



intermolecular contacts, that are clear for experienced investigators, are misunderstood or ignored.

Not all published crystal-structure studies are based on the best attainable diffraction data, but rather on data that are considered to be suitable for the purpose of the study (Thompson, 2019). A good-looking ‘*ORTEP*’ illustration in a chemical journal is often considered as sufficient proof of the reported chemistry. Crystal structures are not always ‘well-behaved’, often related to sub-optimal crystal quality, experimental issues, or disorder or twinning. Structural disorder can be pursued in great detail with constraints and restraints or handled in a way sufficient for the study of the main species of interest, for example using tools such as *PLATON/SQUEEZE* (Spek, 2015) for handling disordered solvents as part of the structure refinement. Readers need to know whether reported geometrical details, such as intermolecular contacts or bond distances, involve restrained or constrained parameters. All this and residual issues should be clearly documented in the structure report.

Reported crystal structures end up in the Cambridge Structural Database (CSD; Groom *et al.*, 2016) as curated but otherwise largely unqualified database entries. It is therefore of the utmost importance that all experimental data, data reduction and refinement details are archived as well. In that way, users of those data can do their own evaluation of the reported crystal structures or use the experimental data for more detailed studies in a different context. Those data may be hard to obtain again, either because of the involved synthesis of the compound of interest or the difficulty in obtaining a particular polymorph again.

During the 1980s, with the growth in the number of structure reports, the need for a standard computer-readable format for the exchange and archival of crystallographic data became clear. Obviously, that would avoid the error-prone retyping of numerical information, such as the refined parameter values, into the desired format from typo-prone manuscripts. It would also make the diffraction data more accessible for processing. Such a standard, called CIF, was proposed by Hall *et al.* (1991). One of the early adopters of that format was George Sheldrick with his new and popular 1997 version of the *SHELXL* (Sheldrick, 2008) structure-refinement program, making CIF widely accepted as a data-archival format, along with the CIF-style structured FCF file with observed and calculated reflection data. Those data were previously available only as printed  $F_{\text{obs}}$ ,  $F_{\text{calc}}$  and  $\sigma(F_{\text{obs}})$  tables. Interestingly, *SHELXL* does not read and write CIFs for its own input and output. For efficiency, it uses for its calculations the non-CIF-structured RES and HKL computer- and human-readable files for the input and storage of refinement details and reflection data, respectively. Those files have also become a *de facto* exchange standard for a number of applications and programs such as the CSD utilities (Groom *et al.*, 2016). Today, the archived material is expected to include not only the structural results, but also the unmerged reflection data (either as a CIF-structured reflection-data loop or as embedded *SHELXL*-style *name.hkl* file) and details of the data reduction and non-standard refinement details (CIF-

structured or as an embedded *SHELXL*-style *name.res* file). Those files can easily be extracted from the CIF with the *shredcif* utility and used for alternative refinements. Most current structure-refinement packages include that information in the CIF file.

The introduction of the CIF standard also opened the way for the automated checking of the archived data for their internal consistency and completeness, which was needed to handle the exploding number of structure reports. The International Union of Crystallography (IUCr) journal *Acta Crystallographica Section C* pioneered automated structure validation as a tool for authors, referees and readers. This project started with flagging missing data and inconsistencies and testing numerical data against expected values. The result was a report consisting of a list of ALERTS, with associated A, B and C levels of importance, for issues that needed to be addressed. It should be clear that ALERTS are not necessarily errors. They might also point to interesting features in a crystal structure. G-level ALERTS, which were introduced later, are mostly informative but are not to be neglected. All ALERTS should be checked by the authors: a set of lower-level ALERTS may in combination point to a serious issue that needs to be addressed.

Currently, *IUCr/checkCIF* includes, in addition to the above-mentioned ALERTS, *PLATON/checkCIF* (Spek, 2003) based ALERTS (starting with ‘PLAT’ followed by a number) implementing some 500 additional tests for issues such as missed symmetry, missed twinning and missed solvent-accessible voids in the crystal structure. This paper is a follow-up to earlier papers on the validation issue and *checkCIF* (Spek, 2003, 2009, 2018). It offers additional background details for a number of more commonly issued ALERTS along with suggestions on how to address them.

Three types of archived information should be available: (a) crystal data such as cell parameters, symmetry, refined model parameters, derived geometry and *R* values – those data (*name.cif*) are checked for completeness, internal consistency and expected parameter values; (b) the final ‘FCF’ file (*name.fcf*) with calculated and observed intensities and associated  $\sigma(I)$ /weight to check the reported final *R* and *S* values along with an analysis-of-variance analysis of the refinement results; (c) the unmerged *h*, *k*, *l*,  $I_{\text{obs}}$ ,  $\sigma(I_{\text{obs}})$  reflection data [CIFs created using *SHELXL* (Sheldrick, 2015b) include this ‘HKL’ file (*name.hkl*) along with the final refinement instruction ‘RES’ file (*name.res*) as embedded data as values of the *\_shelx\_hkl\_file* and *\_shelx\_res\_file* CIF datanames along with associated checksums reported as values of the datanames *\_shelx\_hkl\_checksum* and *\_shelx\_res\_checksum*, respectively]. No archival of the final FCF file is needed in the case of matching calculated and reported checksums for the *name.res* and *name.hkl* files. In the case where a final *name.fcf* is recreated with *SHELXL* from the embedded *name.hkl* and *name.res* files and recalculated, the final *R* and *S* values are compared with those reported in the CIF for consistency. Otherwise the final *name.fcf* file should be archived as well.

The alternative datanames *\_iucr\_refine\_instructions\_details*, *\_iucr\_refine\_reflections\_details* and *\_iucr\_refine\_fcf\_details* are

available to embed refinement and unmerged HKL and FCF data associated with non-*SHELXL*-based refinement programs. *OLEX2* (Dolomanov *et al.*, 2009) offers both a refinement based on *SHELXL* and a *SHELXL*-compatible refinement tool (REFINE). In the former case, standard *SHELXL* *name.cif* and *name.fcf* files are created. In the latter case, the unmerged reflections are included in the CIF with a looped structure and the *name.fcf* file embedded in the *\_iucr\_refine\_fcf\_details* field. The *CRYSTALS* (Betteridge *et al.*, 2003), *JANA* (Petricek & Dusek, 2000) and *WinGX* (Farrugia, 2012) refinement packages follow similar archival schemes. *A Guide to CIF for Authors* is available at [https://www.iucr.org/\\_data/assets/pdf\\_file/0019/22618/cifguide.pdf](https://www.iucr.org/_data/assets/pdf_file/0019/22618/cifguide.pdf).

## 2. Experimental data and data reduction

Most current crystal structure reports are based on data collected with area detectors as opposed to previously used serial detectors. Some of the ALERTS and validation criteria have their origin in the use of serial detectors. They are kept for backward compatibility. Many non-applicable ALERTS are suppressed automatically when obvious from the other information available in the CIF.

