Structure Validation in Chemical Crystallography

with CheckCIF/PLATON

Ton Spek,

National Single Crystal Service Facility, Utrecht University

The Netherlands

Freiburg,17-Sept-2009.

Overview of this Lecture

- Why Crystal Structure Validation ?
- What are the Validation Questions ?
- How is Validation Implemented ?
- What key problems are addressed ?
- Some Examples of Detected Issues.
- Evaluation and Performance.
- Summary.

Why Crystal Structure Validation ?

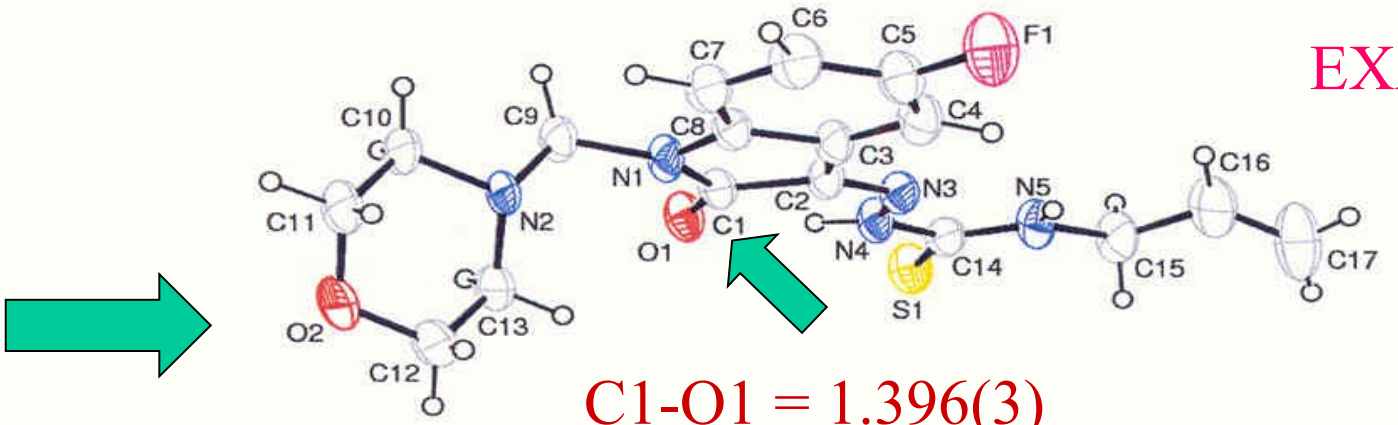
- The explosion of Reported Structure Determinations every year.
- Many analyses are done nowadays Black-Box style by non-specialists.
- There is a limited number of experts/referees trained and available to detect common pitfalls in publications.
- Validation offers a list of ALERTed (i.e. unusual) issues that require special attention of the analyst, the specialist and the referee.
- Validation tries to be helpful and sets quality standards.
- New and sadly: Detection of clear fraud and fraudulent practices.

Just two Examples of problems with entries archived in the CSD

- The CSD is a rich source of chemical information.
- However: An analysis of the ~ 500000 structures in the CSD learns that a not insignificant number of the entries has undetected serious errors.
- Nearly all searches in the CSD for statistical info show outliers that, when inspected closely, can be shown to be erroneous.
- The following two problem cases were detected as part of one such a search for short S…S contacts.

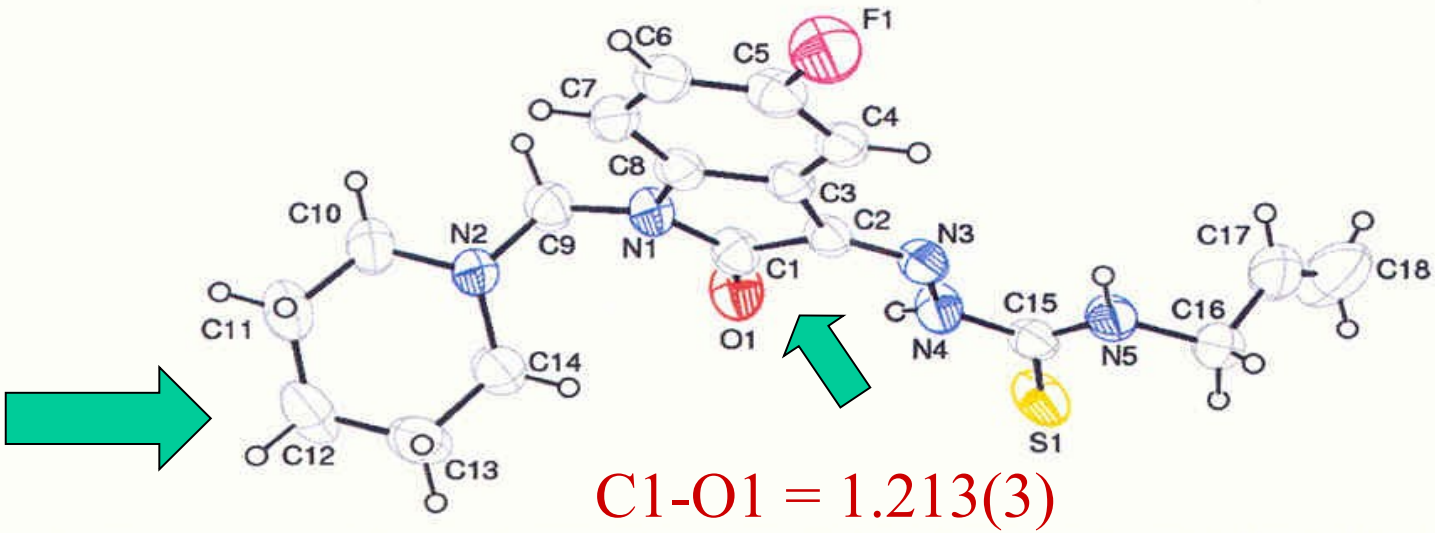
Two Related Structures – Strange Metrical Differences

EXAMPLE 1



C1-O1 = 1.396(3)

Figure 2. Molecular structure of **3e** showing the atom labeling scheme. Displacement ellipsoids are drawn as 30% probability level.



C1-O1 = 1.213(3)

Figure 3. Molecular structure of **3f** showing the atom labeling scheme. Displacement ellipsoids are drawn as 30% probability level.

Huge Geometry Differences !?

Table 2. Selected bond lengths and angles (°, Å) for **3e** and **3f**

3e		**3f**	
C1–O1	1.396(3)	C1–O1	1.213(3)
C1–N1	1.313(3)	C1–N1	1.367(3)
C1–C2	1.612(4)	C1–C2	1.509(4)
C2–N3	1.163(3)	C2–N3	1.291(3)
C2–C3	1.575(4)	C2–C3	1.448(3)
C5–F1	1.567(4)	C5–F1	1.355(3)
C8–N1	1.630(4)	C8–N1	1.409(3)
C9–N2	1.253(3)	C9–N2	1.436(3)
C9–N1	1.478(3)	C9–N1	1.459(3)
C10–N2	1.423(3)	C10–N2	1.463(3)
C10–C11	1.302(4)	C10–C11	1.486(4)
C11–O2	1.481(4)	C11–C12	1.503(5)
C12–O2	1.357(4)	C12–C13	1.516(5)
C12–C13	1.311(4)	C13–C14	1.500(4)
C13–N2	1.532(4)	C14–N2	1.471(3)
C14–N4	1.245(3)	C15–N5	1.323(3)
C14–N5	1.451(4)	C15–N4	1.369(3)
C14–S1	1.864(3)	C15–S1	1.669(3)
C15–C16	1.645(5)	C16–C17	1.440(4)
C15–N5	1.327(3)	C16–N5	1.456(3)
C16–C17	1.238(5)	C17–C18	1.201(5)
N3–N4	1.492(3)	N3–N4	1.341(3)

EXAMPLE 1

There is obviously a problem with 3e:
Where were the referees of this paper ?

TIKRUB

Reference:	Xue-Mei Li, Su-Juan Ye, Xu-Li Tang, Cai-Feng Ding, Shu-Sheng Zhang (2006) *Asian J.Chem.* ,18,491
Formula:	$C_{22}\ H_{27}\ N_3\ O_{11}\ S_1,H_2\ O_1$
Compound Name:	N-(2,3,4,6-Tetra-O-acetyl-β-D-glucopyranosyl) thiocarbamic 2,4-dihydroxybenzoyl hydrazine monohydrate

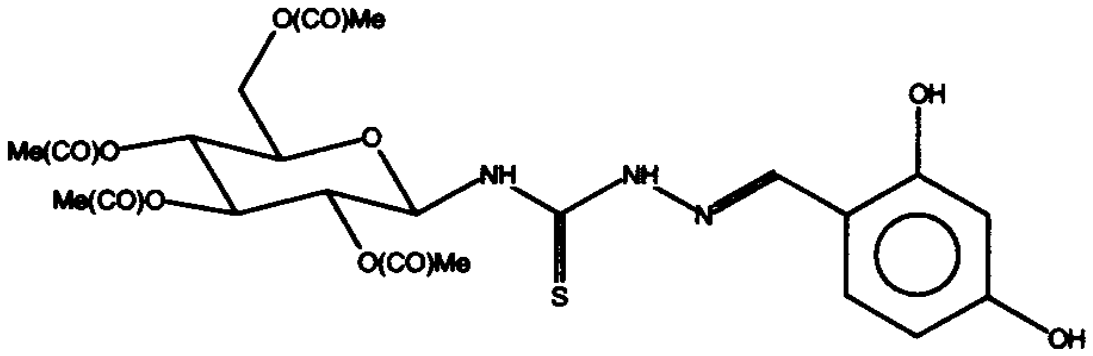
Space Group:	P43212	***Cell:***	***a***	12.020(0)	***b***	12.020(0)	***c*** 38.877(3)
Space Group No.:	96	***(Å,°)***	α	90.00	β	90.00	γ 90.00

