

EXPO: Software Overview and guided Structure Solution and Refinement

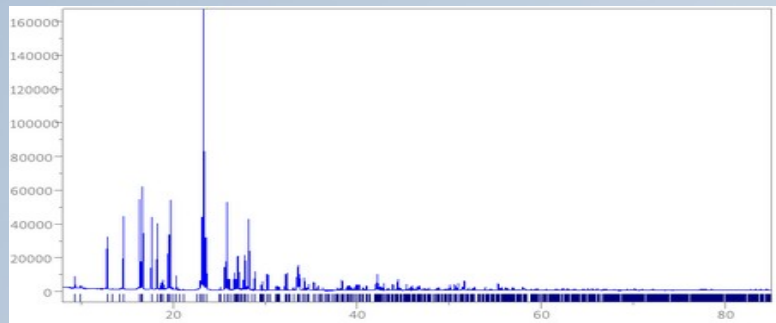
Corrado Cuocci

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

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

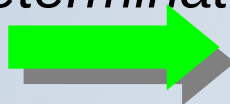


Crystal Structure Determination Process



Experimental powder diffraction pattern

*Indexing &
space group
determination*



unit cell & space group

Structure solution

the biggest challenge

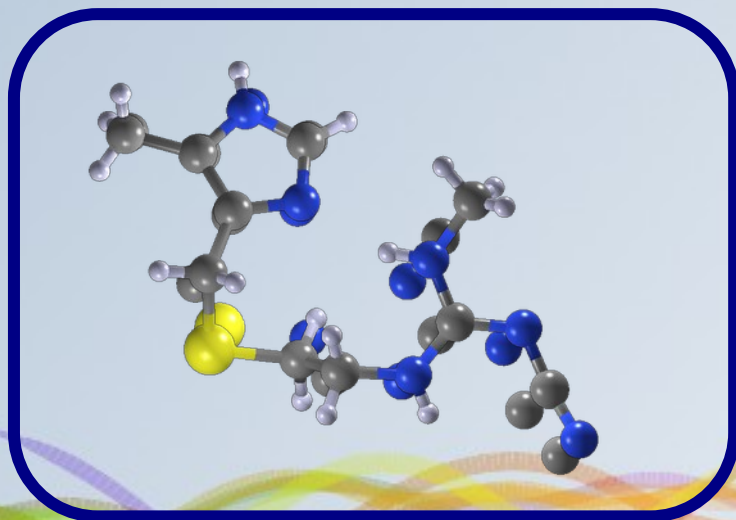


initial structural model

Rietveld method



final crystal structure

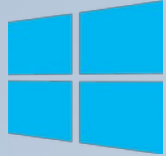


Getting Started with Expo



Download expo software from:

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



Windows: 64-bit: Windows 10, 11

File: expo2-xx.yy.zz-win64.exe



Mac OS X: macOS 12 (Monterey) and later, Apple Silicon (arm64)

File: expo-xx.yy.zz-arm64.dmg



GNU/Linux: Ubuntu 24.04 LTS, Ubuntu 22.04 LTS, other Debian-based distributions, Red Hat-based distributions (e.g., RHEL, Fedora)

File: expo-xx.yy.zz-codename_amd64.deb, expo2-xx.yy.zz-Source.tar.gz



On **Linux** and **macOS**, running DICVOL or McMaille during the indexing procedure requires Wine to be installed.

Getting Started with Expo



Installation notes:

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

The documentation is available at the link:

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

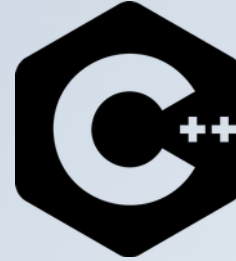
Tutorials:

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

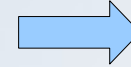
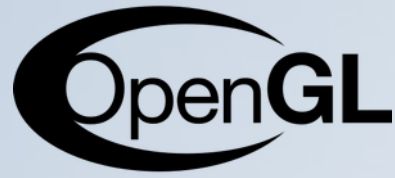
Technical Details



- Program written in Fortran and C++



- Program is built using OpenGL and Qt 6 libraries



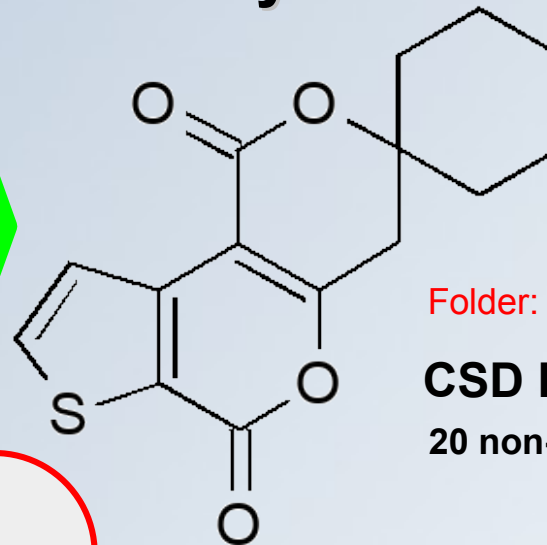
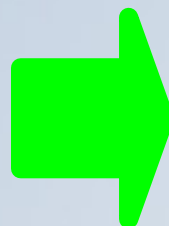
- Licensed under terms of GNU General Public License



Crystal structure determination by direct methods



Laboratory of Industrial and
Synthetic Organic Chemistry
(LISOC)

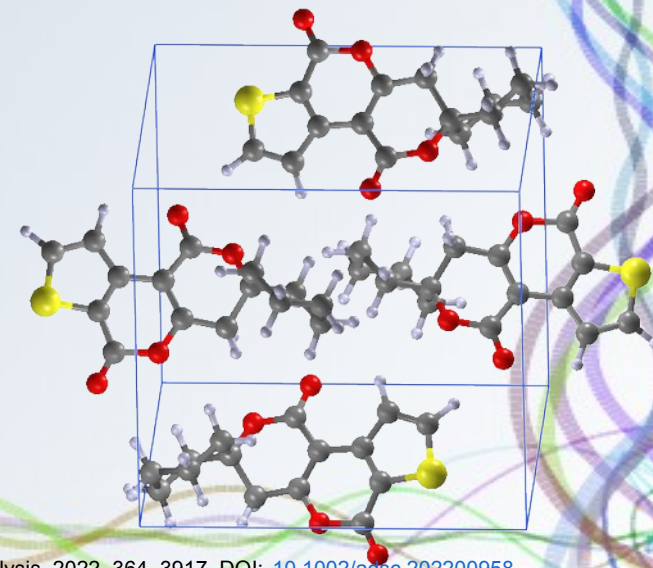
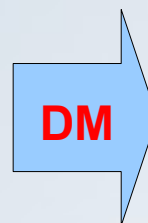
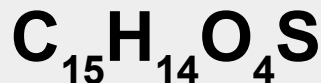
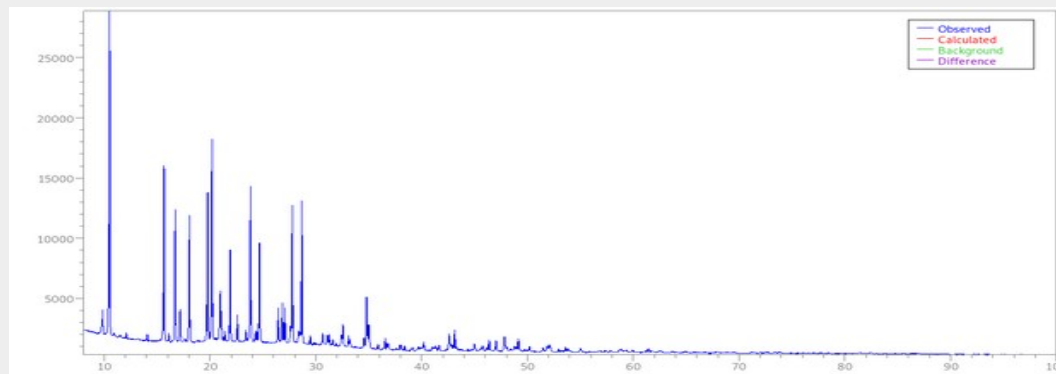


Folder: *examples/cr14*

CSD Entry: HEYHOM

20 non-H atoms

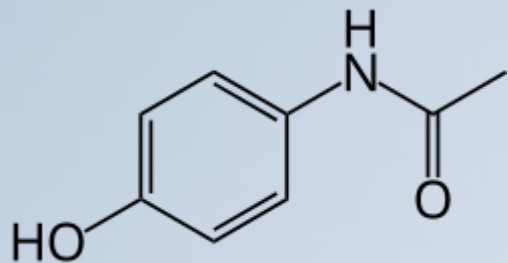
Required information



Structure of the EXPO Input File

The input consists of a sequence of **comments**, **commands** and **directives**. The commands start with a '%' character and directives must follow the related command. EXPO input file can be created using a text editor or by means of the GUI provided by the program.

Crystal structure solution of paracetamol



Space group: $P 2_1/n$

Cell: $a=7.100$ $b=9.380$ $c=11.708$ $\beta=97.42$

Chemical formula: $C_8H_9NO_2$

Cell Formula Units: 4

Data file name: paracetamol.xy

paracetamol.exp

```
!TEST on Paracetamol
%structure paracetamol
%job paracetamol
data
cell 7.100 9.380 11.708 90.0 97.42 90.0
space P 21/n
content (C8H9NO2) 4
pattern paracetamol.xy
%continue
```

Main EXPO Commands

%STRUCTURE *string*

This command is used to specify the project name and to define the names of the output files generated by the program (e.g. *loganin*, *paracetamol_form2*, *sample01*).

%JOB

This is a brief description of the project (e.g. *my beautiful structure, data provided by Mauro Gemmi, sample no. 3*).

%DATA

Specifies the experimental diffraction data and the basic information required for structure solution.

%CONTINUE

Used after %data, this command automatically continues the structure solution workflow, including integrated intensity extraction, structure solution by direct methods, and model optimization.

%FRAGMENT *file*

Used to import a structural model (for Simulated Annealing).

%SANNEAL

Starts the structure solution process using the Simulated Annealing algorithm.

