PLATON SQUEEZE: a tool for the calculation of the disordered solvent contribution to the calculated structure factors

Anthony L. Spek

Crystal and Structural Chemistry, Utrecht University, Padualaan 8, Utrecht 3584CH, The Netherlands

Correspondence e-mail: a.l.spek@uu.nl

The completion of a crystal structure determination is often hampered by the presence of embedded solvent molecules or ions that are seriously disordered. Their contribution to the calculated structure factors in the least-squares refinement of a crystal structure has to be included in some way. Traditionally, an atomistic solvent disorder model is attempted. Such an approach is generally to be preferred, but it does not always lead to a satisfactory result and may even be impossible in cases where channels in the structure are filled with continuous electron density. This paper documents the SQUEEZE method as an alternative means of addressing the solvent disorder issue. It conveniently interfaces with the 2014 version of the least-squares refinement program *SHELXL* [Sheldrick (2015). *Acta Cryst.* C**71**. In the press] and other refinement programs that accept externally provided fixed contributions to the calculated structure factors. The *PLATON* SQUEEZE tool calculates the solvent contribution to the structure factors by back-Fourier transformation of the electron density found in the solvent-accessible region of a phase-optimized difference electron-density map. The actual least-squares structure refinement is delegated to, for example, *SHELXL*. The current versions of *PLATON* SQUEEZE and *SHELXL* now address several of the unnecessary complications with the earlier implementation of the SQUEEZE procedure that were a necessity because least-squares refinement with the now superseded *SHELXL97* program did not allow for the input of fixed externally provided contributions to the structure-factor calculation. It is no longer necessary to subtract the solvent contribution temporarily from the observed intensities to be able to use *SHELXL* for the least-squares refinement, since that program now accepts the solvent contribution from an external file (`.fab` file) if the ABIN instruction is used. In addition, many twinned structures containing disordered solvents are now also treatable by SQUEEZE. The details

of a SQUEEZE calculation are now automatically included in the CIF archive file, along with the unmerged reflection data. The current implementation of the SQUEEZE procedure is described, and discussed and illustrated with three examples. Two of them are based on the reflection data of published structures and one on synthetic reflection data generated for a published structure.

Keywords: SQUEEZE; *PLATON*; embedded solvent; disorder.

1. Introduction

Disordered solvents of crystallization can pose a time-consuming problem for the completion of an otherwise routine crystal structure determination. Solvents often fill packing voids in a crystal structure with no significant interaction with their host structure, and are thus prone to disorder or even rapid loss from the crystal once it is removed from the mother liquor. Unfortunately, the nature of the disordered solvent present in the crystal is also not always known. It can be a mixture of the solvents that were used during the synthesis of the compound of interest and the solvents used for their crystallization. That might be the case in particular when a batch contains only a few good quality crystals that only grow due to the inclusion of a suitable impurity. The problem then is how to model their scattering contribution to the calculated structure factors as part of the least-squares refinement of the structure. Although often not of particular structural interest, their impact on the quality of the part of the structure of interest can be significant, particularly when quantified in terms of the standard uncertainties (s.u.'s) of the atomic coordinates and derived geometric parameter values, the geometric parameter values themselves and the confidence factors $R[F^2 > 2\sigma(F^2)]$, $wR(F^2)$ and S.

Programs such as *SHELXL* (Sheldrick, 2008, 2015), *CRYSTALS* (Betteridge *et al.*, 2003), *JANA* (Petříček *et al.*, 2014) and *OLEX2* (Dolomanov *et al.*, 2009) offer elaborate constraint and restraint tools to handle the least-squares refinement of an atomistic disorder model of the solvent structure. Their use is often the preferred procedure when the nature of the solvent is clear and the devised disorder model meaningful. The complete model should result in an essentially featureless difference electron-density map.

Unfortunately, in many cases, the introduction of a disorder model can be difficult, complex, unsatisfactory and in some cases unfeasible. Examples are unsymmetrical molecules located at or about high-symmetry sites, solvent molecules in infinite channels and solvent mixtures of unknown composition. Continuous electron density in infinite channels and tori, due to the rotation of a group of atoms, cannot be approximated satisfactorily as a sum of Gaussian-shaped atomic densities. In such cases, a hybrid approach can be attempted where the total electron density is split up into a part that can be modelled and refined as usual, and a part of the electron density corresponding to the disordered solvent that is back-Fourier transformed into the solvent contribution to the calculated structure factors.
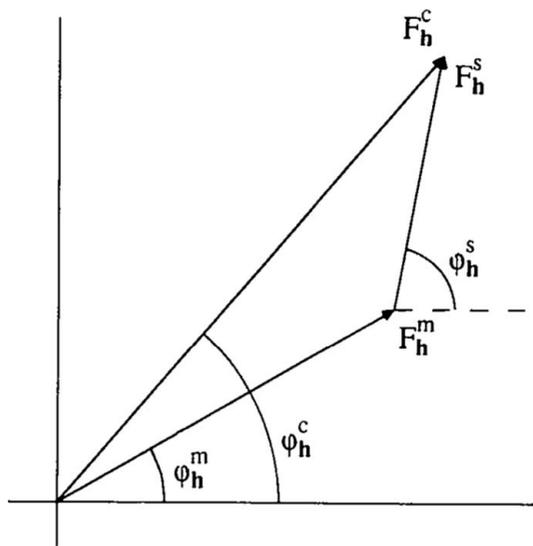
Figure 1
Argand diagram showing the total calculated structure factor F_h^c to be used in the least-squares refinement against the observed $\left|F_h^o\right|^2$, as the vector sum of the host F_h^m and solvent F_h^s contributions. F_h^s is determined iteratively using the SQUEEZE algorithm with the F_h^m contribution fixed, F_h^c set equal to $\left|F_h^o\right|$ with phase φ_h^c and starting from $\varphi_h^s = \varphi_h^m$.

Initially, an unsatisfactorily high $R[F > 5\sigma(F)]$ factor of 0.096 was obtained for the crystal structure of the pharmaceutical compound Salazopyrine® (van der Sluis & Spek, 1990a). Close inspection of contoured difference electron-density maps showed this to be due to unaccounted-for continuous density in infinite channels filled with unknown solvent. This prompted us 25 years ago to investigate the back-Fourier transform approach, based on an idea gleaned from a footnote in a paper by Wehman *et al.* (1988). A prototype proof-of-principle implementation for this, including a number of *ad hoc* programs, was developed and published as the *BYPASS* procedure (van der Sluis & Spek, 1990b), based around the *SHELX76* (Sheldrick, 2008) least-squares refinement program available to us at that time. Application to the aforementioned structure made the final refinement converge at a healthy $R[F > 5\sigma(F)]$ value of 0.045. A more definite and distributable version of the back-Fourier transform procedure was implemented subsequently as the SQUEEZE tool in the program *PLATON* (Spek, 2009), now tailored to work optimally in conjunction with the widely used *SHELXL97* (Sheldrick, 2008) refinement program, but also easy to use within the *CRYSTALS* package (Betteridge *et al.*, 2003).

