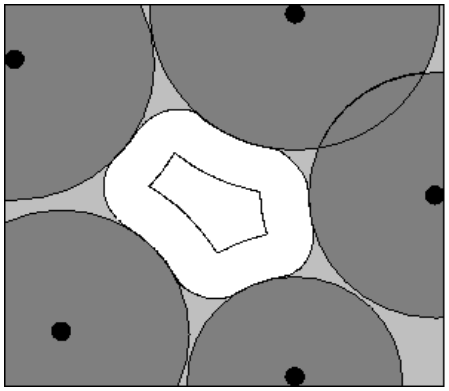
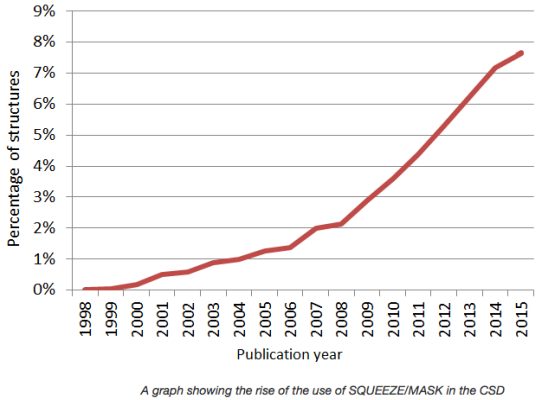
(check)CIF, SHELXL-2014, SQUEEZE





A graph showing the rise of the use of SQUEEZE/MASK in the CSD

Ton Spek

Utrecht University

The Netherlands

SHELX Workshop, ACA-Denver, July 22, 2016

The CIF Standard & Validation

- CIF was created around 1990 by an IUCr committee for data exchange and archival.
- One of the early adopters was SHELXL97.
- Acta Cryst. C pioneered its use for publication data and text entry.
- Acta Cryst. C pioneered automated checking of data consistency and data completeness.
- Today, an IUCr-checkCIF report is an essential requirement for publication in most journals.

FCF-Validation Added

- A SHELX97 style CIF only reports the numerical results of a structure determination (i.e. Space group, model parameters and R-values)
- The associated 'CIF-style' FCF file allows for a detailed analysis of the fit of the structure model (Fcalc) to the reflection data (Fobs)
- Together, the CIF + FCF offer the authors interpretation of the experimental data

Archival of the Experimental Data

- For a proper review and archival for possible follow-up research we would need at least the deposition of the unmerged reflection data.
- Needed to resolve issues concerning missed symmetry, missed twinning, hydrogen atoms, main molecule disorder, disordered solvents etc.
- The 'embedding' mechanism was choosen to include the unmerged reflection data in the CIF as a comment with a proper data name, to be ignored in most applications such as graphics and geometry calculations.

Final .res & .hkl Embedding

- Two general data names were introduced for the inclusion of the refinement and reflection details _iucr_refine_instructions_details and _iucr_refine_reflections_details resp.
- The .res and .hkl are embedded as text between semicolons (i.e. '; <*newline*> <text> <*newline*> ;')
- SHELXL2014 introduced its own equivalents: _shelx_res_file & _shelx_hkl_file along with associated checksums for data integrity.
- Those embedded data should NOT be edited or removed from the CIF.
- Use 'shredcif' or PLATON to extract the .res & .hkl

