PLATON and raw diffraction data opportunities for chemical crystallography publishing

Ton Spek
Utrecht University
The Netherlands
Vienna – ECM32-18-08-2019
What is a Crystal Structure Report?

- A published crystal structure represents its author’s interpretation of the underlying experimental diffraction data.

- Often the main structural result that is presented in chemical journals is just an ORTEP like illustration as ‘proof’ of the chemistry and a CSD reference code for crystallographic details.
The Problem is ..

- The interpretation might be erroneous.
- Qualifiers such as ‘best attainable’ and ‘sufficient for the purpose of the study’ are often lost in the archive or ignored by the users of the structural results.
- (Unreported) constraints and restraints may make the sometimes elaborate discussion of the details of the molecular geometry largely meaningless.
What is needed for a good Report?

• Archival of the experimental data, details of the data reduction, details of the refinement.
• The possibility to repeat the structure analysis in order to investigate claimed unusual results.
• The option to use the experimental data for unrelated research that might require more advanced analysis of the data.
• Note: the experimental data may be unique or not easily obtained again.
The Current Solution

• The CIF standard introduced in the 1990’s gives the option to archive the relevant information electronically as opposed to the need to retype data from archived printed material. [Dick Marsh retyped the Fo/Fc data]
• The development of validation software
• Enforcing the deposition of the relevant data in computer readable format and their validation with an associated validation report
Required Data for Validation

• Currently, both a ‘CIF’ and an ‘FCF’ file are required for full validation (both in CIF format)
• The CIF should contain all experimental details
• THE CIF is tested for consistency, completeness and unusual results
• The FCF lists the observed and calculated intensities and is used to analyse the refinement.
• The FCF file is optional when it can be reconstructed from the info in the CIF as is e.g. the case when the SHELXL20xx refinement tool is used (.res & unmerged .hkl embedded)
An Example with a Problem

• Structure published in a chemical journal in (2017) (it has an issue communicated to me by Dr. Natalie Johnson (CCDC))

• This example illustrates a failed refereeing process of the underlying experimental data.

• The reported refinement results turn out not based on experimental data.

• It is potentially fraudulent but might be due to insufficient experience with the software used.
“... The relative orientation of three contiguous all-carbon quaternary stereocenters was confirmed from the X-Ray Structure of the corresponding $p$-nitrobenzyl ether 36. ...”
What about the Xtal Details?

• There are no experimental details in the paper.
• The paper gives no CSD reference codes in the printed text.
• The journal site offers supplementary material in PDF format including a ‘Table 1’ with selected experimental details (including a CSD reference number).
• The missing details (in CIF format) could be found in the CSD. An FCF could be reconstructed based on the embedded .res & .hkl.
• Did the referee(s) have access to those data?
<table>
<thead>
<tr>
<th>Crystal data and Structure refinement for 36 (CCDC1553045)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Empirical formula</strong></td>
</tr>
<tr>
<td><strong>Formula weight</strong></td>
</tr>
<tr>
<td><strong>Crystal habit, colour</strong></td>
</tr>
<tr>
<td><strong>Crystal size, mm$^3$</strong></td>
</tr>
<tr>
<td><strong>Temperature, $T$</strong></td>
</tr>
<tr>
<td><strong>Wavelength, $\lambda$(Å)</strong></td>
</tr>
<tr>
<td><strong>Crystal system</strong></td>
</tr>
<tr>
<td><strong>Space group</strong></td>
</tr>
<tr>
<td><strong>Unit cell dimensions</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Volume, $V$(Å$^3$)</strong></td>
</tr>
<tr>
<td><strong>Z</strong></td>
</tr>
<tr>
<td><strong>Calculated density, Mg·m$^{-3}$</strong></td>
</tr>
<tr>
<td><strong>Absorption coefficient, $\mu$(mm$^{-1}$)</strong></td>
</tr>
<tr>
<td><strong>$F(000)$</strong></td>
</tr>
<tr>
<td><strong>$\theta$ range for data collection</strong></td>
</tr>
<tr>
<td><strong>Limiting indices</strong></td>
</tr>
<tr>
<td><strong>Reflection collected / unique</strong></td>
</tr>
<tr>
<td><strong>Completeness to $\theta$</strong></td>
</tr>
<tr>
<td><strong>Refinement method</strong></td>
</tr>
<tr>
<td><strong>Data / restraints / parameters</strong></td>
</tr>
<tr>
<td><strong>Goodness-of-fit on $F^2$</strong></td>
</tr>
<tr>
<td><strong>Final $R$ indices [$I&gt;2\sigma(I)$]</strong></td>
</tr>
<tr>
<td><strong>$R$ indices (all data)</strong></td>
</tr>
</tbody>
</table>
10% Probability Displacement Ellipsoid Plot