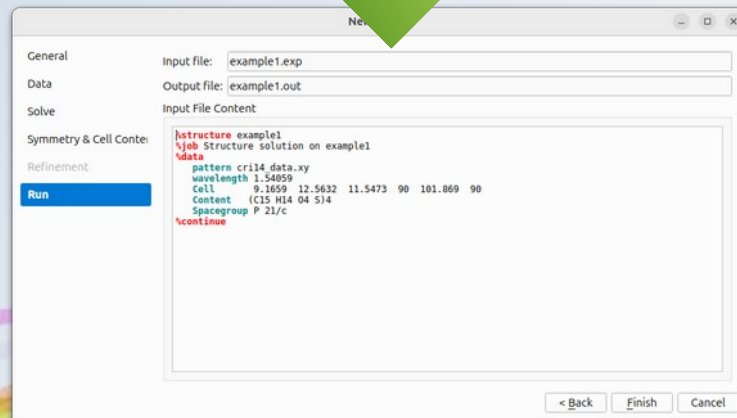
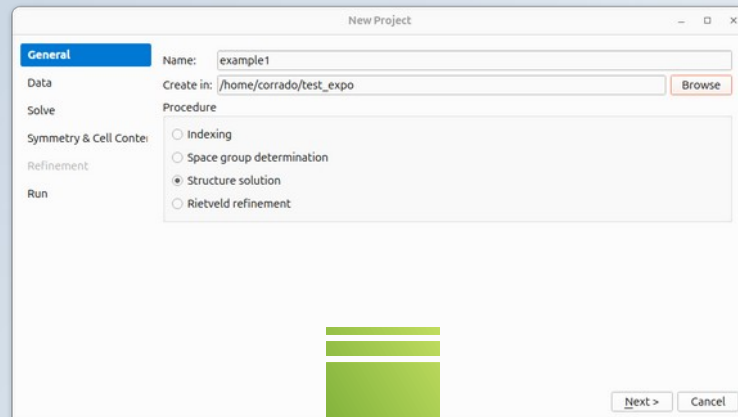
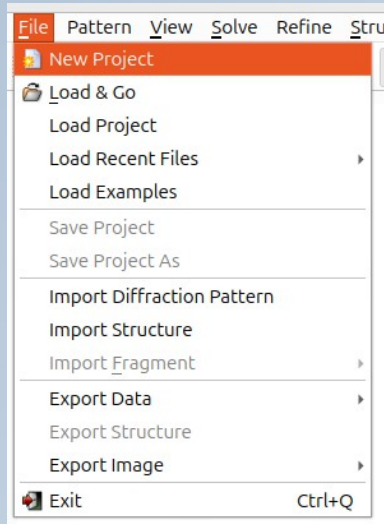
%CRYSTAL

Imports a crystal structure from a CIF file. It is generally used in conjunction with %rietveld to load the structural model to be refined.

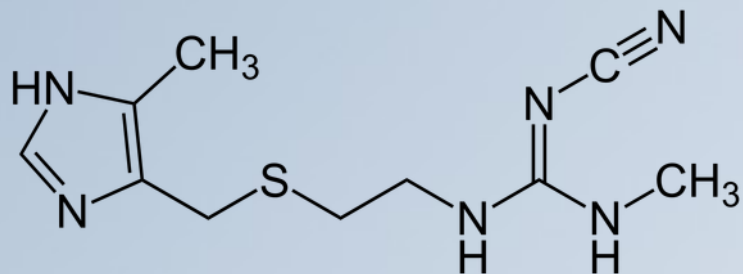
%RIETVELD

Performs Rietveld refinement of a structural model.

How to Make a EXPO Input File with the GUI



Crystal structure determination of cimetidine from powder synchrotron diffraction data

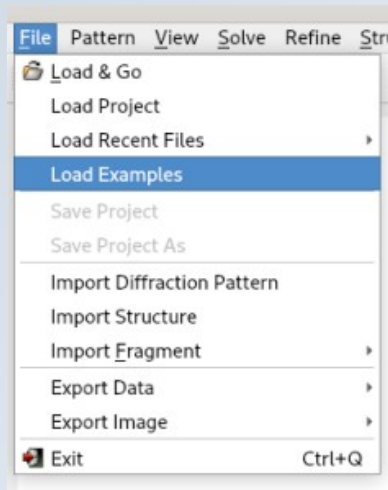
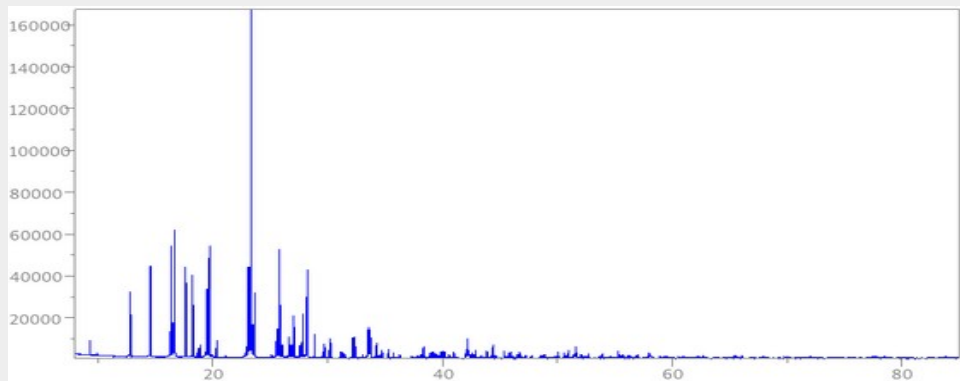


cimetidine

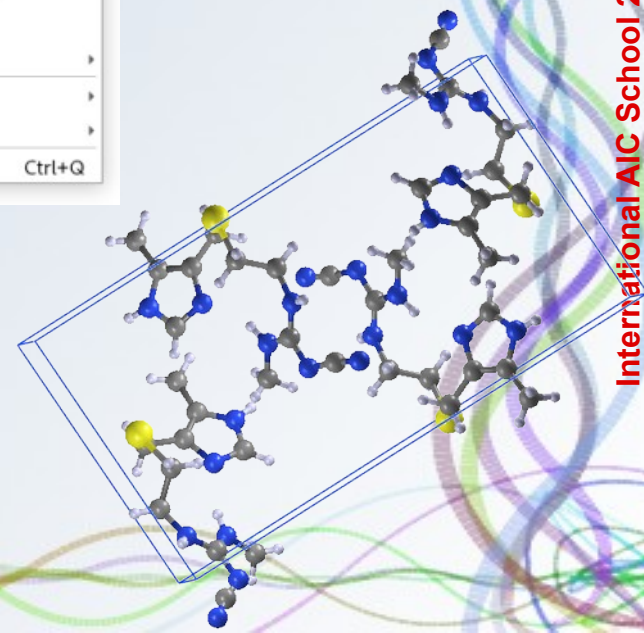
See tutorial 1:

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

Required information



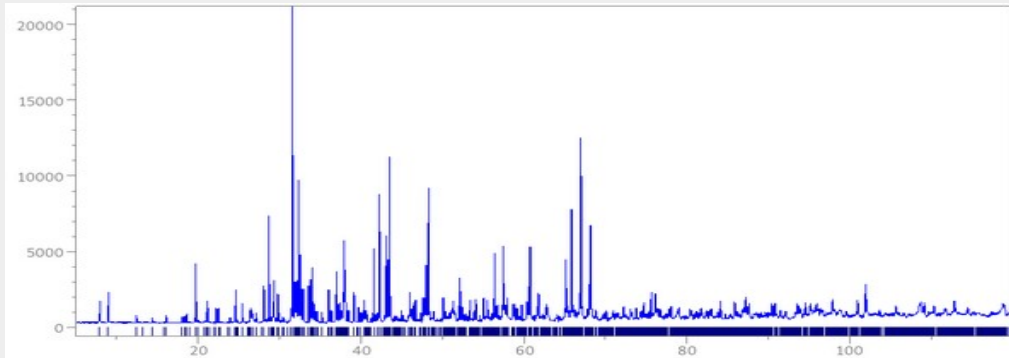
DM



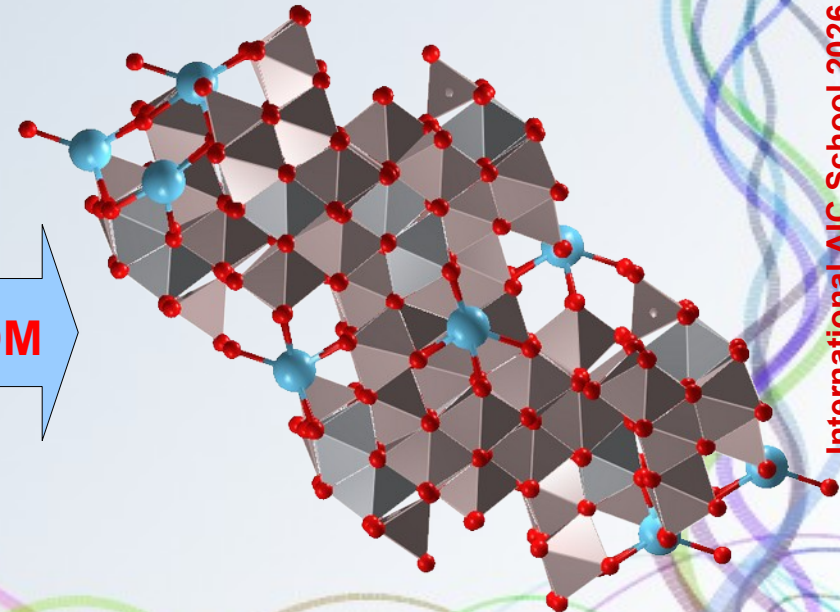
Crystal structure determination of $\text{LaTi}_2\text{Al}_9\text{O}_{19}$

```
%structure LaTi
%job LaTi from Acta Cryst. (2011) .B67,455-460
%data
cell 22.59355 10.99919 9.72968 90 98.5634 90
spacegroup C2/c
content (La Ti2 Al9 O19)8
pattern LaTi.rtv
%continue
```