R-Factor (%):	6.47	***Temperature(K):***	293	***Density(g/cm³):***	1.323

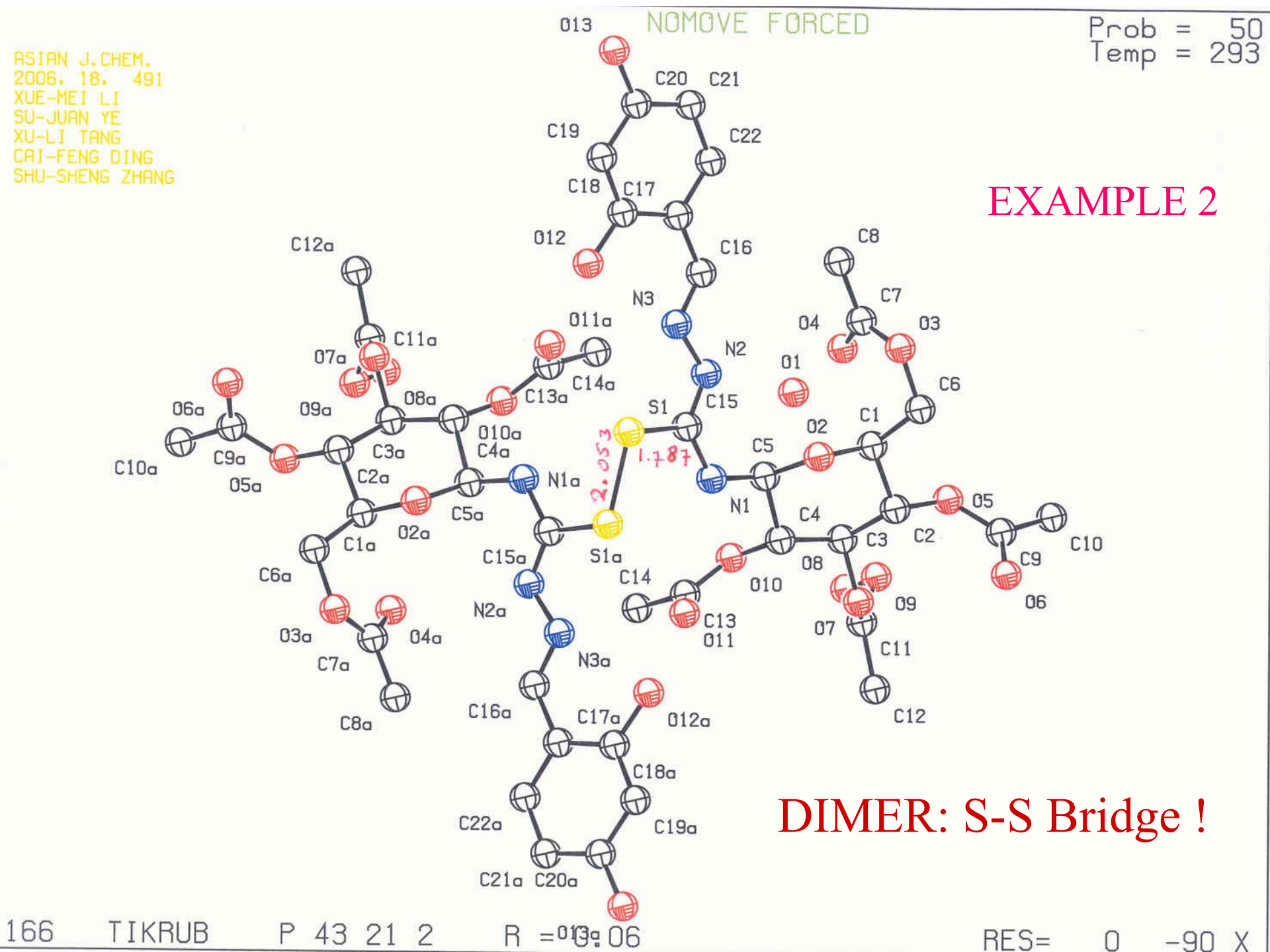
EXAMPLE 2

Reported as Monomer

BUT ➔



H_2O


ASIAN J.CHEM.
2006, 18, 491
XUE-MEI LI
SU-JUAN YE
XU-LI TANG
CAI-FENG DING
SHU-SHENG ZHANG
NOMOVE FORCED
Prob = 50
Temp = 293
EXAMPLE 2
DIMER: S-S Bridge !
166 TIKRUB P 43 21 2 R = 0.06
RES= 0 -90 X

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?

Implementation Problems of Structure Validation Around 1990

- Multiple Data Storage Types (often listing files).
- No Standard Computer Readable Format for data exchange.
- Data entry for publication via retyping in the manuscript.
- Thus: multiple typo's in Published Data.
- CSD Database Archival by Retyping from the published paper.
- Published data often incomplete.
- No easy numerical checking for referees etc.

How is Validation Currently Implemented ?

- The results of a structure analysis are now required to be available in the computer readable CIF format.
- Validation checks can be executed at any time both in-house or through the WEB-based IUCr CHECKCIF server.
- A file (Check.def) defines the issues that are tested with levels of severity and associated explanation and advise.
- Most non-trivial tests are executed by routines in the program PLATON

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.

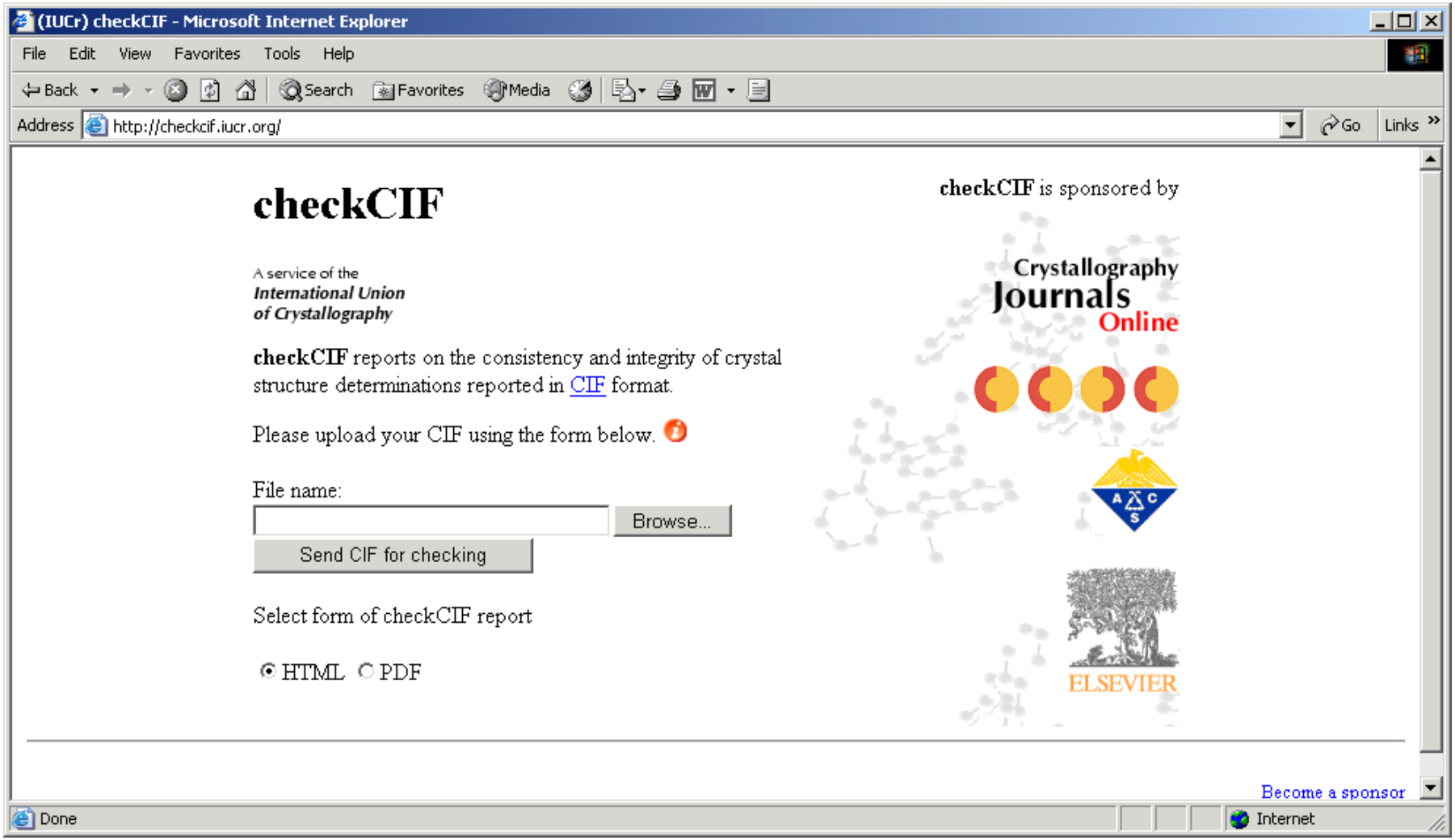
2 - Indicators that the Structure Model
may be Wrong or Deficient.

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

4 – Info, Cosmetic Improvements, Queries and
Suggestions.

Simple Validation Issues

- Many data sets are apparently collected at either 293(2) or 273 K
- Program defaults or values from previous papers are retained.
- Data collected with a CCD system and corrected for absorption with Psi-scans !


(IUCr) checkCIF - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://checkcif.iucr.org/
Go Links
checkCIF
A service of the
International Union
of Crystallography
checkCIF reports on the consistency and integrity of crystal structure determinations reported in CIF format.
Please upload your CIF using the form below.
File name:
Browse...
Send CIF for checking
Select form of checkCIF report
HTML PDF
checkCIF is sponsored by
Crystallography
Journals
Online
ELSEVIER
Become a sponsor
Done
Internet

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.

