Single-crystal structure validation with the program *PLATON*

A. L. Spek

Crystal and Structural Chemistry, Bijvoet Centre for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. Correspondence e-mail: a.l.spek@chem.uu.nl

The results of a single-crystal structure determination when in CIF format can now be validated routinely by automatic procedures. In this way, many errors in published papers can be avoided. The validation software generates a set of `ALERTS` detailing issues to be addressed by the experimenter, author, referee and publication journal. Validation was pioneered by the IUCr journal *Acta Crystallographica Section C* and is currently standard procedure for structures submitted for publication in all IUCr journals. The implementation of validation procedures by other journals is in progress. This paper describes the concepts of validation and the classes of checks that are carried out by the program *PLATON* as part of the IUCr *checkCIF* facility. *PLATON* validation can be run at any stage of the structure refinement, independent of the structure determination package used, and is recommended for use as a routine tool during or at least at the completion of every structure determination. Two examples are discussed where proper validation procedures could have avoided the publication of incorrect structures that had serious consequences for the chemistry involved.

1. Introduction

A single-crystal X-ray study has the unique potential to provide 'solid' knowledge about the three-dimensional structure of molecules and complexes in the crystalline state along with their intermolecular interactions. Much of our current knowledge concerning inorganic and metal-organic compounds is derived from single-crystal studies. Structure determinations that are carried out carefully will generally provide incontrovertible results, upon which subsequent research can be built. Unfortunately, for various reasons, not all structures that end up in the refereed literature and subsequently in databases appear to be correct, either being erroneous only in certain details or containing major errors which lead to the derivation of incorrect conclusions. An excellent paper by Harlow (1996) discusses many examples. The assignment of the correct space group to a structure is one of the most common problems, as has been demonstrated many times by Marsh and others (*e.g.* Marsh & Spek, 2001). Some of the more serious errors include incorrectly assigned atom types and missing or too many hydrogen atoms in a structure, obviously with serious implications for the chemistry involved. Often, papers include a detailed discussion on an interesting feature of a molecular structure that in hindsight turns out to be based on an artefact. Parkin (1993) addresses in detail the illusive 'bond-stretch isomerism' phenomenon.

Technical advances have turned the structure determination of many crystals of sufficient quality into a routine procedure in the hands of an experienced crystallographer. Much of the data collection, structure solution and refinement steps have been automated and often form part of the software package that comes with the data-collection instrument. Current CCD (charge-coupled device) detector-based diffractometer systems are capable of producing up to 1000 data sets per year. Computing power, being a serious bottleneck in the past, is no longer a limiting factor in this field. A modern PC can easily handle all the necessary calculations.

An important and certainly final part of a structure determination should be the validation of the structural results. Traditionally, this was the responsibility of a professional crystallographer and, when submitted for publication, the referees. The unfortunate current situation is that the number of structures that are submitted for publication each year is orders of magnitude larger than the number of experienced crystallographers and referees knowledgeable in crystallography. In addition, referees are often given only limited access to the supporting experimental data or are simply informed that all crystallographic details have been deposited. The Cambridge Structural Database (CSD; Allen, 2002) includes many entries with comment records detailing problems encountered by the data-entry staff during the processing of published structures.

A recent editorial (Eisenberg, 2002) in *Inorganic Chemistry* addresses the issue of properly 'reviewing crystallographic data'. The International Union of Crystallography (IUCr) identified the general problem a decade ago and created as a first step the computer-readable Crystallographic Information File (CIF) standard for reporting, exchanging and archiving

crystal structure data (Hall *et al.*, 1991). Most current software packages can now generate such files. Subsequently, a project (*checkCIF*) was initiated to validate crystal structure data supplied in CIF format. Currently, the electronic validation procedures are fully operational.

2. Validation

Single-crystal structure validation addresses three important issues that can be expressed in the form of the following questions:

Question 1: is the reported information complete?
Question 2: what is the quality of the analysis?
Question 3: is the structure correct?

