



# Advanced Crystallography

## Publication of Crystal Structures

Charles Campana, Ph.D.  
Senior Applications Scientist  
Bruker AXS

# Course Overview



## Advanced Crystallography - Publication of Crystal Structures

- Collection of high-quality intensity data
- Common refinement problems
- Data evaluation and error analysis
- Report generation and display of crystallographic results
- Preparation of Crystallographic Information Framework (CIF) files
- The use of PubCIF and other CIF-checking programs
- Submission of manuscripts



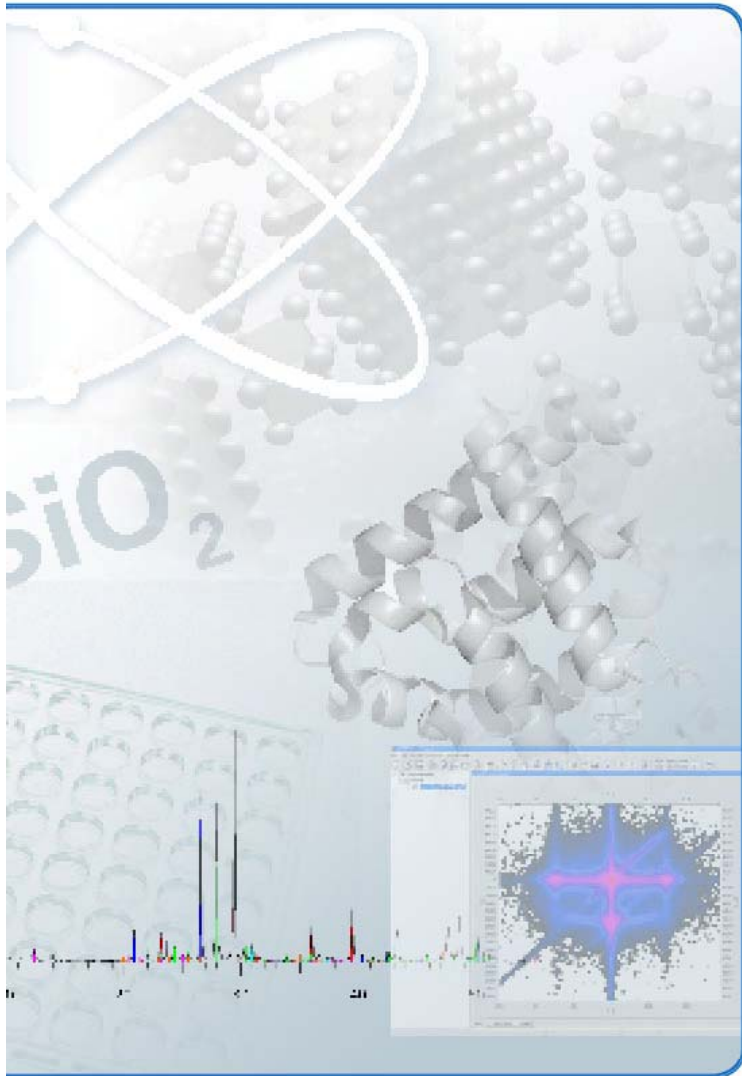
# Introduction



# Why Should You Publish Crystal Structures?

- One of the most important reasons for doing scientific research is to advance the total body of knowledge in each field. This requires that experimental results be described very precisely, using accepted standards for each field, and that the manuscript reviewed by experts in the field prior to acceptance for publication.
- Successful publication of papers in peer-reviewed journals, makes the data and its analysis available to the international scientific community, so that current and future scientists may repeat and extend the research, as new tools and additional information become available.
- Publication of research results gives proper credit to all co-authors and their sponsoring institutions and funding sources.
- Presentation of posters or talks at scientific meetings should be used for preliminary results only and not a substitute for 'real' publication in peer-reviewed scientific journals.

# Crystallographic Databases

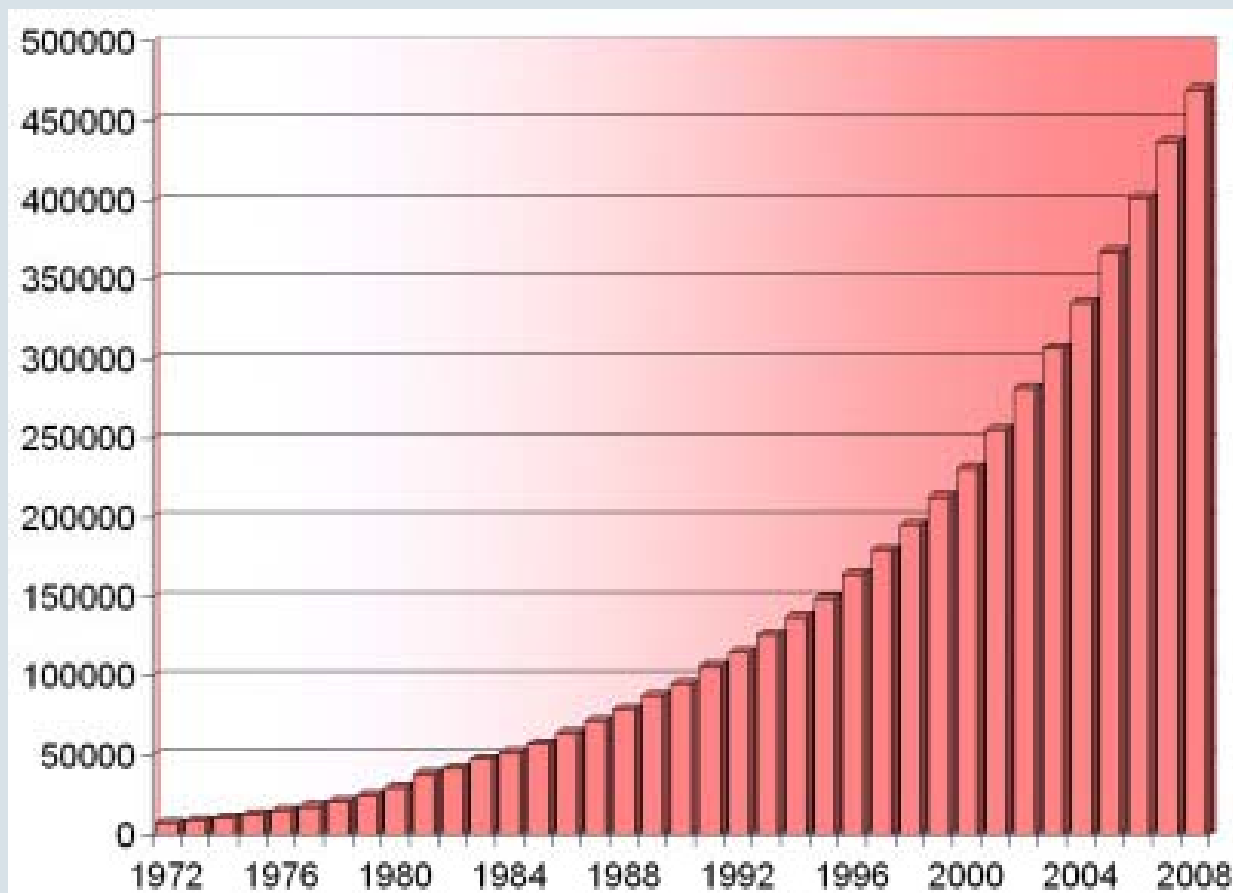




# Growth in Number of Published Crystal Structures

- Crystal structure determination is the method of choice for the analysis of novel chemical compounds from monatomic metals to proteins and viruses, with some 700,000 crystal structures published so far.
- Of these, over 500,000 (70%) are carbon-containing organic and metal-organic small molecules, for which experimental three-dimensional (3D) numerical results are stored in the Cambridge Structural Database (CSD).
- Other crystallographic databases – Protein Data Bank, Mineralogical Data Base, Inorganic Data Base

# Growth in Number of Structures in Cambridge Structural Database



# Growth in Number of Structures in the Protein Data Bank



October 1, 2010

Last Updated: 11 Aug 2010

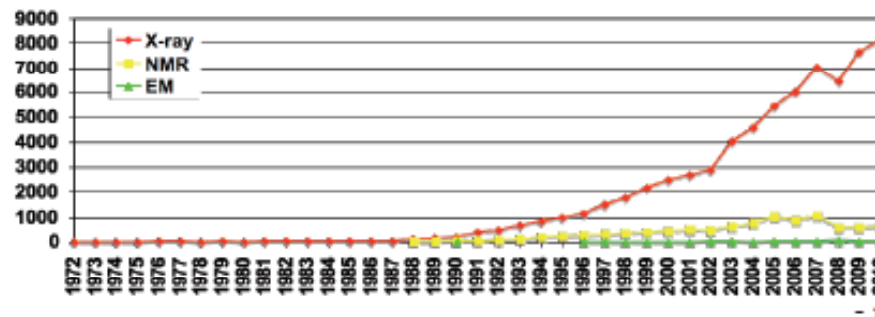
## PDB Depositions

By deposition and processing site  
\*(projected)

Year	Total Depositions	Deposited To			Processed By		
		RCSB	PDBj	EBI	RCSB	PDBj	EBI
2000	2983	2445	10	528	2297	158	528
2001	3286	2673	118	495	2408	383	495
2002	3563	2769	289	505	2401	657	505
2003	4830	3488	673	669	3135	1026	669
2004	5508	3796	900	812	3083	1613	812
2005	6678	4507	1166	1005	3563	2110	1005
2006	7282	5145	1052	1085	4252	1945	1085
2007	8130	5399	1603	1128	4703	2299	1128
2008	7073	5452	648	973	4106	1994	973
2009	8300	6715	527	1058	5069	2173	1058
2010	5415 (*8800)	4291	326	798	3421	1196	798
<b>TOTAL</b>	<b>63048</b>	<b>46680</b>	<b>7312</b>	<b>9056</b>	<b>38438</b>	<b>15554</b>	<b>9056</b>

By experimental method  
\*(projected)

Note: NMR depositions at the RCSB PDB come through the BMRB



# Documentation of a Crystal Structure Analysis





# Laboratory Notebooks

A dedicated laboratory notebook should be used to log all crystal structure determinations.

