Reliable and Highly Accurate Molecular Crystal Structures from a Combination of XRPD and DFT-D

Jacco van de Streek

University of Copenhagen
Two Themes

1. Reliability: is a crystal structure from XRPD correct or not?

2. Accuracy: for a correct crystal structure from XRPD, how accurate are e.g. the bond lengths?
Solved from powder data with DASH, molecular starting geometry from 6-31G** optimisation, correct chemical compound, restrained refinement with TOPAS, no short contacts, no voids, all bond lengths and valence angles within < 3 ESDs (Mogul), all torsion angles as expected, no preferred orientation, zero-point error = 0.025, $B_{iso} = 2.6$, all hydrogen-bond donors and accepted satisfied with perfect geometries, 1.5 Å resolution data, normal background, occupancies 1.0.

No tricks!
Pigment Yellow 181

Solved from powder data with DASH, molecular starting geometry from 6-31G** optimisation, correct chemical compound, restrained refinement with TOPAS, no short contacts, no voids, all bond lengths and valence angles within < 3 ESDs (Mogul), all torsion angles as expected, no preferred orientation, zero-point error = 0.025, $B_{iso} = 2.6$, all hydrogen-bond donors and accepted satisfied with perfect geometries, 1.5 Å resolution data, normal background, occupancies 1.0.

Who would suspect that this structure could possibly be correct?
Dispersion-corrected DFT (DFT-D)

Force fields...
RMS = 0.497 Å
10 sec

Minimised
Experiment
Dispersion-corrected DFT (DFT-D)

Pure DFT...
RMS = 0.833 Å
100 hrs

Force fields...
RMS = 0.497 Å
10 sec

Minimised
Experiment
Dispersion-corrected DFT (DFT-D)

Dispersion-corrected DFT...
RMS = 0.083 Å
100 hrs

Pure DFT...
RMS = 0.833 Å
100 hrs

Force fields...
RMS = 0.497 Å
10 sec

Minimised
Experiment
Reproduction of Crystal Structures

225 “organic only” crystal structures from the August 2008 issue of *Acta Cryst.* E were downloaded (Open Access!)
- All 225 were energy-optimised with unit cell free
- Nett calculation time: one month

Reproduction of Crystal Structures

225 “organic only” crystal structures from the August 2008 issue of *Acta Cryst.* E were downloaded (Open Access!)
- All 225 were energy-optimised with unit cell free
- Nett calculation time: one month

225 experimental high-quality single-crystal structures...
225 energy-minimised structures...
How well are the experimental structures reproduced?

RMS Cartesian Displacement

Average = 0.083 Å

225 structures
Unit cell free
No H-atoms
Example RMS = 0.083 Å

Unit cell free
$\Delta V = -3\%$

- Red: Experimental
- Blue: Minimised
What about Wrong Structures?

- Unit cell free
- No H-atoms

*Acta Cryst.* E test set

Incorrect structures

RMS Cartesian displacement
Pigment Yellow 181

\[
R_{wp} = 1.6 \\
R_p = 1.2 \\
R'_{wp} = 4.5 \\
R'_p = 5.0 \\
\chi^2 = 1.2
\]

Solved from powder data with DASH, molecular starting geometry from 6-31G** optimisation, correct chemical compound, restrained refinement with TOPAS, no short contacts, no voids, all bond lengths and valence angles within < 3 ESDs (Mogul), all torsion angles as expected, no preferred orientation, zero-point error = 0.025, \( B_{iso} = 2.6 \), all hydrogen-bond donors and accepted satisfied with perfect geometries, 1.5 Å resolution data, normal background, occupancies 1.0.

