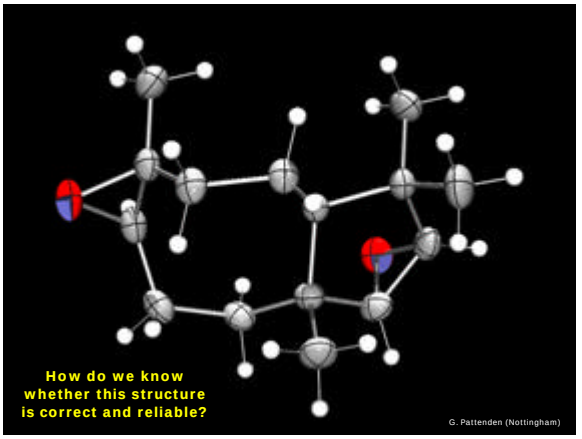
Validation and checking of crystal structures

_refine_ls_number_reflns	6811
_refine_ls_number_parameters	261
_refine_ls_number_restraints	0
_refine_ls_R_factor_all	0.0455
_refine_ls_R_factor_gt	0.0284
_refine_ls_wR_factor_ref	0.0560
_refine_ls_wR_factor_gt	0.0537
_refine_ls_goodness_of_fit_ref	0.923
_refine_ls_restrained_S_all	0.923
_refine_ls_shift/su_max	0.002
_refine_ls_shift/su_mean	0.000



OUTLINE

- Overview validation and checking
- Validation for Acta C, etc.
- Validation for other journals
- The limits of validation

Validation involves comparison against a set of test criteria

- Do cell volume and cell 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

Checking is additional to validation

- Does the structure make sense to you?
- Does the structure look right?
- Do chemically equivalent bonds agree?
- Are all CIF entries complete and correct?

Automated data validation with checkCIF or PLATON

- Checks for
  - CIF construction and syntax errors
  - missing information
  - parameters outside expected norms
  - conformation with convention

ALERT LEVELS

A Serious – attention essential

Item omitted or large deviation from norm

Alert A No crystal dimensions have been given

Alert A Ratio of Tmax / Tmin expected is > 1.30
An absorption correction is required.

Alert A Atom C58A ADP max / min Ratio 18.00

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 < 1.9s
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
- an extensive list may indicate problems

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 neighbors: C1
Alert C D-H without acceptor N2–H2 ?

C1 and N2 should be N and C, respectively

A / B / C indicate the seriousness of the problem

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

ALERT Type 2:
Indicator that the structure model may be wrong / deficient

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

ALERT Type 4:
Cosmetic improvement, query or suggestion

Not all combinations are logical, for example 4 A

Sources of outlier parameters

- Unresolved feature (e.g., untreated disorder)
- Artefact due to limited data quality
- Inadequate procedures (e.g., poor corrections)
- Incorrect structure (e.g., wrong space group)
- A genuinely unusual observation!!

What does validation software do?

- Identifies possible problems *via* ALERTs
- Provides explanations of ALERTs
- Suggests interpretations and possible solutions

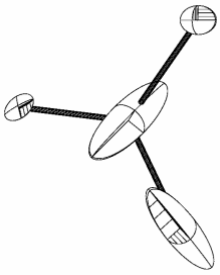
Not just for authors

- referees use it for assessment
- authors need to be aware of this
- how appropriate are IUCr criteria?

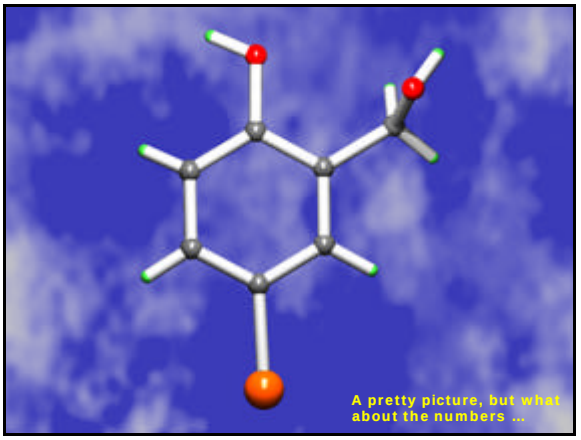
When to validate?

