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 stuctural 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

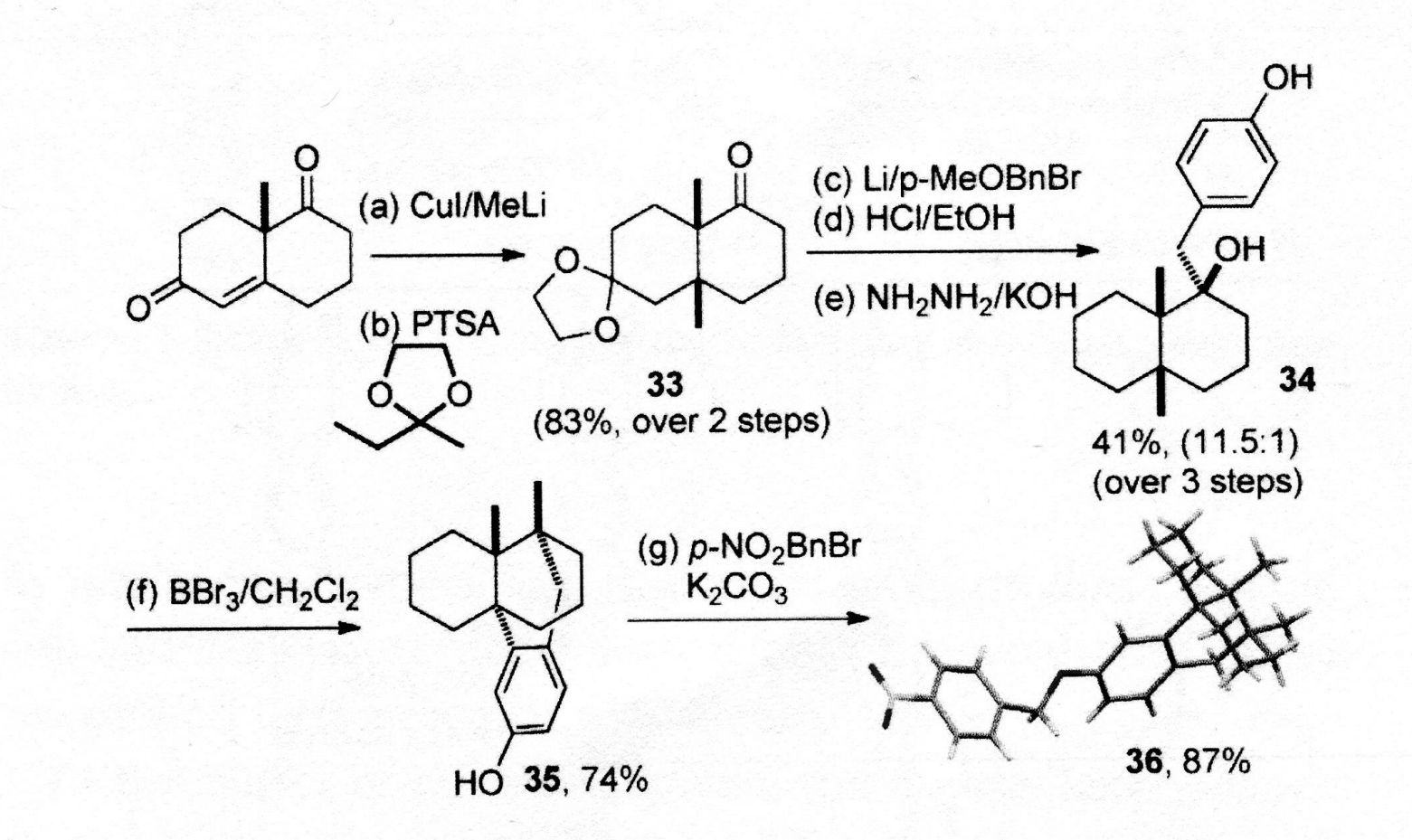
- The CIF standard introduced in the 1990's gives the option to archive the relevant information electronically as opposed to the need to retype data from archived printed material. [Dick Marsh retyped the Fo/Fc data]
- The development of validation software
- Enforcing the deposition of the relevant 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.

