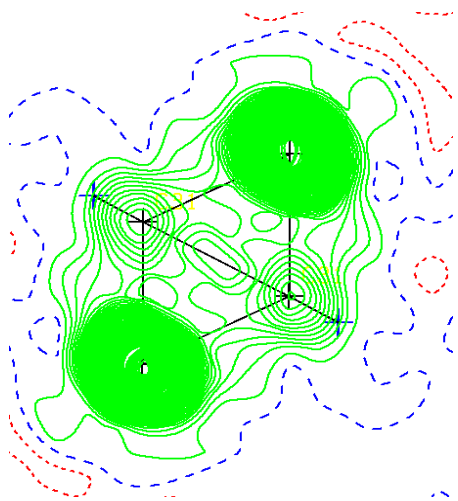
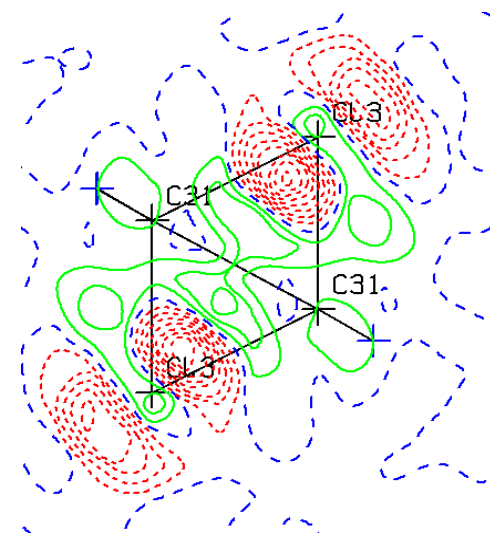


CIF, PLATON-2014, SHELXL-2014, VALIDATION & SQUEEZE



Ton Spek
Utrecht University
The Netherlands



The CIF Data Exchange Standard

- CIF already exists for a quarter of a century
- It was created around 1990 by Syd Hall et al.
- CIF provides a standardized **computer readable intermediate** between structure refinement software and software that makes use of the resulting structural model (Graphics, Geometry etc.) and data collection and refinement data
- In essence a CIF file consists of a set of records with associations between properly defined **data names** and their associated **values** (free format)

EARLY ADOPTERS OF CIF

Refinement Programs

- The, at that time, widely but now rarely used XTAL system program suite (Hall et al.)
- SHELXL92 (The beta follow-up of SHELX-76)

CIF Users

- IUCr/AC-C (Publication, Archival, Checking)
- CCDC/Encifer (Archival/No Retyping)
- Acta: PLATON/Validation (VOIDS/ADDSYM)

The Current Status of CIF

- Most current software can read and/or write CIF
- Hard to keep-up with new CIF items due to new science, changing data collection and refinement techniques (e.g. detectors, constraints/restraints)
- New **official** data names have been defined for powder studies, twinning & constraints/restraints but not generally implemented in existing software
- In contrast, many more **unofficial** data names are currently in use starting with `_shelx_`, `_olex2_` etc.
- Regrettably, up to recently, primary reflection data were not generally included in deposited CIF's

CIF-ISSUES

- Unfortunately, The CIF syntax has its problems (no end loop, no loop nesting etc.)
- The realization that when a CIF that does not include the underlying (primary) reflection data, the information is incomplete and thus poor science.
- Many questions about a structural study can be answered only with the availability of reflection data.
- The reflection data may be unique (or hard to obtain) and possibly needed for follow-up studies.
- The historical practice of Fo/Fc deposition is often unsatisfactory (data averaged, 'extinctions' deleted)

Current Solution

- The refinement program specific **instruction** and **reflection** files are now soon expected to be at least **embedded** as comment (under their proper data names) in the CIF
- SHELXL2014 **automatically embeds** those data with proper checksums for transmission errors etc..
- The **embedded data** in a SHELXL2014 CIF can be easily **extracted** and used for alternative refinement purposes or for the creation of an FCF where needed
- PLATON/SQUEEZE makes extensive use of the above

PLATON Tools

Interface: name.cif, .fcf., .ins, .res, .hkl

- Validation - 'platon -u name.cif'
- SQUEEZE - 'platon -q name.cif'
- ORTEP - 'platon -a name.cif'
- ADDSYM - 'platon -m name.cif'
- Etc.
- GUI - 'platon name.cif' for CALC, H-Bonds, TwinRotMat, Bijvoet etc.

PLATON

A Multipurpose Crystallographic Tool

(C) 1980-2014 A.L.Spek - Version: 270714 [WEB: Jul 24, 2014]

GRAPHICS	GEOM-CALC	VOIDS FLIP	SYMMETRY	ABSORPTION	REPORT	MISC-TOOLS
PlutonAuto	Calc All	Calc Solv	Addsym	MULscanABS	Validation	System-S
Ortep-Plot	Calc Intra	Calc K.P.I	Addsym-EQL	ABSPslScan	Asym-Vlew	fcf2hkl
NewmanPlot	Calc Inter	Squeeze	Addsym-EXT	ABSTempa	FCF-Valid	Expand2P1
Ring-Plots	Calc Coord	Hybrid	Addsym-PLT	ABSGauss	DlfFourier	FCF-Gener
Plane-Plot	Calc Metal	CalcFCFsqd	Addsym-SHX	ABSXtal	ANALofVAR	HKL-Gener
Polyhedra	Calc Geom	Contoursqd	Newsym	ABSSphere	ByvoetPair	HKL-Transf
ContourDlf	Calc Hbond	Solv F3D	Nonsym	ShxAbs	Asym-Expt	Exor-Res
Contour-Fo	Calc TMA	Solv Plot	LePage	AnomDlsVal	Asym-Valid	Anls-Res
AutoMolFlt	L.S.-Plane	CavityPlot	DelRed	AnomDlsPlt	SupplMater	Rename-RES
hkl2Powder	DihedAngle		Molsym	MuPlot	Expect-hkl	Auto-Renum
SimPowderP	AngleLines	Flip Menu	SPGRfromEX		CSD-Cell	Create-spf
RadDistFun	AngLsplLin	Flip Show	Asym		CSD-Quest	Create-res
Patterson	CremerPopl	Flip Patt	ASYMaverFR		StructTidy	Create-clf
ShelxtPlot	BondValenc	Flipper 25	LePageTwln	XtlPlanAgl	StrainAnal	Create-pdb
PlutonAtlv	Hflx - res	Structure?	TwlnRotMat	Xtal Hablt	LocCIF-acc	clf2shelxl

Xtal Data (CIF14) sl2349_shelxl.clf - Set 1(1): sl2349_s
 Refl Data (LIST4) sl2349_shelxl.fcf [FCF] (1): sl2349_s

