

## Crystal structure solution by direct space methods



1/43

## Structure solution

**Structure  
solution**

### Traditional approaches:

- direct methods
- Patterson methods
- charge flipping

### Direct space methods

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

*Popular and widely available in recent years*

2/43

## Comparison

### Traditional approaches

- Do not use chemical knowledge
- Complexity of the problem depends on the number of atoms in the a.u.

### Direct space methods

- Can incorporate a massive amount of prior chemical information
- Complexity of procedure depends on the number of degrees of freedom. Can solve structures too large for being refined

3/43

## Comparison

### Traditional approaches

- Take advantage by using data of higher resolution
- Generally require less time to run

### Direct-Space methods

- High resolution is not needed. 2-2.5 Å resolution is enough
- Take time and patience. For large molecules: faster computer, run overnight

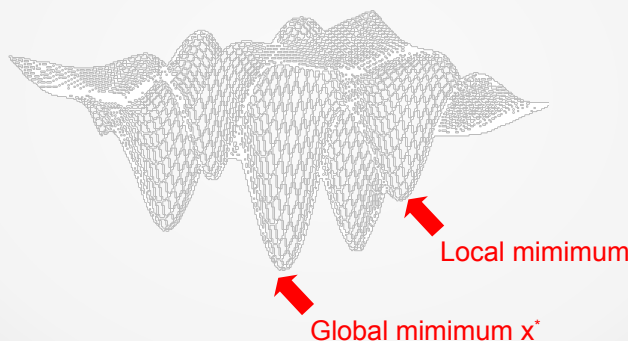
4/43

## 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 ( $\theta$ ,  $\phi$ ,  $\psi$ ), torsion angles ( $\tau_1$ ,  $\tau_2$ , ...  $\tau_n$ ) of molecular fragments




Local optimization methods  Structure refinement

Global optimization methods  Structure solution

5/43

## Global optimization methods

- Deterministic methods
  - Branch and Bound methods*
  - Cutting Plane methods*
  - Interval methods*
  - .....
- Heuristic strategies
  - Genetic Algorithms \**
  - Simulated Annealing \**** 
  - Tabu Search*
  - Ant Colony Optimization*
  - Particle Swarm Optimization \**
  - Bee Algorithms*
  - Firefly Algorithms*
  - Harmony Search*
  - Big Bang-Big Crunch \**
  - .....

*Widely used and with  
the largest impact*

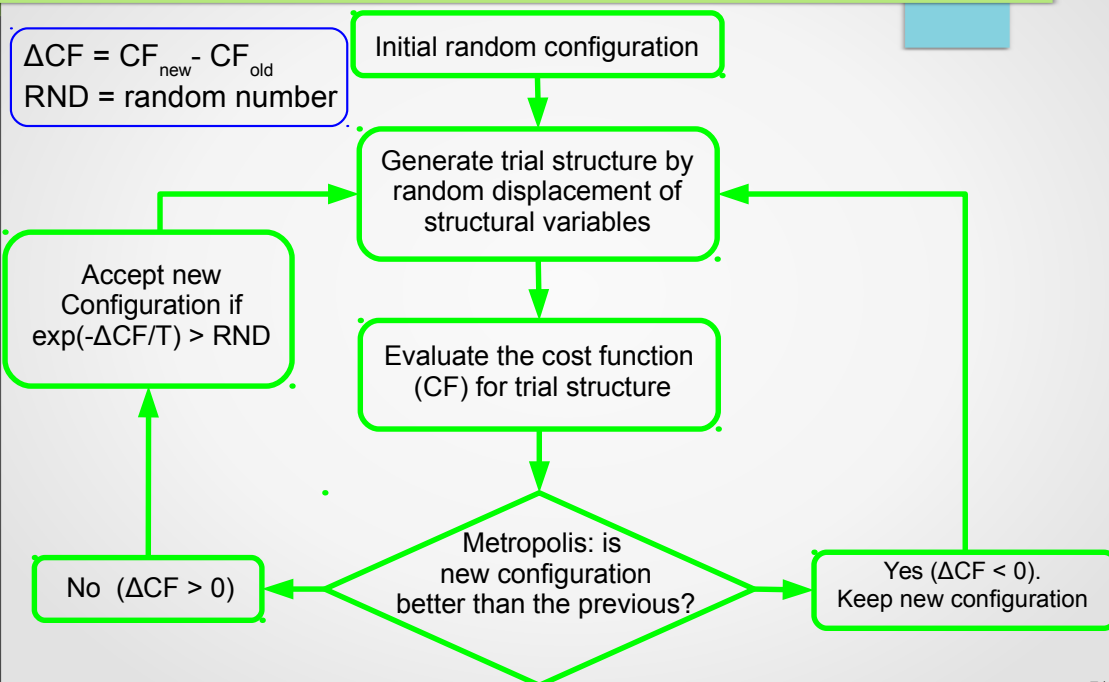
*Various modifications:*

- *parallel tempering*
- *adaptive simulated annealing*

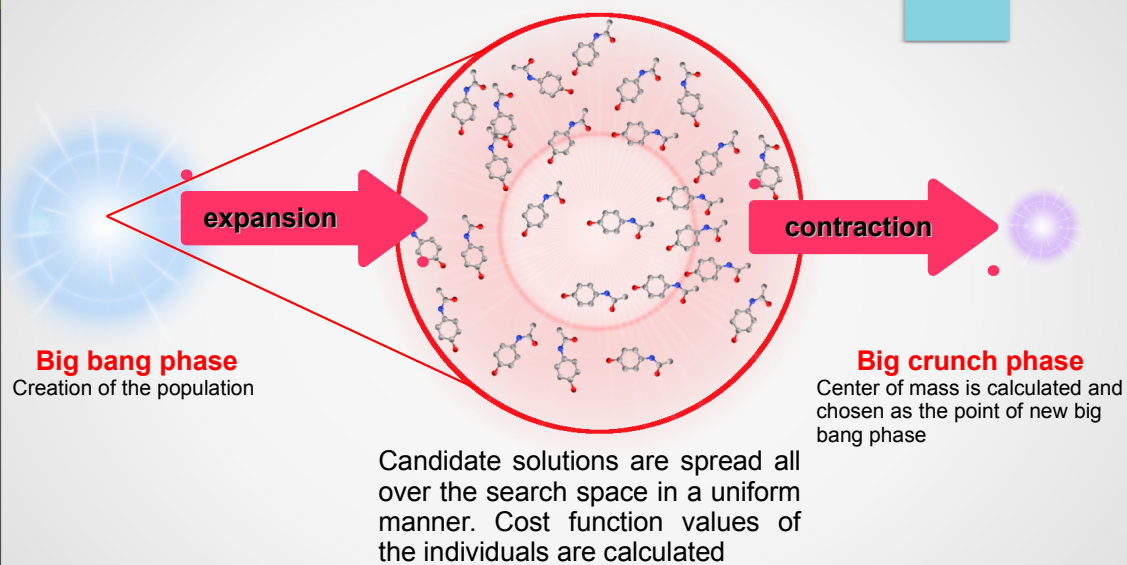
(\*) employed in solving crystal structure