The answer to question 1 should be easy since it essentially involves a checklist of items describing various details of the structure analysis. Many journals provide in their notes for authors such a checklist of details to be provided so that others can evaluate and repeat the study. The CIF standard offers an adequate mechanism to make this information available for subsequent electronic processing. The automatic analysis of the data present in the CIF for completeness and internal consistency is relatively trivial.

Judging the quality of a structure determination, question 2, is more difficult. It very much depends on factors such as crystal quality, data collection hardware, expertise of the investigator and software used. Nevertheless, various values can be calculated that can be compared with some generally accepted standards. Examples are resolution and completeness of the data set, bond precision (see below), convergence of the least-squares refinement and *R* factors. Referees can judge whether the quality of the analysis is sufficient to support the conclusions drawn from it and meet journal standards. A journal might choose to publish only those structures with the highest attainable quality, which can be higher than needed for the purpose of the study but relevant for the follow-up research that might build upon it.

Question 3 is probably the most important and difficult. This is not something to which a computer (program) can provide a complete answer. What a program can do is to report on any unusual feature it detects. It is then up to the investigator/author to consider the issue, take action if necessary, or otherwise comment convincingly on it. Subsequently, the referees must decide on the validity of the arguments offered.

Incorrect structures usually exhibit one or more of three symptoms: (i) unusual bond distances, (ii) unreasonable displacement parameters (ellipsoids) and (iii) impossible intra- and intermolecular contacts (see below). When unusual structures are reported, only a high-quality structure might give the level of confidence needed to proclaim that it indeed has unique features.

3. Validation tests

The IUCr has defined and documented a large number of validation tests (http://journals.iucr.org/services/cif/checking/autolist.html) that are carried out for all structures submitted for publication in IUCr journals. These tests address issues related to questions 1 and 2 detailed above and are readily available through the Web-based IUCr *checkCIF* facility (http://journals.iucr.org/services/cif/checking/check-form.html). The tests include checks for proper refinement and absorption correction procedures. The program *IUCRVAL* (Farrugia, 2000) also incorporates this set of tests and can be implemented locally.

The validation function in the program *PLATON* (Spek, 2002) addresses all three questions. The next section of this paper details the various classes of tests that are implemented in the program.

PLATON can also conduct F_o/F_c CIF validation. Issues addressed include resolution and completeness of the data set. In particular, missing low-order reflection data are reported. However, no details will be presented here.

4. Validation tests implemented in *PLATON*

The tests are given a three-digit number, *nxx*, and fall into ten groupings:

0xx: this range concerns data completeness, consistency and quality tests.

1xx: tests addressing unit-cell and space-group symmetry issues.

2xx: issues related to (an)isotropic displacement parameters.

3xx: tests and reports on intramolecular issues.

4xx: tests and reports on intermolecular issues.

5xx: coordination-related issues.

6xx: issues related to solvent-accessible voids.

7xx: problems with bonds and their associated standard uncertainties.

8xx: validation-software problems

9xx: problems with the reflection data.

In the following, a number of the issues addressed will be discussed in detail.

4.1. Missed symmetry

The assignment of the proper space group to a given structure is not always obvious at the beginning of a structure determination. Often, a preliminary structure can be obtained only in a space group with symmetry that is lower than the actual symmetry. Subsequent analysis should lead to a description in the correct space group. Unfortunately, the latter is not always achieved.

ADDSYM, an extended version of the *MISSYM* algorithm (Le Page, 1987, 1988) is used to search for possibly missed higher crystallographic symmetry in the reported structure. A tentative, more appropriate, space group is suggested. Interestingly, approximations to higher symmetry (pseudosymmetry) occur frequently. For that reason, missed symmetry ALERTS generally require as a follow-up a detailed analysis of the situation 'by hand', with access to the primary reflection data.

An attempt to refine a centrosymmetric structure in a non-centrosymmetric space group generally results in poor geometry due to the (near) singularity of the least-squares normal matrix. Chemically equivalent bonds may differ significantly and displacement parameters generally make little sense in such a case. Proper action includes, apart from leaving out half of the atoms in the model and the addition of an inversion centre, a shift of the structure to the proper origin. A number of reported cases of missed higher symmetry are due to a failure by the authors to apply such a shift, with the result that the subsequent refinement of the centrosymmetric model was reported 'unsuccessful'. For a corrected example of a structure originally published in $P1$ with two crystallographically independent molecules and an unusual *ORTEP* diagram, see Kahn *et al.* (2000a,b).

