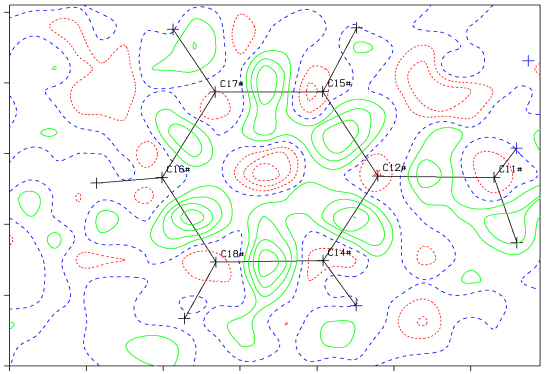


The PLATON checkCIF and SQUEEZE Tools

Ton Spek

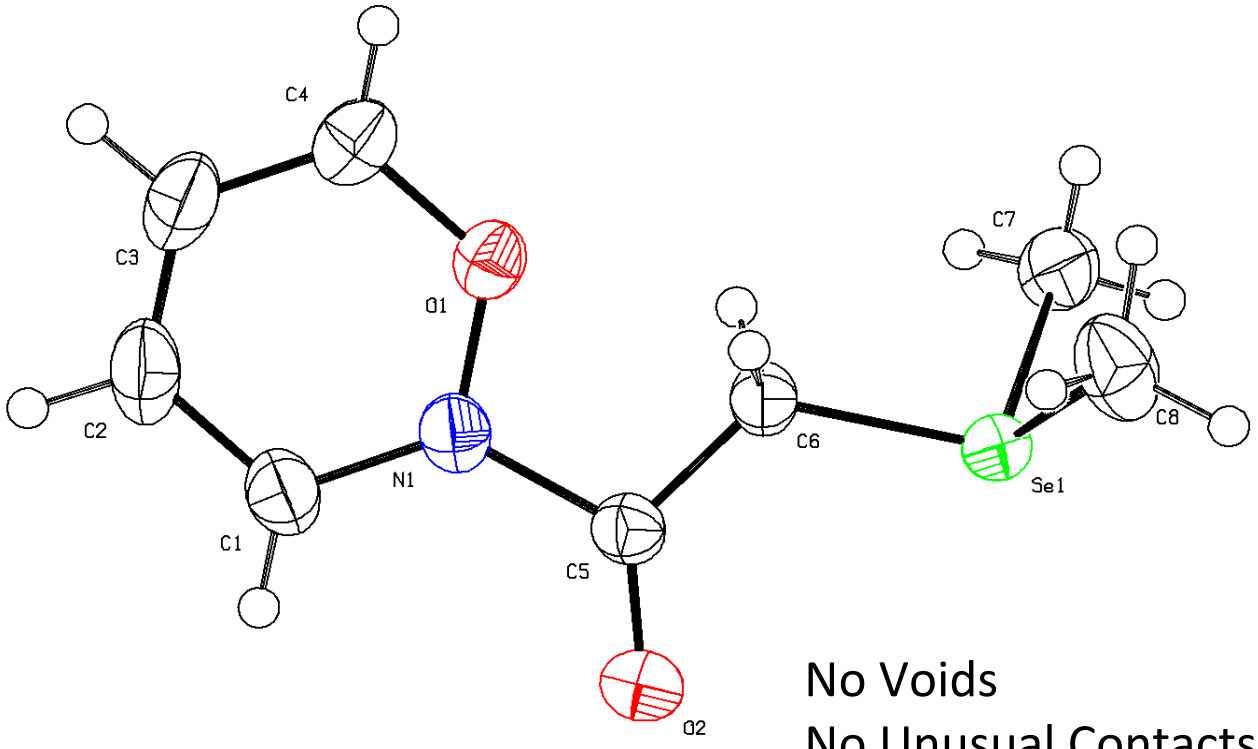
Utrecht University, NL

Rigaku/Oxford Diffraction Usermeeting

Wroclaw, june 20, 2016

QUESTION: Is this Structure Correct ?

