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 ValidationImplemented?
• 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**

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

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

C1-O1 = 1.396(3)

C1-O1 = 1.213(3)
Huge Geometry Differences !?

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

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

<table>
<thead>
<tr>
<th></th>
<th>3e</th>
<th>3f</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1–O1</td>
<td>1.396(3)</td>
<td>C1–O1</td>
</tr>
<tr>
<td>C1–N1</td>
<td>1.313(3)</td>
<td>C1–N1</td>
</tr>
<tr>
<td>C1–C2</td>
<td>1.612(4)</td>
<td>C1–C2</td>
</tr>
<tr>
<td>C2–N3</td>
<td>1.163(3)</td>
<td>C2–N3</td>
</tr>
<tr>
<td>C2–C3</td>
<td>1.575(4)</td>
<td>C2–C3</td>
</tr>
<tr>
<td>C5–F1</td>
<td>1.567(4)</td>
<td>C5–F1</td>
</tr>
<tr>
<td>C8–N1</td>
<td>1.630(4)</td>
<td>C8–N1</td>
</tr>
<tr>
<td>C9–N2</td>
<td>1.253(3)</td>
<td>C9–N2</td>
</tr>
<tr>
<td>C9–N1</td>
<td>1.478(3)</td>
<td>C9–N1</td>
</tr>
<tr>
<td>C10–N2</td>
<td>1.423(3)</td>
<td>C10–N2</td>
</tr>
<tr>
<td>C10–C11</td>
<td>1.302(4)</td>
<td>C10–C11</td>
</tr>
<tr>
<td>C11–O2</td>
<td>1.481(4)</td>
<td>C11–C12</td>
</tr>
<tr>
<td>C12–O2</td>
<td>1.357(4)</td>
<td>C12–C13</td>
</tr>
<tr>
<td>C12–C13</td>
<td>1.311(4)</td>
<td>C13–C14</td>
</tr>
<tr>
<td>C13–N2</td>
<td>1.532(4)</td>
<td>C14–N2</td>
</tr>
<tr>
<td>C14–N4</td>
<td>1.245(3)</td>
<td>C15–N5</td>
</tr>
<tr>
<td>C14–N5</td>
<td>1.451(4)</td>
<td>C15–N4</td>
</tr>
<tr>
<td>C14–S1</td>
<td>1.864(3)</td>
<td>C15–S1</td>
</tr>
<tr>
<td>C15–C16</td>
<td>1.645(5)</td>
<td>C16–C17</td>
</tr>
<tr>
<td>C15–N5</td>
<td>1.327(3)</td>
<td>C16–N5</td>
</tr>
<tr>
<td>C16–C17</td>
<td>1.238(5)</td>
<td>C17–C18</td>
</tr>
<tr>
<td>N3–N4</td>
<td>1.492(3)</td>
<td>N3–N4</td>
</tr>
</tbody>
</table>
Reported as Monomer

BUT ➞
DIMER: S-S Bridge!
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).
- 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!
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:

Send CIF for checking

Select form of checkCIF report

HTML  PDF
**checkCIF/PLATON report (publication check)**

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

**Datablock: I**

<table>
<thead>
<tr>
<th>Bond precision:</th>
<th>C-C = 0.0157 Å</th>
<th>Wavelength=0.71073</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell:</td>
<td>a=7.6336(15)</td>
<td>b=27.725(6)</td>
</tr>
<tr>
<td></td>
<td>c=12.051(2)</td>
<td>g=90</td>
</tr>
<tr>
<td>Temperature:</td>
<td>153 K</td>
<td></td>
</tr>
<tr>
<td>Calculated</td>
<td>Reported</td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>2520.5(9)</td>
<td>2520.6(9)</td>
</tr>
<tr>
<td>Space group</td>
<td>P 21/n</td>
<td>P 21/n</td>
</tr>
<tr>
<td>Hall group</td>
<td>-P 2yn</td>
<td>-P 2yn</td>
</tr>
<tr>
<td>Moiety formula</td>
<td>C2H2 Br0.97 Cl1.03 N2 Ru</td>
<td></td>
</tr>
<tr>
<td>Sum formula</td>
<td>C2H2 Br0.97 Cl1.03 N2 Ru</td>
<td></td>
</tr>
<tr>
<td>Mr</td>
<td>611.69</td>
<td>611.69</td>
</tr>
<tr>
<td>D, g cm-3</td>
<td>1.612</td>
<td>1.612</td>
</tr>
<tr>
<td>z</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Mu (mm-1)</td>
<td>2.290</td>
<td>2.290</td>
</tr>
<tr>
<td>F000</td>
<td>1237.9</td>
<td>1237.9</td>
</tr>
<tr>
<td>F000'</td>
<td>1232.67</td>
<td></td>
</tr>
<tr>
<td>h,k,l, max</td>
<td>9,9,14</td>
<td>9,9,14</td>
</tr>
<tr>
<td>Nref</td>
<td>4452</td>
<td>4449</td>
</tr>
<tr>
<td>Tmin, Tmax</td>
<td>0.725, 0.955</td>
<td>0.581, 0.955</td>
</tr>
<tr>
<td>Correction method= MULTI-SCAN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data completeness= 0.999