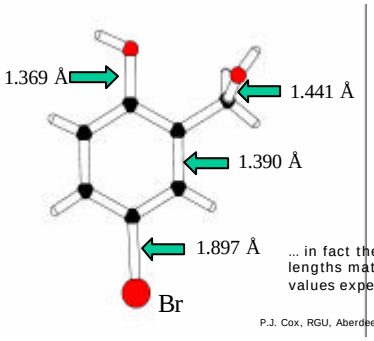
- software for data collection, refinement, etc
  - *should* do its own validation
- use PLATON in final stages of determination
- validate raw CIF from the refinement program
- must validate the *final* version as well
- avoids problems at submission, refereeing, etc

Looking at the structure

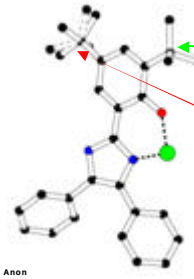


A visual examination can often be revealing: here there are some extreme ellipsoids which are also incompatible with a rigid bond model





Less satisfactory



Ordered *t*-butyl group has all C-C distances around 1.52 Å

Within the disordered group the range is 1.49 to 1.60 Å

Need (better) restraints?

VALIDATION / CHECKING PROCEDURE

1. Check the CIF from refinement using PLATON
2. Augment CIF using e.g. XCIF and enCIFer
3. Re-check the CIF using PLATON or checkCIF
4. Look at ellipsoid plots from several directions
5. Check bond lengths are sensible and consistent
6. After any changes, re-check the CIF

Validation and IUCr Journals

Early 1990's - CIF introduced

- allows automatic creation of tables
- enables full electronic submission / processing
- increases efficiency, faster publication times
- automates many editorial tasks
- improves appearance of the journal

→ permits automated validation ←

Automation of syntax and data checks

- authors get instant, anonymous feedback
- can detect and fix problems before submission
- fewer, shorter revision cycles
- consistent application of acceptance criteria
- editors / referees can focus on science

- RESULT: faster publication times

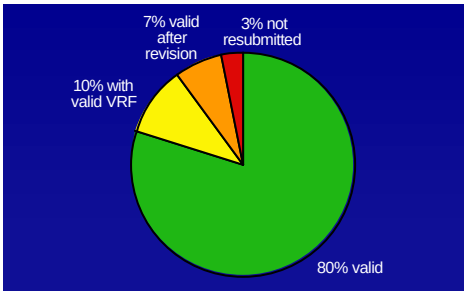
How to get a CIF through

- Give ALL Alerts due consideration
  - appreciate validation criteria
  - criteria are based on normally expected results from routine analyses
  - Why, then, is your structure not routine?
- In any VRF...
  - avoid casual or circular responses
  - show you understand the causes of the outlier
  - explain why it is a true feature of the analysis

Common problems ...

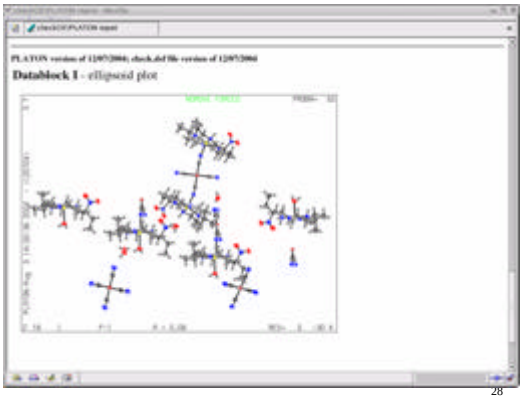
- ✖ Data completeness or resolution too low
- ✖ Maltreatment of H atoms
- ✖ Structure not at convergence
- ✖ Missing or inadequate absorption correction
- ✖ Indications of a poor structure

All Acta C submissions in 2000



checkCIF in 2005

- **the new home of checkCIF:** ***http://checkcif.iucr.org***
- **service sponsored by ACS, CCDC and Elsevier**
- **an ORTEP plot is now included**
- **part of new Acta C/E submission procedures**
- **online upload of all material for Acta C and E papers (CIF + figures/schemes/structure factors)**
- **coming in the next 2-3 months**



Validation and other Journals

Standards, procedures vary widely

- **some journals perform extensive checks**
- **some do only very basic checks**
- **some do none at all**

? so what do authors do ?

Perform your own validation

- ✓ **ensure there are no serious mistakes**
- ✓ **ensure the quality is adequate**
- ✓ **submit a copy of the checking report**

Example – an ACS procedure

- Authors submit the CIF along with the paper
- CIF must contain author names and paper title
- authors must have checked the CIF first
  - the check report may be requested
- reviewers have Web access to the CIF, along with the manuscript and any supplementary data

Other procedures

- Submit paper to journal
  get a code for the paper
  submit CIF under this code
- Submit CIF to CCDC or ICSD
  get deposition number
  include number in paper

If you have a "difficult" structure

- Identify and describe the problem
- Give details of the remedial action taken
- Describe the (successful?) outcome

Where and how?