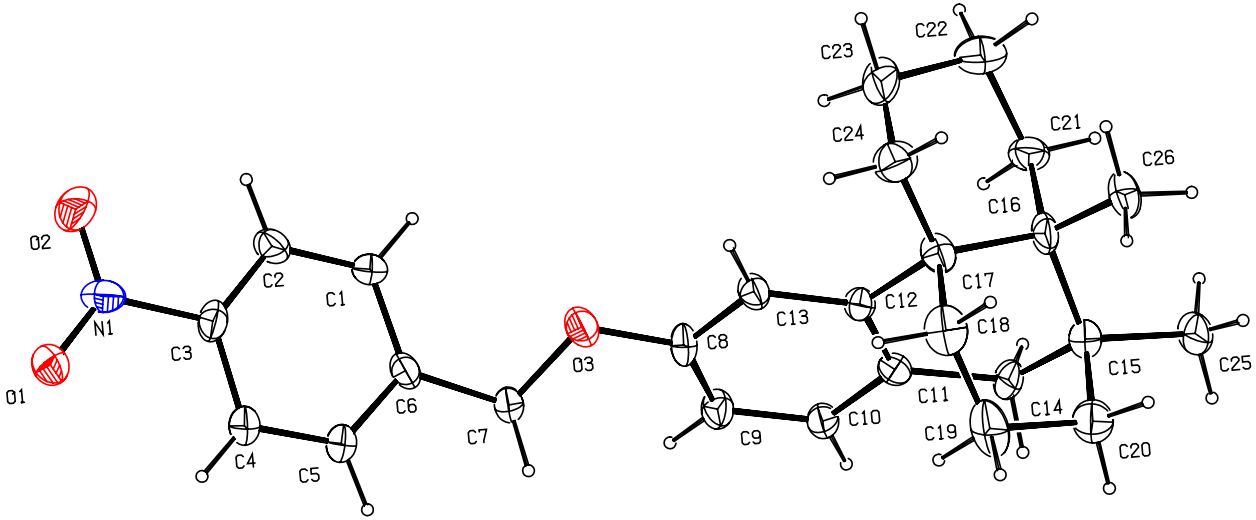


"… The relative orientation of three contiguous all-carbon quaternary stereocenters was confirmed from the X-Ray Structure of the corresponding *p*-nitrobenzyl ether **36**. …"

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 ?

Crystal data and Structure refinement for **36** (CCDC1553045)	
Empirical formula	$C_{26}\ H_{31}\ N\ O_3$
Formula weight	405.52
Crystal habit, colour	needle / whitish
Crystal size, mm^3	0.20 X 0.15 X 0.10
Temperature, T	293(2)K
Wavelength, λ(Å)	0.71073
Crystal system	monoclinic
Space group	*'P 21/n'*
Unit cell dimensions	$a = 7.5010(11)$ Å $b = 14.831(3)$ Å $c = 19.619(5)$ Å $\alpha = 90.00°$, $\beta = 90.424(15)°$, $\gamma = 90.00°$
Volume, V(Å^3)	2182.5(8)
Z	4
Calculated density, $Mg \cdot m^{-3}$	1.139
Absorption coefficient, $\mu(mm^{-1})$	0.080
F(000)	872
θ range for data collection	1.721° to 24.994°
Limiting indices	$-8 \le h \le 8, -17 \le k \le 17, -23 \le l \le 23$
Reflection collected / unique	7670 / 3829 [R(int) = 0.0001]
Completeness to θ	100% (θ = 24.994°)
Refinement method	'SHELXL-2014/7 (Sheldrick, 2014)'
Data / restraints / parameters	3829 / 0 / 274
Goodness–of–fit on F^2	1.055
Final R indices [I>2sigma(I)]	$R1 = 0.1954$, $wR2 = 0.4555$
R indices (all data)	$R1 = 0.1956$, $wR2 = 0.4556$
Largest diff. peak and hole	0.427 and – 0.341 $e \cdot Å^{-3}$



