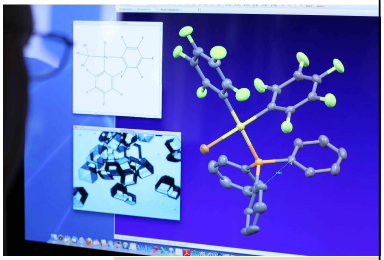
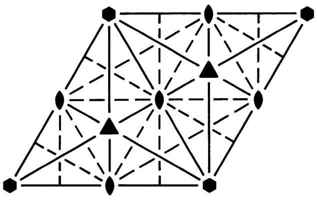
University of Zurich^{UZH}

Department of Chemistry

CIF and
Structure Validation



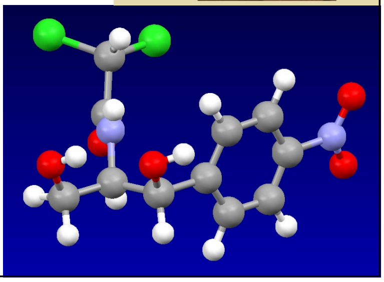
Anthony Linden



7th European Crystallography School

Lisbon

July 10-15, 2022



Crystallographic Information File (CIF)

- A standard for electronic archiving and exchange of crystallographic data
- Original paper describing the CIF syntax and defining many of the data items:
  S.R. Hall, F.H. Allen, I.D. Brown, Acta Crystallogr. **A47**, 655 (1991).
- Current standard reference and definitions:
  International Tables Volume G.

CIF syntax

- Text only file. If using WORD, save as “text only”.
- Line length is 80 characters maximum, no line wrapping allowed. WORD: courier new font, set margins to have line length < 80 characters (16.5 cm)
- Plain text characters only (standard keyboard).

```
Abcdefghijklmnopqrstuvwxyz
ABCDEFGHIJKLMNOPQRSTUVWXYZ
0123456789
!@#$%^&*()_+{}:"~<>?|\-=[];'`,/.
```

CIF syntax

- Special characters (Greek, accents, symbols) cannot be inserted directly, but can be created by special character combinations:
  \a = α
  \%A = Å
  \% = °   (degree symbol)
  \’a = á
- Bold and italics originally not supported, but now may be indicated with HTML markup,

  this text is bold
  <i>italic</i>

CIF syntax

- Every CIF starts with:
  data_xxxxxxx, (usually data_global)
- Every structure data block in the same CIF, if more than one, must begin with a data_xxxx line
- Every data item has its own name and begins with an underscore
  e.g. _exptl_crystal_density_diffrn 2.088

CIF syntax

- In tables, a specific sequence of items repeats. These are preceded by loop_

```
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
Ag1 N19 2.166(4) . ?
Ag1 C30 2.170(4) . ?
```

CIF syntax

- All defined data names MUST have a corresponding entry in the list. If you omit something, the syntax will get out of sync.
  If there is no value, use ?.
  If there is a default value, use a dot.

```
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
Ag1 N19 2.166(4) . ?
Ag1 C30 2.170(4) . ?
```

CIF syntax

- Some items have numerical values, others have text associated with them. Do not mix.

- Correct:

_chemical_melting_point	403
_cell_length_a	22.6241(4)

- Incorrect:

_chemical_melting_point	273-275
_chemical_melting_point	‘273 K’
_cell_volume	2367.5(8)\%A^3^

CIF syntax

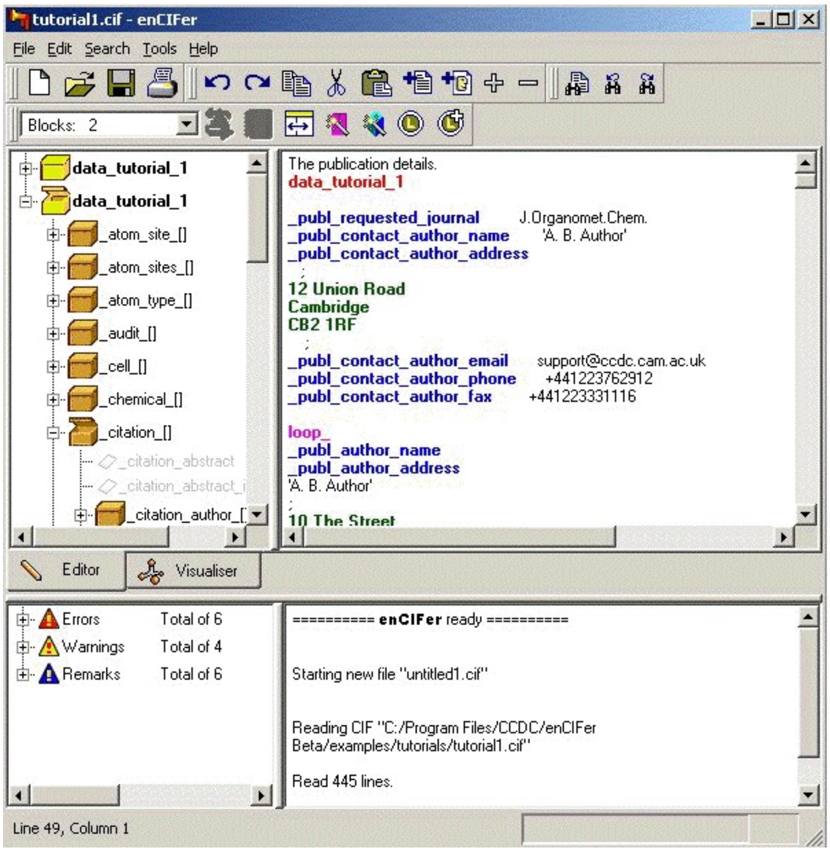
- Text items can be included 3 ways:
- Single word:
  `_exptl_crystal_description        rod`
- Several words all on one line, use quotation marks
  `_computing_data_collection   'COLLECT (Nonius, 2000)'`
- Long text over several lines delimited by **;** before and after the text, always in column one of a line

```
_atom_type_scat_source
; International Tables for Crystallography (1992)
Vol. C, Tables 4.2.6.8 and 6.1.1.4
;
```

Completing a CIF

- SHELXL generated CIFs have many data items missing, mostly related to the data collection experiment.
- Some diffractometers provide the data collection information in CIF format – merge files. Olex^2 can do this for some diffractometers.
- Fill in missing entries with Olex^2, a text editor, or use publCIF, enCIFer
- Be careful when editing a CIF by hand!

enCIFer - www.ccdc.cam.ac.uk



```
The publication details.
data_tutorial_1

_publ_requested_journal       J.Organomet.Chem.
_publ_contact_author_name       'A. B. Author'
_publ_contact_author_address
;
12 Union Road
Cambridge
CB2 1RF
;
_publ_contact_author_email      support@ccdc.cam.ac.uk
_publ_contact_author_phone      +441223762912
_publ_contact_author_fax      +441223331116

loop_
_publ_author_name
_publ_author_address
'A. B. Author'
;
10 The Street
```

```
========== enCIFer ready ==========


Starting new file "untitled1.cif"


Reading CIF "C:/Program Files/CCDC/enCIFer
Beta/examples/tutorials/tutorial1.cif"

Read 445 lines.
```

A useful CIF editor

Checks syntax

Highlights unbalanced markup flags (; ~ ^)

Wizard for inserting publication details or templates

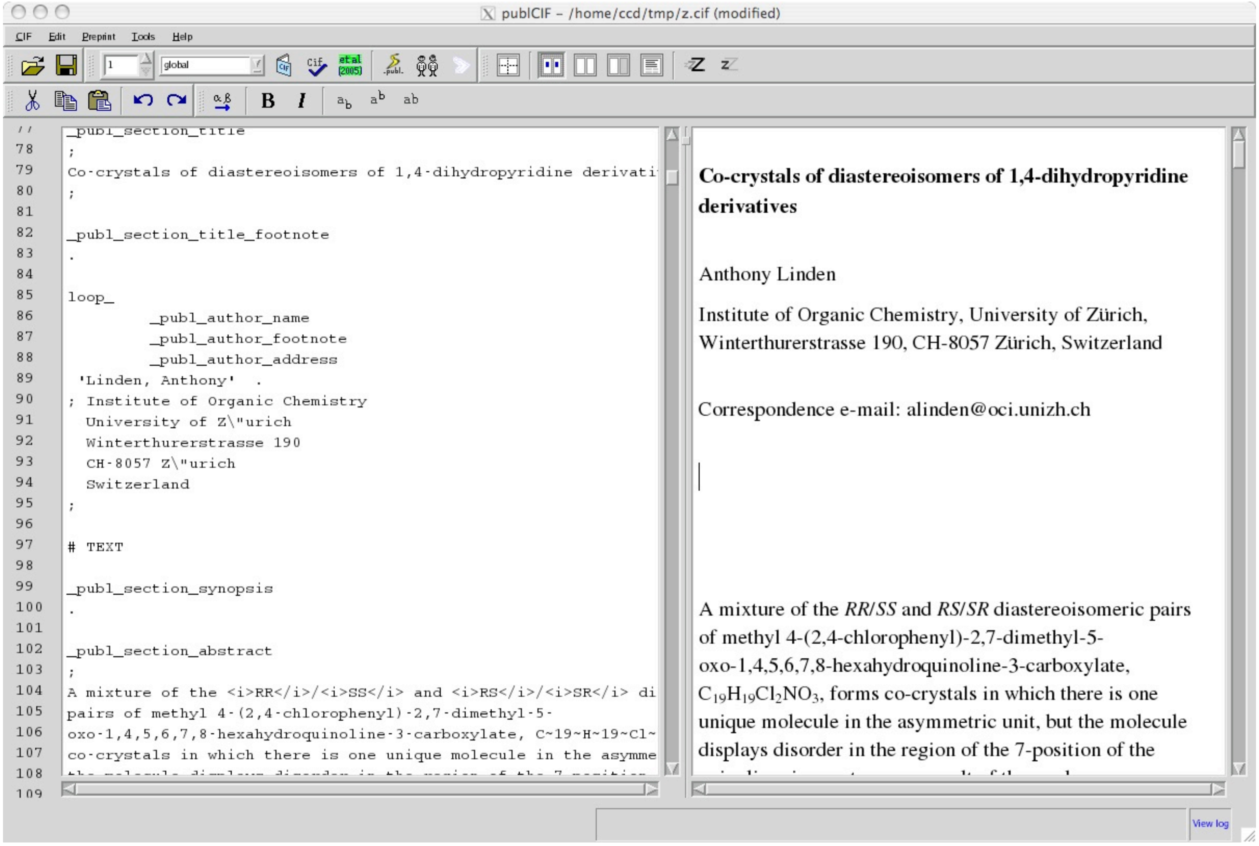
Does not validate data items other than for correct syntax

---

publCIF - journals.iucr.org/services/cif/publcif

- View the raw CIF and the final printed manuscript style simultaneously
- Editing one is immediately reflected in the other
- Wizards for adding missing data items, formatting markup, authors, special tables, etc.
- Database of common references
- Cross-check references with the text citations
- Create your own database of author details
- Run checkCIF from within program

publCIF - journals.iucr.org/services/cif/publcif



```
_publ_section_title
;
Co-crystals of diastereoisomers of 1,4-dihydropyridine derivati
;

_publ_section_title_footnote
.

loop_
          _publ_author_name
          _publ_author_footnote
          _publ_author_address
 'Linden, Anthony'  .
; Institute of Organic Chemistry
  University of Z\"urich
  Winterthurerstrasse 190
  CH-8057 Z\"urich
  Switzerland
;

# TEXT

_publ_section_synopsis
.

_publ_section_abstract
;
A mixture of the <i>RR</i>/<i>SS</i> and <i>RS</i>/<i>SR</i> di
pairs of methyl 4-(2,4-chlorophenyl)-2,7-dimethyl-5-
oxo-1,4,5,6,7,8-hexahydroquinoline-3-carboxylate, C~19~H~19~Cl~
co-crystals in which there is one unique molecule in the asymme
```

Co-crystals of diastereoisomers of 1,4-dihydropyridine derivatives

Anthony Linden

Institute of Organic Chemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Correspondence e-mail: alinden@oci.unizh.ch

A mixture of the *RR*/*SS* and *RS*/*SR* diastereoisomeric pairs of methyl 4-(2,4-chlorophenyl)-2,7-dimethyl-5-oxo-1,4,5,6,7,8-hexahydroquinoline-3-carboxylate, $C_{19}H_{19}Cl_2NO_3$, forms co-crystals in which there is one unique molecule in the asymmetric unit, but the molecule displays disorder in the region of the 7-position of the

---

What is validation?

Comparison against normally expected values or conditions

- Are all the usual information and data present?
- Do related or derived parameters match?
- Do bonded atoms have compatible U^{ij} values?
- Has the refinement converged?
- Is the space group correct?
- Are the assigned atom types correct?
- etc, etc, etc…

Valid-ation

Correct

Appropriate

Defensible

Why do we need *checkCIF?*

checkCIF introduced by the IUCr in 1997

Ongoing development by Ton Spek in *PLATON*

- Throughput of labs exploded in the CCD era
- Nice GUIs, but people often no longer look at output/log files
- Help people avoid simple errors and oversights
- Encourage maintenance of quality standards (best practice)
- Increase publication success rate for authors (less revisions)
- Decrease publication times for journals

Are validation and vigilance still needed?

- Many avoidable mistakes still appear in submitted or published papers
  - Inexperience
  - Complacency
  - Ignoring (lesser) validation alerts
  - Do not understand alerts
  - Blind reliance on *checkCIF* – if there is no alert, it must be OK
  - Conversely, blind reliance by reviewers – if there is an alert, there must be a problem!

---

checkCIF is…

- A tool to help YOU…
  - efficiently check your work
  - avoid blunders
  - follow best practice ideals
  - achieve the best result possible
- Not intended as a hurdle to make life tough
- Not intended to hinder publication of correct results
- Not intended to make you write long explanations for everything – scientists always document (non-routine) experimental procedures, don't they…?
- Also a useful tool for (knowledgeable) reviewers

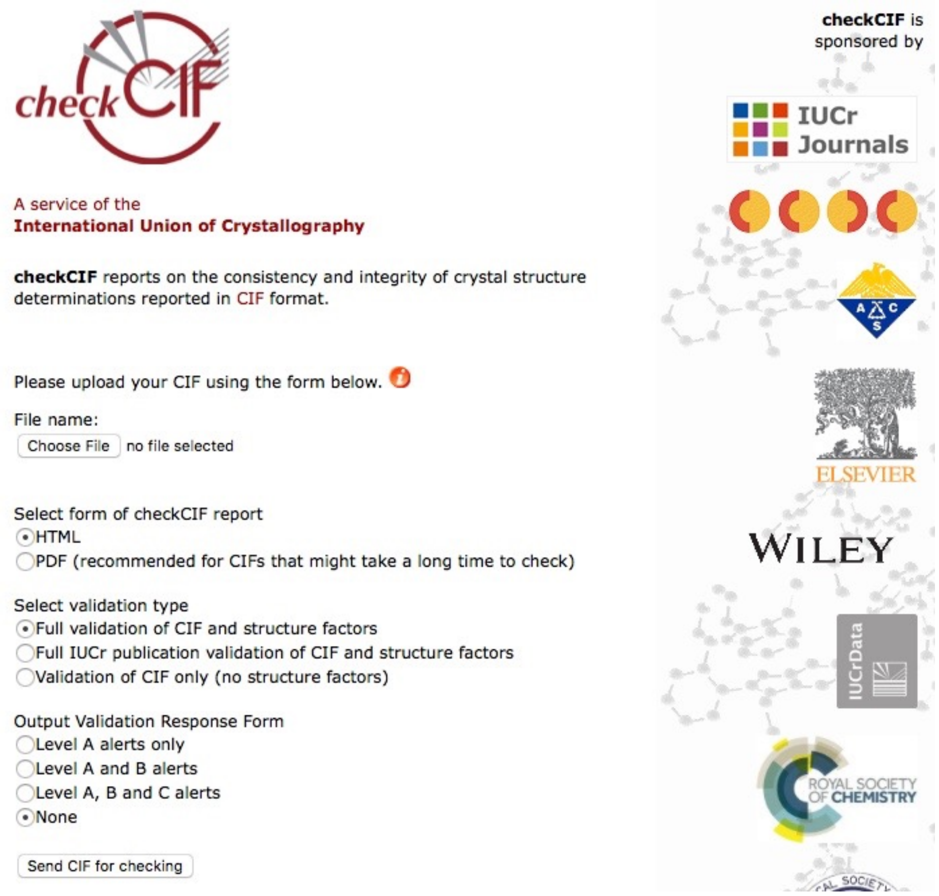
Current *checkCIF* and *PLATON* tests

- CIF syntax, missing information, data consistency and quality
- Unit cell & space-group symmetry
- (An)isotropic displacement parameters
- Intramolecular & intermolecular contacts
- Coordination-related issues
- Solvent-accessible voids
- Consistency of geometric parameters & s.u.s
- Reflection data consistency, completeness, twinning
- and much more…

Sources of outlier parameters

- Incorrect structure (e.g. wrong space group or atom)
- Unresolved feature (e.g. untreated disorder)
- Non-optimal procedures (e.g. poor disorder modelling)
- Artefact resulting from limited data quality
- Special experimental conditions (document them)
- A genuinely unusual observation – worthy of discussion!

checkcif.iucr.org



A service of the
International Union of Crystallography

checkCIF reports on the consistency and integrity of crystal structure determinations reported in CIF format.

checkCIF is sponsored by

Please upload your CIF using the form below.

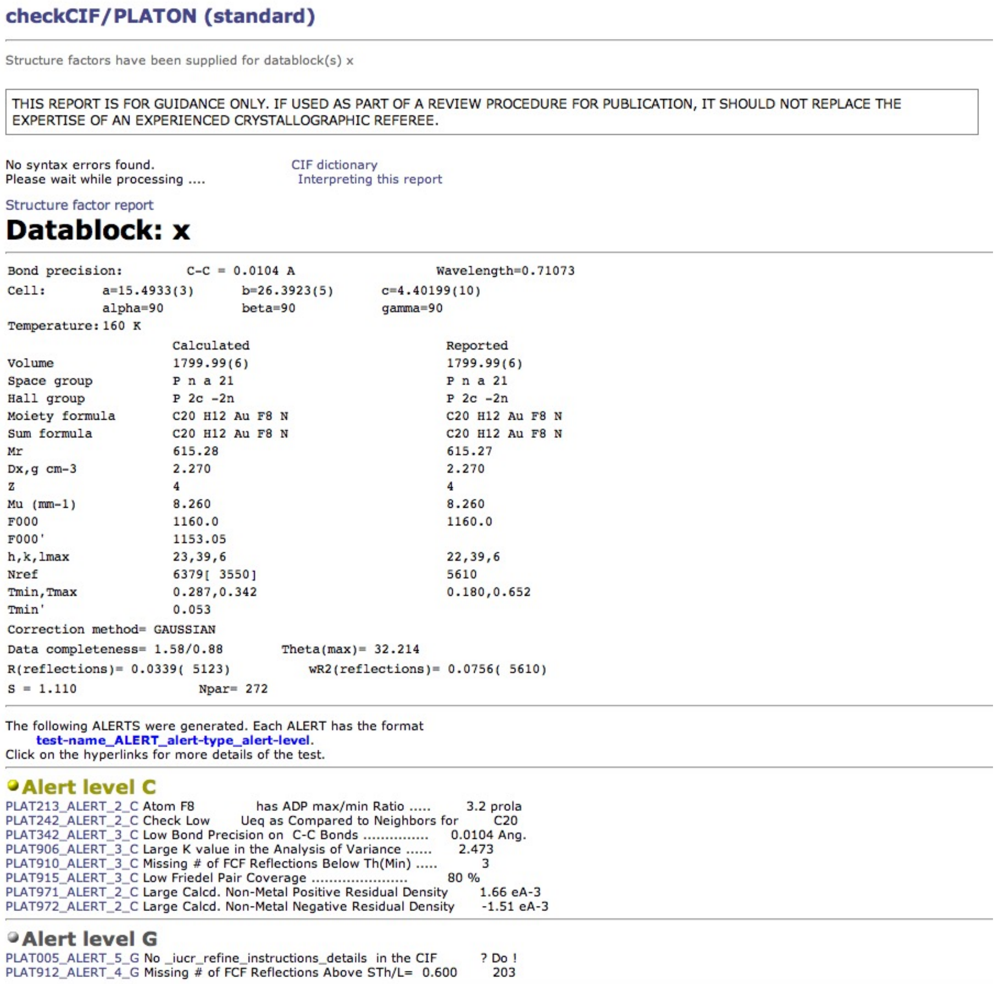
File name:
Choose File no file selected

Select form of checkCIF report
- ● HTML
- ○ PDF (recommended for CIFs that might take a long time to check)

Select validation type
- ● Full validation of CIF and structure factors
- ○ Full IUCr publication validation of CIF and structure factors
- ○ Validation of CIF only (no structure factors)

Output Validation Response Form
- ○ Level A alerts only
- ○ Level A and B alerts
- ○ Level A, B and C alerts
- ● None

Send CIF for checking



checkCIF/PLATON (standard)

Structure factors have been supplied for datablock(s) x

THIS REPORT IS FOR GUIDANCE ONLY. IF USED AS PART OF A REVIEW PROCEDURE FOR PUBLICATION, IT SHOULD NOT REPLACE THE EXPERTISE OF AN EXPERIENCED CRYSTALLOGRAPHIC REFEREE.

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

CIF dictionary
Interpreting this report

Structure factor report

Datablock: x

```
Bond precision:          C-C = 0.0104 A                 Wavelength=0.71073
Cell:          a=15.4933(3)      b=26.3923(5)      c=4.40199(10)
               alpha=90          beta=90           gamma=90
Temperature: 160 K
                        Calculated                     Reported
Volume                  1799.99(6)                     1799.99(6)
Space group             P n a 21                       P n a 21
Hall group              P 2c -2n                       P 2c -2n
Moiety formula          C20 H12 Au F8 N                C20 H12 Au F8 N
Sum formula             C20 H12 Au F8 N                C20 H12 Au F8 N
Mr                      615.28                         615.27
Dx,g cm-3               2.270                          2.270
Z                       4                              4
Mu (mm-1)               8.260                          8.260
F000                    1160.0                         1160.0
F000'                   1153.05
h,k,lmax                23,39,6                        22,39,6
Nref                    6379[ 3550]                    5610
Tmin,Tmax               0.287,0.342                    0.180,0.652
Tmin'                   0.053
Correction method= GAUSSIAN
Data completeness= 1.58/0.88        Theta(max)= 32.214
R(reflections)= 0.0339( 5123)           wR2(reflections)= 0.0756( 5610)
S = 1.110                 Npar= 272
```

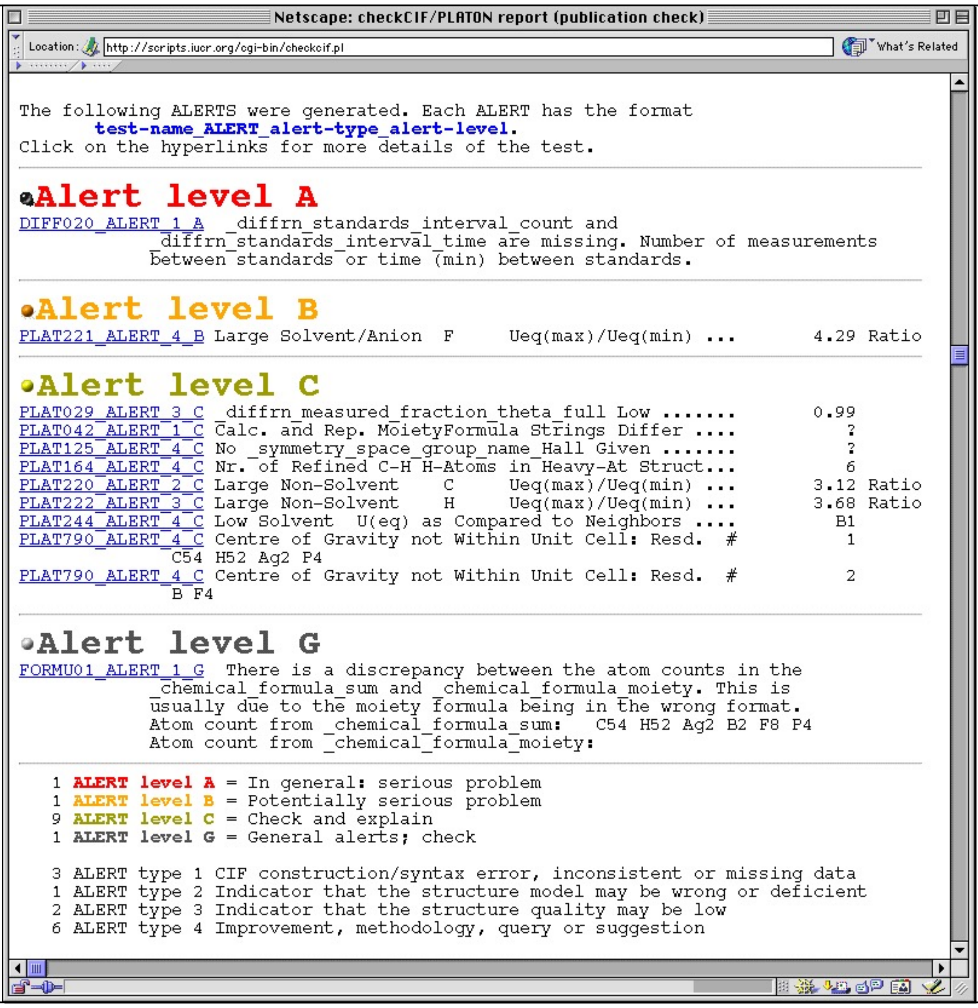
The following ALERTS were generated. Each ALERT has the format
test-name_ALERT_alert-type_alert-level.
Click on the hyperlinks for more details of the test.

Alert level C

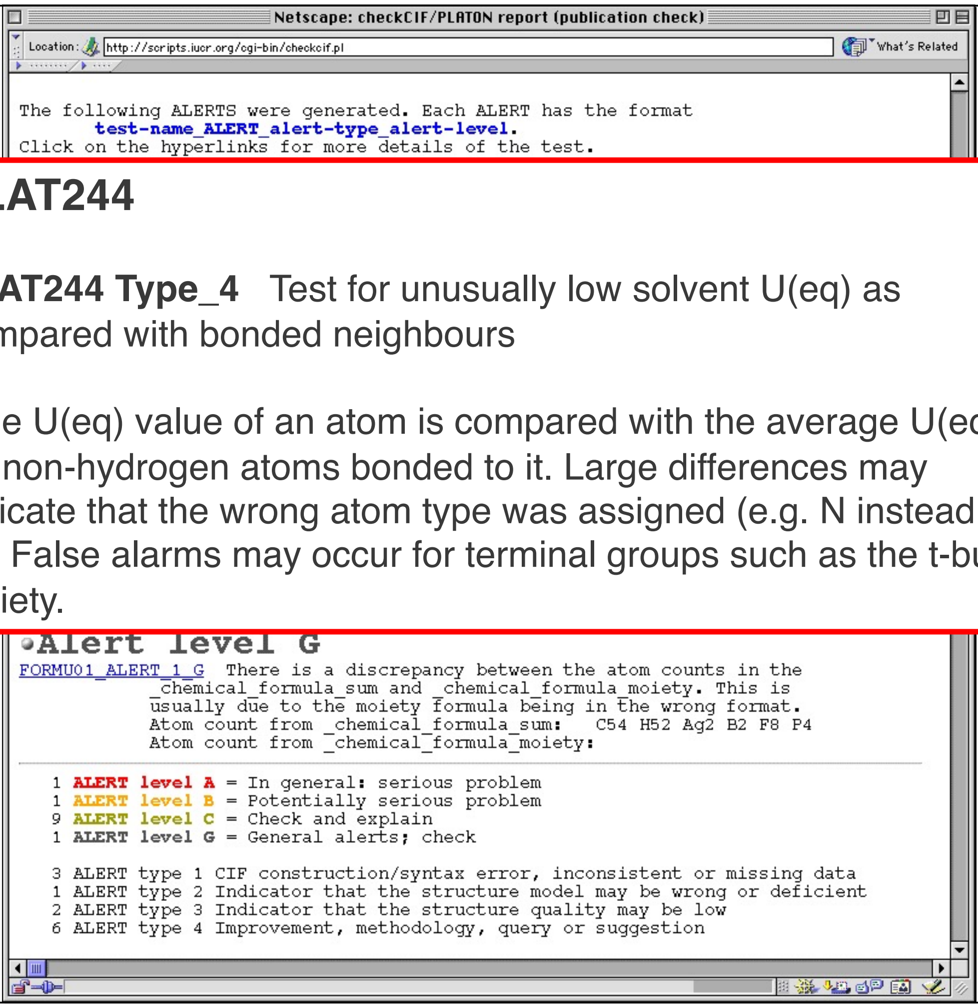
```
PLAT213_ALERT_2_C Atom F8            has ADP max/min Ratio .....        3.2 prola
PLAT242_ALERT_2_C Check Low      Ueq as Compared to Neighbors for       C20
PLAT342_ALERT_3_C Low Bond Precision on  C-C Bonds ...............    0.0104 Ang.
PLAT906_ALERT_3_C Large K value in the Analysis of Variance ......     2.473
PLAT910_ALERT_3_C Missing # of FCF Reflections Below Th(Min) .....         3
PLAT915_ALERT_3_C Low Friedel Pair Coverage ......................     80 %
PLAT971_ALERT_2_C Large Calcd. Non-Metal Positive Residual Density     1.66 eA-3
PLAT972_ALERT_2_C Large Calcd. Non-Metal Negative Residual Density    -1.51 eA-3
```

Alert level G

```
PLAT005_ALERT_5_G No _iucr_refine_instructions_details  in the CIF       ? Do !
PLAT912_ALERT_4_G Missing # of FCF Reflections Above STh/L=  0.600       203
```



```
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 A
DIFF020_ALERT_1_A  _diffrn_standards_interval_count and
             _diffrn_standards_interval_time are missing. Number of measurements
            between standards or time (min) between standards.

Alert level B
PLAT221_ALERT_4_B Large Solvent/Anion  F     Ueq(max)/Ueq(min) ...       4.29 Ratio

Alert level C
PLAT029_ALERT_3_C _diffrn_measured_fraction_theta_full Low .......       0.99
PLAT042_ALERT_1_C Calc. and Rep. MoietyFormula Strings Differ ....          ?
PLAT125_ALERT_4_C No _symmetry_space_group_name_Hall Given .......          ?
PLAT164_ALERT_4_C Nr. of Refined C-H H-Atoms in Heavy-At Struct...          6
PLAT220_ALERT_2_C Large Non-Solvent    C     Ueq(max)/Ueq(min) ...       3.12 Ratio
PLAT222_ALERT_3_C Large Non-Solvent    H     Ueq(max)/Ueq(min) ...       3.68 Ratio
PLAT244_ALERT_4_C Low Solvent  U(eq) as Compared to Neighbors ....         B1
PLAT790_ALERT_4_C Centre of Gravity not Within Unit Cell: Resd.  #          1
             C54 H52 Ag2 P4
PLAT790_ALERT_4_C Centre of Gravity not Within Unit Cell: Resd.  #          2
             B F4

Alert level G
FORMU01_ALERT_1_G  There is a discrepancy between the atom counts in the
             _chemical_formula_sum and _chemical_formula_moiety. This is
            usually due to the moiety formula being in the wrong format.
            Atom count from _chemical_formula_sum:   C54 H52 Ag2 B2 F8 P4
            Atom count from _chemical_formula_moiety:

   1 ALERT level A = In general: serious problem
   1 ALERT level B = Potentially serious problem
   9 ALERT level C = Check and explain
   1 ALERT level G = General alerts; check

   3 ALERT type 1 CIF construction/syntax error, inconsistent or missing data
   1 ALERT type 2 Indicator that the structure model may be wrong or deficient
   2 ALERT type 3 Indicator that the structure quality may be low
   6 ALERT type 4 Improvement, methodology, query or suggestion
```



```
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.
```

PLAT244

PLAT244 Type_4 Test for unusually low solvent U(eq) as compared with bonded neighbours

The U(eq) value of an atom is compared with the average U(eq) for non-hydrogen atoms bonded to it. Large differences may indicate that the wrong atom type was assigned (e.g. N instead of O). False alarms may occur for terminal groups such as the t-butyl moiety.

```
Alert level G
FORMU01_ALERT_1_G  There is a discrepancy between the atom counts in the
             _chemical_formula_sum and _chemical_formula_moiety. This is
            usually due to the moiety formula being in the wrong format.
            Atom count from _chemical_formula_sum:   C54 H52 Ag2 B2 F8 P4
            Atom count from _chemical_formula_moiety:

   1 ALERT level A = In general: serious problem
   1 ALERT level B = Potentially serious problem
   9 ALERT level C = Check and explain
   1 ALERT level G = General alerts; check

   3 ALERT type 1 CIF construction/syntax error, inconsistent or missing data
   1 ALERT type 2 Indicator that the structure model may be wrong or deficient
   2 ALERT type 3 Indicator that the structure quality may be low
   6 ALERT type 4 Improvement, methodology, query or suggestion
```

```
Xtal Data (CIF13   ) x.cif- Set    1(    1): x
Refl Data (LIST4   ) x.fcf [ FCF      ]   (  1): x

http://www.platonsoft.nl/PLATON-MANUAL.pdf
```

```
                       VALIDATION REPORT FOR CURRENT CIF
 #==============================================================================
 # PLATON/CHECK-( 60513) versus check.def version of 040513  for entry: x
 # Data From: x.cif - Data Type: CIF                Bond Precision   C-C = 0.0104 A
 # Refl Data: x.fcf - Data Type: LIST4                                Temp = 160 K
 #                                                           Nref/Npar =  12.3
 # UCL  15.4933(3)  26.3923(5) 4.40199(10)              90            90            90
 # WaveLength 0.71073    Volume Reported     1799.99(6)  Calculated     1799.99(6)
 # SpaceGroup from Symmetry P n a 21   Hall: P 2c -2n
 #                  Reported P n a 21         P 2c -2n
 # MoietyFormula C20 H12 Au F8 N
 #      Reported C20 H12 Au F8 N
 #    SumFormula C20 H12 Au F8 N
 #      Reported C20 H12 Au F8 N
 # Mr          =    615.28[Calc],    615.27[Rep]
 # Dx,gcm-3    =     2.270[Calc],     2.270[Rep]
 # Z           =         4[Calc],         4[Rep]
 # Mu (mm-1)   =     8.260[Calc],     8.260[Rep]
 # F000        =    1160.0[Calc],    1160.0[Rep]  or F000' =   1153.05[Calc]
 # Reported   T Limits: Tmin=0.180                Tmax=0.652  AbsCorr=GAUSSIAN
 # Calculated T Limits: Tmin=0.287 Tmin'=0.053  Tmax=0.342
 # Reported   Hmax= 22, Kmax= 39, Lmax=  6, Nref=  5610          , Th(max)= 32.214
 # Obs in FCF Hmax= 22, Kmax= 39, Lmax=  6, Nref=  5610[  3338], Th(max)= 32.214
 # Calculated Hmax= 23, Kmax= 39, Lmax=  6, Nref=  6379[  3550], Ratio=1.58/0.88
 # Reported   Rho(min) = -1.94, Rho(max) =  1.56 e/Ang**3 (From CIF)
 # Calculated Rho(min) = -2.08, Rho(max) =  1.66 e/Ang**3 (From CIF+FCF data)
 # w=1/[sigma**2(Fo**2)+(0.0303P)**2+  4.2752P], P=(Fo**2+2*Fc**2)/3
 # R= 0.0339(  5123), wR2= 0.0756(  5610), S = 1.110      (From CIF+FCF data)
 # R= 0.0339(  5123), wR2= 0.0756(  5610), S = 1.110      (From FCF data only)
 # R= 0.0339(  5123), wR2= 0.0756(  5610), S = 1.110, Npar= 272, Flack -0.001(6)
 # Number Bijvoet Pairs = 2272            ( 1909 Selected for:  Parsons -0.002(4)
 # P2(tr) 1.000,P3(tr) 1.000,Bijvoet Pair Coverage (Perc) =  80, Hooft -0.006(6)
 #==============================================================================
 For Documentation: http://www.platonsoft.nl/CIF-VALIDATION.pdf
 #==============================================================================

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

INSTRUCTION INPUT via KEYBOARD or LEFT-MOUSE-CLICKS (HELP with RIGHT CLICKS)

>>  Continue (Y/N[Y])
```

