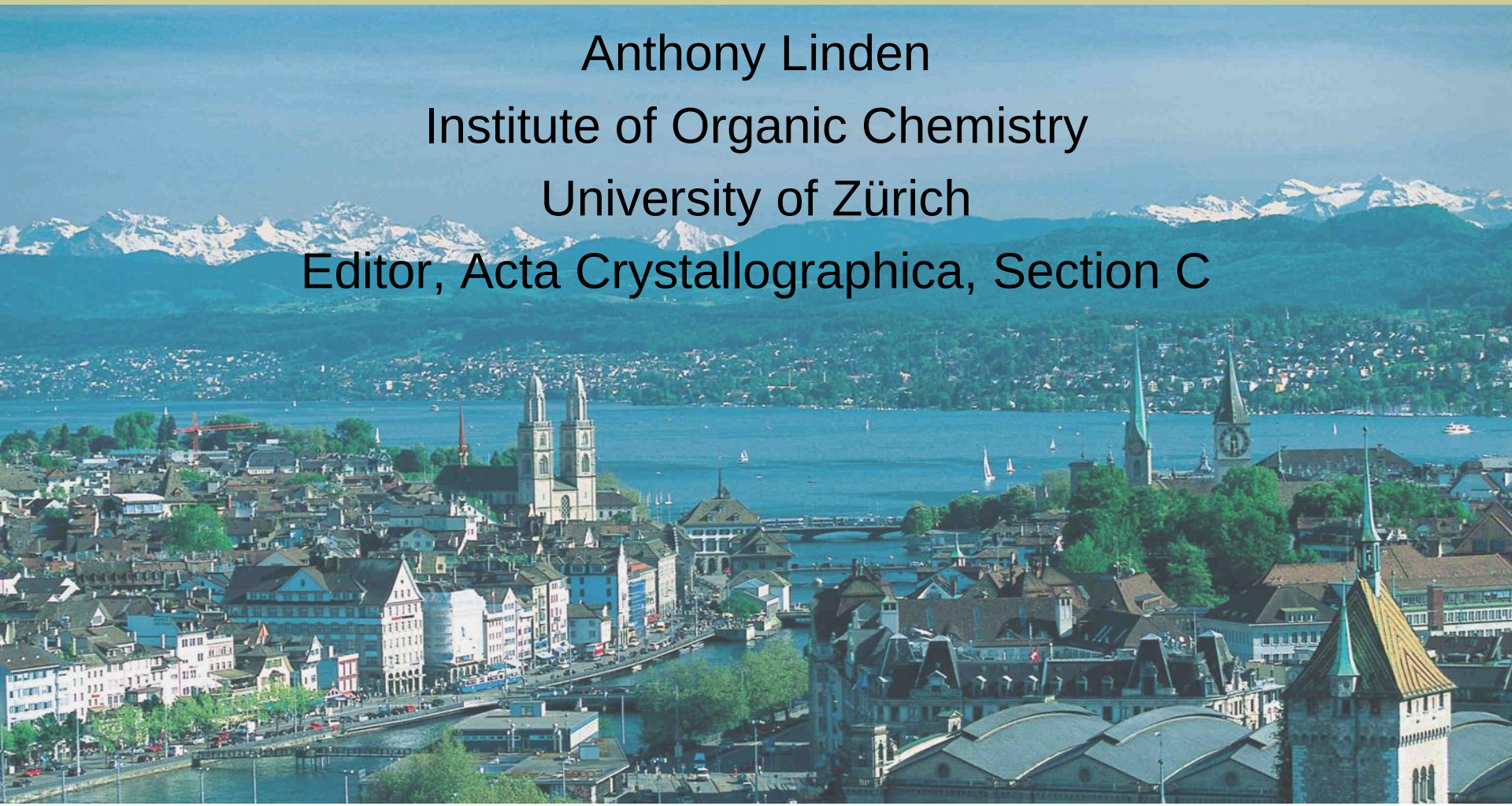
Validating a small-unit-cell structure; understanding *checkCIF* reports

Anthony Linden
Institute of Organic Chemistry
University of Zürich
Editor, Acta Crystallographica, Section C



Anno BC: before CIF

Do we know how life was before the motor car?

Do we know (remember) how life was before CIF?

How were data & results transmitted?

- Different labs used different software & instrumentation
- No common format/content: a lot of manual effort
- Potential for typographical errors
- Mailed sheaves of paper, later floppy disks
- How did we know everything was complete and correct?
- Did we do Validation? Back then we had lots of time…
- Publication took many months

The CIF breakthrough

- Easy to follow 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)*

- Getting software developers on-board early was key
- SHELXL-93 was very timely
- 1994: Adoption of the CIF standard for Acta Cryst. C submissions
- 1997: Beginning of automated validation: *checkCIF*
- Current standard reference: *International Tables Volume G*

Validation – 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; from 1994
- Nice GUIs, but people often no longer look at output/log files
- More non-experts determining structures
- 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; no laborious manual checking

What is validation?

Comparison against *normally expected* values or conditions

- Are all the usual information and data present?
- Do related or derived parameters match?
- Are parameters consistent and logical?
- Are there significant outliers?
- Has the refinement converged?
- Is the space group correct?
- Are the assigned atom types correct?
- etc, etc, etc…

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…

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

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!

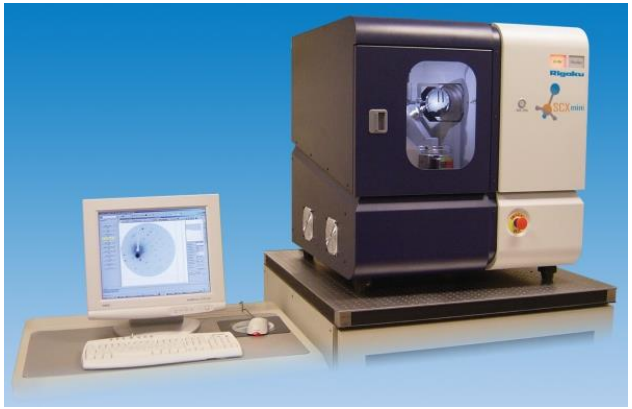
Are validation and vigilance still needed?