The SHELXL2014 ABIN Instruction

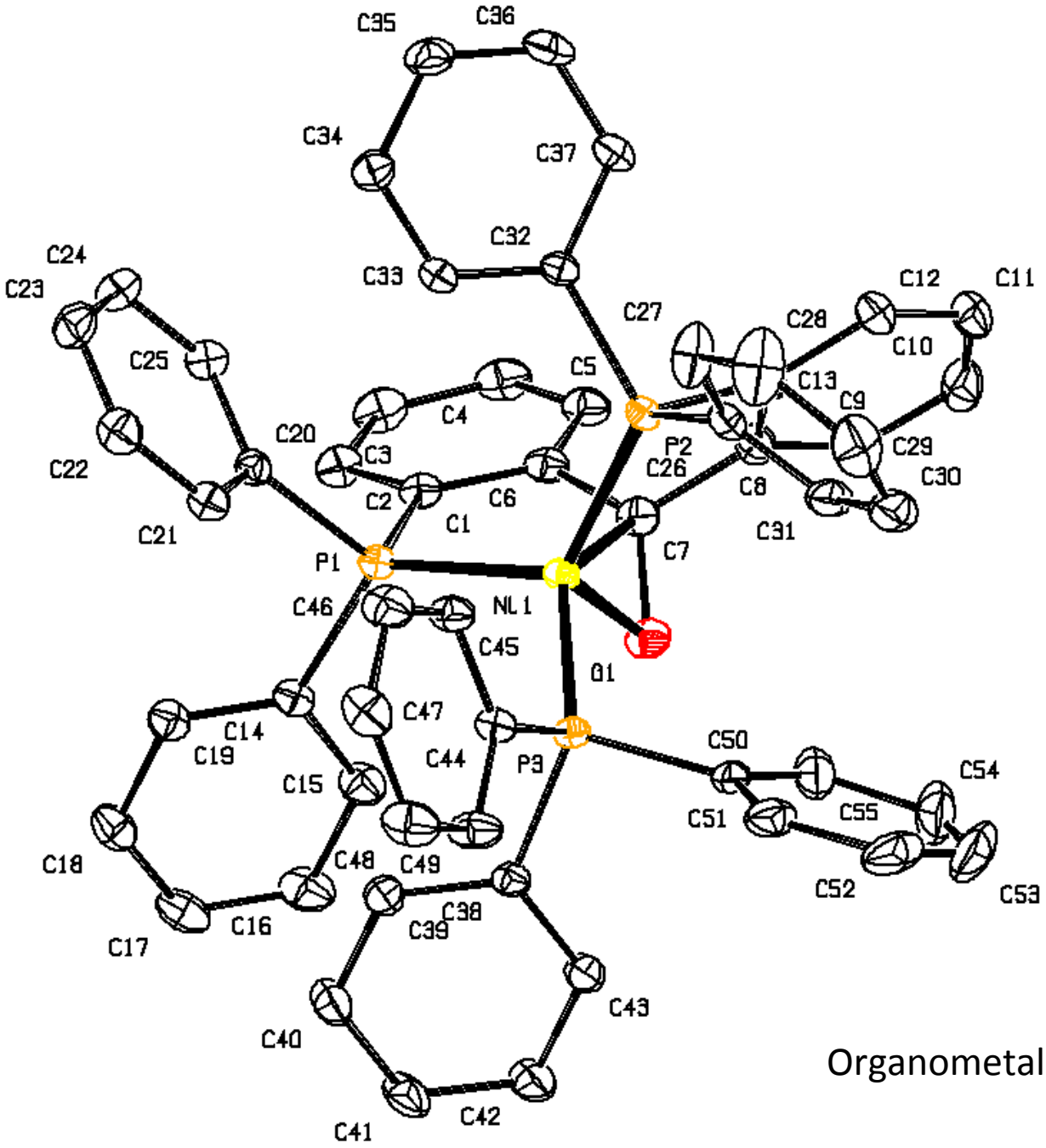
- The total electron density in the unit cell can be split up into two parts, rho1 & rho2, with associated contributions to F_h(calc): F_h (calc) = F_h1 + F_h2.
- F_h1 might be associated with the main molecule of interest and F_h2 with a disordered solvent region.
- Generally, a disorder model takes care of F_h2.
- Optionally, the F_h2 part can be calculated using an external program and read by SHELXL from a .fab file
- The ABIN instruction informs SHELXL2014 to search for and read the external .fab file with H,K,L,A_h2,B_h2.

The Disordered Solvent Problem

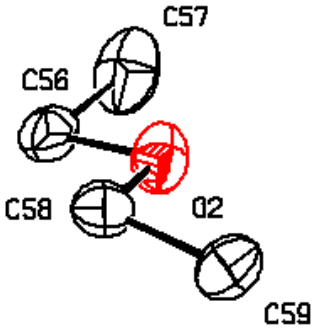
- SHELXL2014 offers an extensive set of options to model and refine disordered solvents. This is the preferred approach in most known solvent disorder cases.
- In cases of multiple unknown solvent mixtures and smeared density, an elaborate disorder model might not work satisfactorily.
- In such cases the SQUEEZE approach with an externally determined solvent contribution might result in a satisfactory main molecule refinement

PLATON/SQUEEZE

- The current implementation of the SQUEEZE tool to handle disordered solvents is the third generation of a method published by us more than 25 years ago.
- Interfacing with SHELXL2014 refinement solves many earlier issues with SHELX76 & SHELXL97 using .res & .hkl data. [e.g. Modification of the observed data]
- Documentation of the recommended procedure: A.L.Spek (2015) Acta Cryst. C71, 9-18
- http://www.platonsoft.nl/PLATON_HOW_TO.pdf
- Example: Comparison of disorder model <> SQUEEZE