Who would suspect that this structure could possibly be correct?
The amide group can be turned over 180°:
- O and N (or NH₂) only 1 electron difference
- Because all hydrogen atoms are moved as well, the infinite chain of hydrogen bonds remains intact.
The amide group can be turned over 180°:
- O and N (or NH₂) only 1 electron difference
- Because all hydrogen atoms are moved as well, the infinite chain of hydrogen bonds remains intact.
DFT-D Minimisation

Alternative 1
+0 kJ/mol

Alternative 2
+23 kJ/mol
Pigment Yellow 181

\[
R'_{wp} = 4.29 \quad (R_{wp} = 1.54) \\
R'_{p} = 4.76 \quad (R_{p} = 1.19) \\
\chi^2 = 1.19
\]


Example: Celecoxib Nicotinamide

3 x 2 x 2 = 12 different possibilities mentioned in paper

Accuracy

In this talk, we only look at the atomic \(x,y,z\) coordinates of molecular crystal structures.

No disorder.

Hydrogen atoms: ?

Peak shape, background etc. are “nuisance parameters”.

Common excuse: “it is only XRPD data, so the fit is not so good”. It is the opposite way round!

Accuracy: single crystal as gold standard, we also use RMS Cartesian displacement with DFT-D
Mogul z-scores

= value in query

SX values (CSD)

Number of ESDs from mean = z-score

Each bond has a z-score
Each angle has a z-score
Accuracy with DFT-D

Validating the crystal structure is done after the Rietveld refinement: it does not influence the Rietveld process.

This is a pity: the DFT-D contains a lot of independent information, can this information be used as part of the Rietveld refinement?

I.e. can the independent information from the DFT-D be merged into the Rietveld refinement to complement the experimental data to make the final result more accurate?
Accuracy with DFT-D

XRPD provides the packing...  ...DFT-D provides the details
Accuracy with DFT-D

Even better than restraints from a single molecule in vacuum: use the bond lengths and bond angles from the DFT-D minimised crystal structure as restraints.

“Polymorph-dependent restraints”.

Only after the structure has been validated as being correct, otherwise you are biasing your refinement.
Accuracy with DFT-D

Average absolute difference over 5,778 bonds from Acta E test set: 0.013 Å (non-hydrogen atoms only)
Accuracy with DFT-D

Polymorph-dependent restraints in TOPAS:

Distance_Restrain( N1  C2,  1.47872, 1.47998`_0.00610, 0, 10000 )
Distance_Restrain( N1  C3,  1.47894, 1.48956`_0.00690, 0, 10000 )
Distance_Restrain( N1  C4,  1.48941, 1.48492`_0.00524, 0, 10000 )
Distance_Restrain( C2  C5,  1.50425, 1.47233`_0.00715, 0, 10000 )
Distance_Restrain( C2  H6,  0.95,       0.96054`_0.01471, 0, 10000 )
Distance_Restrain( C2  H7,  0.95,       0.94072`_0.01347, 0, 10000 )
Distance_Restrain( C3  C8,  1.50403, 1.49550`_0.00524, 0, 10000 )
Distance_Restrain( C3  H9,  0.95,       0.95970`_0.01483, 0, 10000 )
...

Angle_Restrain( C2  N1  C3, 111.15614, 115.12083`_0.35599, 1, 1 )
Angle_Restrain( C2  N1  C4, 112.79224, 112.04806`_0.36718, 1, 1 )
Angle_Restrain( C3  N1  C4, 114.20513, 113.81510`_0.39248, 1, 1 )
Angle_Restrain( N1  C2  C5, 112.35920, 113.55737`_0.35977, 1, 1 )
Angle_Restrain( N1  C2  H6, 111.80674, 113.25174`_1.01145, 1, 1 )
...
Planarity Restraints

The DFT-D tells you which atoms are in the same plane, so the planarity restraints are also based directly on the DFT-D calculations

Flatten( C5 C15 H27 C26 O40 C38 H47, , 4.17658429 \_5.92244831, 0, 100000 )
Hydrogen Atoms

For the hydrogen atoms, restraints are not always sufficient.