High ‘thermal’ motion

Looks reasonable notwithstanding the high R and wR2 values
VALIDATION REPORT FOR CURRENT CIF

# PLATON/CHECK-(120819) versus check.def version 190730, Entry: platon_pl
# Data: 1553045.cif - Type: CIF14   Band Precision C-C = 0.0139 Å
# Refl: 1553045.fcf - Type: LIST4   Temp = 293 K
# Audlt: SHELXL-2014/7
# Refln: SHELXL-2014/7 (SHELDRICK, 2014)
# X-ray MoKa   R(Lnt) = 0.000, wR2/R(Lnt) =*****  Nref/Npar = 14.0
# Cell 7.5010(11) 14.831(3) 19.619(5) 90 90 90 424(15) 90
# Wavelength 0.71073 Volume Reported 2182.5(8) Calculated 2182.5(8)
# SpaceGroup from Symmetry P 21/n  Hall: -P 2yn  monoclinic
#   Reported P 21/n -P 2yn  monoclinic
# MolSymForm C26 H31 N 03
#   Reported C26 H31 N 03
#   SumFormula C26 H31 N 03
#   Reported C26 H31 N 03
#   Mr = 405.52 [Calc], 405.52[Rep]  Volume/NonHAtoms = 18.2
# Dx,gcm-3 = 1.234 [Calc], 1.234[Rep]
# Z = 4 [Calc], 4[Rep]
# Mu (mm-1) = 0.080 [Calc], 0.080[Rep]  Xtal Sz = 0.100×0.150×0.200 mm
# F000 = 872.0 [Calc], 872.0[Rep]  or F000’ = 872.38[Calc]
# Reported T Limts: Tmln=0.986  Tmax=0.992  AbsCorr = ?
# Calculated T Limts: Tmln=0.986  Tmax=0.992 Extl = 0.00600
# Measured HKL: Reported 7670, Embedded 9985
# Reported Hmax = 8, Kmax = 17, Lmax = 23, Nref = 3839  Th(max)= 24.994
# Obs ln FCF Hmax = 8, Kmax = 17, Lmax = 23, Nref = 3839  Th(max)= 24.994
# Calculated Hmax = 8, Kmax = 17, Lmax = 23, Nref = 3839  Ratlo = 1.000
# Reported Rho(mln) = -0.34, Rho(max) = 0.43 e/Ang**3 (From CIF)
# Calculated Rho(mln) = -0.36, Rho(max) = 0.47 e/Ang**3 (From CIF+FCF data)
# w=1/[slama**2(Fo**2)+(0.1670P)**2+ 7.7057P], P=(Fo**2+2*Fc**2)/3
# R= 0.1951( 3811), wR2= 0.4557( 3839), S = 1.055  (From CIF+FCF data)
# R= 0.1951( 3811), wR2= 0.4556( 3839), S = 1.055  (From FCF data only)
# R= 0.1954( 3829), wR2= 0.4556( 3839), S = 1.055, Npar=274

334_ALERT_2_C Small Aver. Benzene C-C Dlst C1 -C6................. 1.37 Ang.
410_ALERT_2_C Short Intra H....H Contact  H13 ........ H24A .......... 1.98 Ang.