Example
P21/c, 150K
R = 0.0386
wR2 = 0.0966
S =1.037



Diethyl Ether
Disordered over
Inversion centre
PART -1

Organometallics (2015), 34,2710-13

PLATON

A Multipurpose Crystallographic Tool

(C) 1980-2016 A.L.Spek - Version: 150716 [WEB:Jul 12, 2016]

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-View	fcf2hkl
NewmanPlot	Calc Inter	Squeeze	Addsym-EXT	ABSTompa	FCF-Valid	Expand2P1
Ring-Plots	Calc Coord	Hybrid	Addsym-PLT	ABSGauss	DlfFourier	FCF-Gener
Plane-Plot	Calc Metal	CalcFCFsqd	Addsym-SHX	ABSXtal	ANALofVAR	HKLF-Gener
Polyhedra	Calc Geom	Contoursqd	Newsym	ABSSphere	ByvoetPair	HKL-Transf
ContourDlf	Calc Hbond	Solv F3D	Nonsym	ShxAbs	Asym-Expct	Exor-Res
Contour-Fo	Calc TMA	Solv Plot	LePage	AnomDlsVal	Asym-Valid	Anis-Res
AutoMolFit	L.S.-Plane	CavityPlot	DelRed	AnomDlsPlt	SupplMater	Rename-Res
hkl2Powder	DihedAngle	WilsonPlot	Molsym	MuPlot	Expect-hkl	Auto-Renum
SlmPowderP	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	LePageTwin	XtlPlanAgl	StrainAnal	Create-pdb
PlutONatlv	Hflx - res	Structure?	TwinRotMat	Xtal Habit	LocCIF-acc	clf2shelxl

Xtal Data (CIF14) x.clf- Set 1(1): I

Refl Data (LIST4) x.fcf [FCF] (1): I

Browser - HELP

http://www.platonsoft.nl/PLATON-MANUAL.pdf http://www.platonsoft.nl/PLATON HOW TO.pdf