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

Browser - HELP

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

>>

PLA-SP2 12

OptionMenu

Print-Level

EPS-Listing

PDF-Listing

PageHeader

NoMolFitInv

KeepMon-I-n

NoExpand

FCF-Calc

PNG

ReflListing

DebugOutput

SetWinSize

Portrait

GenerRandom

EPS HPGL

Auto-Plot

X-LineWidth

Reverse-B&W

Browser

Reset End

Exit

MenuActive

CIF-Validation

- ALERTS come in levels A,B,C & G
- All A,B,C ALERTS should be looked at carefully and acted upon (correct or comment)
- G-ALERTS are info and not necessarily Errors
- The use of constraint/restraints should be commented upon. Their use, just to beautify an ORTEP to hide problems, is poor practice
- IUCr standard is best attainable, not 25 deg
- Next a validation output example based on a SHELXL2014 CIF →

```

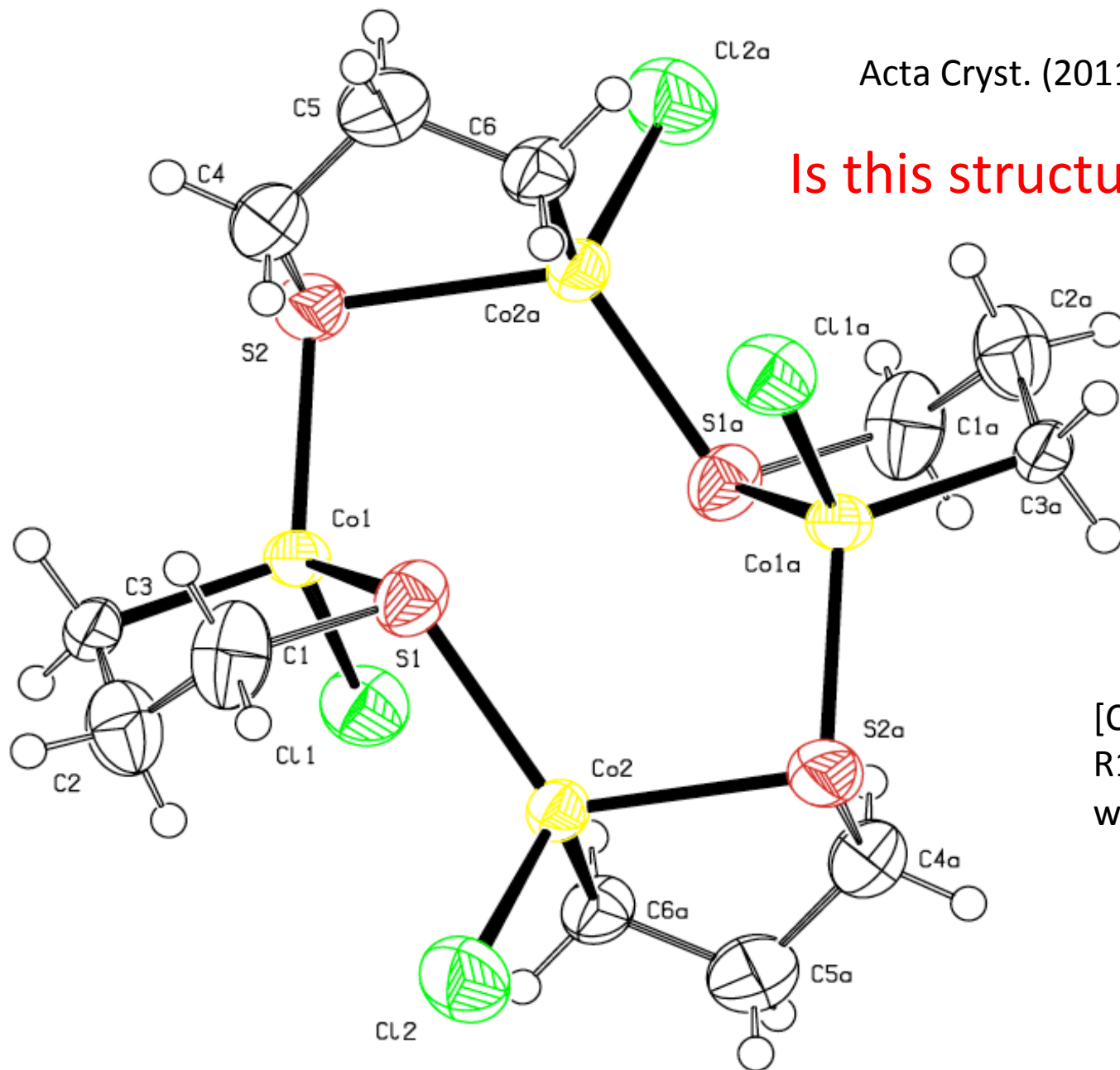
#=====
# PLATON/CHECK-(270714) versus check.def version of 240714 for Entry: si2349_s
# Data: si2349_shelxl.cif - Type: CIF14 Bond Precision C-C = 0.0113 A
# Refl: si2349_shelxl.fcf - Type: LIST4 Temp = 296 K
# X-Ray R(int) = 0.058, wR2/R(int) = 2.1, Nref/Npar = 19.7
# Cell 23.6135(12) 7.8465(3) 16.8693(9) 90 130.440(4) 90
# Wavelength 0.71073 Volume Reported 2378.8(2) Calculated 2378.9(2)
# SpaceGroup from Symmetry C 2/c Hall: -C 2yc monoclinic
# Reported C 2/c -C 2yc monoclinic
# MoietyFormula C12 H24 Cl4 Co4 S4
# Reported ?
# SumFormula C12 H24 Cl4 Co4 S4
# Reported C12 H24 Cl4 Co4 S4
# Mr = 674.07[Calc], 674.07[Rep]
# Dx,gcm-3 = 1.882[Calc], 1.882[Rep]
# Z = 4[Calc], 4[Rep]
# Mu (mm-1) = 3.537[Calc], 3.537[Rep]
# F000 = 1344.0[Calc], 1344.0[Rep] or F000' = 1354.05[Calc]
# Reported T Limits: Tmin=0.675 Tmax=0.683 AbsCorr=?
# Calculated T Limits: Tmin=0.512 Tmin'=0.424 Tmax=0.609
# Reported Hmax= 28, Kmax= 9, Lmax= 20, Nref= 2152, Th(max)= 25.242
# Obs in FCF Hmax= 28, Kmax= 9, Lmax= 20, Nref= 2152[ 2152], Th(max)= 25.242
# Calculated Hmax= 28, Kmax= 9, Lmax= 20, Nref= 2152, Ratio = 1.000
# Reported Rho(min) = -0.62, Rho(max) = 0.70 e/Ang**3 (From CIF)
# Calculated Rho(min) = -0.61, Rho(max) = 0.73 e/Ang**3 (From CIF+FCF data)
# w=1/[sigma**2(Fo**2)+(0.0653P)**2+ 15.5359P], P=(Fo**2+2*Fc**2)/3
# R= 0.0413( 1782), wR2= 0.1237( 2152), S = 1.029 (From CIF+FCF data)
# R= 0.0413( 1782), wR2= 0.1236( 2152), S = 1.029 (From FCF data only)
# R= 0.0413( 1782), wR2= 0.1236( 2152), S = 1.029, Npar= 109
#=====

```

232_ALERT_2_B	Hirshfeld Test Diff (M-X)	Co1	--	S1	..	10.1	su
232_ALERT_2_B	Hirshfeld Test Diff (M-X)	Co1	--	S2	..	10.3	su
232_ALERT_2_B	Hirshfeld Test Diff (M-X)	Co2	--	S1	..	11.3	su
232_ALERT_2_B	Hirshfeld Test Diff (M-X)	Co2	--	S2_a	..	11.3	su
#=====							
030_ALERT_1_C	_diffrn_reflns_number < _reflns_number_total						Please Check
048_ALERT_1_C	MoietyFormula Not Given						Please Do !
052_ALERT_1_C	Info on Absorption Correction Method Not Given .						Please Do !
057_ALERT_3_C	Correction for Absorption Required RT(exp) ...					1.19	Do !
230_ALERT_2_C	Hirshfeld Test Diff for C2 -- C3 ..					6.0	su
242_ALERT_2_C	Low Ueq as Compared to Neighbors for						Co2 Check
341_ALERT_3_C	Low Bond Precision on C-C Bonds					0.0113	Ang.
790_ALERT_4_C	Centre of Gravity not Within Unit Cell: Resd. #						1 Note
	C12 H24 Cl4 Co4 S4						
#=====							
083_ALERT_2_G	SHELXL Second Parameter in WGHT Unusually Large.					15.54	Report
128_ALERT_4_G	Alternate Setting for Input Space Group C2/c						I2/a Note
232_ALERT_2_G	Hirshfeld Test Diff (M-X) Co1 -- Cl1 ..					10.4	su
232_ALERT_2_G	Hirshfeld Test Diff (M-X) Co2 -- Cl2 ..					8.9	su
764_ALERT_4_G	Overcomplete CIF Bond List Detected (Rep/Expd) .					1.14	Ratio
909_ALERT_3_G	Percentage of Observed Data at Theta(Max) still					65	%
#=====							