Each sample should be assigned a unique identification number, which may be cross-referenced to the synthetic notebook which contains detailed information on the preparation, purification and crystallization of the sample. Typical identification numbers include the research group, initials of student or postdoctoral associate, and date of analysis.

The laboratory notebook should contain all important information not included in the electronic data (such as crystal size, color etc.), as well as choices made during refinement.

The laboratory notebook must also contain the name and location of final archive files for each structure.



# Laboratory Notebooks

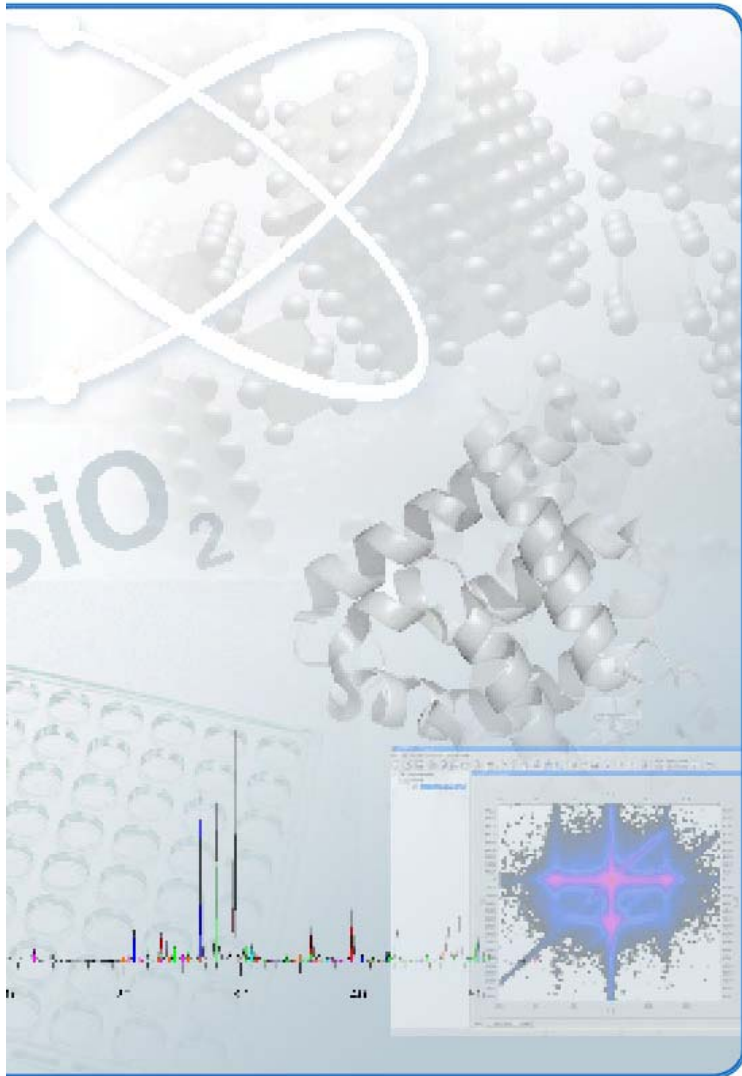
A proper synthetic notebook should include negative as well as positive results. Documentation of unsuccessful attempts may help future avoid the same experimental paths.

Good science requires that successful experiments be documented in sufficient detail that the results may be duplicated and verified, if necessary.

A few of the more competitive research groups require that entries in laboratory notebooks be signed and dated, then reviewed and witnessed by collaborators or supervisors on a daily or weekly basis.

When each investigator leaves the research group, the laboratory notebooks are stored in a secure place, under the control of the research director.

## Collection of High Quality Intensity Data

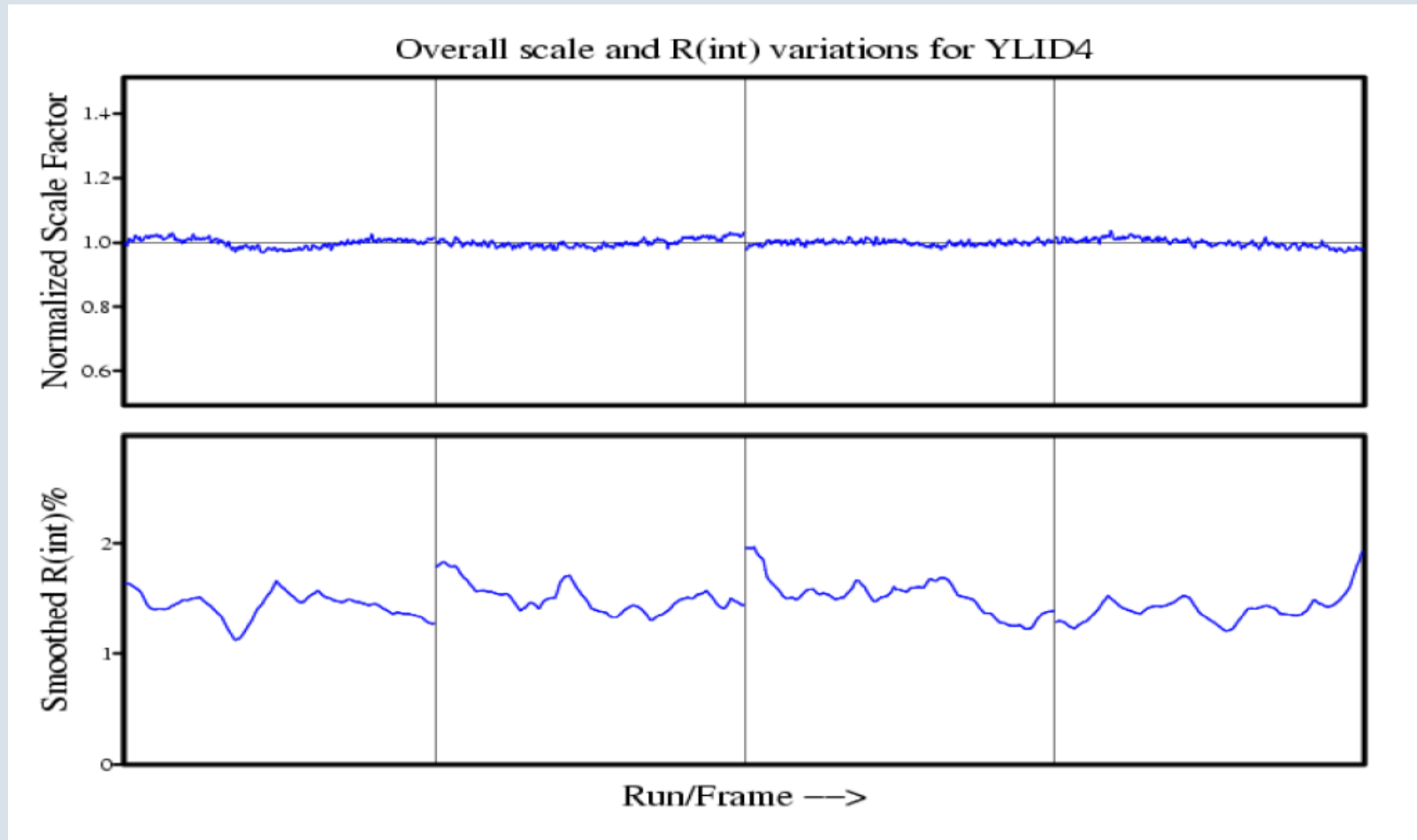




# Collection of High Quality Data

- Choose the data collection options to suit the problem (e.g., X-ray source, radiation type, crystal-to-detector distance, sample temperature, etc.)
- Choose a data collection strategy to balance resolution, completeness, redundancy, exposure time, and total time available for data collection
- Process (integrate) data carefully to the maximum usable resolution (minimum resolution 0.84 Å for small molecules in IUCr journals)
- Scale data and correct data for absorption effects (empirical or face-indexed correction) using the correct Laue class and carefully examine the diagnostic plots

# Diagnostic Plots from Data Reduction





# Crystal Structure Refinement



# Structure Refinement with SHELX / SHELXTL

## SHELX (or SHELXTL):

"A short history of SHELX". Sheldrick, G.M. (2008). Acta Cryst. A64, 112-122

<http://journals.iucr.org/a/issues/2008/01/00/issconts.html>

# Structure Refinement with SHELX / SHELXTL

- George M. Sheldrick, Professor of Structural Chemistry at the Georg-August-Universität Göttingen and part-time programming technician.
- Author of public-domain SHELX and Bruker SHELXTL solution and refinement software and other programs.
- Sheldrick software is used in ca. 70% of all crystal structure refinements.



# SHELX Data File Format (HKLF 4)

```
example.hkl:
0 0 1 14.04 2.27 1
0 0 220748.33 212.60 1
0 0 3 3.91 4.31 1
0 0 4 8452.21 96.84 1
0 0 5 15.19 9.36 1
0 -1 -690345.18 728.10 1
0 0 7 11.09 8.55 1
0 0 8 2957.52 45.64 1
```

Miller indices h k l	Intensity I	Standard deviation of the intensity $\sigma(I)$	batch number (see BASF in the SHELX manual)
0 0 1	14.04	2.27	1
0 0 2	220748.33	212.60	1
0 0 3	3.91	4.31	1
0 0 4	8452.21	96.84	1
0 0 5	15.19	9.36	1
0 -1 -6	690345.18	728.10	1
0 0 7	11.09	8.55	1
0 0 8	2957.52	45.64	1

...  
0 0 0 End of file

- In contrast to the res/ins/cif files, the file \*.hkl has a precise format (i.e., each space matters).
- Data are sorted so that equivalent reflections are together, but they are not automatically merged.