6/43

## Simulated annealing



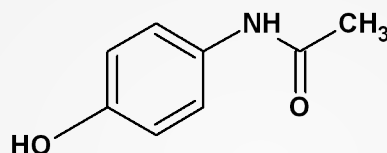
## Big bang-big crunch optimization



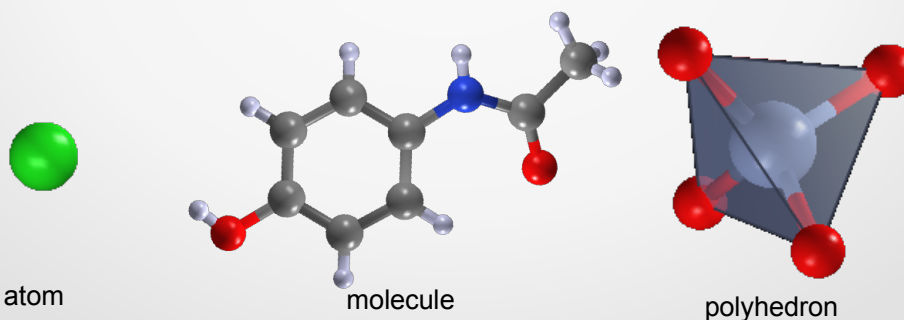
**Hybrid version:** SA is applied on selected individuals before the big crunch

## 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



9/43

## Building starting model

- Check for similar molecules** in databases or in the literature, use tables of standard bond lengths (volume C of International Tables for X-Ray Crystallography)
- Optimize molecular geometry** by computational chemistry programs

10/43

## 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)

11/43

## 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*

**2D molecular structures must be optimized before being used for structure solution**

12/43

## Geometry optimization

### *Three levels of theory*

- Molecular-mechanics force fields
- Semi-empirical methods
- *Ab initio* methods: self-consistent field (SCF), density functional theory (DFT)

*Programs:* MOPAC, Gamess, NWChem, Gaussian, ABINIT, etc.

13/43

## Molecule editor

### *A molecular 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*

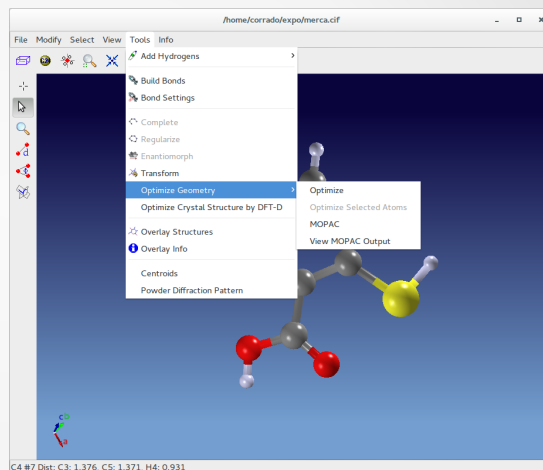
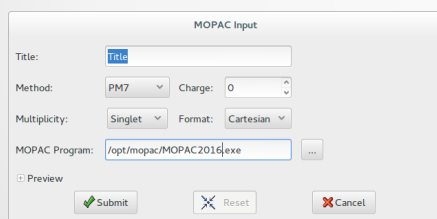
- ACD/ChemSketch - <http://www.acdlabs.com>
- Avogadro - [http://avogadro.openmolecules.net/wiki/Main\\_Page](http://avogadro.openmolecules.net/wiki/Main_Page)
- MarvinSketch - <http://www.chemaxon.com/products/marvin/>
- Gabedit: <http://gabedit.sourceforge.net/>