Acta Cryst. (2011). E67, m576-m577

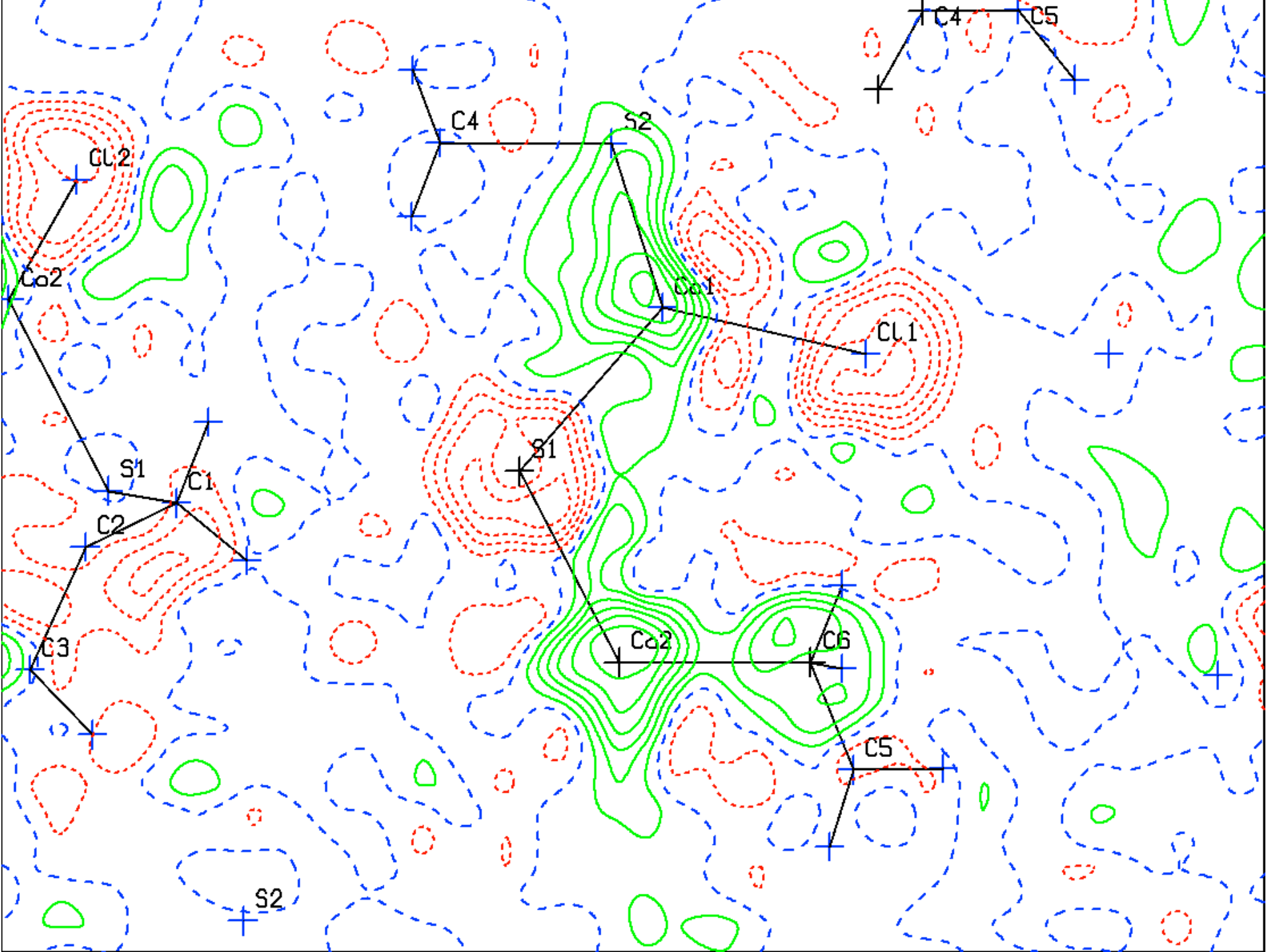
Is this structure correct ?