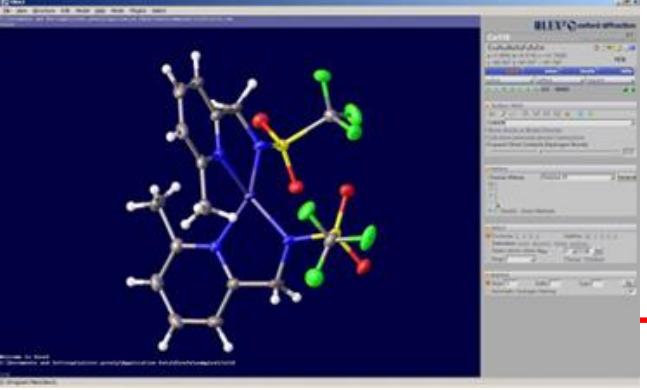
Many avoidable mistakes still appear in submitted or published papers

- 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

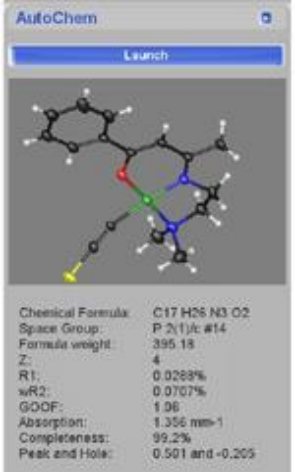
Automation

- New generation of fully-automated diffractometers
- Progress in automatic structure solution & refinement
- Manufacturers promise:
  “No or little crystallographic knowledge required”
  “Routine small molecule structure determination is accessible to students and scientists of other disciplines”

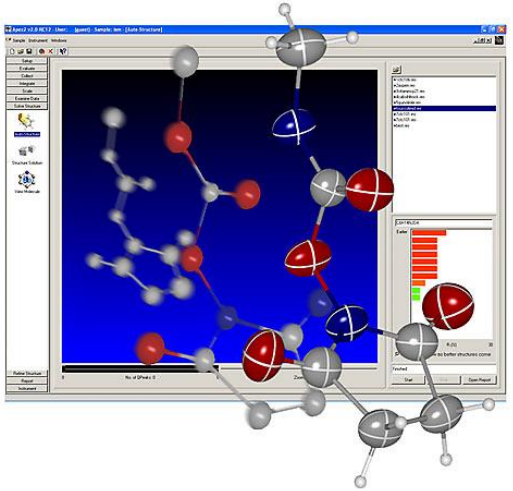




Automation



- Drop in a crystal, push a button, sit back, and …
- Pretty picture without further ado – if there are no alerts, it must be OK, right?
- Can a person with “no crystallographic knowledge” rely on that (yet)?
- Further checking of results seems essential (e.g. element assignments)
- If result is not expected molecule, what then?



checkcif.iucr.org

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:
Choose File no file selected

Select form of checkCIF report
(•) HTML ( ) PDF

Select validation type
(•) Full 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
(•) None

Send CIF for checking

---

checkCIF is sponsored by

Crystallography Journals Online

ELSEVIER

WILEY

ROYAL SOCIETY OF CHEMISTRY

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. CIF dictionary
Please wait while processing .... 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