# Violations of Systematic Absences / Inconsistent Equivalents

```

exemple.lst:
  h  k  l      Fo^2      Sigma      Why rejected
  0  0  1      104.55     5.29      observed but should be systematically absent
  2  0  1       74.15     6.04      observed but should be systematically absent
etc.
20975 Reflections read, of which 685 rejected

-10 <= h <= 10,   -13 <= k <= 13,   -18 <= l <= 18,   Max. 2-theta = 137.96
 23 Systematic absence violations

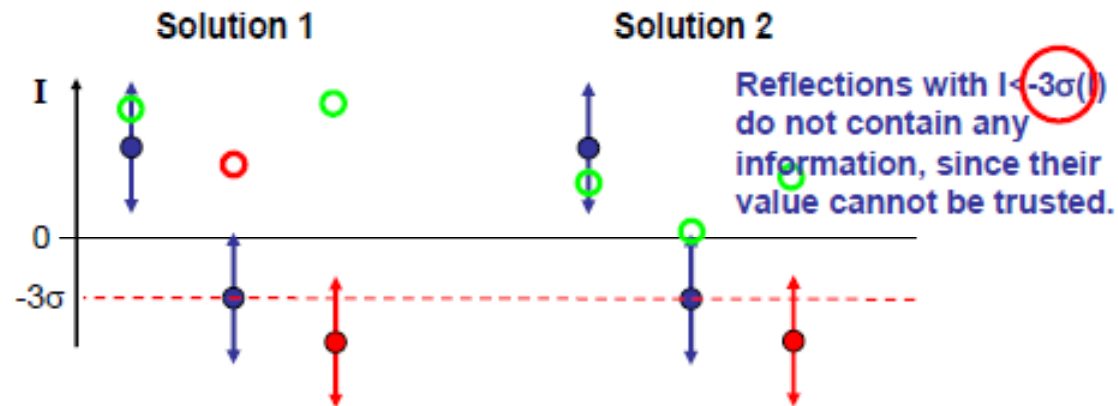
Inconsistent equivalents etc.
  h  k  l      Fo^2      Sigma(Fo^2)  N  Esd of mean(Fo^2)
  2  0  0    127128.20    518.46     4   3194.33
  0  2  0    65677.31    243.31     6   1223.07
etc.
 14 Inconsistent equivalents
2847 Unique reflections, of which 0 suppressed

R(int) = 0.0275   R(sigma) = 0.0111   Friedel opposites merged
  
```

- Rejected reflections:
- Systematic absences
  - Inconsistent equivalents
  - Reflections with  $I < -3 \cdot \sigma(I)$

# Reflections with Negative Intensities

- During the integration of the reflections (peaks), the intensities are determined by  $I = I_{\text{reflection}} - I_{\text{background}}$
- For weak reflections the result can be negative.
- Reflections with  $0 > I > -3\sigma(I)$  contain useful information. ( $I = 0$  is in the margin of errors for these reflections.)
- Reflections with  $I < -3\sigma(I)$  are physically impossible, indicate problems in the determination and are thus omitted.

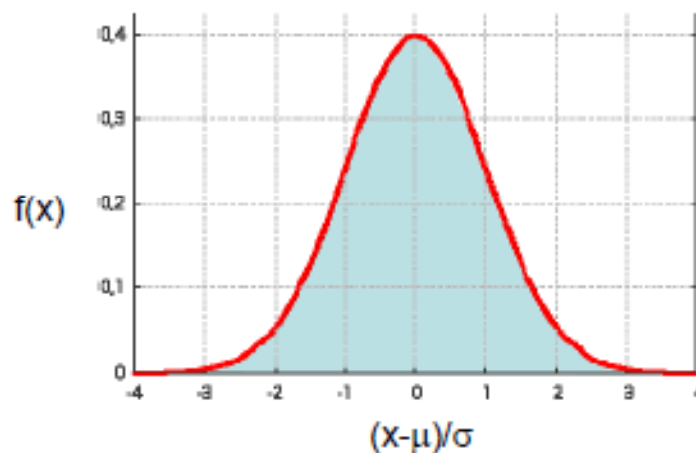


# Standard Deviations and Confidence Levels

A random distribution of errors should follow a Gaussian distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$\mu$ : correct value  
 $\sigma$ : standard deviation



$$\int_{\mu-\sigma}^{\mu+\sigma} f(x) = 0.68 = 68\%$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} f(x) = 0.96 = 96\%$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} f(x) = 0.99 = 99\%$$

**99.7% probability that x is in the interval  $\mu \pm 3\sigma$**

$$a=1.234; \sigma=0.005: a = 1.234(5) \rightarrow a = 1.234 \pm 0.015$$

# Merging of Reflections

Number of reflections measured

```

exemple.lst:
20975 Reflections read, of which 685 rejected

-10 <= h <= 10,   -13 <= k <= 13,   -18 <= l <= 18,   Max. 2-theta = 137.96
  23 Systematic absence violations

Inconsistent equivalents etc.
  h  k  l    Fo^2  Sigma(Fo^2)  N  Estd of mean(Fo^2)
  2  0  0  127128.20  518.46    4  3194.33
  0  2  0   65677.31  243.31    6  1223.07
etc.
  14 Inconsistent equivalents
  2847 Unique reflections, of which 0 suppressed

R(int) = 0.0275   R(sigma) = 0.0111   Friedel opposites merged
  
```

Number of independent reflections measured (not repeated, not related by symmetry)

# Merging of Reflections – R(int)

*R<sub>int</sub>* →

```
exemple.lst:
20975 Reflections read, of which 685 rejected
--
2847 Unique reflections, of which 0 suppressed
R(int) = 0.0275    R(sigma) = 0.0111    Friedel opposites merged
```

$$R_{\text{int}} = \frac{\sum |F_o^2 - F_o^2(\text{mean})|}{\sum F_o^2}$$

- R(int): Merging error (measure of the precision/reproducibility)
- Possible error sources (high R(int) value):
  - Incorrect Laue group
  - Bad or missing absorption correction
  - Crystal decomposition
  - Twinning
  - Goniometer problems (covered reflections, misalignment)

## Merging of Reflections – R(sigma)

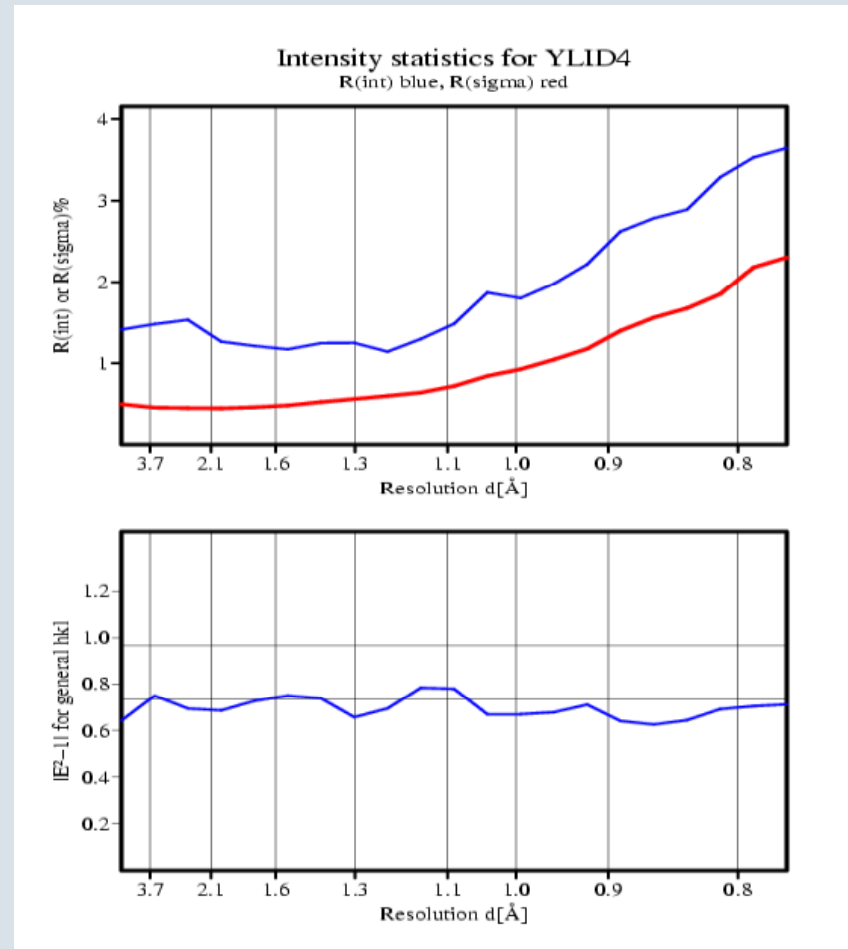
```
exemple.lst:
20975 Reflections read, of which 685 rejected
--
2847 Unique reflections, of which 0 suppressed
R(int) = 0.0275   R(sigma) = 0.0111   Friedel opposites merged
```

$$R_{\sigma} = \frac{\sum \sigma(F_o^2)}{\sum F_o^2}$$

- R(sigma) - Measure of the signal-to-noise ratio
- As a rough approximation, the structure confidence factor R1 cannot be much lower than R(sigma)
- If R(int) >> R(sigma), (e.g., more than 2-3 times), there is a problem

# Diagnostic Plots from Data Reduction

– Rint, Rsigma, [ $E^2-1$ ]



# Confidence Factors – R1 and wR2

Optimised value: 
$$M = \sum w(F_o^2 - F_c^2)^2$$

(The lower M, the better is the agreement of our model with the experimental data.)