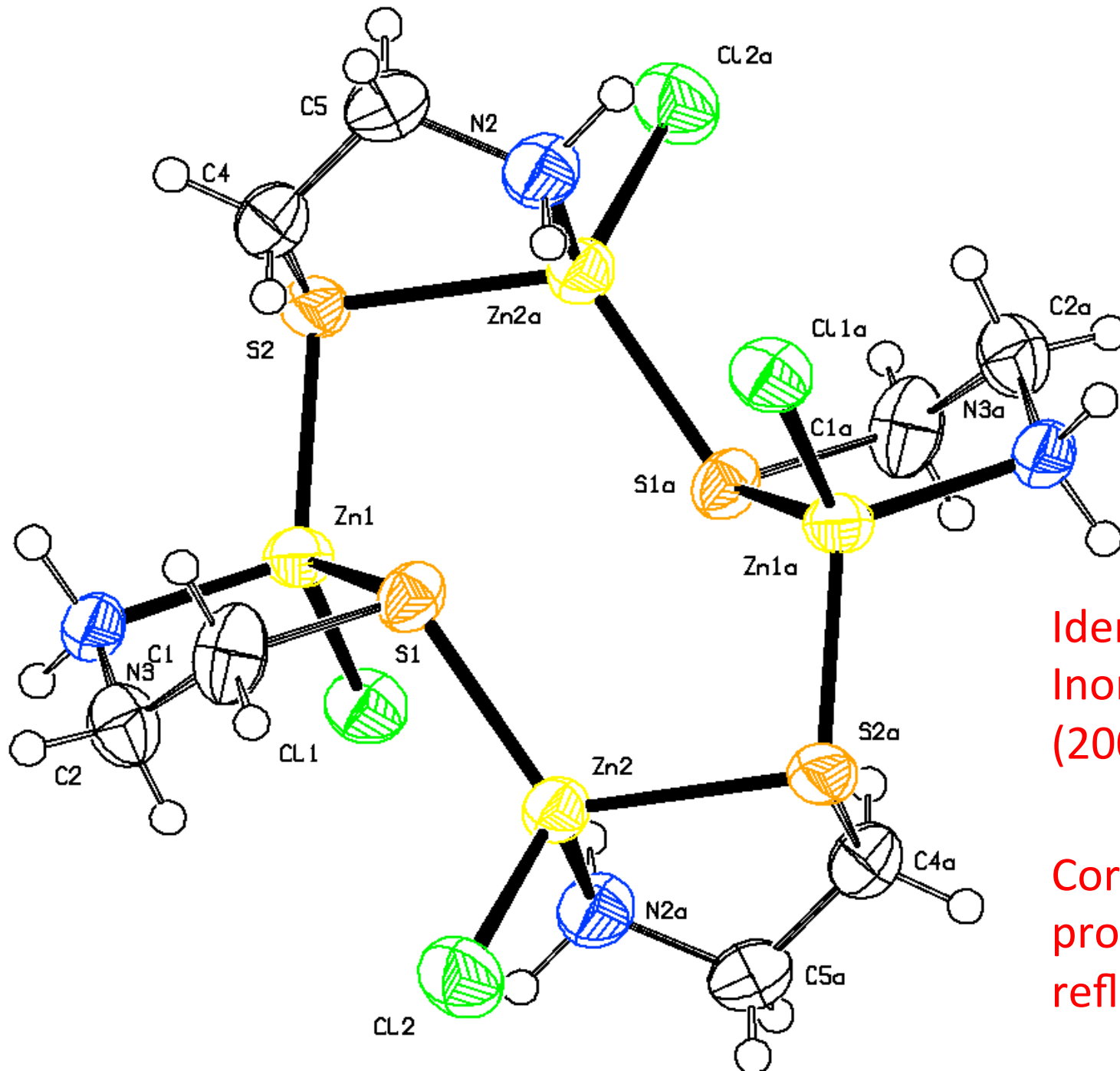


[Co₄Cl₄(C₃H₆S)₄]

R1 = 0.041

wR2 = 0.125





Co → Zn
C3 → N1
C6 → N2
R1 = 0.028
wR2 = 0.072

Identical to:
Inorg. Chem.
(2006) 45, 8318

Correction
proved with
reflection data

FCF-Validation

- Listing of outliers
- Listing of missing reflections
- %observed data versus resolution plot
- Analysis of Variance + Normal Prob. Plot
- TwinRotMat
- Bijvoet Pair Hooft/Parsons analysis
- Difference Fourier peak list (see whether peaks are near heavy atoms or elsewhere)

PLATON-(Version 270714)-Mode=2 FCF-File Validation for:I

For Documentation: <http://www.platonsoft.nl/FCF-VALIDATION.pdf>

Section 1

General Data

Crystal Data From: zl2039.cif
Fo/Fc Data From: zl2039.fcf FCF-TYPE=LIST4
Space Group : P21/n
Wavelength (Ang) : 0.71073
Unit Cell (CIF) : 8.5814 19.4260 10.1180 90.000 96.620 90.000
SHELX WGHT Pars. : 0.0336 0.0000

Section 2

Reflections with $\text{abs}((I(\text{obs}) - I(\text{calc})) / \text{SigW}(I)) > 3.0$

Nr	H	K	L	Theta	I(obs)	I(calc)	Sigma(I)	Ratio	SigW(I)	RatioW
1	8	6	0	20.56	89.20	155.04	20.77	-3.17	21.25	-3.10
2	4	7	0	12.14	102.88	64.92	11.25	3.37	11.55	3.29
3	-1	0	3	6.28	782.69	674.72	26.62	4.06	35.76	3.02
4	0	4	3	7.40	485.61	590.60	24.33	-4.32	30.67	-3.42
5	8	6	4	23.12	-11.16	29.85	13.62	-3.01	13.64	-3.01
6	4	13	5	20.48	528.19	400.54	39.70	3.22	42.40	3.01
7	1	15	7	22.03	94.71	177.93	26.92	-3.09	27.39	-3.04
8	6	2	9	25.37	723.80	1021.52	91.48	-3.25	96.59	-3.08
								Average =	-0.77	-0.79

Missing Reflections (Asym. Refl. Unit) below $\sin(\theta)/\lambda = 0.5$

Nr	H	K	L	$\sin(\theta)/\lambda$	Theta	I(calc)	I(calc)/I(max)
1	0	2	0	0.051	2.10*	23.09	0.00029
2	1	3	0	0.097	3.95	168.56	0.00209
3	-1	5	1	0.148	6.02	307.55	0.00381
4	6	10	1	0.443	18.37	2760.21	0.03422
5	-1	1	2	0.113	4.59	146.91	0.00182
6	0	1	2	0.103	4.19	1183.75	0.01468
7	-1	3	2	0.134	5.47	1582.16	0.01962
8	-4	14	3	0.446	18.49	30.64	0.00038
9	-5	14	4	0.492	20.47	2085.30	0.02585
10	-3	1	7	0.373	15.35	894.30	0.01109
11	-2	3	7	0.363	14.94	93.44	0.00116

** Note: I(max) is the maximum I(obs) encountered in the fcf-file **

Starred Reflections have a Theta below $\Theta(\text{Min}) = 2.28$

From CIF: $\Theta(\text{Min}) = 2.28$

Section 4:

Resolution & Completeness Statistics (Cumulative and Friedel Pairs Averaged)

Theta	$\sin(\theta)/\lambda$	Complete	Expected	Measured	Missing	
20.82	0.500	0.994	1748	1737	11	
23.01	0.550	0.994	2340	2327	13	
25.24	0.600	0.994	3035	3016	19	
26.01	0.617	0.994	3300	3280	20	ACTA Min. Res. ---

Section 5

=====
R-Value Statistics as a Function of Resolution (in Resolution Shell)
=====

Theta	sin(Th)/L	#	R1	wR2	S	Rs	av(I/SigW)	av(I)	av(SigW)
12.38	0.302	375	0.018	0.050	1.136	0.013	21.16	5080.87	189.10
15.68	0.380	392	0.019	0.049	0.917	0.025	16.13	1900.87	82.64
18.02	0.435	378	0.025	0.061	0.953	0.039	13.10	1223.39	65.58
19.90	0.479	382	0.031	0.074	0.964	0.056	10.62	867.64	58.41
21.51	0.516	389	0.036	0.086	0.946	0.069	8.47	681.53	54.41
22.94	0.548	387	0.046	0.113	0.944	0.104	6.29	415.97	46.59
24.22	0.577	382	0.059	0.143	0.971	0.135	5.20	322.79	45.47
25.40	0.603	379	0.064	0.161	0.956	0.158	4.38	278.71	45.43
26.01	0.617	216	0.056	0.159	0.815	0.196	3.73	233.41	46.74

R(sig) = sum(sig(I)) / sum(I) = 0.0383

From FCF: R1 = 0.0307(2523), wR2 = 0.0711(3280), S = 1.001
From CIF: R1 = 0.0308(2523), wR2 = 0.0711(3280), S = 1.001, Npar = 229
No (SHELXL) Optimized Weights: wR2 = 0.0474 , S = 1.385

Intensity Distribution [Decay of I/Sigma(I) versus sin(theta)/lambda]

```

=====
sh  st/l   Ang    #  0.25  1.0   2.0  Percent  Distr. for I .gt. 2.0 * sig(I)
=====
 1  0.301  1.661   375  98.9  97.1  95.7  *****..
 2  0.379  1.318   389  96.1  92.8  88.9  *****...
 3  0.434  1.152   371  97.0  92.2  85.2  *****....
 4  0.478  1.046   382  95.8  89.0  81.7  *****...
 5  0.515  0.971   395  92.2  86.3  75.4  *****...
 6  0.547  0.914   370  90.0  82.2  69.7  *****...
 7  0.576  0.868   382  91.6  80.4  69.6  *****...
 8  0.602  0.830   385  86.0  71.7  59.2  *****...
 9  0.626  0.798   231  83.1  73.2  60.6  *****...

                                     I           I           I
                               Percent Observed:  0           50           100
    
```

Maximum Percentage of Reflections with I .gt. 2*s(I) in any Resolution Shell 96
 .□

SQUEEZE

- SQUEEZE is a tool in PLATON that takes care of the disordered solvent contribution to F_{calc} as part of the L.S. refinement.
- It is assumed that the total electron density can be split up into a part that can be parameterized with the **usual label,x,y,z,population,Uij** model and a part with unimportant substance that can not be parameterized satisfactorily.
- The solvent accessible region is determined automatically and used iteratively as a mask for the back-Fourier transformation of the solvent density in the difference map to F_{calc} (solvent)
- The actual L.S. refinement of the model parameters has to be done with an external program (SHELXL, CRYSTALS)
- SHELXL97 required refinement against $I(\text{obs})$ corrected for the solvent contribution into solvent free data (seen as heresy ...)