### 2.1. Experimental data

It is good practice to collect diffraction data up to at least *copper sphere* resolution, *i.e.*  $\sin(\theta)/\lambda = 1/1.5418 = 0.65 \text{ \AA}^{-1}$ , in order to achieve a good data-to-refined-parameter ratio. That resolution corresponds to  $27.5^\circ$  in  $\theta$  for Mo  $K\alpha$  radiation and a real space resolution of  $0.77 \text{ \AA}$ , *i.e.* about half of the C—C bond distance. The reflection data set is expected to be essentially complete and have a high redundancy (*i.e.* multiple-measured data). The latter allows for the calculation of the  $R_{\text{int}}$  value as a measure of the internal consistency of similar intensity measurements and their use as a basis for multi-scan-based correction for absorption. Sometimes, the above resolution can only be achieved, with a significant fraction of reflections above the noise level, when data are collected at low temperature. Unfortunately, physical restrictions such as data collection with a Cu  $K\alpha$  X-ray source, will allow data to be collected only up to a resolution of about  $0.6 \text{ \AA}^{-1}$ . For that reason, data sets are expected to be at least complete up to that value (*i.e.*  $\sim 25^\circ$  in  $\theta$  for Mo  $K\alpha$ ). This does not imply that meaningful data should be removed from the structure analysis beyond that resolution without a valid reason. Arguments based on limited available instrument time do not count anymore with currently used area detectors. A sensible cut-off value might be a value beyond which there is only noise, not when there is still a significant number of reflections with  $I > 2\sigma(I)$  at the applied cut-off. A poor data-collection strategy may lead to a cusp of missing reflections. Randomly missing reflections, *e.g.* due to overflows, may be acceptable for the least-squares structure refinement. Unfortunately, incomplete data sets may be unavoidable in case of data collection on a low-symmetry crystal in a high-pressure cell. Particular attention is needed for the low-order reflec-

tions when difference-density maps are of interest. Extra efforts, such as a special scan with the detector further away, should be made to include them or remove them from the data set when (partly) obscured by the beamstop. Low-order reflections may carry significant information.

It might be helpful to make a record of special features encountered during diffraction image processing such as unindexed (weak) spots, streaks, diffuse scattering, and split-up or broadened reflection profiles. They might provide a clue to the reason why or when problems arise during structure determination and refinement.

Unit-cell dimensions, as all measured and derived quantities, are to be accompanied with a standard uncertainty. They are commonly based on a least-squares treatment of the setting angles of a number of reflections. With serial detectors, that number was traditionally around 25 whereas with area-detectors, approximate setting angles of many thousands of reflections are used. Standard uncertainties tend to be much smaller in the latter case and often unrealistic when confronted with the variance in the cell dimensions obtained in unit-cell dimensions for crystals from the same or different batches. It is likely best to accept and report standard uncertainties as they come out of the least-squares-fit program and report them along with the number of reflections involved and the  $\theta$  range. Consider them as a measure of internal consistency and do not arbitrarily apply an empirical and undocumented multiplier to them to obtain supposedly realistic values. Otherwise, users of those data might unknowingly apply yet another multiplier. Errors in cell dimensions and their standard uncertainties may have significant effects on the reported derived geometry values. The reported wavelength should have sufficient significant figures. Their values are directly related to the wavelength value used in the calculation of the cell parameters.

### 2.2. Data reduction

**2.2.1. Absorption correction.** Given a crystal with faces that can be indexed and bathed fully in the homogeneous part of an X-ray beam, there are numerical and analytical ways to correct for absorption. This ideal is rarely achieved in current practice. Crystals are rarely bathed homogeneously, as assumed by the above correction algorithms. Instead, the experimental multi-scan technique is used where an absorption surface is constructed using the intensity differences between supposedly identical reflections (either symmetry related or measured on different orientations about the diffraction vector). It is important to report the actual correction range and compare that range with the range expected based on the crystal dimensions. In addition, a significant reflection redundancy is in order for a meaningful correction. Sometimes, more than one absorption correction is concatenated (*e.g.* numerical or analytical correction followed by multi-scan-based correction). This should be reported in the CIF in the special details section.

**2.2.2. Symmetry.** The symmetry of the electron density of a crystal structure can, in the majority of reported cases, be

associated with one of the 230 distinct three-dimensional symmetry patterns. Each of those patterns has a Hermann–Mauguin symbol associated with it, *e.g.*  $P2_1/c$ . Those symbols imply a standard unit cell and origin choice setting, in this case at the inversion center, along with a twofold screw axis and a *c*-glide. Different choices include  $P2_1/a$  and  $P2_1/n$  based on a different choice of the unit cell or even with the origin off the inversion center. Even  $P2_1/b$  is possible if the unique axis is taken to be *a* or *c*. There can be various reasons not to report a structure in the standard setting. Sometimes it is considered beneficial to select the  $P2_1/n$  setting rather than  $P2_1/c$  because that unit cell has a closer to  $90^\circ$   $\beta$  angle, which leads to better convergence and lower correlations between the refined atomic parameters (Feast *et al.*, 2009). In addition, a non-standard choice may be useful when its relation with another structure is to be shown (*e.g.* a redetermination or a phase transition).

Fortunately, the symmetry of a structure, needed for the various computations, is completely described by the set of symmetry operators or the subset of space-group generators that can be expanded into the full set. Those symmetry operators are not easily or uniquely derivable from the Hermann–Mauguin symbols. The alternative Hall symbol (Hall, 1981) has been introduced to record the essential space-group symmetry generators in one symbol and suitable to be used to uniquely derive the full set of symmetry operators from them. As an example, the Hall symbol for  $P2_1/n$  with the origin on the inversion center is  $-P\ 2yn$ , where ‘ $-$ ’ represents the inversion operation  $-x, -y, -z$  and ‘ $2yn$ ’ the screw axis operation  $\frac{1}{2} - x, \frac{1}{2} + y, \frac{1}{2} - z$ . The two symmetry operations are sufficient to create the complete set of symmetry operations of the space group. It should be noted that Hall symbols are not necessarily unique. Different choices of generators may lead to the same set of symmetry operators. Hall symbols can be especially useful as a space-group-symmetry identifier in the case of a non-standard setting where no Hermann–Mauguin symbol has been officially defined. A CIF is expected to report both the Hermann–Mauguin and Hall symbols along with the full set of symmetry operations. *checkCIF* checks for the presence of those data along with their mutual consistency and their consistency with the reported crystal system. It currently ignores the inconsistency between Hermann–Mauguin and Hall symbols in cases of uncommon non-standard settings. Recent refinement programs such as *SHELXL2018* (Sheldrick, 2015b) and later include proper Hall symbols in the CIF for standard space-group settings. *PLATON/checkCIF* (Spek, 2003) will propose in many cases a Hall symbol consistent with the set of supplied symmetry operations to be inserted in the CIF. In other cases, the Hall (1981) paper will give guidance to create a proper Hall symbol. The current *checkCIF* tool cannot handle symmetries of non-three-dimensional structures such as those for incommensurate structures.

### 3. Structure solution

The path from a preliminary structure to a refined structure is full of pitfalls for the unwary. Validation can be helpful in that

process and should not be postponed to the publication stage where it takes much more effort to address problems with a structure, in particular when a new data collection will be needed. Either the IUCr *checkCIF* server or a locally installed version of the program *PLATON/checkCIF* can be used for that. *PLATON* also includes additional utilities to investigate ALERTed issues such as hydrogen-atom location, twinning and ADDSYM reported issues.

#### 3.1. Connected set

Preliminary structure solution does not always result in a logically ordered, sensibly labelled and connected set of atoms in the asymmetric unit. That is often not essential at the refinement stage but is certainly a requirement for a meaningful and professional report. It is good practice to make sure that all species in the asymmetric unit have their center-of-gravity within the unit-cell bounds. Exceptions include small species such as water molecules and counter-ions. Those are best located near their interaction points with the main species. This limits the number of symmetry operations needed in hydrogen-bond tables and offers a cleaner presentation.