The *ADDSYM* algorithm allows a small percentage of atoms to fail the proposed higher symmetry. Although this feature will bring up a number of false ALERTS, it will often catch cases of missed symmetry in poorly refined structures or those with missing, miss-identified or disordered atoms.

4.2. Voids

Solvent-accessible voids (van der Sluis & Spek, 1990) are reported. Such voids might include disordered solvent that went undetected by peak-search algorithms. A common reason might be that the disorder results in density ridges or faint plateaus rather than isolated peaks. Except for some framework structures, crystal structures generally collapse when they have lost solvent molecules of crystallization.

Voids are frequently located at or along symmetry elements. Solvent molecules on those sites are generally highly disordered or fill one-dimensional channels along three-, four- or sixfold axes.

Void ALERTS in combination with ALERTS on short intermolecular contacts may point to molecules that are misplaced with respect to the symmetry elements.

4.3. Displacement ellipsoid tests

A displacement ellipsoid plot (*ORTEP*II; Johnson, 1976) is an excellent validation tool but not suitable as an automatic tool. Fortunately, numerical analogues for visual inspection by an expert are available. The Hirshfeld rigid-bond test (Hirshfeld, 1976) can be indicative of many problems with a structural model. The central idea is that the components of the anisotropic displacement parameters along the bond for two bonded atoms should have approximately the same value. This will generally not be the case when incorrect atom types are assigned to density peaks. Carbon atoms might be nitrogen atoms or oxygen, *etc.*

Elongated ellipsoids are, in general, indicative of unresolved disorder. An attempt should be made to develop a proper model to represent the disorder.

Many systematic errors (including absorption and wrong wavelength) find their way into the displacement parameters, often giving rise to physically impossible non-positive-definite values for the main-axis values.

4.4. Bonds and angles

The values of bonds and angles are checked to determine whether their values fall within expected ranges. Single, double or triple bond types are assigned from the deduced hybridization of the bonded atoms. Bonds that are too short or too long may be caused by unresolved disorder. Failure to assign a proper hybridization type to a carbon atom may be indicative of a missing (hydrogen) atom. The average and the range of C—C bond lengths within a phenyl moiety are compared with the expected value, 1.395 Å. A significant deviation may indicate incorrect cell dimensions (possibly calculated with the wrong wavelength), poor diffraction data or an incorrect refinement model. When the data quality does not support their refinement, methyl H atoms often behave badly, giving rise to unrealistic geometry. In addition, the geometrical data given in the CIF are checked to see that they correspond to the values calculated directly from the atomic coordinate data.

A quality indicator called bond precision is reported and calculated as the average standard uncertainty on C—C bonds.

4.5. Intermolecular contacts

Intermolecular contacts can be very informative in indicating incorrect structures. Obviously, when atoms approach closer than the sum of their van der Waals radii there must be either a missed interaction, such as a hydrogen bond, or their positions are in some way in error. Bumping hydrogen atoms may indicate misplaced hydrogen atoms (*e.g.* two instead of one hydrogen on an sp^2 carbon) or methyl moieties fixed in an inappropriate conformation.

4.6. Hydrogen bonds

As a rule, OH moieties are hydrogen bonded to an acceptor. A test is carried out to find out whether this is indeed the case. Potential H-atom positions lie on a cone. Finding the correct position on this cone can be tricky when the difference electron-density map does not present a single suitable maximum. *SHELXL97* (Sheldrick, 1997) provides an option to find the optimal position by way of an electron-density calculation around a circle. Alternatively, the program *HYDROGEN* (Nardelli, 1999) may also be used to find an optimal position based on geometric and energy considerations.