The Aug. 2015 IUCr/checkCIF reports no serious ALERTS

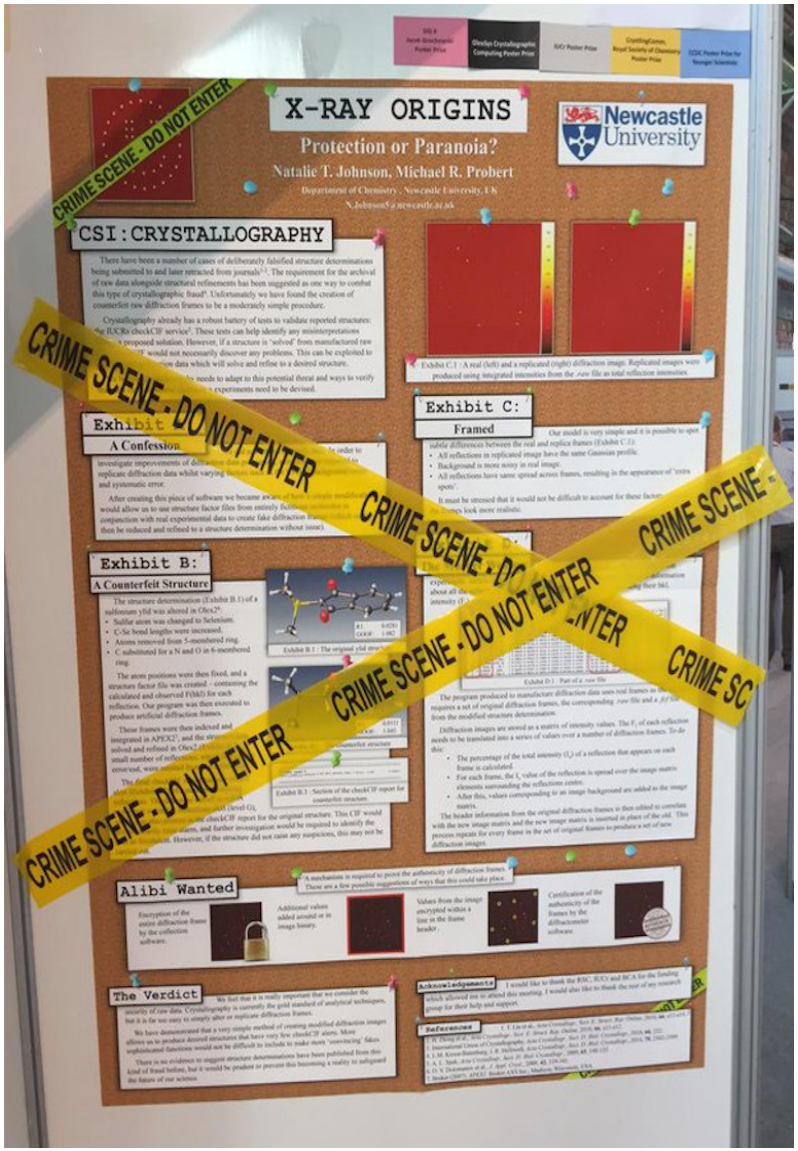


R1 = 0.0111
wR2 = 0.0339
S = 1.042
Rhomin -0.43
Rhomax 0.47

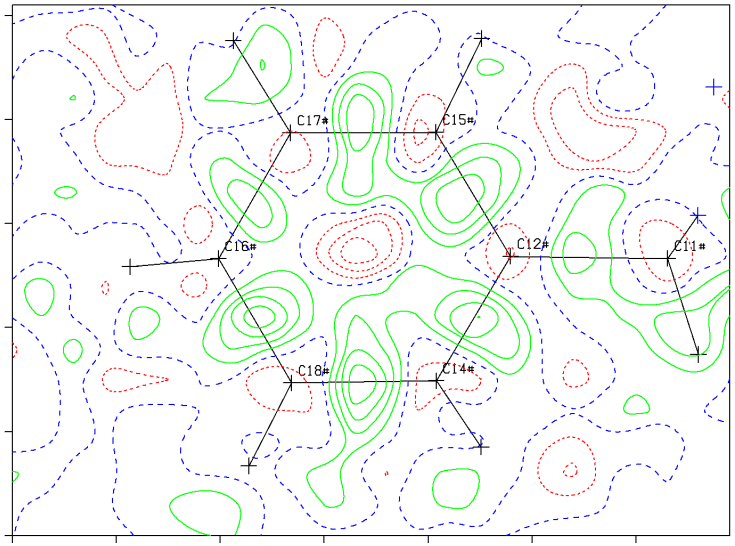
No Voids
No Unusual Contacts
Normal? Difference Density Range