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                  Bond Precision   C-C = 0.0139 A
# Refl: 1553045.fcf - Type: LIST4                                    Temp = 293 K
# Audit:SHELXL-2014/7
# Refin:SHELXL-2014/7 (SHELDRICK, 2014)
# X-ray MoKa              R(int) = 0.000,  wR2/R(int) =*****,  Nref/Npar =  14.0
# Cell  7.5010(11)   14.831(3)   19.619(5)          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
# MoietyFormula C26 H31 N O3
#      Reported C26 H31 N O3
#    SumFormula C26 H31 N O3
#      Reported C26 H31 N O3
# 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 Size =  0.100x0.150x0.200 mm
# F000        =     872.0[Calc],     872.0[Rep]  or F000' =    872.38[Calc]
# Reported   T Limits: Tmin=0.986                 Tmax=0.992  AbsCorr = ?
# Calculated T Limits: Tmin=0.986 Tmin'=0.984  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 in FCF Hmax=  8, Kmax= 17, Lmax= 23, Nref=   3839[  3839], Th(max)= 24.994
# Calculated Hmax=  8, Kmax= 17, Lmax= 23, Nref=   3839         , Ratio  =  1.000
# Reported   Rho(min) = -0.34, Rho(max) =  0.43 e/Ang**3 (From CIF)
# Calculated Rho(min) = -0.36, Rho(max) =  0.47 e/Ang**3 (From CIF+FCF data)
# w=1/[sigma**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
#==============================================================================
For Documentation: http://http://www.platonsoft.nl/CIF-VALIDATION.pdf
#==============================================================================
```

```
334_ALERT_2_C Small Aver. Benzene C-C Dist C1       -C6              1.37 Ang.
410_ALERT_2_C Short Intra H...H Contact  H13      ..H24A     .      1.98 Ang.
                                                    x,y,z  =      1_555 Check
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 Reflection(s) with I(obs) much Smaller I(calc) .       54 Check
939_ALERT_3_C Large Value of Not (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 _refins_number_gt Values Differ by      18 Check
#=============================================================================
066_ALERT_1_G Predicted and Reported Tmin&Tmax Range Identical       ? Check
072_ALERT_2_G SHELXL First  Parameter in WGHT  Unusually Large     0.17 Report
083_ALERT_2_G SHELXL Second Parameter in 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 in CIF Coordinate List Out-of-Sequence ..     C15 Note
796_ALERT_4_G O-Atom in CIF Coordinate List Out-of-Sequence ..     O2 Note
883_ALERT_1_G No Info/Value for _atom_sites_solution_primary .  Please Do !
898_ALERT_4_G Second Reported H-M Symbol in 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 Intensities .......  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_{obs}/sigma(I_{obs}) versus $_{Iobs}$