This paper describes the current SQUEEZE implementation that is based on new functionality included in the 2014 version of *SHELXL* (Sheldrick, 2015). The current versions of *PLATON* SQUEEZE and *SHELXL* now avoid temporary subtraction of the solvent contribution from the observed intensities, which was necessary in order to refine a structure model using the earlier versions of *SHELX(L)*. Although not essential for the end result, because the original observed reflection data could be reinstated along with the calculated structure factors (including the solvent contribution to the calculated structure factors) and the associated $R[F^2 > 2\sigma(F^2)]$,

$wR(F^2)$ and S values, even the temporary modification of the observed reflection data was considered by some as scientific heresy. With an implementation of the *PLATON* SQUEEZE tool in the *CRYSTALS* package (Betteridge *et al.*, 2003), which does allow for the external supply of contributions to the structure factors, it could be shown that both paths give the same result. Relatively recently, an independent implementation of the SQUEEZE concept in the *OLEX2* package (Dolomanov *et al.*, 2009) has become available. All current implementations of the SQUEEZE concept now add the solvent contribution to F_{calc}, while retaining the original experimental input data in the least-squares refinement.

The SQUEEZE tool is only one of the many tools available in the *PLATON* program (Spek, 2003, 2009). Other tools include *checkCIF* for structure validation, TwinRotMat for automatic detection of twinning, ADDSYM for the detection of missed and pseudosymmetry, BIJVOET for absolute structure determination (Hooft *et al.*, 2008, 2010), an extensive assortment of geometric calculations (bonds, angles, torsion angles, least-squares planes, ring puckering and hydrogen bonds, among others) and molecular graphics tools (*ORTEP*, *PLUTON*, contour plots).

2. The SQUEEZE tool

2.1. Theoretical background

The basic idea of SQUEEZE is shown in the Argand diagram in Fig. 1. In this method, the Fourier transform of the total electron density into calculated structure factors F_h^c is approximated as the sum of two separate Fourier transforms, one for the modelled main part of the structure, F_h^m, indicated by m, and one over the solvent region, F_h^s, indicated by s

$$F_h^c = \int_V \rho(\mathbf{r}) \exp\left[2\pi i\left(\mathbf{hr}\right)\right] dV$$
$$= \int_{V_m} \rho^m(\mathbf{r}) \exp\left[2\pi i\left(\mathbf{hr}\right)\right] dV + \int_{V_s} \rho^s(\mathbf{r}) \exp\left[2\pi i\left(\mathbf{hr}\right)\right] dV,$$
$$(1)$$

where V is the total unit-cell volume, V_m the volume of the modelled main part of the structure and V_s the volume of the solvent region.

The modelled main structure electron density, $\rho^m(\mathbf{r})$, is approximated in the usual way as a sum over N individual atomic electron-density distributions

$$\rho^m(\mathbf{r}) = \sum_{j=1}^{N} \rho j(\mathbf{r} - \mathbf{r}_j), \qquad (2)$$

with Fourier transform

$$F_h^m = \sum_{j=1}^{N} f_j \exp\left[2\pi i\left(\mathbf{hr}_j\right)\right], \qquad (3)$$

where the f_j values are assumed to include the temperature factor.

Similarly, $\rho^{\mathrm{s}}(\mathbf{r})$ can be turned into a sum over the electron densities on grid points k (with grid point volume V_{g}) in the solvent-accessible region S and Fourier transformed into a solvent contribution to the F_{h}

$$F_{\mathrm{h}}^{\mathrm{s}} = V_{\mathrm{g}} \sum_{k} \rho(\mathbf{r}_k) \exp\left[2\pi i (\mathbf{h}\mathbf{r}_k)\right]. \qquad (4)$$

Thus, the calculated structure factor $F_{\mathrm{h}}^{\mathrm{c}} = F_{\mathrm{h}}^{\mathrm{m}} + F_{\mathrm{h}}^{\mathrm{s}}$, as shown in the Argand diagram of Fig. 1.

The SQUEEZE algorithm is designed to estimate the amplitude $\left|F_{\mathrm{h}}^{\mathrm{s}}\right|$ and phase $\varphi_{\mathrm{h}}^{\mathrm{s}}$ of $F_{\mathrm{h}}^{\mathrm{s}}$ from the difference electron-density map

$$\Delta\rho(\mathbf{r}) = \frac{1}{V} \sum_{h} \left[k\left|F_{\mathrm{h}}^{\mathrm{o}}\right| \exp\left(i\varphi_{\mathrm{h}}^{\mathrm{c}}\right) - \left|F_{\mathrm{h}}^{\mathrm{m}}\right| \exp\left(i\varphi_{\mathrm{h}}^{\mathrm{m}}\right) \right]$$
$$\times \exp\left[-2\pi i (\mathbf{h}\mathbf{r})\right] + \frac{F_0^{\mathrm{s}}}{V}, \qquad (5)$$

where $\left|F_{\mathrm{h}}^{\mathrm{o}}\right|$ is the observed structure factor and k is the factor required to put $\left|F_{\mathrm{h}}^{\mathrm{o}}\right|$ on the absolute scale of $\left|F_{\mathrm{h}}^{\mathrm{c}}\right|$. F_0^{s} is the solvent part of F_{000} (*i.e.* the number of solvent electrons in the unit cell), as calculated using equation (6) below.

This calculation has to be iterated, since initially $\varphi_{\mathrm{h}}^{\mathrm{c}}$, the phase of the total calculated structure factor that is used to phase the observed structure factor $\left|F_{\mathrm{h}}^{\mathrm{o}}\right|$, has to be equated to $\varphi_{\mathrm{h}}^{\mathrm{m}}$, the phase of the calculated structure factor of the model. In this way, a phase-improved difference electron-density map is obtained (see §4.2).

The number of electrons in the solvent region is calculated as

$$F_0^{\mathrm{s}} = V_{\mathrm{g}} \sum_{k} \rho(\mathbf{r}_k), \qquad (6)$$

where the summation is over all grid points k (with grid point volume V_{g}) in the solvent-accessible region S.

Standard least-squares refinement with the current version of *SHELXL* will be on $\left|F_{\mathrm{h}}^{\mathrm{c}}\right|^2$ against $\left|F_{\mathrm{h}}^{\mathrm{o}}\right|^2$. The parameters of the host m are refined and the contribution of the solvent is kept fixed.