No, The Structure was Deliberately INVENTED

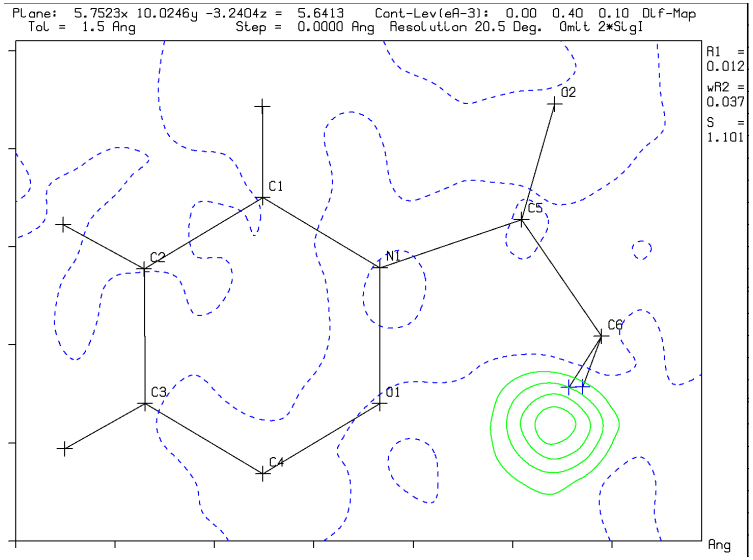
This devious structure was clevery created by Natalie Johnson et al., Newcastle, UK, aiming to beat checkCIF. It was presented as an excellent Poster during the 2015 ECM Congress in Rovinj, Croatia.



But: Every crime leaves its traces …..

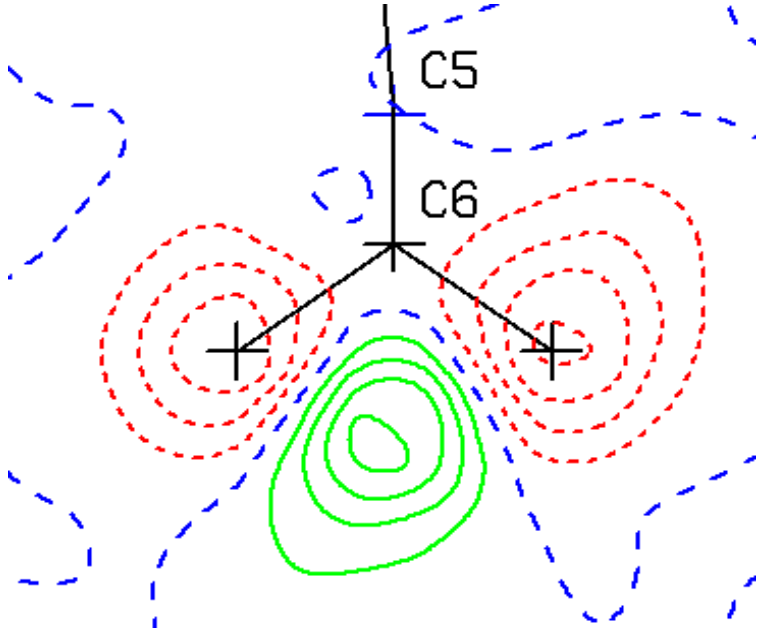


Expected type of difference map



Unusual Actual difference map Density

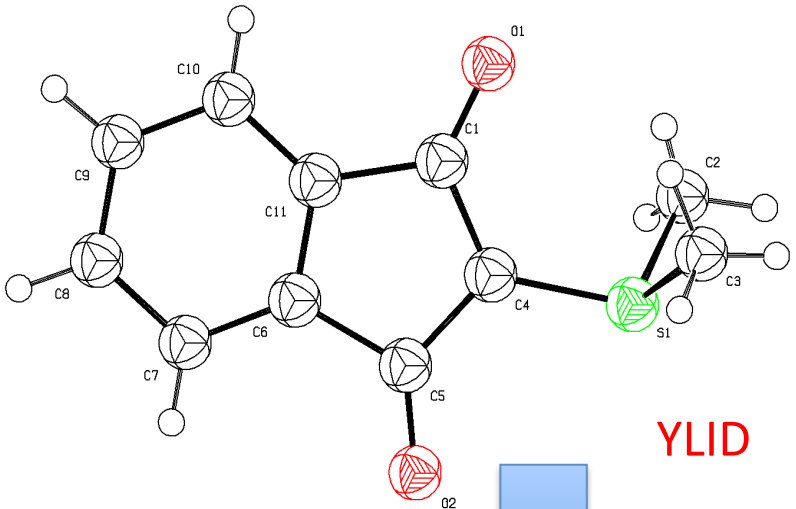
Clear traces of the ‘Crime’ are in the Difference Density Map



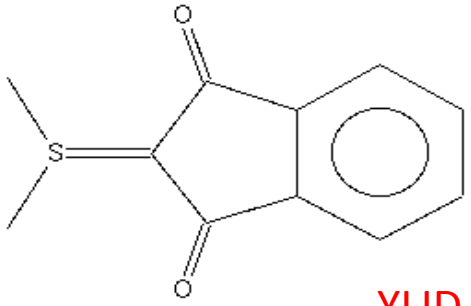
Difference map in the CH2 plane

The CH2 Hydrogen atoms at calculated positions are definitely not in F(obs)

How was Structure ‘Natalie’ created ?

