editorial

Some guidelines for publishing *SHELXL*-generated CIF results in *Acta Crystallographica*

William Clegg

School of Natural Sciences (Chemistry), University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

A large majority of crystal structures submitted for publication in *Acta Crystallographica Section E: Structure Reports Online* have been refined with either the public domain program *SHELXL* (Sheldrick, 1997) or its commercial relative *SHELXTL* (Sheldrick, 2000). This paper is intended to provide help and guidance to authors using these programs and to alert them to various features of the software, including default settings, in the light of actual procedures. This will lead to a consistent presentation of results that conforms to the 'house style' of the journal. A number of these issues are being discussed with Professor George Sheldrick, the author of the software, and it is hoped that they will be addressed in the next release of the programs. In the meantime, this editorial article provides advice that should help to avoid such difficulties. It addresses particularly users of the *SHELX* family of programs, though some of the points will be of relevance to users of other software for refinement and for generation of CIF output for publication. It may be read with advantage by any intending author of papers in *Section C* and other IUCr journals where CIFs are required, as well as for *Section E*. The most serious problems are detected in the combined checking procedures of *checkCIF* and *PLATON* (Spek, 2002, 2003), but there are numerous, more minor points that do not raise *checkCIF* alerts and are currently dealt with editorially.

1. Information on experimental and refinement procedures

It should be remembered that a number of experimental details are inserted in the CIF by the software (not only the refinement program, but also some other components of *SHELXTL*) on the basis of an assumption rather than known fact and that authors should replace incorrect information with details relevant to their own work. Particular examples are the type of X-ray tube and monochromator, diffractometer and scan type. Although not all of these items appear in the HTML and PDF versions of the published article, they are included in the deposited CIF, which is made available to readers as an integral part of the publication; many published articles state that 'φ and ω scans' were used on SMART and other CCD diffractometers, and it is likely that this is by no means always true. Similarly, the software always reports that the structure has been solved by direct methods and that H atoms have been refined by 'mixed' methods (even when the structure contains no H atoms at all!).

Absorption corrections based on azimuthal scan (ψ scan) measurements of a selected set of reflections, commonly used with four-circle diffractometers, should be flagged by the code 'psi-scan', while the use of programs like *SADABS* (Sheldrick, 2002) and *HKL SCALEPACK* (Otwinowski & Minor, 1997), based on procedures described by Blessing (1995, 1997), are correctly flagged as 'multi-scan'. The term 'empirical' is ambiguous and should not be used. *DIFABS* (Walker & Stuart, 1983) and its derivatives, in which absorption is modelled on the basis of a comparison of $|F_o|$ and $|F_c|$ values, is a 'refdelf' method.

Note that the inclusion of a SIZE instruction in a *SHELXL*/*SHELXTL* refinement produces values of minimum and maximum transmission factors in the CIF. These numbers are estimates of the *expected* transmission factors, based

purely on the crystal dimensions and the absorption coefficient (calculated from the unit-cell contents declared in the SFAC and UNIT instructions); they are *not* the experimentally determined values and should be replaced by the values derived or estimated from the actual absorption correction, or reset to question marks if no absorption correction has been applied (in which case the absorption correction method should be 'none').

2. H-atom treatment and disorder

As mentioned above, the item 'mixed' is automatically inserted in the CIF by the refinement software, regardless of the actual treatment used for the H atoms. This means that some H atoms have been constrained, while others were refined freely; such a procedure is often used when $C-H$ groups are easily predictable, but $N-H$ and/or $O-H$ groups are not uniquely determined by the rest of the molecular geometry. In many cases, however, it is not the correct description, and it should often be replaced by either 'All H-atom parameters refined' (for free refinement of all H atoms) or 'H-atom parameters constrained' (when all H atoms are constrained, usually with a riding model); there are some other possible terms, but they are not so widely used. Note that AFIX/HFIX instructions are *constraints*, while instructions such as SAME/SADI/DFIX are *restraints*; the latter act in a quite different way in the refinement, and such H atoms are not correctly described by 'H-atom parameters constrained'. Clearly, the simple menu of CIF terms available is not really adequate, and the term chosen and set in the CIF is intended only as an overall summary. The actual method used should always be described in the text of the `_publ_section_exptl_refinement` section; this need be only a single short sentence for standard procedures, such as routine use of constraints or completely free refinement, but a fuller description will be needed in other cases.