The now superseded 1997 and earlier versions of *SHELXL* required observed reflection data, from which the solvent contribution was subtracted. This can be visualized in Fig. 1 by substituting $F_{\mathrm{h}}^{\mathrm{o}}$ for $F_{\mathrm{h}}^{\mathrm{c}}$ and $F_{\mathrm{h}}^{\mathrm{o}\prime}$ for $F_{\mathrm{h}}^{\mathrm{m}}$. The new 'observed' data are then $\left|F_{\mathrm{h}}^{\mathrm{o}\prime}\right|^2$. After convergence of the refinement, the reverse operation should be done to obtain a proper listing of $\left|F_{\mathrm{h}}^{\mathrm{o}}\right|^2$ against $\left|F_{\mathrm{h}}^{\mathrm{c}}\right|^2$.

2.2. The *PLATON* SQUEEZE procedure combined with *SHELXL* refinement

It is assumed that the nondisordered solvent part of the structure model is complete, including attached H atoms, and with host structure disorder, when present, modelled. The solvent-accessible region that is to be SQUEEZEd should be left empty. The SQUEEZE procedure can be carried out using as input either *name*.cif and *name*.fcf data files or *name*.ins and *name*.hkl data files, where *name* is the chosen name for the data set [these are the usual files, respectively generated

by, or used as input to, the *SHELXL* program (Sheldrick, 2015)]. The former pair of files adds the benefit of the more elaborate merging of the reflection data by *SHELXL*. In the procedures described in the following sections, *PLATON* SQUEEZE and *SHELXL* are run using terminal window commands. Alternatively, leaving out the -q flag on the *PLATON* command line will give access to a graphical option menu, from where SQUEEZE and other related tools such as HYBRID (see §3.3) can be invoked. Detailed information on the SQUEEZE calculations can be found in the listing file (*name*_sq.lis). Selected details are also shown on the terminal display and in the graphics window.

2.2.1. *PLATON* SQUEEZE execution based on .cif and .fcf data. **Step 1**. Refine the solvent-free model to convergence (*i.e.* exclude any solvent that needs to be modelled by *PLATON* SQUEEZE) using the latest *SHELXL* version, starting with the files *name*.ins and *name*.hkl. Include an ACTA instruction to create implicitly a *SHELXL* LIST 4 type structure factor file (*name*.fcf). The result of that calculation will be the files *name*.cif and *name*.fcf. Do not remove the embedded *name*.res and *name*.hkl files from the *name*.cif file as they are used in Step 2 to prepare renamed input files for *SHELXL*. The averaged observed reflection intensities in the *name*.fcf file will only be used in Step 2.

Step 2. Run *PLATON* SQUEEZE based on the *name*.cif and *name*.fcf files produced in Step 1, with the terminal command PLATON -q *name*.cif. The result will be the files *name*_sq.ins, *name*_sq.hkl and *name*_sq.fab. The *name*_sq.fab file includes the solvent contribution to the calculated structure factors (*i.e.* A_{solv} and B_{solv} for each reflection *hkl*). Details of the SQUEEZE calculation are embedded in CIF format at the end of this file as well. Any additional information on the solvents can be inserted here. This should be done before the final refinement cycles, otherwise the *SHELX* checksum on the .fab file content will be compromised. The information in the *name*.fab file is recognized by the *checkCIF* validation software within *PLATON* or at http://checkcif.iucr.org. The *name*_sq.hkl file is a copy of the CIF-embedded *name*.hkl file. The same applies to the *name*_sq.ins file, apart from the insertion of the ABIN instruction (without parameter values) to instruct *SHELXL* to read the .fab file and an update of the L.S. instruction with the estimated number of additional parameters, as described in §4.3. The listing file *name*_sq.lis should be inspected for details of the SQUEEZE calculation.

Step 3. Continue the final *SHELXL* structure refinement in the presence of the files *name*_sq.ins, *name*_sq.hkl and *name*_sq.fab from Step 2 with the terminal command shelxl *name*_sq.

Step 4. Inspect the list files and validate (*i.e.* run PLATON -u *name*_sq.cif). The validation reports will be in the files *name*_sq.chk and *name*_sq.ckf.

2.2.2. *PLATON* SQUEEZE execution based on .ins and .hkl data. Using the *name*.ins plus *name*.hkl files directly, without the intermediate step in which a *name*.cif and *name*.fcf are generated, can be a convenient shortcut for running SQUEEZE without explicit reference to *SHELXL*

feature articles

refinement. The content of the *SHELXL* style `name.ins` file is used to calculate structure factors for the host. The same output files will be produced as in the procedure described in the previous section. The main, but slight, difference is the way that redundant data are merged for the calculation of the solvent contribution to the calculated structure factors. This method should not be used in cases of twinning (see §3.1). In addition, the `name_sqd.ins` and `name_sqd.hkl` files, where the solvent contribution is subtracted from the observed reflection data, are produced for backward compatibility.

2.3. SQUEEZE examples

Two examples of the application of SQUEEZE with data from published structures are discussed.