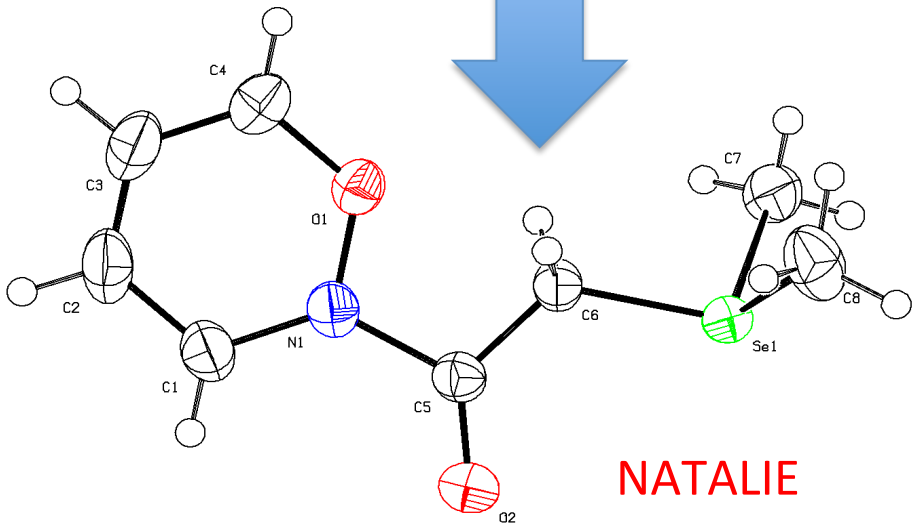


YLID



YLID

‘Starting Material’ (Bruker)



NATALIE

Chemical issue: Se^{+}

No example in the CSD for the 6-membered ring System

Mogul : Se – C outlier

```
# PLATON/CHECK-(140116) versus check.def version of 160106 for Entry: t5_ylid_0m
# Data: natalie.cif - Type: CIF                  Bond Precision   C-C = 0.0027 A
# Refl: natalie.fcf - Type: LIST4                                    Temp = 296 K
# X-ray AgKa              R(int) = 0.015,  wR2/R(int) =  2.3,  Nref/Npar =  10.8
# Cell   9.0257(4)  18.3726(9)   5.9550(3)          90           90           90
# Wavelength 0.56085    Volume Reported     987.49(8)   Calculated     987.49(8)
# SpaceGroup from Symmetry P 21 21 21 Hall: P 2ac 2ab          orthorhombic
#                 Reported P 21 21 21       P 2ac 2ab          orthorhombic
# MoietyFormula C8 H12 N O2 Se
#      Reported C8 H12 N O2 Se
#    SumFormula C8 H12 N O2 Se
#      Reported C8 H12 N O2 Se
# Mr        =    233.15[Calc],    233.15[Rep]
# Dx,gcm-3  =     1.568[Calc],     1.568[Rep]
# Z         =        4[Calc],         4[Rep]
# Mu (mm-1) =     1.999[Calc],     1.999[Rep]
# F000      =     468.0[Calc],     468.0[Rep]  or F000' =    468.99[Calc]
# Reported   T Limits: Tmin=0.704              Tmax=0.745  AbsCorr = MULTI-SCAN
# Calculated T Limits: Tmin=0.726 Tmin'=0.726  Tmax=0.756  Exti    =     0.00000
# Reported   Hmax= 11, Kmax= 22, Lmax=  7, Nref=   2016        , Th(max)= 20.517
# Obs in FCF Hmax= 11, Kmax= 22, Lmax=  7, Nref=   2016[  1196], Th(max)= 20.517
# Calculated Hmax= 11, Kmax= 22, Lmax=  7, Nref=   2025[  1205], Ratio=1.67/1.00
# Reported   Rho(min) = -0.43, Rho(max) =  0.47 e/Ang**3 (From CIF)
# Calculated Rho(min) = -0.43, Rho(max) =  0.42 e/Ang**3 (From CIF+FCF data)
# w=1/[sigma**2(Fo**2)+(0.0239P)**2+  0.1184P], P=(Fo**2+2*Fc**2)/3
# R= 0.0111(  2010), wR2= 0.0339(  2016), S = 1.042      (From CIF+FCF data)
# R= 0.0111(  2010), wR2= 0.0340(  2016), S = 1.044      (From FCF data only)
# R= 0.0111(  2010), wR2= 0.0340(  2016), S = 1.044, Npar=  111, Flack -0.003(2)
# Number Bijvoet Pairs =  820 (100%),       819 Selected for:  Parsons -0.007(1)
# P2(tr) 1.000, P3(tr) 1.000, P3(tw) 0.000, Student-T Nu =   5,  Hooft  0.000(1)
#===============================================================================
     Format: alert-number_ALERT_alert-type_alert-level text

911_ALERT_3_C Missing # FCF Refl Between THmin & STh/L=  0.600          9 Report
977_ALERT_2_C Check the Negative Difference Density on     H6A      -0.37 eA-3
977_ALERT_2_C Check the Negative Difference Density on     H6B      -0.43 eA-3
978_ALERT_2_C Number C-C Bonds with Positive Residual Density          0 Note
#===============================================================================
760_ALERT_1_G CIF Contains no Torsion Angles .................         ? Info
795_ALERT_4_G C-Atom in CIF Coordinate List out of Sequence ..        C4 Note
```