PLA-SP2 12

OptionMenus

Print-Level

EPS-Listing

PDF-Listing

PageHeader

NoMolFitInv

KeepMon-I-n

NoExpand

FCF-Calc

PNG

ReflListing

DebugOutput

SetWindSize

Portrait

GenerSthlMx

GenerRandom

EPS HPGL

Auto-Plot

X-LineWidth

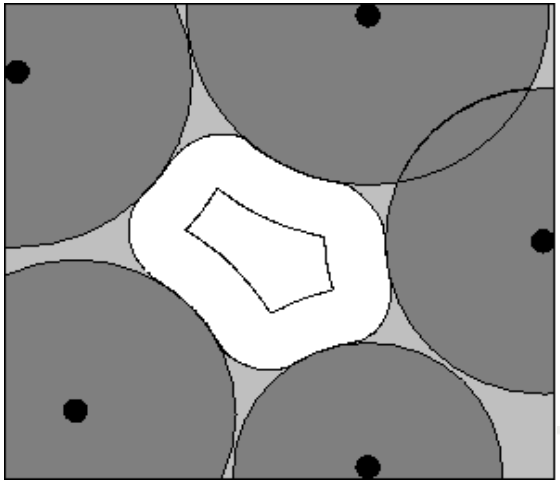
Reverse-B&W

Browser

Reset End

Search for and Analysis of Solvent Accessible Voids in the Structure

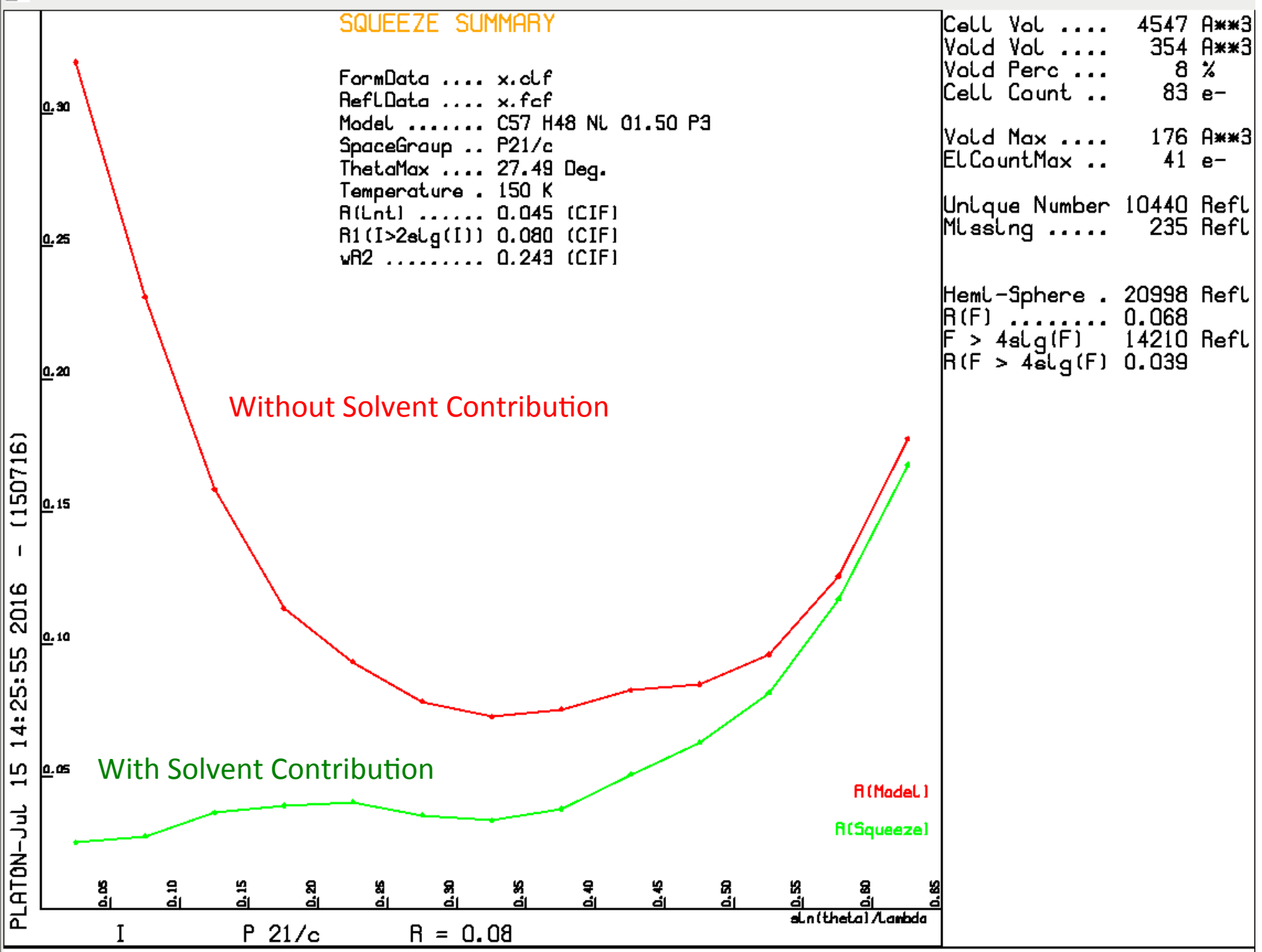
Area	#GridPoint	VolPerc.	Vol(A^3)	X(av)	Y(av)	Z(av)	Eigenvector(frac)			Sig(Ang)
1	22574[3803]	4	177[29.8]	0.500	-0.000	0.500	1	-0.508	1.000	0.709	2.38
							2	0.081	-0.975	1.000	1.49
							3	1.000	0.164	0.311	1.22
2	22577[3803]	4	177[29.8]	0.500	0.500	1.000	1	0.509	1.000	-0.705	2.38
							2	0.057	0.982	1.000	1.48
							3	1.000	-0.159	0.317	1.22



Definition of VOIDS (white area): roll sphere with radius 1.2 A
In this case there are two solvent accessible voids with
Volume 177 A**3 in the unit cell
SQUEEZE uses this area as a mask to recover the density
In the white area from the difference density map by
Iterative back-Fourier transformation into F_h2 (.fab)

PLATON/SQUEEZE

Cycle	R(F)	Nref(Heml)	R(F)>4*sig(F)	Nref	El(Solv)/Cell
1	0.109	20998	0.076	14210	0
2	0.069	20998	0.040	14210	70
3	0.068	20998	0.039	14210	82
4	0.068	20998	0.039	14210	83
5	0.068	20998	0.039	14210	83