2.3.1. Example 1. The following example is based on a structure report of an organometallic compound with dichloromethane as embedded solvent (Pijnenburg *et al.*, 2014). The space group is $P2_1/c$. The low-temperature reflection data set has a resolution of 0.77 Å. Only one low-angle reflection (100) is missing. The structure was published with a disorder model for the dichloromethane molecule, which is disordered over an inversion centre. The results of three test refinements are shown in Fig. 2 (CH_2Cl_2 not included in the refinement; the inset shows the difference-map section, defined by the omitted CH_2Cl_2 coordinates, shown 'in place' in Fig. 3), Fig. 3 (CH_2Cl_2 refined using the published disorder model) and Fig. 4 (refinement result after SQUEEZE), and collected in Table 1. The difference-map section shown in the inset of Fig. 3, depicted next to the disorder model of CH_2Cl_2, clearly shows that the solvent disorder model is not completely satisfactory. There is still significant residual electron density around the solvent Cl atoms. The inset of Fig. 4 shows the phase-enhanced difference electron-density map [*i.e.* calculated using equation (5) in §2.1], to be compared with the
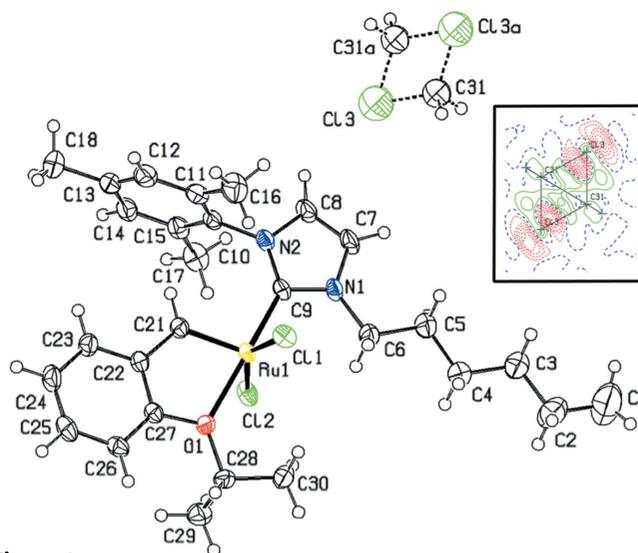
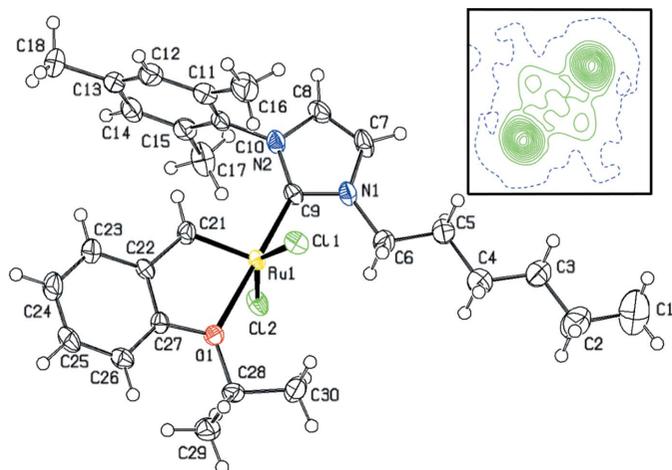**Figure 3**
Displacement ellipsoid plot (50% probability level) of the final refinement result for $2(C_{28}H_{38}Cl_2N_2ORu)\cdot xCH_2Cl_2$ as reported by Pijnenburg *et al.* (2014), with a disorder model for CH_2Cl_2. The inset next to the disordered CH_2Cl_2 shows the residual electron density in the final difference electron-density map at the solvent site. Contour levels are at 0.1 e Å^{-3}.

Table 1
Summary of three different refinement results for $2(C_{28}H_{38}Cl_2N_2ORu)\cdot xCH_2Cl_2$ with disordered solvent CH_2Cl_2.

	No CH_2Cl_2	Disorder model	SQUEEZE
$R[F^2 > 2\sigma(F^2)]$	0.0954	0.0442	0.0327
$wR(F^2)$	0.3111	0.1315	0.0921
S	3.691	1.050	1.060
ρ_{max}, e Å^{-3}	12.15	2.02	1.07
ρ_{min}, e Å^{-3}	−0.77	−1.46	−0.43
Bond precision†	0.0137	0.0058	0.0042
Occupancy, x	0	0.69	0.88
Electron count	0	29	37

† Bond precision is the average C—C bond standard uncertainty (s.u.).

Figure 2
Displacement ellipsoid plot (50% probability level) of the final refinement result for $2(C_{28}H_{38}Cl_2N_2ORu)\cdot xCH_2Cl_2$ (Pijnenburg *et al.*, 2014), without the disordered solvent CH_2Cl_2. The inset shows the residual electron density at the CH_2Cl_2 site, with contour levels at 0.1 e Å^{-3}.

Figure 4
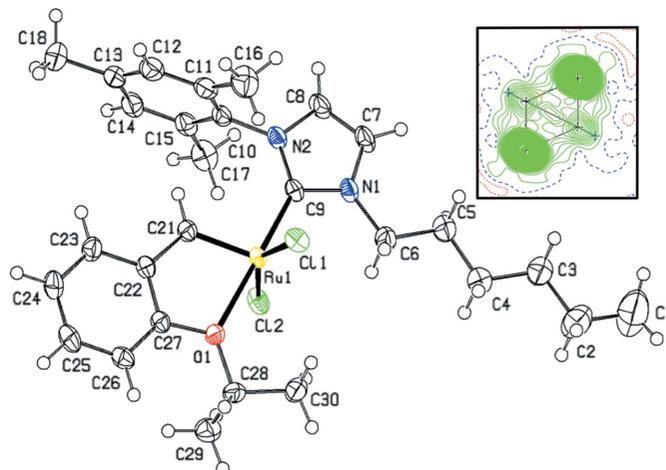Displacement ellipsoid plot (50% probability level) of the final refinement result for $2(C_{28}H_{38}Cl_2N_2ORu)\cdot xCH_2Cl_2$ (Pijnenburg *et al.*, 2014) after using SQUEEZE. The inset shows the SQUEEZE phase-improved difference electron-density map at the CH_2Cl_2 site, with contour levels at 0.1 e Å^{-3}.

Figure 5
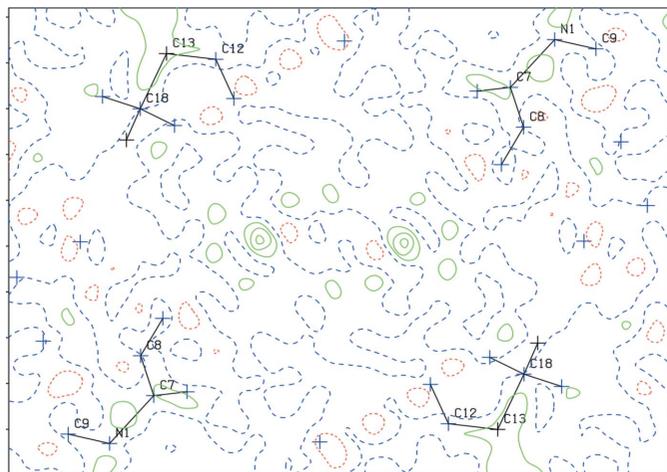Final difference electron-density map at the CH_2Cl_2 site (in the centre of the picture) after SQUEEZE for $2(C_{28}H_{38}Cl_2N_2ORu)\cdot xCH_2Cl_2$, with contour levels at 0.1 e Å^{-3}.

normal difference map shown in Fig. 2. Fig. 5 shows the final difference electron-density-map section through the region of the removed disordered CH_2Cl_2. It is interesting to notice that the displacement ellipsoid plots for the Ru complex molecule in all three refinements are nearly indistinguishable. SQUEEZE calculates an electron count of 37 electrons at the CH_2Cl_2 site, where 42 are expected for full occupancy. From this ratio, a tentative CH_2Cl_2 occupancy of 0.88 can be calculated and compared with the value of 0.69 obtained in the least-squares refinement.