CURRENT PLATON/CHECKCIF VALIDATION REPORT FOR 'NATALIE'

↙ No H-density

↖ No Density on Bonds

What is PLATON

- A collection of SHELX compatible tools bundled in a single program
- The Toolset includes:
  - CheckCIF (Structure Validation)
  - SQUEEZE (Handling Disordered Solvents)
  - ADDSYM (Missed Symmetry)
  - TwinRotMat (Twinning Detection)
  - Bijvoet (Absolute Structure – Hooft(y))

PLATON/CHECKCIF

- Is currently the major validation engine behind the IUCr driven checkCIF validation project of crystal structures.
- It Checks and Reports on (with ALERTS)
  - Completeness of the supplied information
  - Problems with the experimental data
  - Problems with the interpretation of the data
  - Problems with the reported structure model
  - Unusual structural features

Structure Validation now over 20 Years

- The introduction of the CIF standard for data archival made automatic structure checking possible.
- Structure validation provides authors, referees and readers with a list of possibly interesting issues with a structure report that might need to be addressed.
- Currently about 500 tests have been implemented in checkCIF and that number is still increasing on the basis of newly detected issues with supplied CIF's.
- ALERTS are not necessarily errors. They often point at interesting structural features to be discussed.

Some Validation Issues

- A CIF essentially archives the authors interpretation of the underlying experimental diffraction data.
- Archived reflection data are needed for a meaningful evaluation of unusual results and for test calculations.
- Archival of Fo/Fc data (FCF) already solves part of this issue (side effect of its availability: detection and prove of cases of a few hundred seriously frauded structures in Acta Cryst. E)
- Recently: embedding of refinement instructions (res) and unmerged reflection data (hkl) in the CIF urged.
- The embedding of the .res & .hkl into the CIF with SHELXL2014 currently voids the need to supply an FCF.
- IUCr/checkCIF and hkl deposition are now also part of the CSD deposition and archival procedures.