Alert indicators

```
380 ALERT 4 C Likely Unrefined X(sp2)-Methyl Moiety ...... C18
412 ALERT 2 C Short Intra XH3 .. XHn : H19B .. H30A = 1.81 Ang.
720 ALERT 4 C Number of Unusual/Non-Standard Label(s) .... 1
```

Alert numbers 1-5 indicate the type of issue.

Alerts levels A, B, C indicate the severity of the issue.

G is a general issue to check or information, not necessarily an error.

Alert types

```
380 ALERT 4 C Likely Unrefined X(sp2)-Methyl Moiety ...... C18
412 ALERT 2 C Short Intra XH3 .. XHn : H19B .. H30A = 1.81 Ang.
720 ALERT 4 C Number of Unusual/Non-Standard Label(s) .... 1
```

```
ALERT Type 1  = CIF construction/syntax error,
                inconsistent or missing data

ALERT Type 2  = Indicator that the structure
                model may be wrong or deficient

ALERT Type 3  = Indicator that the structure quality
                may be low

ALERT Type 4  = Improvement, methodology, query or
                suggestion

ALERT Type 5  = Informative message, check
```

checkCIF alert levels

✦ A Serious – attention essential

Required item omitted, large deviation from usually expected value, or inconsistent values

Alert A	No _chemical_absolute_configuration info
Alert A	No crystal dimensions have been given
Alert A	Atom C58A ADP max/min Ratio 18.00
Alert A	H…A calc 5.82(3); rep 1.915; dev 3.91 Å
Alert A	Space group symbol does not match sym. ops.

Alert levels

✦B Significant – action needed?

Item is a significant or unexpected outlier

Alert B	The formula has elements in wrong order
Alert B	ADDSYM detects Cc to Fdd2 transformation
Alert B	Refined extinction parameter < 1.9σ
Alert B	Structure contains VOIDS of 130.00 Å^3

Alert levels

✦ C Outside expected norms – examine

May appear trivial, but do not dismiss out of hand

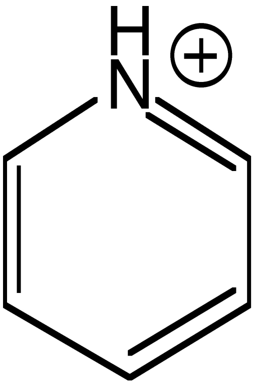
A long list may indicate subtle errors

Alert C Moiety formula not given

Alert C Short inter X...Y contact: O7...C1 = 2.96 Å

Alert C Low U(eq) as compared to neighbours: C1

Alert C D-H without acceptor N2–H2 ?



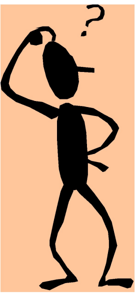
Alert C Short inter X...Y contact: O7...C1 = 2.96 Å

Alert C Low U(eq) as compared to neighbours: C1

Alert C D-H without acceptor N2–H2 ?

}

Don’t dismiss C alerts
Some things hard to find manually



```
Alert A   Angle   Calc   120.6(12), Rep   116.0(17), Dev.  3.83 Sigma
                  C8   -N83  -H831    1.555   1.555   1.555

Alert B   Angle   Calc   118.2(17), Rep   121.8(17), Dev.  2.12 Sigma
                  H831 -N83  -H832    1.555   1.555   1.555

Alert C   D-H     Calc   9.701(16), Rep     0.990              Sym ops?
                  O12  -H1      1.555   1.555
```

Update CIFs carefully and thoroughly!

Alert levels

✦G General issues to check

Not necessarily an error

A reminder prompt, in case there is an oversight

Do the results concur with (chemical) expectation?

```
ALERT G
Atom count from _chemical_formula_sum:   C46 H54 N4 O26 Ti1
Atom count from the _atom_site data:     C46 H41 N4 O26 Ti1
WARNING: H atoms missing from atom site list. Intentional?

ALERT_1_G Confirm the Absolute Configuration of C1:    S
```

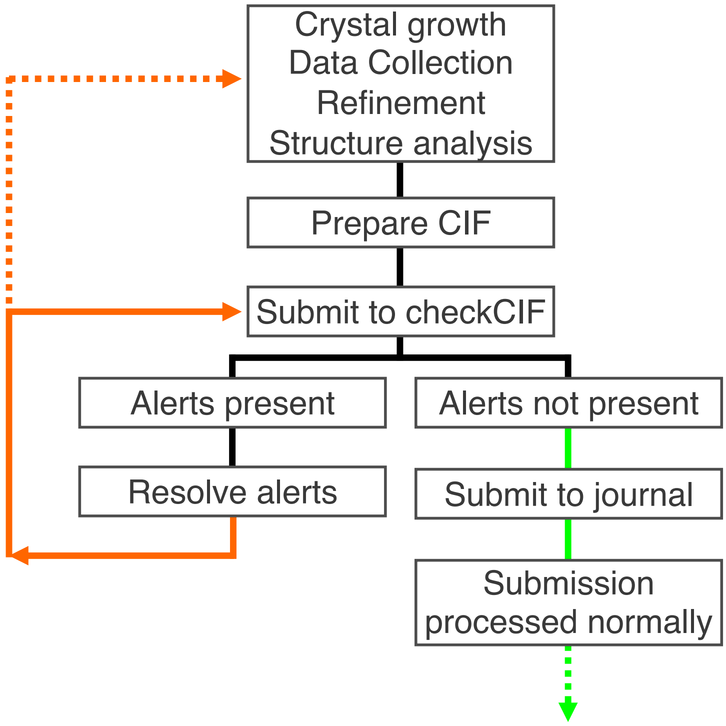
Vigilance – additional to validation

- Does the structure make sense to you?
- Does the structure look right and is it geometrically logical?
- Can you rationalize structure with the expected or plausible chemistry, etc.
- Don’t force (restrain) a structure to be that which it is not.
- Does the geometry agree with similar structures in databases?
- Unusual geometry or other features are rarely a new property – more likely to be the effect of an inadequacy of the model
- Look critically at the output files (e.g. .lst file)
- ALWAYS revalidate when you make any change to a CIF

Possible limits to validation

- Test not (yet) implemented
- Test not practical
- Error not a validation issue
- Mistake cannot be detected from data in CIF
- Nonsense entries in the CIF

Authors working with checkCIF



Crystal growth
Data Collection
Refinement
Structure analysis

Prepare CIF

Submit to checkCIF

Alerts present

Resolve alerts

Alerts not present

Submit to journal

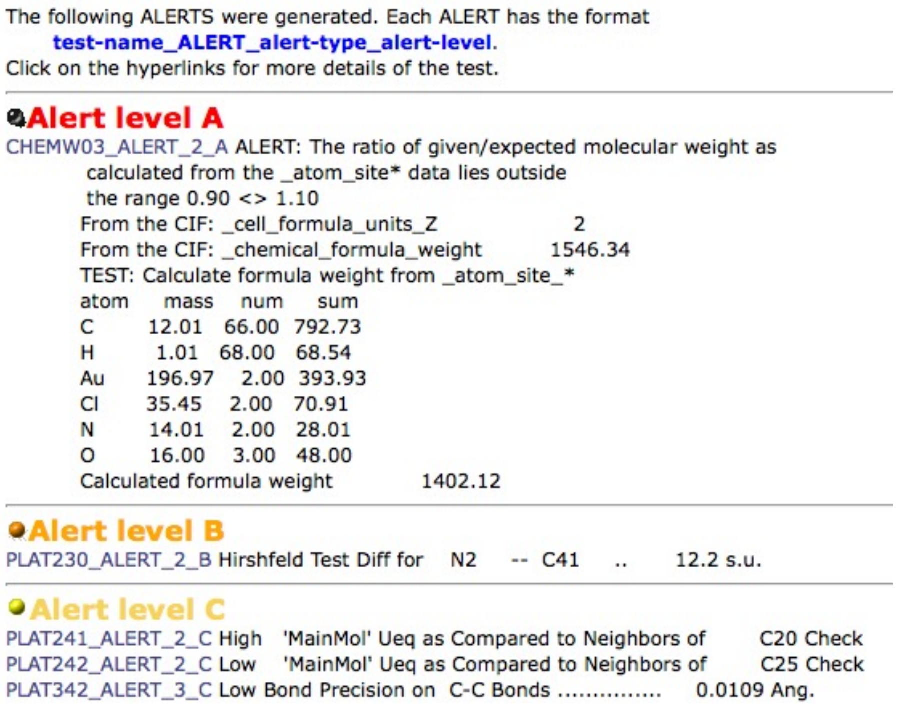
Submission processed normally

Still getting A (or B) alerts?