2.3.2. Example 2. The crystal structure of the ionic coordination complex $[Mn(C_{15}H_{11}N_3)_2](NO_3)_2\cdot H_2O$ (Rompel *et al.*, 2004) was published without a disorder model for the NO_3^- anion. The authors reported some residual electron density in the vicinity of the nitrate anion. A difference electron-density map indeed confirmed the suspected nitrate disorder. This structure report was selected to see whether SQUEEZE could address this disorder. It was also considered to be a good test to see whether the expected number of electrons for a necessarily fully occupied nitrate site could be recovered with the electron count in the 'solvent'-accessible region.

The structure was published in the tetragonal space group $I4_1/a$, with $R[F^2 > 2\sigma(F^2)] = 0.050$, $wR(F^2) = 0.153$, $S = 1.033$, mean $\sigma(C—C) = 0.004$ Å and $\rho_{max} = 1.14$ e Å^{-3}. The second weight parameter value is 14.4, which usually indicates some unresolved model or data issues. In fact, it is noted that many displacement ellipsoids point in a 'preferred' direction. There are eight low-order missing reflections.

A SQUEEZE calculation, with the nitrate anion removed from the structure model, followed by *SHELXL* refinement, converged at $R[F^2 > 2\sigma(F^2)] = 0.039$, $wR(F^2) = 0.112$, $S = 1.029$, mean $\sigma(C—C) = 0.003$ Å and $\rho_{max} = 0.69$ e Å^{-3}. The number of electrons recovered from the NO_3^- voids amounts to 30 electrons per void of 43 Å^3, a slight underestimation of the expected 32 electrons for NO_3^-. Estimates were used for the difference-map coefficients of the eight missing reflections,

as described in §§3.2 and 4.2. The second optimized weight parameter value went down to 6.87 but the 'preferred' displacement ellipsoid direction remained about the same. The latter is not serious but might point to an additional unresolved data problem, unrelated to the resolved nitrate disorder.

3. Additional new SQUEEZE features

3.1. SQUEEZE and twinning

The current *SHELXL* release (*SHELXL2014/7*) allows for the output of a detwinned $F_o^2/F_c^2/\sigma(F_o^2)$ reflection file (*i.e.* a LIST 8 type `.fcf` file). This makes it possible, in cases for which detwinning succeeds, to SQUEEZE twinned structures. Again, the detwinned data are used only for the generation of the `.fab` file. Least-squares refinement will be based, as usual, on the `.res` and `.hkl` data embedded in the `.cif`. The required changes to the procedure described in §2.2 are minimal. An additional LIST 8 instruction should be included in the `name.ins` file in Step 1. Both BASF/TWIN and BASF/HKLF 5 twin refinements are accommodated in this way. The other steps of the SQUEEZE procedure are unchanged. It might be necessary to repeat the SQUEEZE procedure, for example, when the refined value of the twin domain ratio has changed significantly (see §3.3).

3.2. Missing reflections and outliers

The previous implementation of the SQUEEZE procedure, tailored for refinement using *SHELX97*, assumed that an essentially complete data set has been supplied. This is of particular importance when the reported electron count obtained by integration of the electron density found in the solvent region of the difference map is to be used to estimate the number of solvent molecules involved. Missing strong low-order reflections make the estimated value of the electron count unreliable for that purpose, although the calculation of the solvent contribution to the structure factors still works reliably. With today's two-dimensional detector systems, it is sometimes more difficult to measure complete data sets, in particular as far as low-angle reflections are concerned, compared with data collections in the past using serial detectors. This is mainly related to restrictions caused by beam-stop configurations. A proper data-collection strategy should include additional scans, possibly with a longer crystal-to-detector distance, as part of the data-collection protocol. As an alternative, a computational solution for the missing-reflection problem is discussed in §4.2.

Intensity outliers, such as those due to secondary extinction, pose a similar problem to that discussed above. Omitting these observations from the `name.fcf` file and substituting them automatically by estimates as part of the difference electron-density-map calculation (§4.2) is probably the best option. Of course, those reflections affected by extinction should be kept in the `name.hkl` file and a correction for extinction should be attempted in the least-squares refinement.
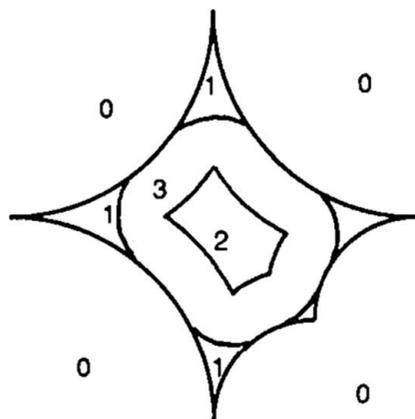
Figure 6
A two-dimensional cartoon of the definition of the solvent-accessible region in a crystal structure. Region 0 represents the van der Waals volume of the atoms of the host structure. The outline of region 2 defines the location of the centre of a hypothetical atom with radius 1.2 Å that just touches the van der Waals surface of the host. The region enclosed by rolling the hypothetical sphere with its centre over the outline of region 2 defines the solvent-accessible region (2 + 3). Regions that are inaccessible for solvents are indicated as 1.

3.3. Recycling of the SQUEEZE procedure

The procedure described in the *BYPASS* paper (van der Sluis & Spek, 1990b) involved two nested optimization cycles. The inner cycle described in that paper coincides with the SQUEEZE procedure described above. An outer optimization cycle then repeated the inner SQUEEZE cycle with the supposedly slightly changed positional and displacement parameters of the host structure. In practice, it turned out that this outer cycle refinement had little effect and is thus not needed in well defined applications of SQUEEZE. However, in view of the ample availability of computing resources, and in particular for application under less well defined circumstances, an option was implemented to automate this outer optimization cycle as well. This is now available as the PLATON -qn *name*.cif terminal window command, where n is the number of outer cycles. Alternatively, the HYBRID option in the graphical menu can be used instead of the SQUEEZE graphical menu option. The nested two-cycle optimization option might be needed in cases with troublesome detwinning, as well as in cases where the scattering contribution of the solvent to the calculated structure factors is relatively large (see the example in §3.3.1).

3.3.1. HYBRID test. For this test example, a synthetic data set was generated with the HKLF-GENER utility in *PLATON*, based on the atomic coordinates of the JORFEB entry (Calvert *et al.*, 1992) in the Cambridge Structural Database (CSD; Allen, 2002). This structure, $C_8H_2Cl_4N_2O_7$·$C_4H_{10}O$, is triclinic, with space group $P1$, and the solvent is diethyl ether. The generated data set includes a complete set of Friedel pairs [Flack parameter (Flack, 1983) value = 0] and has a resolution of 0.77 Å. Some noise was added to the reflection data to avoid numerical instability of the *SHELXL* refinement. The starting confidence factors are $R[F^2 > 2\sigma(F^2)] =$

0.001, $wR(F^2) = 0.003$ and $S = 0.293$ for the complete structure. Least-squares refinement of this structure after removal of the diethyl ether solvent converges at $R[F^2 > 2\sigma(F^2)] = 0.177$ and $wR(F^2) = 0.435$. With the subsequent SQUEEZE calculation, the contribution of 36 electrons was recovered, compared with the expected 42 for diethyl ether. Subsequent refinement with *SHELXL* converged at $R[F^2 > 2\sigma(F^2)] = 0.0276$, $wR(F^2) = 0.0763$ and $S = 1.068$, with a Flack value of 0.04 (6). The alternative HYBRID calculation converged after a few cycles to $R[F^2 > 2\sigma(F^2)] = 0.0196$, $wR(F^2) = 0.0533$ and $S = 1.023$, with a Flack value of 0.03 (4). The number of recovered electrons increased to 41.

4. Computational details

All SQUEEZE calculations are based on F_{obs} and F_{calc} values and are done in the triclinic system. Nonprimitive Bravais lattices are not transformed to a primitive lattice. The supplied reflection data are expanded to half a sphere of Friedel-averaged reflections. F_{obs} values for reflections with negative measured intensities have been set to zero.

4.1. Solvent-accessible region

Solvent-accessible regions (SARs) in a structure are defined on a grid of approximately 0.2 Å. It is assumed that the lattice solvents have only van der Waals contacts to the host crystal structure. It is also assumed that the major contacts involve H atoms with a van der Waals radius of 1.2 Å. The solvent region can then be defined (Fig. 6) as the volume enclosed when a sphere of radius 1.2 Å is rolled over the surface of the host. Thus, a void in a structure should at least have a volume of $4\pi(1.2)^3/3 = 7.2$ Å^3 to be relevant. With this procedure, cusp-shaped spaces between van der Waals surfaces that cannot accommodate such a sphere are not included in the solvent-accessible volume. Such spaces can constitute of the order of 30% of the unit-cell volume when the solvent-accessible volume is zero. The search is done over the whole unit cell. All voids are detected individually and reported with their location, volume and shape. This algorithm works well in general, although strongly hydrogen-bonded water-molecule sites might escape detection. This is not a problem, however, because they will be detected easily anyway as part of the structure solution. The void surface can be displayed with the SOLV PLOT graphical menu option.

The void map, as detailed above, has to be mapped onto a second grid with a grid size of the order of 0.3 Å, which is used as a mask on the electron density for the back-Fourier transform calculation. That calculation is done with a fast Fourier transform base-2 algorithm. The number of grid points, 2^N, in each of the three map directions is chosen with a value of N such that $2^N > 2m + 1$, where m is h_{max}, k_{max} or l_{max}.

4.2. Phase-improved difference electron-density map and F_{000}

The difference electron-density-map calculation needs to be recycled in order to obtain meaningful values for the solvent

contribution to the calculated structure factors and the electron count in the voids (F_{000}) (see §2.1). This applies to non-centrosymmetric structures in particular. Electron-density peaks will appear at only half height in the first difference map (Lipson & Cochran, 1966). The first difference map is calculated with F_{000} set to zero. Subsequently, all grid point densities outside the SARs are set to zero before back-Fourier transformation. This procedure is repeated until convergence. The value of F_{000} is reset each time to the electron count in the current SAR. Convergence is considered to be reached when both F_{000} and the R value become stable.

Missing Fourier coefficients to be used in the next difference-map calculation are given the value obtained from the back-Fourier transform of the previous difference-map optimization cycle. This approach appears to work reasonably well. It was tested with examples where measured reflections with strong difference-map contributions were deliberately left out of the SQUEEZE calculation. The reason why this works well must be that all density in the difference electron-density map outside the solvent-accessible volume is set to zero prior to the back-Fourier transform.

4.3. The number of additional 'refinement' parameters

Standard uncertainties on parameters determined by least-squares depend, among other factors, on the number of refined parameters. When SQUEEZE is used as part of the structure refinement, the issue arises about how many additional parameters are to be counted for the solvent contribution. The number of refined model parameters is smaller than when a disorder model is refined, and the data-to-parameter ratio is thus larger. These numbers are likely correct when the voids have lost their content and the structure has survived. It is less obvious how many additional parameters are involved with the application of SQUEEZE. An estimated value can be supplied as the NEXTRA parameter value on the *SHELXL* L.S. instruction record. By default, *PLATON* SQUEEZE will estimate this number with the expression NEXTRA $= (En)/(Zm)$, where E is the number of recovered electrons in the unit cell, Z is the number of asymmetric units, n is the number of parameters usually refined for a CH_2 fragment (*i.e.* 9) and m is the number of electrons in a CH_2 fragment (*i.e.* 8). This formula has the pleasing property that it vanishes when there is no residual density found in the solvent-accessible region of the structure.

A referee proposed an interesting idea to derive s.u.'s for the solvent F_{calc} (or A_{calc} and B_{calc}) values as a better alternative to the NEXTRA value problem. Those values might be derived from the difference map 'quality' or its contributing Fourier coefficients. The solvent contribution s.u.'s thus obtained could then be included in the weights assigned to the observed reflection data in the least-squares refinement. Implementation of this idea would require an extension of the definition of the contents of the .fab file, since it would need to include the s.u.'s and not only the solvent A_{calc} and B_{calc} entries. An alternative would be the debatable change of the

$\sigma(F^2)$ values of the observed data, reflecting the solvent contribution s.u.'s.

5. Practical issues

5.1. SQUEEZE *versus* a constrained/restrained solvent disorder model

The development of an atomistic model of the disordered solvent is generally to be preferred wherever possible, in particular when the disorder can be described easily with constraint and restraint tools, such as those available in *SHELXL*. In that way, a crystal structure is completely characterized. Examples are solvents such as toluene disordered over an inversion centre. Cases such as those where a tetrahydrofuran molecule is disordered about a threefold axis are more difficult to model satisfactorily (Knotter *et al.*, 1989). Even when multiple disorder components (PARTs in *SHELXL* language) have been used, one can still end up with significant unaccounted-for residual density in a difference electron-density map. Sometimes the nature of the solvent mixture present in the voids of the structure is unclear. The structures of metal–organic frameworks (MOFs) are notorious examples.

The time invested in devising an unsatisfactorily parametrized disordered solvent model is not always considered to be worth the effort. This applies in particular in the context of a routine (service) structure determination intended to characterize the chemistry of the main component in the crystal. The detailed structure of the embedded solvent molecules is usually already known. That structure can certainly not be determined more accurately from the disorder model and is rarely relevant for the main component of interest in the crystal structure.