906_ALERT_3_C Large K Value in the Analysis of Variance ........ 67.406 Check
906_ALERT_3_C Large K Value in the Analysis of Variance ........ 2.163 Check
906_ALERT_3_C Large K Value in the Analysis of Variance ........ 8.790 Check
906_ALERT_3_C Large K Value in the Analysis of Variance ........ 5.402 Check
906_ALERT_3_C Large K Value in the Analysis of Variance ........ 3.031 Check
906_ALERT_3_C Large K Value in the Analysis of Variance ........ 2.099 Check
918_ALERT_3_C Reflections with I(obs) much Smaller I(calc) ........ 54 Check
939_ALERT_3_C Large Value of Nat (SHELXL) Weight Optimized S ........ 46.63 Check
978_ALERT_2_C Number C-C Bonds with Positive Residual Density...... 0 Info
992_ALERT_5_C Repd & Actual_reflns_number_gt Values Differ by........ 18 Check

#=================================
066_ALERT_1_G Predicted and Reported Tmtn&Tmax Range Identical .......... ? Check
072_ALERT_2_G SHELXL Frst Parameter ln WGHT Unusually Large ........ 0.17 Report
083_ALERT_2_G SHELXL Second Parameter ln WGHT Unusually Large .... 7.71 Why ?
199_ALERT_1_G Reported _cell_measurement_temperature ........ (K) .... 293 Check
200_ALERT_1_G Reported _diffrn_ambient_temperature ........ (K) .... 293 Check
795_ALERT_4_G C-Atom ln CIF Coordinate List Out-of-Sequence .. C15 Note
796_ALERT_4_G O-Atom ln CIF Coordinate List Out-of-Sequence .... 02 Note
883_ALERT_1_G No Info/Value for _atom_site_solution.prior ... Please Do !
898_ALERT_4_G Second Reported H-M Symbol ln CIF Ignored .......... ! Check
909_ALERT_3_G Percentage of I>2sig(I) Data at Theta(Max) Still .... 99% Note
949_ALERT_5_G Unusual Experimental Reflection Sigma(I)’s .......... Please Check
961_ALERT_5_G Dataset Contains no Negative Intensity Values ......... Please Check
991_ALERT_5_G HKL data created with PLATON/generate ............... ! Note

#=================================

ALERT_Level and ALERT_Type Summary
==================================

2 ALERT_Level_A = Most Likely a Serious Problem - Resolve or Explain
8 ALERT_Level_B = A Potentially Serious Problem - Consider Carefully
33 ALERT_Level_C = Check. Ensure it is Not caused by an Omission or Oversight
13 ALERT_Level_G = General Info/Check that it is not Something Unexpected
Inspection of the hkl data

• Diederichs Plot: $I_{\text{obs}}/\sigma(I_{\text{obs}})$ versus $I_{\text{obs}}$
  1 – expected
  2 – for this structure
Also:
Plot $\sigma(I_{\text{obs}})$ versus $\sqrt{I_{\text{obs}}}$
Example of a Normal Diederichs Plot
I/\Sigma (I) versus LOG10(I) PLOT - (Diederichs Plot)


Data .................. 1553045.fcf
Device Type .......... SUPERNOVA, SINGLE SOURCE AT OFFSET), EOS
Data Collection ..... CRYSTALISPRO, AGILENT TECHNOLOGIES
Data Reduction .... CRYSTALISPRO, AGILENT TECHNOLOGIES
Absorb Details ?
Structure Refinement SHELXL-2014/7 (SHELDRICK, 2014)
Based on the CIF embedded unmerged set of calculated ‘observed’ data
What might have happened?

• No details of the data processing are available, however there are traces in the CIF data that suggest that the structure was originally solved in Pn and subsequently, using PLATON/ADDSYM, transformed to P2_1/n.

• Another PLATON tool, stricktly intended for testing purposes only, was apparently used to create ‘observed data’ for the final refinement, substituting the real observed data.
Thus ...

• The currently archived data in the CSD are largely useless for this structure.
• This example might make a case for (automated) archival of the diffraction images along with the availability of evaluation tools.
Archival of Diffraction Images?

• The solution of some problems with a structure may require the need to go back to the (archived) diffraction images.
• Additional spots may indicate twinning, super cells, disorder [Integration erroneous]
• Sometimes re-integration will be needed
• Alternatively, qualitative info about the images might be sufficient to resolve questions.
• It might be helpful to archive some synthetic precession images
Synthetic Precession Images

0kl

1kl
Finally

• An (automatically created) validation report of the image processing with details about streaks, additional (weak) diffraction spots, diffuse scattering etc. might be very helpful to understand structure analysis problems.
Thanks!

a.l.spek@uu.nl

http://www.cryst.chem.uu.nl/spek