- Is there a sound scientific basis for the outlier?
- Insert Validation Response Form (VRF) into CIF
- Give clear short but scientifically sound comments, *i.e*. properly document your experiment.
- Show you understand the causes of the outlier
- Use also _exptl_special_details or _refine_special_details

- Editors and reviewers then know you are aware

Initial checkCIF report



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 A

```
CHEMW03_ALERT_2_A ALERT: The ratio of given/expected molecular weight as
           calculated from the _atom_site* data lies outside
           the range 0.90 <> 1.10
          From the CIF: _cell_formula_units_Z              2
          From the CIF: _chemical_formula_weight        1546.34
          TEST: Calculate formula weight from _atom_site_*
          atom    mass    num    sum
          C      12.01   66.00  792.73
          H       1.01   68.00   68.54
          Au    196.97    2.00  393.93
          Cl     35.45    2.00   70.91
          N      14.01    2.00   28.01
          O      16.00    3.00   48.00
          Calculated formula weight           1402.12
```

Alert level B

```
PLAT230_ALERT_2_B Hirshfeld Test Diff for    N2    --  C41    ..      12.2 s.u.
```

Alert level C

```
PLAT241_ALERT_2_C High   'MainMol' Ueq as Compared to Neighbors of       C20 Check
PLAT242_ALERT_2_C Low    'MainMol' Ueq as Compared to Neighbors of       C25 Check
PLAT342_ALERT_3_C Low Bond Precision on  C-C Bonds ..............    0.0109 Ang.
```

Scroll down to…

Validation response form

Please find below a validation response form (VRF) that can be filled in and pasted into your CIF.

```
# start Validation Reply Form
_vrf_CHEMW03_NV1610
;
PROBLEM: ALERT: The ratio of given/expected molecular weight as
RESPONSE: ...
;
# end Validation Reply Form
```

```
# start Validation Reply Form
_vrf_CHEMW03_NV1610
;
PROBLEM: ALERT: The ratio of given/expected molecular weight as
RESPONSE: SQUEEZE used to handle disordered solvent, but the
sum and moiety formulae, formula weight, F(000), density and mu all
include the total estimated solvent content of the crystal.
See _refine_special_details
;
# end Validation Reply Form
```

New checkCIF report with responses

Alert level A

CHEMW03_ALERT_2_A ALERT: The ratio of given/expected molecular weight as calculated from the _atom_site* data lies outside the range 0.90 < > 1.10

```
From the CIF: _cell_formula_units_Z                  2
From the CIF: _chemical_formula_weight         1546.34
TEST: Calculate formula weight from _atom_site_*
atom     mass     num     sum
C        12.01    66.00  792.73
H         1.01    68.00   68.54
Au      196.97     2.00  393.93
Cl       35.45     2.00   70.91
N        14.01     2.00   28.01
O        16.00     3.00   48.00
Calculated formula weight            1402.12
```

Author Response: SQUEEZE used to handle disordered solvent, but the sum and moiety formulae, formula weight, F(000), density and mu all include the total estimated solvent content of the crystal. See _refine_special_details.

A good argument

Alert B ADDSYM detects additional (pseudo) symmetry element: I

Author Response:
This additional symmetry element does not hold true for one of the ether bridges, as discussed in the text.

A less good answer

Alert A < 85% complete (θ max?)

Author Response:

Hemisphere of data collected

Space group $P2_1/n$, Nonius FAST system

- Does not address reason for missing data
- Is it –

  inherent geometrical limitations?

  oversight in experiment or data reduction?

With a “difficult” structure

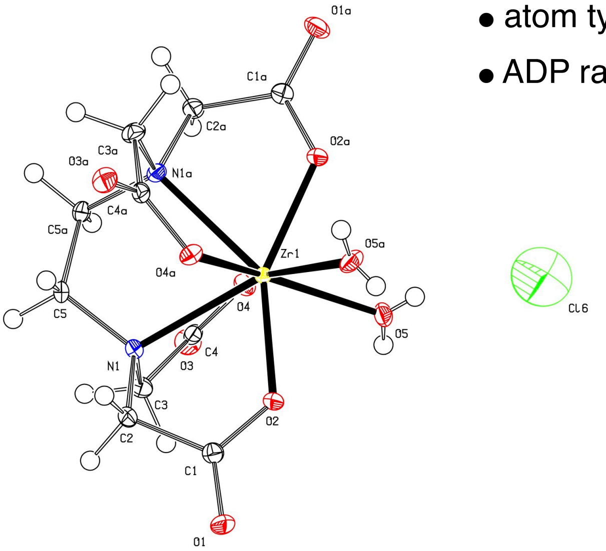
- Identify and describe the problem
- Give details of remedial action taken
- Describe the (successful?) outcome
- Use:
  - experimental section of the paper
  - add text to the CIF, e.g. _refine_special_details

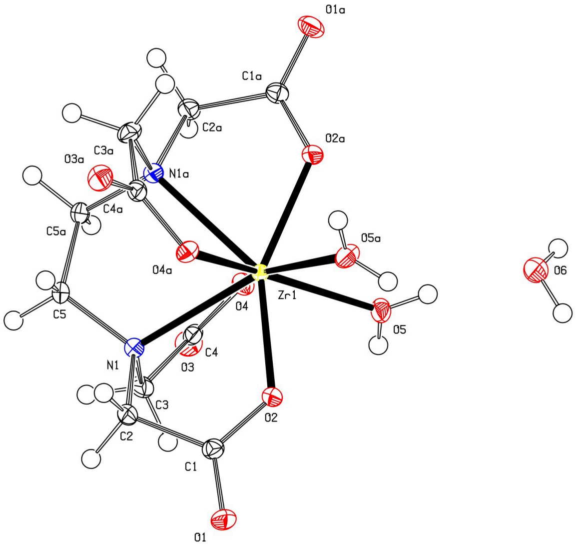
Test not implemented

Example: High ADPs on isolated atom

Not detected by

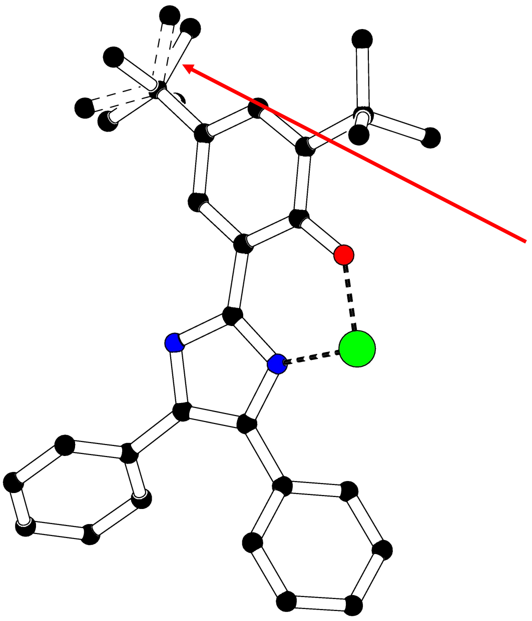
- rigid bond test
- atom type test
- ADP ratio test