14/43

## Geometry optimization and EXPO2014

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

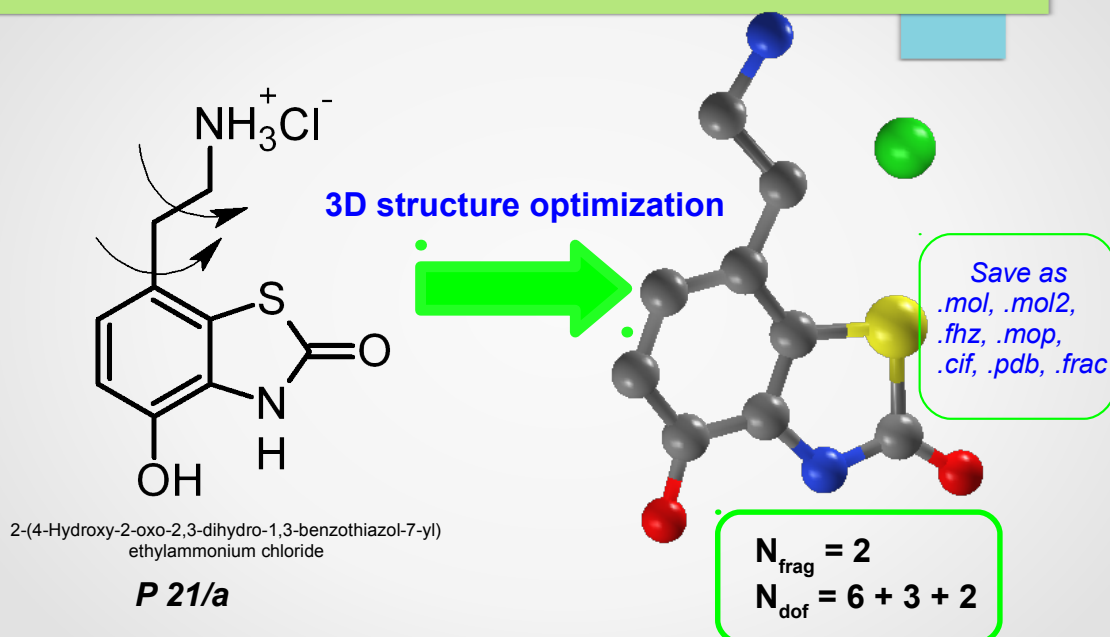
### Graphical interface of MOPAC



- Able to process input and output files of the most common quantum-chemistry packages