But: M increases with the number of reflections and with their intensity. It is thus structure dependent, with well diffracting structures with high redundancy giving the highest M values. We thus need a structure independent value.

Confidence factor, Residual, R-factor:

R2 → 
$$wR_2 = R_w(F^2) = \sqrt{\frac{\sum w(F_o^2 - F_c^2)^2}{\sum w(F_o^2)^2}}$$

For statistical reasons, refinement against  $F^2$  gives R-factors approximately twice as high than those for refinement against  $F$ . To facilitate comparison (and to increase acceptance of the new method) SHELXTL calculates also the R-factor based on  $F$ .

$$R_1 = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|} \quad \leftarrow R1$$

## Goodness-of-Fit - S

The GoF or GooF is another value which describes the quality of our model:

$$GooF = S = \sqrt{\frac{\sum w(F_o^2 - F_c^2)^2}{N_{Ref.} - N_{Par.}}}$$

$N_{Ref.}$ : number of independent reflections,  $N_{Par.}$ : number of parameters

- In contrast to the R-factor, which also depends on the signal-to-noise ratio, S is relatively independent from the noise.
- S should be around 1 for a good structure
  - S > 1: bad model or bad data/parameter ratio
  - S < 1: model is better than the data - problems with the absorption correction, outlier reflections at low resolution
- Caution – S is strongly affected by the weighting scheme!

# Criteria for Good Structures

SHELXL calculates 4 confidence values:

- **wR2 (all data)** Refinement against  $F^2$  requires a correct weighing scheme
- **wR2 (observed data,  $I > 2\sigma(I)$ )**
- **R1 (all data)** The weighing schemes optimised for refinement against  $F^2$  cannot be used for the calculation of R1.
- **R1 (observed data,  $I > 2\sigma(I)$ )**

The important values are **wR2 (all data)** (since we do the refinement with all data) and **R1 (observed data)**, for comparison with the old method.

	Good	Acceptable	Problematic	Really problematic
R1	< 5%	< 7%	>10%	>15%
wR2	< 12%	< 20%	>25% (ou > 2*R1)	>35%
S	0.9-1.2	0.8-1.5	<0.8 ou >2	<0.6 ou >4

# Residual Electron Density

**XP: info Q1**

**test.cif:**

```
_refine_diff_density_max 0.425  
_refine_diff_density_min -0.560
```

*How to find the residual electron density?*

**test.lst:**

```
Electron density synthesis with coefficients Fo-Fc  
Highest peak 0.42 at 0.2140 0.0398 0.5135 [ 0.40 A from C2 ]  
Deepest hole -0.56 at 0.1424 0.4014 0.6865 [ 0.53 A from F3 ]
```

If our model is good, we should have described all electrons in our structure. Thus the remaining electron density should be zero.

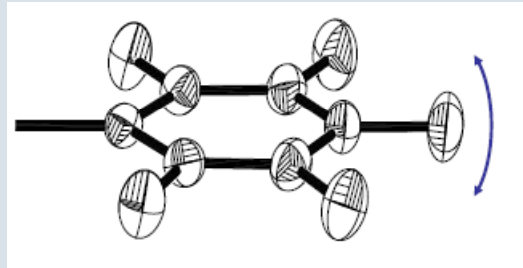
Acceptable values for residual electron density:

- For light atom structures (H – F) :  $< 0.5 \text{ e}^-/\text{\AA}^3$
- For heavy atom structures : 10% of the electrons of the heavy atom per  $\text{\AA}^3$  in a distance smaller 1.2  $\text{\AA}$  from the heavy atom. (Fourier truncation errors)
- Accumulation of electron density on special positions

Sources of errors:

- Bad absorption correction
- Disorder

# Thermal Ellipsoids



- With the exception of a wrong space group, most other problems of a structure are more visible in the thermal parameters than in the atom positions.
- In general:
  - Values of the thermal displacement should be comparable for comparable atoms.
  - The displacement should be in agreement with the thermal vibration of lowest energy.

# Common Refinement Problems



# Large Correlation Matrix Elements

**test.lst:**

Largest correlation matrix elements

0.853 U11 Fe1 / OSF

0.728 U11 S2 / U11 Fe1

0.524 U11 S1 / U11 Fe1

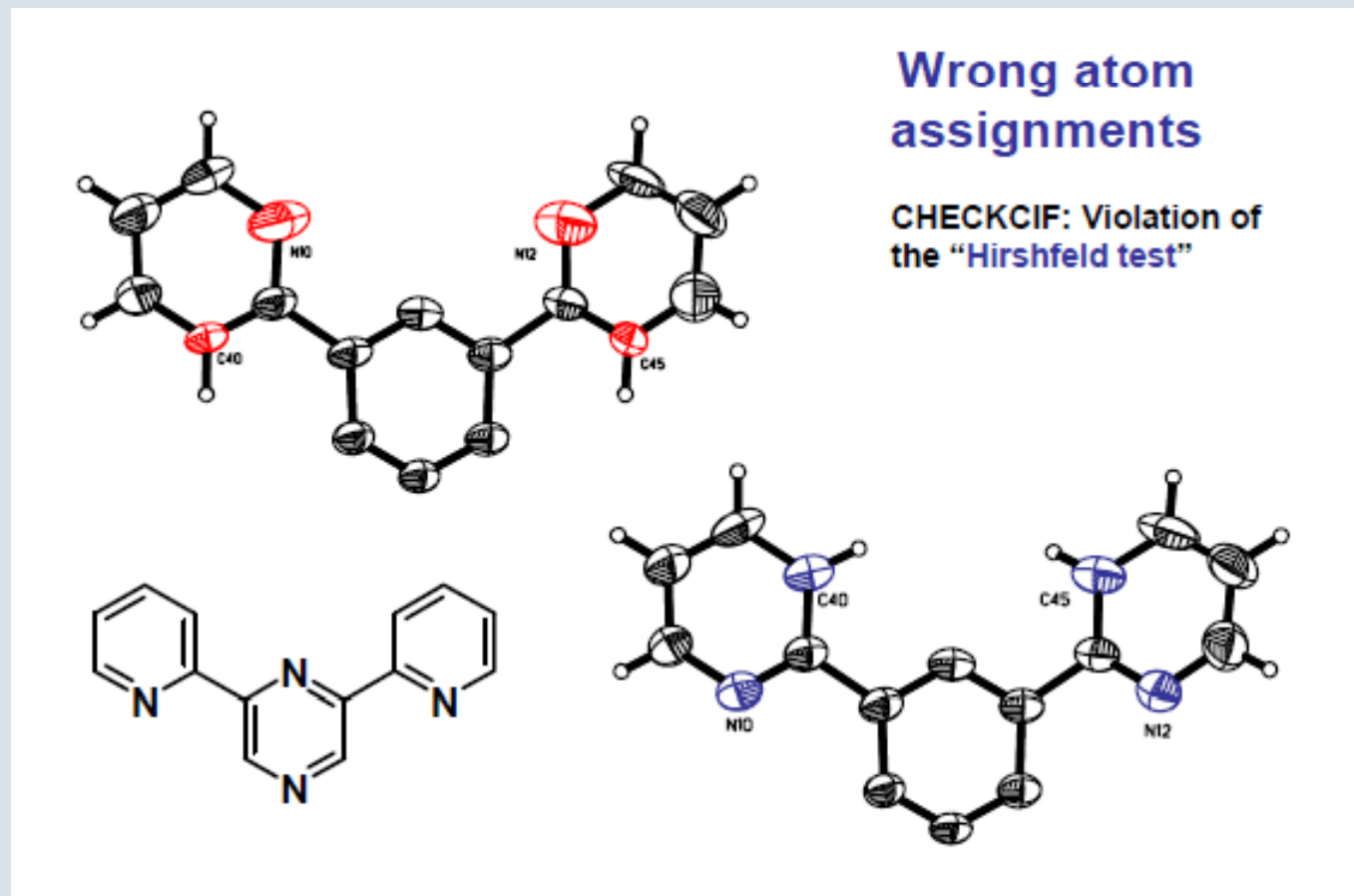
0.771 U11 S2 / OSF

0.588 U11 S1 / OSF

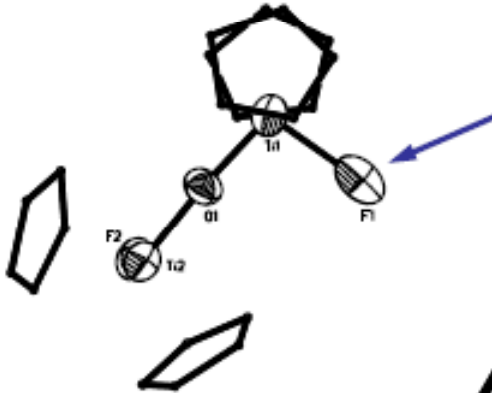
0.543 U12 S1 / U23 S1

- Values  $> 0.5$  for the correlation matrix elements indicate that some parameters in the refinement are dependent on each other.
- Some correlations are acceptable, for example, between the thermal parameters of the heavy atom and the overall scale factor or between the  $U_{xy}$  of the same atom.
- Attention: A high number of correlations  $> 0.5$  between multiple atoms might indicate a missed symmetry! (wrong space group).