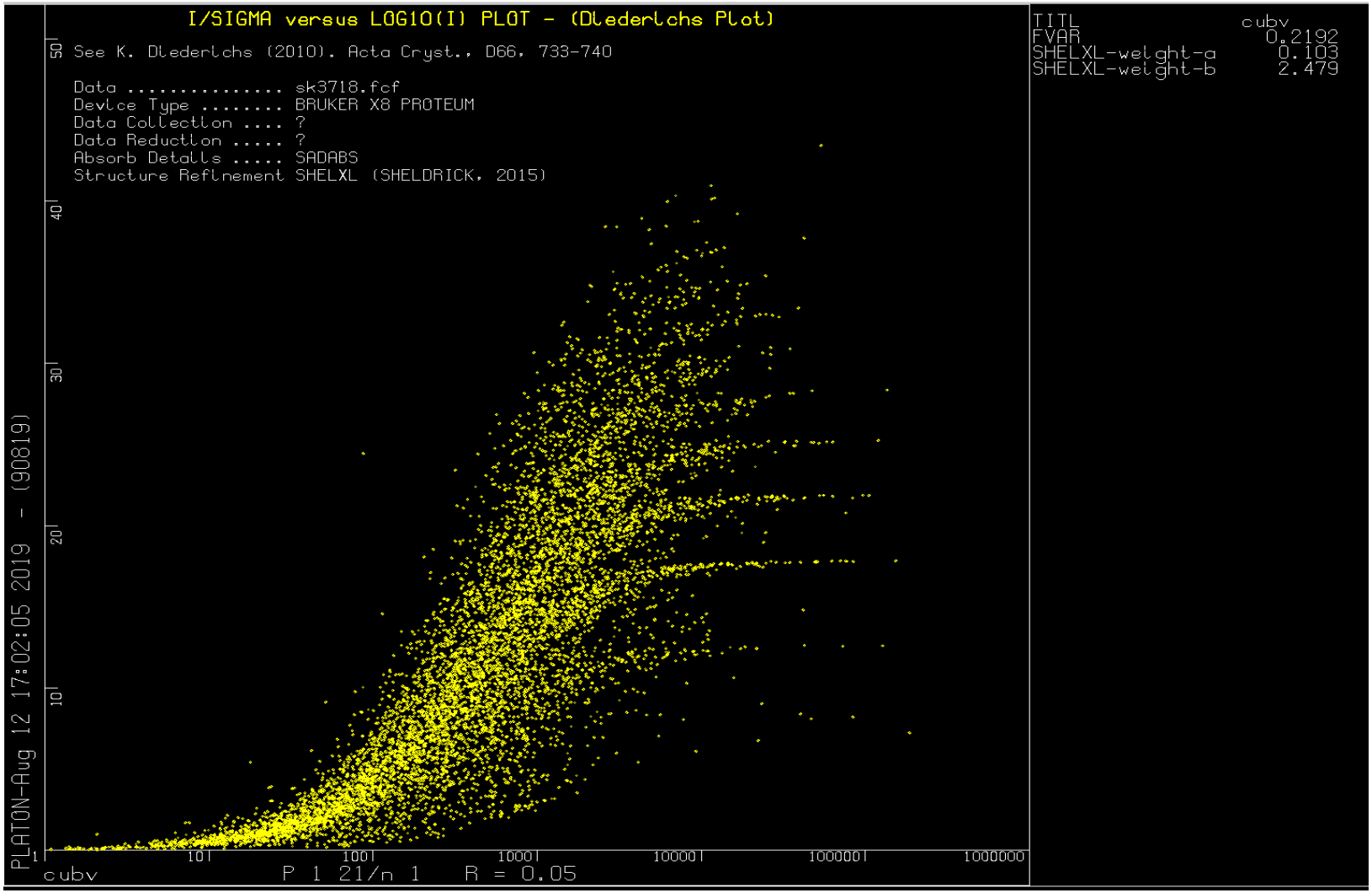
1 – expected

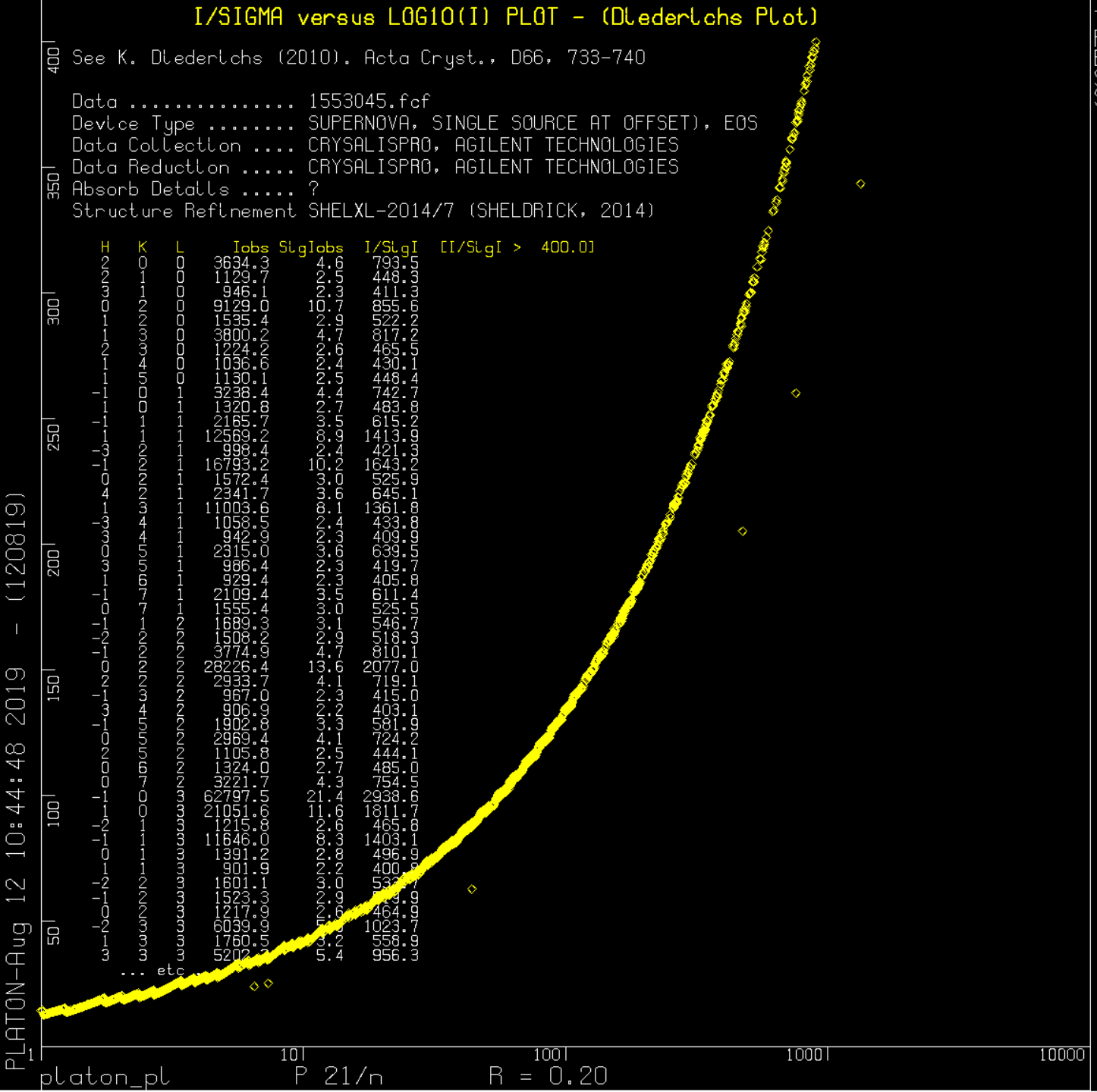