The analysis of structures containing hydroxy moieties obtained from the Cambridge Structural Database (CSD; Allen, 2002) gives many examples where hydroxy hydrogen atoms are in the wrong position. For a corrected example, see Körner *et al.* (2000a,b).

4.7. Connectivity

The CIF is assumed to contain a set of atomic coordinates that do not require the application of symmetry operations to connect them into chemically complete molecules and ions. The exception is where a molecule possesses crystallographic symmetry, but the asymmetric fragment should still be a

connected set. Checks are performed to identify isolated atoms. An isolated transition metal probably points to a misinterpreted identity. Isolated hydrogen atoms might need a symmetry operation to bring them into a bonding position or their bond distances might be outside the expected range. Isolated oxygen atoms generally indicate missing attached hydrogen atoms on a water molecule. Single-bonded metal atoms are also flagged since they probably represent the assignment of an incorrect atom type. A mismatch of site-occupation factors in disordered structures can also lead to connectivity alerts during validation.

Connected sets of atoms should generally have their centre of gravity within the bounds of the unit cell. Molecules or ions with their centres outside the base unit cell can sometimes arise following the inversion of a chiral structure without applying an origin shift, or from the initial set of atomic coordinates generated by the structure solution software.

4.8. Disorder

Several tests address the issue of disorder. Reported disorder can be real or an artefact resulting from poor experimental procedures. With the point detector of a serial diffractometer, it may happen that the reflection data are collected for a direct-space subcell only. The resulting structure may then be described with a 50:50 disorder model. Alternatively, an average structure that exhibits unusual displacement ellipsoids, falsely suggesting high rigid-body motion, may result (for an example, see Spek, 1993). Reported disorder is especially suspicious when the structure at hand diffracts well at high diffraction angles.

Pseudo-symmetry, particularly in cases where the structure contains real and pseudo centres of symmetry, may result in partially disordered structures when described with respect to the pseudo-symmetry element. For an example, see Spek (1993; note that Figs. 3 and 4 therein should be interchanged).

Occupancy parameters cannot have values larger than 1.0. Site-occupancy parameters in partially disordered side chains should make sense. An atom closer to the end of a chain cannot have occupancy higher than the occupancy of the previous atom in the chain (unless coinciding with an atom of the other disordered form).

Occupancy is often misunderstood for atoms on special positions. Traditionally, refinement programs (*e.g.* *SHELXL97*; Sheldrick, 1997) include site-symmetry in what is often called population parameters. Thus, a fully occupied (*i.e.* occupation = 1.0) position of an atom on a twofold axis is assigned a population parameter of 0.5. The value in the CIF for the occupancy should be 1.0 in this case. In this regard, investigators should be aware of the distinction between the two CIF data names `_atom_site_occupancy` and `_atom_site_symmetry_multiplicity`.

4.9. Completeness and consistency

Many checks address missing, incomplete or inconsistent data issues. Most of the IUCr checks published on the Internet are included (http://journals.iucr.org/services/cif/checking/ autolist.html). As a general rule, the data set should be complete up to at least $\sin(\theta)/\lambda = 0.6 \text{ Å}^{-1}$. The actual number of observed reflections is compared with the number to be expected for the stated resolution. Incomplete data sets may be caused by improper selection of the asymmetric reflection unit on a serial diffractometer or an improper set of scans with an area detector, which results in a cusp of missing data.

Several tests check for the proper application and implementation of the Flack parameter (Flack & Bernardinelli, 1999, 2000) for the determination and reporting of the absolute structure.

The numerical values of bonds, angles, torsion angles and hydrogen bonds as reported in the CIF are checked for consistency with corresponding values calculated from the coordinate data. Failures are often caused by inconsistent symmetry codes associated with the atoms, or are indicative that an old CIF has been updated with atomic coordinates from a fresh refinement without the corresponding revised geometric data being included. The reported standard uncertainties on geometry items, which are generally derived using the full covariance matrix for the parameters involved, should at least resemble the standard uncertainties calculated on the basis of the variances reported for the coordinates.