# Incorrect Atom Assignments

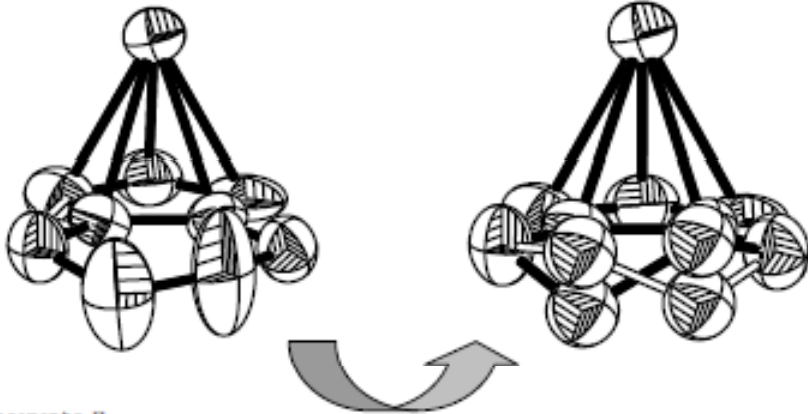


# Disorders



Displacement parallel to a bond greater than perpendicular displacement  
 $\Rightarrow$  disorder, two atoms with different bond lengths sharing the same position, here F et Cl.

CHECKCIF: Violation of the “Hirshfeld test”



```

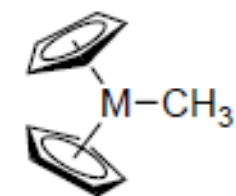
*.lst:
Principal mean square atomic displacements U
[-]
0.3098 0.0893 0.0464 C4 may be split into 0.6218 0.2673 0.2408 and 0.6118 0.2471 0.2666
0.3100 0.0924 0.0392 C5 may be split into 0.5976 0.3191 0.3424 and 0.5834 0.3017 0.3597
    
```

# Unusual Bond Lengths and / or Bond Angles

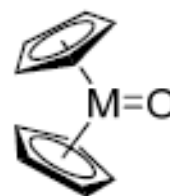


**Verify if the obtained geometry is reasonable**

For example, M-CH<sub>3</sub> versus M=O:



*synthesized*



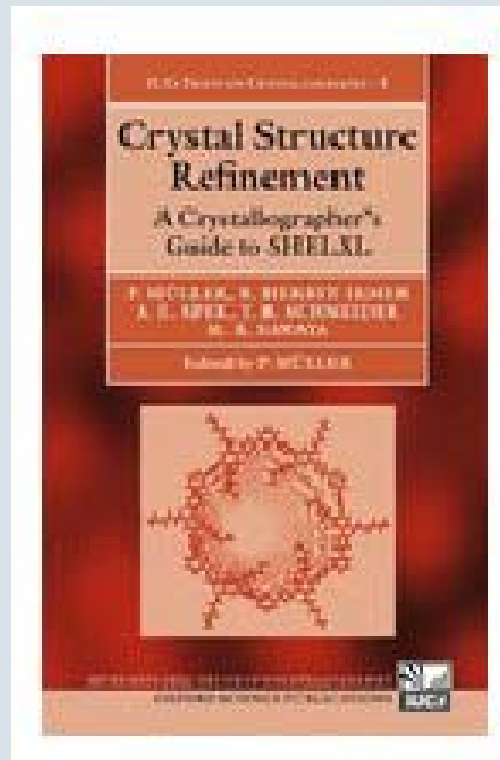
*crystallized*

**Sources for bond distances and angles values:**

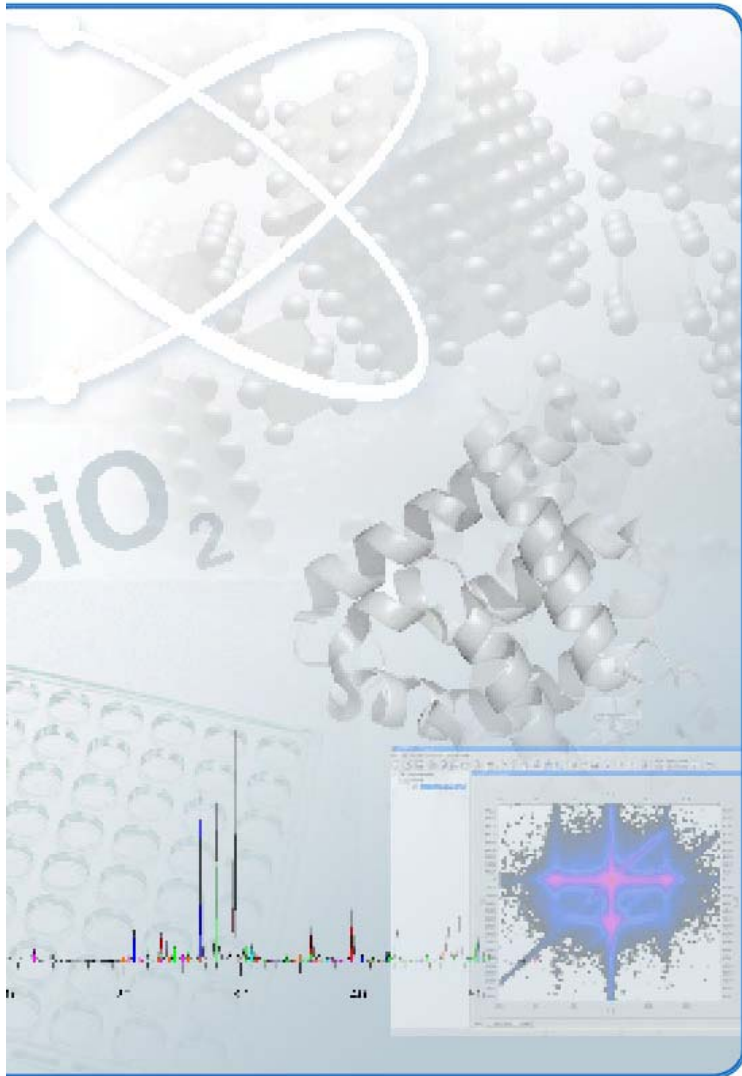
- Cambridge data base
- A. G. Orpen, L. Brammer, F. H. Allen, O. Kennard, D. G. Watson, R. Taylor, *J. Chem. Soc. Dalton Trans.* **1989**, S1-S83.
- F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, R. Taylor, *J. Chem. Soc. Perkin Trans. II* **1987**, S1-S19.

# Recommended Reference for Problem Structures

Müller, P. *Crystal Structure Refinement*. 1st ed. New York, NY: Oxford University Press, 2006. ISBN: 0198570767



## Report Generation and Display of Structure





# Final Crystal Structure Refinement

- Correct SFAC and UNIT instructions
- SIZE and TEMP instructions
- L.S. Instruction with sufficient number of cycles for convergence
- BOND (\$H) instruction
- FMAP 2 and PLAN for final difference map
- ACTA instruction
- CELL and ZERR instructions containing unit cell constants from final integration with proper sigma's
- CONF instruction ( torsion angles)
- RTAB instruction
- HTAB (and EQIV) instructions for important hydrogen bonds
- MPLA instructions for important molecular planes
- WPDB instruction (write PDB file)



# Production of Final Crystal Diagrams

- Bruker SHELXTL – XP Program
- Bruker APEX2 – XSHELL Program
- PLATON – ORTEP Program
- CCDC - Mercury Program
- OLEX2 - ORTEP Program
- WinGX – ORTEP Program
- publCIF – JMOL Program

# Data Evaluation and Error Analysis



# PLATON Program

- **PLATON Reference:**

A.L. Spek, "PLATON, a multipurpose crystallographic tool." Utrecht University, Utrecht, The Netherlands, 2001,  
<http://www.cryst.chem.uu.nl/platon/>

- **PLATON** is a **SHELX**-compatible structure viewing and analysis toolkit software package by Ton Spek that has a wide variety of functionality for checking, validation, structure drawing,, etc.

# PLATON Programs



**PLATON**  
A Multipurpose Crystallographic Tool  
(C) 1980-2010 A.L. Spek - 40M-Version: 80710

GRAPHICS	GEOM-CALC	VOIDS FLIP	SYMMETRY	ABSORPTION	REPORT	MISC-TOOLS
PLUTONauto	Calc ALL	Calc Solv	ADD3YM	MULscanABS	ValidatIon	SYSTEM-3
ORTEP/ADP	Calc Intra	Calc K.P.I	ADDSYM-EQL	ABSPslScan	ASYM-VIEW	FCF2HKL
NewmanPlot	Calc Inter	SQUEEZE	ADDSYM-EXT	ABSTempa	FCF-Valid	Expand2P1
Ring-Plots	Calc Coord	CALC-FCF	ADDSYM-PLT	ABSGauss	Dl fFouler	FCF-Gener
Plane-Plot	Calc Metal	Contour-SQ	ADDSYM-SHX	ABSXtal	ANALofVAR	HKL-Gener
Polyhedra	Calc Geom	SOLV F3D	NEWSYM	ABSSphere	ByvoetPatr	HKL-Transf
ContourDf	Calc Hbond	SOLV PLOT	NONSYM	SHXABS	ASYM-EXPCT	EXOR-RES
Contour-Fo	Calc TMA	CavlttyPlot	LePage		ASYM-Valld	ANIS-RES
AutoMolFlt	L.S.-PLANE		DelRed		SupplMater	Rename-RES
HKL2Powder	DihedAngle		MOLSYM		EXPECT-HKL	Auto-Renum
SlmPowderP	AngleLines	FLIP MENU	SPGRfromEX		CSD-CELL	SPF -eld
RadDlstFun	AngLsplLLn	FLIP SHOW	ASYM		CSD-QUEST	SHELXL-res
Patterson	CremerPopl	FLIP PATT	ASYMaverFR		StructTidy	CIF -acc
	BondValenc	FLIPPER 25	LePageTwn		StralnAnal	PDB -pdb
PLUTONatl v	HFIX - RES	STRUCTURE?	TwnRotMat	Xtal Habit	LocCIF-acc	CIF2SHELXL