15/43

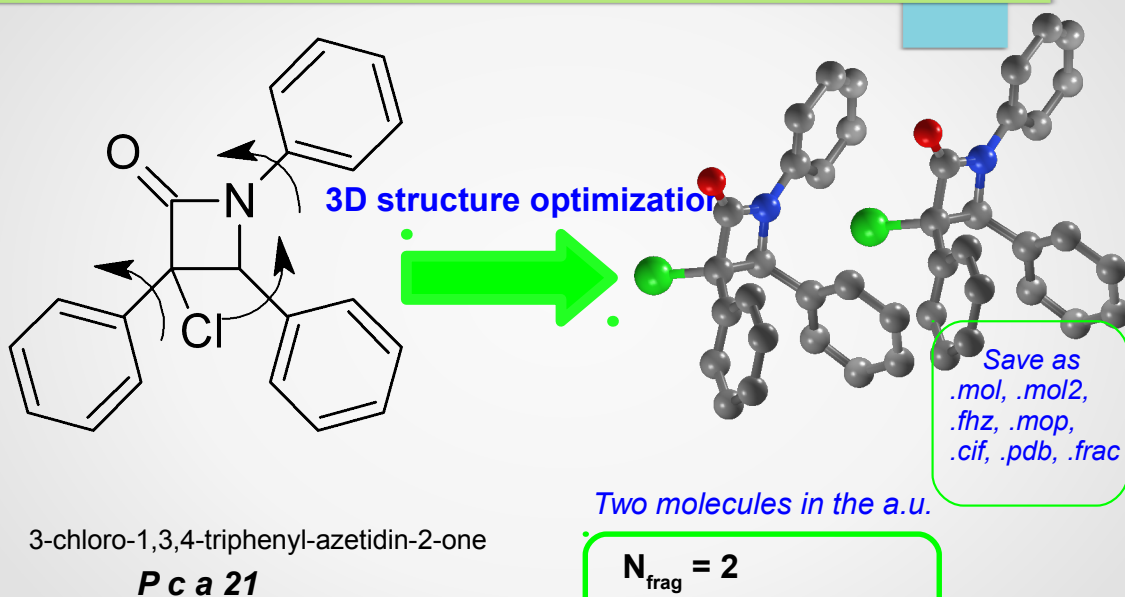
## Building starting model: example 1



16/43

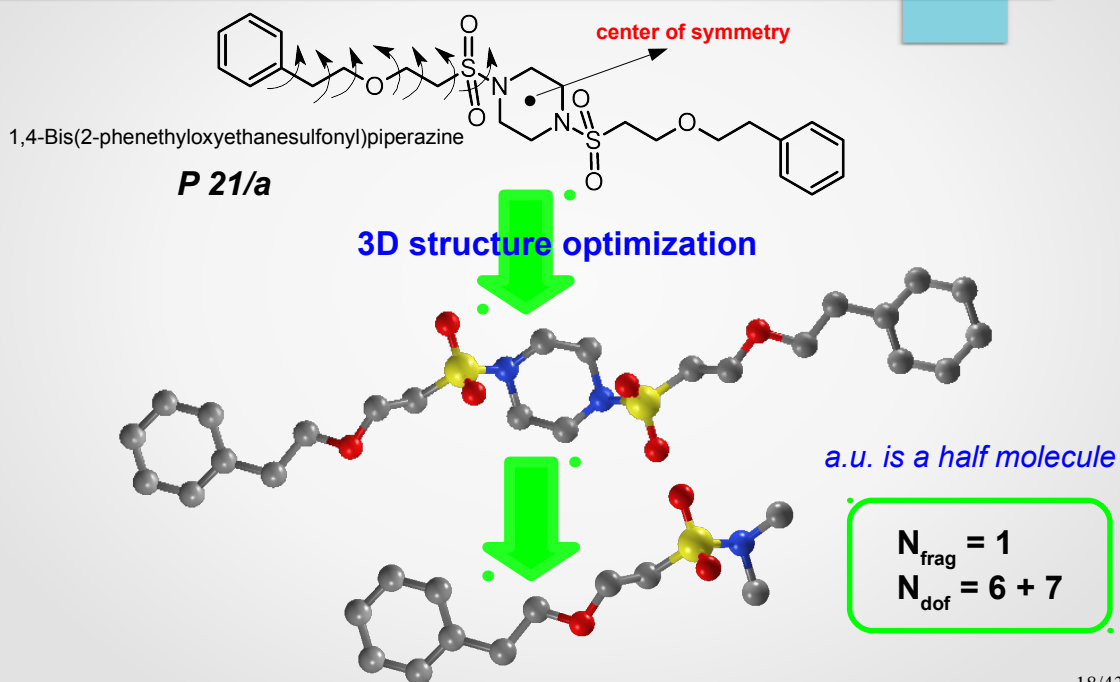


## Building starting model: example 2



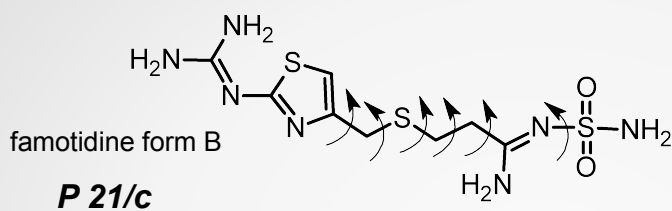
17/43

## Building starting model: example 3

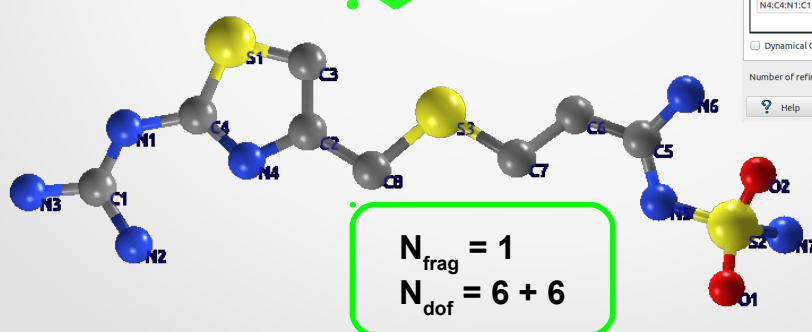


18/43

## Building starting model: example 4



3D structure optimization



SA conditions: External DOF Internal DOF Anti-bump

Torsion	Refine	Value	Lower	Upper
C4N1:C1:N2	<input type="checkbox"/>	-0.14	-180.00	180.00
N6:C5:N5:S2	<input 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:C4N1:C1	<input type="checkbox"/>	1.99	-180.00	180.00

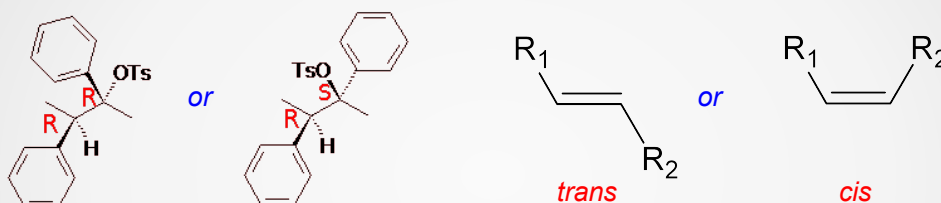
☐ Dynamical Occupancy Correction ☒ Atomic Parameters

Number of refined parameters: 13

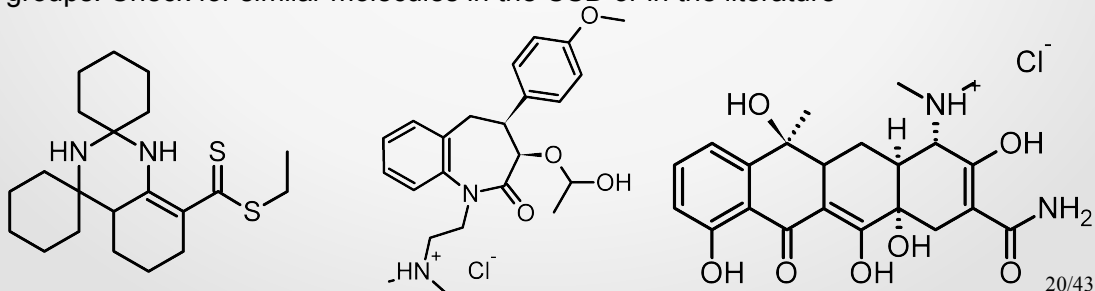
19/43

## Building starting model

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



- 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



20/43

## Building starting model

- Check  $V_{\text{cell}}/V_{\text{mol}}$

$$V_{\text{mol}} = \text{number of non-hydrogen atoms} \cdot 18 \quad (18 \text{ \AA}^3 \text{ rule})$$

or

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

$n_i$  = number of atoms of the  $i^{\text{th}}$  type in the structure  
 $v_i$  = volume contribution (in  $\text{\AA}^3$ ) for the  $i^{\text{th}}$  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  $\approx 15 - 20 \text{ \AA}^3$

*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*

21/43

## Cost function

- Weighted-profile R value ( $R_{\text{wp}}$ )

$$CF = R_{\text{WP}} = \sqrt{\frac{\sum_i w_i (y_{\text{obs}}(\theta_i) - y_{\text{calc}}(\theta_i))^2}{\sum_i w_i y_{\text{obs}}(\theta_i)^2}}$$

*Calculated in the range  $2\theta_0 - f \cdot \text{FWHM} < 2\theta_i < 2\theta_0 + f \cdot \text{FWHM}$ , with  $f=1$*

*There is no need to extract the intensities  $I_{hkl}$*

- Bragg intensity R value ( $R_{\text{B}}$ )

$$CF = R_{\text{B}} = \frac{\sum_h |I_h^{\text{obs}} - I_{hkl}^{\text{calc}}|}{\sum_h I_{hkl}^{\text{exp}}}$$

*$I_h^{\text{obs}}$  are the integrated intensities by Le Bail decomposition method*

*The advantage of this approach is that it offers a reduction in the computation time*