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
#================================================================================
```

The CIF Standard Solution

- CIF-Standard Proposal for Data Archival and Exchange:

  S.R. Hall, F.H. Allen, I.D. Brown (1991). Acta Cryst. A47, 655-685.
- Pioneered and Adopted by the International Union for Crystallography and Syd Hall (XTAL-System)
- Early adoption by the author of the now most used software package SHELXL97 (G.M.Sheldrick)
- Most current software now reads & writes CIF

CIF File Structure

- Both Computer and Human Readable Ascii encoded file
- Free Format
- Mostly 80 columns wide (maximum 2048)
- Parsable in units (Data names and Values)
- Data Order Flexible
- Dataname and Value associations
- loops

Constructs

- **data_*name***
  where *name* the choosen identifier of the data
- Data associations e.g.
  _ cell_length_a 16.6392(2)
- Repetition (loop)
  loop_
  __symmetry_equiv_pos_as_xyz
  'x, y, z'
  '-x, y+1/2, -z'

Construct for Text

- Text can be included between semi-columns
- Used for Acta Cryst. Section C & E submissions
- Example

_publ_section_comment

;

This paper presents to the best of our knowledge the first example of a very important MOF contruct.

;

(IUCr) Crystal Structure Communications Online author services - Iceweasel

File Edit View History Bookmarks Tools Help

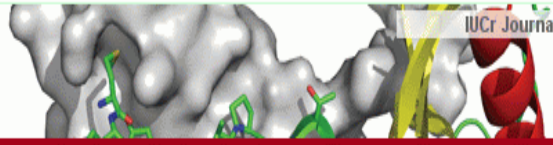
http://journals.iucr.org/c/services/authorservices.html

Most Visited Getting Started Latest Headlines

Crystal Structure Communications Online

Journals Online | supplementary files
search | contact us
help | terms of use
subscribe | site index

International Union of CRYSTALLOGRAPHY



IUCr Journals | International Tables | World Directory

iucr | journals | books | news | education | people | resources

Home

Enter all or part of a CIF dataname to search for its definition in a CIF Dictionary.

Search term

_exptl_crystal_color

Which dictionary? Core CIF

Submit Query Reset

Dictionary Lookup Example

Notes

1. Searches are case-insensitive.
2. To search on category descriptions in DDL1 dictionaries (pseudo-datanames that include square brackets), precede the brackets with backslash characters (*e.g.* `_publ_\[\]`)

The International Union of Crystallography is a non-profit scientific union serving the world-wide interests of crystallographers and other scientists employing crystallographic methods.

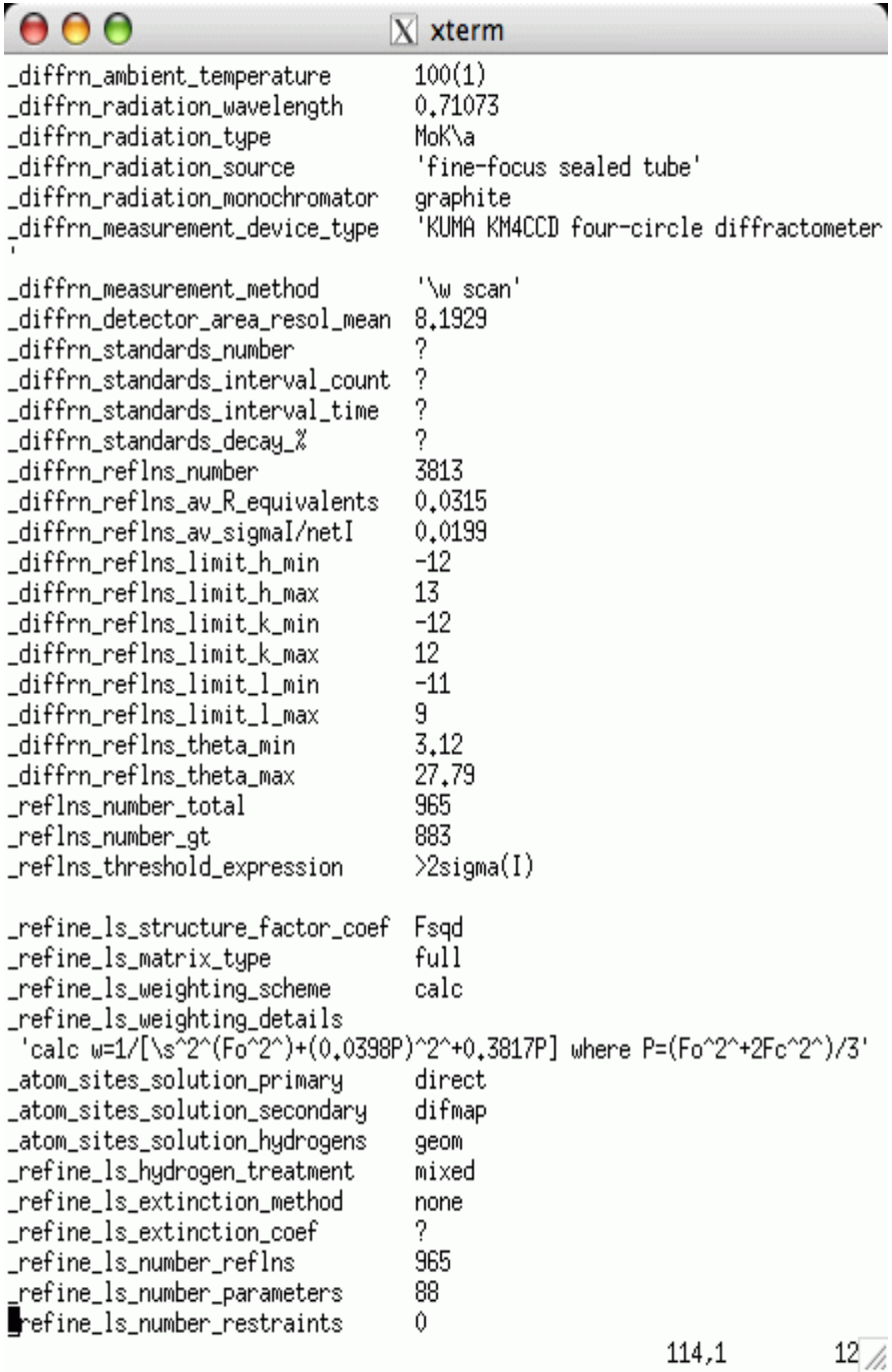
CIF Example File

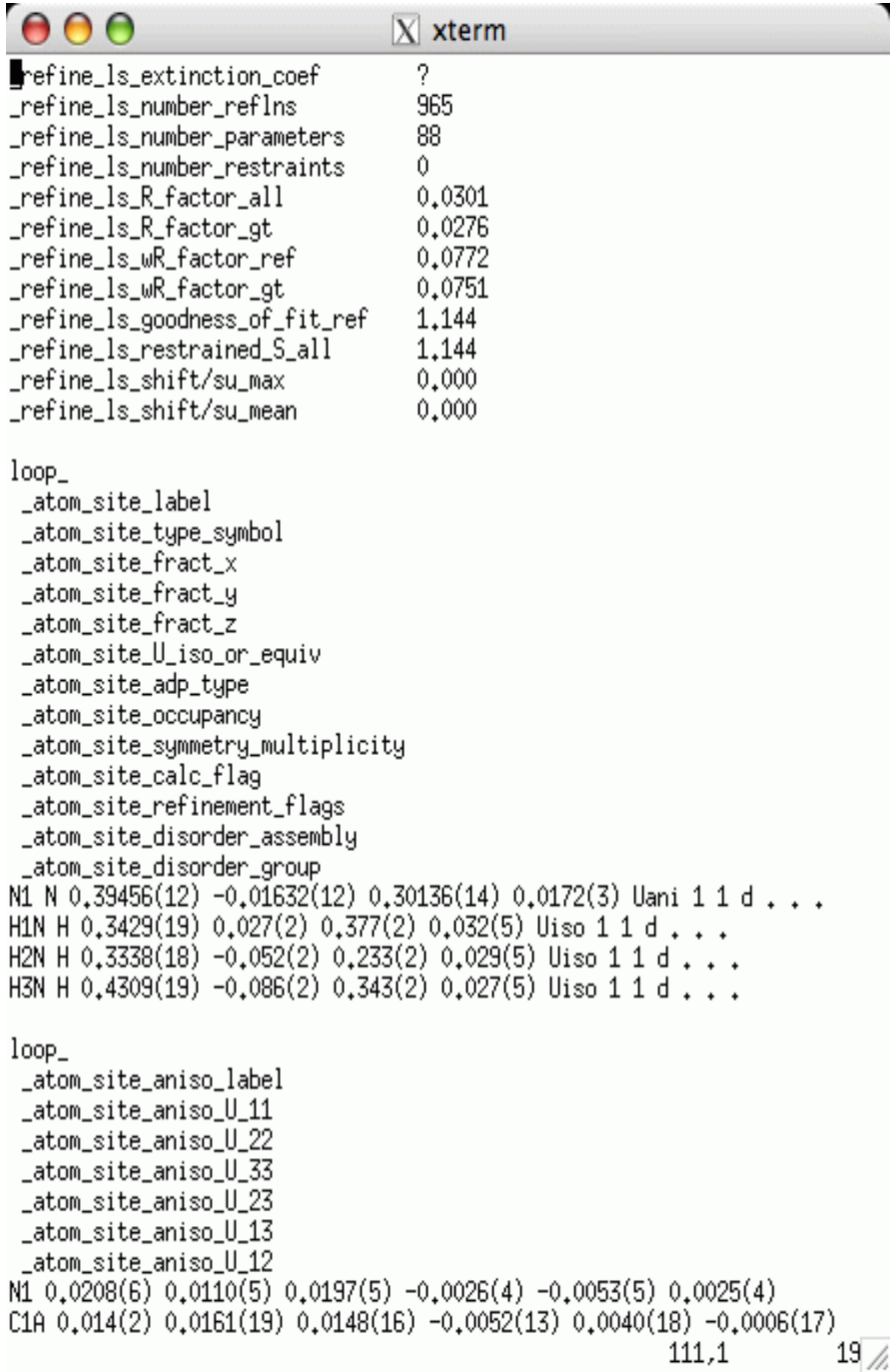
xterm

```
data_100K
_audit_creation_method            SHELXL-97
_chemical_name_systematic
;
Butane-1,4-diaminium sulfate
;
_chemical_name_common             'putrescinium sulfate'
_chemical_melting_point           ?
_chemical_formula_moiety          '2(C2 H7 N), O4 S'
_chemical_formula_sum              'C4 H14 N2 O4 S'
_chemical_formula_weight          186.23
loop_
 _atom_type_symbol
 _atom_type_description
 _atom_type_scat_dispersion_real
 _atom_type_scat_dispersion_imag
 _atom_type_scat_source
 'C'  'C'   0.0033   0.0016
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
 'H'  'H'   0.0000   0.0000
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
 'N'  'N'   0.0061   0.0033
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
 'O'  'O'   0.0106   0.0060
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
 'S'  'S'   0.1246   0.1234
 'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
_symmetry_cell_setting            orthorhombic
_symmetry_space_group_name_H-M    'P n m a'
_symmetry_space_group_name_Hall   '-P 2ac 2n'
loop_
 _symmetry_equiv_pos_as_xyz
 'x, y, z'
 '-x+1/2, -y, z+1/2'
 'x+1/2, -y+1/2, -z+1/2'
 '-x, y+1/2, -z'
 '-x, -y, -z'
 'x-1/2, y, -z-1/2'
 '-x-1/2, y-1/2, z-1/2'
 'x, -y-1/2, z'
                                   1,1           To
```

xterm

```
 'x, y, z'
 '-x+1/2, -y, z+1/2'
 'x+1/2, -y+1/2, -z+1/2'
 '-x, y+1/2, -z'
 '-x, -y, -z'
 'x-1/2, y, -z-1/2'
 '-x-1/2, y-1/2, z-1/2'
 'x, -y-1/2, z'

_cell_length_a                    9.9722(4)
_cell_length_b                    9.4675(4)
_cell_length_c                    8.6532(4)
_cell_angle_alpha                 90.00
_cell_angle_beta                  90.00
_cell_angle_gamma                 90.00
_cell_volume                      816.96(6)
_cell_formula_units_Z             4
_cell_measurement_temperature     100(1)
_cell_measurement_reflns_used     2655
_cell_measurement_theta_min       2.0340
_cell_measurement_theta_max       27.6992
_exptl_crystal_description        'prismatic'
_exptl_crystal_colour             'colourless'
_exptl_crystal_size_max           0.6
_exptl_crystal_size_mid           0.4
_exptl_crystal_size_min           0.4
_exptl_crystal_density_meas       ?
_exptl_crystal_density_diffrn     1.514
_exptl_crystal_density_method     'not measured'
_exptl_crystal_F_000              400
_exptl_absorpt_coefficient_mu     0.370
_exptl_absorpt_correction_type    none
_exptl_absorpt_correction_T_min   ?
_exptl_absorpt_correction_T_max   ?
_exptl_absorpt_process_details    ?