A CIF gives both an explicit list of symmetry operators and the short Hermann–Mauguin symbol. The latter symbol can be ambiguous and may lead to confusion when not given in association with the set of symmetry operators. It is therefore suggested that the Hall symbol (Hall, 1981) be included as well. Several programs generate the set of symmetry operators from this symbol. Conventionally, the twofold screw axis for space group $P2_1$ will run along the b axis. However, under certain circumstances, such as in the description of a reversible $P2_1/c \leftrightarrow P2_1$ phase transition, it may be convenient to keep the screw axis $\frac{1}{4}c$ off the b axis. The Hall symbol was introduced as a way to avoid this choice of origin ambiguity. The standard and alternative settings are then specified as 'P 2yb' and 'P 2ybc', respectively.

5. Implementation

The validation tests in *PLATON* form an integral part of the automatic geometry-analysis option of the program when a data file in CIF format is used as the input file. A scratch file is created containing data from which a validation report is generated. The actual validation is carried out against an external file named CHECK.DEF. This file contains editable validation criteria and associated printable explanatory text. The generated validation report presents a list of ALERTS. Based on the criteria in the CHECK.DEF file, they are classified as Level A, B or C ALERTS. Level A ALERTS should generally be taken very seriously. The related problem should either be resolved or discussed/explained convincingly as an exception. Level C ALERTS draw attention to non-standard issues and should be inspected to see whether they can be ignored or should be acted upon. Level B ALERTS are generally intermediate. Large numbers of Level C and Level B ALERTS

should not be ignored, as collectively they could indicate serious problems with the structure.

The ALERTS are also classified into the following four categories.

ALERT Type 1: CIF construction/syntax errors, inconsistent or missing data. This type of ALERT should be easy to address before publication.

ALERT Type 2: indicator that the structure model may be wrong or deficient. This type of ALERT should be resolved completely at the time of the analysis or commented on when published.

ALERT Type 3: indicator that the structure quality may be low. The follow-up on this will depend on the purpose of the structure determination and the publication policy of the journal.

ALERT Type 4: improvement, methodology, query or suggestion. Issues to be considered and acted upon when appropriate or necessary for the purpose of the analysis at hand.