Symmetry operations on coordinates are generally represented by a code of the type **s\_uvw**, where **s** refers to the sequence number of the symmetry operation in the previously defined list of symmetry operators in the CIF and **u**, **v** and **w** for unit-cell translations. The symmetry code for the primary species is represented by 1\_555 and that for a species moved by two units in the *a*-axis direction and one unit in the  $-c$ -axis direction by 1\_754. This way of representation of symmetry codes has its origin in a similar translation encoding used in the molecular graphics program *ORTEP* (Johnson, 1976) with its origin in the 1960s. The main drawback of this type of encoding is that only translations in the range  $-5$  to  $4$  can be represented with this code. This can be a problem with polymeric structures and long molecules that stretch over multiple unit cells. A proper use of the location of a species within the unit-cell bounds, close to the origin, might be essential in those cases. Care should be taken when, for example, hydrogen-bond table information is copied from external programs such as *PLATON*. The symmetry codes and sequence must be identical to those defined in the CIF.

#### 3.2. Missed symmetry

The applicable symmetry of a crystal structure is not always obvious. It depends on issues such as the radiation used in the experiment. Neutrons may see a different symmetry than X-rays. The former may ‘see’ magnetic spin-structure symmetry where the latter just indicate electron-density symmetry. The majority of X-ray structure reports are based on the space- and time-averaged electron density with a three-dimensional translation lattice, corresponding to sampling of the diffraction data at discrete points in diffraction (reciprocal) space. All diffraction outside those discrete points with possible information on order/disorder, superlattices and incommensurate behavior is then ignored.

All three-dimensional structures can be described in space group  $P1$ . Some currently used structure-solution programs such as *SHELXT* do just that and search for higher symmetry at a later stage. In practice, symmetry relations, compatible with the translation lattice, can be found between atomic positions within a certain tolerance. Collectively, they form the applicable space group. Alternatively, or in support, similar relations can be found in reciprocal space (as implemented in *SIR* and *SHELXT*).

*checkCIF* uses the *PLATON/ADDSYM* routine, an extended implementation of the *MISSYM* algorithm (Le Page, 1987, 1988) to check the reported space-group symmetry or to propose a possible higher-symmetry description within default tolerance settings. Those defaults are set at values needed to catch both higher proposed symmetry with poor data sets or interesting pseudo-symmetry. For that reason, hydrogen atoms and disordered atoms are ignored in the *ADDSYM* analysis. Also, the difference in atom types is ignored in order to catch atom-type mis-assignments (carbon for nitrogen, *etc.*). In addition, a certain percentage of atoms is allowed not to conform with the proposed higher symmetry. For that reason, *checkCIF* space-group ALERTS should always be checked, *e.g.* with the *PLATON/ADDSYM* tool using different tolerance settings. Examples of missed symmetry can be found in Marsh & Spek (2001).

The set of coordinates tested by *ADDSYM* for higher symmetry necessarily represents only part of the information present in the experimental reflection data. Cases of pseudo-symmetry need detailed inspection of the reflection data, including significant intensity values for supposedly systematic extinctions. Pseudo-symmetry easily comes together with twinning. In those cases, the purely coordinate-based *ADDSYM* algorithm needs to be supplemented by or alternatively tested with the reflection data. Good examples for that are described in detail by Clegg (2019). Eventually, refinement based on the proposed higher-symmetry options should provide the final verdict. Sometimes, even the diffraction data do not carry the necessary information to make a final decision between *e.g.*  $P1$  and  $P\bar{1}$  in the absence of knowledge of enantiopurity of the crystal.

### 3.3. Twinning detection

*checkCIF* includes a test for twinning. Both missed and possibly already applied but unreported twinning is suggested, along with an estimated twinning fraction and estimated  $R$ -value drop when applied. Reported cases of twinning can be non-merohedral, pseudo-merohedral and merohedral. The first two should already be indicated during the image-processing and data-reduction stage as additional or extended diffraction spots. Merohedral twinning might only show up during the structure solution and refinement stage with unexpectedly high  $R$  values or apparent systematic absences not consistent with any known space group.

The routine used in *checkCIF*, *TwinRotMat*, is based on the observation that twinning results in a significant number of reflections with  $I_{\text{obs}} \gg I_{\text{calc}}$  for the untwinned model refine-

ment. Those reflections are overlapped by a strong reflection with (approximately) the same diffraction  $\theta$  angle from a rotation-related twin-component lattice. Analysis of multiple such cases may lead to proposed candidates for the twinning operation at hand. Twinning detection with *TwinRotMat* may obviously be hampered when applied to reflection files where all strongly deviating reflections have already been removed from the refinement.

The *TwinRotMat* tool in a local implementation of the program *PLATON* can be used for more detailed analysis of the suggested twinning operations. Optionally, an HKLF5-style reflection file can be created for handling non-merohedral twinning with the *SHELXL* refinement program. Unfortunately, for the creation of such a file, empirical criteria have to be applied concerning overlapping reflections. In many cases, it is therefore often better to create HKLF5 files at the image-integration stage.

## 4. Refinement

The experimental electron density in a crystal structure is approximated by a set of structure model parameters. The most commonly used form is the atom-in-molecule (AIM) model, assuming spherical atomic scattering factors. The parameter set includes the positional parameters of the atoms, the anisotropic displacement parameters on the non-hydrogen atoms and isotropic displacement parameters for the hydrogen atoms. Some additional parameters are experiment-related such as the scaling factor of the observed diffraction data and correction factors for twinning and extinction. More involved models with non-spherical scattering factors, anharmonic displacement parameters or incommensurate structure models will not be considered here.

### 4.1. Constraints and restraints

Most of the model parameters can be refined. An exception involves space groups with a floating origin such as  $P1$  and  $P2_1$ . That situation is usually handled by either fixing one or more of the positional parameters of one of the atoms, or better, limiting the shift of the center-of-gravity of the atoms or similar. There are also symmetry restrictions on refineable coordinates and anisotropic displacement parameters of atoms lying on special positions. Most current structure-refinement programs handle those issues automatically.

Additional parameter restrictions may be needed, subject to the resolution, the reflection data-to-parameter ratio or in the case of disorder in the structure. Restrictions come in two varieties (Müller *et al.*, 2006). Constraints fix parameters to a fixed value or to the value of another (refineable) parameter. Restraints (sometimes confusingly called 'soft constraints') are 'springs' of variable strength to draw parameters to desired parameter values, subject to the counter force from the experimental data.

Common restrictions include the fixing of carbon-bonded hydrogen atoms on calculated positions and riding them on their carrier atom in the refinement. This will give a better

data-to-parameter ratio and avoids hydrogen atoms that refine to unrealistic distances due to noisy diffraction data. The positions of hydrogen atoms bonded to non-carbon atoms are often less predictable because their location depends on the presence of available acceptors in their environment. Their positions should be confirmed in a difference-density map (*vide infra*) and refined whenever possible as proof of their validity (Fábry, 2018).

Disorder resulting in overlapping atomic sites will require the application of constraints and/or restraints in order to keep the refinement stable and the model realistic.

The application of satisfactory constraints and restraints to model part of a structure may result in a more reliable model of the ordered part of interest of a structure. This applies in particular for the refinement of target molecules along with only weakly interacting disordered solvents. However, when the interest lies in the disordered part of a structure, there is always the question about what additional information is obtained beyond what is already known or included explicitly; see also Immirzi (2009) and Spek (2018).