_exptl_special_details
;

;
-- INSERT --                      53,42         5
```


xterm
_diffrn_ambient_temperature 100(1)
_diffrn_radiation_wavelength 0.71073
_diffrn_radiation_type MoK\a
_diffrn_radiation_source 'fine-focus sealed tube'
_diffrn_radiation_monochromator graphite
_diffrn_measurement_device_type 'KUMA KM4CCD four-circle diffractometer'
_diffrn_measurement_method '\w scan'
_diffrn_detector_area_resol_mean 8.1929
_diffrn_standards_number ?
_diffrn_standards_interval_count ?
_diffrn_standards_interval_time ?
_diffrn_standards_decay_% ?
_diffrn_reflns_number 3813
_diffrn_reflns_av_R_equivalents 0.0315
_diffrn_reflns_av_sigmaI/netI 0.0199
_diffrn_reflns_limit_h_min -12
_diffrn_reflns_limit_h_max 13
_diffrn_reflns_limit_k_min -12
_diffrn_reflns_limit_k_max 12
_diffrn_reflns_limit_l_min -11
_diffrn_reflns_limit_l_max 9
_diffrn_reflns_theta_min 3.12
_diffrn_reflns_theta_max 27.79
_reflns_number_total 965
_reflns_number_gt 883
_reflns_threshold_expression >2sigma(I)
_refine_ls_structure_factor_coef Fsqd
_refine_ls_matrix_type full
_refine_ls_weighting_scheme calc
_refine_ls_weighting_details
'calc w=1/[\s^2^(Fo^2^)+(0.0398P)^2^+0.3817P] where P=(Fo^2^+2Fc^2^)/3'
_atom_sites_solution_primary direct
_atom_sites_solution_secondary difmap
_atom_sites_solution_hydrogens geom
_refine_ls_hydrogen_treatment mixed
_refine_ls_extinction_method none
_refine_ls_extinction_coef ?
_refine_ls_number_reflns 965
_refine_ls_number_parameters 88
_refine_ls_number_restraints 0
114,1 12


xterm
_refine_ls_extinction_coef ?
_refine_ls_number_reflns 965
_refine_ls_number_parameters 88
_refine_ls_number_restraints 0
_refine_ls_R_factor_all 0.0301
_refine_ls_R_factor_gt 0.0276
_refine_ls_wR_factor_ref 0.0772
_refine_ls_wR_factor_gt 0.0751
_refine_ls_goodness_of_fit_ref 1.144
_refine_ls_restrained_S_all 1.144
_refine_ls_shift/su_max 0.000
_refine_ls_shift/su_mean 0.000
loop_
_atom_site_label
_atom_site_type_symbol
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_U_iso_or_equiv
_atom_site_adp_type
_atom_site_occupancy
_atom_site_symmetry_multiplicity
_atom_site_calc_flag
_atom_site_refinement_flags
_atom_site_disorder_assembly
_atom_site_disorder_group
N1 N 0.39456(12) -0.01632(12) 0.30136(14) 0.0172(3) Uani 1 1 d . . .
H1N H 0.3429(19) 0.027(2) 0.377(2) 0.032(5) Uiso 1 1 d . . .
H2N H 0.3338(18) -0.052(2) 0.233(2) 0.029(5) Uiso 1 1 d . . .
H3N H 0.4309(19) -0.086(2) 0.343(2) 0.027(5) Uiso 1 1 d . . .
loop_
_atom_site_aniso_label
_atom_site_aniso_U_11
_atom_site_aniso_U_22
_atom_site_aniso_U_33
_atom_site_aniso_U_23
_atom_site_aniso_U_13
_atom_site_aniso_U_12
N1 0.0208(6) 0.0110(5) 0.0197(5) -0.0026(4) -0.0053(5) 0.0025(4)
C1A 0.014(2) 0.0161(19) 0.0148(16) -0.0052(13) 0.0040(18) -0.0006(17)
111,1 19

xterm

```
loop_
 _geom_bond_atom_site_label_1
 _geom_bond_atom_site_label_2
 _geom_bond_distance
 _geom_bond_site_symmetry_2
 _geom_bond_publ_flag
N1 C1A 1.396(6) . ?
N1 C1B 1.588(5) . ?
N1 H1N 0.93(2) . ?

loop_
 _geom_angle_atom_site_label_1
 _geom_angle_atom_site_label_2
 _geom_angle_atom_site_label_3
 _geom_angle
 _geom_angle_site_symmetry_1
 _geom_angle_site_symmetry_3
 _geom_angle_publ_flag
C1A N1 H1N 105.8(13) . . ?
C1B N1 H1N 120.3(13) . . ?
C1A N1 H2N 110.6(12) . . ?
C1B N1 H2N 111.8(12) . . ?

loop_
 _geom_torsion_atom_site_label_1
 _geom_torsion_atom_site_label_2
 _geom_torsion_atom_site_label_3
 _geom_torsion_atom_site_label_4
 _geom_torsion
 _geom_torsion_site_symmetry_1
 _geom_torsion_site_symmetry_2
 _geom_torsion_site_symmetry_3
 _geom_torsion_site_symmetry_4
 _geom_torsion_publ_flag
N1 C1A C2A C2B 176.1(3) . . . 8_565 ?
N1 C1B C2B C2A 69.1(4) . . . 8_565 ?
C1A C2A C2b C1B -75.7(4) . . 8_565 8_565 ?
```

— INSERT — 1,1 To

xterm

```
bash-3.2$ ssh xray11
Password:
Linux xray11.chem.uu.nl 2.6.26-1-686 #1 SMP Wed Nov 26 1

The programs included with the Debian GNU/Linux system a
the exact distribution terms for each program are descri
 _geom_hbond_atom_site_label_A
 _geom_hbond_distance_DH
 _geom_hbond_distance_HA
 _geom_hbond_distance_DA
 _geom_hbond_angle_DHA
 _geom_hbond_site_symmetry_A
N1 H1N O2  0.93(2) 1.89(2) 2.8105(15) 171.5(17) .
N1 H2N O2  0.91(2) 2.00(2) 2.8925(15) 165.5(17) 2_554
N1 H2N O1  0.91(2) 2.385(19) 3.0099(15) 125.7(15) 2_554
N1 H3N O3  0.83(2) 2.00(2) 2.8326(15) 175.9(18) 5_656

_diffrn_measured_fraction_theta_max    0.943
_diffrn_reflns_theta_full              25.00
_diffrn_measured_fraction_theta_full   1.000
_refine_diff_density_max    0.264
_refine_diff_density_min   -0.603
_refine_diff_density_rms    0.060


data_293K

_audit_creation_method            SHELXL-97
_chemical_name_systematic
;
Butane-1,4-diaminium sulfate
;
_chemical_name_common             'putrescinium sulfate'
_chemical_melting_point           ?
_chemical_formula_moiety          '2(C2 H7 N), O4 S'
_chemical_formula_sum
 'C4 H14 N2 O4 S'
_chemical_formula_weight          186.23
```

CIF Completion

- CIF files are mostly created by the refinement program (e.g. SHELXL).
- Missing data can be added with a Text Editor, The Program **enCIFer** (from the CCDC) or **publCIF** (From the IUCr).
- The syntax can be checked with a locally installed version of the program **enCIFer** (Freely Available: www.ccdc.cam.ac.uk).

sk3290.cif – enCIFer

File Edit Search Tools Help

Blocks: 3

```
_atom_type_description
_atom_type_scat_dispersion_real
_atom_type_scat_dispersion_imag
_atom_type_scat_source
'C'  'C'   0.0033   0.0016
'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
'H'  'H'   0.0000   0.0000
'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
'N'  'N'   0.0061   0.0033
'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
'O'  'O'   0.0106   0.0060
'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
'S'  'S'   0.1246   0.1234
'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'

_symmetry_cell_setting          orthorhombic
_symmetry_space_group_name_H-M   'P n m a'
_symmetry_space_group_name_Hall  '-P 2ac 2n'

loop_
 _symmetry_equiv_pos_as_xyz
'x, y, z'
'-x+1/2, -y, z+1/2'
'x+1/2, -y+1/2, -z+1/2'
'-x, y+1/2, -z'
'-x, -y, -z'
'x-1/2, y, -z-1/2'
'-x-1/2, y-1/2, z-1/2'
'x, -y-1/2, z'