Atom is probably O of H_2O rather than Cl

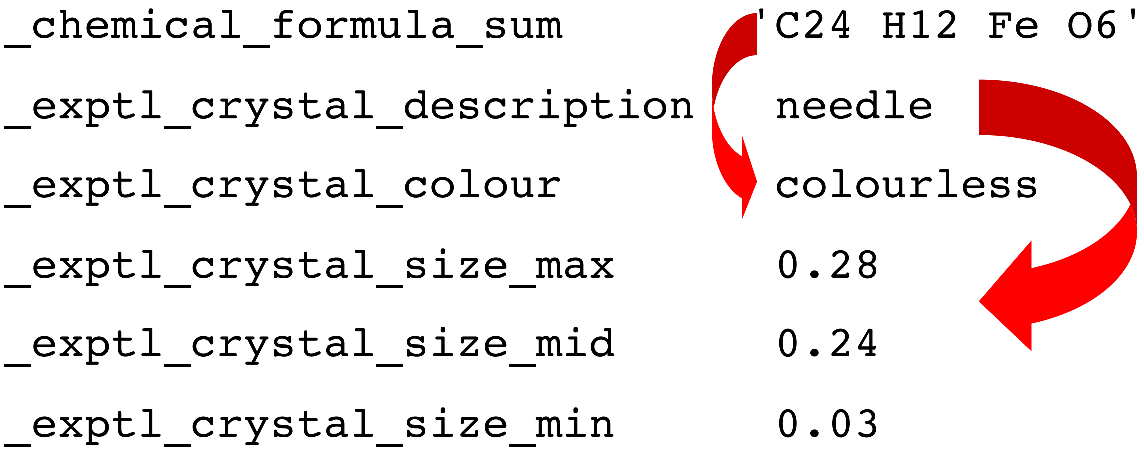
Test not practical



C-C range is 1.49 to 1.60 Å

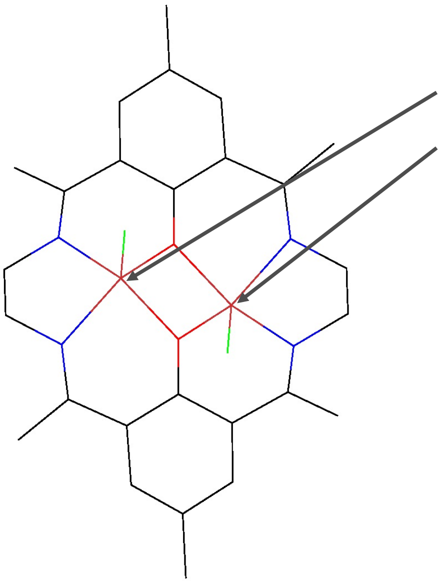
However, C-C single bonds are found within this range

Error not a validation issue



```
_chemical_formula_sum            'C24 H12 Fe O6'
_exptl_crystal_description        needle
_exptl_crystal_colour             colourless
_exptl_crystal_size_max           0.28
_exptl_crystal_size_mid           0.24
_exptl_crystal_size_min           0.03
```

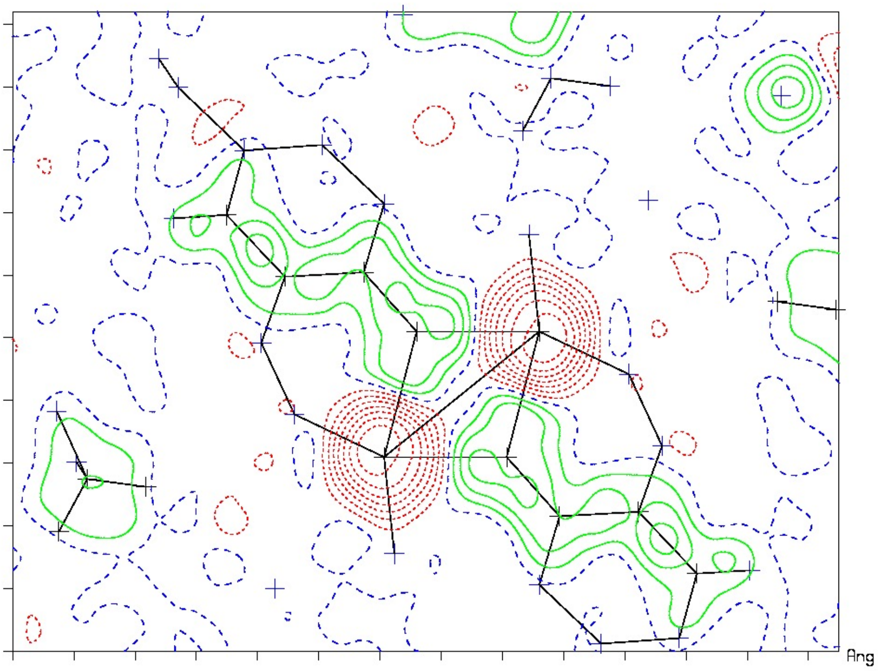
Error not detectable from CIF data



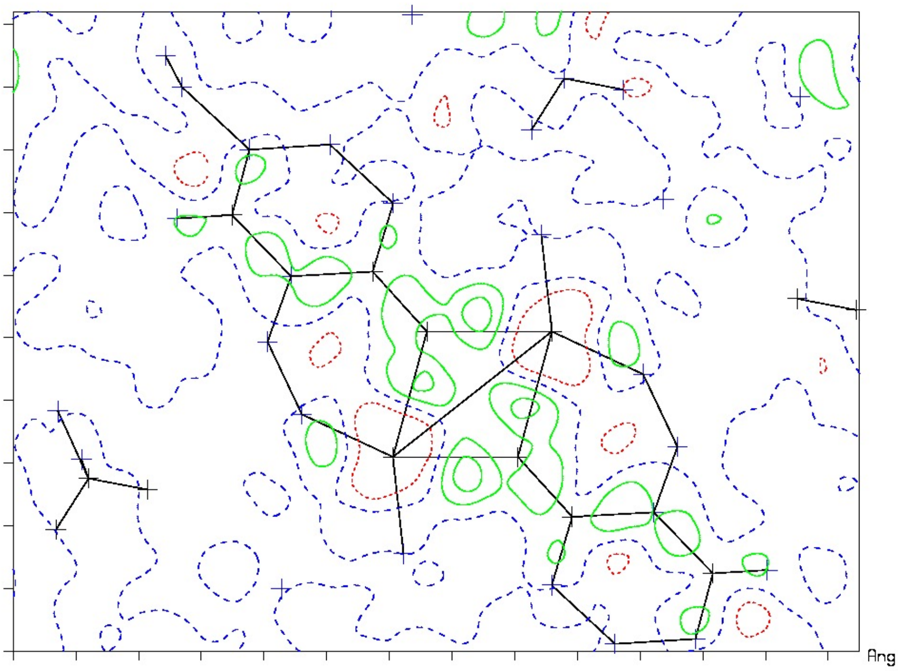
Prior chemical information: complex is either Ru / Ru or Ru / Zn

Refinement as Ru / Ru gave R_1 = 0.064; unusual five-coordinate Ru geometry

Difference map with Ru / Ru model ($R_1 = 0.064$)



Difference map with Zn / Zn model ($R_1 = 0.022$)

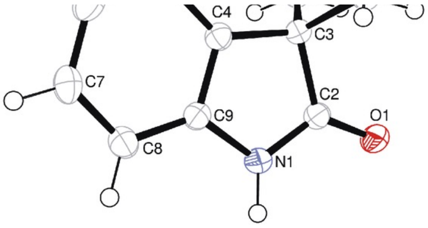


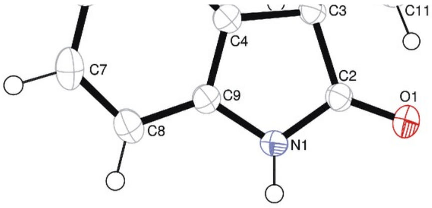
Nonsense entries in the CIF

see Clegg, W. (2003). *Acta Cryst.* E**59**, e2-e5

```
_diffrn_ambient_temperature                  293(2)
_diffrn_radiation_wavelength                 0.69010
_diffrn_radiation_type                       synchrotron
_diffrn_radiation_source        'fine-focus sealed tube'
_diffrn_radiation_monochromator              graphite
_diffrn_measurement_device_type
               'SMART 1k on Daresbury SRS Station 9.8'
```

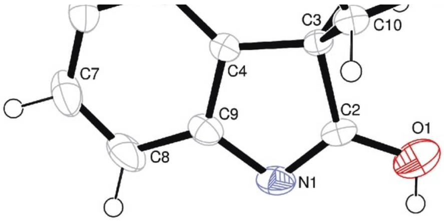
Misassigned element

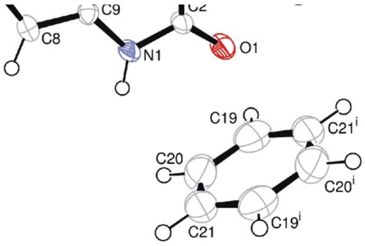




Four related lactams. One is a "rarely seen imidic acid tautomer"