1. Briefly in any experimental footnote
2. At the top of the CIF
   use _refine_special_details
3. In any other Supplementary Data

Example of text

_refine_special_details
; Disorder was identified in one of the tetrafluoroborate anions. All the F atoms were affected and two orientations were identified.

Similarity restraints were applied to B-F distances, and to F-B-F angles. All F atoms were refined isotropically. The occupancies of each group of four partially-occupied F atoms were refined competitively using a free variable. Each F atom was found to be disordered over two equally occupied sites, as shown by the final group occupancies of 0.506(12) and 0.494(12).

In the final model the range of B-F distances was 1.31(2)-1.42(2) Angstroms and the F-B-F angles spanned 105(2)-112(2) degrees. No difference Fourier peak in the region exceeds 0.6 e/A**3.
;

The limits of validation

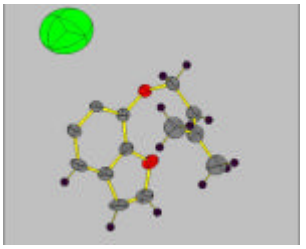
(automatic validation will not catch every problem)

Possible limits to validation

- × test not (yet) implemented
- × test not practical
- × error not a validation issue
- × error cannot be detected from data in CIF
- × nonsense entries in the CIF

Test not implemented

Example: High ADPs on isolated atom

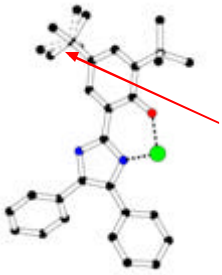


Not detected by

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

Atom is probably O rather than Cl

Test not practical



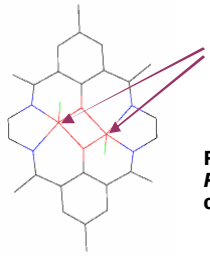
C-C range is 1.49 to 1.60 Å

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

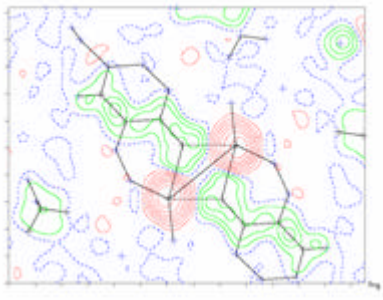
Error not detectable from CIF data



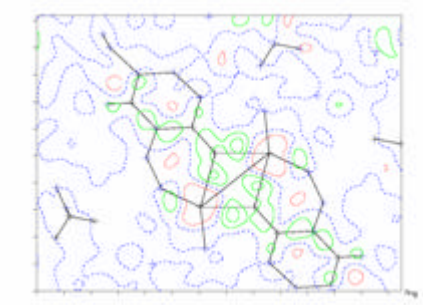
Prior chemical information: Complex is either Ru/Ru or Ru/Zn

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

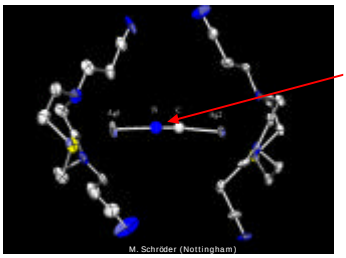
Difference map with Ru/Ru model (R_1 = 0.064)



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



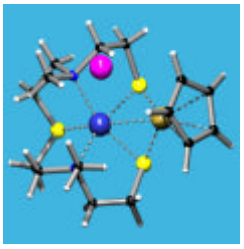
Other examples



Ag-CN-Ag link,
Ag-NC-Ag link
or disordered?

could only be resolved using D*F maps*

Model complexes for [NiFe] hydrogenase



need good data to distinguish Ni and Fe by refinement

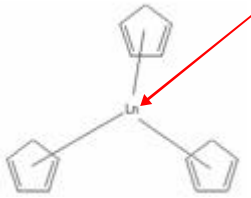
Ni (Z = 28) *vs* Fe (Z = 26)

DZ/Z = 2/28

$R_1 < 4\,\Delta Z$?

Use D*F* maps to confirm and when data are poor

Lanthanide complexes



Ln = Er, Tm or Yb ?

Z = 68, 69, 70

Similar co-ordination

Similar geometry parameters

Crystallography is not much good at distinguishing these metals.

Nonsense entries in the CIF*

```
_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'
```

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

Unsuitable SHELX(T)L-97 defaults ?

- space group notation
- diffractometer
- T = 293 K ?
- absorption correction
- total data collected
- index limits

- R_{int}
- H atom treatment
- weighting scheme
- precision
- structure solution