Automated Crystal Structure Validation

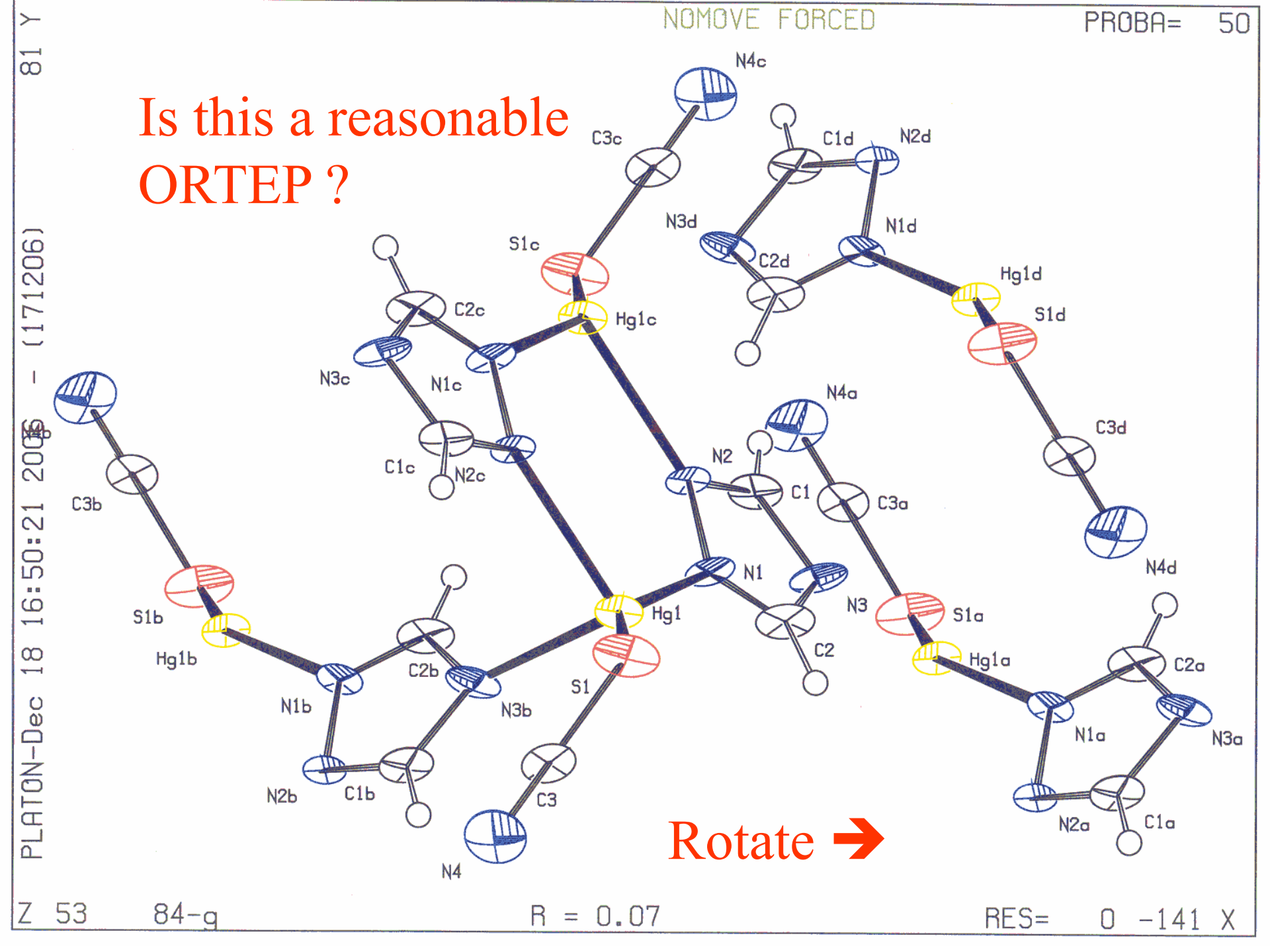
Ton Spek,

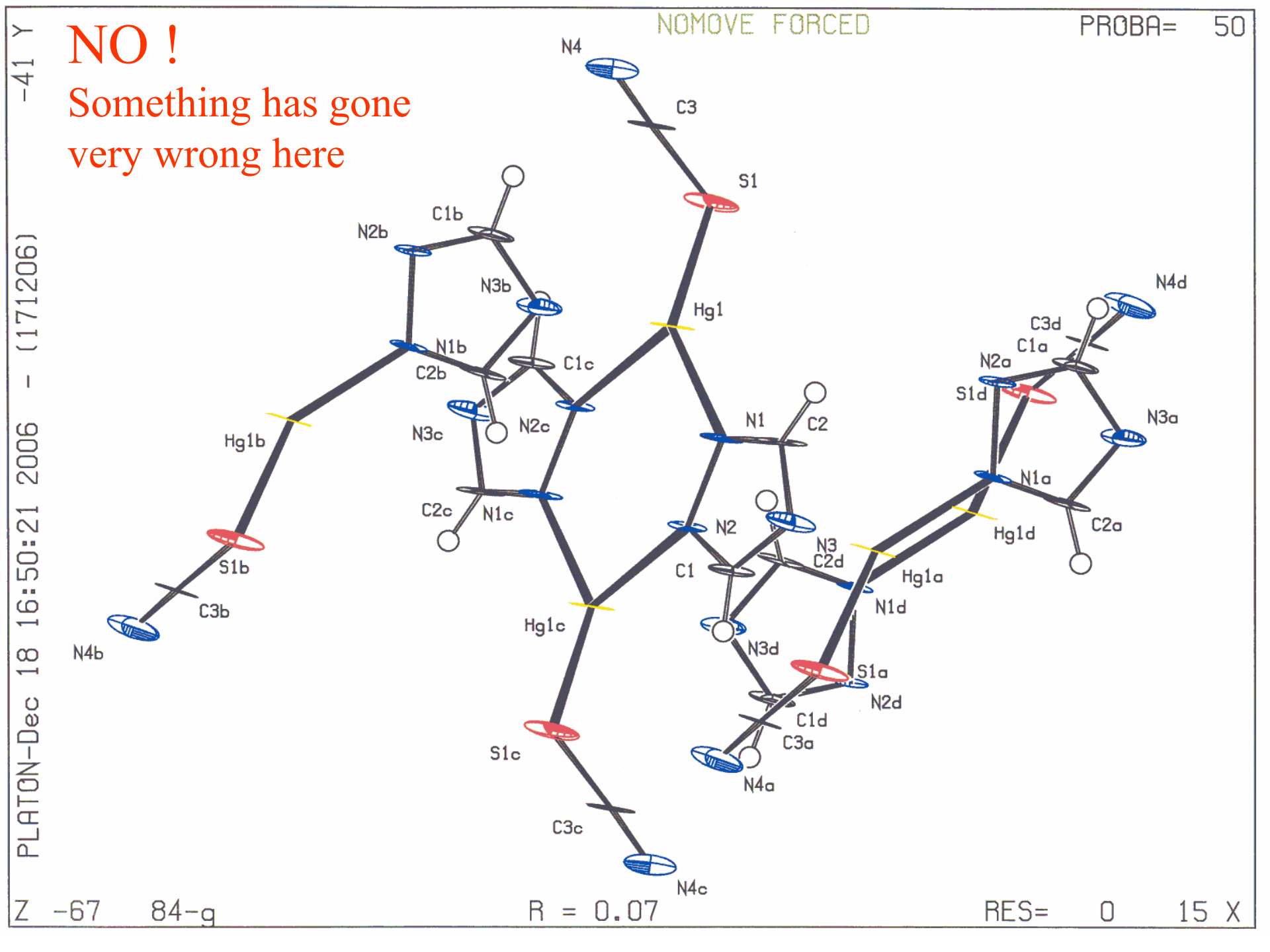
National Single Crystal Facility,
Utrecht University,
Utrecht, The Netherlands

Platon Workshop
Chicago, 24-July-2010.

Why Automated Structure Validation ?

- It is easy to miss problems with a structure as a busy author or as a referee
- Increasingly: Black-Box style analyses done by non-experts
- Limited number of referees & experts available
- It is easy to hide problems with a ball-and-stick style illustration
- Sadly, Fraudulous results and structures published and in the CSD
- Even an ORTEP can hide problems =>


NOMOVE FORCED
PROBA= 50
Is this a reasonable ORTEP ?
Rotate
PLATON-Dec 18 16:50:21 2006 - (171206)
Z 53 84-g
R = 0.07
RES= 0 -141 X
N4c
C3c
C1d
N2d
N3d
N1d
C2d
Hg1d
S1d
S1c
C2c
Hg1c
N3c
N1c
C1c
N2c
C3d
N4d
N4a
N2
C1
C3a
N1
N3
S1a
Hg1a
C2a
N3a
N1a
N2a
C1a
C2
Hg1
S1
C3b
S1b
Hg1b
C2b
N3b
N1b
N2b
C1b
C3
N4