_cell_length_a                  9.9722(4)
_cell_length_b                  9.4675(4)
_cell_length_c                  8.6532(4)
_cell_angle_alpha
_cell_angle_beta                90.00
_cell_angle_gamma               90.00
_cell_volume                    816.96(6)
_cell_formula_units_Z           4
_cell_measurement_temperatur    100(1)
_cell_measurement_reflns_used   2655
_cell_measurement_theta_min     2.0340
```

Error detected with PROGRAM enCIFer

Missing Data

Editor | Visualiser

- Errors — Total of 1
  - Line 439 Data name _cell_angle_alpha immediately followed by another data name
- Warnings — Total of 7
- Remarks — - none

Total errors: 0; warnings: 9; remarks: 0

Checking CIF ...

Total errors: 1; warnings: 10; remarks: 0

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.

Examples of Correctable Issues

- Following are some examples of the type of problems addressed.

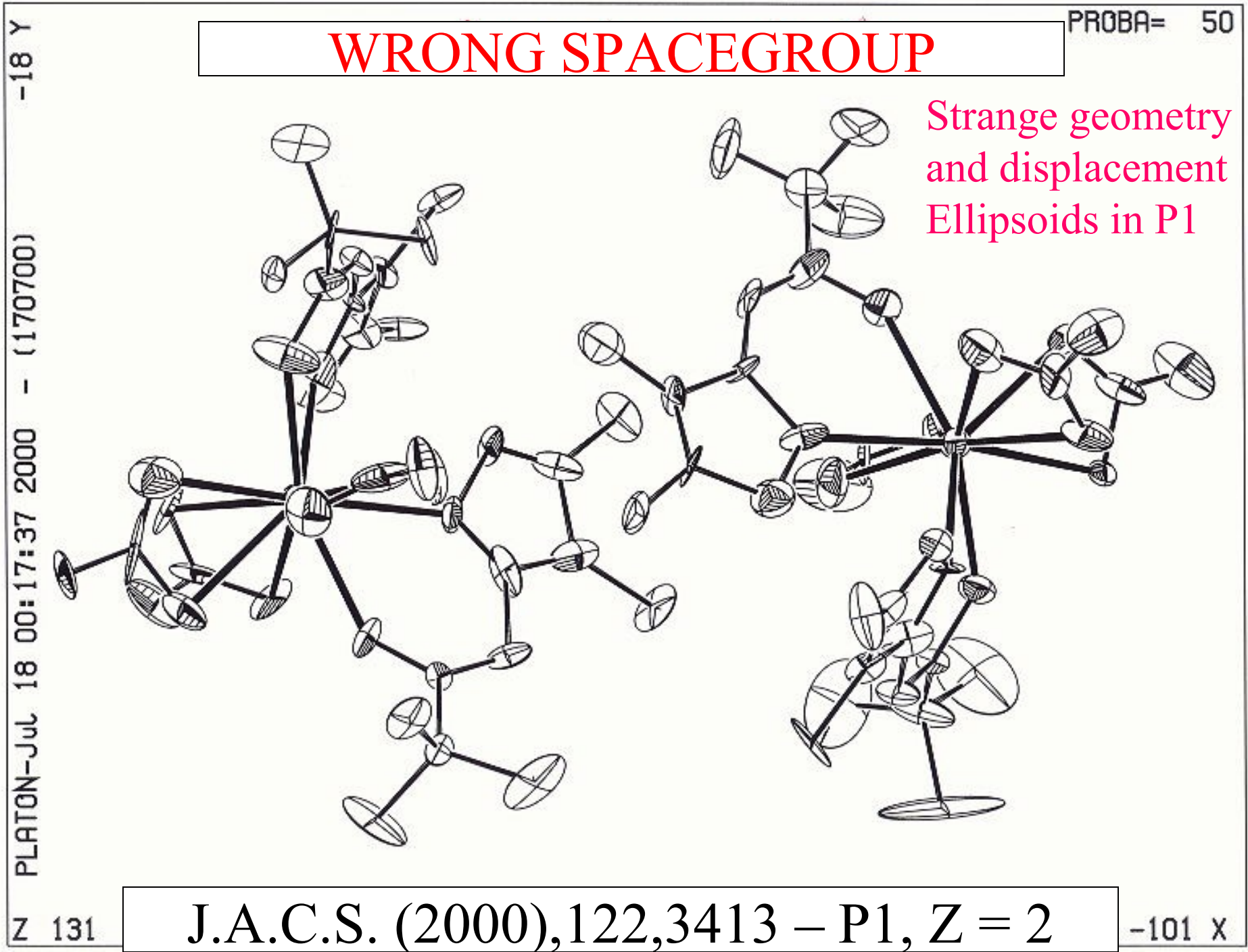
1 – Refinement in the Wrong Space group.

2 – Wrong Atom Type Assignment.

3 – Misplaced H-Atoms.

4 – Missing H-Atoms.

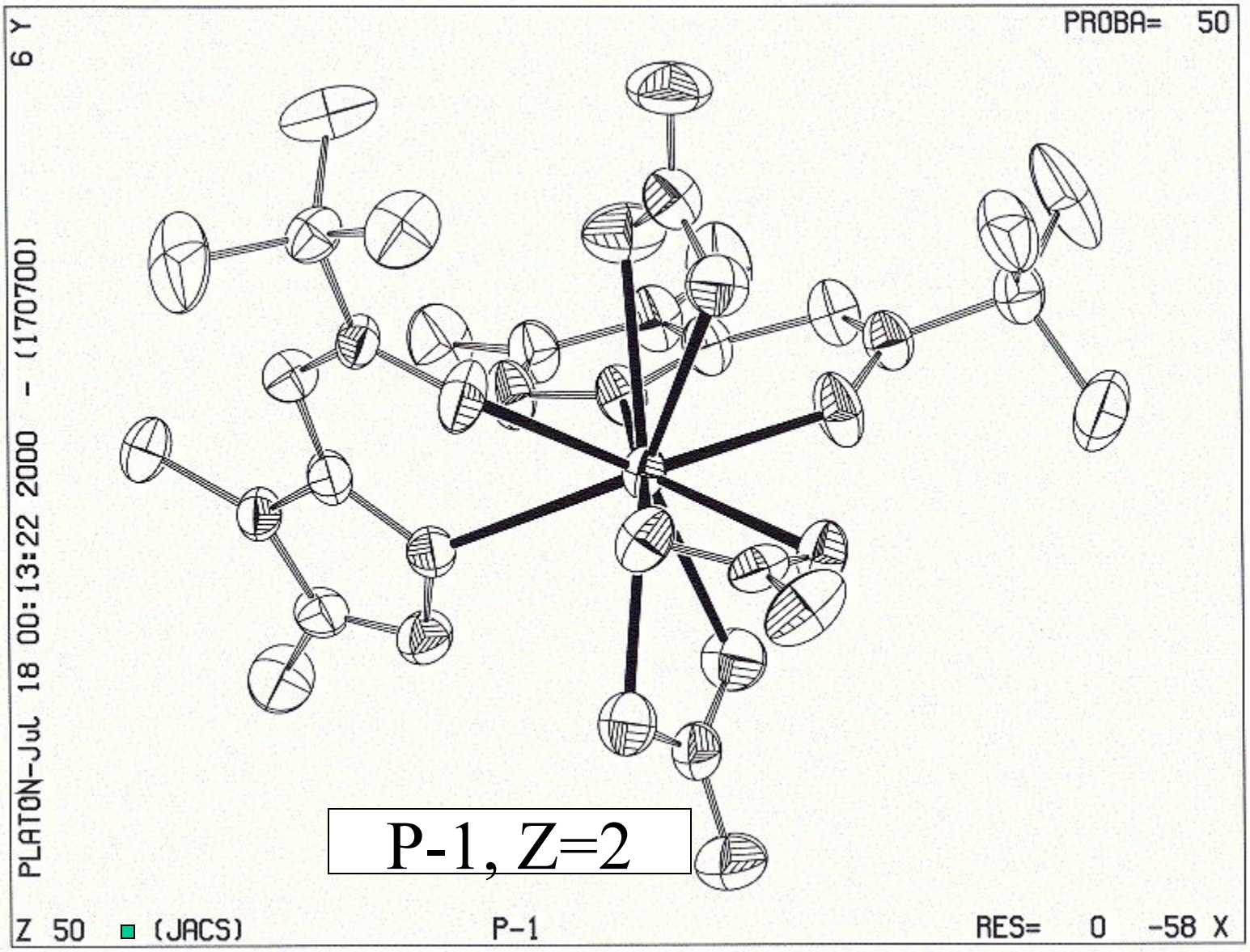
WRONG SPACEGROUP



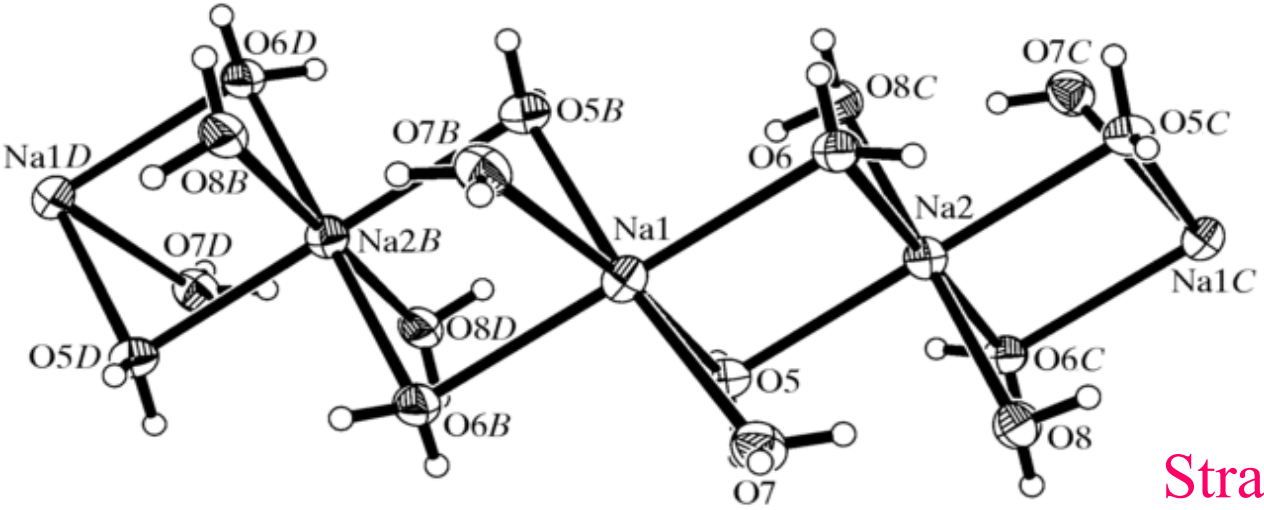
Strange geometry and displacement Ellipsoids in P1

J.A.C.S. (2000),122,3413 – P1, Z = 2

CORRECTLY REFINED STRUCTURE

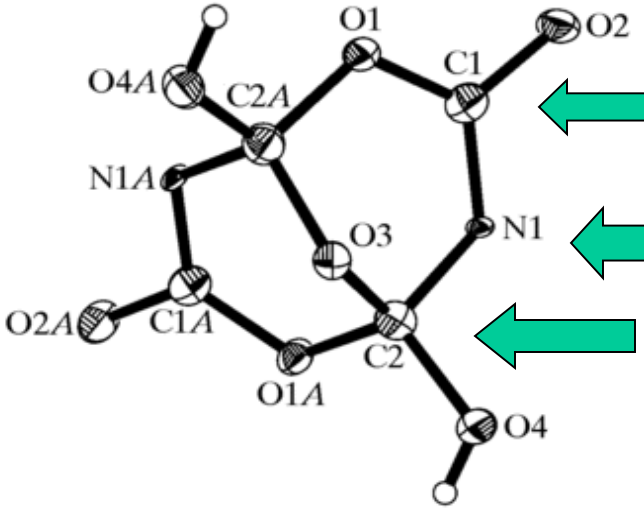


Published with Wrong Composition



Unexpected Result !

Strange Ellipsoids



C → B

N → O

C → B

Corrected Structure BORAX !

=> Retracted

Searching for structures with a Methyl Moiety bridging two metals …

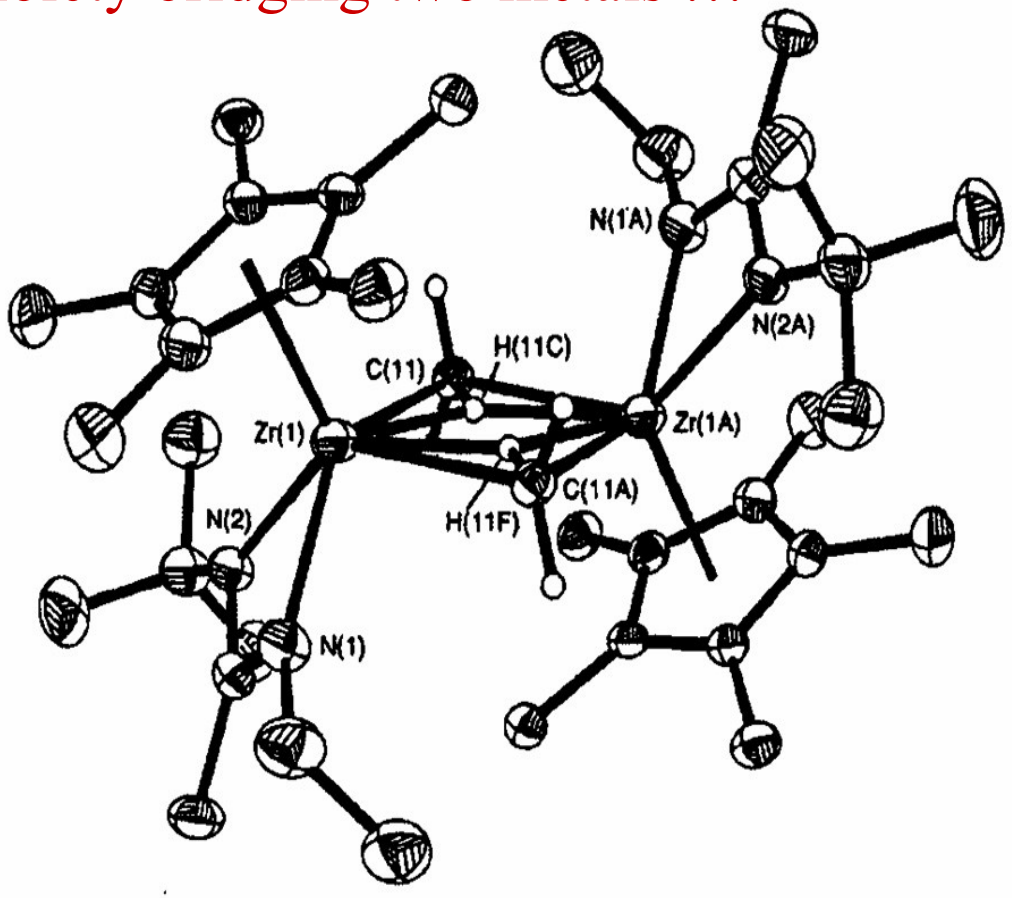


Figure 2. Molecular structure of **2** (30% thermal ellipsoids). The two $[B(C_6F_5)_4]$ anions and all but the bridging methyl hydrogen atoms, represented by spheres of arbitrary size, have been omitted for the sake of clarity. For each bridging methyl group, only the labeled hydrogen atom has been crystallographically located with the other two being placed in logical positions.

Structure of a strange CH_3 Bridged Zr Dimer

Paper has been cited

47 times !

So can we believe this structure?

The Referees did …!

But …

H .. H = 1.32 Ang. !

HOT STRUCTURE – FAST LANE PUBLICATION

THE NEWSMAGAZINE OF THE CHEMICAL WORLD ONLINE | HOW TO LOG IN | HOW TO REACH US | SITE MAP

CHEMICAL & Engineering News

Search Go!

HOME | THIS WEEK'S CONTENTS | C&EN CLASSIFIEDS | JOIN ACS

Back Issues
2002 Go!

Safety Letters
Chemcyclopedia

ACS Members can sign up to receive C&EN e-mail newsletter.
Join

Join ACS

SCIENCE & TECHNOLOGY

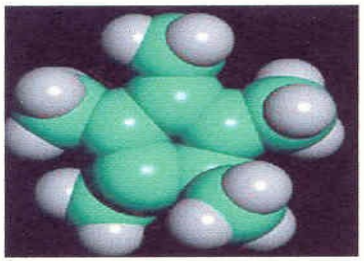
April 29, 2002
Volume 80, Number 17
CENEAR 80 17 p. 30
ISSN 0009-2347

ELUSIVE CARBOCATION ISOLATED AS A SOLID

Pentamethylcyclopentadienyl cation is found to be a stable singlet with a distorted structure

RON DAGANI

Chemists like to study molecules they can store in a bottle on the shelf. If they can't put the molecules in a bottle, they will, of course, study them any way they can--in solution, in the gas phase, in a frozen matrix. But there's a particular satisfaction in bottling a molecule that no one thought could be bottled.