22/43

## Simulated annealing options

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

SA conditions | External DOF | Internal DOF | Anti-bump | Pattern

**General conditions**

Cost function: R weighted profile

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

Random seed: 1

Nr. of runs: 10

**Simulated annealing options**

Starting temperature: 10.000 ☒ automatic

Number of moves: 60 ☒ automatic

Temperature reduction factor: 0.900

Randomize parameters

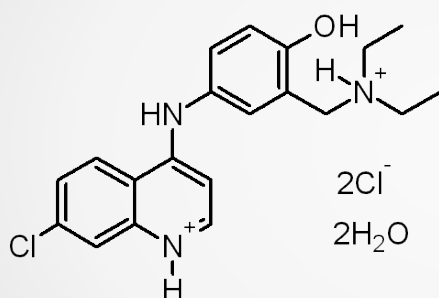
Number of refined parameters: 8

Help Quit Execute

23/43

## H atoms

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



2Cl<sup>-</sup>  
2H<sub>2</sub>O

amodiaquium dichloride dihydrate

**Number of atoms**

**External DoFs**

**Internal DoFs**

**Time**

**Success (%20 runs)**

**RMSD**

**with H atoms**

57

6+6+6+3+3=24

9

25h 59m

40

0.0116

**without H atoms**

29

6+3+3+3+3=18

6

10h 53m

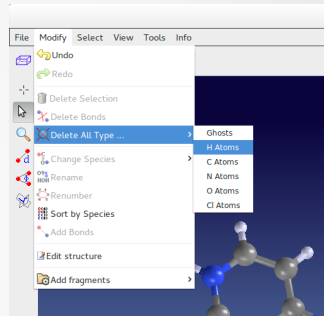
30

0.0294

24/43

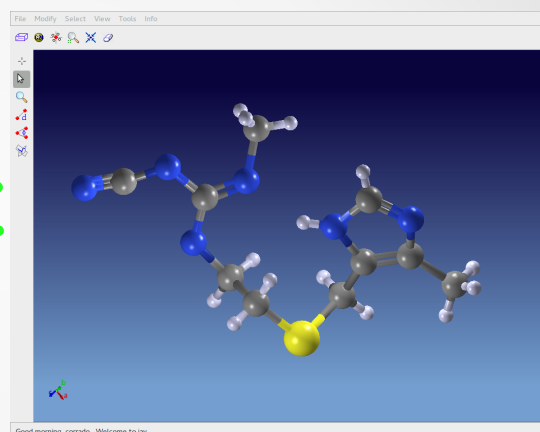
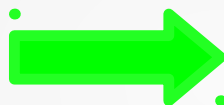
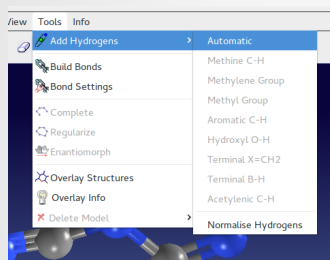
## H atoms

- Eliminating the H atoms reduces the number of atoms and DoFs, decreasing the time to evaluate CF for each trial structure
- How to delete the hydrogens?
- The interpretation of the correctness of crystal structure is far easier when the H atoms are present
- Collectively, their contribution to the overall scattering can be significant and they should be included in the molecular model whenever possible



25/43

## Hydrogen calculation



Hydrogen atoms are positioned geometrically

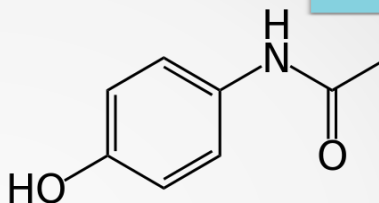
26/43

## Using expo2014 for direct-space solution

### By input file (\*.exp)

**File > Load & Go**

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



### Command-line usage

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

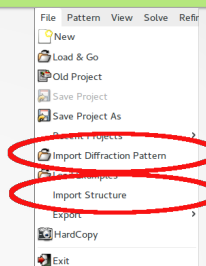
27/43

## Using expo2014 for direct-space solution

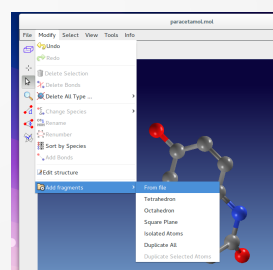
### By graphical interface

**File > Import Diffraction Pattern**

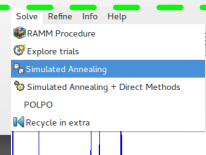
**File > Import Structure**



**Modify > Add Fragments**



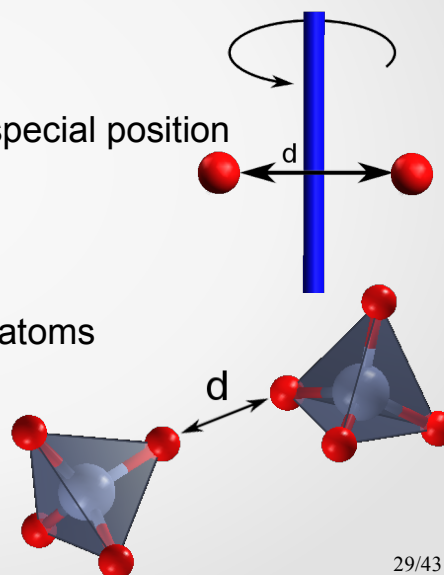
**Solve > Simulated Annealing**



28/43

## 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

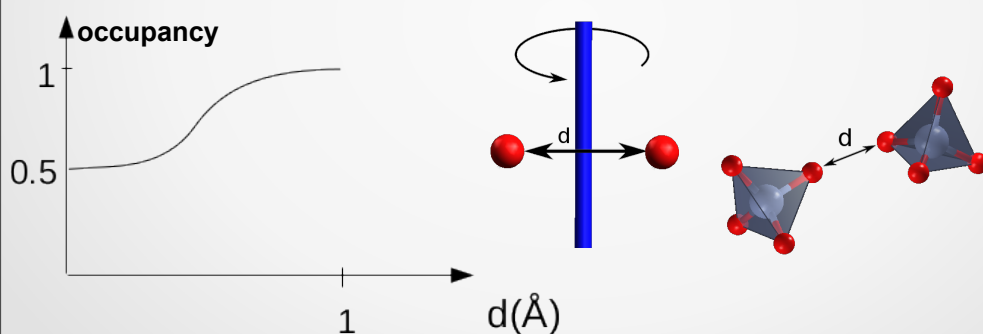


29/43

## 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

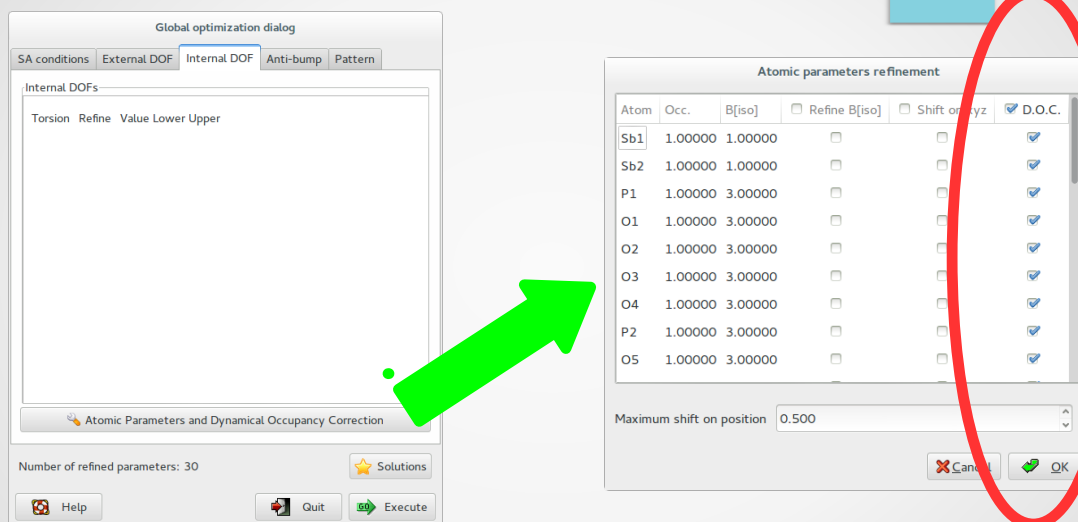
$$\text{Occupancy}^{-1} = 1 + \sum_{\text{neighbour}} |d_{\min} - d_i| \quad d_{\min} = 1 \text{ \AA}$$



DOC is able to merge the excess atoms automatically

30/43

## Dynamical occupancy correction (DOC)



Global optimization dialog

SA conditions External DOF Internal DOF Anti-bump Pattern

Internal DOFs

Torsion Refine Value Lower Upper

Atomic Parameters and Dynamical Occupancy Correction

Number of refined parameters: 30

Solutions

Help Quit Execute

Atomic parameters refinement

Atom	Occ.	B[iso]	Refine B[iso]	Shift on xyz	D.O.C.
Sb1	1.00000	1.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Sb2	1.00000	1.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
P1	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
O1	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
O2	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
O3	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
O4	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
P2	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
O5	1.00000	3.00000	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

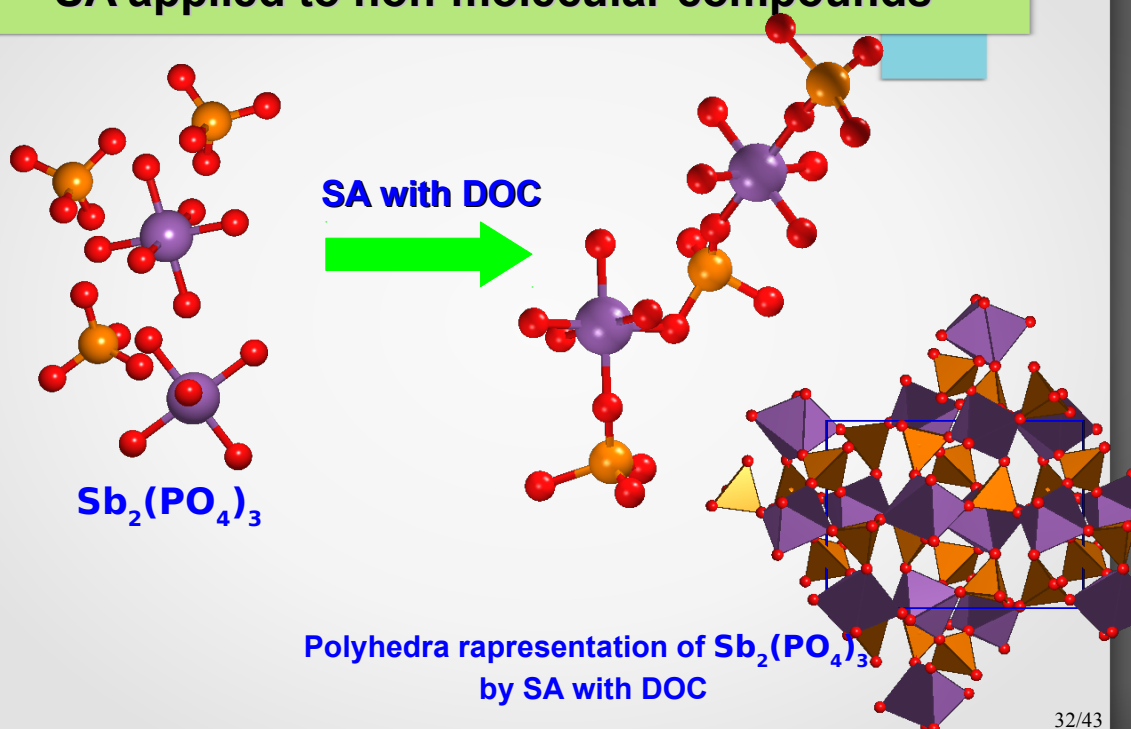
Maximum shift on position 0.500

Cancel OK

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

31/43

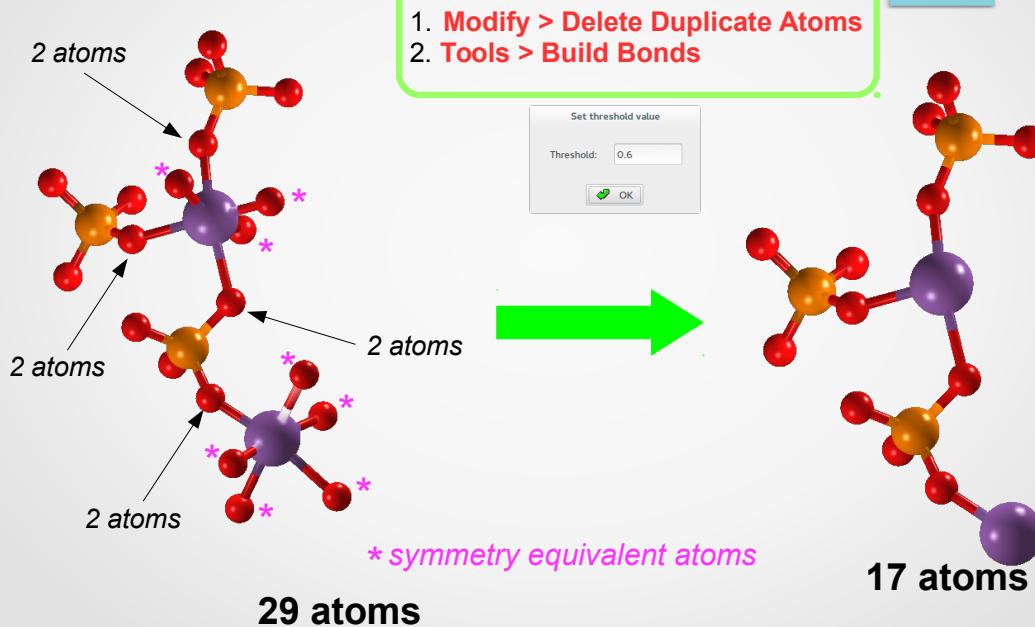
## SA applied to non-molecular compounds