NO !
Something has gone very wrong here
NOMOVE FORCED
PROBA= 50
-41 Y
PLATON-Dec 18 16:50:21 2006 - (171206)
Z -67 84-g
R = 0.07
RES= 0 15 X
N4
C3
S1
C1b
N2b
N3b
Hg1
N1b
C2b
C1c
N3c
N2c
Hg1b
N1
C2
C2c
N1c
N2
N3
C2d
C1
Hg1a
N1d
S1b
C3b
N4b
Hg1c
N3d
S1a
N2d
C1d
C3a
N4a
S1c
C3c
N4c
N4d
C3d
C1a
N2a
S1d
N3a
N1a
C2a
Hg1d

The CIF Standard Solution

- The IUCr Automated Structure Validation project was pioneered and ‘pushed’ by Syd Hall, at that time section editor of Acta Cryst C. by:
- The creation of the CIF Standard for data archival and exchange (Hall et al., (1991) Acta Cryst., A47, 655-685.
- Having CIF adopted by Sheldrick for SHELXL93
- Making CIF the Acta Cryst. submission standard
- Setting up early CIF checking procedures for Acta
- Inviting me to include PLATON checking tools such as ADDSYM and VOID search.

WHAT ARE THE VALIDATION QUESTIONS ?

Single Crystal Structure Validation addresses three simple but important questions:

1 – Is the reported information complete?

2 – What is the quality of the analysis?

3 – Is the Structure Correct?

How is Validation Currently Implemented ?

- Validation checks on CIF data can be executed at any time, both in-house (PLATON/CHECK) or through the WEB-based IUCr CHECKCIF server.
- A file, **check.def**, defines the issues that are tested (currently more than 400) with levels of severity and associated explanation and advise. (www.cryst.chem.uu.nl/platon/CIF-VALIDATION.pdf)
- Most non-trivial tests on the IUCr CheckCIF server are executed with routines in the program PLATON. (Identified as PLATxyz)

VALIDATION ALERT LEVELS

CheckCIF/PLATON creates a report in the form of a list of ALERTS with the following ALERT levels:

- ALERT A – Serious Problem
- ALERT B – Potentially Serious Problem
- ALERT C – Check & Explain
- ALERT G – Verify or Take Notice

VALIDATION ALERT TYPES

1 - CIF Construction/Syntax errors,
Missing or Inconsistent Data.

2 - Indicators that the Structure Model
may be Wrong or Deficient.

3 - Indicators that the quality of the results
may be low.

4 – Info, Cosmetic Improvements, Queries and
Suggestions.

PLATON/CHECK CIF + FCF Results

```
==================================================================================
>>> The Following Model and Quality ALERTS were generated     - (Acta-Mode) <<<
==================================================================================
     Format: alert-number_ALERT_alert-type_alert-level text

148_ALERT_3_B su on the      a  - Axis is Too Large (x 1000) .          10 Ang.
148_ALERT_3_B su on the      b  - Axis is Too Large (x 1000) .           8 Ang.
148_ALERT_3_B su on the      c  - Axis is Too Large (x 1000) .          30 Ang.
230_ALERT_2_B Hirshfeld Test Diff for    O1     --  C1       ..       26.27 su
230_ALERT_2_B Hirshfeld Test Diff for    O2     --  C1       ..       11.10 su
242_ALERT_2_B Check Low       Ueq as Compared to Neighbors for          C1
420_ALERT_2_B D-H Without Acceptor       O1     -   H1      ...         ?
#==================================================================================
230_ALERT_2_C Hirshfeld Test Diff for    C1     --  C2       ..        6.87 su
242_ALERT_2_C Check Low       Ueq as Compared to Neighbors for         C11
911_ALERT_3_C Missing # FCF Refl Between THmin & STh/L=  0.595         11
913_ALERT_3_C Missing # of Very Strong Reflections in FCF ....          1
#==================================================================================
860_ALERT_3_G Note: Number of Least-Squares Restraints .......         2
#==================================================================================
#==================================================================================

==================================================================================
>>> The Following Improvement and Query ALERTS were generated - (Acta-Mode) <<<
==================================================================================
926_ALERT_1_B Reported and Calculated   R1 * 100.0 Differ by .       -0.81
927_ALERT_1_B Reported and Calculated  wR2 * 100.0 Differ by .       -2.26
#==================================================================================
042_ALERT_1_C Calc. and Reported MoietyFormula Strings  Differ          ?
790_ALERT_4_C Centre of Gravity not Within Unit Cell: Resd.  #          2
              C8 H6 O4
928_ALERT_1_C Reported and Calculated    S value   Differ by .       -0.13
#==================================================================================
128_ALERT_4_G Non-standard setting of Space-group P21/c   ....      P21/n
#==================================================================================
```

checkCIF/PLATON report (publication check)

No syntax errors found.
Please wait while processing ....