MISSHAPEN A space-filling model of the pentamethylcyclopentadienyl cation shows its distorted, nonplanar ring.

Chemistry professor Joseph B. Lambert of Northwestern University knows that satisfaction. Earlier this month, he and graduate student Lijun Lin reported the first isolation and X-ray structural characterization of a cyclopentadienyl cation--specifically, the pentamethyl-substituted $C_5Me_5^+$ cation [*Angew. Chem. Int. Ed.*, **41**, 1429 (2002)].

Related People
Joseph B. Lambert
Vitaly Rassolov
Ronald Breslow

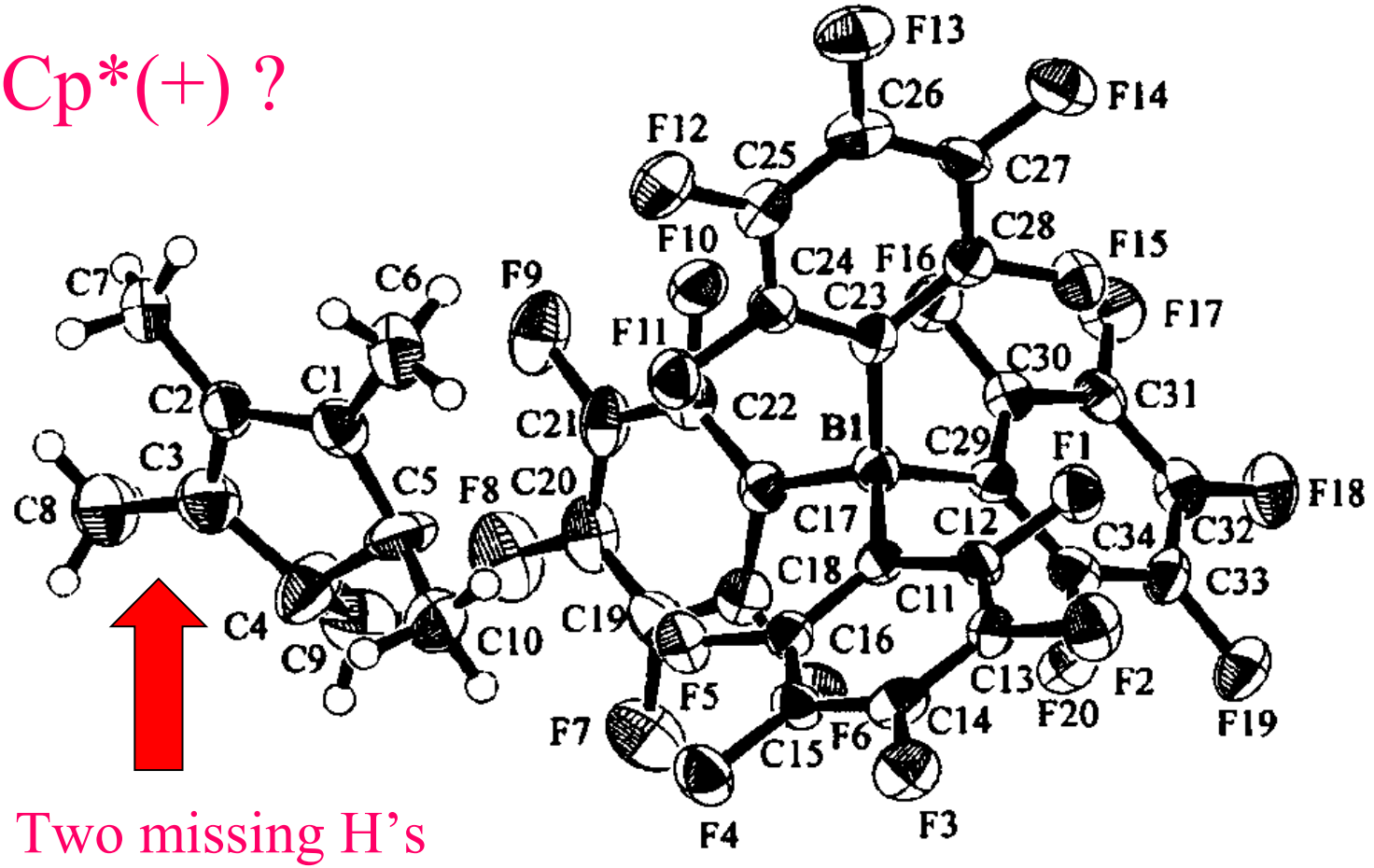
E-mail this article to a friend
Print this article
E-mail the editor

Cp*(+) !! . ?

THE STABLE PENTAMETHYLCYCLOPENTADIENYL CATION

J.B.Lambert et al. Angew. Chem. Int. Ed. 2002, 41, 1429-1431

Cp*(+) ?



No ! Two missing H's

Figure 1. The crystal structure of pentamethylcyclopentadienyl tetrakis-(pentafluorophenyl)borate. There is no covalent bonding between the cation on the left and the anion on the right.

NOT SO HOT AFTER ALL !!

Editors Note in the next issue of Angewandte Chemie

CORRIGENDUM

Note from the Editors: unfortunately the results reported in the communication "The Stable Pentamethylcyclopentadienyl Cation" by Joseph B. Lambert et al. in issue 8/2002 (pp. 1429–1431) must be corrected. Guy Bertrand et al. quickly discovered that not the pentamethylcyclopentadienyl cation but the pentamethylcyclopentenyl cation was prepared and characterized (the corresponding communication will be published in issue 13, and will appear earlier on the *Angewandte Chemie* homepage).

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 (not generally yet unfortunately).
- Most major chemical journals currently have now some form of a validation scheme implemented.
- But, does it solve all problems … ?

Problems to be Addressed

- Synthetic Chemist View: 'Addressing Crystallographic Details holds up the Publication of Important Chemistry' (but see previous example in Angew. Chemie !)
- Interesting Author Question in response to referee issue:

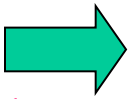
  What does it mean "Space group Incorrect"
- Crystallographic Education (beyond Pushbutton training and Black Box operation) is getting scarce nowadays.
- Sadly: Referees who do not understand or do not know how to respond adequately to ALERTS
- Recently: The need to Detect Fraud and Fraudulous manipulation ….

Note on Editing the CIF

- The Idea of editing the CIF is to add missing (experimental) information to the CIF.
- However: Some authors have now been found to polish away less nice numerical values.
- This leaves traces and is generally detected sooner or later by the validation software and is not good for the scientific career of the culprit…
- The recently implemented FCF-Checking now addresses this issue in even more detail.

Reflection CIF (FCF)

Cell Data Should correspond with CIF data



Terminal — ssh — 73×43