Xtal Data (CIF ) vltac.cif- Set 1 ( 1): vltac  
 Refl Data (SHELXL ) vltac.fcf [ NO-DIRC ] ( 1): vltac

<http://www.crysl.chem.uu.nl/epk/platon/PLATON-MANUAL.pdf>

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

Browser - HELP

PLATON 10  
 OptionMenu  
 NoMove  
 Join-Expand  
 Organic  
 Round  
 Parentheses  
 Label-Alias  
 R/S-Determ  
 Norm-H-bond  
 NoSymm  
 NoDisorder  
 ListARU RCell  
 ListCellSymm  
 ListAtoms  
 ListBonds  
 ListFlagRadi  
 Exclude H  
 MinQPeakHgt  
 MinQPeakDis  
 Q-Peak-Incl  
 KeyInstruct  
 Prev Next  
 SAVE-InstrS  
 ENTRY-LIST  
 Reset End  
 Exit  
 MenuActive



# PLATON - CALC ALL

- **Check of Final Crystal Structure Refinement**

- Unit Cell and Space Group Validation

- Analysis of hydrogen bonding

- Calculation of least-squares planes

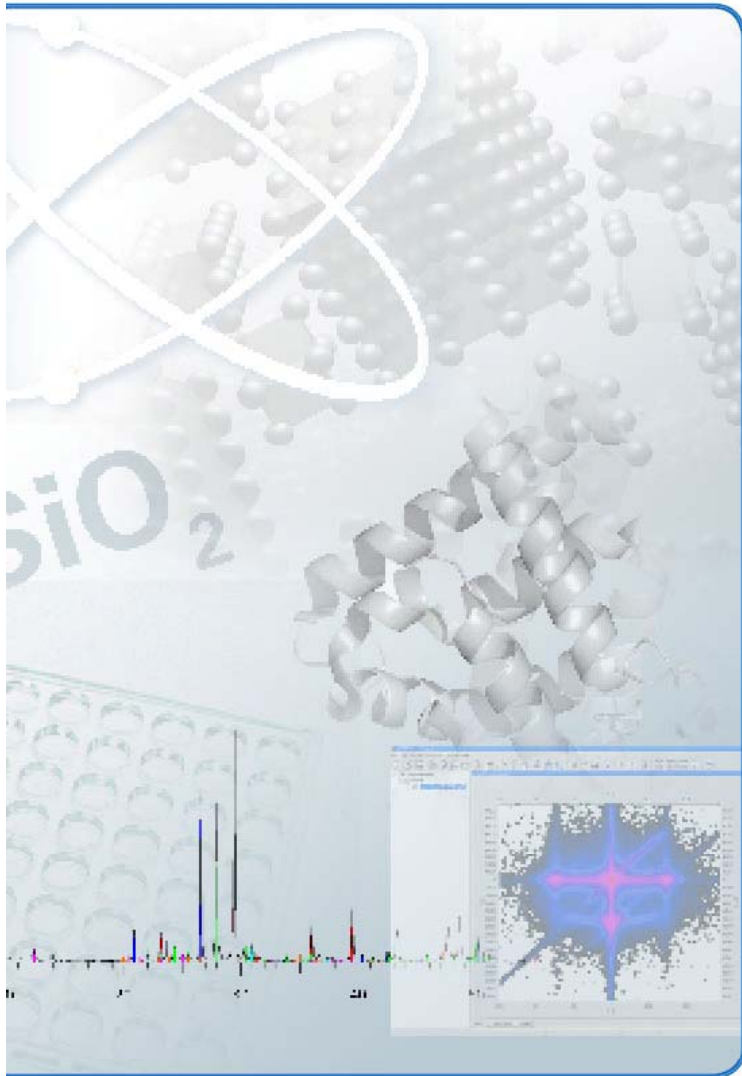
- Analysis of solvent accessible voids

- Hirshfeld analysis

- Librational analysis

- **Automatic generation of an extensive molecular geometry analysis report**

# Crystal Structure Reports





# Typical Crystal Structure Report

## ■ Data Collection

- Source of sample and conditions of crystallization.
- Habit, color, and dimensions of the crystal.
- Formula and formula weight.
- Unit cell parameters and volume with esd's. The number of data and  $\theta$  range of data used to determine the cell parameters.
- Crystal type and space group.
- Z, density, and linear absorption coefficient.
- Instrument and temperature of data collection and cell parameter determination.
- # of data collected, # unique [ $R(int)$ ].
- Absorption correction details.

## ■ Structure Solution

- Method and program(s) used for structure solution.



# Typical Crystal Structure Report

## ■ Structure Refinement

- Method and program(s) for refinement.
- # of data refined, # restraints, # parameters.
- Weighting scheme.
- $R_1$ (observed data),  $wR_2$ (all data), and  $S$  values.
- Final maximum absolute value of the shift/error.
- Maximum and minimum of the final difference electron density map.

## ■ Tables and Figures

- Positional parameters and isotropic or equivalent displacement parameters.
- Bond distances and angles.
- Anisotropic displacement parameters.
- Structure factor tables (often required for review but discarded by the journal).
- Torsion angles(optional).
- Least-squares planes(optional).
- Hydrogen bond geometry(optional).
- A labeled figure showing the displacement ellipsoids.
- A packing diagram showing relevant intermolecular interactions.



# APEX2 – Generate Report Option (HTML)

Report | Atoms | Bond lengths | Bond angles | Torsion angles | Hydrogen bonds | Authors | Miscellaneous

## Crystal Structure Report for YLID4

A translucent pale yellow sphere-like specimen of  $C_{11}H_{10}O_2S$ , approximate dimensions 0.350 mm x 0.350 mm x 0.350 mm, was used for the X-ray crystallographic analysis. The X-ray intensity data were measured on a Bruker Kappa APEX II CCD system equipped with a Triumph curved crystal monochromator and a Mo fine-focus tube ( $\lambda = 0.71073 \text{ \AA}$ ).

### Table 1: Data collection details for YLID4.

Axis	dx/mm	2 $\theta$ /°	$\omega$ /°	$\phi$ /°	$\chi$ /°	Width/°	Frames	Time/s	Wavelength/Å	Voltage/kV	Current/mA	T <sub>exp</sub> /s
Omega	59.941	-28.00	332.00	0.00	54.78	-0.50	366	30.00	0.71073	50	30.0	29
Omega	59.941	-28.00	332.00	90.00	54.78	-0.50	366	30.00	0.71073	50	30.0	29
Omega	59.941	-28.00	332.00	180.00	54.78	-0.50	366	30.00	0.71073	50	30.0	29
Omega	59.941	-28.00	332.00	270.00	54.78	-0.50	366	30.00	0.71073	50	30.0	29

A total of 1464 frames were collected. The total exposure time was 12.20 hours. The frames were integrated with the Bruker SAINT software package using a narrow-frame algorithm. The integration of the data using an orthorhombic unit cell yielded a total of 11859 reflections to a maximum  $\theta$  angle of  $27.61^\circ$  (0.77 Å resolution), of which 2306 were independent (average redundancy 5.143, completeness = 99.9%,  $R_{\text{int}} = 1.34\%$ ,  $R_{\text{sig}} = 0.88\%$ ) and 2271 (98.48%) were greater than  $2\sigma(F^2)$ . The final cell constants of  $a = 5.9641(2) \text{ \AA}$ ,  $b = 9.0422(3) \text{ \AA}$ ,  $c = 18.4015(6) \text{ \AA}$ , volume =  $992.37(6) \text{ \AA}^3$ , are based upon the refinement of the XYZ-centroids of 9708 reflections above  $20 \sigma(I)$  with  $4.426^\circ < 2\theta < 55.21^\circ$ . Data were corrected for absorption effects using the multi-scan method (SADABS). The ratio of minimum to maximum apparent transmission was 0.937. The calculated minimum and maximum transmission coefficients (based on crystal size) are 0.9040 and 0.9040.

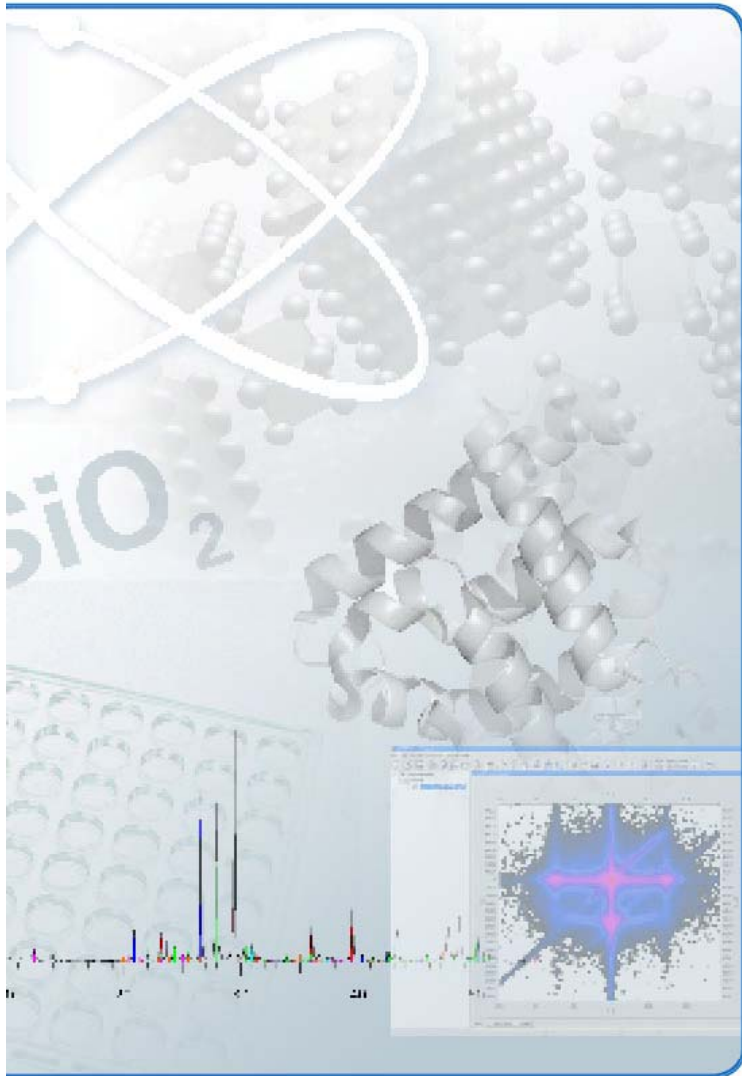
The structure was solved and refined using the Bruker SHELXTL Software Package, using the space group  $P 2_1 2_1 2_1$ , with  $Z = 4$  for the formula unit,  $C_{11}H_{10}O_2S$ . The final anisotropic full-matrix least-squares refinement on  $F^2$  with 168 variables converged at  $R_1 = 2.30\%$ , for the observed data and  $wR_2 = 6.43\%$  for all data. The goodness-of-fit was 1.075. The largest peak in the final difference electron density synthesis was  $0.212 \text{ e}/\text{\AA}^3$  and the largest hole was  $-0.155 \text{ e}/\text{\AA}^3$  with an RMS deviation of  $0.034 \text{ e}/\text{\AA}^3$ . On the basis of the final model, the calculated density was  $1.380 \text{ g}/\text{cm}^3$  and  $F(000)$ , 432 e<sup>-</sup>.