SQUEEZE SUMMARY
FormData .... x.clf
ReflData .... x.fcf
Model C57 H48 N4 O1.50 P3
SpaceGroup .. P21/c
ThetaMax .... 27.49 Deg.
Temperature . 150 K
R(Int) ..... 0.045 (CIF)
R1(I>2sig(I)) 0.080 (CIF)
wR2 ..... 0.243 (CIF)
Cell Vol 4547 A**3
Void Vol 354 A**3
Void Perc ... 8 %
Cell Count .. 83 e-
Void Max 176 A**3
ElCountMax .. 41 e-
Unique Number 10440 Refl
Missing ..... 235 Refl
Heml-Sphere . 20998 Refl
R(F) 0.068
F > 4sig(F) 14210 Refl
R(F > 4sig(F) 0.039
Without Solvent Contribution
With Solvent Contribution
R(Model)
R(Squeeze)
0.30
0.25
0.20
0.15
0.10
0.05
0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65
sin(theta)/Lambda
PLATON-Jul 15 14:25:55 2016 - (150716)
I P 21/c R = 0.08

S U M M A R Y

```
:: Total (Fo-Fc)map Electron Count Outside Voids =       0
================================================================
HOW TO PROCEED with L.S. refinement after running SQUEEZE:
================================================================
SHELXL201n: Continue refinement in the presence of the three
files name_sq.ins, name_sq.hkl & name_sq.fab
Additional info on name_sq.sqf , name_sq.sqz & name_sq.lis
================================================================
Citation: Spek, A.L. (2015). Acta Cryst. C71, 9-18.
================================================================
```

- The _sq.ins file is the original .res (from .cif) + ABIN Instruction

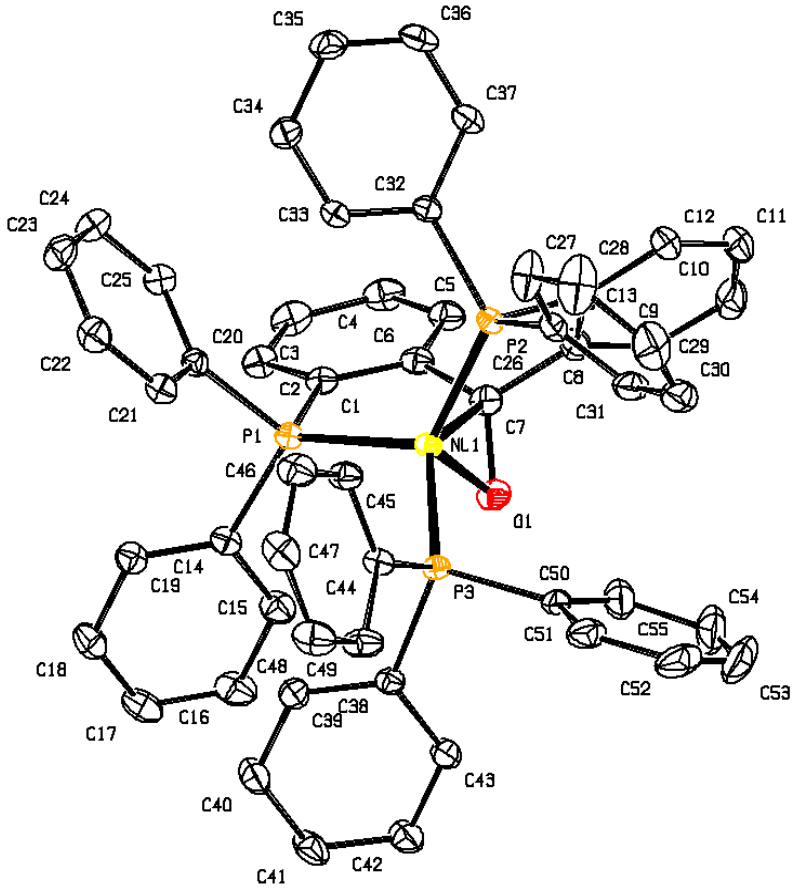
- The _sq.hkl file is the original .hkl (from .cif)

- The _sq.fab file (created by SQUEEZE) includes after the last reflection info about the SQUEEZE job i.e._sq.sqf & _sq.sqz