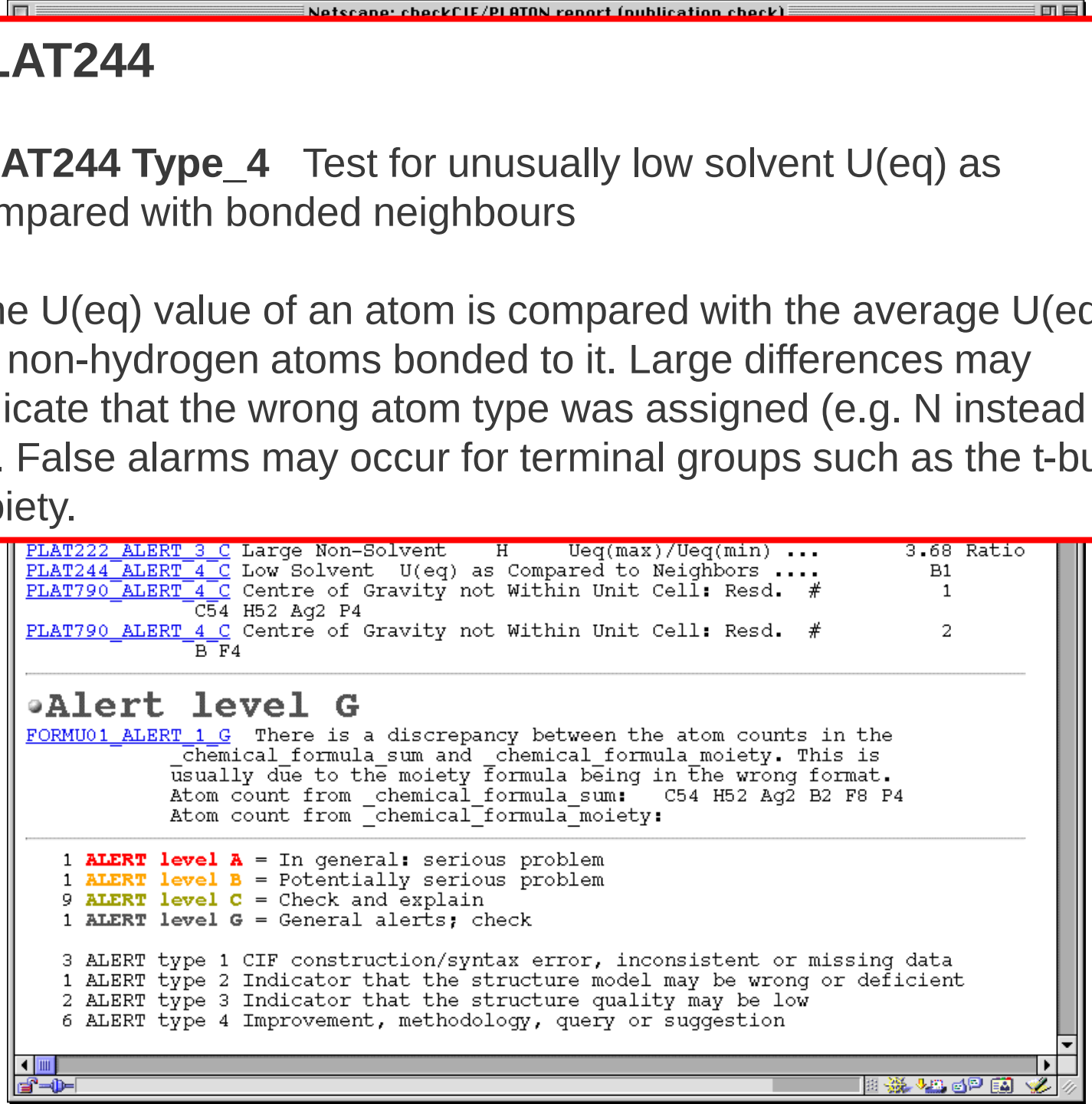
Alert level G

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

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.



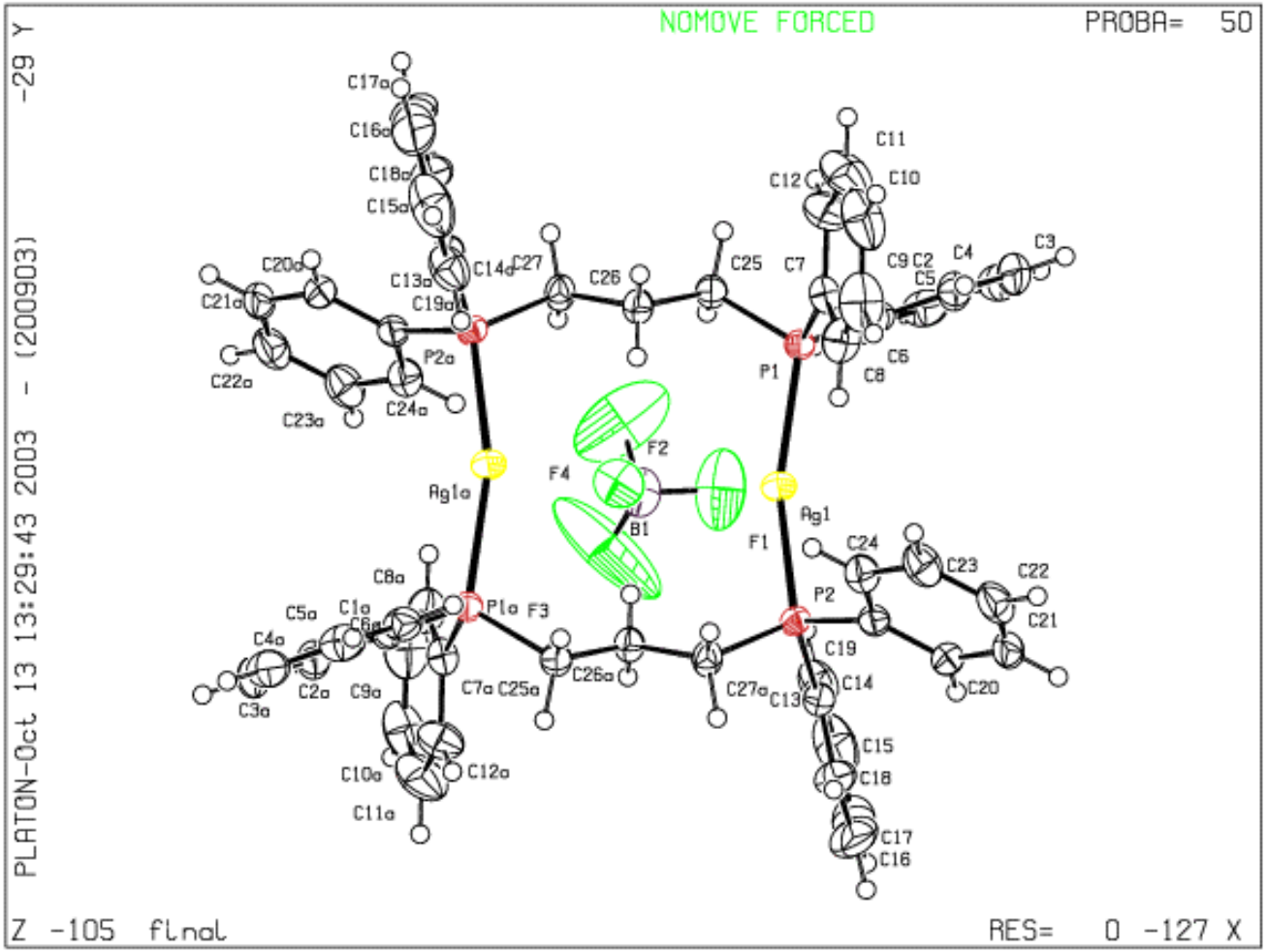
Netscape: checkCIF/PLATON report (publication check)