CIF dictionary
Interpreting this report

Datablock: I

Bond precision: C-C = 0.0157 A Wavelength=0.71073
Cell: a=7.6336(15) b=27.725(6) c=12.051(2)
alpha=90 beta=98.80(3) gamma=90
Temperature: 153 K

	Calculated	Reported
Volume	2520.5(9)	2520.5(9)
Space group	P 21/n	P 21/n
Hall group	-P 2yn	-P 2yn
Moiety formula	C28 H32 Br0.97 Cl1.03 N2 Ru	C28 H32 Br0.97 Cl1.03 N2 Ru
Sum formula	C28 H32 Br0.97 Cl1.03 N2 Ru	C28 H32 Br0.97 Cl1.03 N2 Ru
Mr	611.69	611.69
Dx,g cm-3	1.612	1.612
Z	4	4
Mu (mm-1)	2.290	2.290
F000	1237.9	1237.9
F000'	1232.67	
h,k,lmax	9,33,14	9,33,14
Nref	4452	4449
Tmin,Tmax	0.726,0.955	0.581,0.955
Tmin'	0.546	

Correction method= MULTI-SCAN
Data completeness= 0.999 Theta(max)= 25.030
R(reflections)= 0.0891(3288) wR2(reflections)= 0.2547(4449)
S = 1.037 Npar= 306

The following ALERTS were generated. Each ALERT has the format
test-name_ALERT_alert-type_alert-level.
Click on the hyperlinks for more details of the test.

Alert level B

PLAT083_ALERT_2_B SHELXL Second Parameter in WGHT unusually Large. 37.00

Alert level C

RFACR01_ALERT_3_C The value of the weighted R factor is > 0.25
Weighted R factor given 0.255
PLAT301_ALERT_3_C Main Residue Disorder ..... 6.00 Perc.

Which Key Validation Issues are Addressed

- Missed Space Group symmetry (“being Marshed”)
- Wrong chemistry (Mis-assigned atom types).
- Too many, too few or misplaced H-atoms.
- Unusual displacement parameters.
- Hirshfeld Rigid Bond test violations.
- Missed solvent accessible voids in the structure.
- Missed Twinning.
- Absolute structure
- Data quality and completenes.

Some Relevant ALERTS

Wrong atom type assignments generally cause:

- Serious Hirshfeld Rigid Bond Violation ALERTS
- Larger than expected difference map minima and maxima.
- wR2 >> 2 * R1
- High values for the SHELXL refined weight parameter

Evaluation and Performance

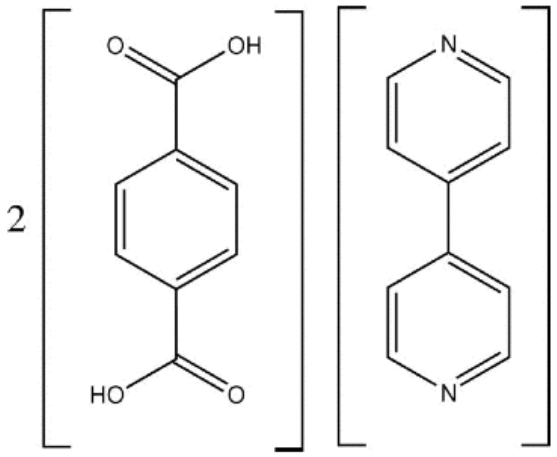
- The validation scheme has been very successful for Acta Cryst. C & E in setting standards for quality and reliability.
- The missed symmetry problem has been solved for the IUCr journals (unfortunately not generally yet: There are still numerous ‘Marshable’ structures).
- Most major chemical journals currently have now some form of a validation scheme implemented.
- Recently included: FCF validation

