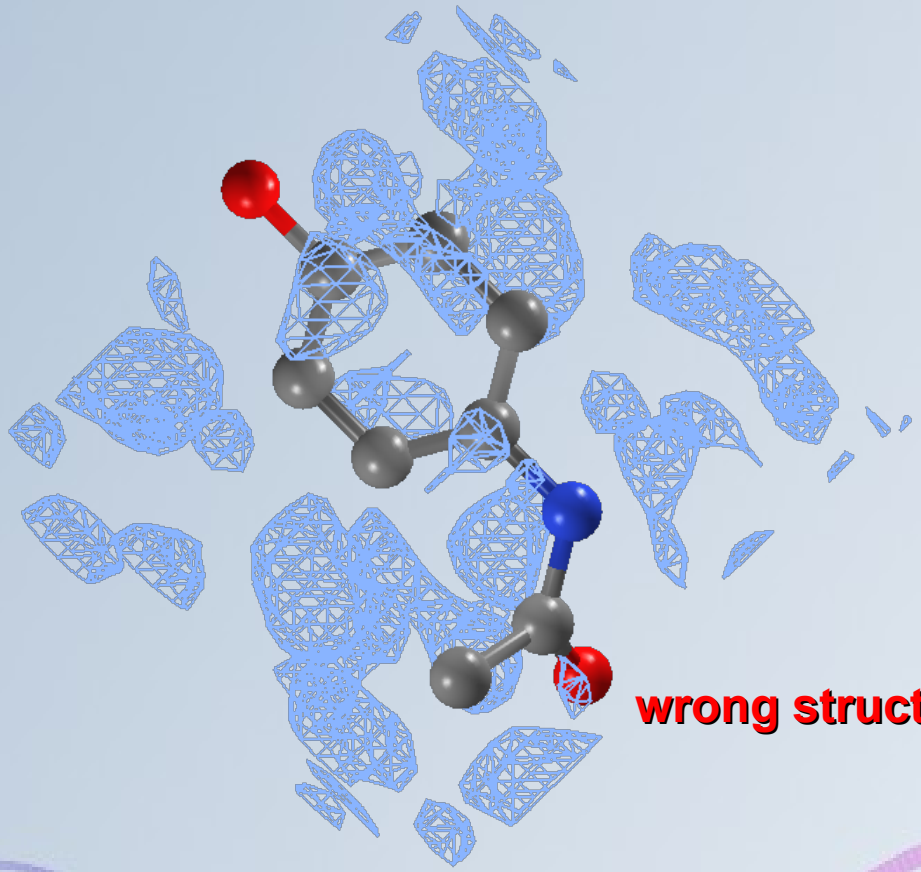
- Agreements factors
- Visual match between calculated and observed profile
- Reproducibility of solution