Some refinement programs include parameter value shift-limiting constraints or restraints. *SHELXL* includes the *DAMP* instruction to control extreme parameter shifts of refined parameters that are poorly defined by the reflection data. This tool might be useful in the preliminary stages of the structure refinement but should be removed from the final refinement cycles and replaced by fixing still troublesome parameters to reasonable values. A strong damping factor may erroneously give the illusion of a converged structure with unreasonably low s.u.'s on the refined parameters.

All applied constraints and/or restraints should be reported in the experimental section and explained in the discussion sections of a paper.

#### 4.2. SQUEEZE

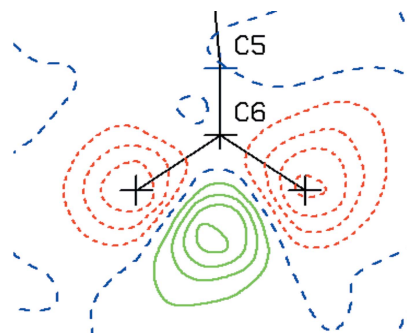
An alternative way to address the disordered-solvent problem is the use of the SQUEEZE algorithm. With that technique the solvent-accessible volume in a structure is identified and used to calculate the contribution of the scattering material found in that volume to the total calculated structure factor (Spek, 2015). The exact content of the disordered solvent volume is not determined explicitly with this method, but can sometimes be guessed from the reported total electron count of the integrated density in that volume, along with information about the solvents used during the synthesis and crystallization. The quality of the reported electron count is dependent on the quality, resolution and completeness of the diffraction data and the way the ordered part of the structure is modelled satisfactorily. Improper handling of low-angle data or omitting those data may seriously hamper the usefulness of the electron count. Disordered solvents contribute in particular to low-angle data. The missing data are estimated, but that is only partly satisfactorily. Inclusion of reflections that are hampered by the beamstop can pose a serious problem.

The disordered solvent contribution to the calculated structure factors as obtained with the current implementation of the SQUEEZE algorithm does not include resonant scattering contributions for *e.g.* disordered  $\text{CHCl}_3$ , making the absolute-structure determination of a light-atom main molecule of interest less reliable. A workaround was reported by Cooper *et al.* (2017).

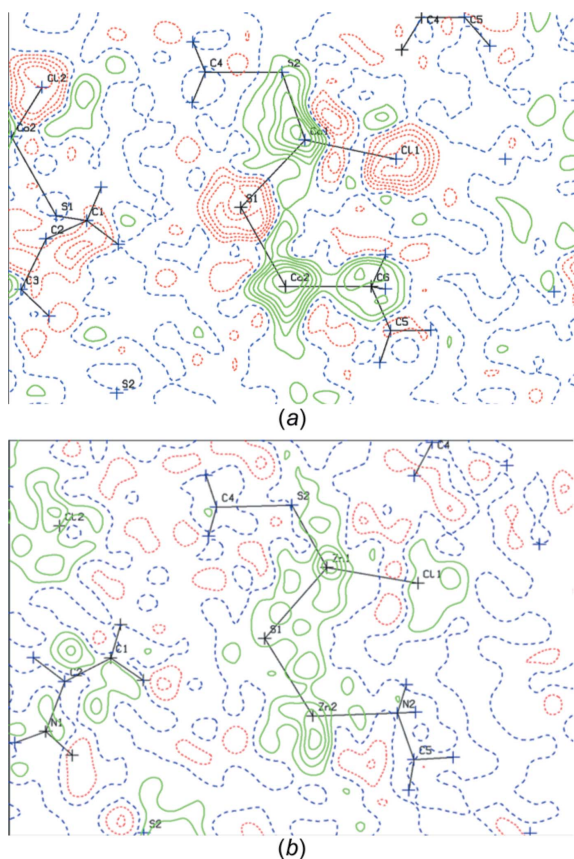
The SQUEEZE tool is implemented in *PLATON*. Relevant details of the SQUEEZE treatment are included at the end of the *name.cif* embedded *name.fab* file. An alternative, and slightly different, implementation of the SQUEEZE concept is available in the *OLEX2* software (Dolomanov *et al.*, 2009).

#### 4.3. Difference density maps

An  $F_{\text{calc}}$ -phased  $|F_{\text{obs}}| - F_{\text{calc}}$  difference-density map offers an extremely useful tool to detect problems with a structure determination. Such a map should be close to featureless with similar positive and negative density excursions of less than  $\sim \pm 0.5 \text{ e } \text{Å}^{-3}$ . It might show regions of positive density about 1 Å from light atoms such as carbon, nitrogen and oxygen, generally an indication for a missed hydrogen-atom site. Negative densities on a calculated hydrogen-atom site may indicate misplaced hydrogen atoms, *i.e.* a location where there is no density to be modelled. A common problem are the hydrogen-atom positions on a C–CH<sub>3</sub> fragment that often need to be rotated by 60° around the C–C bond from their current calculated positions. The difference map section through the three hydrogen-atom sites will then show six alternating negative and positive density minima and maxima. A wrong hybridization instruction for the calculation of the hydrogen-atom positions may show up as shown in Fig. 1. Other examples are ring substituents where C–H and N sites have to be interchanged and hydrogen atoms on N atoms that are calculated on planar  $sp^2$  locations rather than with tetrahedral  $sp^3$  geometry. Positive or negative densities on non-hydrogen-atom sites may indicate wrong atom-type assignments (Fig. 2). Validation ALERTS will generally point to those problems. Thoughtful examination of the model (and possible consultation with synthetic chemists) will help to resolve them.



**Figure 1**  
Example of hydrogen atoms wrongly calculated in  $sp^2$  rather than  $sp^3$  geometry positions. Wrongly calculated hydrogen-atom positions show up as negative (red) densities in the difference density map and correct places as positive (green) densities. Contours are drawn with  $0.1 \text{ e } \text{Å}^{-3}$  intervals.



**Figure 2**  
 (a) Difference density map for a structure with wrong atom-type assignments. (b) Difference map with correct atom-type assignments. Contours are drawn with  $0.1 \text{ e } \text{Å}^{-3}$  intervals.

Other issues in the difference-density map may easily escape attention or be explained away as being due to poor data when one just relies on the maximum and minimum residual density excursions reported by the refinement software. Some programs do not report difference-map density on atom sites. Contoured two- and three-dimensional difference-density maps offer a more detailed perspective. Significant approximately spherical density in a difference-density map cannot be ignored. Such density is usually interpretable as artificial (due to wrong assumed symmetry) or real structural disorder, substitutional disorder (e.g. a mixture of Cl and  $\text{CH}_3$  substituents) or missed twinning. Disordered solvent sites, in particular when occupied by a mixture of solvents or solvents (incommensurately ordered) in infinite channels (Fig. 3), may be missed when one relies on peak-search algorithms that assume ellipsoidal density shapes, e.g. density ridges are easily missed. Algorithms, such as the one used in *checkCIF*, are available to detect those regions in the structure as solvent-accessible voids. The SQUEEZE algorithm (van der Sluis & Spek, 1990; Spek, 2015) can be used to both investigate the content of those voids and, in the case where significant density is found, used to take blurred positive density into account.