SQUEEZE & SHELXL2014

- SHELXL2014 now has a command, ABIN, that instructs to read an externally supplied **.fab** file with fixed contributions to the calculated structure factors
- PLATON/SQUEEZE supplies such a **.fab** file with the disordered solvent contribution.
- The new SHELXL2014 CIF file with the **embedded .res and .hkl** makes the use of SQUEEZE much simpler.
- The recommended input files needed for a SQUEEZE job are a **.cif and .fcf** from a converged SHELXL2014 refinement job
- SQUEEZE generates **_sq.ins, _sq.hkl & _sq.fab** for SHELXL2014
- Info on the use of SQUEEZE for archival and validation is automatically embedded in the CIF

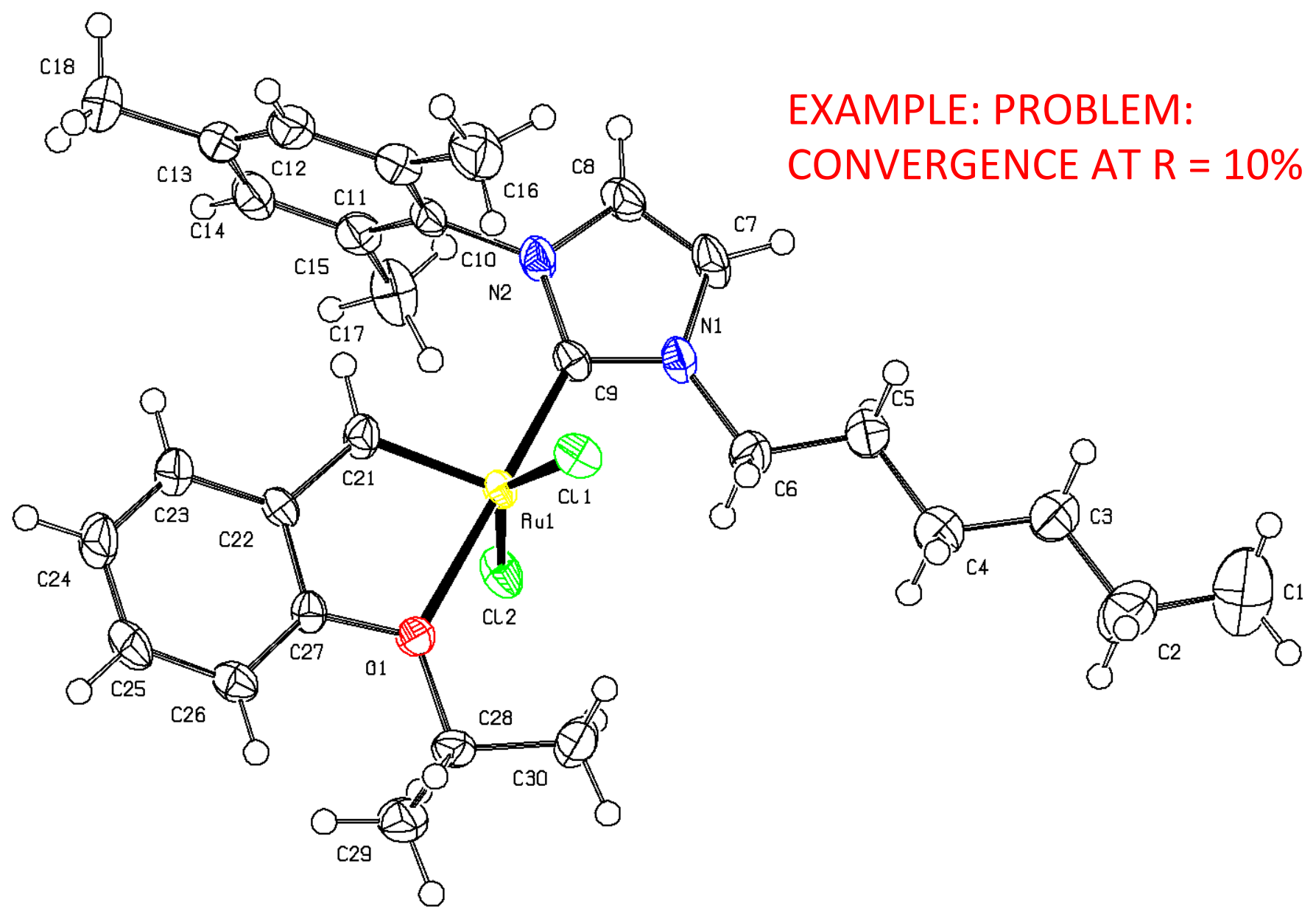
How to SQUEEZE with SHELXL2014

1. Refine non-solvent model (Include ACTA, and in case of twinning, LIST 8 record(s) in the .ins).
2. Run SQUEEZE, based on the *.cif* & *.fcf* from **1** as `'platon -q name.cif'`.
3. Continue SHELXL2014 refinement with the files *name_sq.ins*, *name_sq.hkl* & *name_sq.fab* from **2** as `'shelxl name_sq'`
4. Inspect the *.lis* & *.lst* files and Validate