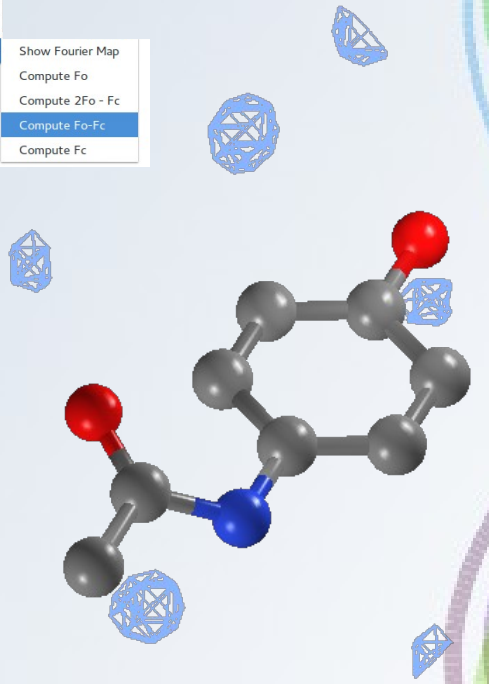
Assessing the solution

- Display difference Fourier map





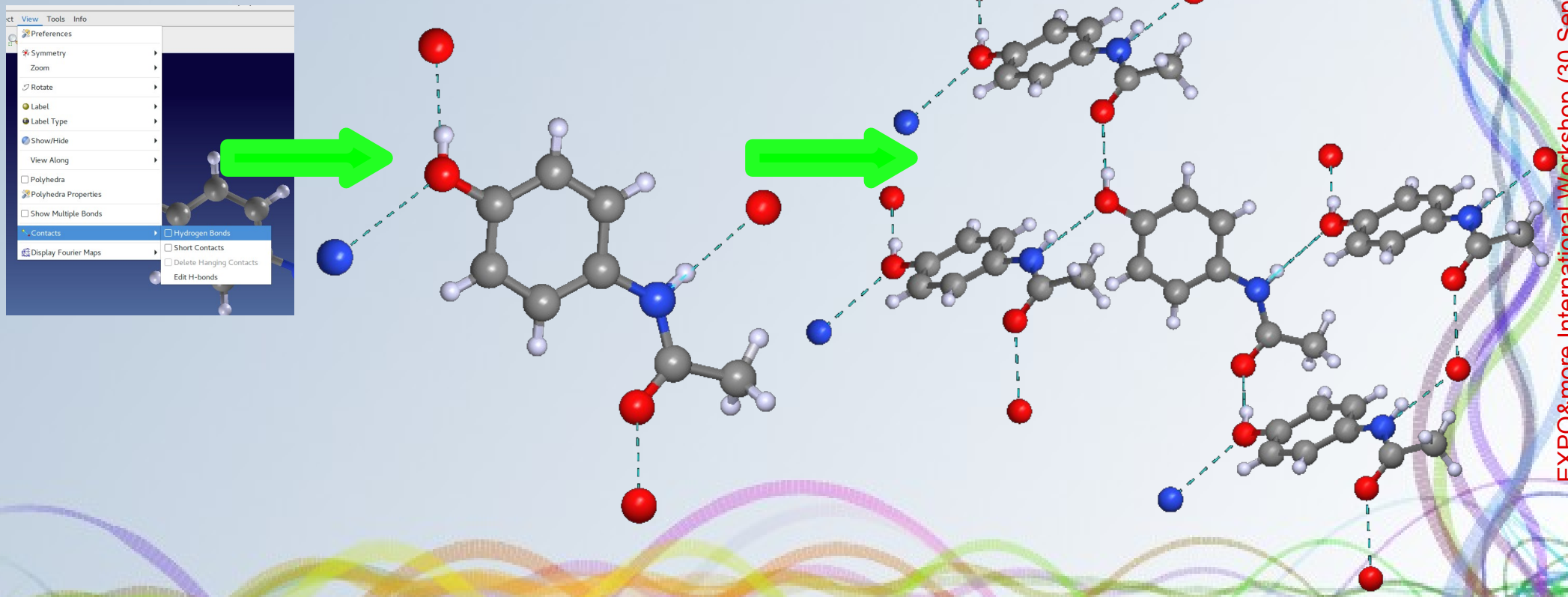
wrong structure



correct structure

Assessing the solution

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



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 package*

When Structure Solution Fails

- Poor quality diffraction pattern

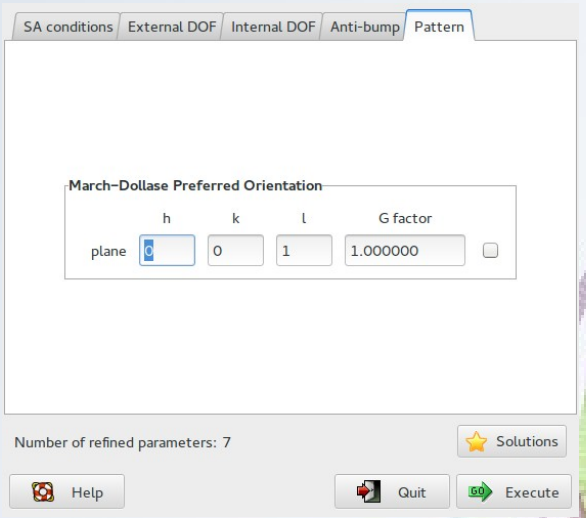
Solution:

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

- Systematic problem in powder diffraction data
  - *Preferred orientation*
  - *Kα2 contributions*

Solution:

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

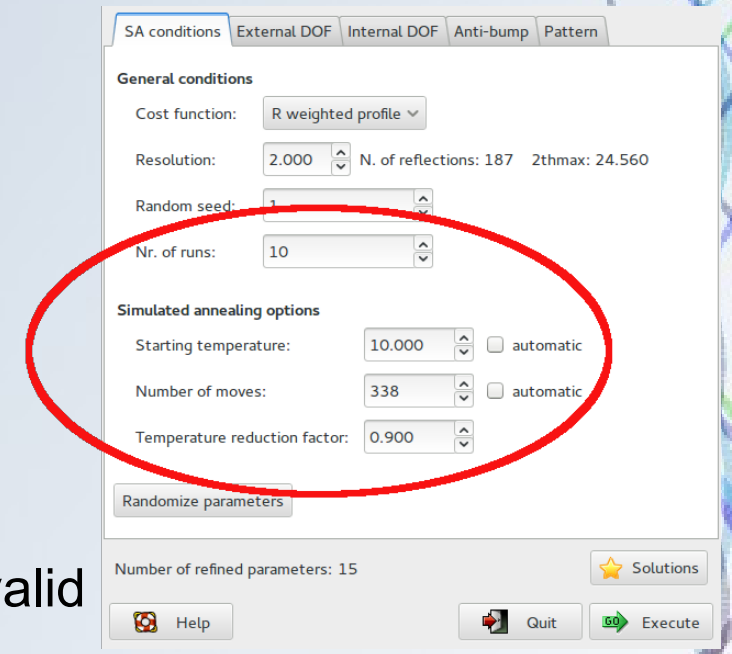


When Structure Solution Fails

- For complex structure (internal DOF > 10) the default SA conditions can be not sufficient

Solution:

- *Increase the number of moves (`niter` directive) and/or runs (`nrun` directive)*
- *Try with slower temperature reduction*



- The assumptions about thermal factors are invalid

Solution:

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

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*

Multiple phase powder diffraction patterns

```
%job Sample_ID_1h
%structure sample1h

%data
pattern cpd-1h.dat
wave 1.54056 1.54439 0.5

%crystal
name Al2O3
cell 4.75920   4.75920   12.992   90   90   120
space R-3c
fragment atoms Al O
doc

%crystal
name CaF2
cell 5.4649   5.4649   5.4649  90   90   90
space F m3m
fragment atoms Ca F
doc

%crystal
name ZnO
cell  3.2501   3.2501   5.2071   90   90   120
space P 63mc
fragment atoms Zn O
doc

%sannel
```

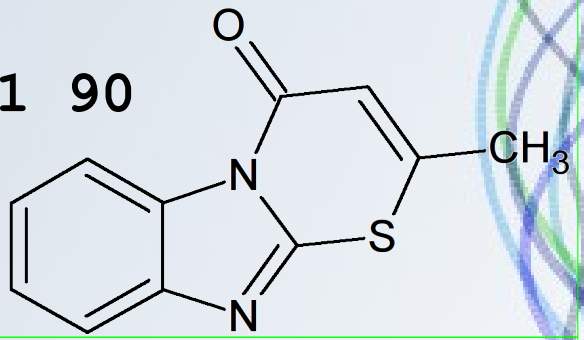
J. Appl. Cryst. (2001). 34, 409-426

Multiple phase powder diffraction patterns

```
%data
pattern L1.TXT
wavelength 1.54056
```

```
%crystal
name phase1
cell 7.4512 10.1767 13.2955 90 94.481 90
space P21/c
fragment  phase1_mopac_noHbis.mol
```

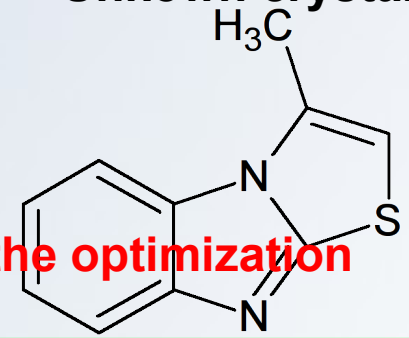
Known crystal phase




```
%crystal
name phase2
file pahse2 1single.cif
fixed
```

Fix the known phase during the optimization

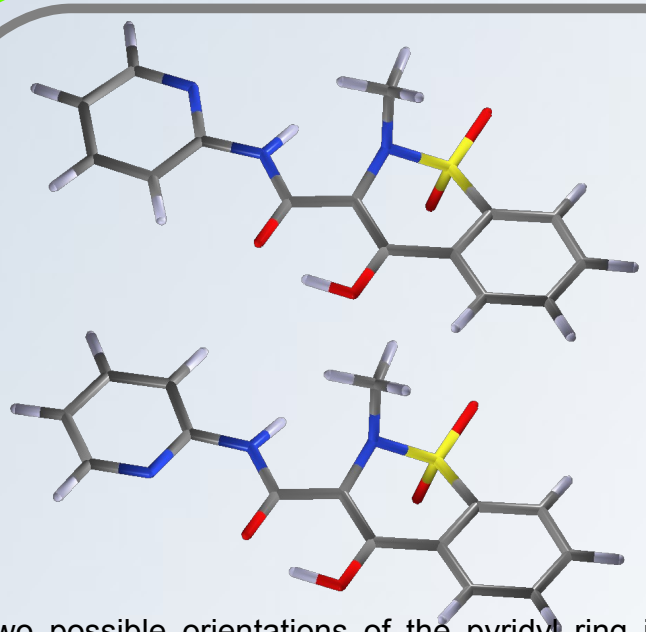
Unkown crystal phase




```
%sannel
```

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)

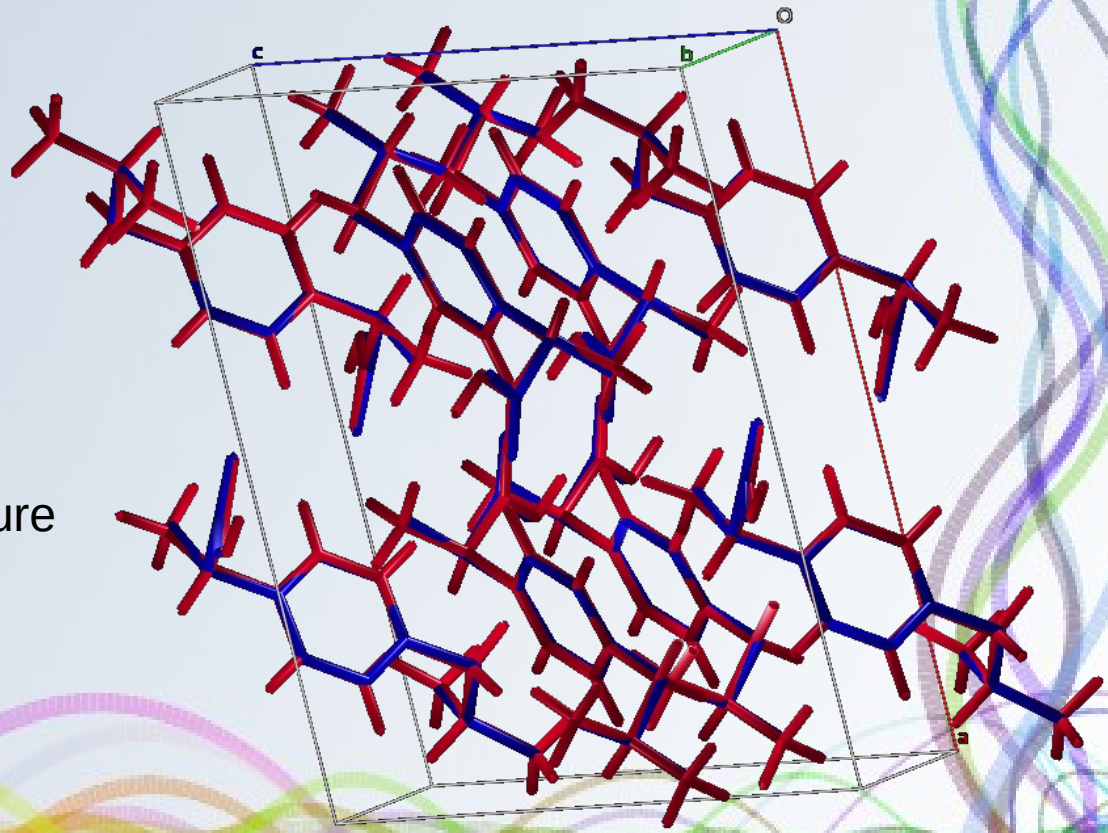
Assessing the solution with DFT-D

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

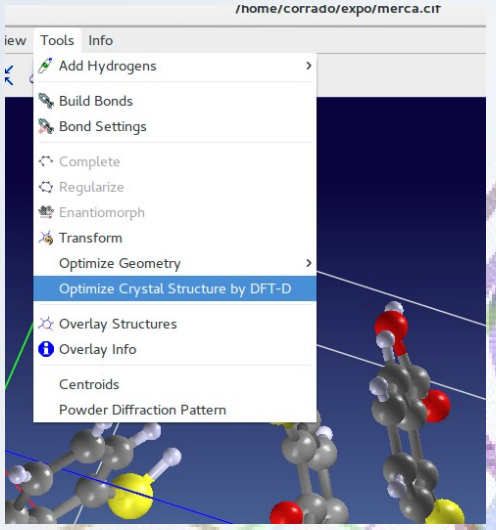
*Jacco ven de Streek *et al.* Validation of molecular crystal structures *Acta Cryst.* (2010). B**66**, 544–558

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



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

Other collaborators
R. Mallamo, F. Ciriaco

Computer resources
CINECA under ISCRA initiative, project HP10CQK3FF