```
#
# h,k,l, Fc-squared, Fo-squared, sigma(Fo-squared) and status flag
#
data_s4223a
_shelx_title ' s4223a - SHELXL'
_shelx_refln_list_code          4
_shelx_F_calc_maximum      687.61
_exptl_crystal_F_000      3008.00
_refIns_d_resolution_high  0.7696

loop_
 _symmetry_equiv_pos_as_xyz
 'x, y, z'
 'x, -y, z+1/2'
 'x+1/2, y+1/2, z'
 'x+1/2, -y+1/2, z+1/2'

_cell_length_a    28.5187
_cell_length_b    11.2036
_cell_length_c    19.4833
_cell_angle_alpha  90.000
_cell_angle_beta  101.351
_cell_angle_gamma  90.000

_shelx_F_squared_multiplier     1.000

loop_
 _refln_index_h
 _refln_index_k
 _refln_index_l
 _refln_F_squared_calc
 _refln_F_squared_meas
 _refln_F_squared_sigma
 _refln_observed_status
   3   1 -25     3134.30     3933.22    696.84 o
   5   1 -25     9060.39     9137.18    788.13 o
   7   1 -25      308.45      307.35    552.52 o
   9   1 -25     1727.00     2193.83    543.48 o
  11   1 -25     1591.10     1053.00    587.08 o
   6   2 -25    11770.63    12530.49    809.23 o
   8   2 -25     4369.66     3470.53    667.51 o
  -4   0 -24    19927.68    20382.10    939.86 o
"s4223a.fcf" 13988L, 698435C                          1,1           Top
```

FCF-VALIDATION

- Check of CIF & FCF data Consistency
- 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.
- Analysis of the details of the Difference Map for unreported features.

Sloppy 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 substituted deliberately to create a ‘new publishable’ structure.

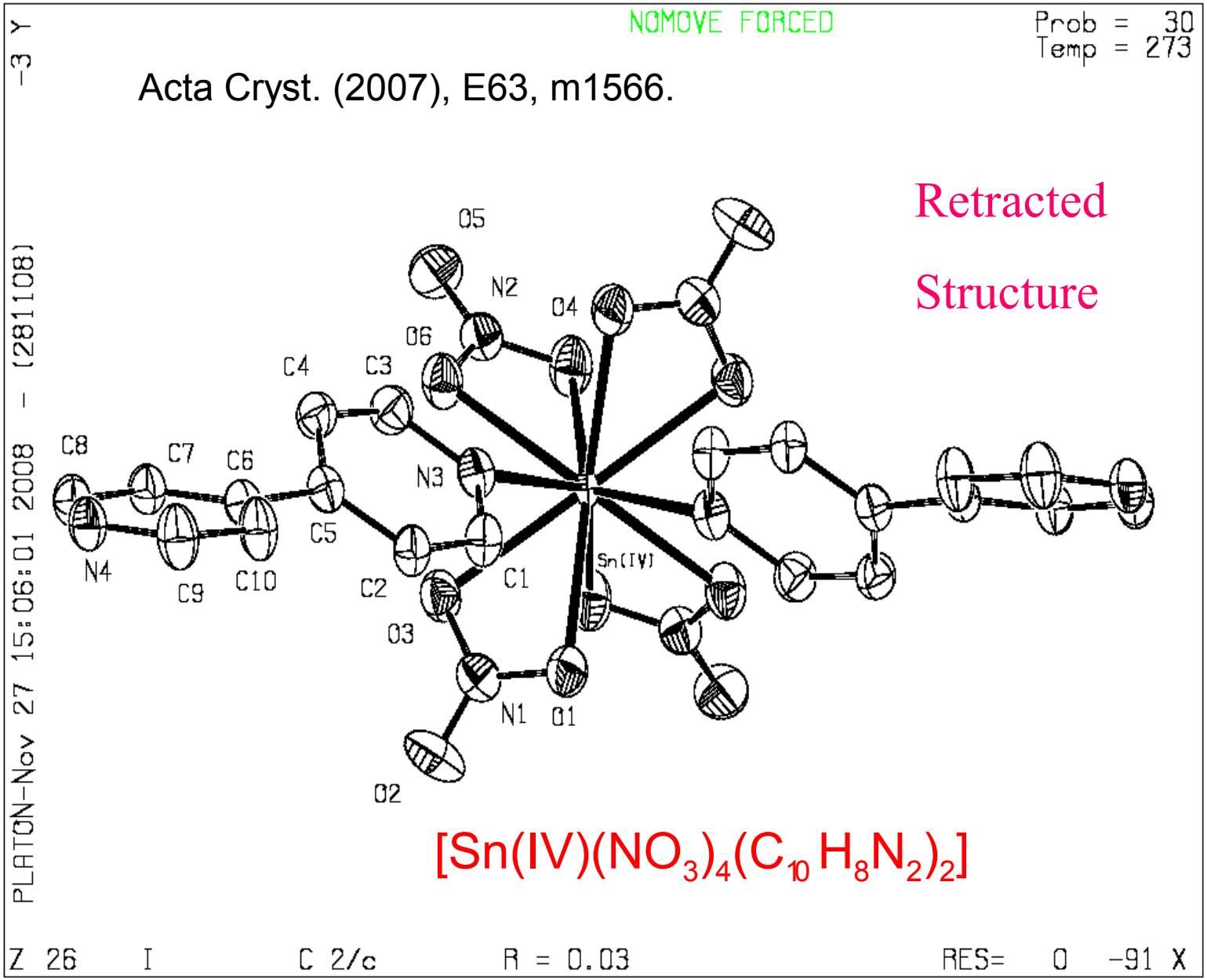
The need of serious validation by knowledgeable Referees

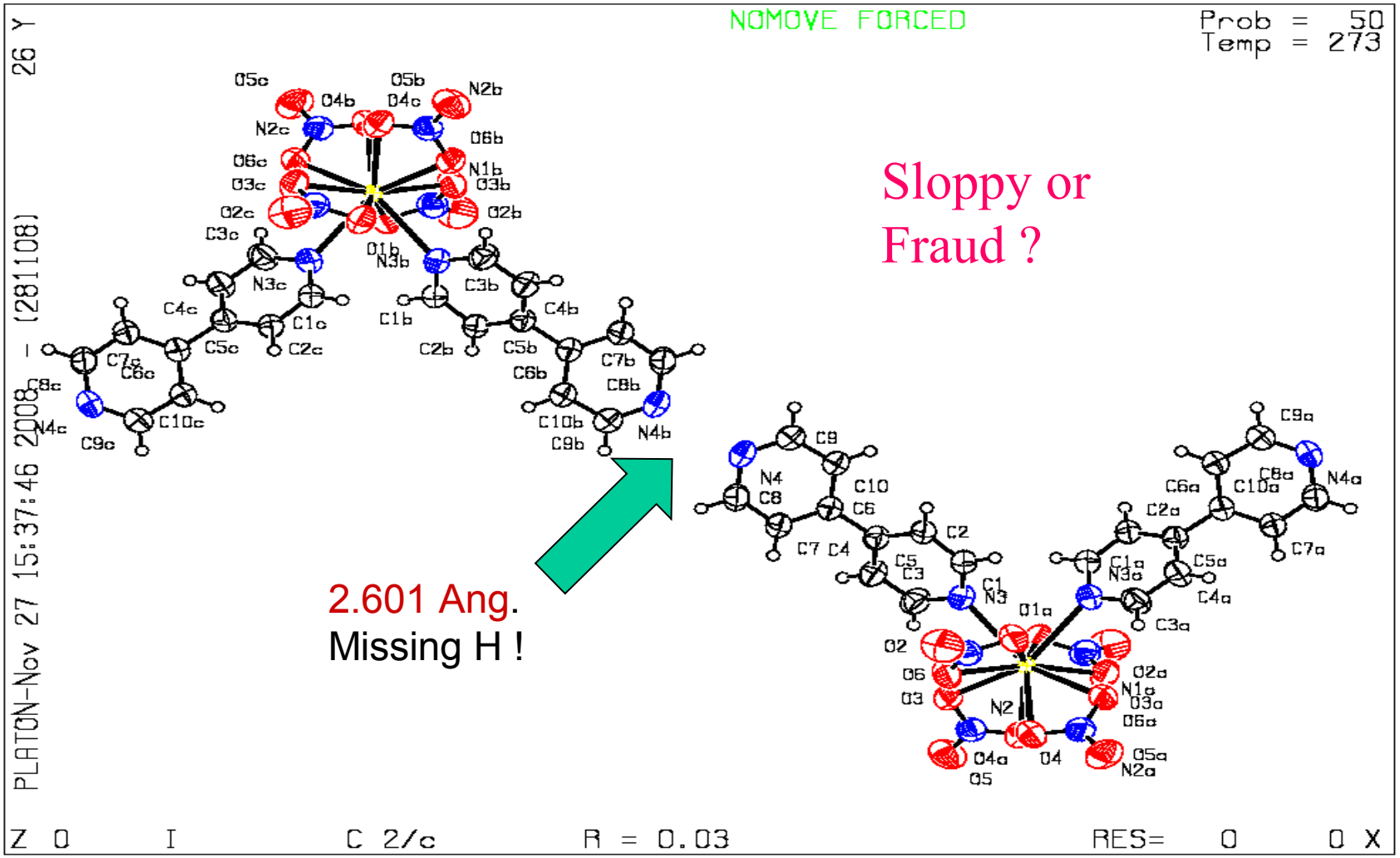
- The validation issues and tools are probably best illustrated by an analysis of a few fraudulous papers that reached the recent literature and (unfortunately) the CSD.
- Early warning signs are generally: troublesome displacement parameters and unusual short inter-molecular contacts.

Some Relevant ALERTS

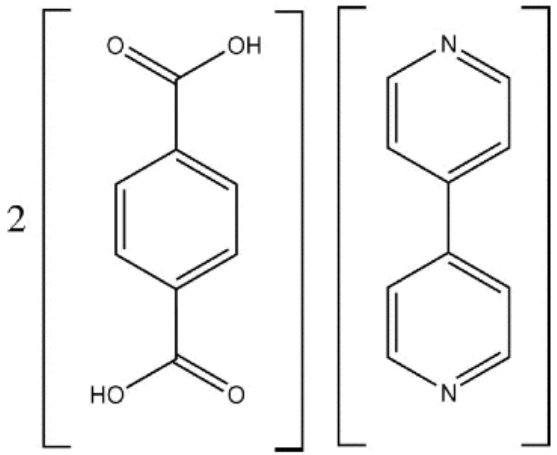
Wrong atom type assignments generally cause:

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

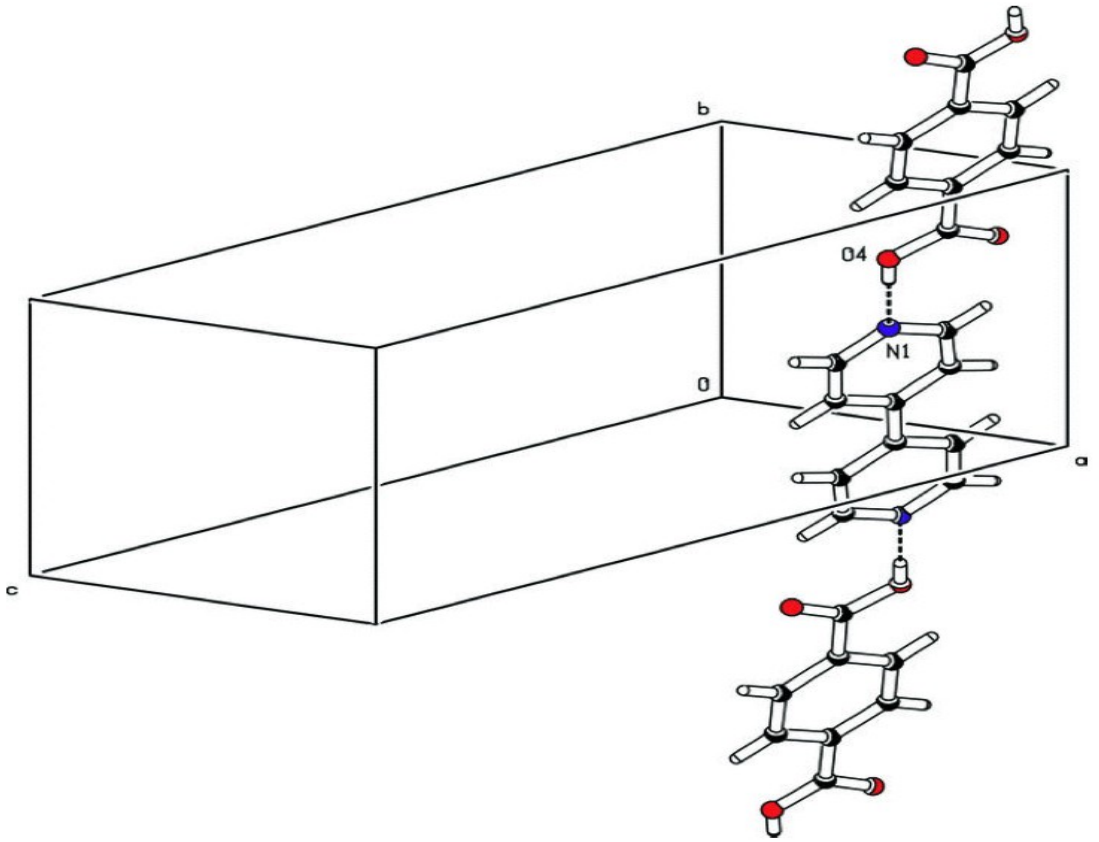

NOMOVE FORCED
Prob = 30
Temp = 273
-3 Y
PLATON-Nov 27 15:06:01 2008 - (281108)
Acta Cryst. (2007), E63, m1566.
Retracted
Structure
O5
N2
O4
O6
C4
C3
C8
C7
C6
N3
C5
N4
C9
C10
C2
C1
Sn(IV)
O3
N1
O1
O2
[Sn(IV)(NO3)4(C10H8N2)2]
Z 26 I C 2/c R = 0.03 RES= 0 -91 X



Missing H in bridge & Sn(IV) => Lanthanide(III)

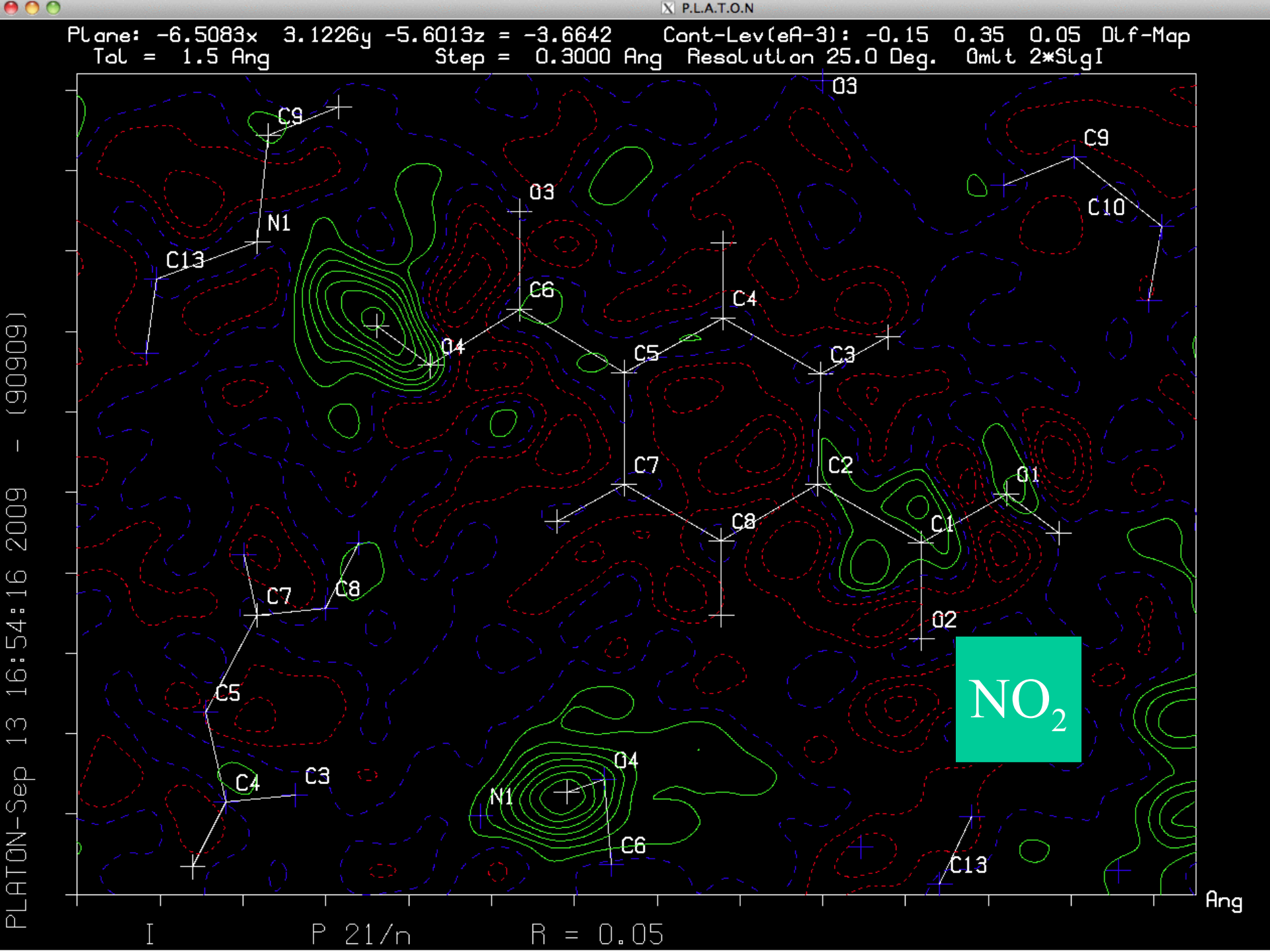


Published structure is claimed to form an infinite hydrogen bonded chain



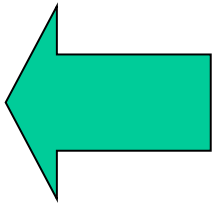
However: This structure does not include a dicarboxylic acid but the previously published para-nitrobenzoic acid.

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. Mult 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

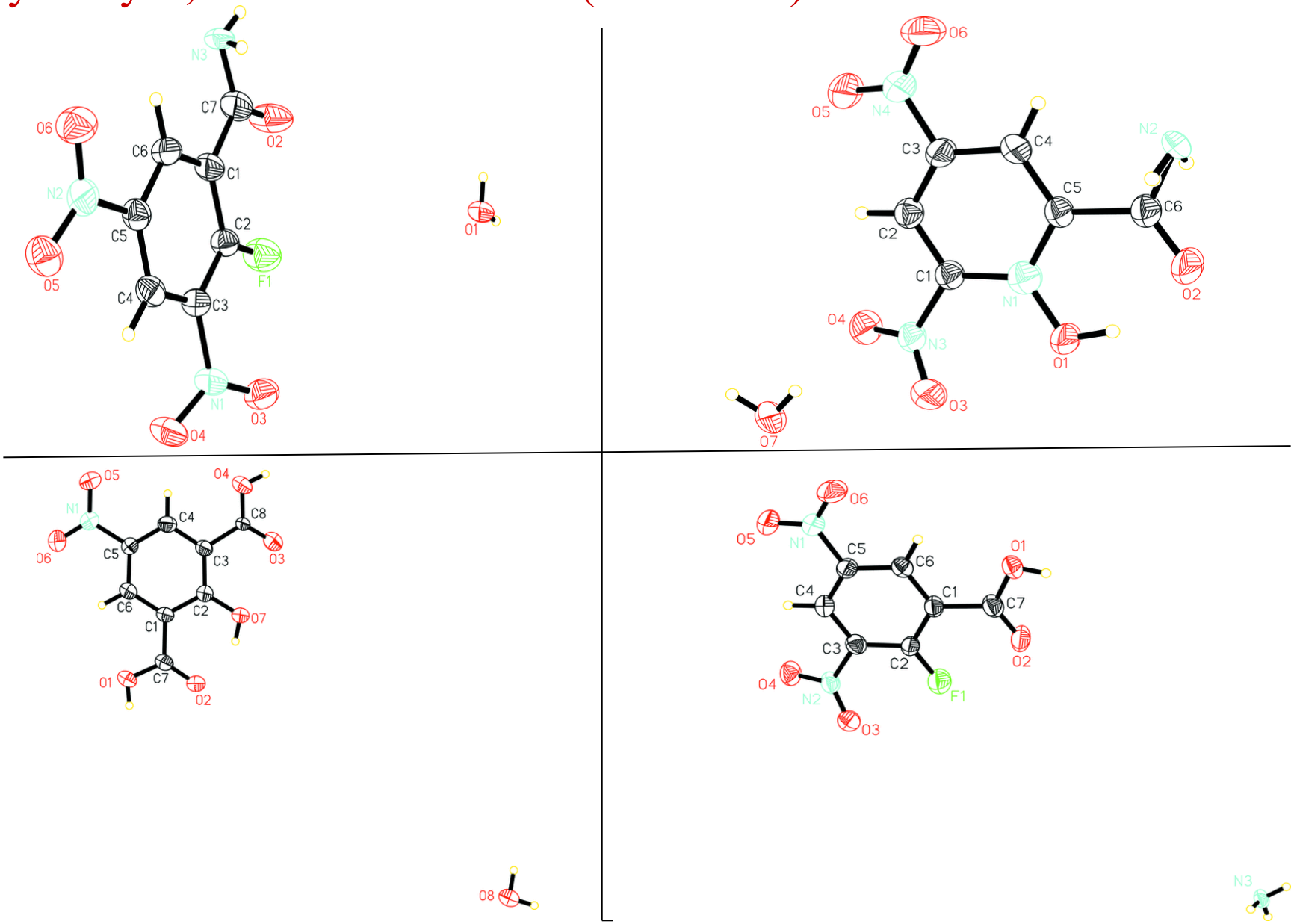
```
=================================================================================
>>> The Following Model and Quality ALERTS were generated     - (Acta-Mode) <<<
=================================================================================
     Format: alert-number_ALERT_alert-type_alert-level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