2 – for this structure

Also:

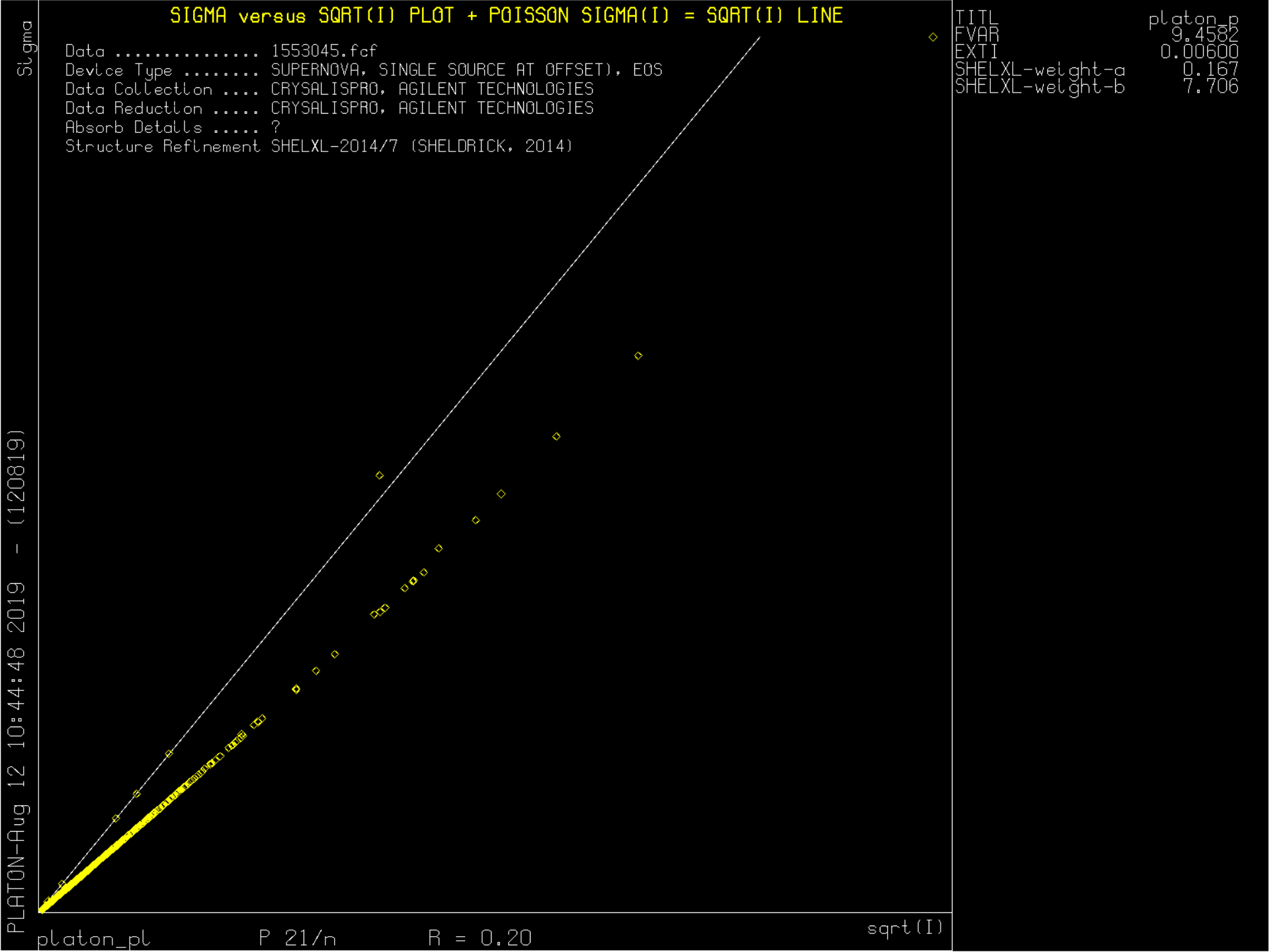
Plot sigma(I_{obs}) versus sqrt(I_{obs})