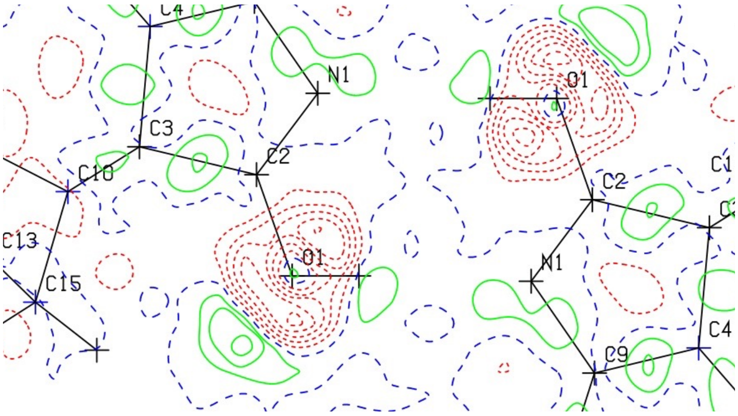
R = 0.059, wR2 = 0.177, S = 1.067





230_ALERT_2_B Hirshfeld Test Diff for O1 -- C2 .. 11.83 su

```
Peaks list
Q1     0.54    1.07 O1
Q2     0.28    0.77 C3
Q3     0.26    0.73 C3
Q4     0.25    0.76 C10
```



Contoured difference maps are very useful – easy in *PLATON*

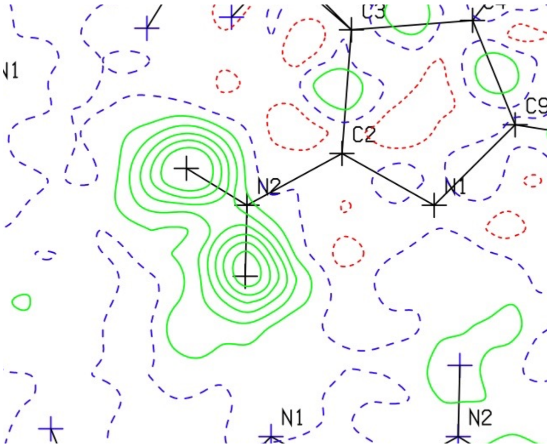
Refine as an amine

R = 0.046, wR2 = 0.117
(formerly R = 0.059)

No relevant alerts

```
Q1 0.22    0.77 C3
```

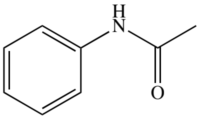
N2 is pyramidal.
Do not fall into the trap of thinking it is planar (imine) and use AFIX 93!



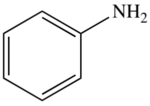
Now the chemist has work to do!

Geometry of –NH and –NH_2 groups

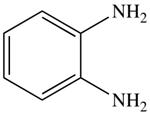
Amides: planar



Phenylamines: usually planar



Phenydiamines: one of the
amine groups may be pyramidal



Validation is not usually revealing.

Be careful about auto-calculation of H with amines and hydroxy groups.

Test the H-atom positions: refine the H-atoms, or refine their Uiso values.

Look at contoured difference maps.

Missing H atom

The issue raises only a G alert

```
343_ALERT_2_G Check sp? Angle Range
in Main Residue for .. C18
```

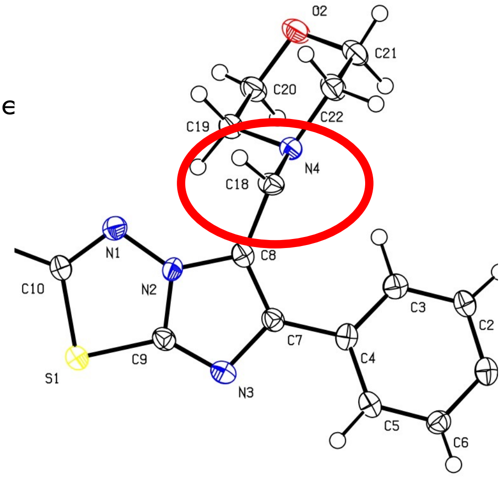
Largest peak: 0.84 e/Å^3

H-atoms from diff. map and refined.

So one H was missed, but…

No mismatched formula!

Author claims that structure is fine because
there is no serious *checkCIF* alert



LOOK at and understand the structure AND the chemistry

What's wrong here?

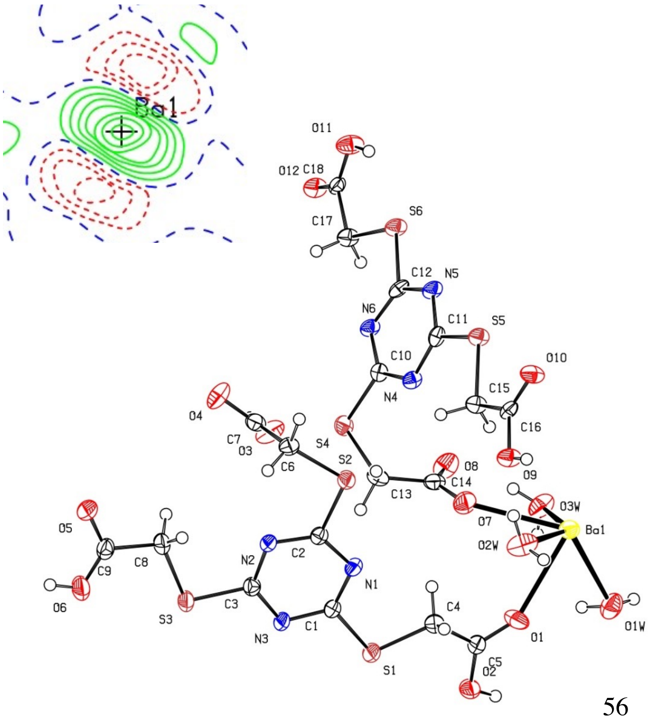
R = 0.047, wR = 0.088,
shift/error 0.000

Looks reasonable visually, but…

Large peak near Ba: 3.4 e/Å^3
A alert

Hard to be sure about hydroxy & water H-atoms – diff. maps quite noisy

C alert about poor Ba-O-H angles for one water



Refine again

R = 0.035, down from 0.047 !!

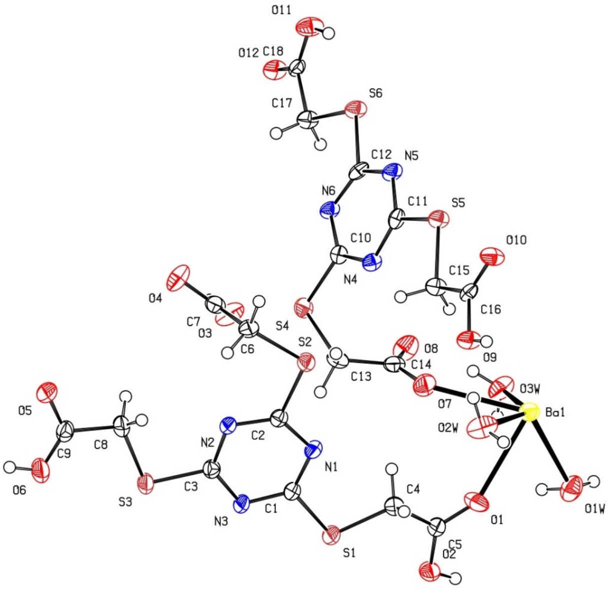
Ba ellipsoid shrinks (large shift/error)

Residual peaks gone

Diff. maps clean and H atoms clear

Using element other than Ba did not reproduce author's "converged" result

Why???



Improper use of DAMP 0 0

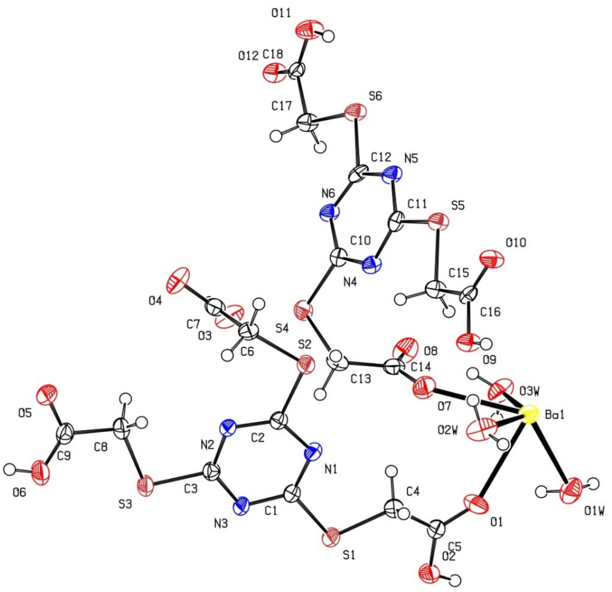
Structure had not yet converged

DAMP 0 0 sets shifts to zero!

Use ONLY for GCLS refinement
AFTER full convergence
(to generate s.u.s)

NEVER for L.S. refinement

GCLS should not usually be
needed for final refinement of
small-molecule structures



Summary

- *CheckCIF* is a tool for YOU
- Be vigilant – do not rely solely on *checkCIF*
- Structure factor validation is also very important

For proper review, referees need the fcf files!

How many journals require their submission?

How many wrong structures are missed because a journal
does not require structure factor submission?

SHELXL-2014 solved that problem by embedding the
.res and .hkl files. Don't delete the hkl data in a CIF.