Better solution: energy-minimise hydrogen positions with non-hydrogens and unit cell kept fixed.
Example: Piroxicam III

Max. Mogul z-scores:
- Bonds: 2.3
- Angles: 3.7
- RMSCD: 0.10 Å
- $\chi^2 = 1.30$

Lab data
No PO

Reliable *and* Accurate

Some example of crystal structures from the literature that can be corrected with DFT-D

and

for which DFT-D provides the polymorph-dependent restraints for the Rietveld refinement
Glipizide (2005)

The pyrazyl ring can be turned over 180°: N and C (or CH) only 1 electron difference
Ambiguity mentioned in paper

Glipizide (2005)

RMSCD 0.72 Å

+4.56 kcal/mol

RMSCD 0.13 Å

0 kcal/mol
RMS Cartesian Displacement

Unit cell free
No H-atoms

0.13 Å
0.72 Å
Glipizide Corrected

Max. Mogul z-scores:
- Bonds: 2.1
- Angles: 1.9
- RMSCD: 0.13 Å

$\chi^2 = 1.11$

Lab data

No PO

* = NaCl
Clarithromycin Monohydrate (2012)

Maximum *Mogul* z-scores:
Bonds: 7.4
Angles: 3.2
Voids/Z ($\text{H}_2\text{O} = 21$ Å³)
*Mercury*: 40 Å³
*Hofmann*: 55 Å³
RMSCD: ?
Synchrotron

Clarithromycin Trihydrate Corrected

Clarithromycin "monohydrate"

Clarithromycin trihydrate

Synchrotron data, y-axis: $\sqrt{l}$
Clarithromycin Trihydrate

Maximum *Mogul* z-scores:
Bonds: 5
Angles: 8
Synchrotron
“Clarithromycin” Trihydrate

Maximum *Mogul* z-scores:

Bonds: 5
Angles: 8

Synchrotron

One of the stereo-centres is wrong:
this is not Clarithromycin
Clarithromycin Trihydrate Corrected

Maximum *Mogul* z-scores:
Bonds: 1.4, Angles: 3.4
RMSCD: 0.14 Å
Synchrotron

$\chi^2 = 1.44$, no PO

Where does DFT-D Enter the Process?

1. To give a better starting molecular geometry
2. Validate the crystal structure
3. Feed back the energy-minimised crystal structure as polymorph-dependent restraints
4. Energy-minimise the hydrogen atoms, keeping the unit-cell parameters and the positions of the non-hydrogen atoms fixed
DFT-D

Which functional? Which dispersion correction?

For *energies*, these questions are critical.

For *structures* (coordinates / unit-cell parameters): it does not matter.

PBE, PW91, BLYP, B3LYP, Neumann & Perrin, Grimme 2006, Grimme 2010 give very similar results.
Limitations...

Temperature effects
Metals
Disorder
Hydrogen atoms (salt *versus* co-crystal)
Hydrogen Atoms

Prilocaine
Hydrogen Atoms
Hydrogen Atoms

RMS = 0.42 Å  
+4.8 kcal/mol

RMS = 0.11 Å  
0.0 kcal/mol
RMS Cartesian Displacement

Unit cell free
No H-atoms

0.11 Å
0.42 Å
Virtual Beamline Pilot

Funding from Villum Foundation for hardware / software

Permission from Avant-garde Materials Simulation and the University of Vienna

Molecular XRPD structures in IUCr journals only

Your crystal structures are energy-minimised with DFT-D free of charge as part of the review process
Conclusions

- DFT-D calculations can validate crystal structures determined from XRPD data.
- DFT-D calculations can provide polymorph-dependent restraints for crystal structures determined from XRPD data.
- DFT-D calculations can accurately position the hydrogen atoms in crystal structures determined from XRPD data.
- Limitations: $T$ effects, metals, disorder, H atoms
Acknowledgements

Marcus Neumann – Avant-garde Materials Simulation
Martin U. Schmidt – University of Frankfurt