R(reflections)= 0.0891 (3288)  \[\text{wR2(reflections)}= 0.2547 (4449)\]

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**

- **PLATON ALERT 28** SHELXL Second Parameter in WGT unreasonably Large. 37.00

**Alert level C**

- **PFAACO ALERT 3** The value of the weighted R factor is > 0.25
  - Weighted R factor given 0.255
  - Weighted R factor given 0.605
- **PLATON 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: Resid. #
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:
• 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’
This paper presents to the best of our knowledge the first example of a very important MOF construct.
Enter all or part of a CIF dataname to search for its definition in a CIF Dictionary.

Search term
exptl_cystal_color

Which dictionary? Core CIF

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

Dictionary Lookup Example
CIF Example File

data_100K
_audit_creation_method SHExL-97
_chemical_name_systematic ;
_Butane-1,4-diaminium sulfate ;
_chemical_name_common 'putrescinum sulfate'
_chemical_melting_point ?
_chemical_formula_moiety '2(C2 H7 N), 04 S'
_chemical_formula_sum 'C4 H14 N2 04 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_dispersion_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

Exptl :

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 1.514
_exptl_crystal_density_diffn
_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
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>_refine_ls_extinction_coef</td>
<td>?</td>
</tr>
<tr>
<td>_refine_ls_number_reflns</td>
<td>965</td>
</tr>
<tr>
<td>_refine_ls_number_parameters</td>
<td>88</td>
</tr>
<tr>
<td>_refine_ls_number_restraints</td>
<td>0</td>
</tr>
<tr>
<td>_refine_ls_R_factor_all</td>
<td>0.0301</td>
</tr>
<tr>
<td>_refine_ls_R_factor_gt</td>
<td>0.0276</td>
</tr>
<tr>
<td>_refine_ls_wR_factor_ref</td>
<td>0.0772</td>
</tr>
<tr>
<td>_refine_ls_wR_factor_gt</td>
<td>0.0751</td>
</tr>
<tr>
<td>_refine_ls_goodness_of_fit_ref</td>
<td>1.144</td>
</tr>
<tr>
<td>_refine_ls_restrained_S_all</td>
<td>1.144</td>
</tr>
<tr>
<td>_refine_ls_shift/su_max</td>
<td>0.000</td>
</tr>
<tr>
<td>_refine_ls_shift/su_mean</td>
<td>0.000</td>
</tr>
</tbody>
</table>

loop_

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>_atom_site_label</td>
<td></td>
</tr>
<tr>
<td>_atom_site_type_symbol</td>
<td></td>
</tr>
<tr>
<td>_atom_site_frick_x</td>
<td></td>
</tr>
<tr>
<td>_atom_site_frick_y</td>
<td></td>
</tr>
<tr>
<td>_atom_site_frick_z</td>
<td></td>
</tr>
<tr>
<td>_atom_site_Uiso_or_equiv</td>
<td></td>
</tr>
<tr>
<td>_atom_site_adp_type</td>
<td></td>
</tr>
<tr>
<td>_atom_site_occupancy</td>
<td></td>
</tr>
<tr>
<td>_atom_site_symmetry_multiplicity</td>
<td></td>
</tr>
<tr>
<td>_atom_site_calc_flag</td>
<td></td>
</tr>
<tr>
<td>_atom_site_refinement_flags</td>
<td></td>
</tr>
<tr>
<td>_atom_site_disorder_assembly</td>
<td></td>
</tr>
<tr>
<td>_atom_site_disorder_group</td>
<td></td>
</tr>
<tr>
<td>NI N 0.39456(12)</td>
<td></td>
</tr>
<tr>
<td>HLN H 0.3429(19)</td>
<td></td>
</tr>
<tr>
<td>HCK H 0.3338(18)</td>
<td></td>
</tr>
</tbody>
</table>

loop_

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>_atom_site_aniso_label</td>
<td></td>
</tr>
<tr>
<td>_atom_site_aniso_U11</td>
<td></td>
</tr>
<tr>
<td>_atom_site_aniso_U22</td>
<td></td>
</tr>
<tr>
<td>_atom_site_aniso_U33</td>
<td></td>
</tr>
<tr>
<td>_atom_site_aniso_U23</td>
<td></td>
</tr>
<tr>
<td>_atom_site_aniso_U13</td>
<td></td>
</tr>
<tr>
<td>_atom_site_aniso_U12</td>
<td></td>
</tr>
<tr>
<td>NI 0.0208(6)</td>
<td></td>
</tr>
<tr>
<td>CIA 0.014(2)</td>
<td></td>
</tr>
</tbody>
</table>

calc_w=1/[s^2+(Fo^2)+(0.0398P)^2+0.3817P] where P=(Fo^2+2Fc^2)/3
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

_loop_
  _geom_bond_atom_site_label_1
  _geom_bond_atom_site_label_2
  _geom_bond_atom_site_label_3
  _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 C2A C2B 169.1(4) .. 8.565 ?
C1A C2A C2b C1B -75.7(4) .. 8.565 8.565 ?