32/43



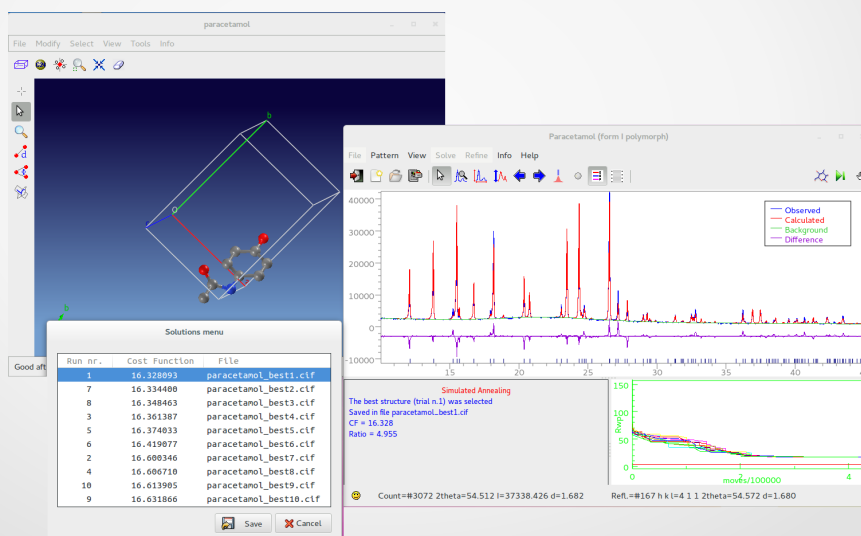
## Delete duplicate atoms



33/43

## Assessing the solution

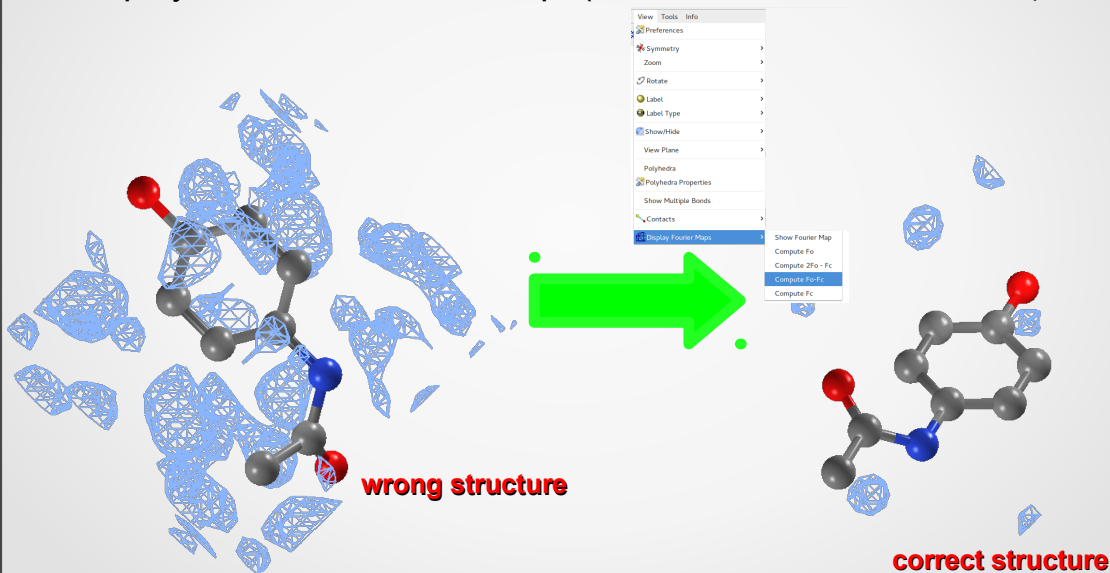
- Agreements factors
- Visual match between calculated and observed profile
- Reproducibility of solution



34/43

## Assessing the solution

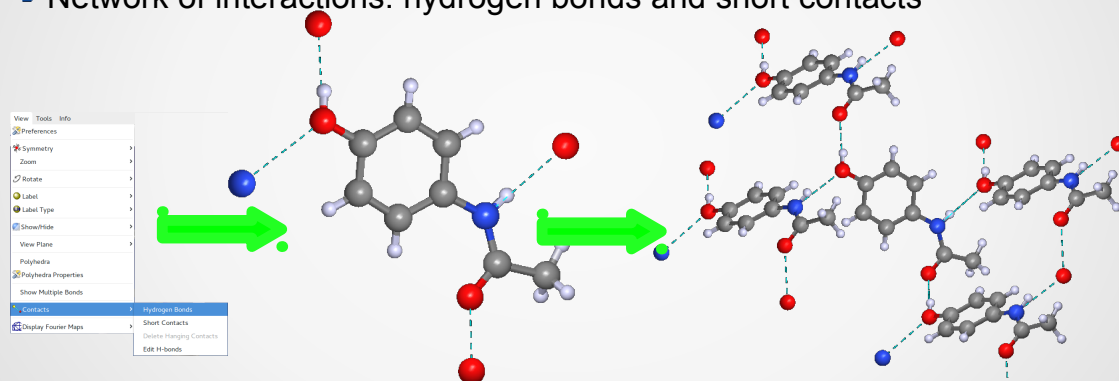
- Display difference Fourier map (at least 1.5 Å resolution data)