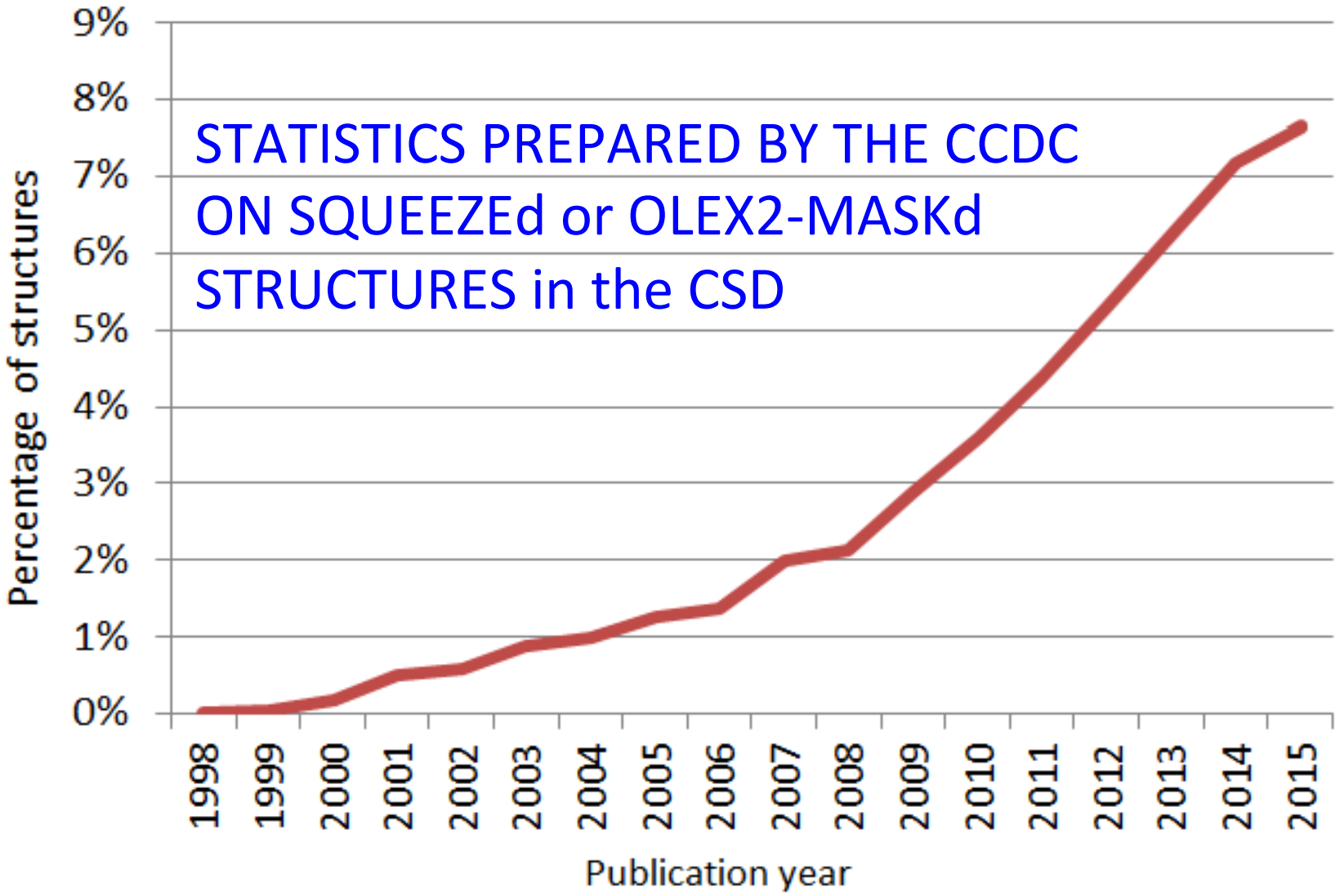
FCF-Validation adds:

- Analysis of the quality of the refinement
- Analysis of the difference map (peaks)
- Detection of void content (SQUEEZE)
- Detection of missing reflections
- Detection of outliers
- Detection of missed twinning
- Check of the absolute structure

Benefits of the SHELXL2014 standard embedding of .res, .hkl (&.fab)

- No need to supply an FCF file along with the CIF as part of a data deposition. The FCF can be created from the embedded data.
- When supplied, the FCF should be 'LIST 4' or 'LIST 8' type
- SHELXL2014 CIF + FCF are the recommended files to run the SQUEEZE tool since the original reflection data are retained in the CIF

The SQUEEZE Tool



PLATON/SQUEEZE

- The current implementation of the SQUEEZE tool to handle disordered solvents is the third generation of a method published by us more than 25 years ago.
- Interfacing with SHELXL2014 refinement solves many earlier issues with SHELX76 & SHELXL97 using .res & .hkl data. [e.g. Modification of the observed data]
- Documentation of the recommended procedure: A.L.Spek (2015) Acta Cryst. C71, 9-18
- http://www.platonsoft.nl/PLATON_HOW_TO.pdf

The Disordered Solvent Problem

- The calculated structure factor Fc can be spit into two parts: Fc = Fc(model) + Fc(solvent)
- Fc(solvent) can be parametrized with an (elaborate) disorder model and refined along with the other model parameters.
- Fc(solvent) can also be approximated with the SQUEEZE tool and used as a fixed contribution to the structure factors in the refinement.
- In simple cases, the first approach is preferred

The PLATON/SQUEEZE Tool

- SQUEEZE, as implemented in PLATON, analyses the content of solvent accessible VOID(s) in a crystal structure. (Q: are the voids empty ?).
- The VOID content will generally involve (heavily) disordered solvent(s) that might be difficult to parameterize meaningfully (e.g. unknown solvents).
- The solvent contribution to the calculated structure factors is approximated by Fourier transformation of the density in the VOID(s) as part of the least-squares refinement of the model parameters. (.fab)
- SQUEEZE does not refine the Fc(model)

The Proper use of SQUEEZE

- It is important that the final CIF archives both the details of the SQUEEZE calculation and the unmerged reflection data. In that way, the calculations can be reconstructed and/or alternative refinement models attempted.
- SHELXL2014 offers all what is needed for that.
- SQUEEZE uses the model parameters taken from *.cif* and merged observed structure factors from the LIST 4 or LIST 8 *.fcf* to calculate solvent F(calc) on *.fab*.
- Final SHELXL refinement will be based on the CIF embedded *.res*, *.hkl* files along with the *.fab* file.