FCF-VALIDATION

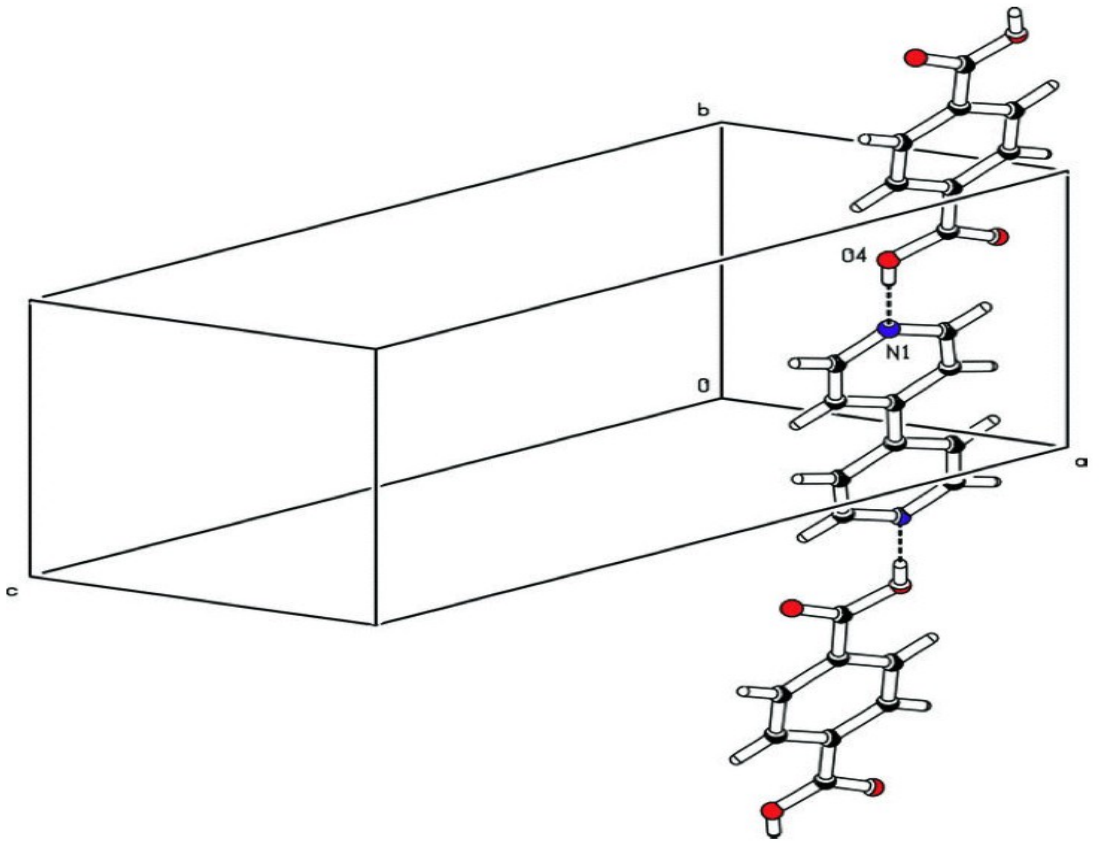
- Check of the CIF & FCF data Consistency (including R-values, cell dimensions)
- Check of Completeness of the reflection data set.
- Automatic Detection of ignored twinning
- Detection of Applied Twinning Correction without having been Reported in the paper.
- Validity check of the reported Flack parameter value against the Hooft parameter value.
- Analysis of the details of the Difference Density Fourier Map for unreported features.

Sloppy, Novice or Fraudulent ?

- Errors are easily made and unfortunately not always discernable from fraud.
- Wrong element type assignments can be caused as part of an incorrect analysis of an unintended reaction product.
- Alternative element types can be (and have been) substituted deliberately to create a ‘new publishable’ structures.
- Reported and calculated R-values differing in the first relevant digit !?
- FCF Validation is the tool to sort out this type of issues.