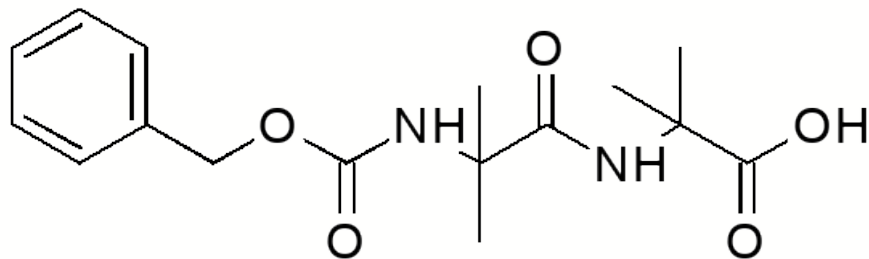
Required information



DM

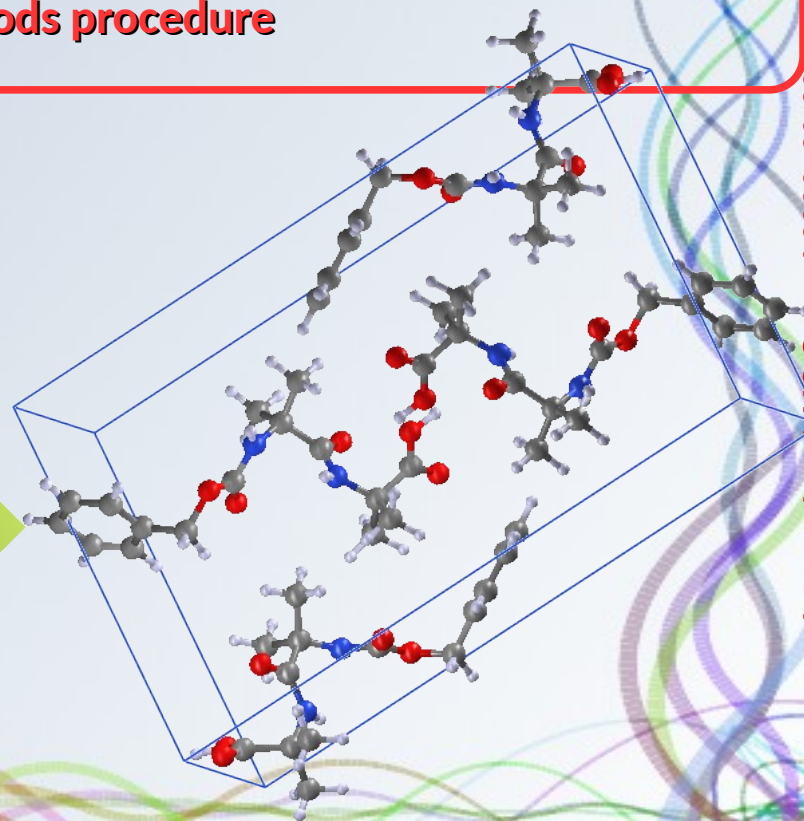
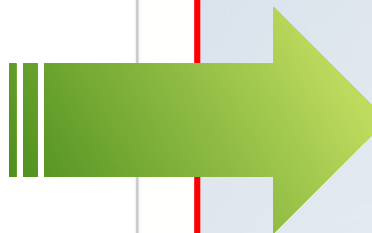
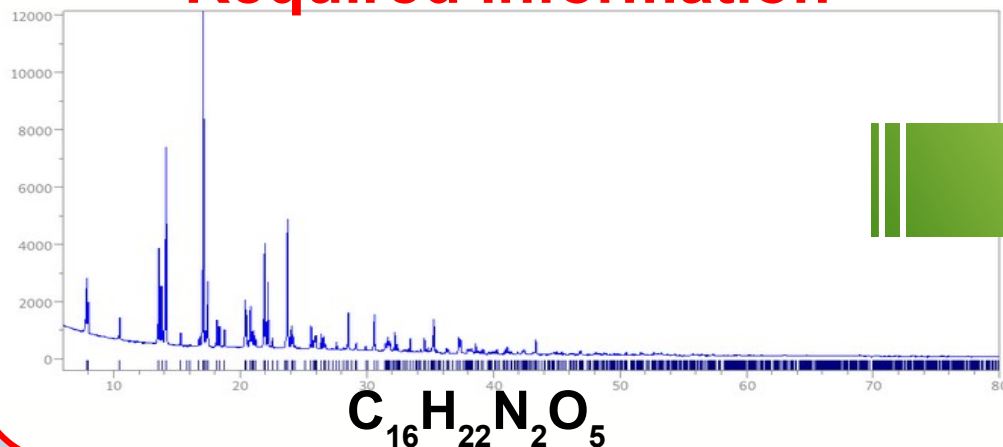


Crystal structure solution from powder diffraction data exploring all trials

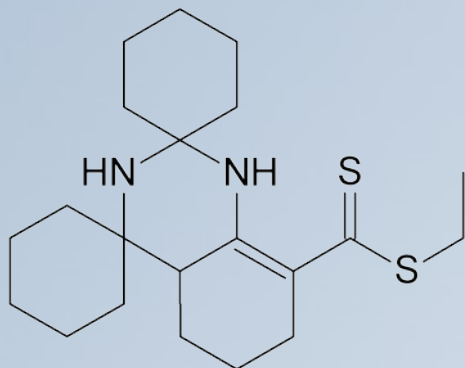


All trials were explored during the Direct Methods procedure

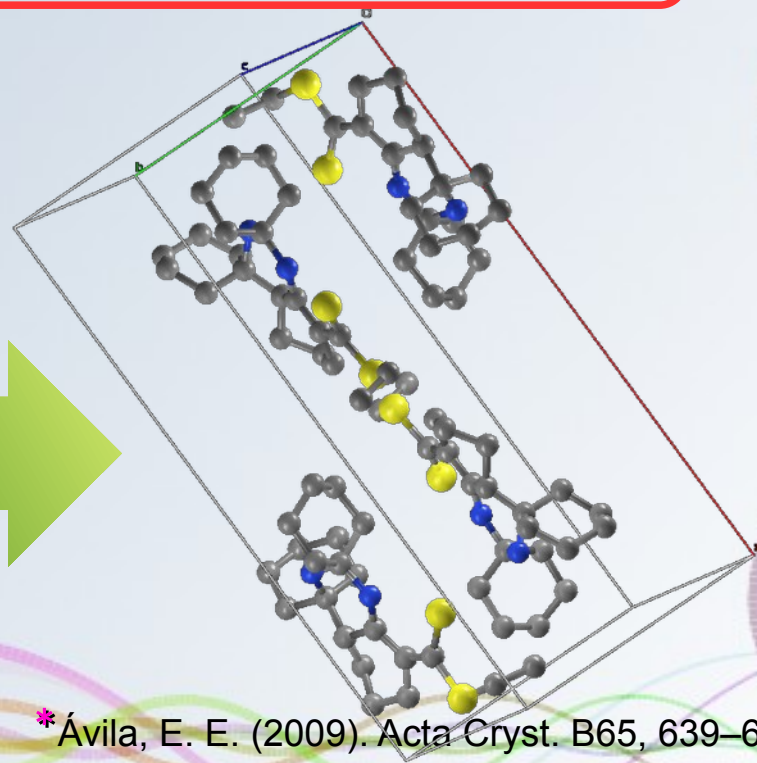
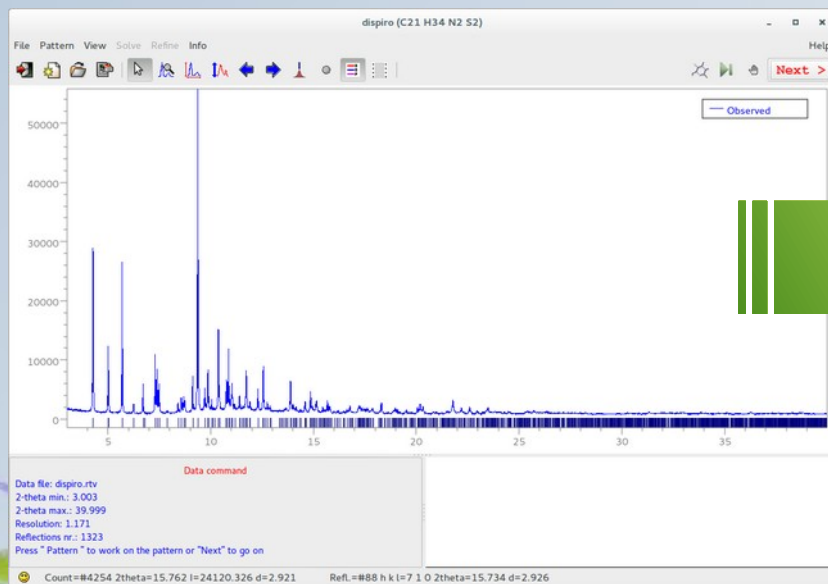
Required information