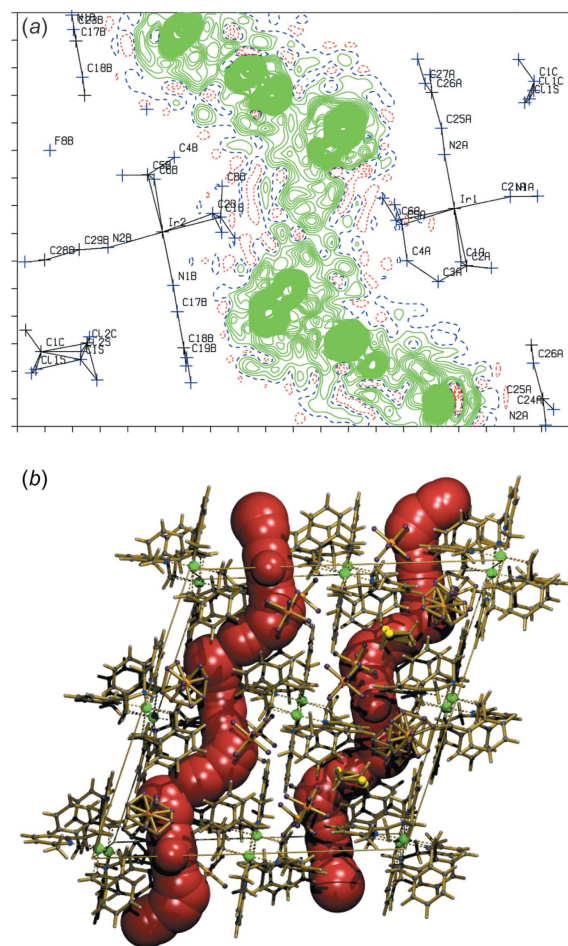
A special type of residual difference density concerns density near heavy atoms, usually within  $\sim 1 \text{ Å}$ . These are

often artifacts related to improper correction for absorption and similar systematic effects. They should not be confused with or explained as arising from positive and negative ripples around heavy atoms, related to resolution truncation, as might be seen in  $F_{\text{obs}}$  maps (Fig. 4a) but essentially absent in the associated difference-density map (Fig. 4b). Absorption artifacts usually show up as a symmetrical pattern of positive and negative peaks around the heavy atoms, correlated with the external shape of the crystal.

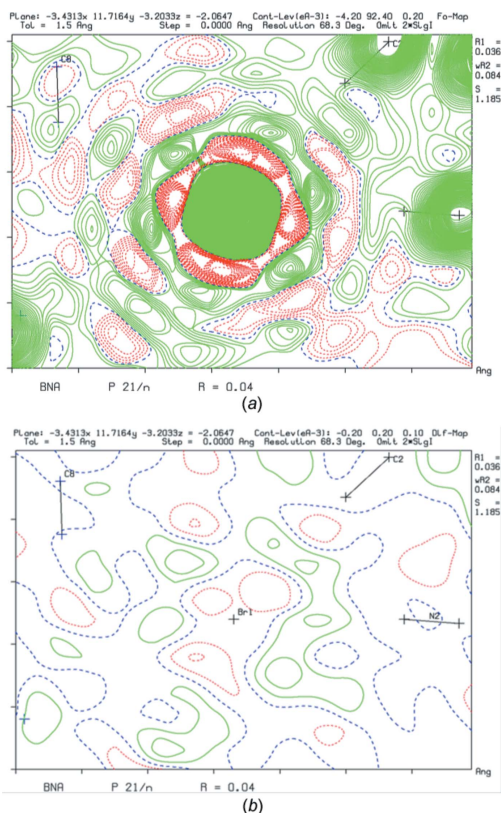
Structure reports for organic compounds based on good data and refined with the AIM model usually have their largest residual electron density on bonds (Fig. 5). Poor data result in noisy difference maps.

#### 4.4. Convergence quality criteria

A low  $R$  value is frequently used as a criterion for a good structure. Unfortunately, that goal is often achieved in structure reports by limiting the resolution of the data used in the refinement or by removing outliers without proper justification. Other parameters such as the data-to-parameter ratio,  $wR_2$  and  $S$  values and weighting parameter values have to be considered as well. A refinement program such as *SHELXL*



**Figure 3**  
 (a) Example of a structure with infinite channels filled with unresolved residual density and (b) display of the infinite channel, artificially filled with spheres.



**Figure 4**  
(a) Truncation ripples in an  $F_{\text{Obs}}$  map and (b) an essentially clean associated difference-density map.

optimizes two parameter values in the reflection-weighting expression to achieve an  $S$  value close to 1.0. Failure to come close to that value might be indicative of unresolved issues (in the data or the model). In addition large values, in particular for the second parameter, in the *SHELXL* weighting expression may convey a similar message. A non-Gaussian error distribution of the observed and calculated intensities may be visualized with a normal probability plot (included in the *name.ckf* listing). The central part of the plot should be straight and plot not too S-shaped, indicating many outliers.

An even more detailed analysis can be obtained with the analysis-of-variance where the fit of the model with the data is analysed as a function of variables such as resolution and  $F_{\text{calc}}/F_{\text{calc}}(\text{max})$ . Significant deviations of the scaling  $K$  and GooF values from 1.0 may indicate improper integration issues and or structure-model errors. Details can be found in the *name.ckf* file for the associated ALERTS.

A sufficient reflection data ( $N$ ) to refined parameter ( $p$ ) ratio is expected. The number of reflections in this ratio is the number of Laue group averaged reflections. The *checkCIF* test criteria for the  $N:p$  ratio are set on values that can be met with good data for both centrosymmetric and non-centrosymmetric structures.

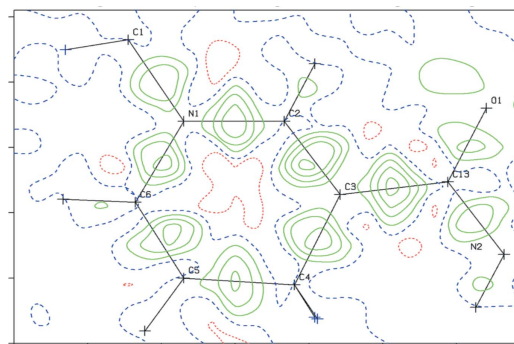
#### 4.5. Absolute structure

Non-centrosymmetric structures are not always enantiopure. A twinning model was introduced by Flack (1983) to

take this into account with a refined Flack  $x$  parameter describing the mixture fraction. As many Bijvoet pairs as possible will be needed for a reliable refinement of that parameter. Usually, as a first step to establish a proper value for the Flack parameter, the non-centrosymmetric structure is refined without this refineable parameter included. Subsequently a post-refinement analysis is performed with the observed and calculated Bijvoet differences to estimate the value of the Flack parameter with associated estimated standard uncertainty. Various estimates for that are in use such as the Parsons *et al.* (2013) and Hooft *et al.* (2008, 2009, 2010) parameter values. Based on this analysis, subsequent refinement will be in order, including the Flack  $x$  parameter, when the estimated values deviate significantly from zero, including proper inversion of the coordinates when greater than 0.5. Criteria for determining valid absolute-structure assignments are give by Flack & Bernardinelli (2000).