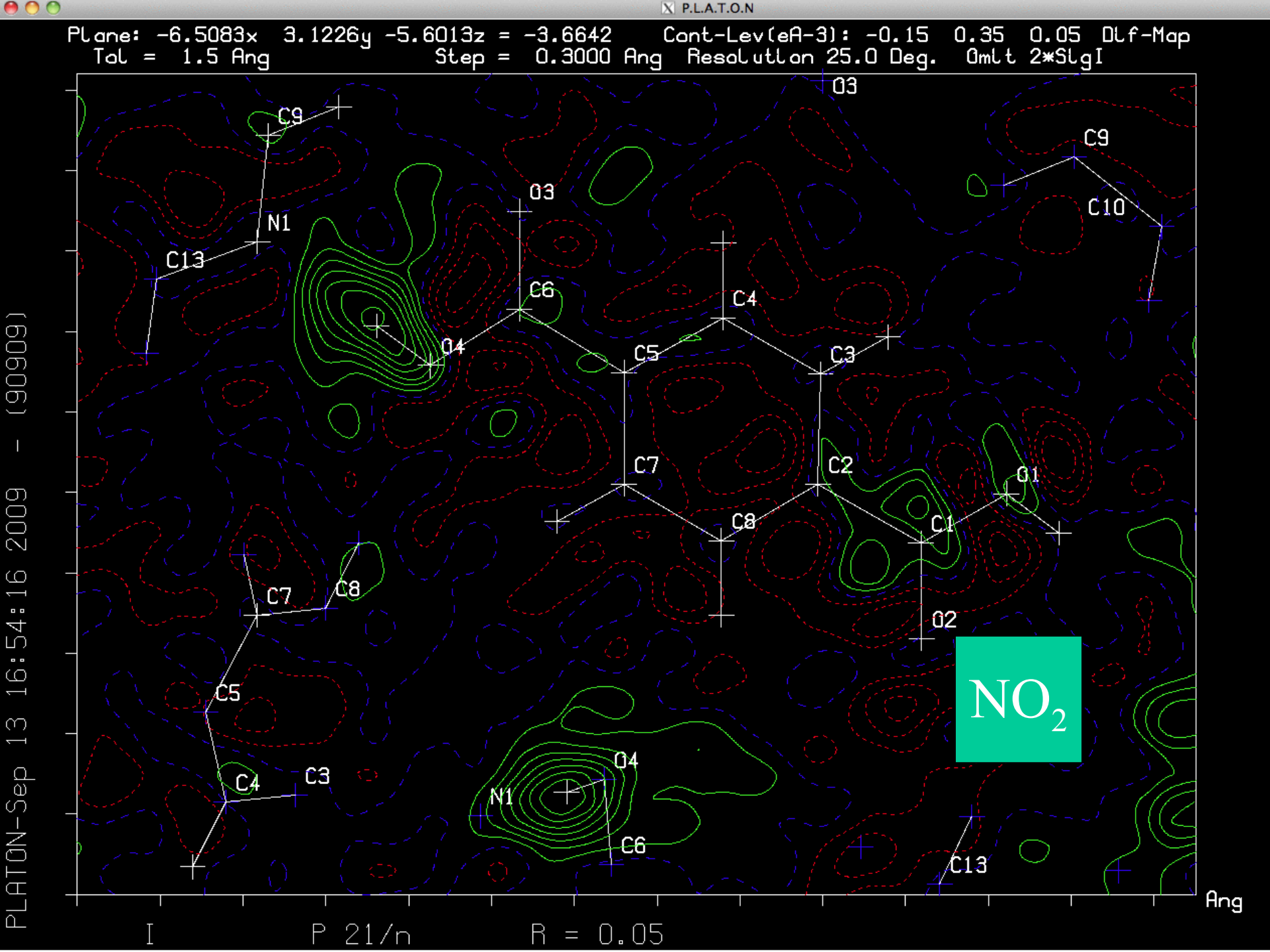


Published structure is claimed to form an infinite hydrogen bonded chain



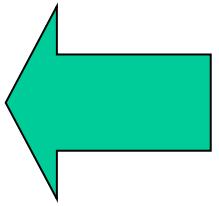
However: This structure does not include a dicarboxylic acid but the previously published para-nitrobenzoic acid.

PROOF: Difference map calculated without the 2 carboxylic H-atoms


P.L.A.T.O.N
Plane: -6.5083x 3.1226y -5.6013z = -3.6642 Cont-Lev(eA-3): -0.15 0.35 0.05 Dlf-Map
Tol = 1.5 Ang Step = 0.3000 Ang Resolution 25.0 Deg. Omit 2*SigI
PLATON-Sep 13 16:54:16 2009 - (90909)
O3
C9
C10
N1
C13
C6
C4
O4
C5
C3
C7
C2
O1
C8
C1
O2
NO_2
Ang
I P 21/n R = 0.05