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

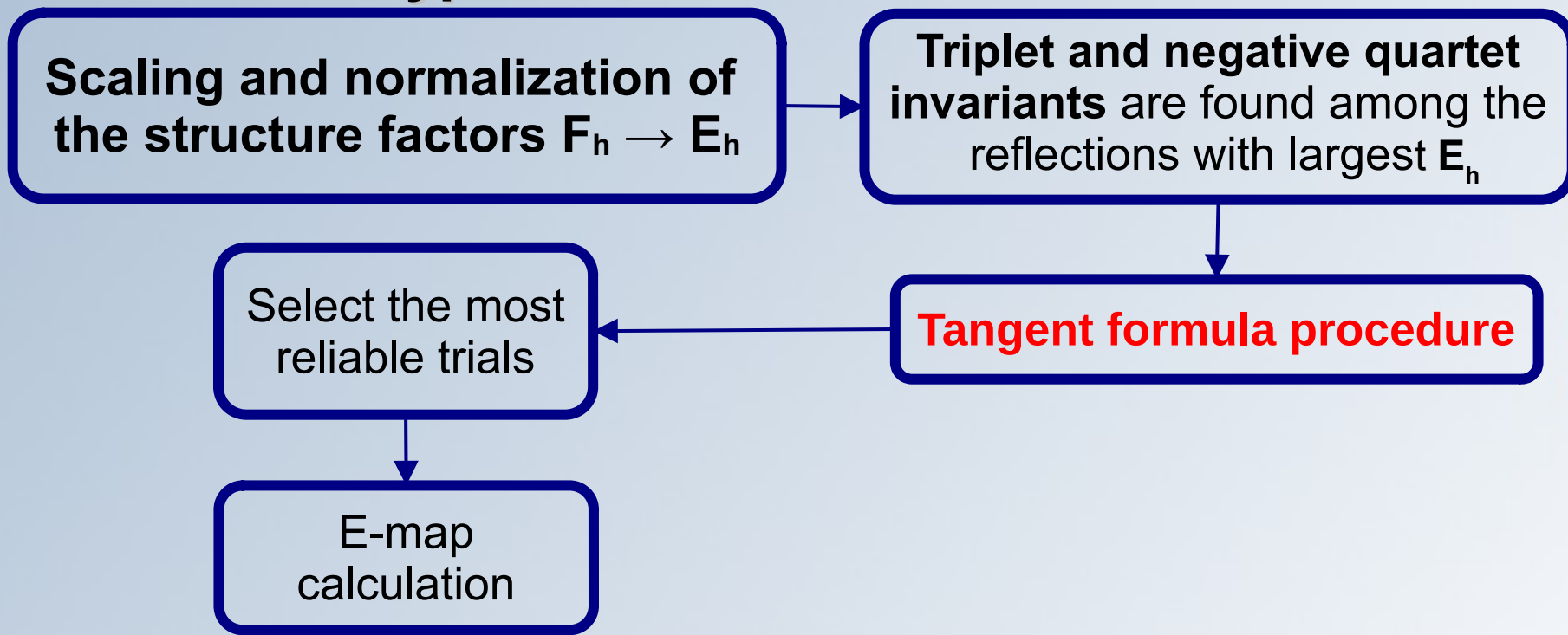


All trials were explored during the Direct Methods procedure



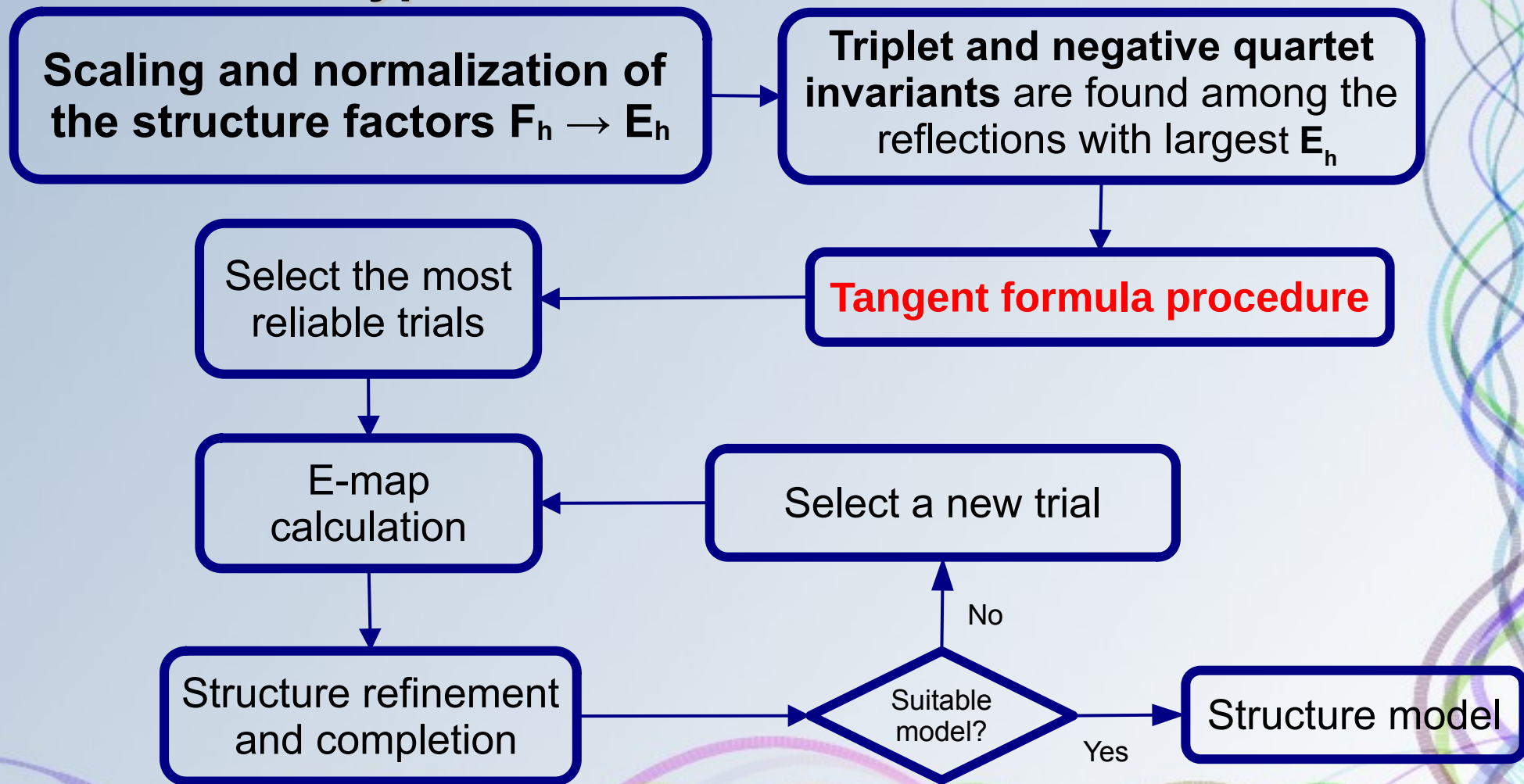
*Ávila, E. E. (2009). Acta Cryst. B65, 639–646

A Typical Direct Methods Procedure



$$\rho(\mathbf{r}) = FT^{-1}[w_h E_h] = \frac{1}{V_{cell}} \sum_{\mathbf{h}} w_h |E_h| \exp(i\varphi_h) \exp(-2\pi i \mathbf{h} \cdot \mathbf{r})$$

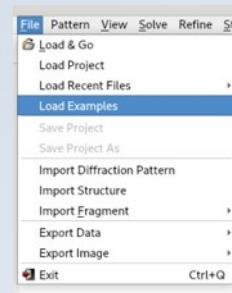
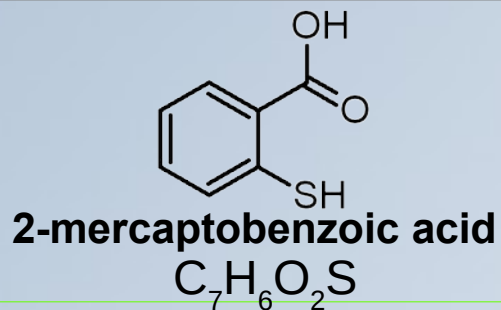
A Typical Direct Methods Procedure



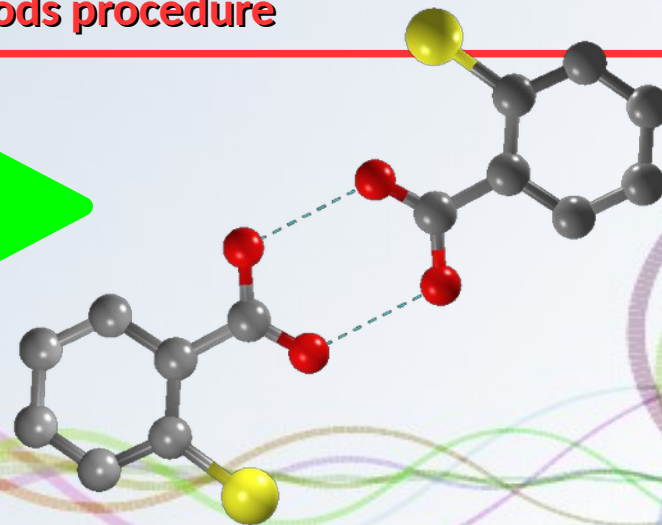
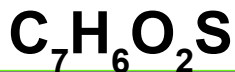
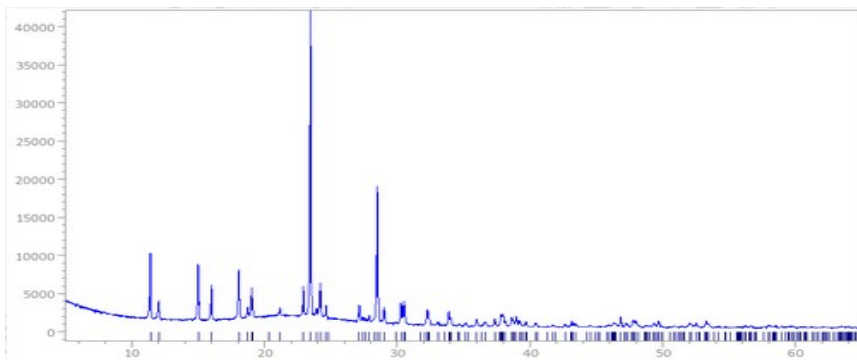
Crystal structure solution of 2-Mercaptobenzoic acid compound by direct methods

See tutorial 2:

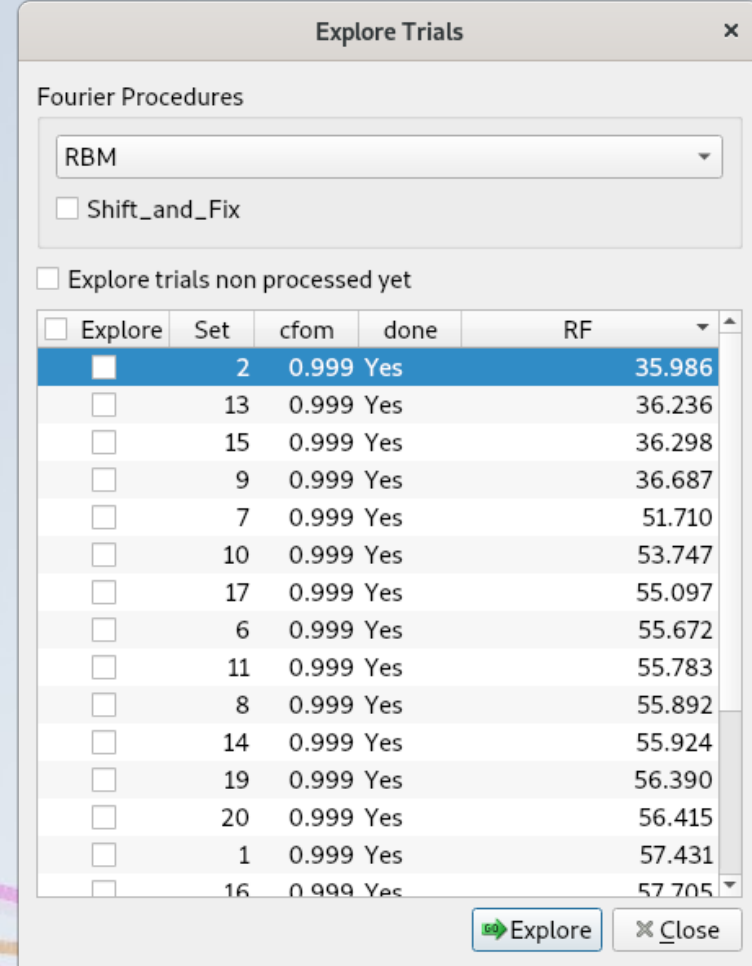
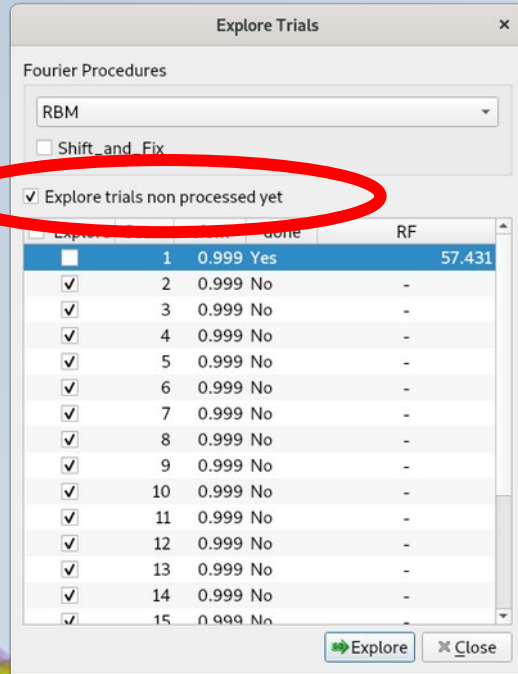
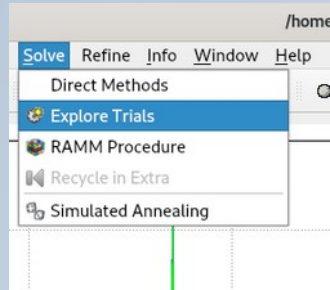
<https://www.ba.ic.cnr.it/softwareic/expo/tutorials/>



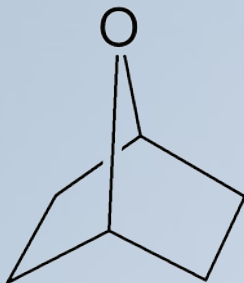
All trials were explored during the Direct Methods procedure



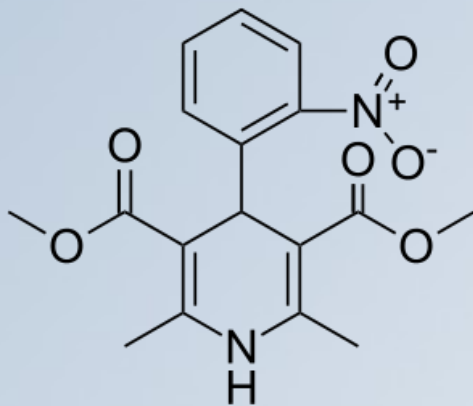
Exploring trials



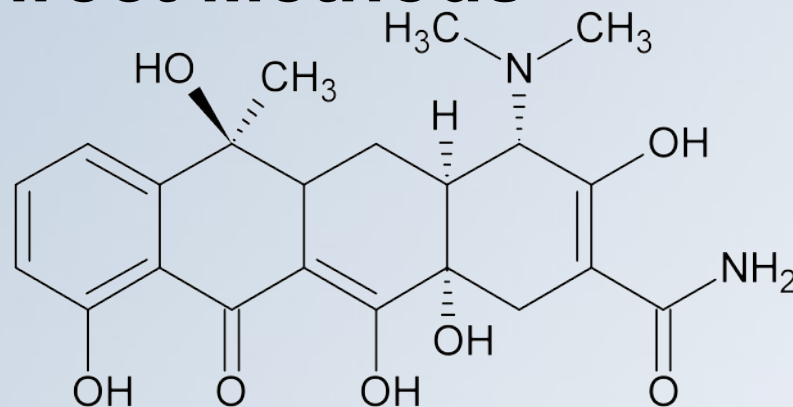
Limits of Direct Methods



phase III of oxanorbornane
 $Z'=4$, *synchrotron data*

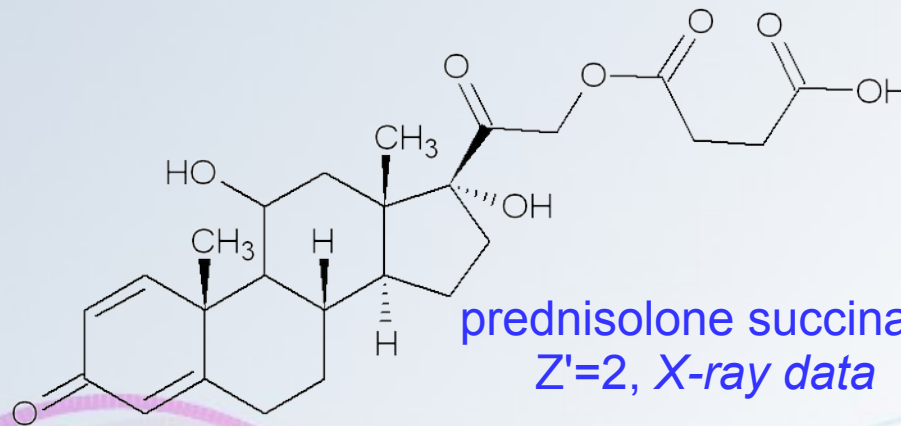


C polymorph of nifedipine
 $Z'=2$, *synchrotron data*



tetracycline hydrochloride
 $Z'=1$, *synchrotron data*

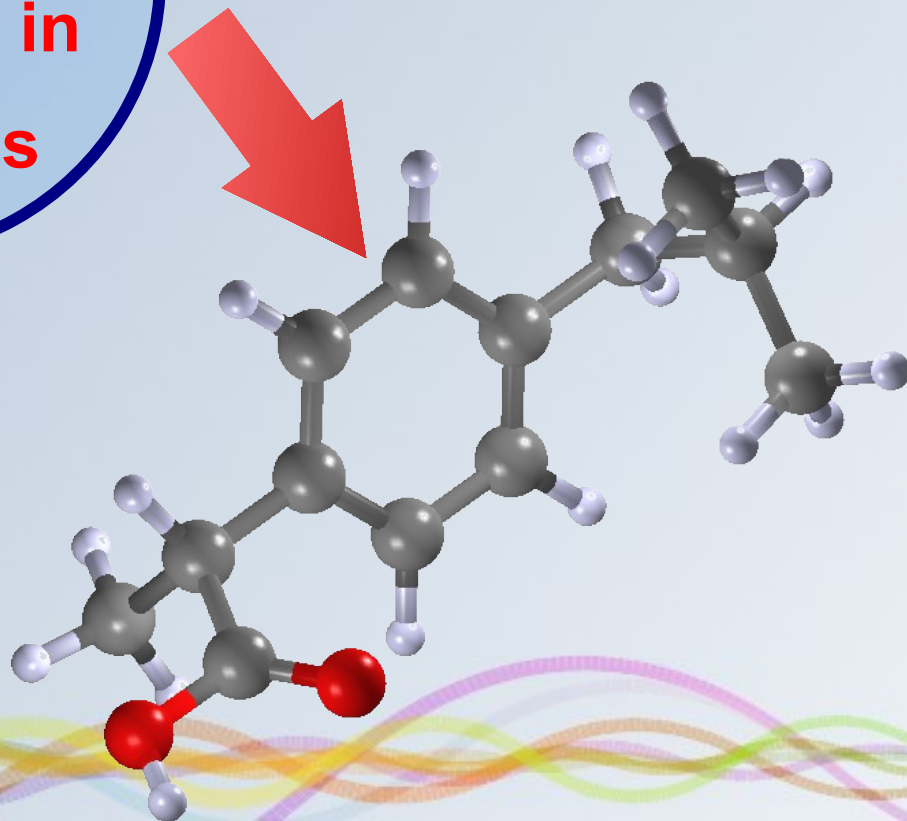
HCl



prednisolone succinate
 $Z'=2$, *X-ray data*

Building starting model

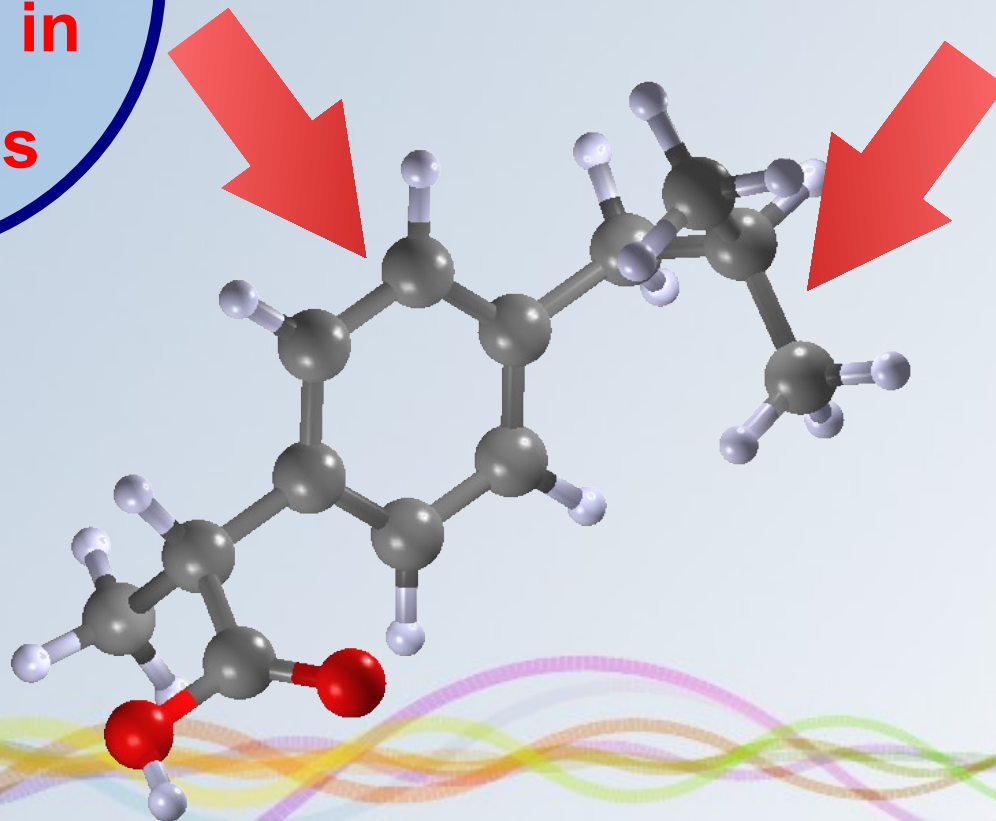
Check for
similar
molecules in
databases



Building starting model

**Check for
similar
molecules in
databases**

**Optimize
Molecular
geometry by
computational
chemistry**



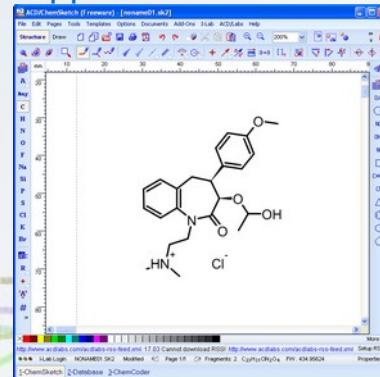
Molecule editor

A molecule editor allows

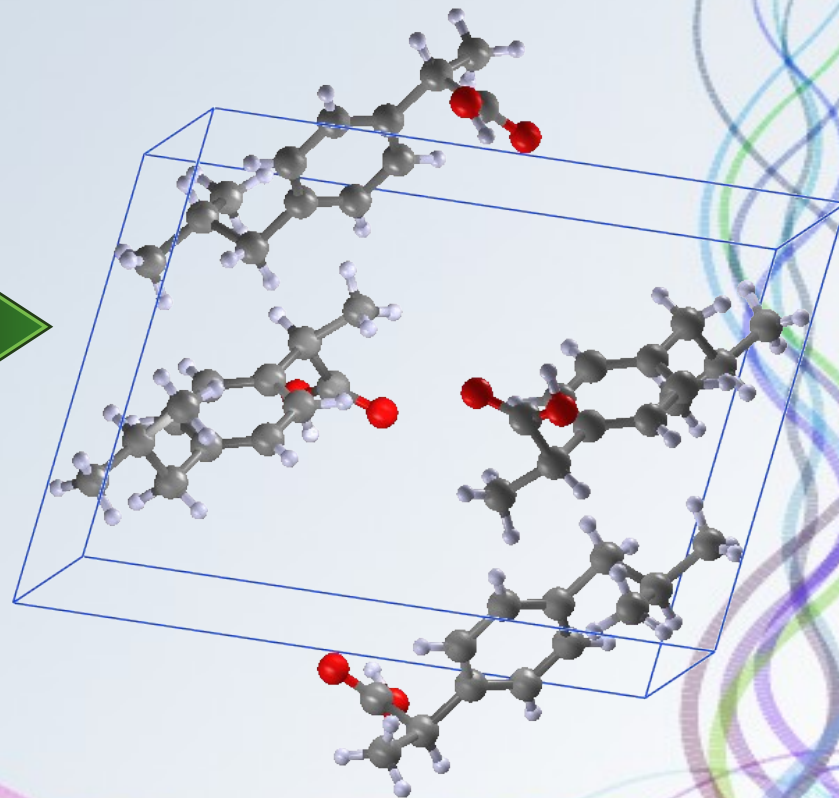
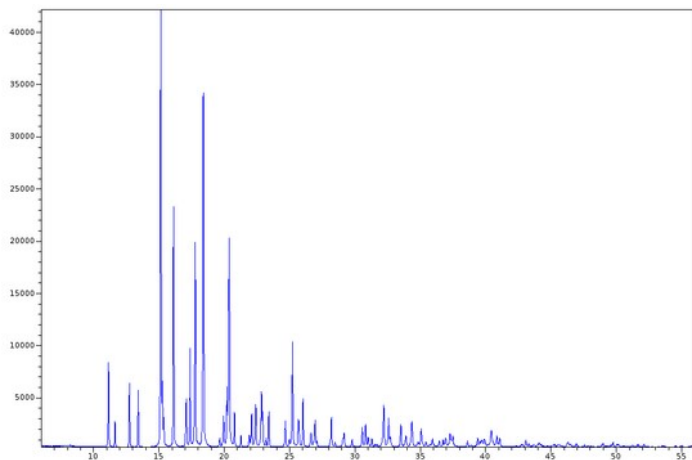
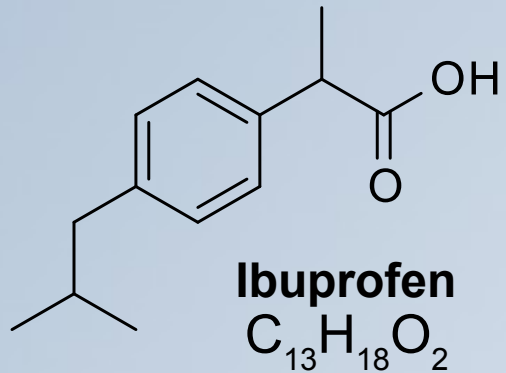
- Sketch molecules in 2D or 3D format
- Optimize the geometry by force fields method
- Create input file for the quantum-chemistry calculations
- Read output files of the most common computational packages

Some free available software

- ACD/ChemSketch <https://www.acdlabs.com/resources/free-chemistry-software-apps/chemsketch-freeware>
- Avogadro http://avogadro.openmolecules.net/wiki/Main_Page
- MarvinSketch <http://www.chemaxon.com/products/marvin/>
- Gabedit <http://gabedit.sourceforge.net/>



Crystal structure solution of ibuprofen by direct space method

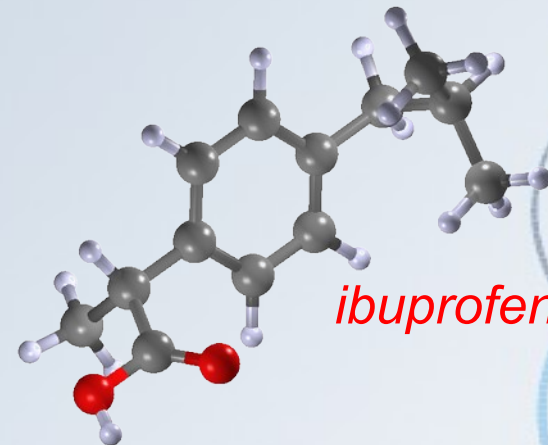


Using expo2014 for direct-space structure solution

- By input file (*.exp)

File → Load and Go

```
%Structure ibuprofen
%Job ibuprofen - C13H18O2
%Data
  Cell 14.6682 7.8933 10.7323 90 99.4092 90
  SpaceGroup p 21/c
  Pattern ibuprofen.dat
  Wavelength 1.405
%fragment ibuprofen.mol
%sanneal
```



- Command-line usage

expo ibuprofen.exp

or

expo ibuprofen.exp -nogui

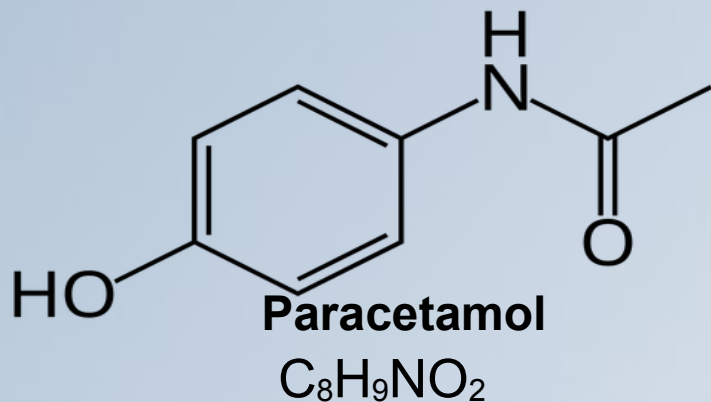
or

expo ibuprofen.exp -auto

Usage of the command %fragment

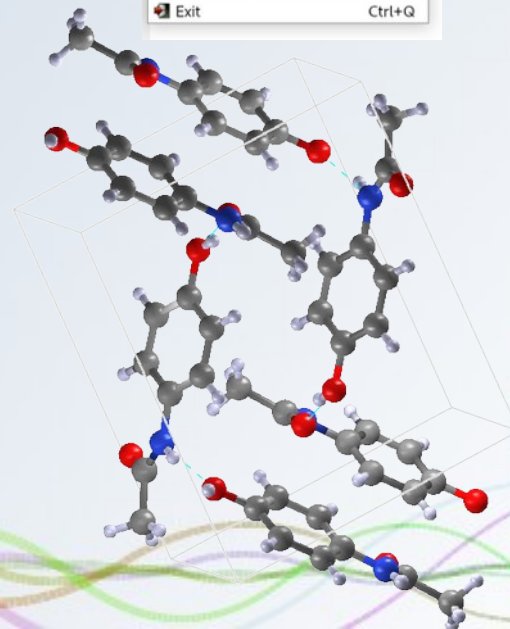
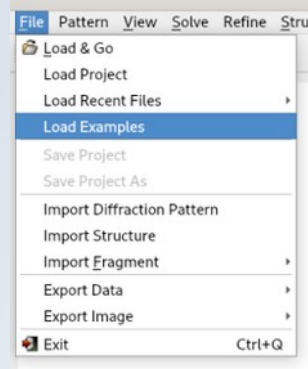
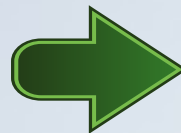
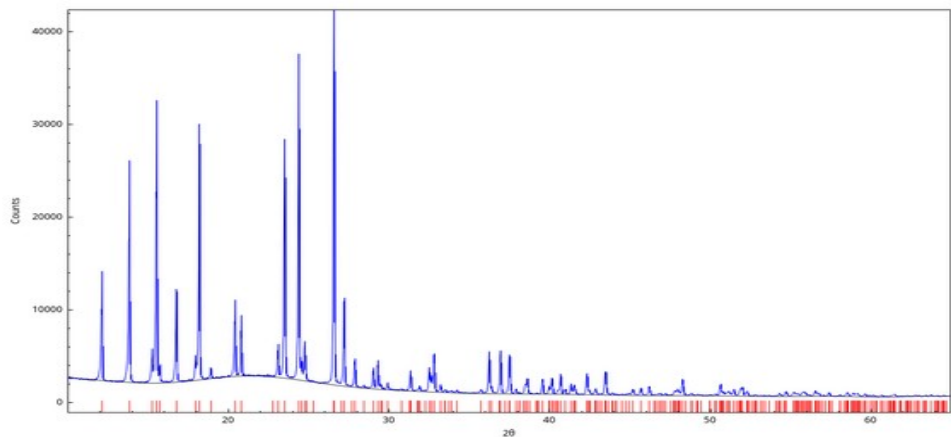
```
%fragment tetra AtC AtV [dist]
%fragment octa AtC AtV [dist]
%fragment square AtC AtV [dist]
%fragment cube AtC AtV [dist]
%fragment trigonal AtC AtV [dist]
%fragment prism_tetra AtC AtV [dist]
%fragment prism_trig AtC AtV [dist]
%fragment icosah AtC AtV [dist]
%fragment atoms chem_formula
%fragment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

Crystal structure solution of paracetamol by direct space method



See tutorial 3:

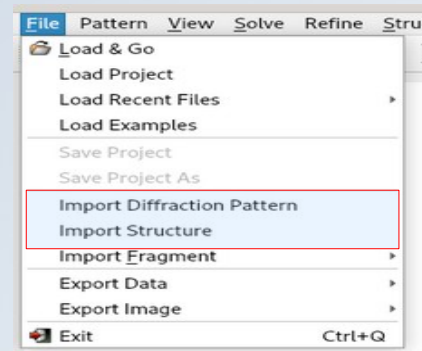
<https://www.ba.ic.cnr.it/softwareic/expo/tutorials/>



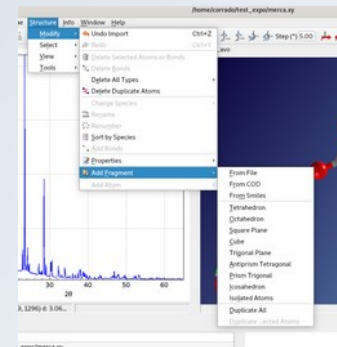
Using EXPO for direct-space solution

- By graphical interface

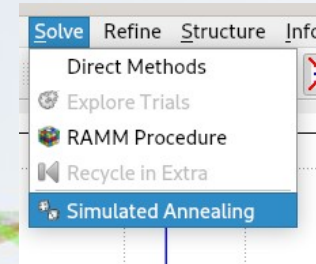
- File → Import Diffraction Pattern
- File → Import Structure



- Structure → Modify → Add Fragments

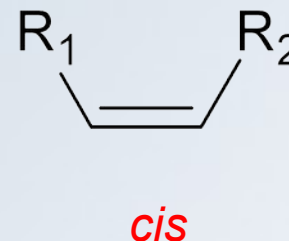
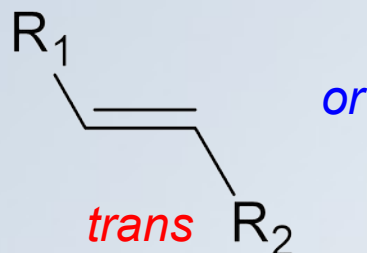
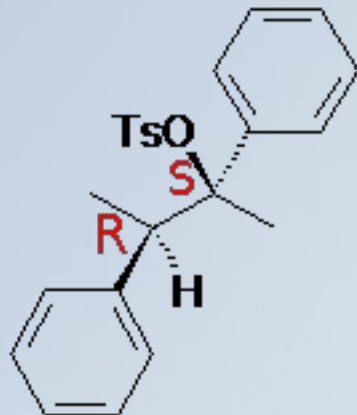
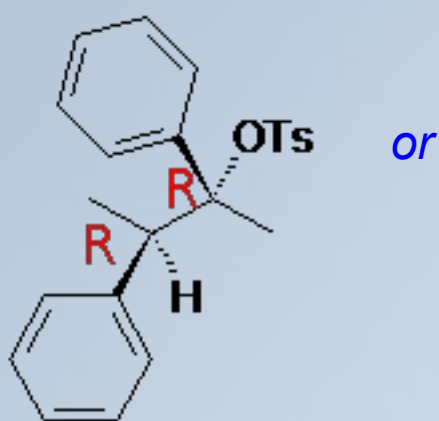


- Solve → Simulated Annealing

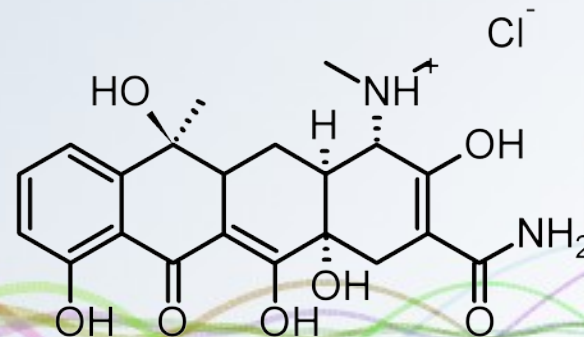
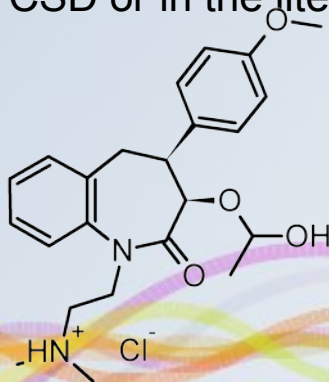
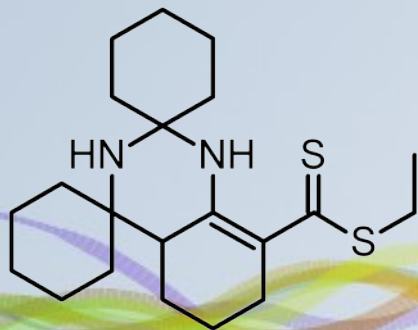


Building starting model

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



- 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 (approach with restraint)

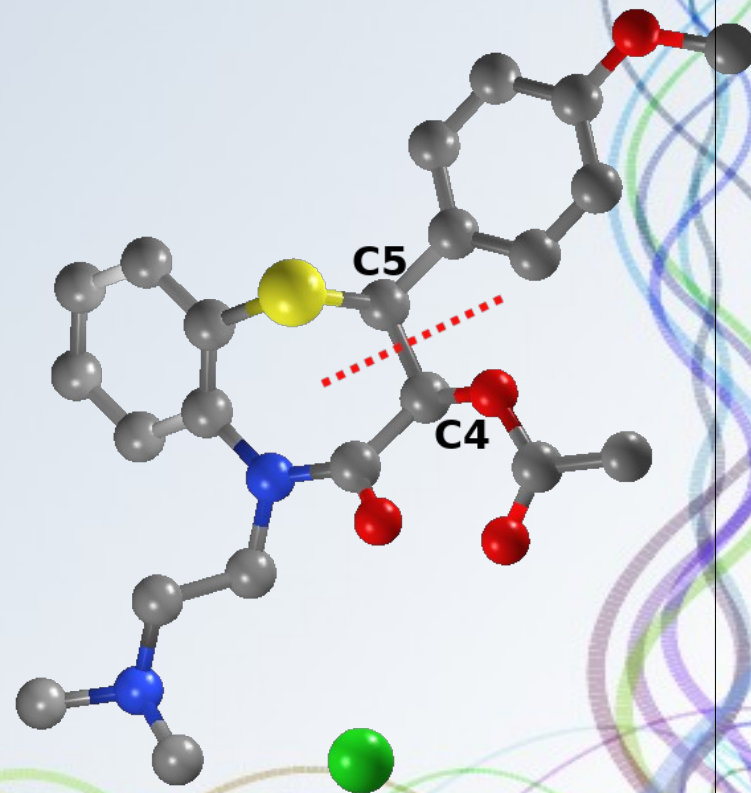
File: diltia_res.exp

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

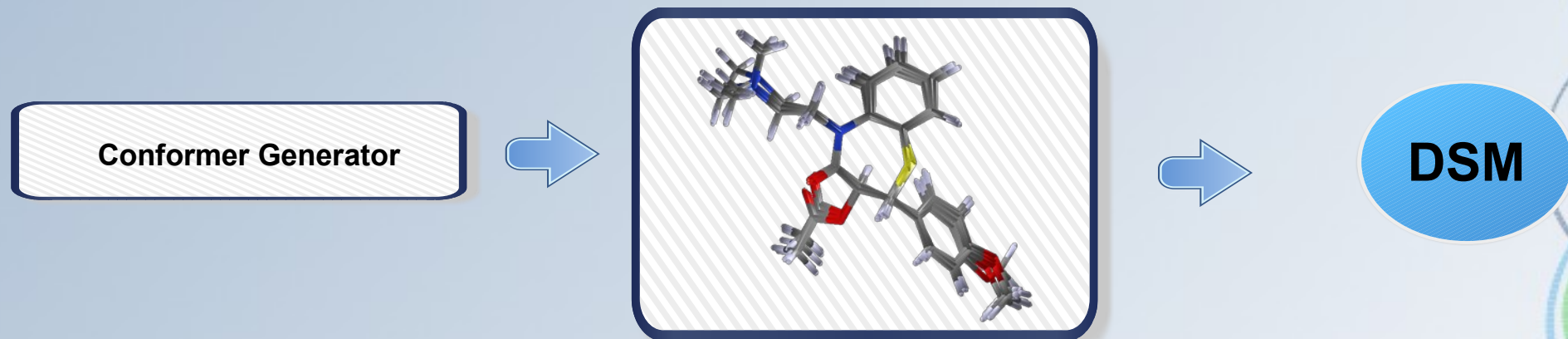
%Data
Cell      42.190   9.075   6.037   90   90   90
SpaceGroup p 21 21 21
Pattern    pd_0029.xy
Wavelength 1.54056

%fragment diltiazem_nw_break.mol
%fragment atoms Cl
deletehydro

%sannel
nrun 100
niter 1000
rest C4 C5
```



