The Crystallographic Information File (CIF) Description and Usage

Ton Spek,
Bijvoet Center for
Biomolecular Research
Utrecht University
Leiden, 27-Jan.-2009.

Overview of this Lecture

- An Overview of the Status of Single Crystal X-Ray Structure Determination
- The What, Why and How of CIF
- CIF Usage and Applications

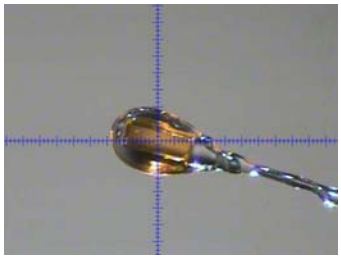
A Single Crystal X-Ray Study Involves

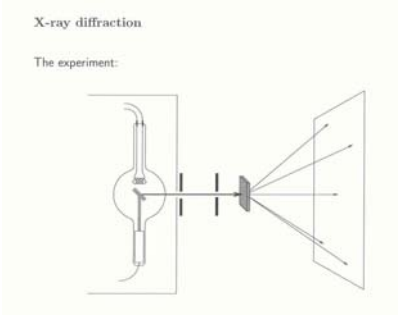
- A Structure Query (What did I make?)
- Needed: A Single Crystal (0.2 mm)
- Collection of X-Ray Diffraction Data
- Solution of the Phase Problem
  (To get a Preliminary Model)
- Structure Model Parameter Refinement
- Interpretation of the Result/Geometry Analysis
- Validation of the Analysis
- Report and (Co)Publication

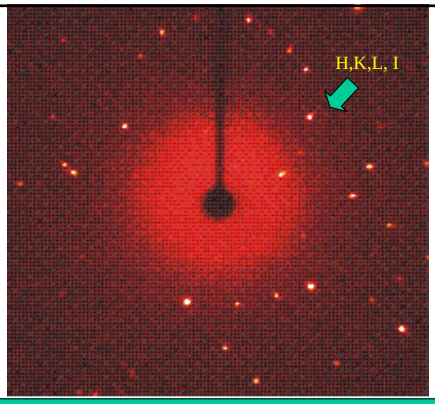
Crystal Requirements

- Preferably a single crystal
  (Sharp Extinctions under Polarized Light)
- Block rather than slim long needle
- Fresh from mother liquor
- Unstable crystals covered with inert oil
  Data collection under cold N_2 stream.
- Twins and split crystals possible but best avoided when possible.

Example of an Oil-mounted Crystal



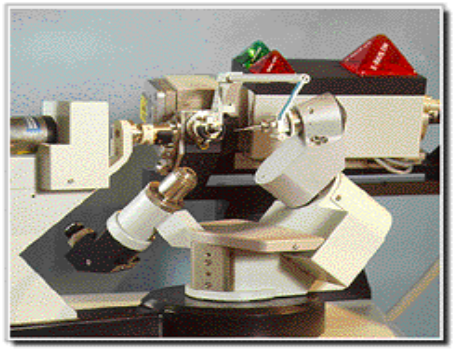




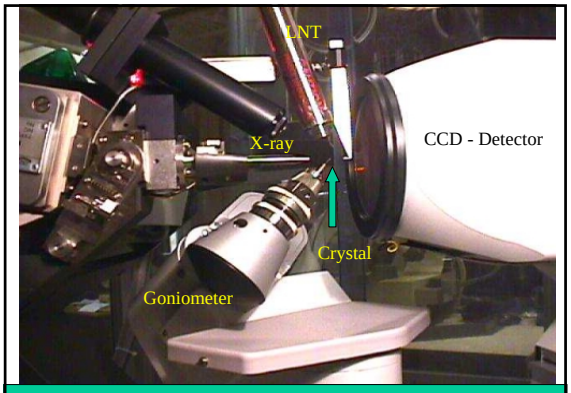
Diffraction Pattern of a crystal rotated over 1 degree

Short History of Data Collection

- 1912 – von Laue et al. Experiment – X-ray Film
- 1913 – Bragg – Diffractometer + Detector
- Tot ~ 1965 Film (Weissenberg, etc.)
- ~1960-1995 – Serial Detector Diffractometer
  ~ 50 datasets / year
- 1995 – present – Image plate/ CCD Detectors
  a few hours Collection Time for a Routine Structure
- ~2007 – Digital Detectors (AXIOM, Pilatus etc.)
  new options: shutterless, low noise etc.



X-Ray source, Goniometer & Serial Detector



X-ray source, goniometer + crystal, N2-cooling and CCD Detector

Central Formulae

- Diffraction Spots: $\mathbf{2d_{hkl}\ \sin\theta_{hkl} = n\lambda}$
- Electron Density Map (3D Fourier Map)
  $\mathbf{\rho_{x,y,z} = 1/V\ \Sigma_{hkl}\ F_{hkl}\ exp\{-2\pi i(hx + ky + lz)\}}$
- Structure Factor (Model)
  $\mathbf{F_{hkl}(calc) = \Sigma_j\ f_j\ T_{j,hkl}\ exp\{2\pi i(hx_j+ky_j+lz_j)\}}$
- Least Squares Model Refinement
  Minimize: $\mathbf{\Sigma_{hkl}\ [w_{hkl}(|F_{hkl}(obs)|^2-|F_{hkl}(calc)|^2)]^2}$
- Convergence Criteria: **R1, wR2, S**

Solution of the Phase Problem

- Early Method: Trial & Error (Salts such as NaCl, Silicates etc.)
- Patterson Methods (Heavy Atom)
- Direct Methods (SHELXS, DIRDIF, SIR)
- New: Charge Flipping (Ab-initio)

- Phase Problem Solved! Given reasonable data.

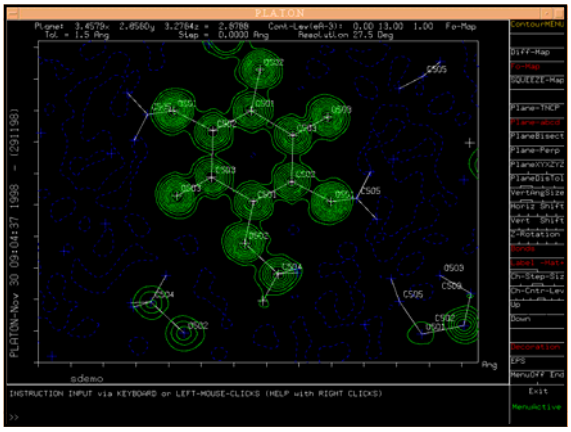
3D-Fourier Map

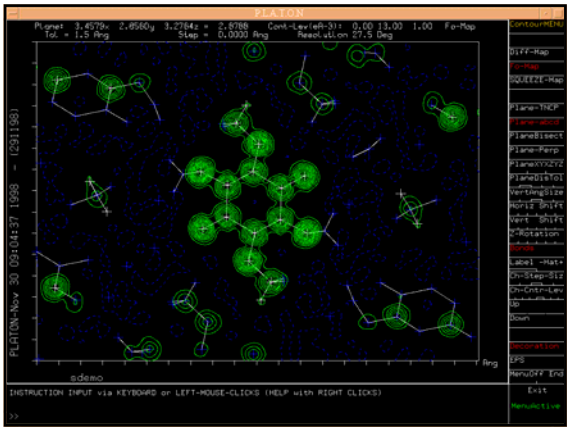
- Given the Diffraction Data and (Approximate) Phases a 3D Electron Density Map can be Calculated.

$\rho_{x,y,z} = 1/V \Sigma_{hkl} F_{hkl} \exp\{-2\pi i(hx + ky + lz)\}$

$F_{hkl} = |F_{hkl}|\exp(\phi_{hkl})$

- Following is a section through such a map





Interpretation in Terms of Atoms

- Position of highest density => Position x,y,z
- Deviation of the density shape from the ideal atomic electron density => Thermal motion parameters:
- Isotropic: U(iso) or
- Anisotropic: U^{11},U^{12},U^{13},U^{22},U^{23},U^{33}

(Displacement Parameters) => ORTEP

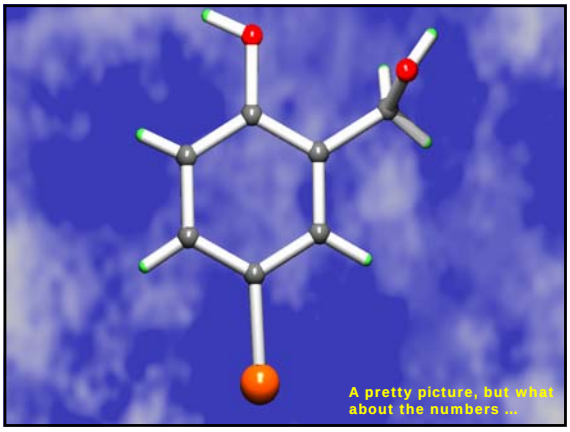
Note: ORTEP does NOT represent the electron distribution.

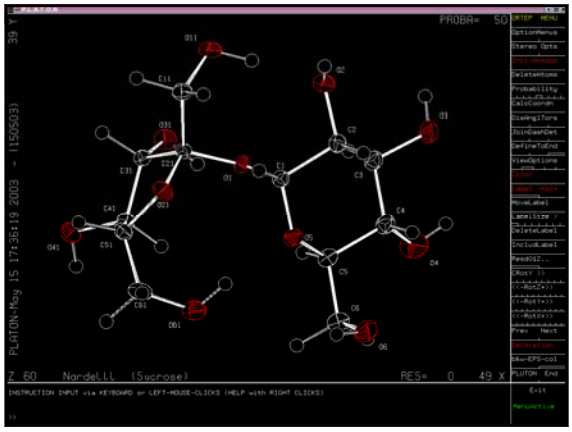
Interpretation in Terms of Bonds

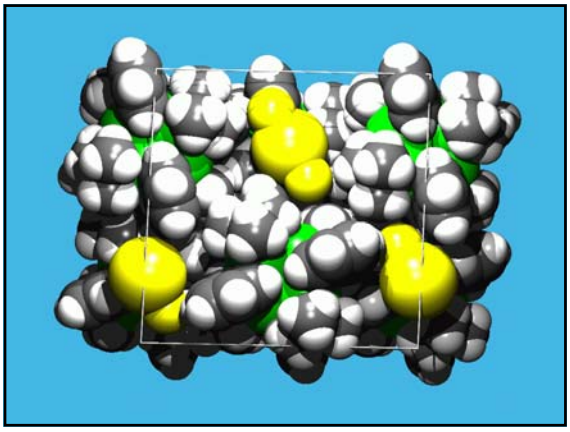
- Bonds between atoms of type A and B are assigned on the basis of atomic covalent radii with: d < R(A) + R(B) + 0.4
- 'Crystallographic Bonds' are not necessarily Chemical Bonds.
- Van der Waals Radii are used to detect isolated molecular species or short contacts.

Display Options

- **Ball-and-Stick**
  Simpel but may hide problems with a structure.
- **ORTEP**
  Often preferred because it visualizes most model parameters and possible problems.
- **CPK**
  Spacefilling PLOT illustrating the shape etc







Parameter Model Refinement

- Translate the 3D electron density in terms of position and temperature parameters
- Non- Lineair Least Squares Method
  Min: $\Sigma_{hkl}\,[w_{hkl}(|F_{hkl}(obs)|^2-|F_{hkl}(calc)|^2)]^2$
- Time consuming: disorder, weak data

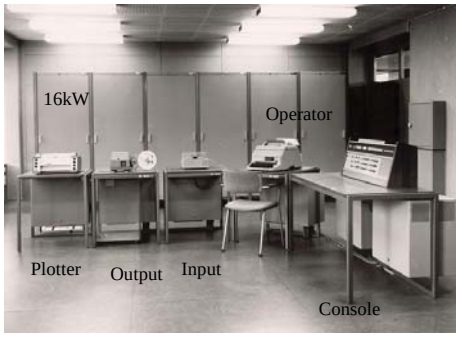
Analysis

Interpretation (geometry, intermolecular interactions etc.)

- Structure Validation
- (Co)Publikation
- Results to Cambridge Crystallographic Database CSD (~500000)

Computing in the Past

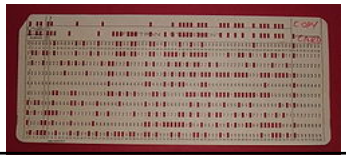
- University Mainframe
- Mainly Batch Mode (Long Waits)
- Storage on Magnetic Tape
- Piles of Lineprinter Output
- Primitive Computer Graphics
- No Internet



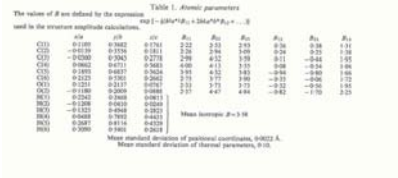
~1966, Electrologica X8 ALGOL60 'Mainframe' (<1MHz)

Flexowriter for the creation and editing of programs and data

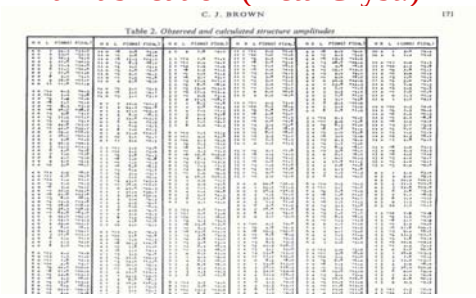
Data Storage in the Past



Archival of Model Parameters in a Publication (Acta Cryst.)



Archival of Reflection Data in a Publication (Acta Cryst.)



Computing Today

- Personal Workstation (MS-Windows or Linux)
- No queues for Computing Facilities
- Good graphics & Hardcopy
- Unlimited Disc Storage
- Internet (exchange, information)
- Automation

Announced Aug 2007: Tabletop 'Black Box' – Smart X2S

Mount crystal and load

Structure ?

Problems Around 1990

- Multiple Data Storage Types
- No Standard Computer Readable Format
- Data Entry of Published data via Retyping.
- No easy numerical checking for referees etc.
- CSD Database Archival by Retyping from the paper
- Multiple typo's in Published Data

Solution

- CIF-Standard Proposal
- S.R. Hall, F.H. Allen, I.D. Brown (1991). Acta Cryst. A47, 655-685.
- Pioneered and Adopted by the International Union for Crystallography
- Adopted by the author of the most used software package SHELXL (G.M.Sheldrick)

What is CIF ?

From: http://ww1.iucr.org/cif/index.html

CIF

The acronym CIF is used both for the *Crystallographic Information File*, the data exchange standard file format of Hall, Allen & Brown (1991) (see Documentation), and for the *Crystallographic Information Framework*, a broader system of exchange protocols based on data dictionaries and relational rules expressible in different machine-readable manifestations, including, but not restricted to, Crystallographic Information File and XML.

Official Entry Point for Definition and Details

Specifications

Formal specification of the Crystallographic Information File format

- Version 1.1 overview
- File syntax
- Common semantic features

Ancillary notes

- Placeholders for unknown and inapplicable values
- Format of date/time strings
- Format of telephone/fax numbers
- Format of proper names
- Recommended abbreviations within data names
- Reserved prefixes for local dictionary extensions

Documentation on CIF

- A selection of papers and articles.

Software

- A selection of libraries, applications and utilities.

Dictionary Definition Languages

Attributes of canonical data descriptors in CIF are described by a machine-readable definition language (DDL).

- Simple Dictionary Definition Language (DDL1)
- Relational Dictionary Definition Language (DDL2) Enhanced methods-rich Dictionary Definition Language (DDL∞) (draft)

Current CIF Dictionaries

Canonical data descriptors and their attributes are presented in machine-readable data dictionaries.

Descriptors for small-molecule, inorganic and other small-unit-cell structures

- Core dictionary (coreCIF) [ASCII | HTML | More information]
- Powder dictionary (pdCIF) [ASCII | HTML | PDF | More information]
- Modulated and composite structures dictionary (mmCIF)

Practical Approach

- We ignore here the scary details that are not relevant in the current context
- We will Discuss the File structure
- We will look at its relevance for publication
- We will discuss software to edit and check the CIF file
- We will look at software that uses CIF as Input.

File Structure

- Both Computer and Human Readable Ascii encoded file
- Free Format
- Mostly 80 colums wide
- Parsable in units
- Data Order Flexible
- Dataname and Value associations

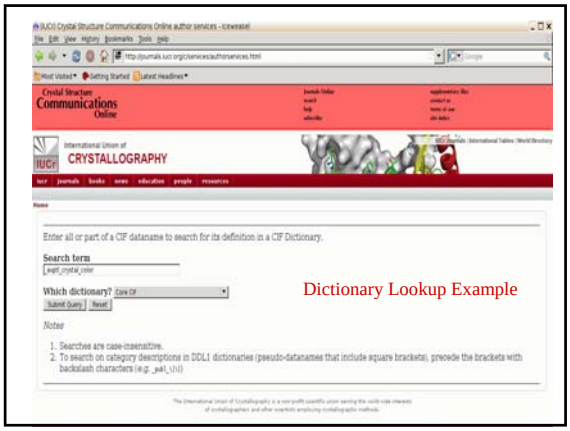
Constructs

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

Construct for Text

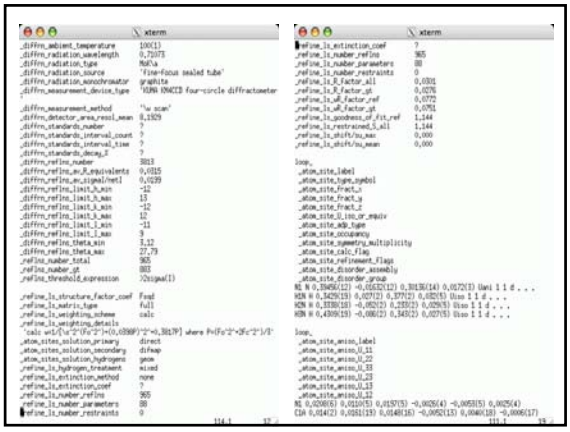
- Text can be included between semi-columns
- Used for Acta Cryst. Section C & E
- Example

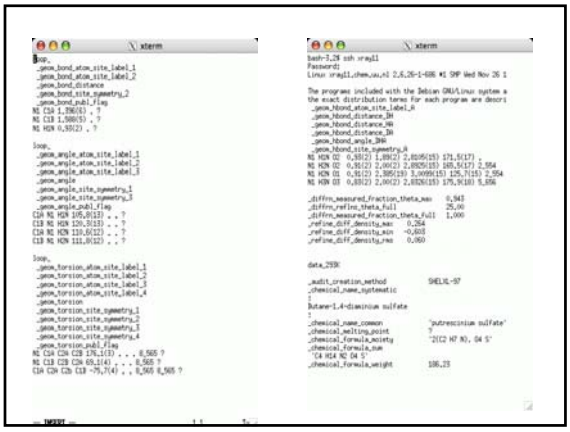
_publ_section_comment
;
This paper presents the first example
of a very important compound.
;



Dictionary Lookup Example

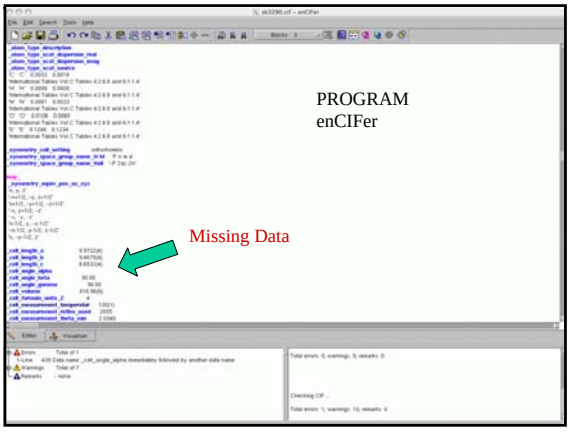
CIF Example File

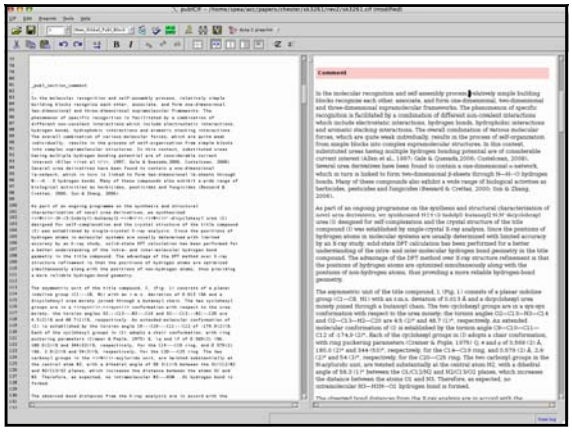




CIF Completion

- CIF Files are created by the refinement program (e.g. SHELXL)
- Missing Date can be added with a Text Editor, 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)



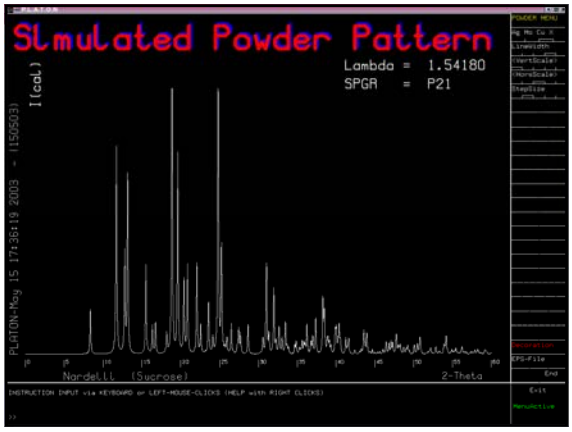


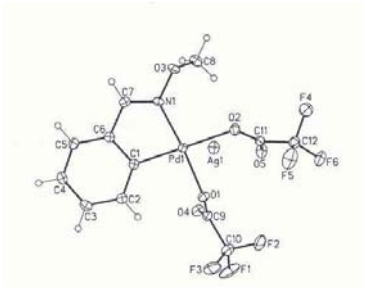
Note on Editing the CIF

- The Idea of editing the CIF is to add missing information to the CIF.
- Some Acta Cryst. authors have been found to polish away less nice numerical values. This leaves traces and is generally detected by the validation software and not good for the career of the culprit…

CIF Applications

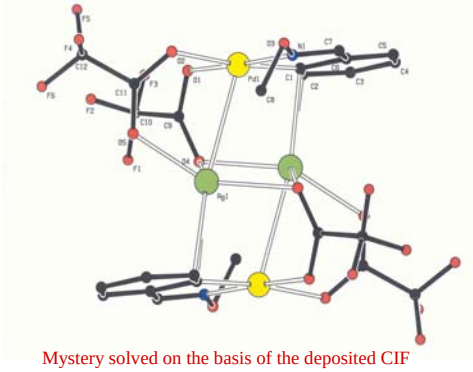
- Data Archival
- Deposition to the CSD (=> CSD number)
- Supplementary Material for Publication
- Input for Geometry and Graphics Software e,g. Mercury (from CCDC) and PLATON
- Standard Format for publications (Structure Communications) in Acta Cryst. Sections C & E.
- Structure Validation





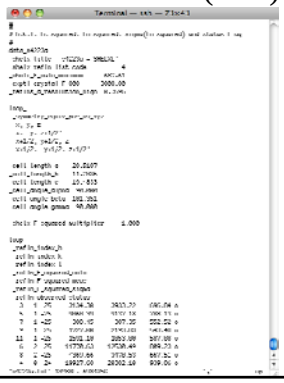
Naked Silver ?

Angew. Chem. 47 (2008) 9462



Mystery solved on the basis of the deposited CIF

Reflection CIF (FCF)



Calculations on Published Structures

- CIF data for a published structure can be obtained from the CCDC
- FCF Data are generally only retrievable from the IUCr website for Acta Cryst. Papers
- PLATON has a tool to re-create .ins and .hkl files for re-refinement with SHELXL
- Useful to investigate difference maps for more details.

Structure Validation

- Pioneered by the IUCr
- Currently most journals have implemented a validation scheme.
- Papers:
- A.L.Spek (2003). J. Appl. Cryst. 36, 7-13.
- A.L.Spek (2009). Acta Cryst. D65, 148-155.

Why Crystal Structure Validation

- The explosion of structure determinations
- An analyses of the nearly 500000 structures in the CSD learns that a significant number are in error
- Many analyses are done today by non-specialists
- Limited number of experts to detect pitfalls
- Validation provides a list of issues that need special attention of the analyst, specialist and referee.
- Validation sets quality standards.

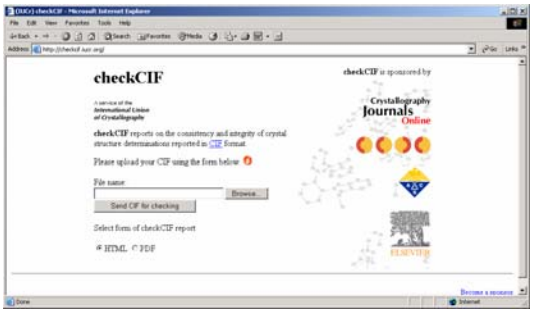
VALIDATION QUESTIONS

Single crystal 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?

How is Validation Implemented

- Computer readable structure analysis results in CIF format (Syd Hall & George Sheldrick)
- A file (Check.def) defines the issues that are tested with levels of severity and associated explanation and advise.
- The tests are executed by the program PLATON
- The tests can be executed both in-house or through the WEB-based IUCr CHECKCIF server.



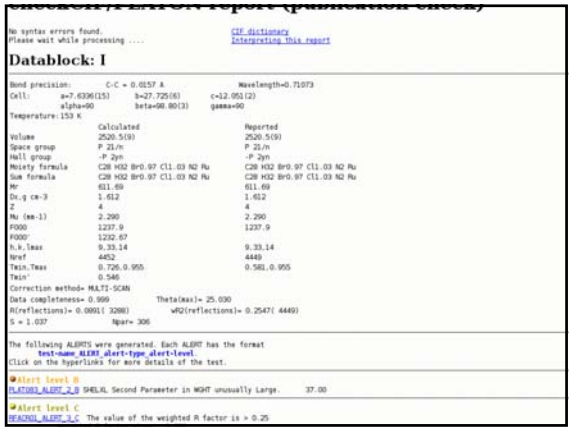
ALERT LEVELS

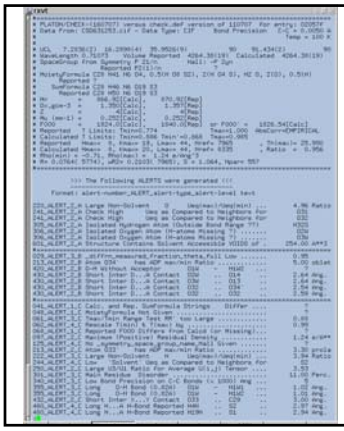
CheckCif Report in terms of a list of ALERTS

- ALERT A – Serious Problem
- ALERT B – Potentially Serious Problem
- ALERT C – Check & Explain
- ALERT G – Verify or Take Notice

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 - Cosmetic Improvements, Queries and
Suggestions.





EXAMPLE OF PLATON GENERATED ALERTS FOR A RECENT PAPER PUBLISHED IN J.Amer.Chem.Soc. (2007)

Attracted special attention in Chemical and Engineering News

(Referees obviously did not Bother)

Which Key Issues are Addressed

- Missed symmetry ("being Marshed")
- Wrong chemistry (Misassigned atom types)
- Too many, too few or misplaced H-atoms
- Missed solvent accessible voids in the structure
- Missed Twinning
- Absolute structure
- Data quality and completeness

FCF-VALIDATION

Forthcoming:

Automatic twinning detection as part of the IUCr CheckCif procedure

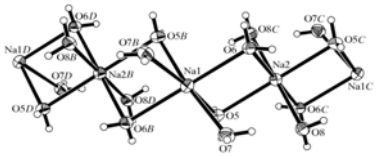
- Detection of ignored twinning
- Detection of Applied Twinning Correction without being reported

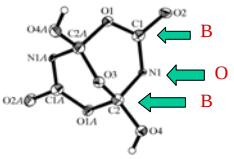
(Already available via PLATON/Check)

Examples

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

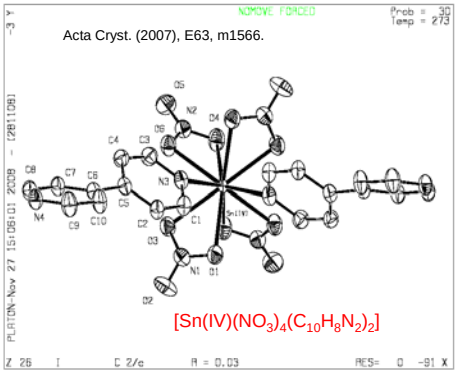
Published with Wrong Composition



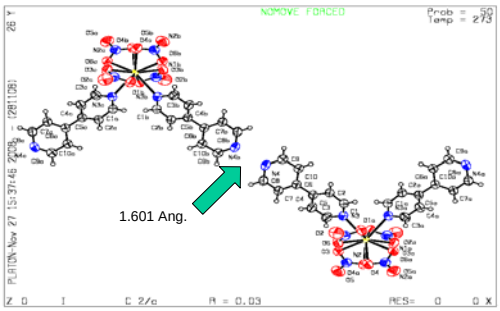


BORAX !

Acta Cryst. (2007), E63, m1566.



$[Sn(IV)(NO_3)_4(C_{10}H_8N_2)_2]$



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

THE STABLE PENTAMETHYLCYCLOPENTADIENYL CATION

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

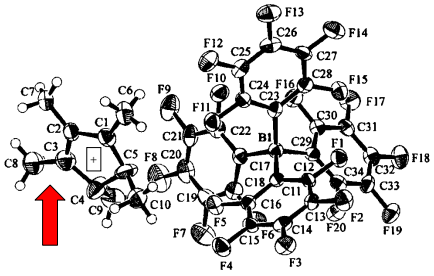
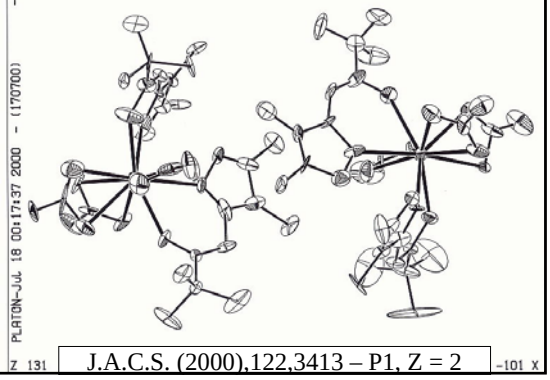


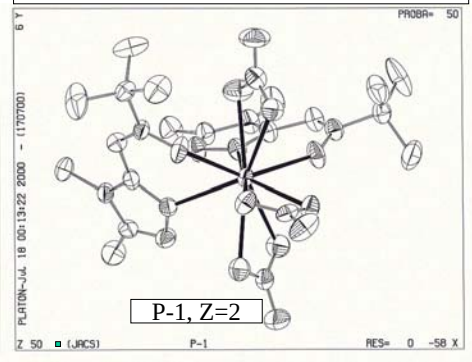
Figure 1. The crystal structure of pentamethylcyclopentadienyl tetrakis-(pentafluorophenyl)borate. There is no covalent bonding between the cation on the left and the anion on the right.

WRONG SPACEGROUP



J.A.C.S. (2000),122,3413 – P1, Z = 2

CORRECTLY REFINED STRUCTURE



P-1, Z=2

HOT STRUCTURE – FAST LANE PUBLICATION


CHEMICAL & Engineering News

SCIENCE & TECHNOLOGY

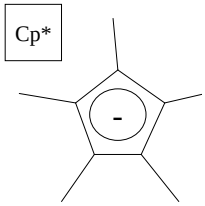
April 29, 2002

ELUSIVE CARBOCATION ISOLATED AS A SOLID

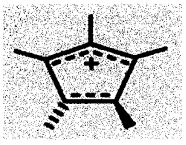
Pentamethylcyclopentadienyl cation is found to be a stable singlet with a distorted structure

Pentamethylcyclopentadienyl

Common Anionic Form



"New" Cationic Form



Assigned Bond types

Enthusiastic Last Paragraph of C&E-News

… The Northwestern chemists are now exploring the reactivity of C5Me5+. "You can sit down and write lots of interesting reactions on paper," Lambert says, and it will be interesting to see if the molecule reacts as expected. But Considering this cation's track record, it might be safer to expect more surprises. [!!]

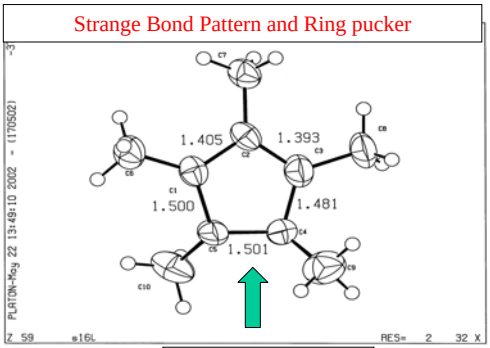
NOT SO HOT AFTER ALL !!

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).

Strange Bond Pattern and Ring pucker



Double Bond ?

D

PACKING EFFECTS ??

Corrected Structure

J.N. Jones et al.,Chem. Comm. 2002,1520-1521

Additional H-Atoms

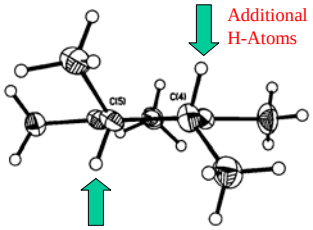


Fig. 2 Side-on view of the pentamethylcyclopentenyl cation showing the hydrogen atoms attached to C(4) and C(5) and the *trans* arrangement of CH_3 groups attached to these carbon atoms.

Concluding Remarks

- The CIF standard makes it possible to easily do follow-up calculations for published structures
- The available information is more complete
- http://www.cryst.chem.uu.nl for more information