Entries from the CSD (Allen, 2002), either in FDAT or CIF format, can also be validated. This can be useful as a prescreening tool before attempting a statistical analysis. Many outliers turn out to represent unresolved problems with the associated database entry.

```
#==============================================================================
# PLATON/CHECK-(200902) versus check.def version of 200902   for entry: (I)
# Data From: skxxxxx.cif - Data Type: CIF       Bond Precision  C-C = 0.0079 A
#
# Cell  10.9776(1)   14.0734(2)   17.3262(2)        90         90         90
# WaveLength 0.71073   Volume  Reported   2676.76(5)  Calculated   2676.76(5)
# SpaceGroup from Symmetry P b c m         Hall: -P 2c 2b
#              Reported P b c m                    ?
# MoietyFormula C27 H28 N2 O6 S, Cu
#      Reported C27 H28 N2 O6 S, Cu
#    SumFormula C27 H28 Cu N2 O6 S
#      Reported C27 H28 Cu N2 O6 S
# Mr        =      572.13[Calc],      572.11[Rep]
# Dx,gcm-3  =        1.420[Calc],        1.420[Rep]
# Z         =            4[Calc],            4[Rep]
# Mu (mm-1) =        0.938[Calc],        0.938[Rep]
# F000      =      1188.0[Calc],      1188.0[Rep]  or F000' =   1190.25[Calc]
# Reported  T limits: Tmin=0.657          Tmax=0.884 'EMPIRICAL'
# Calculated T limits: Tmin=0.694 Tmin'=0.694  Tmax=0.861
# Reported   Hmax= 12, Kmax= 16, Lmax= 20, Nref= 2423      , Th(max)= 25.00
# Calculated Hmax= 13, Kmax= 16, Lmax= 20, Nref= 2452( 2452), Ratio= 0.99( 0.99)
# R= 0.0697( 1962), wR2= 0.2215( 2423), S = 1.056, Npar= 172
#==============================================================================

        >>> The Following ALERTS were generated <<<
        -------------------------------------------
    Format: alert-number_ALERT_alert-type_alert-level text

307_ALERT_2_A Isolated Metal Atom (Unusual !) ..............    Cu(1)
430_ALERT_2_A Short Inter D...A Contact: O(3)   .. O(3)   =     2.42 Ang.
761_ALERT_1_A CIF Contains no X-H Bonds ....................    ?
762_ALERT_1_A CIF Contains no X-Y-H or H-Y-H Angles .......     ?
#==============================================================================
242_ALERT_2_B Check Low    U(eq) as Compared to Neighbors ..    C(2)
430_ALERT_2_B Short Inter D...A Contact: O(2)   .. N(1)   =     2.87 Ang.
#==============================================================================
028_ALERT_3_C _diffrn_measured_fraction_theta_max  Low .....    0.99
048_ALERT_1_C MoietyFormula Not Given ......................    ?
062_ALERT_4_C Rescale T(min) & T(max) by ...................    0.97
125_ALERT_4_C No _symmetry_space_group_name_Hall Given .....    ?
141_ALERT_4_C su on a - Axis Small or Missing (x 100000) ...    10 Ang.
142_ALERT_4_C su on b - Axis Small or Missing (x 100000) ...    20 Ang.
143_ALERT_4_C su on c - Axis Small or Missing (x 100000) ...    20 Ang.
220_ALERT_2_C Large Non-Solvent    C   Ueq(max)/Ueq(min) .     2.88 Ratio
341_ALERT_3_C Low Bond Precision on C-C bonds (x 1000) Ang..    8
#==============================================================================

         ALERT_Level and ALERT_Type Summary
         ===================================
   4 ALERT_Level_A = In General: Serious Problem
   2 ALERT_Level_B = Potentially Serious Problem
   9 ALERT_Level_C = Check & Explain

   3 ALERT_Type_1 CIF Construction/Syntax Error, Inconsistent or Missing Data.
   5 ALERT_Type_2 Indicator that the Structure Model may be Wrong or Deficient.
   2 ALERT_Type_3 Indicator that the Structure Quality may be Low.
   5 ALERT_Type_4 Improvement, Methodology, Query or Suggestion.
#==============================================================================
```

Figure 1
Example of a PLATON validation report for an erroneous structure. The associated commentary is not shown. ALERTS 307 and 430 strongly indicate that the reported structure is wrong.

6. An example

An (edited) example of a PLATON validation report is given in Fig. 1. It reports on an anonymous CIF that was submitted for publication to Acta Crystallographica Section C. The structure was supposedly a somewhat unusual coordination compound of composition Cu^{2+}.Ligand^{2-}, with an R value of 0.06 and a reasonable ORTEP plot. Validation indicated several problems, eventually leading to a revised ionic structure with composition Ligand$^+$.Br$^-$ (i.e. no copper at all!). ALERT number 307 points out that the copper ion is not coordinated by suitable oxygen functions of the ligand, which is of course extremely unusual. ALERT 430 indicates a missing hydrogen atom on N1 and a missing hydrogen atom in the short acid bridge between two carboxyl moieties. As it turned out, HBr was used in the course of the preparation of the intended copper coordination complex. Interestingly, structure determination and refinement proceeds uneventfully with currently available software packages on the basis of the supposed false composition.

7. Discussion

The correct assignment of element types to electron-density maxima can be a serious pitfall (Mueller, 2001). Nitrogen and oxygen are often interchangeable in ring systems. Misinterpretation can have important chemical consequences. A good example of this problem is the structure determination of a multi-ring anticancer marine natural product (Lindquist et al., 1991). It was shown recently by Li, Burgett et al. (2001) that one of the oxygen atoms in one of the five-membered rings should have been identified as N—H. Obviously, such a misassignment has serious consequences when one attempts to synthesize such a compound from scratch (Li, Jeong et al., 2001). Interestingly, this misassignment could have been identified easily had the intermolecular contacts been examined carefully. The packing analysis would have revealed an unusually short O···O=C contact of 2.85 (1) Å. Such a contact is only realistic when hydrogen bonding is involved. The correct interpretation is N—H···O=C. The validation software in PLATON automatically flags the problem as an unusual intermolecular contact, much shorter than the sum of the van der Waals radii.