27 Y
PLATON-Jul 30 15:28:22 2014 - (290714)

NOMOVE FORCED

Prob = 50
Temp = 150

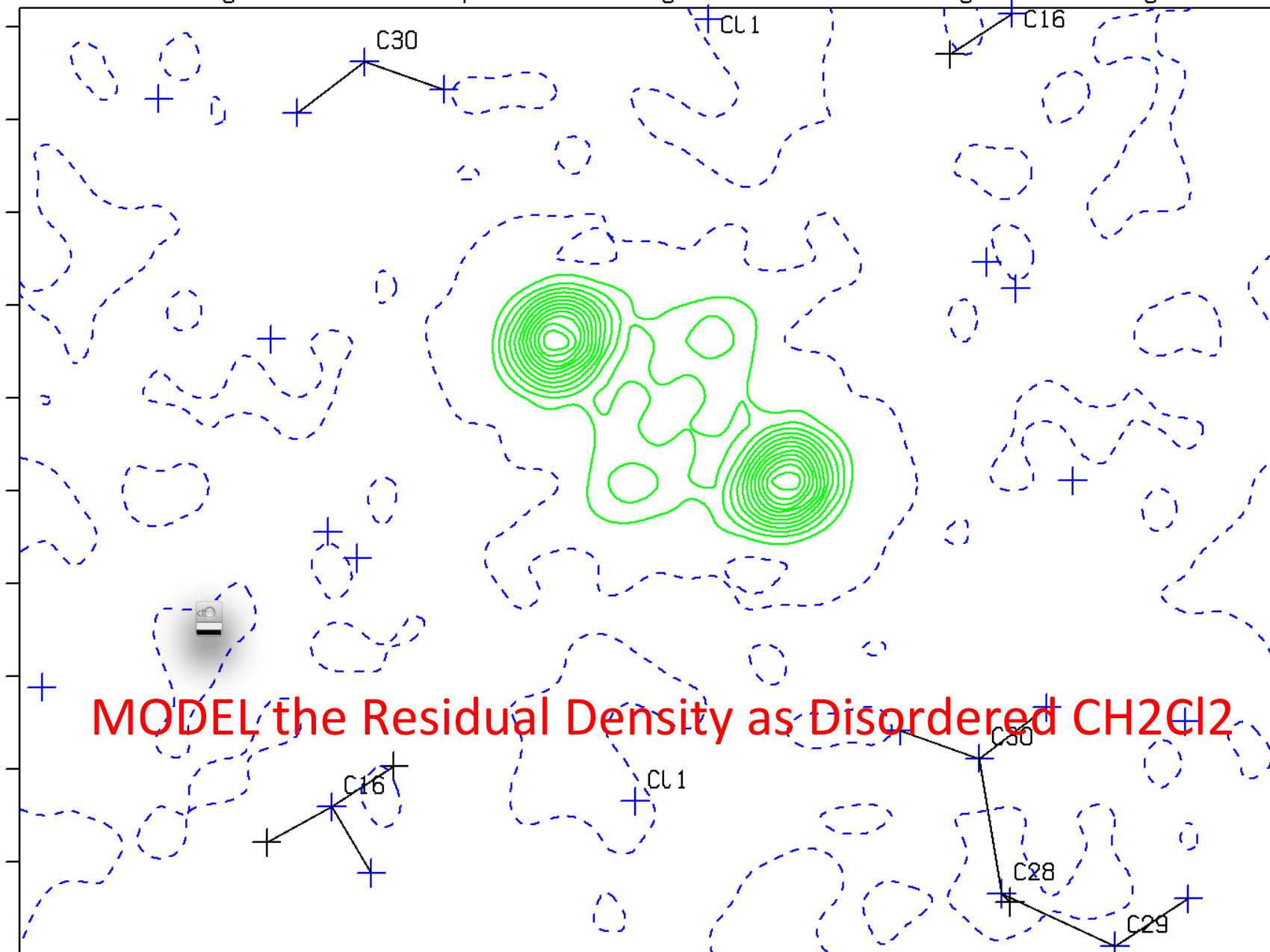


EXAMPLE: PROBLEM:
CONVERGENCE AT R = 10%

PLATON-Jul 30 15:09:26 2014 - (290714)

Plane: -2.6897x -1.6405y-14.6170z = -2.1651 Cont-Lev(eA-3): 0.00 12.00 1.00 Dlf-Map
Tol = 1.5 Ang Step = 0.3000 Ang Resolution 27.5 Deg. Omit 2*StgI

R1 = 0.092
wR2 = 0.301
S = 3.560



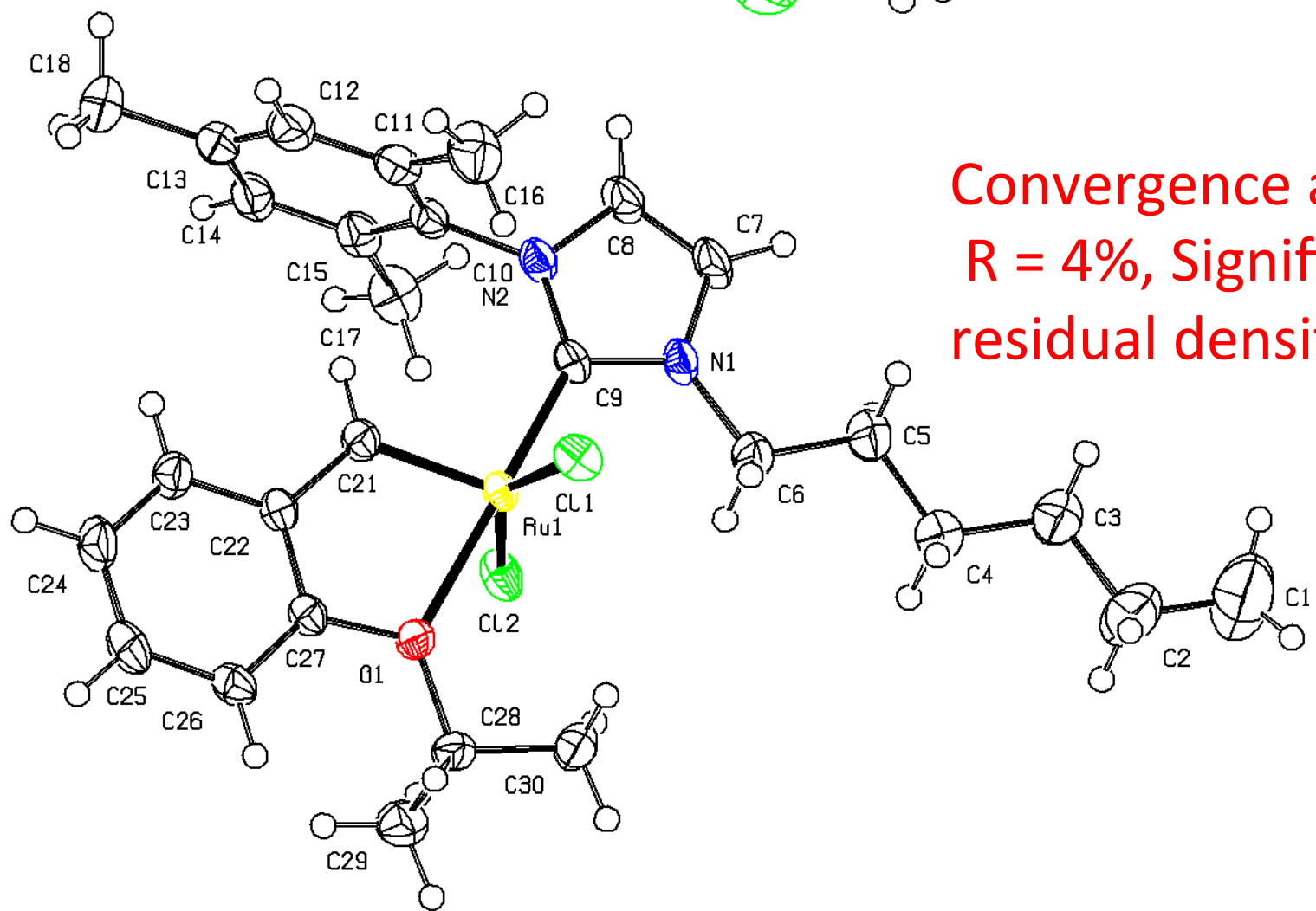
MODEL the Residual Density as Disordered CH₂Cl₂