Note: PLATON/SQUEEZE does NOT refine the Model Parameters

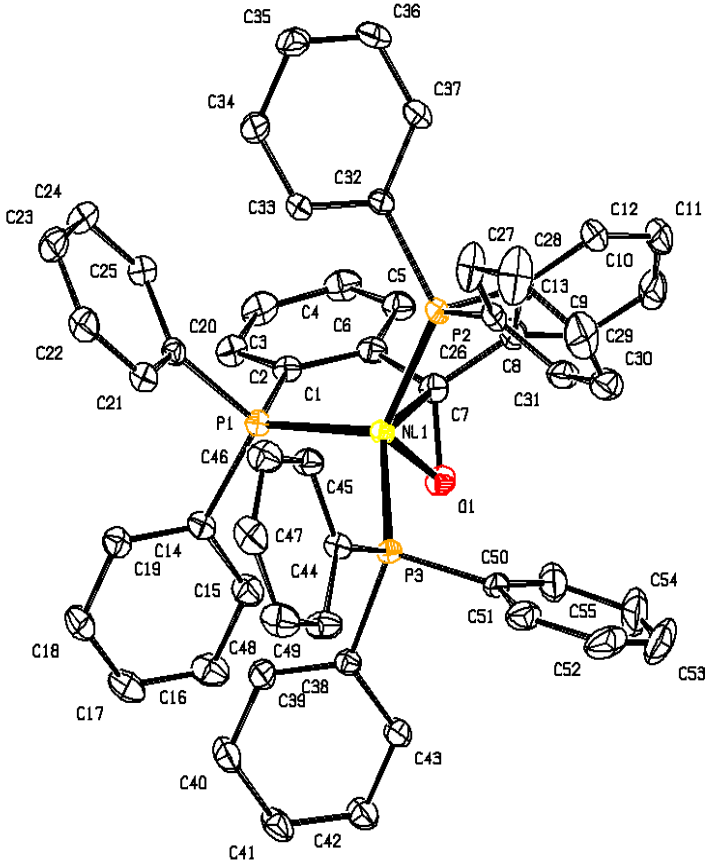
Comparison

$R_1 = 0.0386$, $wR_2 = 0.0966$
S = 1.037, 42 electrons
C-C BP = 0.0036 Angstrom



Disorder Model Diethyl Ether

$R_1 = 0.0383$, $wR_2 = 0.0960$
S = 1.044, 41 electrons
C-C BP = 0.0035 Angstrom



Squeeze Model DiethylEther

The Proper use of the SQUEEZE Tool

- It is important that the final CIF archives both the details of the SQUEEZE calculation and the unmerged reflection data.
- The SQUEEZE details are appended to the .fab file
- SHELXL2014 offers, by embedding the .res, .hkl & .fab data, all what is needed for that.
- In that way, the calculations can be reconstructed and/or alternative refinement models attempted.

Summary of SQUEEZE + SHELXL2014

1. Refine a non-solvent model with name.ins & name.hkl (Include ACTA record, NO LIST 6) .
2. Run PLATON/SQUEEZE, based on *name.cif* & name.*fcf* from **1** as 'platon –q name.cif'.
3. Continue SHELXL refinement with the files *name_sq.ins*, name_sq.hkl & *name_sq.fab* from **2** as '(shel)xl name_sq'
4. Inspect the .lis & .lst files and Validate