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_moietiy '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).
Error detected with PROGRAM enCIFer
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 completeness.
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.
Strange geometry and displacement Ellipsoids in P1
CORRECTLY Refined STRUCTURE

P-1, Z=2
Published with Wrong Composition

Unexpected Result!

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(C₆F₅)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 CH₃ Bridged Zr Dimer

Paper has been cited

47 times!

So can we believe this structure?

The Referees did …!

But …

H .. H = 1.32 Ang. !
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 C₅Me₅⁺ cation [Angew. Chem. Int. Ed., 41, 1429 (2002)].
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
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 fraudulent 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
• $wR2 \gg 2 \times R1$
• High values for the SHELXL refined weight parameter
Retracted Structure

\[ \text{[Sn(IV)(NO}_3\text{)}_4(\text{C}_{10}\text{H}_8\text{N}_2)_2]} \]
2.601 Ang.
Missing H !

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

Sloppy or Fraud ?
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.
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

<table>
<thead>
<tr>
<th>Alert Number</th>
<th>Alert-Number_ALERT</th>
<th>Alert-Type</th>
<th>Alert-Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>148_ALERT_3_B</td>
<td>su on the a</td>
<td>Axis</td>
<td>Too Large</td>
<td>(x 1000) 10 Ang.</td>
</tr>
<tr>
<td>148_ALERT_3_B</td>
<td>su on the b</td>
<td>Axis</td>
<td>Too Large</td>
<td>(x 1000) 8 Ang.</td>
</tr>
<tr>
<td>148_ALERT_3_B</td>
<td>su on the c</td>
<td>Axis</td>
<td>Too Large</td>
<td>(x 1000) 30 Ang.</td>
</tr>
<tr>
<td>230_ALERT_2_B</td>
<td>Hirshfeld Test</td>
<td>Diff For</td>
<td>01 - C1</td>
<td>26.27 su</td>
</tr>
<tr>
<td>230_ALERT_2_B</td>
<td>Hirshfeld Test</td>
<td>Diff For</td>
<td>02 - C1</td>
<td>11.10 su</td>
</tr>
<tr>
<td>242_ALERT_2_B</td>
<td>Check Low</td>
<td>Ueq as Compared to Neighbors for C1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>420_ALERT_2_B</td>
<td>D-H Without Acceptor</td>
<td>Diff For</td>
<td>01 - H1</td>
<td>?</td>
</tr>
<tr>
<td>230_ALERT_2_C</td>
<td>Hirshfeld Test</td>
<td>Diff For</td>
<td>C1 - C2</td>
<td>6.87 su</td>
</tr>
<tr>
<td>242_ALERT_2_C</td>
<td>Check Low</td>
<td>Ueq as Compared to Neighbors for C11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>911_ALERT_3_C</td>
<td>Missing # FCF Refl Between THmin &amp; STh/L= 0.595</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>913_ALERT_3_C</td>
<td>Missing # of Very Strong Reflections in FCF</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>860_ALERT_3_G</td>
<td>Note: Number of Least-Squares Restraints</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

The Following Improvement and Query ALERTS were generated - (Acta-Mode) <<<

<table>
<thead>
<tr>
<th>Alert Number</th>
<th>Alert-Number_ALERT</th>
<th>Alert-Type</th>
<th>Alert-Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>926_ALERT_1_B</td>
<td>Reported and Calculated</td>
<td>R1 * 100.0</td>
<td>Differ by</td>
<td>0.81</td>
</tr>
<tr>
<td>927_ALERT_1_B</td>
<td>Reported and Calculated</td>
<td>wR2 * 100.0</td>
<td>Differ by</td>
<td>2.26</td>
</tr>
<tr>
<td>042_ALERT_1_C</td>
<td>Calc, and Reported Moiety</td>
<td>Formula Strings</td>
<td>Differ</td>
<td>?</td>
</tr>
<tr>
<td>790_ALERT_4_C</td>
<td>Centre of Gravity not Within Unit Cell: Resid.</td>
<td>#</td>
<td>2</td>
<td>C8 H6 04</td>
</tr>
<tr>
<td>928_ALERT_1_C</td>
<td>Reported and Calculated</td>
<td>S value</td>
<td>Differ by</td>
<td>0.13</td>
</tr>
<tr>
<td>128_ALERT_4_G</td>
<td>Non-standard setting of Space-group P21/c</td>
<td>P21/n</td>
<td>........</td>
<td></td>
</tr>
</tbody>
</table>

---


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?
Bogus Variations (with Hirshfeld ALERTS) on the Published Structure 2-hydroxy-3,5-nitrobenzoic acid (ZAJGUM)
Comparison of the Observed data for two ‘isomorphous’ compounds.

The Only Difference Is the SCALE!

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: