

## 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

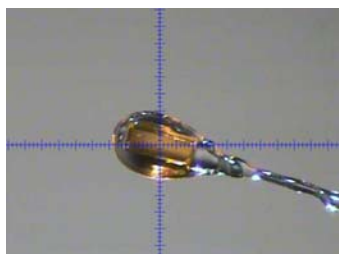
## A Single Crystal X-Ray Study Involves

- 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

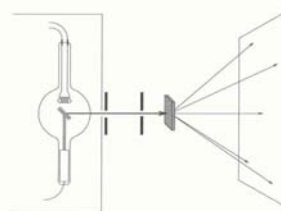
- Preferably a single crystal (Sharp Extinctions under Polarized Light)
- Block rather than slim long needle
- Fresh from mother liquor
- Unstable crystals covered with inert oil  
Data collection under cold N<sub>2</sub> stream.
- Twins and split crystals possible but best avoided when possible.

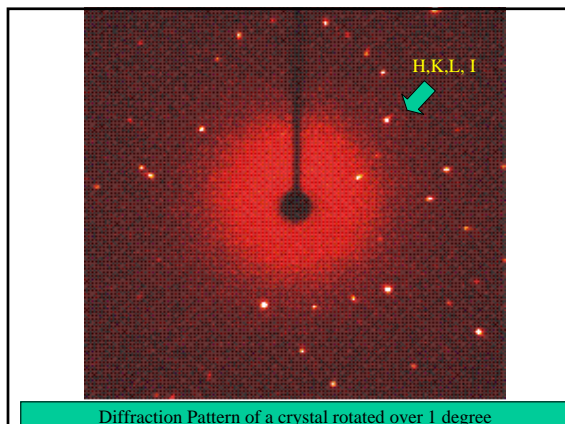
## Example of an Oil-mounted Crystal



X-ray diffraction

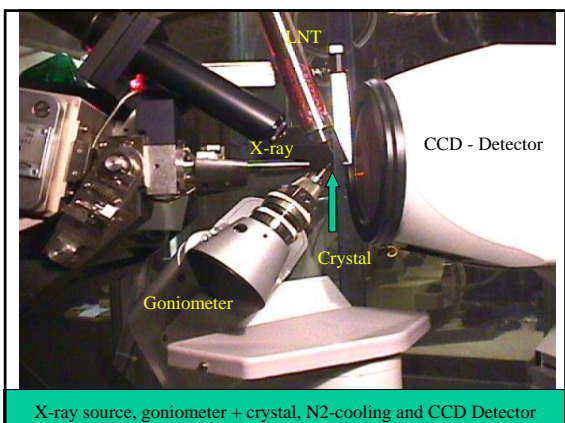
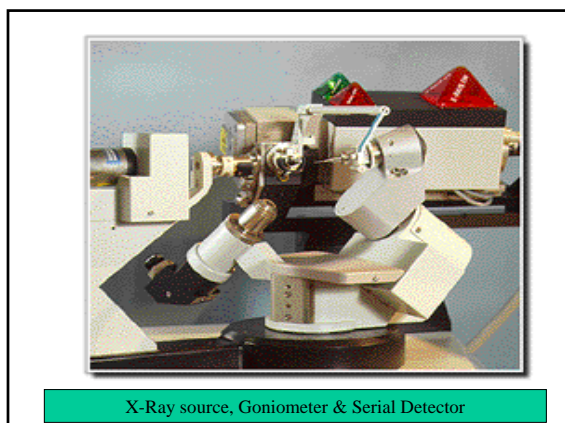
The experiment:





### 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.



### Central Formulae

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

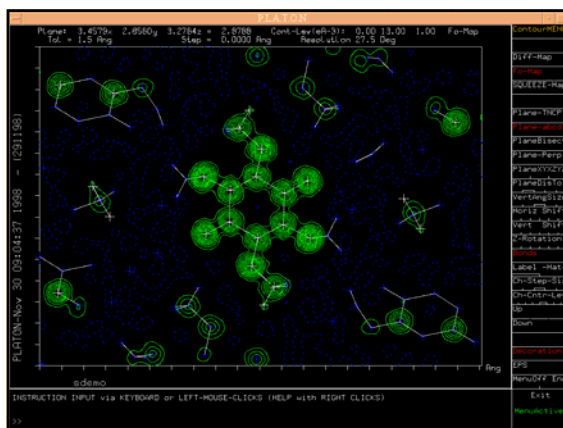
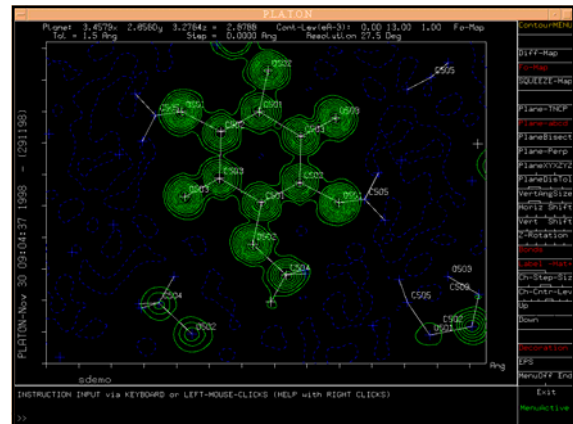
## 3D-Fourier Map

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

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

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

- Following is a section through such a map



## Interpretation in Terms of Atoms

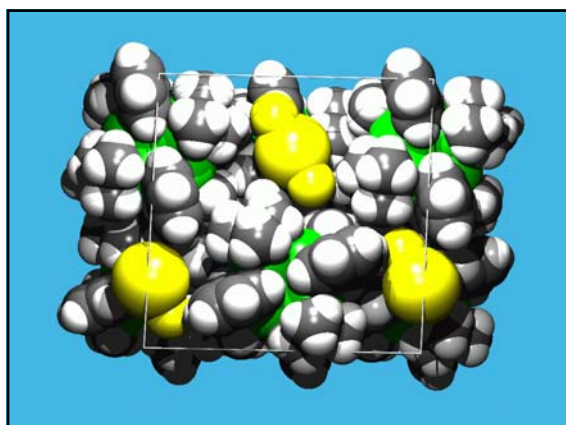
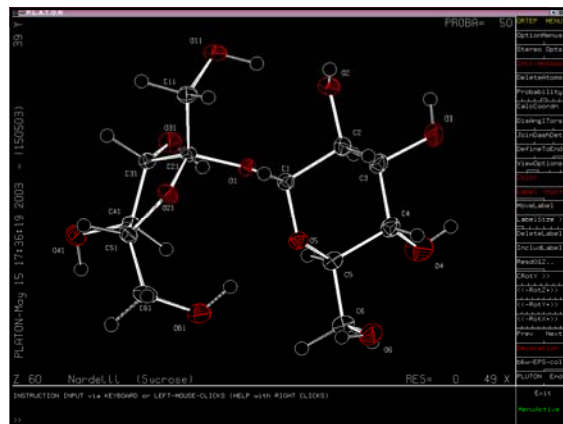
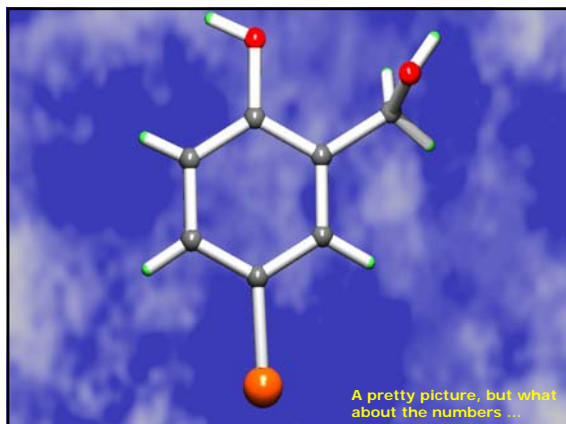
- Position of highest density => Position x,y,z
- Deviation of the density shape from the ideal atomic electron density => Thermal motion parameters:
  - Isotropic:  $U(\text{iso})$  or
  - Anisotropic:  $U^{11}, U^{12}, U^{13}, U^{22}, U^{23}, U^{33}$
 (Displacement Parameters) => ORTEP
- Note: ORTEP 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**  
Simple 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- Linear Least Squares Method  
Min:  $\sum_{hkl} [w_{hkl} (|F_{hkl}(\text{obs})|^2 - |F_{hkl}(\text{calc})|^2)]^2$
- Time consuming: disorder, weak data

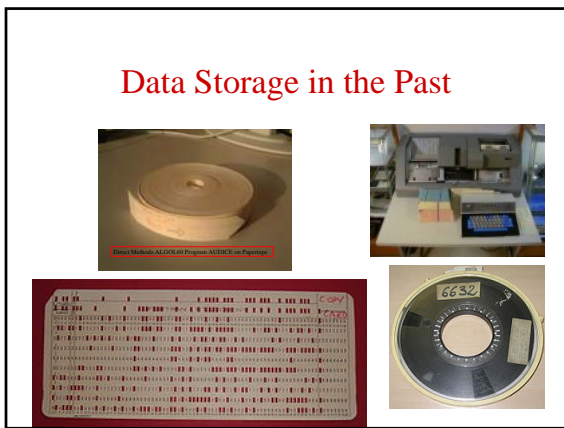
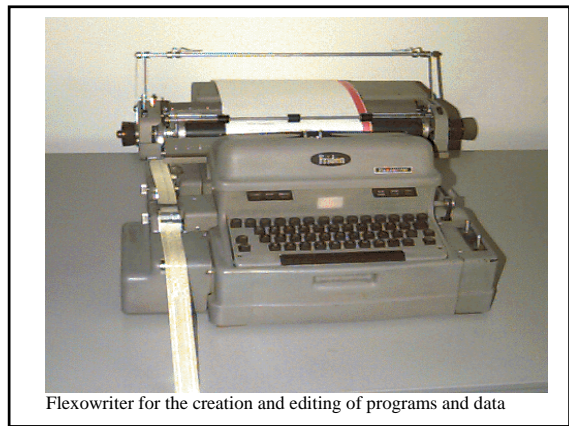
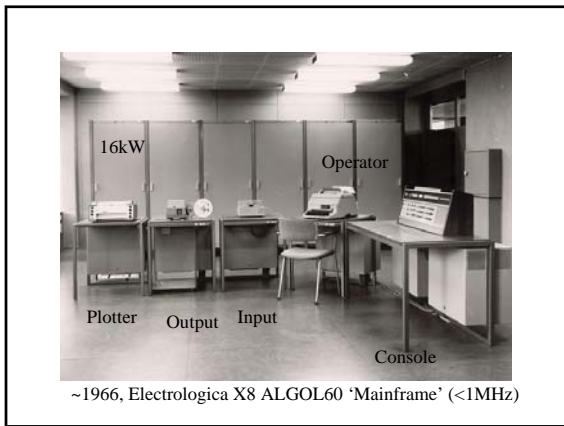
### Analysis

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



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

The values of  $\beta$  are defined by the expression  $\beta = (2\theta)^2 \sin^2 \theta$

Table 1. Atomic parameters

	$x/a$	$y/b$	$z/c$	$B_{11}$	$B_{22}$	$B_{33}$	$B_{12}$	$B_{13}$	$B_{23}$
C121	0.0000	0.0000	0.1761	2.25	2.53	2.53	0.26	0.26	0.14
C221	-0.0139	0.1762	0.1761	2.25	2.53	2.53	0.26	0.26	0.14
C321	0.0000	0.1762	0.3522	2.26	2.52	2.50	0.27	0.27	0.14
C421	0.0002	0.0711	0.5451	0.69	4.11	1.15	-0.04	-0.04	0.04
C521	0.0002	0.0007	0.5626	3.91	6.52	3.83	-0.04	-0.04	0.04
C621	0.2123	0.1761	0.2642	0.73	3.77	3.99	-0.17	-0.01	0.17
C721	0.1764	0.1761	0.0000	3.31	2.71	2.71	-0.22	-0.02	0.00
C821	0.2242	0.2669	0.0411	0.69	4.07	4.04	-0.02	-0.10	0.25
C921	-0.1208	0.0000	0.0000						
C1021	-0.1208	0.0468	0.0271						
C1121	0.0000	0.0000	0.0000						
C1221	0.0000	0.0000	0.0000						
C1321	0.0000	0.0000	0.0000						
C1421	0.0000	0.0000	0.0000						
C1521	0.0000	0.0000	0.0000						

Mean standard deviation of positional coordinates, 0.002 Å.  
Mean standard deviation of thermal parameters, 0.05.

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

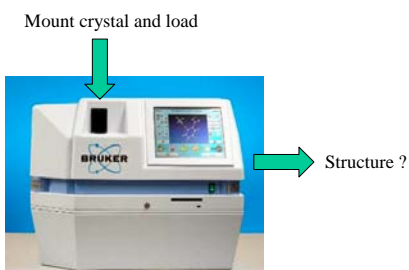
C. J. BROWN

Table 2. Observed and calculated structure amplitudes

h	k	l	Observed	Calculated	h	k	l	Observed	Calculated
0	0	0	100	100	0	0	1	100	100
0	0	2	100	100	0	0	3	100	100
0	0	4	100	100	0	0	5	100	100
0	0	6	100	100	0	0	7	100	100
0	0	8	100	100	0	0	9	100	100
0	0	10	100	100	0	0	11	100	100
0	0	12	100	100	0	0	13	100	100
0	0	14	100	100	0	0	15	100	100
0	0	16	100	100	0	0	17	100	100
0	0	18	100	100	0	0	19	100	100
0	0	20	100	100	0	0	21	100	100
0	0	22	100	100	0	0	23	100	100
0	0	24	100	100	0	0	25	100	100
0	0	26	100	100	0	0	27	100	100
0	0	28	100	100	0	0	29	100	100
0	0	30	100	100	0	0	31	100	100
0	0	32	100	100	0	0	33	100	100
0	0	34	100	100	0	0	35	100	100
0	0	36	100	100	0	0	37	100	100
0	0	38	100	100	0	0	39	100	100
0	0	40	100	100	0	0	41	100	100
0	0	42	100	100	0	0	43	100	100
0	0	44	100	100	0	0	45	100	100
0	0	46	100	100	0	0	47	100	100
0	0	48	100	100	0	0	49	100	100
0	0	50	100	100	0	0	51	100	100
0	0	52	100	100	0	0	53	100	100
0	0	54	100	100	0	0	55	100	100
0	0	56	100	100	0	0	57	100	100
0	0	58	100	100	0	0	59	100	100
0	0	60	100	100	0	0	61	100	100
0	0	62	100	100	0	0	63	100	100
0	0	64	100	100	0	0	65	100	100
0	0	66	100	100	0	0	67	100	100
0	0	68	100	100	0	0	69	100	100
0	0	70	100	100	0	0	71	100	100
0	0	72	100	100	0	0	73	100	100
0	0	74	100	100	0	0	75	100	100
0	0	76	100	100	0	0	77	100	100
0	0	78	100	100	0	0	79	100	100
0	0	80	100	100	0	0	81	100	100
0	0	82	100	100	0	0	83	100	100
0	0	84	100	100	0	0	85	100	100
0	0	86	100	100	0	0	87	100	100
0	0	88	100	100	0	0	89	100	100
0	0	90	100	100	0	0	91	100	100
0	0	92	100	100	0	0	93	100	100
0	0	94	100	100	0	0	95	100	100
0	0	96	100	100	0	0	97	100	100
0	0	98	100	100	0	0	99	100	100
0	0	100	100	100	0	0	101	100	100

- ### 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



## 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 dictionary strings
- Format of telephone 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**  
Attribution of canonical data descriptors in CIF are described by a machine-readable definition language (DDL).

- Simple Dictionary Definition Language (DDL.1)
- Relational Dictionary Definition Language (DDL.2) (Advanced)
- method-rich Dictionary Definition Language (DDL.3) (draft)

**Current CIF Dictionaries**  
Canonical data descriptors and their attributes are presented in machine-readable data dictionaries.

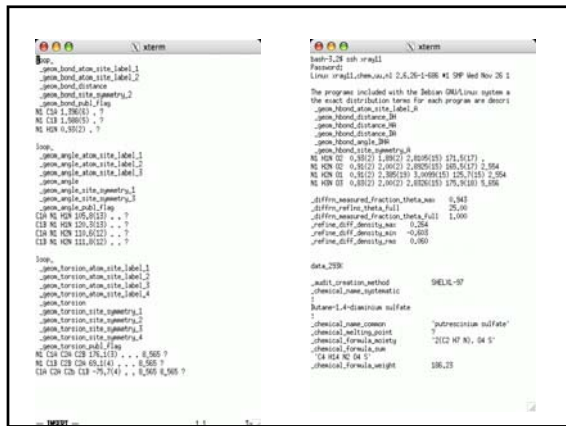
**Descriptors for small-molecule, inorganic and other small-unit-cell structures**

- Core dictionary (coreCIF) [ ASCII | HTML | More information ]
- Powder dictionary (powCIF) [ ASCII | HTML | PDF | More information ]
- Multibuild and composite structures dictionary (mcCIF)

## Practical Approach

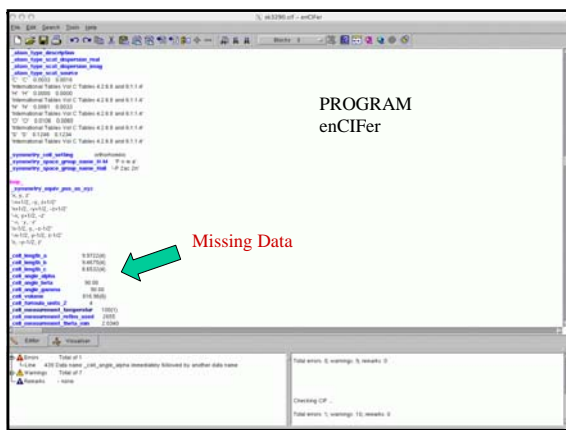
- We ignore here the scary details that are not relevant in the current context
- We will Discuss the File structure
- We will look at its relevance for publication
- We will discuss software to edit and check the CIF file
- We will look at software that uses CIF as Input.



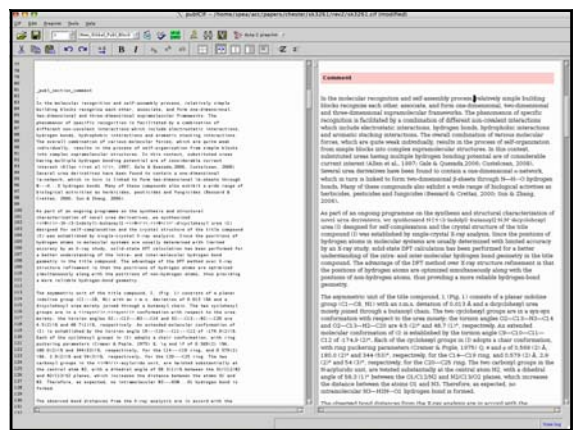


## CIF Completion

- CIF Files are created by the refinement program (e.g. SHELXL)
- Missing Data 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](http://www.ccdc.cam.ac.uk))



PROGRAM enCIFer



## 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

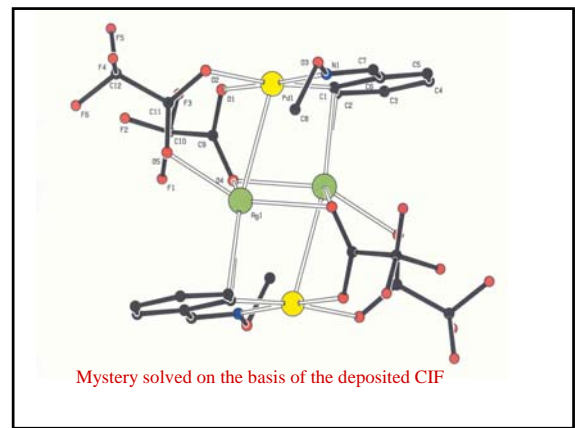
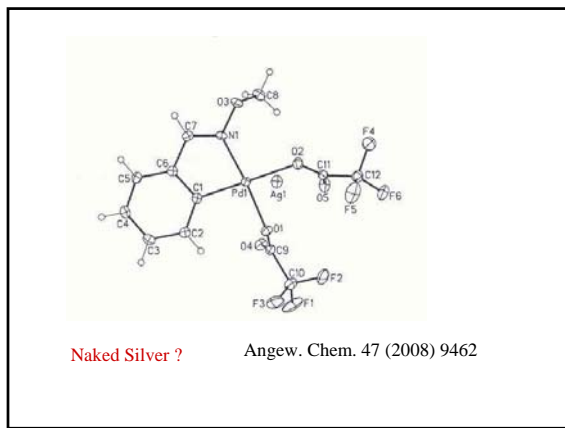
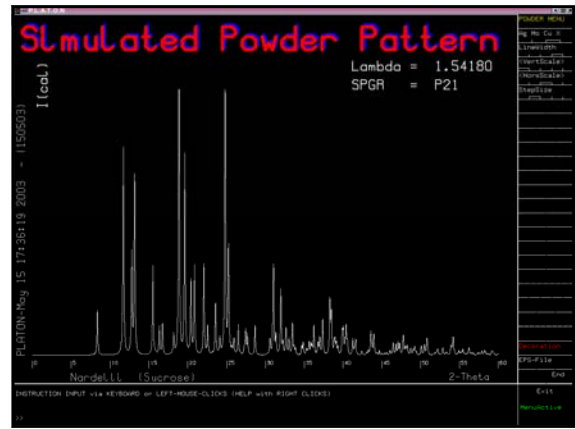


**PLATON**  
A Multipurpose Crystallographic Tool  
(C) 1990-2009 A.L. Spek - 40th-Version: 240908

GRAPHICS	GEOM-CALC	VOIDS	FLIP	SYMMETRY	ABSORPTION	REPORT	MISC-TOOLS
PLUTONato	CALC ALL	CALC SOLV	ADDSYM	DELrefABS	VALIDATION	SYSTEM-S	
DATEP/ADP	CALC INTRA	CALC K.P.	ADDSYM-EOL	ABSPeLScan	ASYM-VIEW	FCF2HKL	
NEWMAN	CALC INTER	SQUEEZE	ADDSYM-EXT	ABSTempa	FCF-VALID	EXPAND-P1	
RING-PLOTS	CALC COORD	CALC-FCF	ADDSYM-PLT	ABSGauss	SUPPLEMENT	FCF-GENER	
PLANE-PLAT	CALC METAL	CONTOUR-SQ	ADDSYM-SHX	ABSXLaL	ANALoFVRr	HKL-GENER	
POLYHEDRA	CALC GEOM	SOLV F30	NEWSYM	ABSSphere	ByvoePLoLr	HKL-TRANSF	
CONTOUR-DF	CBOND	SOLV PLAT	NONSYM	MULaonABS	ASYM-EXPTCT	EXOR-RES	
CONTOUR-FC	CAV	THRA	CAVITY-PLT	LEPAGE	ASYM-VALID	RNIS-RES	
RotatolFL	PLANE			DELRED	DLFFoVler	RENAME-RES	
HKL2Powder	hedAngle			MOLSYM	EXPECT-HKL	PDB -pdb	
SLmPowderP	AngleLines	FLIP MENU	SPGRfromEX		CSD-CELL	SPF -eLd	
rodLstFun	AnglePLLn	FLIP SHOW	ASYM		CSD-QUEST	SHELXL-res	
PATTERSON	CremerPopL	FLIP PATT	ASYMoverFR		StructTLdy	CIF -ooc	
BondValenc	FLIPPER	ZSLPageLwLn			Struol.nAnd	AUTO-RENUM	
PLUTONatL	HP1X - RES	STRUCTURE?	LeLnAndMat	XLaL	Habl.t	CIF-LOCAL	CIF2SHELXL

Ctrl Data (CIF) e22088.cIF Set (111)  
Ref Data (SHELXL) e22088.hKL (NO-DIAG) (111)

Browser - HELP



**Reflection CIF (FCF)**

```

cell length a 11.066
cell length b 11.066
cell length c 10.810
angle alpha 90.0
angle beta 90.0
angle gamma 90.0
cell origin 00.000

```

**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.

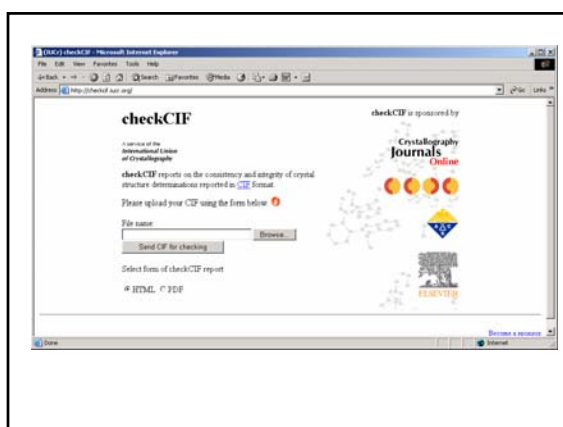
## VALIDATION QUESTIONS

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.

No syntax errors found. [CIF dictionary](#)  
Please wait while processing..... [Interpreting this report](#)


**Datablock: 1**

```

Bond precision:  C-C = 0.0157 Å    wavelength=0.71073
Cell:            a=7.4395(15)      b=27.725(6)      c=12.051(2)
                 alpha=90       beta=96.80(13)  gamma=90
Temperature: 153 K
          Calculated      Reported
Volume        2520.5(9)      2520.5(9)
Space group   F 227m          F 227m
Hall group    -F 22m          -F 22m
Moiety formula C28 H52 Br6 O7 Cl1 O3 N2 Ru C28 H52 Br6 O7 Cl1 O3 N2 Ru
Sum formula   C28 H52 Br6 O7 Cl1 O3 N2 Ru C28 H52 Br6 O7 Cl1 O3 N2 Ru
Mr            611.69         611.69
Dcalc cm-3  1.612             1.612
Z             4             4
Ru (eq-1)     2.260          2.260
Rint         1237.9         1237.9
Rsigma       1232.67        1232.67
h, k, l less  9,33,14        9,33,14
Observed      4452          4449
Train/Test    0.726, 0.955   0.563, 0.955
Rtest       0.546
Correction method= MLTI-SCAN    Theta(max)= 25.030
Data completeness= 0.999        Theta(99)= 25.030
R(Reflections)= 0.0961( 3286)    wR(Reflections)= 0.2547( 4449)
S = 1.037                Gpar= 306

The following ALERTS were generated. Each ALERT has the format:
Alert-name,ALERT-type,Alert-level.
Click on the hyperLinks for more details of the test.

Alert level 0
PLATON_ALERT_3_B SHELL Second Parameter in WGHT unusually Large    37.00
Alert level C
PLATON_ALERT_3_C The value of the weighted R factor is > 0.25
    
```



**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 Marshded")
- 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:**

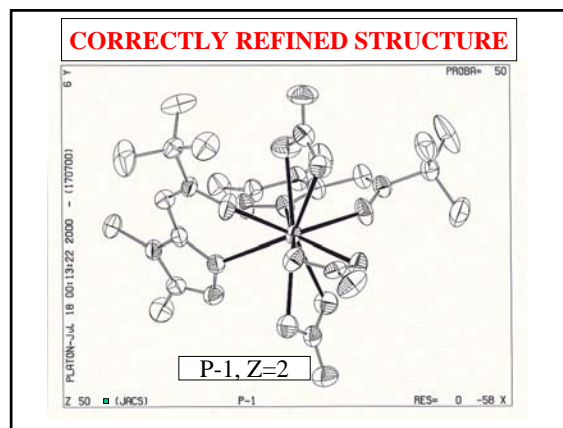
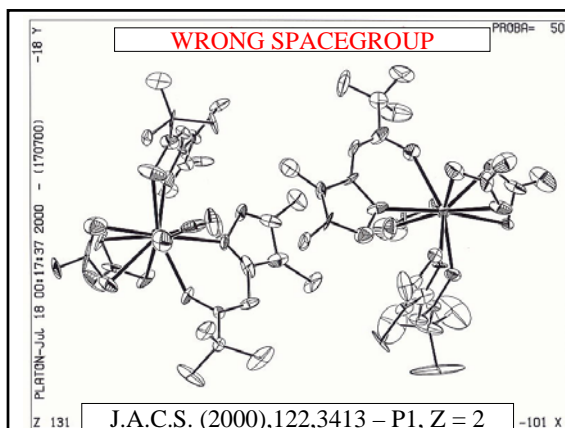
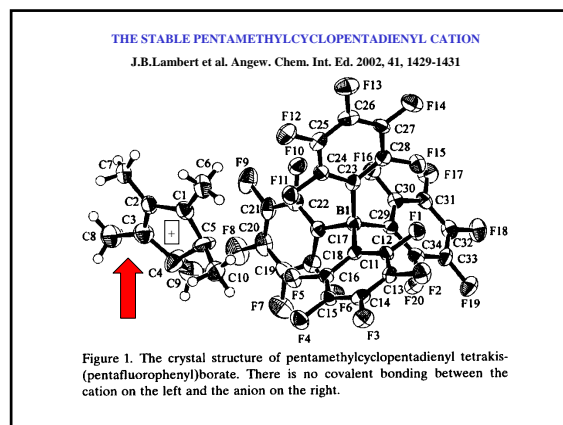
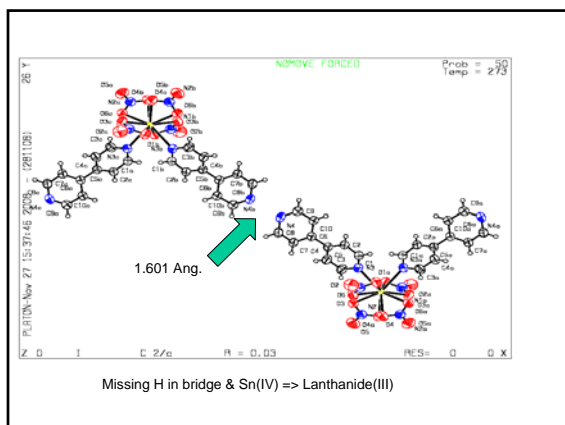
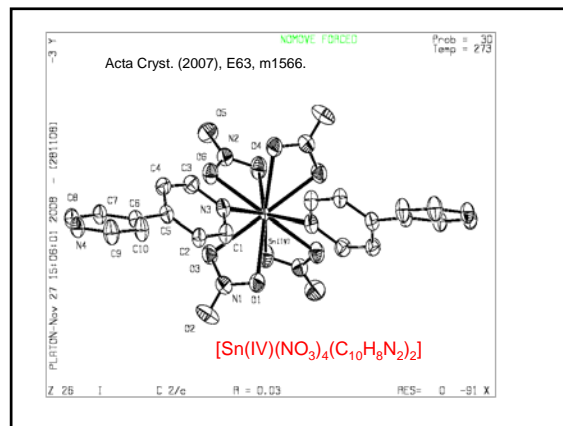
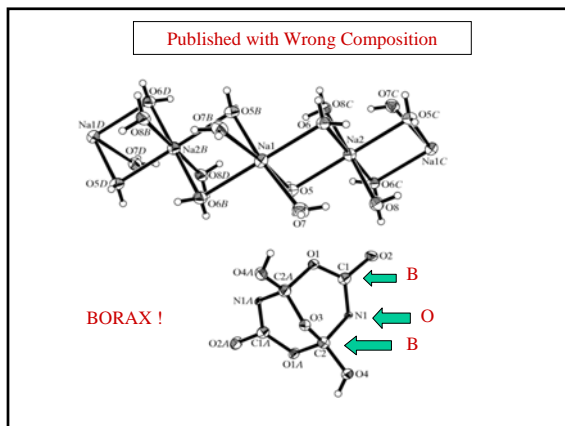
Automatic twinning detection as part of the IUCr CheckCif procedure

- 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.



**HOT STRUCTURE – FAST LANE PUBLICATION**

**CHEMICAL**  
Engineering News

Table of Contents  
 • Editor's Introduction  
 • Cover Story  
 • Industry News  
 • Business  
 • Technology  
 • Education  
 • Career  
 • Classifieds  
 • Advertisements  
 • News & Events  
 • Back Issues  
 • Contact Us

**SCIENCE & TECHNOLOGY**  
 April 29, 2002  
 Volume 30, Number 17  
 ISSN 0891-9524

**ELUSIVE CARBOCATION ISOLATED AS A SOLID**  
 Pentamethylcyclopentadienyl cation is found to be a stable singlet with a distorted structure

**NEWSMAKERS**  
 Chemists like to study molecules they can store in a bottle on the shelf. If they can't put the molecules in a bottle, they will, of course, study them any way they can—in solution. In the gas phase, in a matrix, or even in a particular modification of hosting a molecule that no one else could be held.

**NEWSMAKERS** A space-filling model of the pentamethylcyclopentadienyl cation shows the distorted, nonplanar ring.

Chemistry professor Joseph B. Lambert of Northwestern University knows that structure. Earlier this month, he and graduate student J. Guy Bertrand reported the first isolation and X-ray structural characterization of a cyclopentadienyl cation—specifically, the pentamethyl-substituted  $C_5Me_5^+$  cation (*Science*, **296**, 466, 467, 4623 (2002)).

**Pentamethylcyclopentadienyl**

Common Anionic Form      “New” Cationic Form

$Cp^*$

Assigned Bond types

**Enthusiastic Last Paragraph of C&E-News**

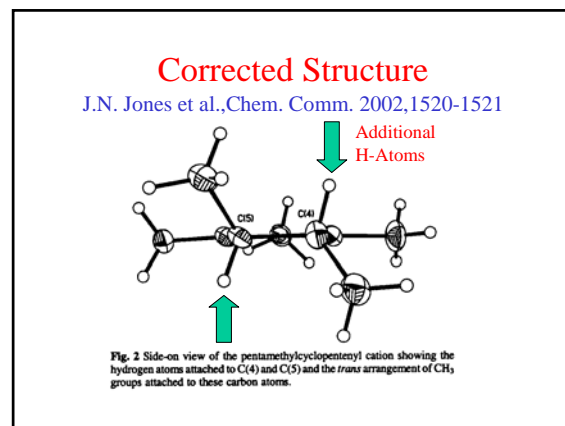
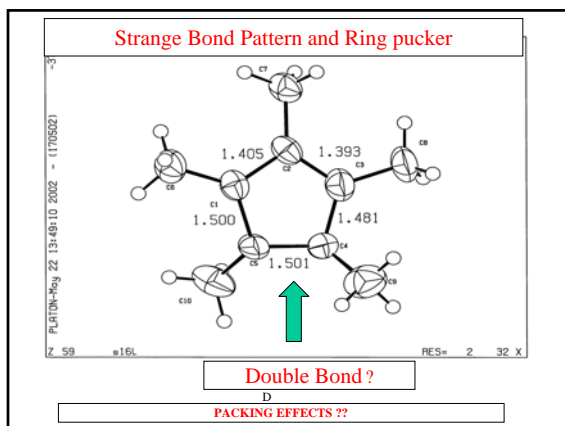
... The Northwestern chemists are now exploring the reactivity of  $C_5Me_5^+$ . “You can sit down and write lots of interesting reactions on paper,” Lambert says, and it will be interesting to see if the molecule reacts as expected. But Considering this cation’s track record, it might be safer to expect **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 pentamethylcyclopentyl cation was prepared and characterized (the corresponding communication will be published in issue 13, and will appear earlier on the *Angewandte Chemie* homepage).



## 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