The SQUEEZE approach can be an efficient and effective alternative for routine structure determinations with problematic but irrelevant solvent disorder. The same applies to anions that are often extremely disordered, such as PF_6^-. Its main purpose can be to bring down the R value as proof that an unaccounted-for solvent is the main reason for a high R factor. The main concern is not to overstretch its application to conditions with poor or limited data sets. The SQUEEZE procedure needs a sufficient ratio of reflection data to least-squares parameters to avoid over-fitting. The procedure followed should be well documented in the final structure report and associated CIF archive. This would involve the reporting of the number of voids per unit cell, their volume, their shape, the number of electrons per void and some estimate of the likely solvent content this might correspond to. *SHELXL* automatically retains the original reflection data (unless deliberately removed by the authors). Deposition and archiving of the reflection data is good scientific practice. In that way, nothing is lost. Follow-up calculations with the archived data are still possible when details of the host structure and/or disordered solvent become of interest for reasons other than those of the original authors. This applies in particular when the crystals or experimental data are

difficult to obtain again. The CCDC now accepts and archives CIFs with embedded reflection data and these can be downloaded free of charge.

5.2. Requirements and restrictions

SQUEEZE is designed for 'small-molecule structures' and is most effective when based on a complete and reliable data set with sufficient resolution. Low-temperature data are strongly advised for better resolution and to avoid loss of solvent during data collection. Systematic errors in the reflection intensities should be taken care of. This applies in particular for low-angle reflections that are (partly) obscured by the beam stop or affected by secondary extinction. These reflections may seriously affect the value of the electron count of the electron density in the disordered solvent region, but not the outcome of the SQUEEZE procedure. The host structure should be completely modelled, including H atoms and any disorder without unresolved residual electron density, because of its impact on the difference map in the solvent region. SQUEEZE cannot properly handle cases of coupled disorder affecting both the host and the solvent region (*i.e.* when space in the average structure is occupied partially by both the host and the solvent). The presence of significant anomalous scatterers in the solvent region *cannot* be used for the determination of the absolute structure of a light-atom host if SQUEEZE has been used: the SQUEEZE calculations are necessarily done with Friedel-averaged data that have been corrected for anomalous dispersion contributions from the host structure. One of the reasons for this is that a complete set of Friedel pairs would be needed for the calculation of the difference electron-density map, a requirement not strictly needed for the least-squares refinement. The subsequent *SHELXL* refinement, of course, includes the anomalous scattering contribution of the host structure. The same Friedel-averaged solvent contribution is added to both Bijvoet-pair related reflections. The effect will be a higher s.u. on the Flack parameter value than when an atomistic solvent model is refined. Also, there should be no unresolved charge-balance issues that might affect the conclusions about the chemistry involved, such as the valency of the metal in the host part of the structure, if SQUEEZE removes undetected counter-ions. Using SQUEEZE as part of the MOF soaking method (Inokuma *et al.*, 2013), where the interest lies in the guest region as opposed to the host region, can be very challenging, is not recommended and should be done with extreme care when attempted.

5.3. Structure validation

checkCIF will suppress certain validation messages when it detects details about the use of SQUEEZE in the CIF. Unfortunately, the current CIF data definitions for `_chemical_formula_sum` and `_chemical_formula_moiety`, and related quantities such as the linear absorption coefficient (μ value) and the molecular weight, are not fully adequate when reporting details of SQUEEZEd solvents. *checkCIF* will report calculated values for the moiety formula, sum formula,

M_r, D_x, μ and F_{000} that are based only on the model parameters. The reported and calculated values should currently be compared manually for consistency. IUCr journals advise authors to include as much as possible of the available information about the SQUEEZEd solvents and their estimated quantities in the `_chemical_formula_sum`, `_chemical_formula_moiety` and derived quantity data entries. Details about the use of the procedure should also be included in the `_platon_squeeze_details` or `_exptl_special_details` sections of the CIF and in the experimental section of the manuscript.

6. Conclusions

The SQUEEZE approach for the handling of disordered solvents in a least-squares structure refinement works well, given sufficient and reliable experimental data. The geometry of the main part of the structure is often at least as reliable and accurate as the geometry achieved with a parameterized solvent disorder model. It should be noted that the examples given in §2.3 fulfil the above criteria. The original *BYPASS* method (van der Sluis & Spek, 1990*b*) was developed with a small solvent-to-host ratio in mind. As it happens, once tools are available they will also be used under less optimal or unintended conditions. SQUEEZE is currently also applied in cases where the structure has large voids but only a limited number of reflection data are available, which might be a concern. Fortunately, the SQUEEZE procedure turns out to be rather robust, in that it generally converges to lower R values and improves the geometry of the host structure.

Several test calculations to investigate possible limitations of the SQUEEZE technique have been done with synthetic normal resolution 'observed' data. One test involved a structure with two independent organic molecules where one of the two molecules is SQUEEZEd out. It resulted in a reasonably refined structure of the other molecule, although with a few percent higher R value than the theoretical value of zero percent. Tests aimed at investigating the effect of the resolution of the data show slowly increasing R factors with a diminishing resolution of the data. More tests are planned, including the leverage effect of the solvent contribution to the structure factors on the model.

Although the refinement of an atomistic disorder model might be the preferred procedure, it may be useful to follow both routes and compare the results. The phase-improved difference electron-density map obtained with SQUEEZE could also provide a key for the refinement of an improved solvent disorder model.

Smeared residual electron density in voids in a structure is not always detected by peak search routines that assume three-dimensional Gaussian-shaped densities for peak fitting. An example is the structure determination of (−)-crebanine (Duangthongyou *et al.*, 2011), where the authors indeed reported 'empty' large voids (maximum residual electron density = 0.55 e Å^{-3}) that, on close inspection, are found to be solvent-filled infinite channels. Voids in a structure are easily visualized with the *CAVITY* tool (Mugnoli, 1992), as imple-
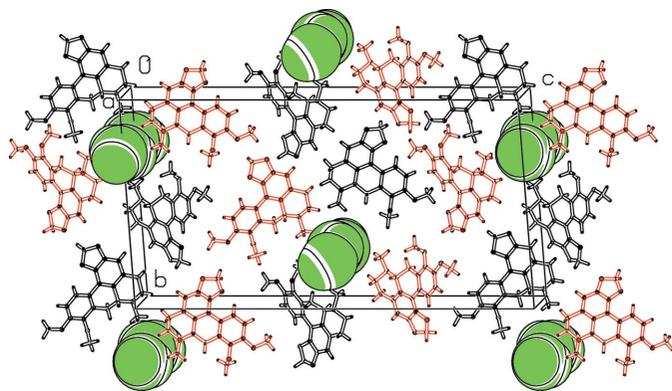
Figure 7
PLATON CAVITY illustration (Mugnoli, 1992) of the structure of (−)-crebanine, viewed down the *a* axis. The green spheres, with a contact radius of at least 1.2 Å, represent the infinite solvent-accessible channels in this structure.