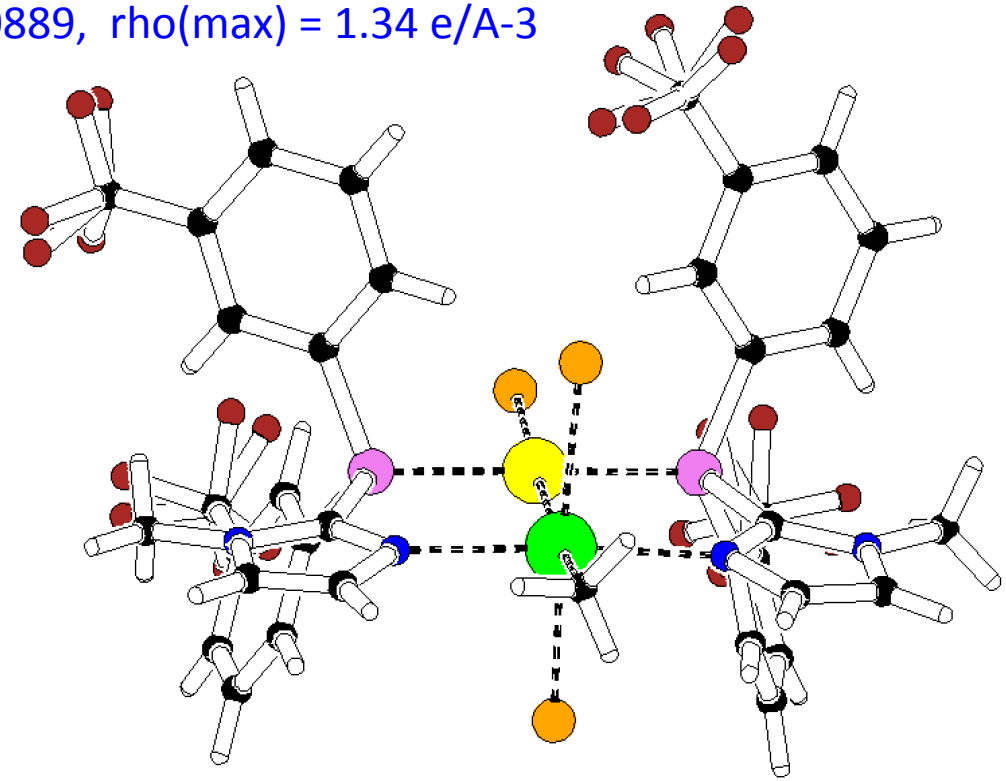
SQUEEZE Disordered Solvent + Twinning

- Step 1: SHELXL2014 refinement based a name.ins (that should include 'ACTA', 'LIST 8', 'BASF' and 'HKLF 5' [or 'TWIN'] records) and a name.hkl file
- Step 2: Run SQUEEZE with the name.cif and name.fcf files produced in Step 1 (i.e. run: platon –q name.cif)
- Step 3: Continue SHELXL refinement with the files name_sq.ins, name_sq.hkl and name_sq.fab produced by PLATON in step 2 → name_sq.cif & name_sq.fcf

SQUEEZE-2016 EXAMPLE [Chem.Eur.J. (2015) 21, 1765]

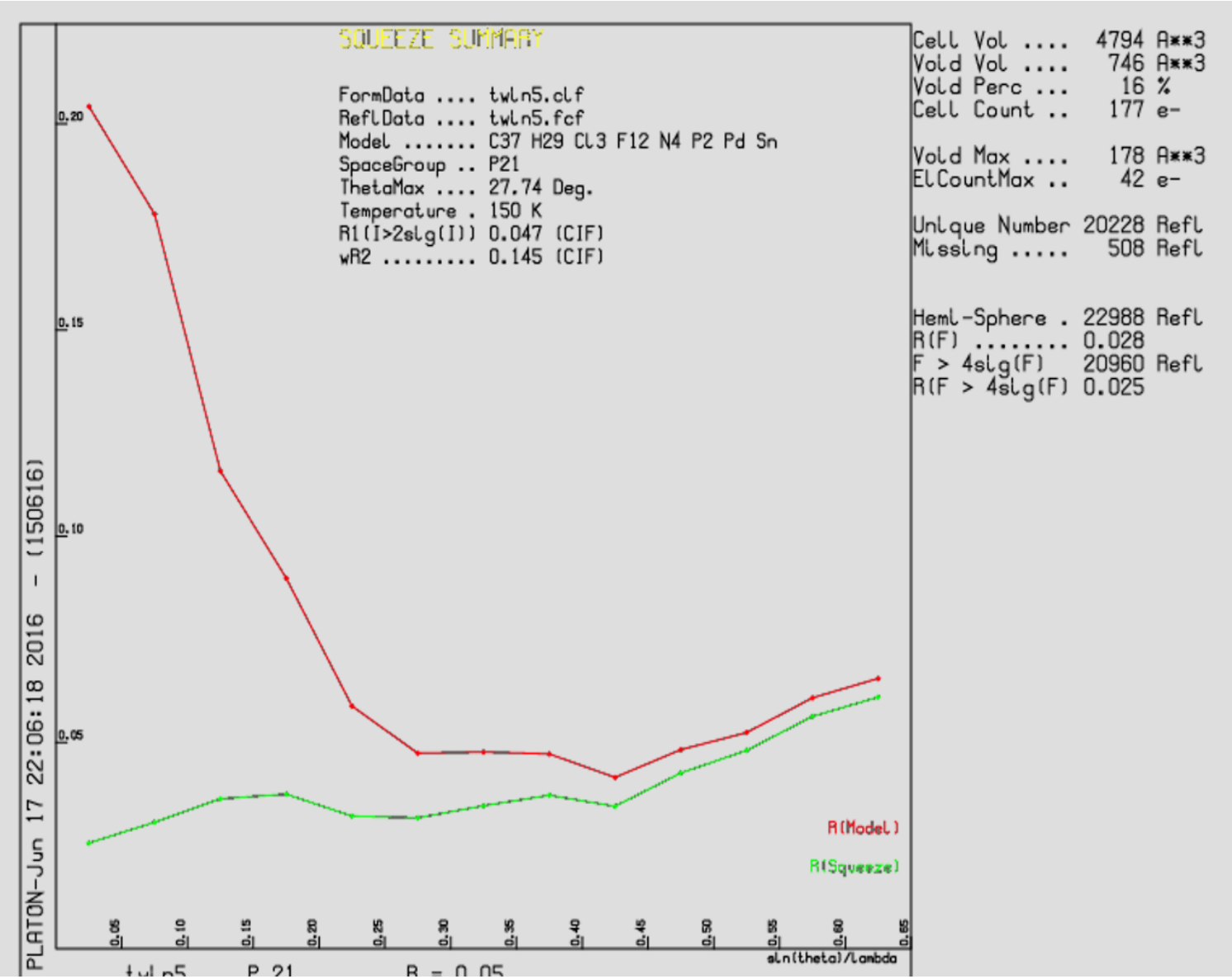
Acetonitril Model: R = 0.0323, wR2 = 0.0889, rho(max) = 1.34 e/A-3