Example of a Normal Diederichs Plot





TITL platon_p
FVAR 9.4582
EXTI 0.00600
SHELXL-weight-a 0.167
SHELXL-weight-b 7.706


SIGMA versus SQRT(I) PLOT + POISSON SIGMA(I) = SQRT(I) LINE
Data 1553045.fcf
Device Type ..... SUPERNOVA, SINGLE SOURCE AT OFFSET), EOS
Data Collection .... CRYSALISPRO, AGILENT TECHNOLOGIES
Data Reduction ..... CRYSALISPRO, AGILENT TECHNOLOGIES
Absorb Details ..... ?
Structure Refinement SHELXL-2014/7 (SHELDRICK, 2014)
TITL platon_p
FVAR 9.4582
EXTI 0.00600
SHELXL-weight-a 0.167
SHELXL-weight-b 7.706
Sigma
PLATON-Aug 12 10:44:48 2019 - (120819)
platon_pl
P 21/n
R = 0.20
sqrt(I)

Top of ASYM output listing

```
   H  K  L     <I>   <SIG> ILT     I  &  SIG       I  &  SIG
-------------------------------------------------------------------------

********************* Spacegroup Extinctions ***************

   0  1  0      4363    66 1      4363    66
   0  3  0       916    30 1       916    30
   0  5  0      3947    63 1      3947    63
   0  7  0     21093   145 1     21093   145
   0  9  0      3806    62 1      3806    62
   0 11  0       484    22 1       484    22
   0 13  0       232    15 1       232    15
   0 15  0       938    31 1       938    31
   0 17  0       399    20 1       399    20

********************* ZONE  L =    0 ************************

   2  0  0    314761   397 1    314746   561    314775   561
   4  0  0      7584    62 1      7584    87      7584    87
   6  0  0      1453    27 1      1455    38      1450    38
   8  0  0      1300    25 1      1300    36      1299    36
   1  1  0      5100    51 1      5116    72      5084    71
   2  1  0    100485   224 1    100486   317    100484   317
   3  1  0     84379   205 1     84379   290     84378   290
   4  1  0       795    20 1       797    28       793    28
   5  1  0      7747    62 1      7746    88      7747    88
   6  1  0       517    16 1       517    23       517    23
   7  1  0       394    14 1       394    20       394    20
   8  1  0       447    15 1       447    21       446    21
   0  2  0    732490   856 1    732490   856
   1  2  0    136120   261 1    136112   369    136127   369
   2  2  0     37119   136 1     37112   193     37125   193
   3  2  0     21554   104 1     21564   147     21544   147
   4  2  0     10202    71 1     10213   101     10191   101
```

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 e thstructure 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.