Figure 8
A contoured difference electron-density section in the $(x, y, 0)$ plane, depicting one of the continuous density channels shown in Fig. 7. Contour levels are at 0.1 e $Å^{-3}$.

mented in *PLATON/PLUTON*. Fig. 7 shows the infinite channels in the (−)-crebanine structure, depicted as a chain of spheres with a minimum contact radius to the host of 1.2 Å. Fig. 8 depicts a contoured section of the continuous electron density in the difference density map. The volume of the channel within the bounds of the unit cell amounts to 131 $Å^3$. The electron count in this channel is 33 electrons.

It should be noted that a SQUEEZE-type algorithm could easily be included as part of a least-squares refinement program. The route chosen by the developer of *SHELXL* is more flexible with its `.fab` file input facility for externally provided contributions to the calculated structure factors. This allows for easy interfacing and testing of alternative approaches to addressing disordered solvent problems and for other applications. Alternatively, *SHELXL* refinement can also be called from within *PLATON* SQUEEZE. This is actually implemented as the HYBRID tool in *PLATON*.

Inspection of the 2014 release of the CSD indicates that the use of the earlier versions of SQUEEZE has been reported in more than 13 600 structure publications. This is certainly an underestimate; not all cases where SQUEEZE was used have been detected by the CCDC staff. This might be due in part to authors not including the relevant information in the CIF, despite the advice in the *PLATON* SQUEEZE program output listings. The current *SHELXL* version (*SHELXL2014/7*) will now do this automatically by default. In this way, both the information on the refinement procedure used and the unmerged data for future alternative calculations are archived.

7. Program availability

SQUEEZE is implemented as a tool in the *PLATON* program (Spek, 2003, 2009), which also includes the *checkCIF* tool that is used as part of the IUCr *checkCIF* facility (http://checkcif.iucr.org). The native Fortran source is available for the UNIX platform (Linux, Mac OS X), where it depends for its graphics on the libraries of the X-Windows system. Copies of the source code, simple compilation instructions for using the freely available GNU gfortran compiler and additional infor-
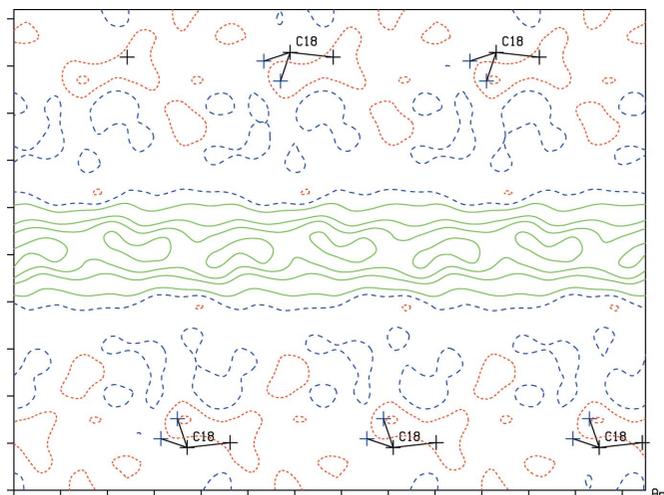
mation are available from http://www.platonsoft.nl. A *PLATON* executable for the Microsoft Windows platform is available from http://www.chem.gla.ac.uk/~louis/software/platon. *PLATON* displays its version date at the top of its main menu. The current version of the program on the download server mentioned above is also shown when the computer is connected to the internet and the 'curl' utility is installed. The UNIX source of the program will be downloaded by clicking on the current version info.

PLATON is a research program. It is regularly updated with new features based on our own research and that of its users. The communication of ideas and enquiries on issues encountered has been most helpful for its development. New ideas and error reports are welcome, where relevant, with reference to the program version being used and preferably with the associated data.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K. & Watkin, D. J. (2003). *J. Appl. Cryst.* **36**, 1487.
Calvert, J. L., Gordon, J. L. M., Hartshorn, M. P., Robinson, W. T. & Wright, G. J. (1992). *Aust. J. Chem.* **45**, 713–719.
Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. (2009). *J. Appl. Cryst.* **42**, 339–341.

Duangthongyou, T., Makarasen, A., Techasakul, S., Chimnoi, N. & Siripaisarnpipat, S. (2011). *Acta Cryst.* E**67**, o402.

Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.

Hooft, R. W. W., Straver, L. H. & Spek, A. L. (2008). *J. Appl. Cryst.* **41**, 96–103.

Hooft, R. W. W., Straver, L. H. & Spek, A. L. (2010). *J. Appl. Cryst.* **43**, 665–668.

Inokuma, Y., Yoshioka, S., Ariyoshi, J., Arai, T., Hitora, Y., Takada, K., Matsunaga, S., Rissanen, K. & Fujita, M. (2013). *Nature*, **495**, 461–466.

Knotter, D. M., van Koten, G., van Maanen, H. L., Grove, D. M. & Spek, A. L. (1989). *Angew. Chem. Int. Ed. Engl.* **28**, 34–35.

Lipson, H. & Cochran, W. (1966). *Crystalline State*, Vol. III, p. 318. London: Bell.

Mugnoli, A. (1992). *Z. Kristallogr. Suppl.* **6**, 530.

Petříček, V., Dušek, M. & Palatinus, L. (2014). *Z. Kristallogr.* **229**, 345–352.

Pijnenburg, N. J. M., Tomás-Mendivil, E., Mayland, K. E., Kleijn, H., Lutz, M., Spek, A. L., van Koten, G. & Klein Gebbink, R. J. M. (2014). *Inorg. Chim. Acta*, **409**, 163–173.

Rompel, A., Bond, A. D. & McKenzie, C. J. (2004). *Acta Cryst.* E**60**, m1759–m1760.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Sheldrick, G. M. (2015). *Acta Cryst.* C**71**, 3–8.

Sluis, P. van der & Spek, A. L. (1990a). *Acta Cryst.* C**46**, 883–886.

Sluis, P. van der & Spek, A. L. (1990b). *Acta Cryst.* A**46**, 194–201.

Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.

Spek, A. L. (2009). *Acta Cryst.* D**65**, 148–155.

Wehman, E., Van Koten, G., Jastrzebski, J. T. B. H., Rotteveel, M. A. & Stam, C. H. (1988). *Organometallics*, **7**, 1477–1485.