#### 5. Validation

Full validation with the program *PLATON* is done on both the *name.cif* file, containing the experimental and refinement data, and the *name.fcf* file, containing the final observed, calculated and  $\sigma$  values of the reflections used in the refinement. The *name.cif* file includes, where possible, both the detailed refinement instructions and the unmerged reflection data. A *name.cif* file created with a recent version of *SHELXL* includes all data necessary to recreate the final *name.fcf* file, thus eliminating the need to explicitly provide that file. The *name.fcf* file is used for a detailed analysis of the refinement. The result of that analysis is written to the *name.ckf* file. The ALERTS derived from the analysis of the *name.cif* and *name.fcf* data, based on criteria detailed in a file named *check.def*, are written to the *name.chk* file. By default, a built-in version of that file is used. *name.cif* files created with recent *SHELXL* versions include checksums meant to ensure the credibility of the embedded *name.res*, *name.hkl* and, in the case of a SQUEEZED structure, *name.fab* data as tested with the *SHELX shredcif* utility. Those embedded data should not be edited in order to avoid the IUCr/*checkCIF* requirement to supply the final *name.fcf* file.



**Figure 5**  
Difference-density map with residual density maxima on bonds for a good structure refined with spherical atomic scattering factors (AIM model). Contours are drawn with  $0.1 \text{ e}^{-3}$  intervals.



Validation ALERTS should not be ignored or worked around. Level A ALERTS that cannot be resolved should be accompanied by a VRF record with an explanation why that ALERT can be ignored in the case at hand. However, it is even better to refer to some explanatory text in the report.

The CIF file to be validated may be a concatenation of one or more structural CIFs, each preceded by the data-block identifier *data\_ID*, where *ID* should be distinct (e.g. I, II, etc.). An additional CIF-structured block *data\_global* with non-structural information (e.g. journal information and publication text for *Acta Cryst. C* and *E*) can be included (preferably as the first block). Similarly, the associated FCF blocks (when not recreatable from the CIF file data), with identical *data\_ID*, should be concatenated into one *name.fcf* file for *Acta Cryst.* submissions.

### 5.1. Formula

The expressions for the *sum formula* and *moiety formula* are well defined in the case of isolated molecular species in general positions with associated *Z* and molecular weight values, consistent with the provided chemical scheme. In cases where (some) of the atoms are located on special positions or with polymeric structures, things are less well defined. *checkCIF* offers a suggestion that may be overruled where more appropriate. The main requirement is consistency with associated data.

Situations where part of the structure remains uninterpreted in terms of explicit atomic parameters need special treatment. *PLATON/checkCIF* accepts the specification of information about the content of the ‘solvent’ voids to be specified between square brackets [], for example ‘[+solvent]’ when the nature of the solvent (mixture) is unclear or ‘[+toluene]’. The contribution within the square brackets is not included in the calculated molecular weight and density as reported in the validation report.

### 5.2. Hirshfeld test

The Hirshfeld rigid-bond test (Hirshfeld, 1976) was originally not introduced as a validation test but turned out to be also useful as a validation tool. Organic molecules were the focus of its original use. The notion is that the components of the displacement parameters of the two atoms in a covalent bond have approximately equal opposite values along that bond. This notion was extended to the generally weaker coordination bonds to a metal and transformed into a validation test. In this test, the Hirshfeld difference divided by its associated s.u. is compared with empirical threshold values. Large deviations may point to erroneously mis-assigned atom types. This test turned out to be instrumental in discovering published structures with deliberately changed atom types.

### 5.3. Resonant scattering factors

Refinement programs such as *SHELXL* include resonant scattering factor values ( $f'$  and  $f''$ ) for most atom types and common X-ray sources (Cu  $K\alpha$ , Ag  $K\alpha$ , Mo  $K\alpha$  etc.). Values for other wavelengths (such as those for a synchrotron

data collection) will need to be user supplied. The  $f'$  and  $f''$  values found in the *name.cif* file are checked against those interpolated from internal tables in *PLATON*. *PLATON/checkCIF* uses data published by Brennan & Cowan (1992) for that purpose.

### 5.4. Occupancy and multiplicity

The *occupancy* factor as documented in the CIF is sometimes confused with the *site-occupancy factor* (SOF) as used in refinement programs such as *SHELXL* that include a weighting factor for atoms in special positions. For example, a SOF value of 0.5 for a fully occupied atom located at an inversion center should be reported with an *occupancy* of 1.0 in the CIF.

There are two, sometimes confusing, ways to specify the atom-site multiplicity associated with the occupation in the CIF: *\_atom\_site\_site\_symmetry\_multiplicity* (*ssm*) and *\_atom\_site\_site\_symmetry\_order* (*sso*). Their relation is  $sso = nsym/ssm$ , where *nsym* = number of symmetry operations of the space group. *SHELXL* reports *sso* values in the CIF rather than the *ssm* values given in the *International Tables for Crystallography, Volume A*. SOF and occupancy are related by  $SOF = occupancy/sso$ .

### 5.5. Short intermolecular contacts

Distances between atoms in the different molecular species in a crystal structure are expected to be approximately equal to or greater than the sum of the van der Waals radii of the two atoms involved. *checkCIF* will flag cases where this distance is shorter.

Short H···H contacts may be related to hydrogen atoms placed in incorrect calculated positions. This might be due to a wrong hybridization assignment or putting the CH<sub>3</sub> moiety in the wrong staggered or eclipsed orientation or not carrying out a rigid-rotating group refinement for the methyl group. Other not uncommon cases are the mis-assignment of a non-hydrogen atom carrying a hetero atom (e.g. S or N) and a C–H in a hetero ring moiety. Exchanging the assignments should solve the issue.

Short contacts involving disordered atom sites should be inspected in detail to ensure that they are physically possible and are not artefacts of the disorder model.

Short O···O or O···N contacts generally point to a missing hydrogen atom in this contact. This usually involves a smeared H atom in the middle of a short acid hydrogen bond, which may or may not be symmetrical (Fábry, 2018).

A valid exception concerns short halogen···halogen interactions (see e.g. Mukherjee & Desiraju, 2014). They are flagged as interesting information on the structure.

### 5.6. FCF validation

Details for reflection-related ALERTS in the *name.chk* file can be found in the CKF file (*name.ckf*).  $|I_{obs}| - |I_{calc}|$  outliers are listed along with information on missing reflections. The listing of the percentage of observed data with  $I_{obs} > 2\sigma(I_{obs})$  as a function of  $\sin(\theta)/\lambda$  can be useful to optionally determine

a noise cut-off level for reflections to be excluded from the refinement. The full *SHELXL*-style analysis-of-variance table is included for details of the associated ALERTS concerning deviations from expected values. Finally there are details listed for the check for twinning, absolute structure and a detailed analysis of the peaks in the final difference map.

### 5.7. The unmerged reflection file

An archived CIF is expected to include the unmerged reflection data on which a study is based. That file can be used to investigate unresolved issues concerning the correct space-group assignment. A merged reflection file has lost information on the quality and data consistency of systematic absences and indications for the lack of a centre of symmetry and absolute structure.

Unfortunately, the unmerged reflection file is already several steps away from the primary experimental data: the diffraction images. Data reduction (integration, correction for absorption *etc.*) is usually poorly documented and reported, being largely a black-box process using proprietary diffractometer software packages. Information about scattering effects outside the integrated diffraction spots and their shape is easily lost but often relevant to understand the origin of a poor structure refinement. The FAIR (findable, accessible, interoperable and reusable; Helliwell, 2019) initiative addresses this issue.

A structure report can also be based on massaged or artificial data, as exemplified by data archived with the CSD. Those cases need not always to be fraudulent but rather related to poor chemical knowledge and inexperience with the available options in software packages. Software packages may include options to create artificial reflection data meant to be used for test purposes only.