s4315a P 21/c R = 0.10

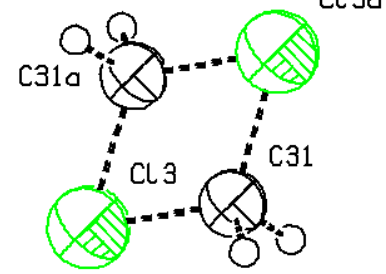
Ang

NOMOVE FORCED

Prob = 50
Temp = 150



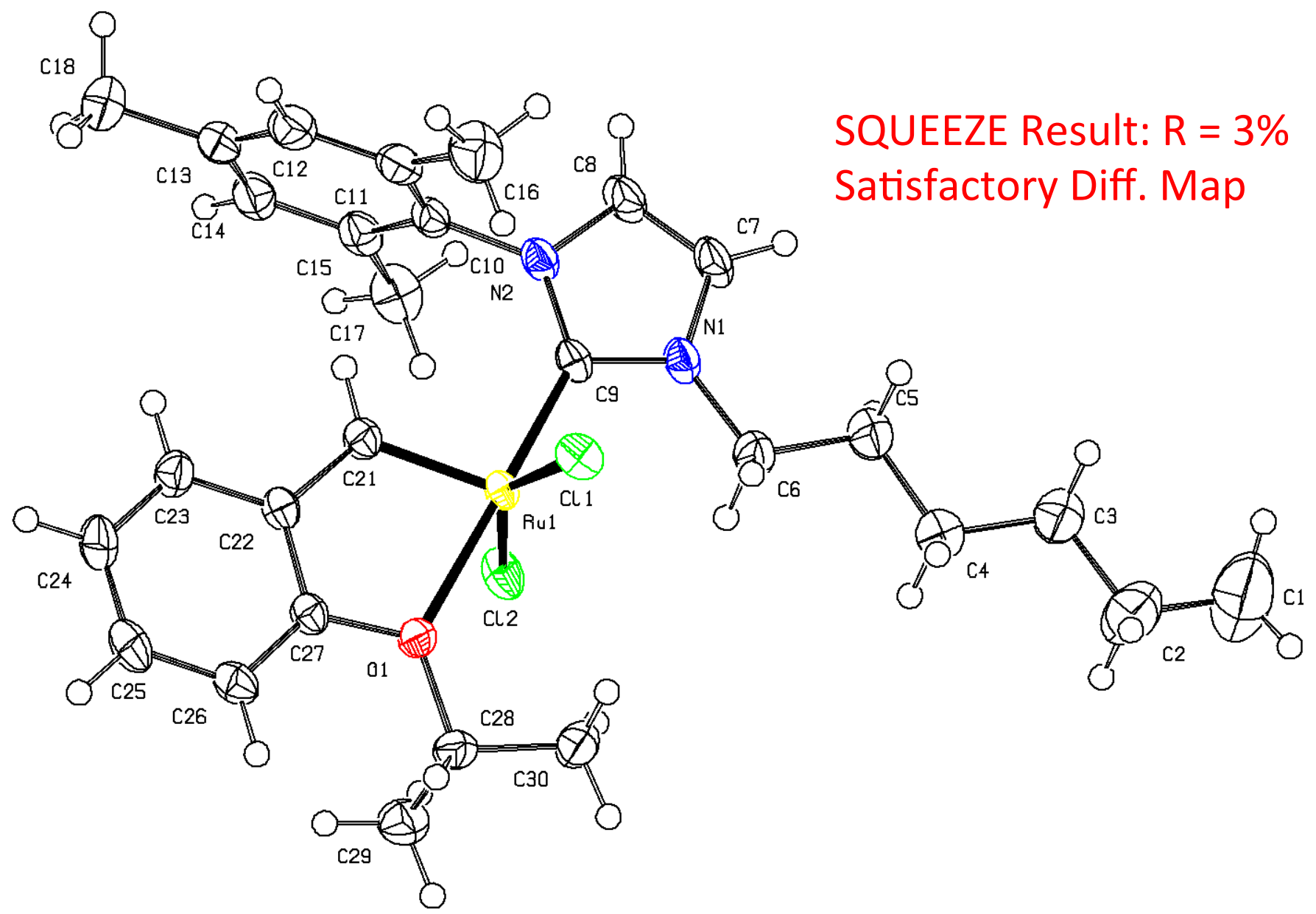
Convergence at
R = 4%, Significant
residual density



27 Y
PLATON-Jul 30 15:19:32 2014 - (290714)

NOMOVE FORCED

Prob = 50
Temp = 150



SQUEEZE Result: R = 3%
Satisfactory Diff. Map

Result Summary

□

P21/c	2(C28 H38 Cl2 N2 O Ru, x(C H2 Cl2)		
	none	0.69(CH2Cl2)	squeeze
R	0.0954	0.0442	0.0327
wR2	0.3111	0.1315	0.0918
S	3.690	1.050	1.064
rho min/max	-0.80, 12.61	-1.59, 2.02	-0.43, 1.07
BP (Ang)	0.0137	0.0058	0.0042