How to SQUEEZE with SHELXL2014

1. Refine a non-solvent model with name.ins & name.hkl (Include ACTA record, NO LIST 6) .
2. Run PLATON/SQUEEZE, based on *name.cif* & name.*fcf* from **1** as 'platon –q name.cif'.
3. Continue SHELXL refinement with the files *name_sq.ins*, name_sq.hkl & *name_sq.fab* from **2** as 'shelxl name_sq'
4. Inspect the .lis & .lst files and Validate

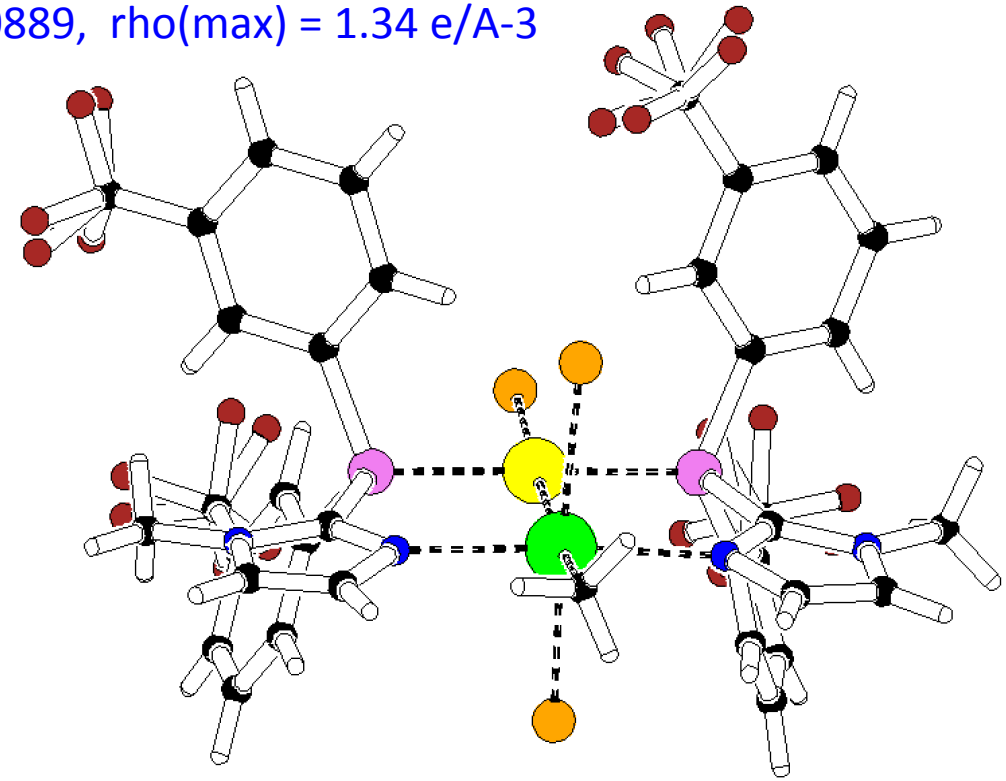
SQUEEZE Disordered Solvent + Twinning

- Step 1: SHELXL refinement based a name.ins (that should include 'ACTA', 'LIST 8', 'BASF' and 'HKLF 5' [or 'TWIN'] records) and a name.hkl file
- Step 2: Run SQUEEZE with the name.cif and name.fcf files produced in Step 1 (i.e. run: platon –q name.cif)
- Step 3: Continue SHELXL refinement with the files name_sq.ins, name_sq.hkl and name_sq.fab produced by PLATON in step 2 ➔ name_sq.cif & name_sq.fcf
- Note: The name_sq.fab file contains the solvent contribution to the SF and the details of SQUEEZE.