Space Group $P2_1$
Z = 4, Z' = 2
60:40 Twin
Twin axis: (0 0 1)
150 K
TWINABS hklf5 data
Acetonitril solvate



Step 1 (SHELXL2014) ➔ R1 = 0.047, wR2 = 0.1445
Step 2 (SQUEEZE) ➔ 177 electrons found in unit cell
Step 3 (SHELXL2014) ➔ R1 = 0.0275, wR2 = 0.0679, S = 1.064

Effect of on R(F) before and after SQUEEZE as a function of sin(theta)/lambda

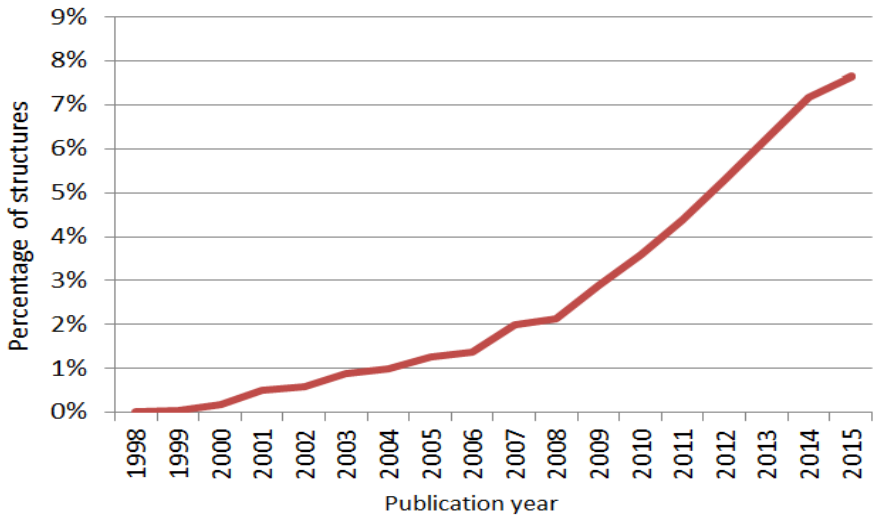


Requirements

- There should be no residual unresolved density in the discrete model region of the structure because of its impact on the difference map in the solvent region. (may invalidate el. Count)
- The data set should be reasonably complete and with sufficient resolution [i.e. sin(theta)/ lambda >0.6].
- 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 reported electron count in the solvent region is meaningful only with the supply of a complete and reliable reflection data set.
- The SQUEEZE technique can not handle properly cases of coupled disorder effecting both the model and the solvent region.
- The solvent region is assumed not to contain significant anomalous scatterers (Friedels averaged)



A graph showing the rise of the use of SQUEEZE/MASK in the CSD

Reported SQUEEZE Usage Statistics as Prepared by the CCDC

Thank you !

a.l.spek@uu.nl

More info:www.platonsoft.nl

(including this powerpoint presentation)