Template File: C:\bn\src\report\apex2.zip

Template Sections:

- Run list
- Structure view (in saved report only)
- Packing view (in saved report only)
- Sample and crystal data
- Data collection and structure refinement
- Atomic coordinates
- Bond lengths
- Bond angles
- Torsion angles
- Anisotropic displacement parameters
- Hydrogen atomic coordinates
- Hydrogen bonds

# Preparation of CIF and PDB Files





# Crystallographic Information Framework

- The International Union of Crystallography is the sponsor of the **Crystallographic Information Framework**, a standard for information interchange in crystallography.
- The acronym **CIF** is used both for the *Crystallographic Information File*, the data exchange standard file format of Hall, Allen & Brown (1991), 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.
- The standard reference describing the Crystallographic Information Framework:  
*International Tables for Crystallography Volume G: Definition and exchange of crystallographic data* (2005), edited by Sydney Hall & Brian McMahon.  
Dordrecht: Springer.

# APEX2 – Generate Report Option (Acta CIF)



Report Atoms Bond lengths Bond angles Torsion angles Hydrogen bonds Authors Miscellaneous

_cell_length_a	5.9641(2)
_cell_length_b	9.0422(3)
_cell_length_c	18.4015(6)
_cell_angle_alpha	90
_cell_angle_beta	90
_cell_angle_gamma	90
_cell_volume	992.37(6)
_cell_formula_units_Z	4
_cell_measurement_reflns_used	9708
_cell_measurement_theta_min	2.2130
_cell_measurement_theta_max	27.6053
_cell_measurement_temperature	296.(2)
_exptl_crystal_description	sphere
_exptl_crystal_colour	'translucent pale yellow'
_exptl_crystal_size_max	0.350
_exptl_crystal_size_mid	0.350
_exptl_crystal_size_min	0.350
_exptl_crystal_density_diffrn	1.380
_exptl_crystal_density_meas	?
_exptl_crystal_density_method	'not measured'
_exptl_crystal_F_000	432
_exptl_absorpt_coefficient_mu	0.294
_exptl_absorpt_correction_type	multi-scan
_exptl_absorpt_process_details	'SADABS V2008/1 (Bruker AXS)'
_exptl_absorpt_correction_T_min	0.85
_exptl_absorpt_correction_T_max	0.90
_exptl_special_details	
;	
:	

Template File:  
C:\bn\src\report\acta.cif

Template Sections:  
No sections available.



# PDB File Format and mmCIF Format

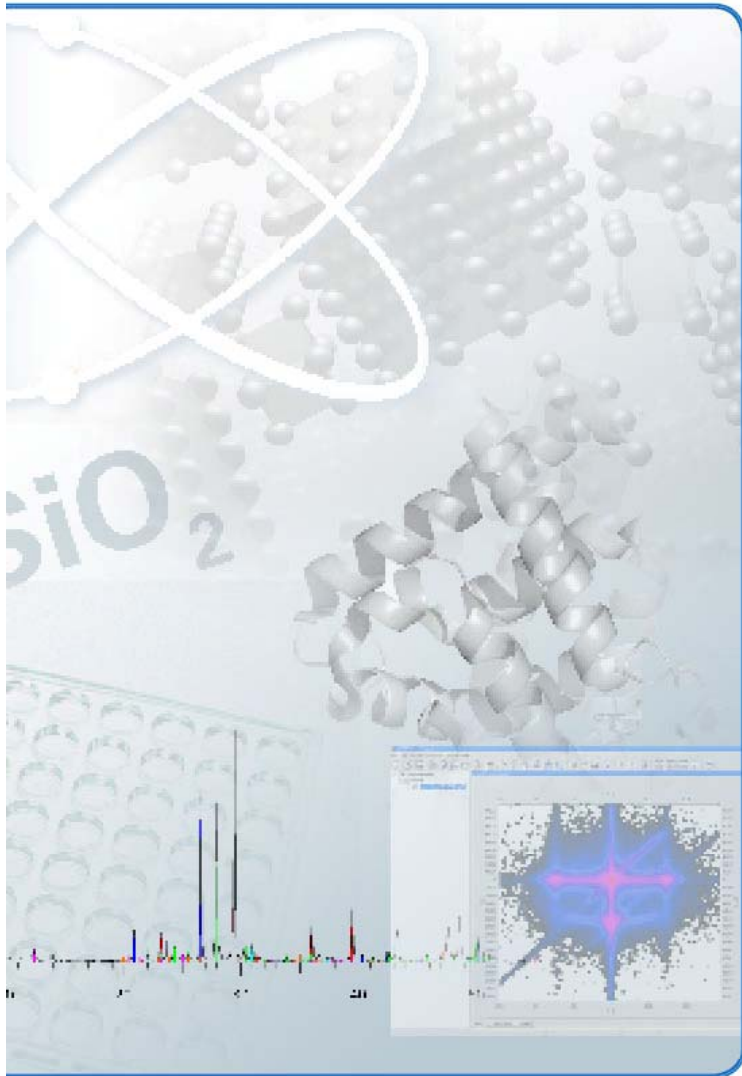
## ■ PDB File Format

- The Protein Data Bank (PDB) format provides a standard representation for macromolecular structure data derived from X-ray diffraction and NMR studies. This representation was created in the 1970's and a large amount of software using it has been written.
- Documentation describing the PDB file format is available from the wwPDB at: <http://www.wwpdb.org/docs.html>.

## ■ mmCIF File Format and PDB Exchange Dictionary

- The Protein Data Bank (PDB) uses macromolecular Crystallographic Information File (mmCIF) data dictionaries to describe the information content of PDB entries.
- Further information and related resources are available at: <http://mmcif.pdb.org/>.

# Use of Validation Programs



# checkCIF Tools

- CIF dictionary  
<http://www.iucr.org/resources/cif/dictionaries>
- Details of checkCIF/PLATON tests  
<http://journals.iucr.org/services/cif/datavalidation.html>
- Download CIF editor (pubCIF) from the IUCr  
<http://journals.iucr.org/services/cif/pubcif/>
- Download CIF editor (enCIFer) from the CCDC  
[http://www.ccdc.cam.ac.uk/free\\_services/encifer/downloads/enCIFer\\_1.4/download.php4](http://www.ccdc.cam.ac.uk/free_services/encifer/downloads/enCIFer_1.4/download.php4)
- Full publication check  
<http://journals.iucr.org/services/cif/checking/checkfull.html>



# checkCIF Reports

- The automated report contains three types of alerts:
  - ALERT level A = In general: serious problem**
  - ALERT level B = Potentially serious problem**
  - ALERT level C = Check and explain**
- It is impossible to explain all possible checkCIF errors here. If you use the checkCIF routine on the site [www.iucr.org](http://www.iucr.org), you can click on the alert to obtain more information.
- If you not understand a checkCIF error, seek advice from a crystallographer. **Do not ignore a checkCIF error.**
- Try to eliminate all problems. If not, comment on the error, explaining where this alert is coming from and why we can ignore it or cannot do anything about it.

# PLATON checkCIF Report

## checkCIF/PLATON report (publication check)

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

[CIF dictionary](#)  
[Interpreting this report](#)

### Datablock: sucrose

```

Bond precision:      C-C = 0.0031 A           Wavelength=0.71073
Cell:               a=7.7110(2)  b=8.6597(2)  c=10.8032(2)
                   alpha=90      beta=102.9843(13) gamma=90
Temperature: 90 K

          Calculated           Reported
Volume      702.94(3)          702.94(3)
Space group  P 21              P 21
Hall group   P 2yb             P 2yb
Moiety formula C12 H22 O11      C12 H22 O11
Sum formula  C12 H22 O11      C12 H22 O11
Mr           342.30            342.30
Dx,g cm-3    1.617             1.617
Z            2                  2
Mu (mm-1)    0.145             0.145
F000         364.0             364.0
F000'        364.28
h,k,lmax     10,11,14          9,11,14
Nref         1701[ 3190]       1698
Tmin,Tmax    0.974,0.983       0.972,0.983
Tmin'        0.971
Correction method= MULTI-SCAN
Data completeness= 1.00/0.53  Theta(max)= 27.450
R(reflections)= 0.0311( 1557)  wR2(reflections)= 0.0715( 1698)
S = 1.066           Npar= 217
  
```

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

```

STRVA01_ALERT_4_C      Flack test results are meaningless.
                       From the CIF: _refine_ls_abs_structure_Flack 0.100
                       From the CIF: _refine_ls_abs_structure_Flack_su 1.000
PLAT089_ALERT_3_C      Poor Data / Parameter Ratio (Zmax .LT. 18) ..... 7.82
PLAT417_ALERT_2_C      Short Inter D-H..H-D      H8A .. H9A .. 2.13 Ang.
PLAT032_ALERT_4_C      Std. Uncertainty in Flack Parameter too High ... 1.00
PLAT153_ALERT_1_C      The su's on the Cell Axes are Equal (x 100000) 20 Ang.
  