Many interpretation errors, some with disastrous consequences, fall in the category 'missing hydrogen atoms'. Otto et al. (2002) discussed another recent example of this problem where it was shown that a supposedly unique compound (Lambert et al., 2002) containing a Cp*(+) moiety was more standard once two missing hydrogen atoms were included. Again, the standard validation procedures sent out the proper alerts.

If the authors of publications fail to validate their results or do not recognize the implications of the validation alerts, such misinterpretations should at least be picked up during the refereeing process. Unfortunately, too many unusual structural features are 'explained' as being due to the poor quality

of the crystal or with the catchall explanation of 'packing effects', as was the case in the last example.

Pseudo-symmetry can give rise to structures which initially appear to be plausible, but which have atoms or molecules misplaced with respect to the true symmetry. Several such cases can be found in the literature. R factors can be misleadingly low in such cases. A recent example has been reported by Bowes *et al.* (2002). In this case, a number of validation alerts led the authors to the true structure.

8. When to use *PLATON* data validation

It is much easier to detect and correct overlooked problems with a crystal structure during or immediately at the end of an analysis than during the publication process. The project is still fresh in ones mind, interest has not yet waned and the sample or additional crystals are readily available or can be prepared if the validation process suggests a serious deficiency that can only be rectified with a fresh data collection. The publication of the results has often to wait for months or years before related investigations have been completed. If the validation is only conducted as a pre-publication check, it may be quite difficult to resolve any problems because the sample is no longer available and the person who prepared the material has departed *etc.* Aside from this, a misinterpreted (preliminary) crystal structure may cause the chemist concerned to spend many months investigating a supposedly unexpected reaction product, where early validation could have avoided the fruitless effort.

The original data validation project initiated by *Acta Crystallographica* (*checkCIF*) was intended as a pre-publication tool. However, the *PLATON* validation checks can be run easily at any stage of the structure refinement and are recommended for use as a routine tool at the completion of every structure determination, if not after every significant change to the structural model. In this way, any aspects requiring attention can be treated promptly and efficiently before the final report is generated.

The CIF standard makes validation independent of the actual software package used for the structure determination.

9. Concluding remarks

The utility and applicability of automated data validation for the investigator is obvious. Even more importantly, the referee of a paper that reports and builds on supporting crystallographic evidence now has the tools to judge a submitted paper adequately and efficiently without having to wade through extensive supplementary material. They can have access to an automatically generated list of issues to be addressed. In case of doubt or for any other good reason they may even do their own calculations with the data, given that the reflection data have been made available as well. Of course, this requires that the data be deposited electronically as a CIF format file at the time the manuscript is submitted, either with the journal concerned or with the Cambridge Crystallographic Data Centre. Unfortunately, only IUCr

journals archive the reflection data and make them available on the Web for future use.

Automated structure validation was pioneered by *Acta Crystallographica Section C*. Structural papers and associated data are now accepted electronically and in CIF format only. It turns out to be very effective. Several other major journals are now implementing (or considering the implementation of) structure validation in their procedures.

Validation sets standards that are not just based on low R values. A validated structure that does not generate serious ALERTS can be considered 'routine' in the hands of its investigator. As a side effect, validation tests often point to unusual and interesting features in a structure (*e.g.* pseudo-symmetry) that merit further investigation and discussion. The only difficulty is that some ALERTS (*e.g.* missed symmetry) will require the experience of a professional in order to sort out their implications.

Validation is a learning process. *PLATON* currently implements more than 200 tests, but there is scope for the implementation of additional tests, particularly for inorganic structures. New tests often have their origin in problems encountered with real-world CIFs submitted to *Acta Crystallographica*.