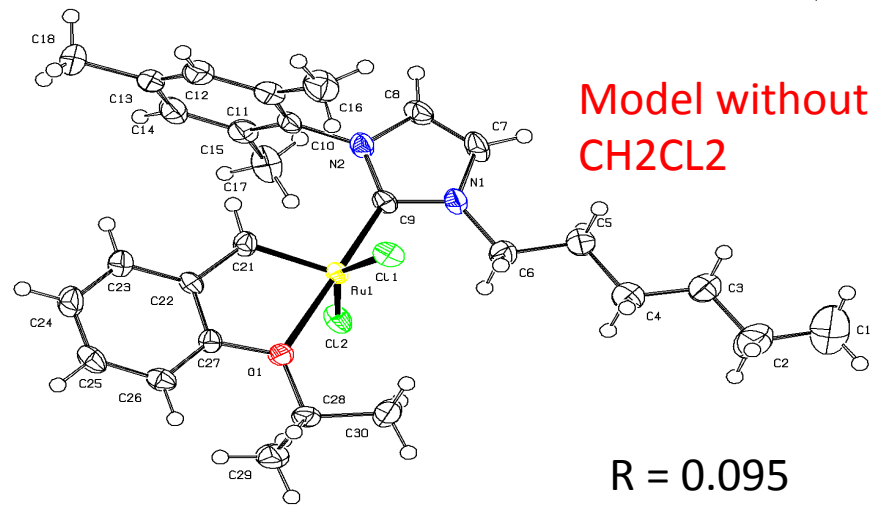
~

~

27 Y
PLATON-Jul 30 15:28:22 2014 - (290714)
Z -46 s4315a P 21/c R = 0.10 RES= 0 2 X

NOMOVE FORCED

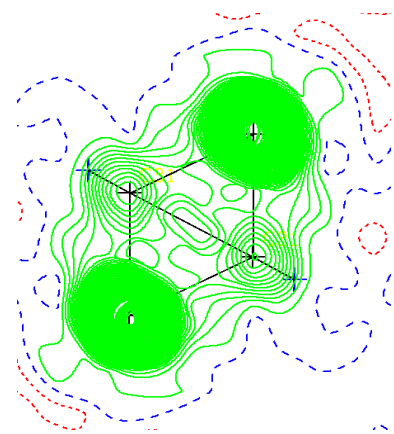
Prob = 50
Temp = 150



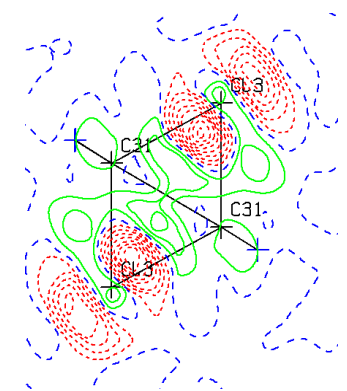
Model without
CH2Cl2

R = 0.095

Difference Maps based on Modelled CH2Cl2



Diff. Map +/- CH2Cl2

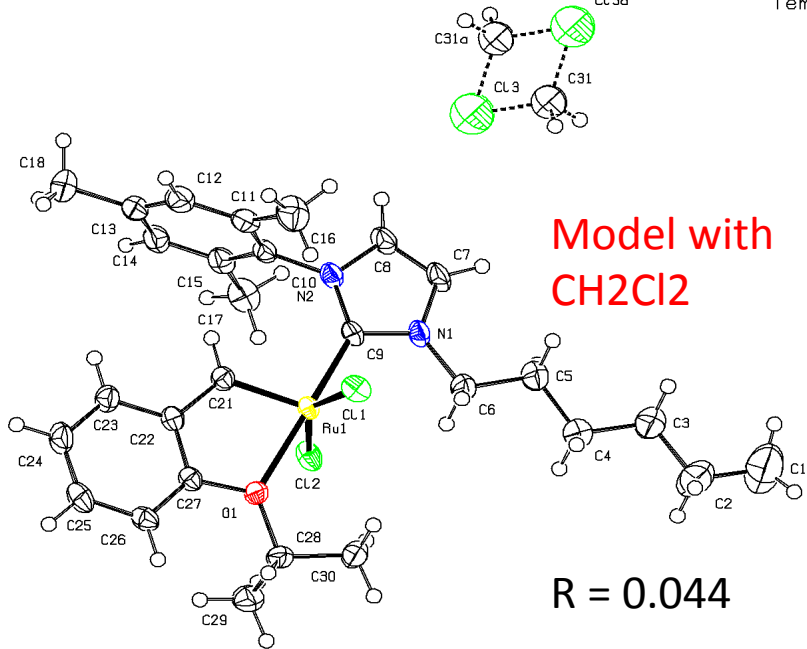


Diff.Map + CH2Cl2

27 Y
PLATON-Jul 30 15:13:47 2014 - (290714)
Z -46 s P 21/c R = 0.04 RES= 0 2 X

NOMOVE FORCED

Prob = 50
Temp = 150



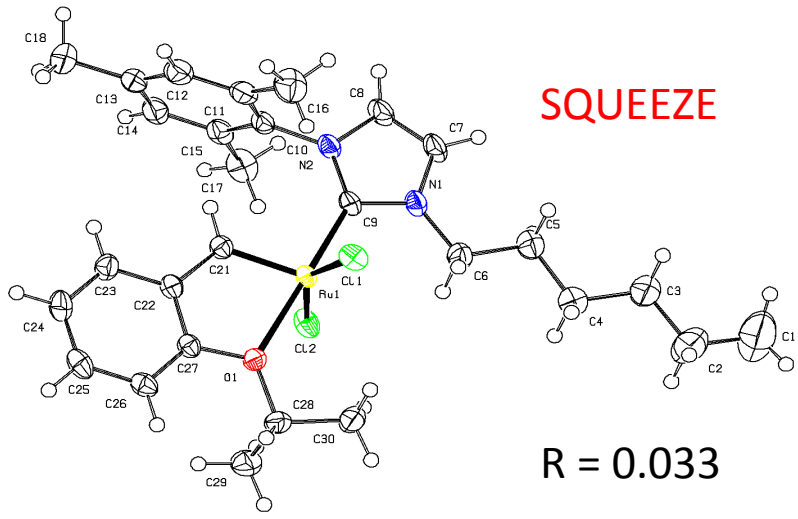
Model with
CH2Cl2

R = 0.044

27 Y
PLATON-Jul 30 15:19:32 2014 - (290714)
Z -46 s4315a_sq P 21/c R = 0.03 RES= 0 2 X

NOMOVE FORCED

Prob = 50
Temp = 150



SQUEEZE

R = 0.033

The 'NEXTRA' Issue

- One of the issues that needs to be addressed is the number of additional parameters to be added in the calculation associated with SQUEEZE on the L.S. Command. The default value is set to: $(E \times n) / (Z \times m)$ where
- E = the number of recovered electrons in the unit cell
- Z = the number of asymmetric units
- m = the number of electrons in a CH₂ fragment (=8)
- n = the number of parameters usually refined for a CH₂ fragment (=9)
- This formula has the nice property that it vanishes when there is no residual density in the void

SQUEEZE Requirements

- There should be **no residual unresolved density** in the parameterized model region of the structure because of its impact in the solvent region.
- The data set should be **reasonably complete** and with **sufficient resolution** [i.e. $\sin(\theta)/\lambda > 0.6$]. Low temperature data help a lot.
- There should be no **unresolved charge balance** issues that might effect the chemistry involved (e.g. The valency of a metal in the ordered part of the structure)

Limitations

- The SQUEEZE technique can not handle cases where the solvent region is **not constant** but varies due to (coupled) disorder in the discrete part of the model.
- The solvent region is assumed not to contain significant anomalous scatterers (**Friedels averaged**)
- Designed for 'small molecule structures'
- The use of SQUEEZE as part of the Fujita et al. MOF soaking method where the interest lies in the solvent region is very tricky and not recommended

SQUEEZE & TWINNING

Thanks to Shelxl2014, SQUEEZE can now handle
the combination of solvent disorder and
twinning via the LIST 8 option
(Both TWIN/BASF & BASF/HKLF 5)

See

SQUEEZE lecture in MS96.005

Thanks

to

George Sheldrick for his excellent major 2014

update to

SHELXL (and of course SHELXT)

and the

many users who gave their time for useful
comments, suggestions and bug reports

(with supporting CIF)

Informal Theory of the SQUEEZE Procedure

$$F_H = \int_V \rho(\vec{r}) e^{2\pi i \vec{H} \cdot \vec{r}} dV$$

$$= \int_V (\rho^M(\vec{r}) + \rho^S(\vec{r})) e^{2\pi i \vec{H} \cdot \vec{r}} dV$$

$$\rho^M(\vec{r}) = \sum_{j=1}^N \rho(\vec{r} - \vec{r}_j)$$

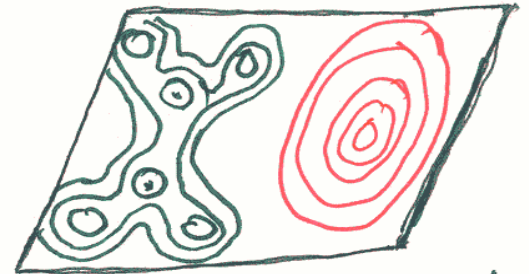
$$F_H^M = \sum_{j=1}^N f_j e^{2\pi i \vec{H} \cdot \vec{r}_j}$$

$\approx \Delta \rho(\vec{r}_k)$ on grid

$$F_H^S = V_g \sum_S \Delta \rho(\vec{r}_k) e^{2\pi i \vec{H} \cdot \vec{r}_k}$$

$$F_H^C = F_H^M + F_H^S \Rightarrow$$

$$\rho_H^S = \rho_H^M$$

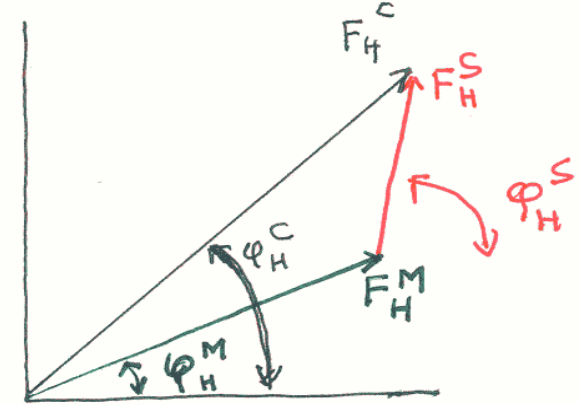


M = Ordered

S = Solvent

Iterate (Initially)

$$\Delta \rho(\vec{r}) = \frac{1}{V} \sum_H (|F_H^C| e^{i\varphi_H^C} - |F_H^M| e^{i\varphi_H^M}) e^{-2\pi i \vec{H} \cdot \vec{r}} + \frac{F_H^S}{V}$$



Solvent Free
ElectronCount

$$F_H^0 : F_H^{0'} = |F_H^0| e^{i\varphi_H^0} - |F_H^S| e^{i\varphi_H^S}$$

$$: V_g \sum_S \Delta \rho(\vec{r}_k)$$