Combining conformational analysis with DSM



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

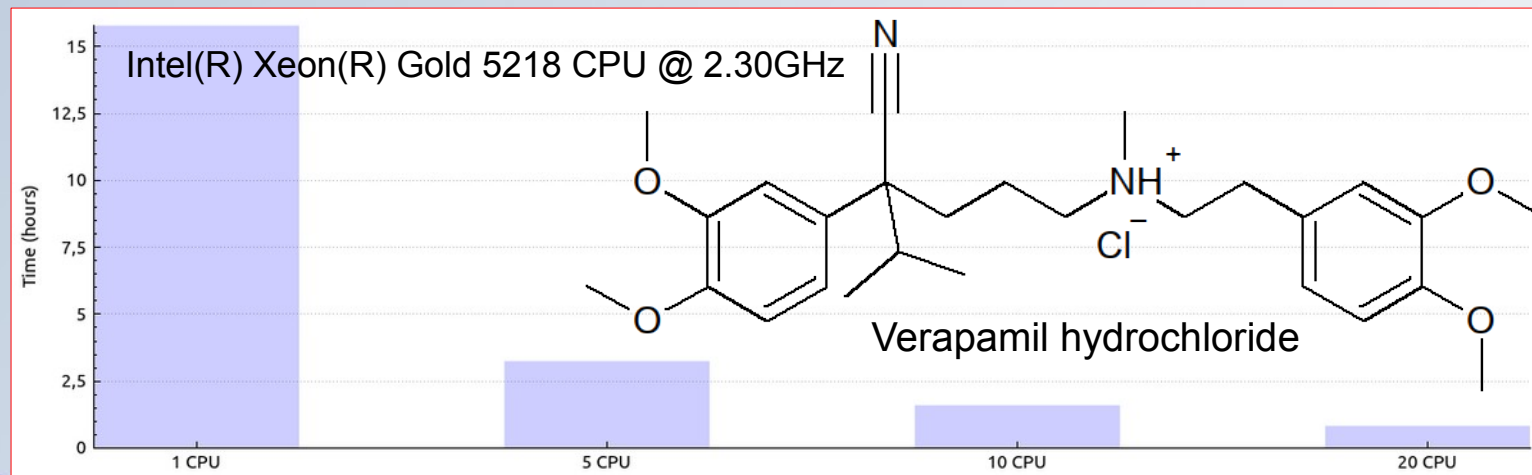


- CORINA
- MOE
- OMEGA

Running the Parallel Version of Expo

- Computer with multi-core CPUs and Linux environment.
- MPI installed.
- Run Expo by using the launcher `mpirun` with the appropriate options.

```
mpirun -np 20 expo input_file.exp
```



Structure solution of selexipag form I

```
%structure vohvia
```

```
%data
```

```
pattern VOHVIA.cif
```

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

```
space P21/c
```

```
%fragment vohvia_mopac.mol
```

```
%fragment vohvia_mopac.mol
```

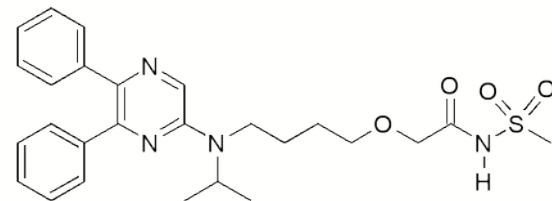
```
deletehydro
```

```
%sannel
```

```
nrun 100
```

```
niter 2000
```

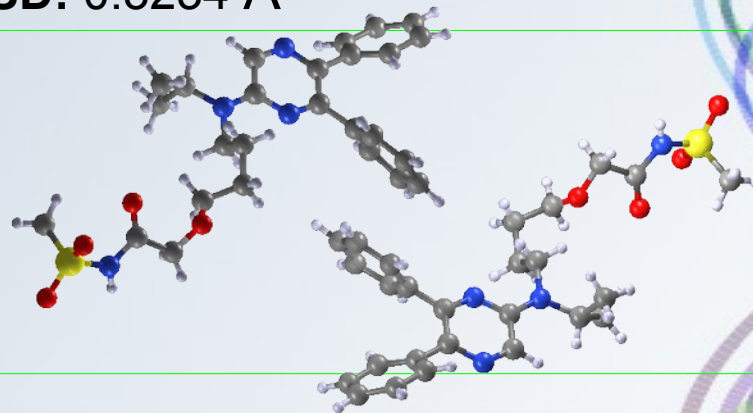
```
resm 2.5
```



DOF: 6+6+13+13 = 38

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

RMSD: 0.3234 Å



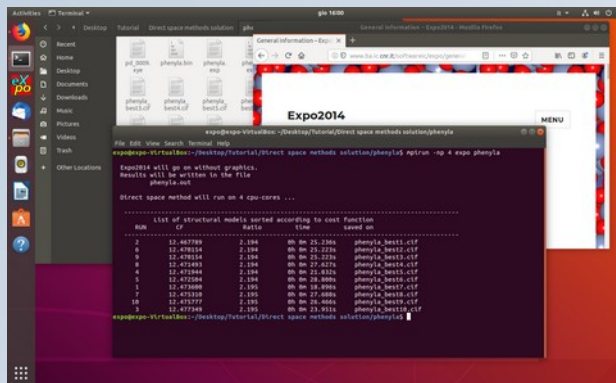
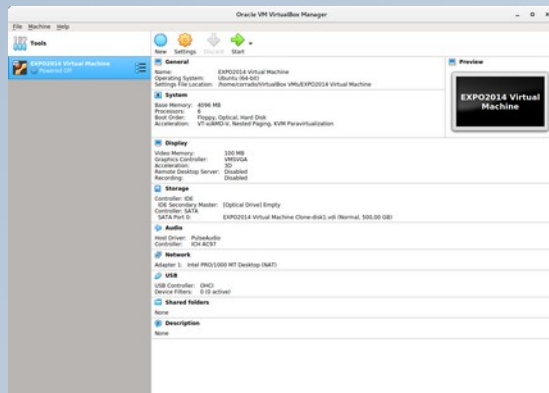
Reference

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

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

Cryst. Growth Des. 2019, 19, 8

Running parallel version of Expo



Windows Subsystem for Linux (WSL)

<https://learn.microsoft.com/en-us/windows/wsl/install>



```
docker run --rm \
-e PUID="$(id -u)" -e PGID="$(id -g)" \
-v "$PWD:/work" \
expo:<version> mpirun -np <NUM_TASK> expo <FILE.exp>
```

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*

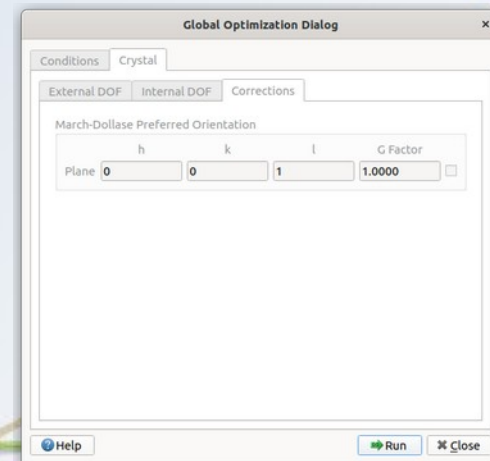
- Systematic problem in powder diffraction data

- *Preferred orientation*
- *Ka2 contributions*



Solution:

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



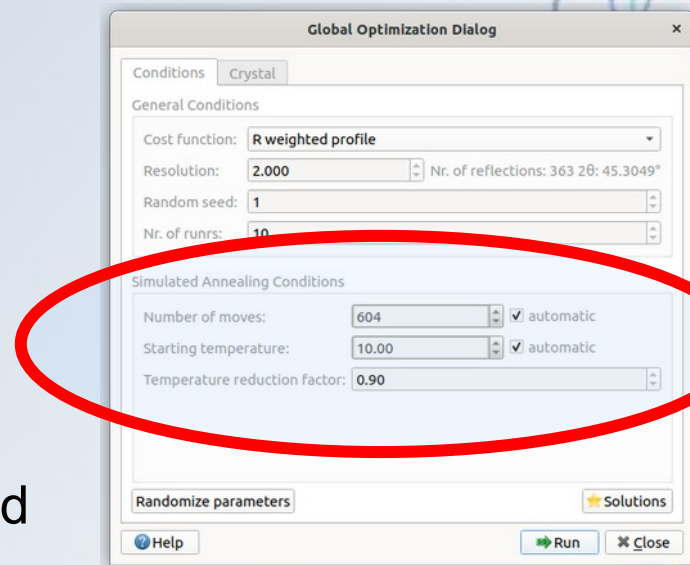
When Structure Solution Fails

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



Solution:

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



- The assumptions about thermal factors are invalid



Solution:

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

When Structure Solution Fails

- Space group and cell are not correct



Solution:

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

Contact, software download and info

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

Thank you for your kind attention