The *PLATON* (Spek, 2002) validation software, both as source code and as executable, is freely available for academics and runs on both Unix/Linux and Microsoft Windows platforms. Access is also freely available as part of the Web-based IUCr *checkCIF* facility (http://journals.iucr.org/services/cif/checking/checkform.html). Investigators are urged to use such facilities and address all unusual issues prior to submitting a paper for publication, if not on the completion of a structure determination. Proper use of these methods can speed up the refereeing process and lead to quicker publication.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Bowes, K. F., Ferguson, G., Glidewell, C., Low, J. N. & Quesada, A. (2002). *Acta Cryst.* C**58**, o551–o554.
Eisenberg, R. (2002). *Inorg. Chem.* **41**, 1995.
Farrugia, L. J. (2000). *IUCRVAL*. University of Glasgow, Scotland. (http://www.Chem.gla.uk/~louis/software/iucrval.)
Flack, H. D. & Bernardinelli, G. (1999). *Acta Cryst.* A**55**, 908–915.
Flack, H. D. & Bernardinelli, G. (2000). *J. Appl. Cryst.* **33**, 1143–1148.
Hall, S. R. (1981). *Acta Cryst.* A**37**, 517–525.

Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Cryst.* A**47**, 655–685.

Harlow, R. L. (1996). *J. Res. Natl Inst. Stand. Technol.* **101**, 327–339.

Hirshfeld, F. L. (1976). *Acta Cryst.* A**32**, 239–244.

Johnson, C. K. (1976). *ORTEPII*. Oak Ridge National Laboratory, Tennessee, USA.

Kahn, M. L., Sutter, J.-P., Golhen, S., Guionneau, P., Quahab, L., Kahn, O. & Chasseau, D. (2000*a*). *J. Am. Chem. Soc.* **122**, 3413–3421.

Kahn, M. L., Sutter, J.-P., Golhen, S., Guionneau, P., Quahab, L., Kahn, O. & Chasseau, D. (2000*b*). *J. Am. Chem. Soc.* **122**, 9566.

Körner, F., Schurmann, M., Preut, H. & Keiser, W. (2000*a*). *Acta Cryst.* C**56**, 74–75.

Körner, F., Schurmann, M., Preut, H. & Keiser, W. (2000*b*). *Acta Cryst.* C**56**, 1056.

Lambert, J. B., Lin, L. & Rassolov, V. (2002). *Angew. Chem. Int. Ed.* **41**, 1429–1431.

Le Page, Y. (1987). *J. Appl. Cryst.* **20**, 264–269.

Le Page, Y. (1988). *J. Appl. Cryst.* **21**, 983–984.

Li, J., Burgett, A. W. G., Esser, L., Amezcua, C. & Harran, P. G. (2001). *Angew. Chem. Int. Ed.* **40**, 4770–4773.

Li, J., Jeong, S., Esser, L. & Harran, P. G. (2001). *Angew. Chem. Int. Ed.* **40**, 4765–4770.

Lindquist, N., Fenical, W., Van Duyne, G. D. & Clardy, J. (1991). *J. Am. Chem. Soc.* **113**, 2303–2304.

Marsh, R. E. & Spek, A. L. (2001). *Acta Cryst.* B**57**, 800–805.

Mueller, P. (2001). Thesis, University of Göttingen, Germany. (http://webdoc.sub.gwdg.de/diss/2001/mueller_peter/.)

Nardelli, M. (1999). *J. Appl. Cryst.* **32**, 563–571.

Otto, M., Scheschkewitz, D., Kato, T., Midland, M. M., Lambert, J. B. & Bertrand, G. (2002). *Angew. Chem. Int. Ed.* **41**, 2275–2276.

Parkin, G. (1993). *Chem. Rev.* **93**, 887–911.

Sheldrick, G. M. (1997). *SHELXL97*. University of Göttingen, Germany.

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

Spek, A. L. (1993). *Crystallographic Computing 6*, edited by H. D. Flack, L. Parkanyi & K. Simon, pp. 123–131. IUCr/Oxford University Press.

Spek, A. L. (2002). *PLATON. A Multipurpose Crystallographic Tool*. Utrecht University, The Netherlands. (http://www.cryst.chem.uu.nl/platon.)