Location : http://scripts.iucr.org/cgi-bin/checkcif.pl

What's Related

PLATON version of 20/09/2003; check.def file version of 12/09/2003

Datablock final - ellipsoid plot



P L A T O N

A Multipurpose Crystallographic Tool

GRAPHICS	GEOM-CALC	VOIDS FLIP	SYMMETRY	ABSORPTION	REPORT	MISC-TOOLS
PLUTONauto	Calc All	Calc Solv	ADDSYM	MULscanABS	Validation	SYSTEM-S
ORTEP/ADP	Calc Intra	Calc K.P.I	ADDSYM-EQL	ABSPstScan	ASYM-VIEW	fcf2hkl
NewmanPlot	Calc Inter	SQUEEZE	ADDSYM-EXT	ABSTompa	FCF-Valid	Expand2P1
Ring-Plots	Calc Coord	CalcFCF-SQ	ADDSYM-PLT	ABSGauss	DifFourier	FCF-Gener
Plane-Plot	Calc Metal	Contour-SQ	ADDSYM-SHX	ABSXtal	ANALofVAR	HKL-Gener
Polyhedra	Calc Geom	Solv F3D	NEWSYM	ABSSphere	ByvoetPair	HKL-Transf
ContourDif	Calc Hbond	Solv Plot	NONSYM	SHXABS	ASYM-EXPCT	EXOR-RES
Contour-Fo	Calc TMA	CavityPlot	LePage	AnomDisVal	ASYM-Valid	ANIS-RES
AutoMolFit	L.S.-PLANE		DelRed	AnomDisPlt	SupplMater	Rename-RES
HKL2Powder	DihedAngle		MOLSYM	MuPlot	EXPECT-HKL	Auto-Renum
SimPowderP	AngleLines	FLIP MENU	SPGRfromEX		CSD-CELL	Create-spf
RadDistFun	AngLsplLin	Flip Show	ASYM		CSD-QUEST	Create-res
Patterson	CremerPopl	Flip Patt	ASYMaverFR		StructTidy	Create-clf
ShelxtPlot	BondValenc	FLIPPER 25	LePageTwin	XtlPlanAgl	StralnAnal	Create-pdb
PLUTONatlv	HFIX - RES	STRUCTURE?	TwinRotMat	Xtal Habit	LocCIF-acc	clf2shelxl

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

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

PLATON 10
OptionMenus
NoMove
Join-Expand
Organic
Round
Parentheses
Label-Alias
R/S-Determ
Norm-H-bond
NoSymm
NoDisorder
LstARU RCel
LstCellSymm
ListAtoms
ListBonds
LstFlagRadi
Exclude H
MinQPeakHgt
MinQPeakDis
Q-Peak-Incl
KeyInstruct
Prev Next
SAVE-InstrS
ENTRY-LIST
Reset End

WORKING

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])

PLATON 10
OptionMenus
NoMove
Join-Expand
Organic
Round
Parentheses
Label-Alias
R/S-Determ
Norm-H-bond
NoSymm
NoDisorder
LstARU RCel
LstCellSymm
ListAtoms
ListBonds
LstFlagRadi
Exclude H
MinQPeakHgt
MinQPeakDis
Q-Peak-Incl
KeyInstruct
Prev Next
SAVE-InstrS
ENTRY-LIST
Reset End
Exit
MenuActive

Alert indicators

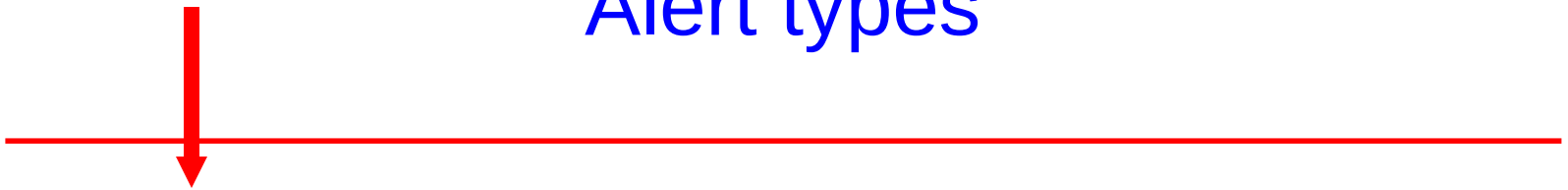
```
380 ALERT 4 C Likely Unrefined X(sp2)-Methyl Moiety ..... C18
412 ALERT 2 C Short Intra XH3 .. XHn : H19B .. H30A = 1.81 Ang.
720 ALERT 4 C Number of Unusual/Non-Standard Label(s) .... 1
```


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 crystal dimensions have been given
Alert A	No _chemical_absolute_configuration info
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 $Å^3$

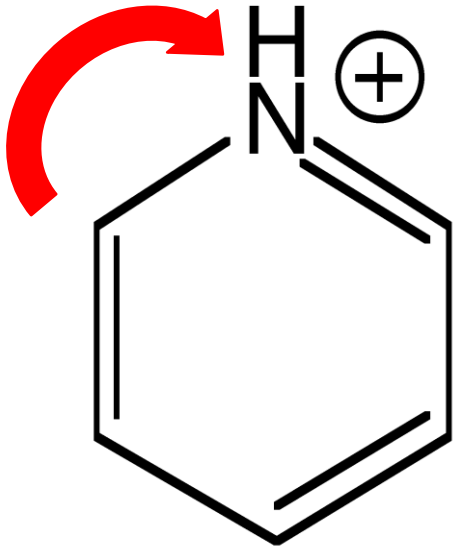
Alert levels

☞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 ?