```



# Full Publication Check (IUCr)

- **On-Line checkCIF:**

<http://checkcif.iucr.org/>

- **This version of checkCIF includes checks on:**

- CIF syntax and construction
- Cell and geometry details
- Space-group symmetry
- Anisotropic displacement parameters
- Structure factors

# Comments on checkCIF Reports

## `_refine_special_details`

```
;  
Refinement of  $F^2$  against ALL reflections. The weighted R-factor wR and  
goodness of fit S are based on  $F^2$ , conventional R-factors R are based  
on F, with F set to zero for negative  $F^2$ . The threshold expression of  
 $F^2 > 2\sigma(F^2)$  is used only for calculating R-factors(gt) etc. and is  
not relevant to the choice of reflections for refinement. R-factors based  
on  $F^2$  are statistically about twice as large as those based on F, and R-  
factors based on ALL data will be even larger.
```

Comments on remaining CHECKCIF errors:

Hirshfield test violations for atoms C24, C25 and C26 are caused by a  
strong vibration of a complete phenyl substituent.

The large difference between  $U_{eq}(max)/U_{eq}(min)$  is caused by the presence  
of strongly vibrating phenyl groups and strongly bound Cp atoms in the  
structure.

```
;
```

# Example of checkCIF Report

The following ALERTS were generated.

Each ALERT has the format `test-name_ALERT_alert-type_alert-level`.


Click on the hyperlinks for more details of the test.

## Alert level A

[PLAT761\\_ALERT\\_1\\_A](#) CIF Contains no X-H Bonds ..... ?

[PLAT762\\_ALERT\\_1\\_A](#) CIF Contains no X-Y-H or H-Y-H Angles ..... ?

Redo the refinement  
with BOND \$H instead  
of BOND



## Alert level B

[PLAT230\\_ALERT\\_2\\_B](#) Hirshfeld Test Diff for C24 - C25 .. 7.36 su

## Alert level C


[CRYSC01\\_ALERT\\_1\\_C](#) The word below has not been recognised as a standard identifier. : plates

[CRYSC01\\_ALERT\\_1\\_C](#) No recognised colour has been given for crystal colour.

[PLAT222\\_ALERT\\_3\\_C](#) Large Non-Solvent H Ueq(max)/Ueq(min) ... 3.08 Ratio

[PLAT230\\_ALERT\\_2\\_C](#) Hirshfeld Test Diff for C25 - C26 .. 5.93 su

Correct manually  
in the CIF file



`.cif:`

`_exptl_crystal_description`

`red`

`_exptl_crystal_colour`

`plates`



# Production of Acta E Preprint with publCIF

```
1 data_global
2
3
4 _publ_section_exptl_refinement
5 ;
6 (type here to add refinement details)
7 ;
8 #Added by publCIF
9
10
11
12
13 _publ_section_exptl_prep
14 ;
15 The title compound was prepared as illustrated in Fig. 5. An oven-dried glass
16 pressure vessel containing a magnetic stir bar was charged with palladium(II)
17 acetate (0.034 g, 0.152 mmol), 1,2-dibromo-4,5-dimethylbenzene (2 g, 7.58
18 mmol), 2-dicyclohexylphosphino-2',6'-dimethoxybiphenyl (0.124 g, 0.303 mmol),
19 phenylboronic acid (2.77 g, 22.7 mmol) and powdered, anhydrous potassium
20 phosphate (11.26 g, 53.0 mmol). Dry THF (20 ml) was added and N2 gas was
21 bubbled through the resulting mixture for 15 min. The glass pressure vessel
22 was sealed with a Teflon cap and heated at 75 °C for 20 h with stirring. The
23 reaction mixture was allowed to cool to room temperature after which the
24 mixture was diluted with diethyl ether (30 ml) and washed with water three
25 times. The organic layer was dried over magnesium sulfate and concentrated at
26 reduced pressure. The crude product was purified by flash column
27 chromatography on silica gel using hexane-chloroform (80/20 v/v) as
28 eluent. The title compound, 1,2-dimethyl-4,5-diphenylbenzene, was obtained in
29 76% isolated yield. 1H NMR (400 MHz, CDCl3) δ 2.35 (s, 6H), 7.11-7.21 (m,
30 12H); 13C NMR (100 MHz, CDCl3) δ 19.6 (CH3), 126.4 (CH), 128.0 (CH),
31 130.1 (CH), 132.1 (CH), 136.0 (C), 138.2 (C), 141.7 (C). An X-ray grade
32 crystal was obtained by slow evaporation of a dichloromethane solution.
33 ;
34 #Added by publCIF
35
36
37
```

**Acta E preprint**

**1,2-Dimethyl-4,5-diphenylbenzene determined on a Bruker SMART X2S benchtop crystallographic system**

**Jonathan B. Briggs, Mikael D. Jazdyk and Glen P. Miller\***

Department of Chemistry and Materials Science Program, University of New Hampshire, Durham, New Hampshire, 03824-3598, USA

Correspondence e-mail: [glen.miller@unh.edu](mailto:glen.miller@unh.edu)

**Abstract**

The title compound, C<sub>20</sub>H<sub>18</sub>, has triclinic (*P*<sub>1</sub>) symmetry and two molecules in the unit cell. The phenyl substituents at C4 and C5 of molecule A are twisted away from the plane defined by the central benzene ring by 131.9 (2)° and -52.7 (2)°, respectively. The phenyl substituents at C4 and C5 of molecule B are twisted by 48.1 (2)° and -130.5 (2)°, respectively. The crystal structure is stabilized by a pair of intramolecular C(aryl, *sp*<sup>2</sup>)-H···π interactions as well as several intermolecular C(methyl, *sp*<sup>3</sup>)-H···π interactions.

# Submission of Manuscripts



# Choice of Journals

- IUCr Journals ([Acta Crystallogr. A](#), [B](#), [C](#), [D](#), [E](#); [J. Appl. Cryst.](#))
- ACS Journals ([J. Am. Chem. Soc.](#), [Organometallics](#), [Inorg. Chem.](#), [J. Org. Chem.](#), [Biochemistry](#))
- International Journals ([Angew. Chem.](#), [Inorg. Chim. Acta](#), [J. Chem. Crystallogr.](#))
- Mineralogy Journals ([Am. Mineral.](#), [Can. Mineral.](#) )
- General Scientific Journals ([Science](#), [Nature](#))



# Publication Guidelines

- When preparing a crystal structure report for publication, be sure to check the Notes for Authors section of the selected journal.
- Authors unfamiliar with preparing crystal structure reports should look through recent issues of the target journal to find articles with similar crystal structures.
- Emulating the format of the crystal structure reports in these articles will greatly reduce the effort needed.



# Acknowledgements and Co-authorship

- There are no 'hard rules' regarding co-authorship versus simple acknowledgements for individuals who have helped with a crystal structure determination. However, the following is accepted practice:
  - If the structure was routine and you have carried out most of the structure determination yourself, but have received advice from a staff crystallographer or another individual, a simple acknowledgement is sufficient.
  - If the structure was non-routine (e.g., twinning, disorder, etc.) and the final structure involved a substantial amount of time and technical expertise from another individual, he (she) should be included as a co-author.
- All co-authors must be given an opportunity to review the manuscript and they must agree to co-authorship prior to manuscript submission.
- All financial support from sponsoring institutions and funding agencies must be acknowledged in the paper.



# Crystallographic Archives

When a crystal structure analysis has been completed, it must be archived for a minimum of 5 years. A proper archive should allow another individual to recover relevant data, even in your absence.

This archive must include:

- A copy of the final report
- Final solution and refinement files (e.g., .ins, .res, .hkl, .cif, .lst files)
- Experimental results (images etc)



# Summary

# Review and Summary



## Advanced Crystallography - Publication of Crystal Structures

- Collection of high-quality intensity data
- Common refinement problems
- Data evaluation and error analysis
- Report generation and display of crystallographic results
- Preparation of Crystallographic Information Framework (CIF) files
- The use of PubCIF and other CIF-checking programs
- Submission of manuscripts



# Acknowledgements



# Acknowledgements

- Prof. Frank Schaper, U. of Montreal
- Dr. Doug Powell, U. of Oklahoma
- Dr. Peter Mueller, MIT





# Where can I obtain help to publish my structures?

## ■ General users:

- Primary support should come from the site where the data was collected
- Attend Workshops and Summer Schools (e.g., ACA Summer School, Canadian Chemical Crystallography Workshop)
- Visit websites of departmental crystallographers (e.g., U of Oklahoma, MIT, U of Kentucky, etc.)

## ■ For Bruker users:

- Join the Bruker Users' Group (email [Iguzei@chem.wisc.edu](mailto:Iguzei@chem.wisc.edu))
- Visit [www.Bruker-AXS.com](http://www.Bruker-AXS.com) and request information
- Attend Bruker workshops and User Meetings
- Email to me at [Charles.Campana@Bruker-AXS.com](mailto:Charles.Campana@Bruker-AXS.com)