35/43

## Assessing the solution

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



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

Two programs:

- CASTEP (<http://www.castep.org/>),
- VASP (<https://www.vasp.at/>)

36/43

## Imposing restraints

$$CF = CF + \sum_i w_i \text{MAX}(0.0, |d_{\text{target}_i} - d_{AB_i}| - \text{tol}_i)^2$$

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

$d_{\text{target}_i}$  = ideal distance

$\text{tol}_i$  = permitted tolerance

$w_i$  = user supplied weight

*Directive:*

**res A B  $d_{\text{target}}$  tol w**

*$d_{\text{target}}$  and tol are optional:*

**res A B**

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

37/43

## Imposing anti-bumping restraints

$$CF = CF + \sum_i w_{ij} (d_{ij}^{\text{min}} - d_{ij}^{\text{model}})^{2k}$$

$$k = 2$$

$$d_{\text{model}} < d^{\text{min}}$$

$$d_{ij}^{\text{min}} = \epsilon (R_i^0 + R_j^0)$$

$$R^0 = \text{van der Waals radius}$$

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

SA conditions External DOF Internal DOF Anti-bump

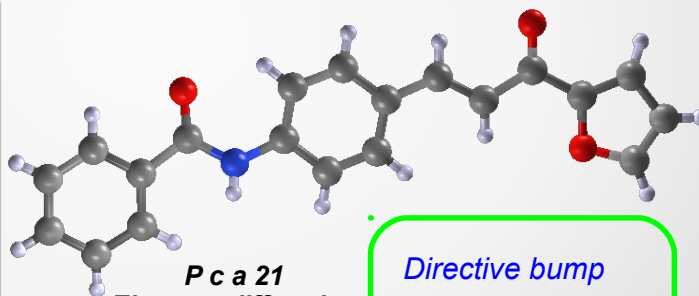
☐ None ☒ Only for C,N,O,S ☐ Select from list

Scale 1.000

List of anti-bump restraints

Atom 1	Atom 2	Target distance	Activate
H	H	1.962	<input type="checkbox"/>
H	C	2.511	<input type="checkbox"/>
H	N	2.376	<input type="checkbox"/>
H	O	2.349	<input type="checkbox"/>
C	C	3.060	<input checked="" type="checkbox"/>
C	N	2.925	<input checked="" type="checkbox"/>
C	O	2.898	<input checked="" type="checkbox"/>
N	N	2.790	<input checked="" type="checkbox"/>
N	O	2.487	<input checked="" type="checkbox"/>
O	O	2.462	<input checked="" type="checkbox"/>

Number of refined parameters: 13



*Directive bump*

.....  
%annealing  
bump  
.....

38/43

## When Structure Solution Fails

### Starting model is incorrect:

- chemical formula is wrong
- bond distances and angles 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 the ring conformation
- Improve your model with CSD or building package

39/43

## When Structure Solution Fails

### Poor quality diffraction pattern



### Solution:

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

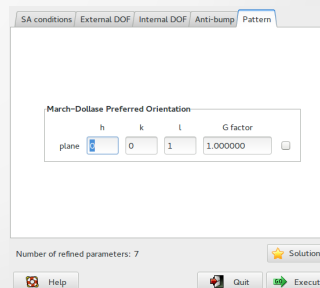
### Systematic problems in powder diffraction data

- Preferred orientation
- Ka2 contributions



### Solution:

- Collect new data
- Refine preferred orientation parameter



40/43

## When Structure Solution Fails

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



### Solution:

- Increase the number of moves and/or runs
- 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

41/43

## 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

42/43

**Thank you**  
for your attention

43/43

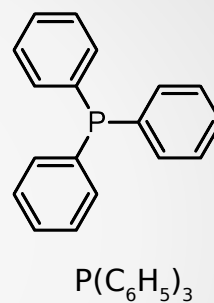
## Exercise 1 – structure determination of $P(C_6H_5)_3$

*Difficult level: easy*

- Perform all the steps of the crystal structure determination of the Triphenylphosphine  
Wavelength: 1.000972 (synchrotron radiation).  
Data file: exercise 1/pph3.xy

### Tips

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



- Compare the final model with the published model `pph3_pub.cif` in the folder `exercise 1/solution` (**Tools** → **Overlay structures**)
- Save the final model as `pph3.mol`, in mol file format

44/43

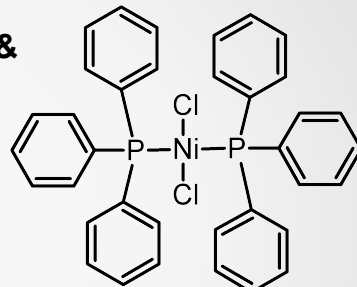
## Exercise 2 – structure determination of $C_{36}H_{30}Cl_2NiP_2$

Difficulty level: hard

### Part 1

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

**Tip**  
correct space group is  $P 2_1/c$



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

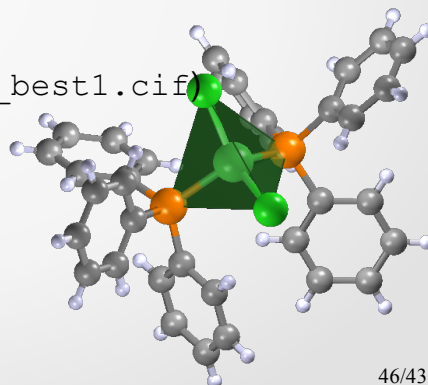
45/43

## Exercise 2 – structure determination of $C_{36}H_{30}Cl_2NiP_2$

Difficulty level: hard

### Part 2

- Solve the structure using the real space technique:
  - Import the `pph3.mol` file created in the exercise 1 (**File** → **Import**)
  - Add one Ni and one Cl in random position (**Modify** → **Add fragments** → **Isolated Atoms**)
  - Run Simulated Annealing (**Solve** → **Simulated Annealing**)
  - Apply DOC on Ni atom
- Compare the best solution of SA (`nickel_best1.cif`) with the `nickel_part1.cif`
- Use the polyhedra tool to visualize the tetrahedral coordination of Ni



46/43