Alert levels

☞G General issues to check

Not necessarily an error

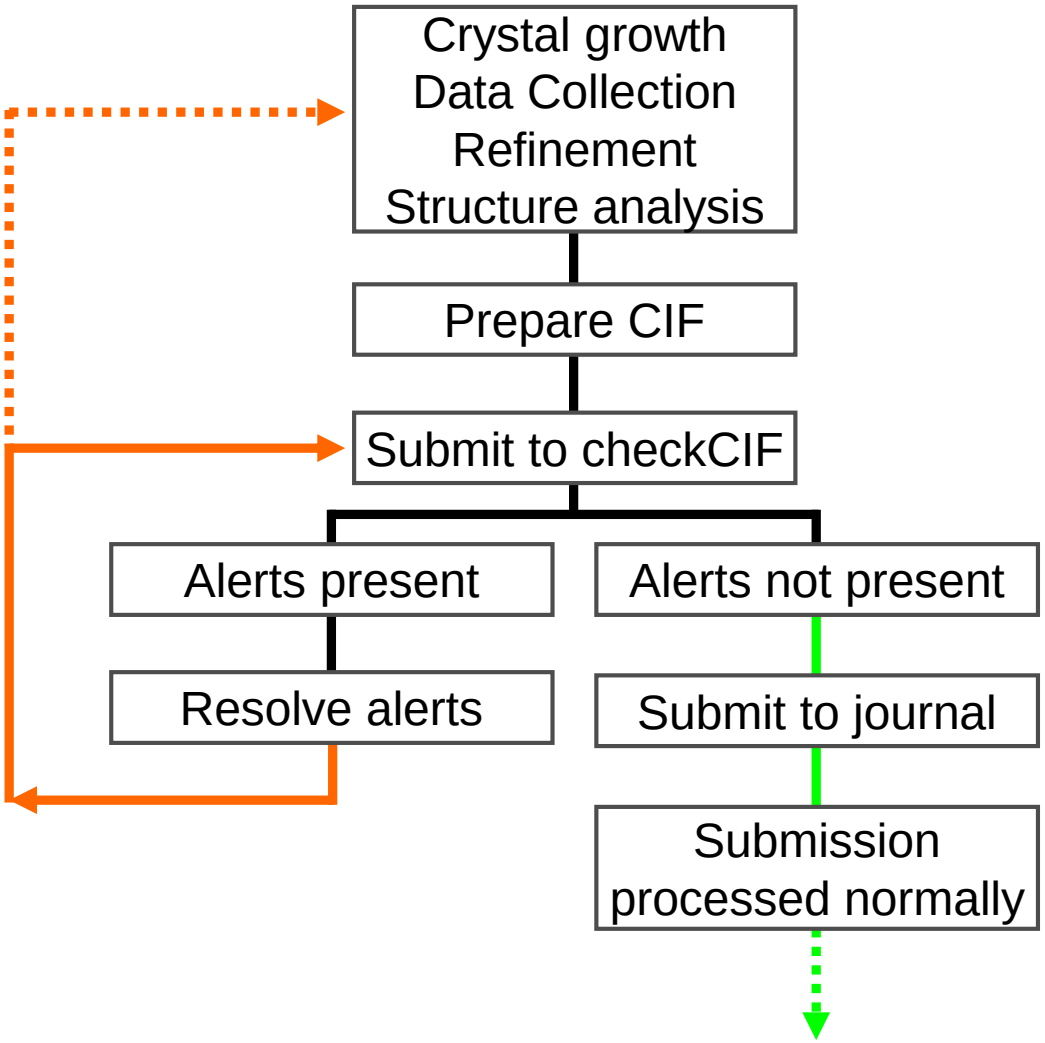
A reminder prompt, in case there is an oversight

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
```

Authors working with checkCIF

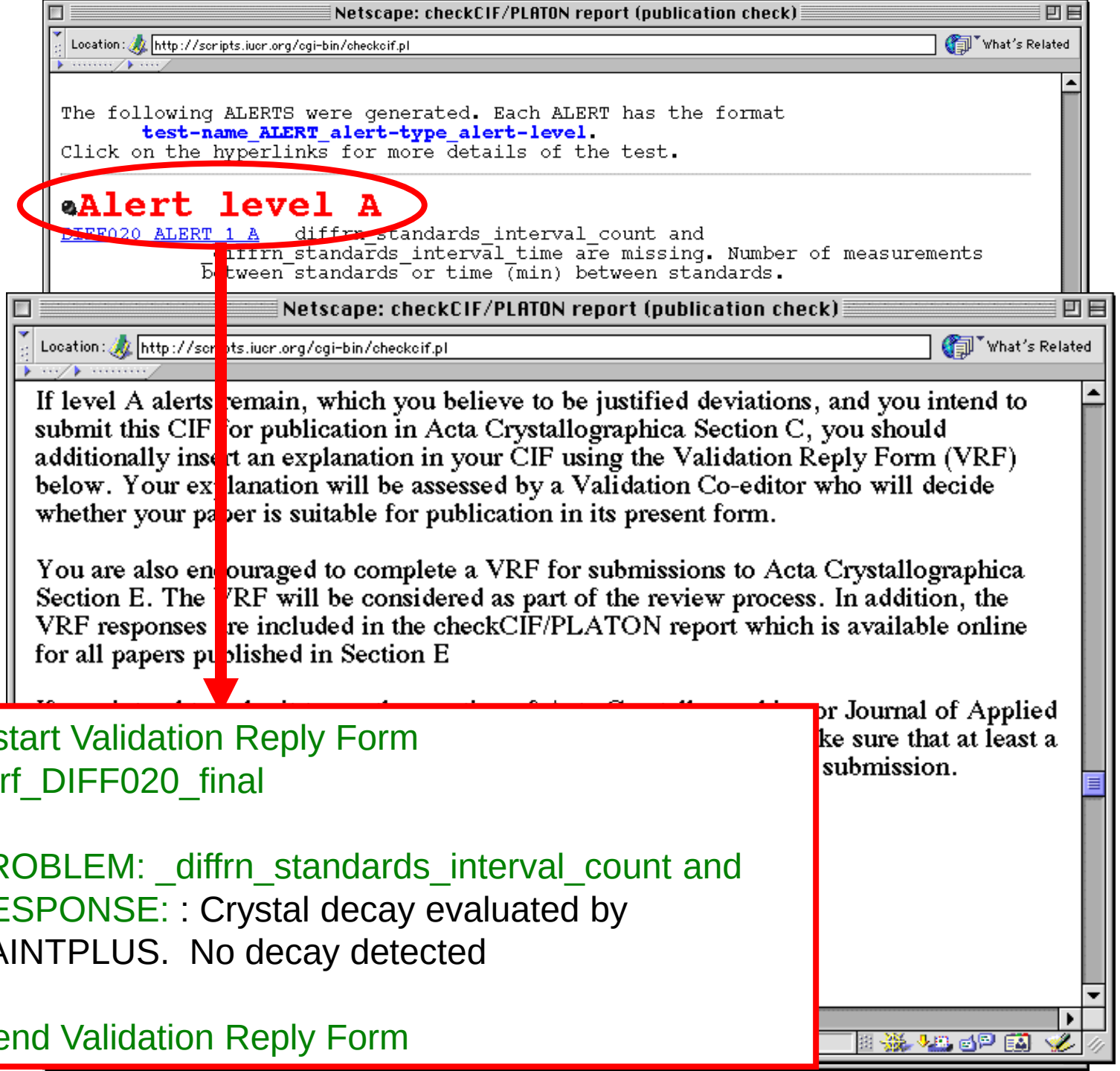


How to treat validation reports

- Procedure is straightforward
  - Give ALL alerts due consideration
  - Appreciate validation criteria
  - Criteria are based on normally expected results from routine analyses
  - If not an oversight, why is your structure not routine?
- Benefits
  - Significantly reduces errors in results
  - Improves efficiency in the publication process

Still getting A (or B) alerts?

- Is there a sound scientific basis for the outlier?
- Insert Validation Response Form (VRF) into CIF
- Use a brief, considered response if outlier justified
- Avoid casual or circular responses
- Show you understand the causes of the outlier
- Explain why it is a true feature of the analysis
- Also use _publ_exptl_refinement, _exptl_special_details or _refine_special_details


Netscape: checkCIF/PLATON report (publication check)
Location: http://scripts.iucr.org/cgi-bin/checkcif.pl
What's Related
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.
Netscape: checkCIF/PLATON report (publication check)
Location: http://scripts.iucr.org/cgi-bin/checkcif.pl
What's Related
If level A alerts remain, which you believe to be justified deviations, and you intend to submit this CIF for publication in Acta Crystallographica Section C, you should additionally insert an explanation in your CIF using the Validation Reply Form (VRF) below. Your explanation will be assessed by a Validation Co-editor who will decide whether your paper is suitable for publication in its present form.
You are also encouraged to complete a VRF for submissions to Acta Crystallographica Section E. The VRF will be considered as part of the review process. In addition, the VRF responses are included in the checkCIF/PLATON report which is available online for all papers published in Section E
or Journal of Applied
ke sure that at least a
submission.
start Validation Reply Form
_vrf_DIFF020_final
;
PROBLEM: _diffrn_standards_interval_count and
RESPONSE: : Crystal decay evaluated by
SAINTPLUS. No decay detected
;
end Validation Reply Form

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.

Author Response: Crystal decay evaluated by SAINTPLUS. No decay detected.

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. # C54 H52 Ag2 P4	1	
PLAT790_ALERT_4_C	Centre of Gravity not Within Unit Cell: Resd. # B F4	2	

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

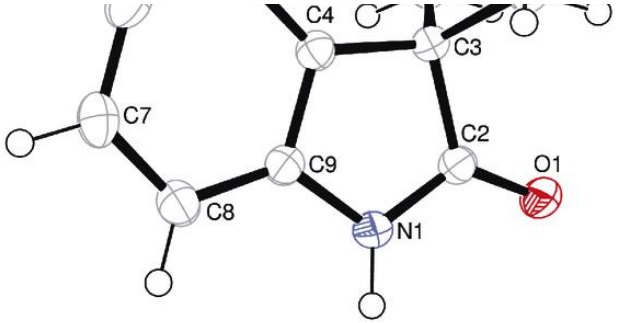
Possible limits to validation

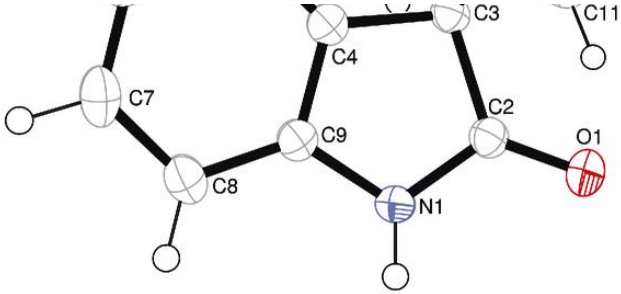
- Test not (yet) implemented
- Test not practical
- Error not a validation issue
- Mistake cannot be detected from data in CIF
- Nonsense entries in the CIF

Vigilance – additional to validation

- Does the structure make sense to you?
- Does the structure look right and is it geometrically logical?
- Must be able to rationalise 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)

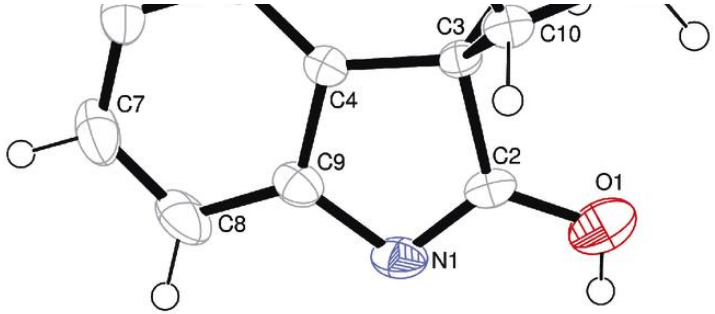
Misassigned element

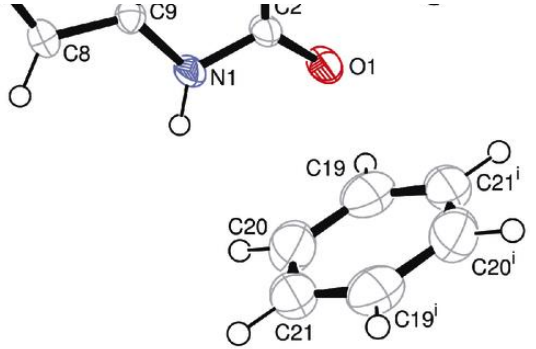




Four related lactams. One is a “rarely seen imidic acid tautomer”

R = 0.059, wR2 = 0.177, S = 1.067

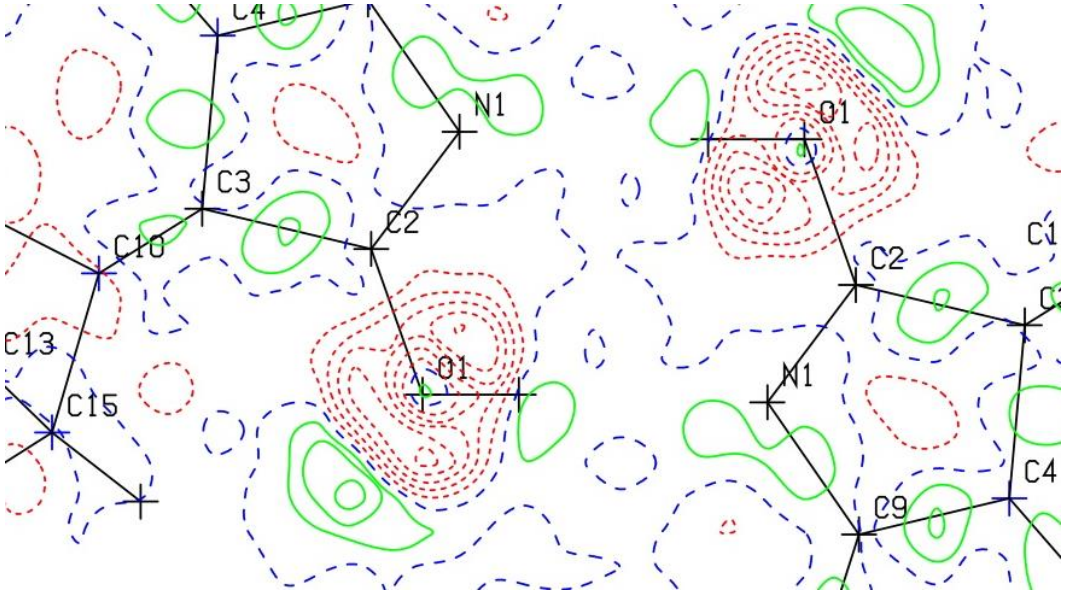




230_ALERT_2_B Hirshfeld Test Diff for O1 -- C2 .. 11.83 su

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



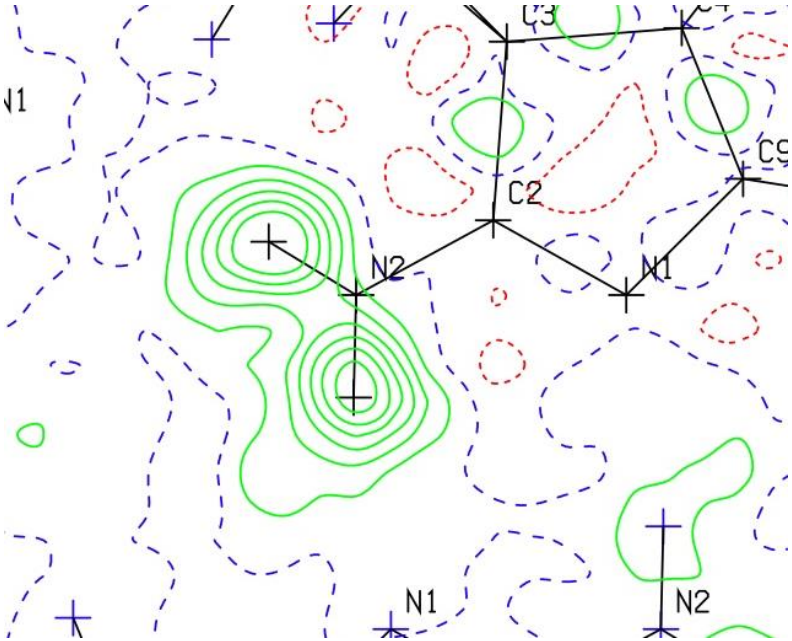
Contoured difference maps are very useful – easy in *PLATON*

Refine as an amine

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

No relevant alerts

```
Q1  0.22      0.77  C3
```



Now the chemist has work to do!

Consistency with known chemistry & geometry – missing H atom

The issue raises only a G alert

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

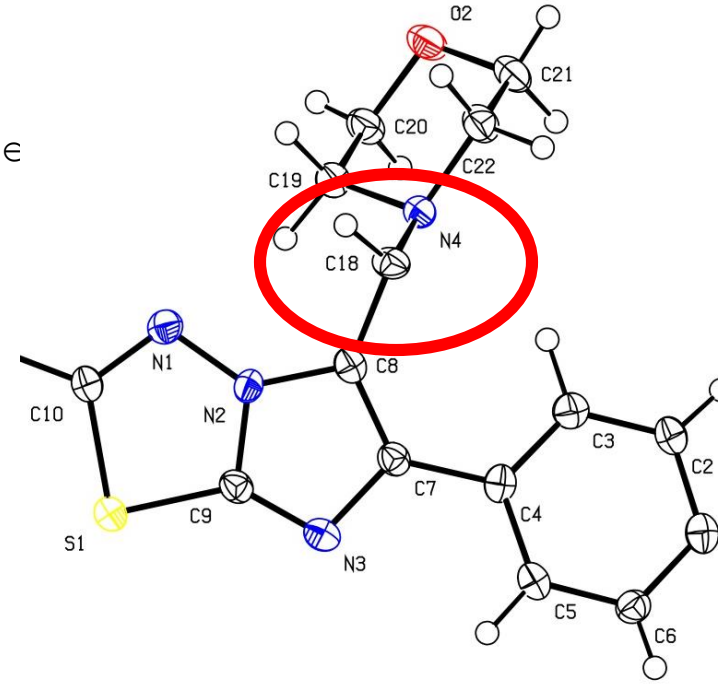
Largest peak: 0.84 e/Å^3

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

Structure Factor Validation

What can fcf validation detect?

- Mismatch between the data block names in the CIF and .fcf file
- Mismatch between cell parameters in the CIF and .fcf file
- The .fcf file is not from the refinement that produced the CIF
- Incomplete updating of a CIF (e.g. weighting scheme)
- Overlooked twinning
- Atomic coordinates transformed, but not the U^{ij}
- Incorrect element assignment (supplements other tests)
- Element reassignment without re-refining
- Modifying atomic and displacement parameters in the CIF (cheating!)

The .fcf file is not from the same refinement as the CIF

A water molecule was omitted from the refinement used to generate the .fcf file, but the finished model is in the CIF