There are clear ALERTS ! But apparently ignored

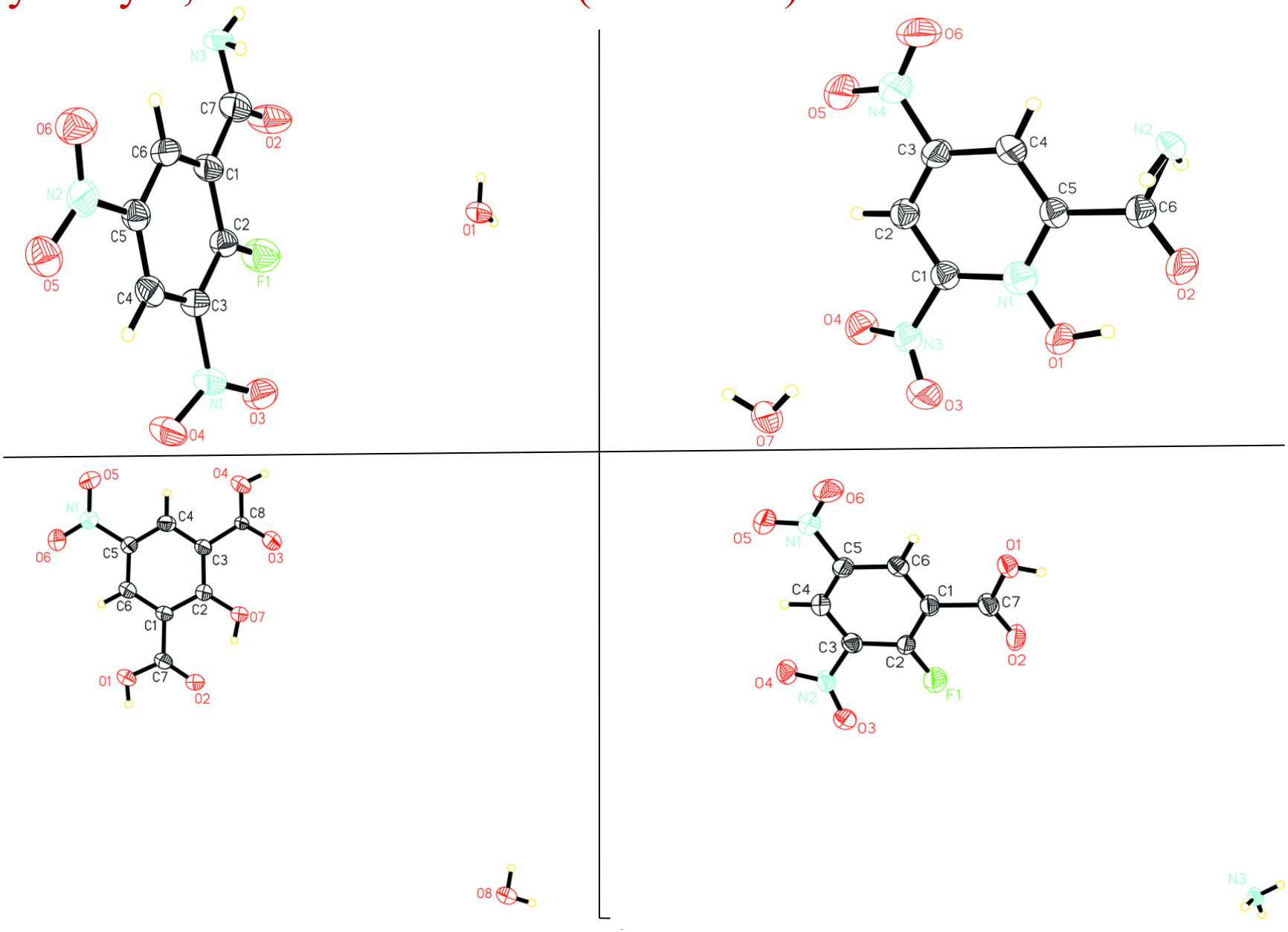
```
=================================================================================
>>> The Following Model and Quality ALERTS were generated     - (Acta-Mode) <<<
=================================================================================
     Format: alert-number_ALERT_alert-type_alert-level text

148_ALERT_3_B su on the      a  - Axis is Too Large (x 1000) .          10 Ang.
148_ALERT_3_B su on the      b  - Axis is Too Large (x 1000) .           8 Ang.
148_ALERT_3_B su on the      c  - Axis is Too Large (x 1000) .          30 Ang.
230_ALERT_2_B Hirshfeld Test Diff for    O1     --  C1       ..       26.27 su
230_ALERT_2_B Hirshfeld Test Diff for    O2     --  C1       ..       11.10 su
242_ALERT_2_B Check Low       Ueq as Compared to Neighbors for          C1
420_ALERT_2_B D-H Without Acceptor       O1     -   H1      ...          ?
#=================================================================================
230_ALERT_2_C Hirshfeld Test Diff for    C1     --  C2       ..        6.87 su
242_ALERT_2_C Check Low       Ueq as Compared to Neighbors for         C11
911_ALERT_3_C Missing # FCF Refl Between THmin & STh/L=  0.595          11
913_ALERT_3_C Missing # of Very Strong Reflections in FCF ....          1
#=================================================================================
860_ALERT_3_G Note: Number of Least-Squares Restraints .......         2
#=================================================================================
#=================================================================================

=================================================================================
>>> The Following Improvement and Query ALERTS were generated - (Acta-Mode) <<<
=================================================================================
926_ALERT_1_B Reported and Calculated   R1 * 100.0 Differ by .       -0.81
927_ALERT_1_B Reported and Calculated  wR2 * 100.0 Differ by .       -2.26
#=================================================================================
042_ALERT_1_C Calc. and Reported MoietyFormula Strings  Differ          ?
790_ALERT_4_C Centre of Gravity not Within Unit Cell: Resd.  #          2
              C8 H6 O4
928_ALERT_1_C Reported and Calculated    S value   Differ by .       -0.13
#=================================================================================
128_ALERT_4_G Non-standard setting of Space-group P21/c   ....      P21/n
#=================================================================================
```