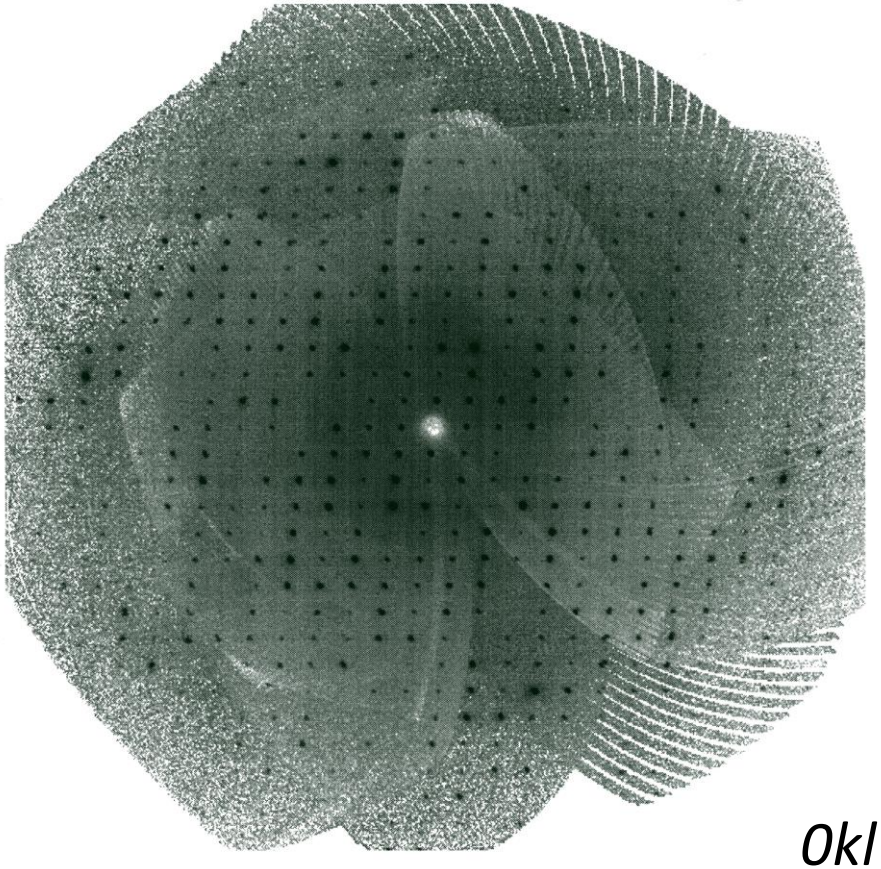
Thus …

- 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.

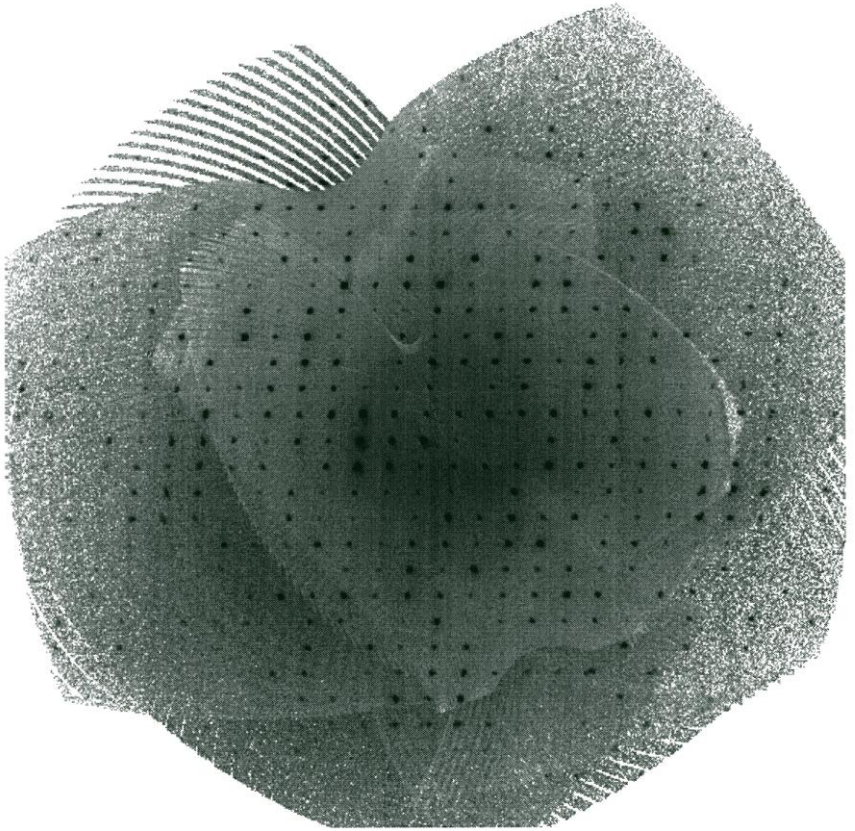
Archival of Diffraction Images?

- 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