```
Reported   Rho(min) = -0.34, Rho(max) =  0.36 e/Ang**3 (From CIF)
Calculated Rho(min) = -1.18, Rho(max) = 10.08 e/Ang**3 (From CIF+FCF data)
w=1/[sigma**2(Fo**2)+(0.0393P)**2+  0.0941P], P=(Fo**2+2*Fc**2)/3

R= 0.1442(  1215), wR2= 0.2787(  1385), S = 4.255      (From CIF+FCF data)
R= 0.2189(  1215), wR2= 0.5046(  1385), S = 7.612      (From FCF data only)
R= 0.0329(  1215), wR2= 0.0800(  1385), S = 1.081, Npar= 126

973_ALERT_2_A Large Calcd. Positive Residual Density on     V1    10.08 eA-3
971_ALERT_2_B Large Calcd. Non-Metal Positive Residual Density     3.14 eA-3
921_ALERT_1_A R1  * 100.0 in the CIF and FCF Differ by .......     -18.60
922_ALERT_1_A wR2 * 100.0 in the CIF and FCF Differ by .......     -42.46
923_ALERT_1_A S    values in the CIF and FCF Differ by .......      -6.53
925_ALERT_1_A The Reported and Calculated Rho(max) Differ by .     9.72 eA-3
926_ALERT_1_A Reported and Calculated   R1 * 100.0 Differ by .     -11.13
927_ALERT_1_A Reported and Calculated  wR2 * 100.0 Differ by .     -19.87
928_ALERT_1_A Reported and Calculated    S value   Differ by .      -3.17
```

Improper editing of a CIF

- Atomic coordinates transformed through a symmetry operation other than inversion, but not the U^{ij}

  – always re-refine and generate a new CIF, avoid piecemeal cut/paste or hand-editing of the CIF itself.

- Element reassignment without re-refining

- Modifying atomic and displacement parameters in the CIF (to hide things)

Such manipulations lead to mismatches of R-factors, goodness-of-fit and residual electron density.

CIF is only as good as its user acceptance

- CIF needs to be practical and transparent for users
- *“CIF is too hard to understand / use / edit / work with”*
- *publCIF* and *enCIFer* are very useful tools, but some authors do not want to have to learn yet another program !!
- Authors want to / are capable of using Word and only Word…
- Is the average person dealing with structures less computer savvy than 25 years ago?
- After nearly 20 years, the text parts of Acta Cryst. C papers can once again be submitted as Word documents

Proliferation of non-standard and undocumented data-names

```
_shelx_estimated_absorpt_T_min
_shelx_estimated_absorpt_T_max
_shelx_res_file
_shelx_res_checksum
_shelx_hkl_file
loop_
_cell_oxdiff_twin_id
_cell_oxdiff_twin_matrix_11
_cell_oxdiff_twin_matrix_12
_cell_oxdiff_twin_matrix_13
_cell_oxdiff_twin_matrix_21
_cell_oxdiff_twin_matrix_22
_cell_oxdiff_twin_matrix_23
_cell_oxdiff_twin_matrix_31
_cell_oxdiff_twin_matrix_32
_cell_oxdiff_twin_matrix_33
1  1.0000  0.0000 0.0000 0.0000  1.0000  0.0000 0.0000  0.0000 1.000
2 -0.9999 -0.0001 0.0002 0.0023 -1.0005 -0.0017 0.6363 -0.0011 0.995
```

Responsibilities of COMCIFS?

- Actively encourage software developers to adopt CIF or remain with standard CIF dictionary items
- Maintenance of the CIF standard must be ongoing
- CIF items and their definitions need to be kept up to date with developments in the field (e.g. twinning, *SQUEEZE*). Proper validation without items for now frequently used procedures is difficult.
- Existing CIF definitions for small molecules need a careful overhaul after 20 years good service (e.g. “absorption” items)
- Timely inclusion of new items that cater to 98% of cases is better than extended discussions over the last 2%

Summary

- *CheckCIF* is a tool for authors, practitioners and reviewers
- Be vigilant – do not rely solely on *checkCIF*
- Structure factor validation is also very important
- We couldn't do this without a data interchange standard – CIF

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?

checkCIF development: Ton Spek (Utrecht) & Mike Hoyland (IUCr Chester office)