*checkCIF* includes a number of checks that should flag issues with the data that need to be investigated in more detail. An example is the ALERT that reports that there is no residual density on bonds in the difference map of a ‘too good-looking’ structure (assuming refinement with spherical atomic scattering factors). A  $1/\sigma(I_{\text{obs}})$  versus  $^{10}\log(I_{\text{obs}})$  plot (Diederichs, 2010) or a  $\sigma(I_{\text{obs}})$  versus  $\sqrt{I_{\text{obs}}}$  plot may be useful to check for and illustrate the use of fabricated ‘experimental’ reflection data.

### 6. A PLATON/checkCIF validation example

Fig. 6 illustrates a *PLATON/checkCIF* validation report, as obtained with a local *checkCIF* implementation, with a number of issues to be addressed. Fig. 6a offers a summary of some relevant information such as experimental data and refinement results. The residual-density ranges, as reported in the *name.cif* file, are compared with those calculated by *PLATON* from the refined parameters in the *name.cif* file and the merged reflection data in the *name.fcf* file. Those values are expected to be closely identical within rounding errors. Three sets of *R*, *wR*<sub>2</sub> and *S* values are calculated and compared. The first set is again calculated from the *name.cif*

and *name.fcf* data. The second set is based on the observed and calculated reflection data in the *name.fcf* file, along with the weighting parameters taken from the *name.cif* file. The third set are the corresponding values as reported in the *name.cif* file. Where relevant, three types of ‘Flack parameter’ values are reported, being the Flack *x* value reported in the *name.cif* file, a *PLATON*-based estimate of the Parsons *z* value and a *PLATON*-based estimate of the Hooft *y* parameter value.

Fig. 6b lists the ALERTS for this example. The A and B level ALERTS all refer to missing data and should be easy to address. So are the first three C level ALERTS. ALERT #977 indicates that atom H9 is located in a location not supported by the data. The #007 ALERT implies that the H-atom positions are not necessarily to be trusted for detailed hydrogen-bond analysis. ALERT #012 indicates that the embedded *name.res* file is corrupt or edited. A non-zero value of ALERT #978 can be taken as a measure of the quality of the difference density map of a refined structure (assuming that spherical atomic scattering factors were used).

```

=====
                    VALIDATION REPORT FOR CURRENT CIF
=====
# PLATON/CHECK (241119) versus check.def version 191115, Entry: ambf
# Data: ambf.cif - Type: CIF16                      Band Preclon  C-C = 0.0020 Å
# Refl: ambf.fcf - Type: LIST4                      Temp = 293 K
# Audit: SHELXL-2016/6
# Refin: SHELXL-2016/6 (SHELDRICK, 2016)
# X-ray MoKα R(Lnt) = 0.012, wR2/R(Lnt) = 9.5, Nref/Npar = 9.0
# Cell 7.6421(10) 7.8306(8) 11.049(3) 90 90 661.2(12)
# Wavelength 0.71073 Å Volume Reported 661.21(19) Calculated 661.2(12)
# SpaceGroup from Symmetry P 21 21 21 Hall: P 2ac 2ab orthorhombic
# Reported ? or F000 = 352.28(Calc.) orthorhombic
# MolecFormula C4 H5 O6, H4 N
# Reported ?
# SumFormula C4 H9 N O6
# Reported C16 H36 N4 O24
# Mr 167.12(Calc.), 668.49[Rep] Volume/NonHatoms = 15.0
# Dx, gcm-3 = 1.679(Calc.), 1.679[Rep]
# Z 4(Calc.), 4[Rep]
# Nu (mm-1) = 0.161(Calc.), 0.161[Rep] Xtol Size = 0.400x0.300x0.200 mm
# F000 = 352.0(Calc.), 352.0[Rep] or F000 = 352.28(Calc.)
# Measured HKL: Reported 5240, Embedded 5264
# Reported Hmax= 9, Kmax= 10, Lmax= 14, Nref= 1521 Th(max)= 27.495
# Obs Ln FCF Hmax= 9, Kmax= 10, Lmax= 14, Nref= 1521(C 903), Th(max)= 27.496
# Calculated Hmax= 9, Kmax= 10, Lmax= 14, Nref= 1526( 903), Ratio=1.68/1.00
# Reported Rho(mLn) = -0.37, Rho(max) = 0.28 e/Ång*3 (From CIF)
# Calculated Rho(mLn) = -0.36, Rho(max) = 0.27 e/Ång*3 (From CIF+FCF data)
# w=1/(Lsigma**2*(Fmax**2)+(0.1000P)**2), P=(Fmax**2+Fc**2)/3
# R = 0.0308( 1501), wR2 = 0.1112( 1521), S = 1.039 (From CIF+FCF data)
# R = 0.0308( 1501), wR2 = 0.1111( 1521), S = 1.038 (From FCF data only)
# R = 0.0308( 1501), wR2 = 0.1112( 1521), S = 1.039, Npar= 100, Flack = 0.08(14)
# Number BL vsot, Polre = 618 (100%), 601 Selected for: Parsons -0.07(16)
# P2(tl) 1.000, P3(tl) 1.000, P3(tw) 0.000, Student-T Nu = 99, Hoaft -0.07(12)
=====
For Documentation: http://http://www.platonsoft.nl/CIF-VALIDATION.pdf
=====
    
```

(a)

```

>>> The Following Improvement and Query ALERTS were generated - (Acta-Mode) <<<
=====
Format: alert-number_ALERT_alert-type_alert-level_text
=====
067_ALERT_1_A MaxLmng Dimension Less Min. Xtol Dimension ..... Please Check
183_ALERT_1_A MlssLng _cell_measurement_reflns_used Value ..... Please Do I
184_ALERT_1_A MlssLng _cell_measurement_theta_min Value ..... Please Do I
185_ALERT_1_A MlssLng _cell_measurement_theta_max Value ..... Please Do I
693_ALERT_1_A MlssLng _exptl_crystal_description Value ..... Please Do I
035_ALERT_1_B chemLcal_absolute_configuration Info Not Given Please Do I
=====
048_ALERT_1_C MolecFormula Not Given (or Incomplete) ..... Please Check
052_ALERT_1_C Info on Absorption Correction Method Not Given Please Do I
126_ALERT_4_C No symmetry space group name_Hall Given ..... Please Do I
911_ALERT_3_C MlssLng FCF Refl. Between Th(Ln) & STH/L = 0.600 6 Report
913_ALERT_3_C MlssLng # of Very Strong Reflections Ln FCF ... 5 Note
977_ALERT_2_C Check Negative Difference Density on H9 -0.36 eÅ-3
=====
007_ALERT_5_C Number of Unrefined Donor-H Atoms ..... 7 Report
012_ALERT_1_G N.O.K. _shxl_res_checksum Found Ln CIF ..... Please Check
045_ALERT_1_G Calculated and Reported Z Differ by a Factor ... 4.00 Check
199_ALERT_1_G Reported _cell_measurement_temperature ..... (K) 293 Check
200_ALERT_1_G Reported _diffrn_ambient_temperature ..... (K) 293 Check
791_ALERT_4_G Model has ChiralLty at C3 (Chiral SPGR) A Verify
978_ALERT_2_G Number C-C Bonds wLth PosLtlve ResLdual DensLty. 2 Info
=====
ALERT_Level and ALERT_Type Summary
=====
5 ALERT_Level_A = Most Lkely a Serious Problem - Resolve an Expln
1 ALERT_Level_B = A PotentLly Serious Problem - Consider CarefLly
6 ALERT_Level_C = Check, Ensure Lt Is Not caused by an OmlsLsn or OvsrLgt
8 ALERT_Level_G = General Info/Check that Lt Is not Somethng Unexpectd
=====
12 ALERT_Type_1 CIF ConstructLion/Syntax Error, Inconsistent or MlssLng Data.
2 ALERT_Type_2 Indicator that the Structure Model may be Wrong or DefLct.
    
```

(b)

Figure 6 *PLATON/checkCIF* report: (a) structural data summary and (b) ALERT listing.