The Ultimate Shame

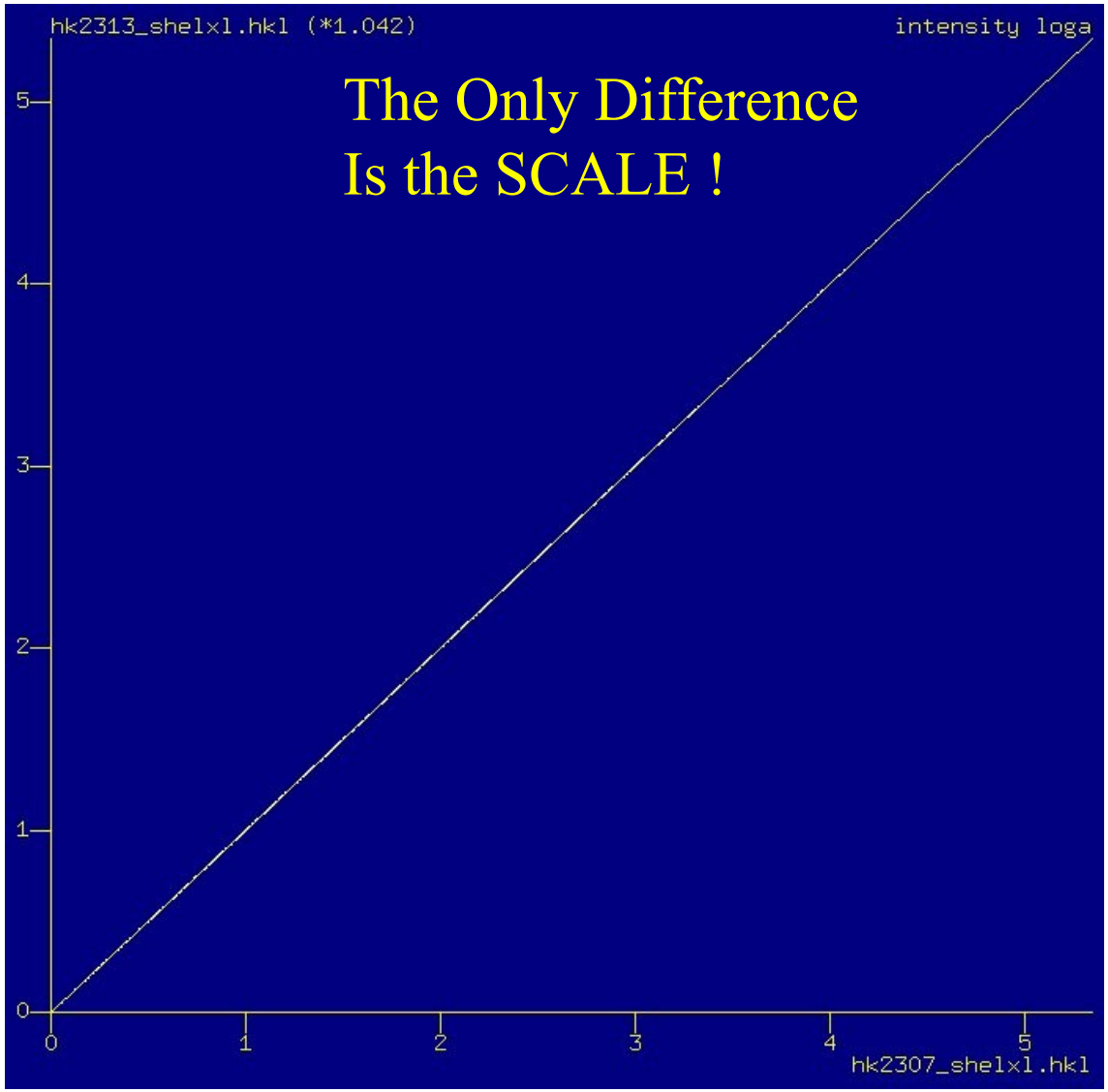
- Recently a whole series of ‘isomorphous’ substitions was detected for an already published structure.
- Similar series have now been detected for coordination complexes (Transition metals and lanthanides)
- How could referees let those pass ?
- Over 100 structures now retracted
- Fraud detected by looking at all papers of the same authors of a ‘strange’ structure (and their institutions)

BogusVariations (with Hirshfeld ALERTS) on the Published Structure 2-hydroxy-3,5-nitrobenzoic acid (ZAJGUM)



Comparison of the Observed data for two ‘isomorphous’ compounds.

Tool: platon –d name1.fcf name2.fcf



Conclusion
The Same
Data !

SLOPPY

Or

FRAUD ?

Info

www.cryst.chem.uu.nl/platon/CIF-VALIDATION.pdf
www.cryst.chem.uu.nl/platon/FCF-VALIDATION.pdf

- Papers on structure validation:
- A.L.Spek (2003). J. Appl. Cryst. 36, 7-13.
- A.L.Spek (2009). Acta Cryst. D65, 148-155.