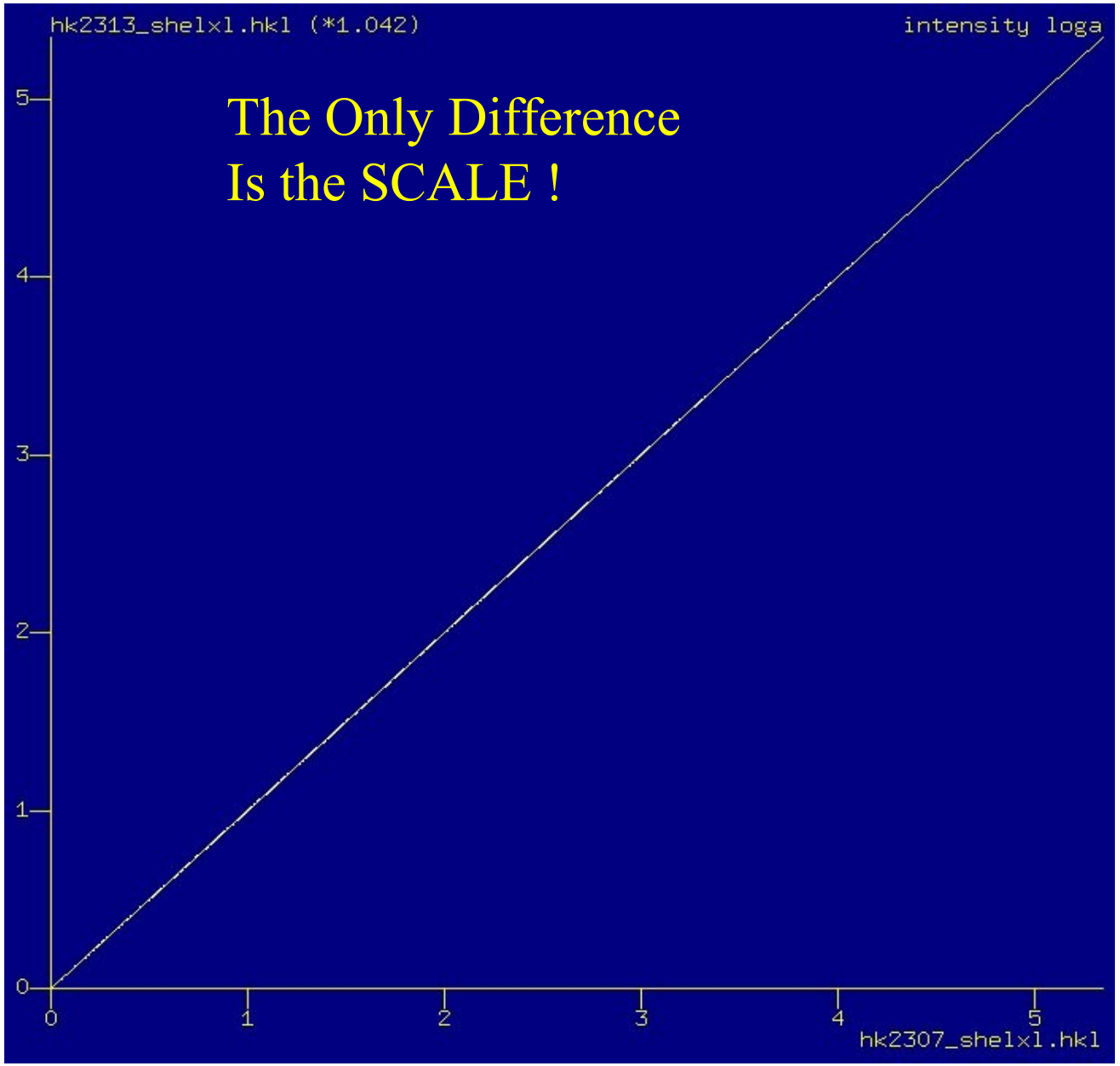
The Ultimate Shame

- Recently a whole series of ‘isomorphous’ substutions was detected for an already published structure.
- Similar series have now been detected for coordination complexes
- How can referees let those pass ?

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



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



Same
Data !

SLOPPY

Or

FRAUD ?

Summary & Conclusions

Validation Procedures:

- May save a lot of Time in Checking, both by the Investigators and by the Journals (referees).
- Often surface problems that only an experienced crystallographer might be able to address.
- May point at Interesting Structural Features (Pseudo-Symmetry, short Interactions etc.) to be investigated/discussed.
- Set Quality Standards (Not just on R-Value).
- May provide Proof of a GOOD or Fraud structure.

Thanks !

For your attention

www.cryst.chem.uu.nl/ppp/freiburg-2009.ppt

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