## 7. Concluding remarks

*checkCIF* offers a machine-generated report in the form of ALERTS. It does not give a 'good structure' or 'bad structure' (or 'publish' or 'reject') verdict. That is left to human beings, either referees or users of the reported results. The value of a structure report lies in the added reliable scientific information. A well-documented poor-quality structure may be fine if supporting part of the reported chemistry in a chemical journal, but less suitable for publication in a crystallographic journal such as *Acta Crystallographica* that aims to publish the best attainable quality of the reported structure, the latter in view of future use of the data in unrelated research such as non-AIM model refinement.

Not all reported issues with a structure report can be resolved with access to the reflection data. Some can be resolved only with access to the diffraction images. Work on the archiving of those data is underway (Kroon-Batenburg *et al.*, 2017).

Current IUCr/*checkCIF* validation procedures mainly cover single-crystal structure reports refined with the atoms-in-molecule (AIM) model. Refinements against powder diffraction data (e.g. *GSAS*; Toby & Von Dreele, 2013) and non-AIM model-based refinements [e.g. *IDEAL* (Lübben *et al.*, 2019) and *OLEX2/HARt* (Fugel *et al.*, 2018)] are currently only partly covered. The same applies to incommensurate structures (*JANA*; Petricek & Dusek, 2000) and structure determinations based on electron diffraction. Those will need the development of specialized validation procedures of the associated experimental, refinement and interpretation of the reported results.

## 8. *checkCIF* software accessibility

The IUCr CIF/FCF validation facility is available through <https://checkcif.iucr.org> along with related tools.

A Microsoft Windows executable for local use is available from <http://www.chem.gla.ac.uk/louis/software/platon>.

Alternatively, a local copy of the source code with implementation instructions for LINUX and MacOS implementation of *PLATON/checkCIF* is available from <http://www.platonsoft.nl/xraysoft>. That version depends on an interface to the X-Windows system. A special version of the software is available without that dependency. Both the full and special versions of *PLATON* allow for easy keyboard instructions to run *checkCIF*. The keyboard instruction *platon -u name.cif* will produce a full *checkCIF* report on the files *name.chk* and *name.ckf*. The alternative command *platon -U name.cif* will not include the short explanations of the ALERTed-for issues.

## Acknowledgements

Many of the currently implemented set of ALERTS and associated criteria are based on issues detected with CIFs submitted for publication in *Acta Cryst.* journals. I am also indebted to many who supplied ideas, comments and in particular CIFs with new issues to be addressed.

## References

- Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K. & Watkin, D. J. (2003). *J. Appl. Cryst.* **36**, 1487.
- Brennan, S. & Cowan, P. L. (1992). *Rev. Sci. Instrum.* **63**, 850–853.
- Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2003). *J. Appl. Cryst.* **36**, 1103.
- Clegg, W. (2019). *Acta Cryst.* **E75**, 1812–1819.
- Cooper, R. I., Flack, H. D. & Watkin, D. J. (2017). *Acta Cryst.* **C73**, 845–853.
- Diederichs, K. (2010). *Acta Cryst.* **D66**, 733–740.
- Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. (2009). *J. Appl. Cryst.* **42**, 339–341.
- Fábrý, J. (2018). *Acta Cryst.* **E74**, 1344–1357.
- Farrugia, L. J. (2012). *J. Appl. Cryst.* **45**, 849–854.
- Feast, G. C., Haestier, J., Page, L. W., Robertson, J., Thompson, A. L. & Watkin, D. J. (2009). *Acta Cryst.* **C65**, o635–o638.
- Flack, H. D. (1983). *Acta Cryst.* **A39**, 876–881.
- Flack, H. D. & Bernardinelli, G. (2000). *J. Appl. Cryst.* **33**, 1143–1148.
- Fugel, M., Jayatilaka, D., Hupf, E., Overgaard, J., Hathwar, V. R., Macchi, P., Turner, M. J., Howard, J. A. K., Dolomanov, O. V., Puschmann, H., Iversen, B. B., Bürgi, H.-B. & Grabowsky, S. (2018). *IUCrJ*, **5**, 32–44.
- Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst.* **B72**, 171–179.
- Hall, S. R. (1981). *Acta Cryst.* **A37**, 517–525.
- Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Cryst.* **A47**, 655–685.
- Helliwell, J. R. (2019). *Struct. Dyn.* **6**, 054306. doi: 10.1063/1.5124439. eCollection 2019 Sep.
- Hirshfeld, F. L. (1976). *Acta Cryst.* **A32**, 239–244.
- 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. (2009). *Acta Cryst.* **A65**, 319–321.
- Hooft, R. W. W., Straver, L. H. & Spek, A. L. (2010). *J. Appl. Cryst.* **43**, 665–668.
- Immirzi, A. (2009). *J. Appl. Cryst.* **42**, 362–364.
- Johnson, C. K. (1976). *ORTEP II*. Report ORNL -5138. Oak Ridge National Laboratory, Tennessee, USA.
- Kroon-Batenburg, L. M. J., Helliwell, J. R., McMahon, B. & Terwilliger, T. C. (2017). *IUCrJ*, **4**, 87–99.
- Le Page, Y. (1987). *J. Appl. Cryst.* **20**, 264–269.
- Le Page, Y. (1988). *J. Appl. Cryst.* **21**, 983–984.
- Lübben, J., Wandtke, C. M., Hübschle, C. B., Ruf, M., Sheldrick, G. M. & Dittrich, B. (2019). *Acta Cryst.* **A75**, 50–62.
- Marsh, R. E. & Spek, A. L. (2001). *Acta Cryst.* **B57**, 800–805.
- Mukherjee, A. & Desiraju, G. R. (2014). *IUCrJ*, **1**, 49–60.
- Müller, P., Herbst-Immer, R., Spek, A. L., Schneider, T. R. & Sawaya, M. R. (2006). *Crystal Structure Refinement*. Oxford University Press.
- Palatinus, L. & Chapuis, G. (2007). *J. Appl. Cryst.* **40**, 786–790.
- Parsons, S., Flack, H. D. & Wagner, T. (2013). *Acta Cryst.* **B69**, 249–259.
- Petricek, V. & Dusek, M. (2000). *JANA2000*. Institute of Physics, Czech Academy of Sciences, Prague, Czech Republic.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Sheldrick, G. M. (2015a). *Acta Cryst.* **A71**, 3–8.
- Sheldrick, G. M. (2015b). *Acta Cryst.* **C71**, 3–8.
- Sluis, P. van der & Spek, A. L. (1990). *Acta Cryst.* **A46**, 194–201.
- Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.
- Spek, A. L. (2009). *Acta Cryst.* **D65**, 148–155.
- Spek, A. L. (2015). *Acta Cryst.* **C71**, 9–18.
- Spek, A. L. (2018). *Inorg. Chim. Acta*, **470**, 232–237.
- Thompson, A. L. (2019). *Crystallogr. Rev.* **25**, 3–53.
- Toby, B. H. & Von Dreele, R. B. (2013). *J. Appl. Cryst.* **46**, 544–549.