SQUEEZE-2016 Example: Coordination Compound

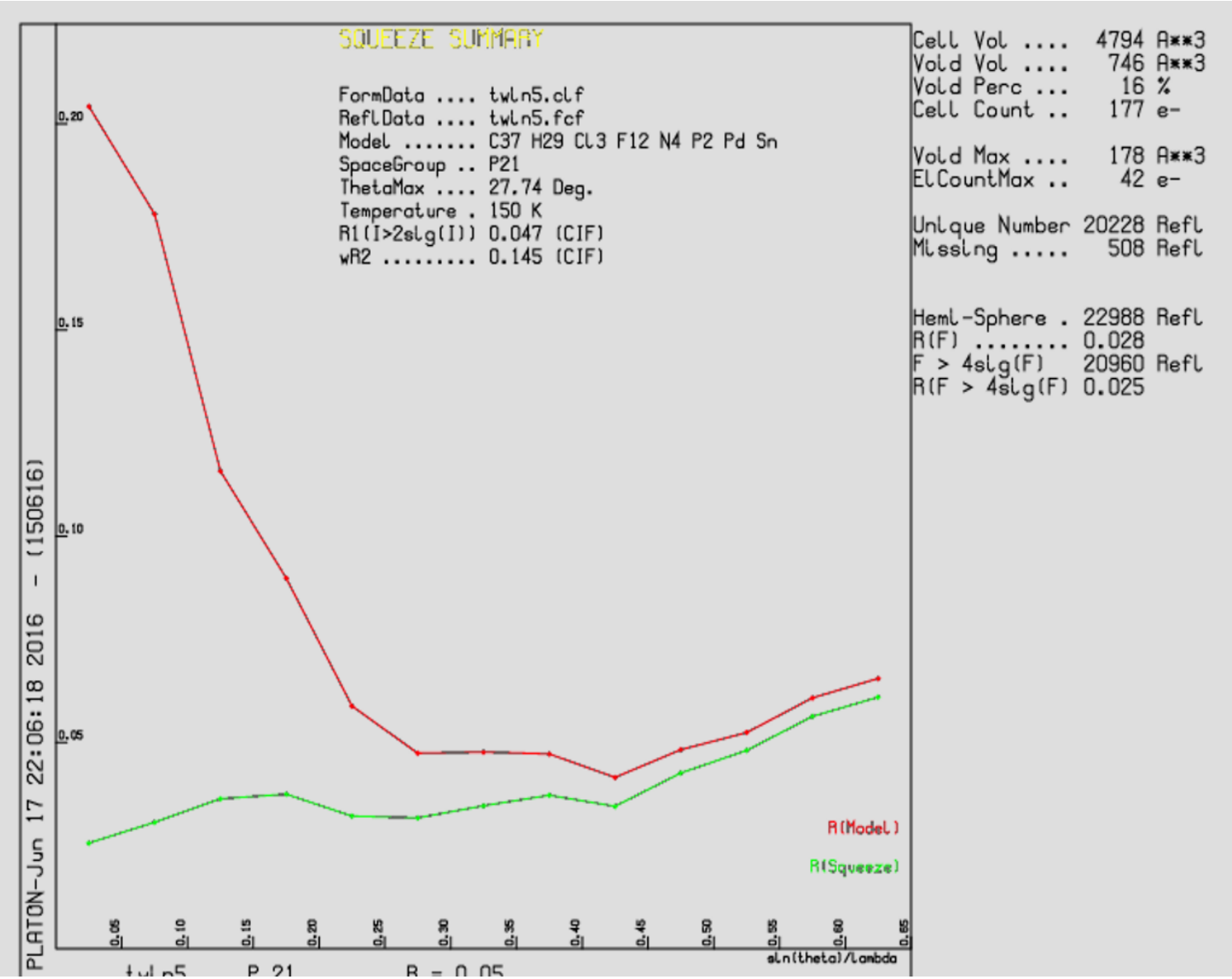
Acetonitril Model: R = 0.0323, wR2 = 0.0889, rho(max) = 1.34 e/A-3

Space Group $P2_1$
Z = 4, Z' = 2
60:40 Twin
Twin axis: (0 0 1)
150 K
TWINABS hklf5 data
Acetonitril solvate



Step 1 (SHELXL2014) ➔ R1 = 0.047, wR2 = 0.1445
Step 2 (SQUEEZE) ➔ 177 electrons found in unit cell
Step 3 (SHELXL2014) ➔ R1 = 0.0275, wR2 = 0.0679, S = 1.064

Effect of on R(F) before and after SQUEEZE as a function of sin(theta)/lambda



Requirements

- There should be no residual unresolved density in the discrete model region of the structure because of its impact on the difference map in the solvent region.
- The data set should be reasonably complete and with sufficient resolution [i.e. sin(theta)/ lambda >0.6].
- There should be no unresolved charge balance issues that might effect the chemistry involved (e.g. The valency of a metal in the ordered part of the structure)

Limitations

- The reported electron count in the solvent region is meaningful only with the supply of a complete and reliable reflection data set.
- The SQUEEZE technique can not handle properly cases of coupled disorder effecting both the model and the solvent region.
- The solvent region is assumed not to contain significant anomalous scatterers (Friedels averaged)

Thank you !

Please send suggestions and examples (with data) of (annoying) issues to:

a.l.spek@uu.nl

More info:www.platonsoft.nl

(including this powerpoint presentation)