Likewise, any structural disorder needs to be described for the benefit of readers. There may be some description in the *Comment* section, depending on the importance of this feature of the structure, but there should generally be an account of the modelling of the disorder in the `_publ_section_exptl_refinement` section; this should include the values of refined occupancy factors.

The various `_special_details` sections of the CIF are available for inclusion of any information that could be of interest and value to readers but is not for the HTML or PDF published version.

It is recommended that the full list of bond lengths and angles, including those involving H atoms, should always be present in the CIF (by use of the BOND $H instruction), even when H atoms have been constrained; this allows readers to see what assumptions have been made, and provides Co-editors with an opportunity to assess them. Note that the refinement program outputs constrained $X-H$ distances to four decimal places, although most refined bond lengths are precise to only three places (and refined $X-H$ distances usually to only two places); it is advisable to truncate these values in the CIF before submission.

Hydrogen bonds are readily generated by HTAB instructions, and the information thus generated in the CIF is automatically converted into a table in the publication. In order to select some entries for publication, and suppress others while leaving them in the CIF, the CIF data name `_geom_hbond_publ_flag` should be inserted as an extra line immediately before the first entry in the list of hydrogen bonds, and each successive line in the list should be flagged with either `yes` or `no` (or a period) in the same way as the other geometry lists in the CIF. Angles at H atoms in this list not carrying a standard uncertainty (s.u.) because they involve constrained H atoms should be rounded to an integer value, since this is the precision typically achieved when such H atoms are freely refined.

One potential source of serious errors is the generation of hydrogen-bonding information in a different program (such as *PLATON*) for subsequent pasting back into the *SHELXL*-generated CIF. Unfortunately, the two programs often store the space-group symmetry operations in a different order, and so the symmetry codes produced may not be compatible with those defined in the CIF. If this procedure is used instead of the HTAB instruction of *SHELXL*, the symmetry codes must be carefully checked and amended to be consistent. This error is detected by the standard *checkCIF*/*PLATON* checking procedures, which authors are asked to use before submitting a manuscript.

3. Issues of precision and rounding

One problematic feature of the CIF-generation facilities of *SHELXL* is the conversion of any s.u. of 1 into a value of 10, with the concomitant addition of a zero digit on the end of the corresponding parameter value. This occurs most often in the unit-cell parameters. Thus, a cell angle of 95.437°, for example, with its s.u. given as 0.001° in the ZERR instruction, appears in the CIF as 95.4370(10) instead of 95.437(1). The appearance of greater experimental precision thus created is entirely spurious. Although the journal has a policy of preferring s.u. values to be between 2 and 19 in the final digit of refined

and derived parameters (the so-called 'rule of 19'), an honest s.u. of 1 is far preferable to an invented 10! Authors are requested to check such values carefully, particularly in their unit-cell parameters, before submitting a CIF for publication. An unfortunate consequence of the much greater efficiency and automation provided by the CIF publication mechanism is that details like this are more easily overlooked, leading to the potential introduction of misleading information.

Note that X-ray wavelengths are always padded to 5 decimal places and crystal dimensions are truncated to 2 in the CIF output of *SHELXL* and *SHELXTL*. This is of importance to synchrotron users, for whom the wavelength is usually not known to such precision, and for whom crystals may indeed be on the micron scale. Appropriate corrective action is required during editing of the CIF for publication.

The term '+ 0.00000P', frequently appearing in refinement weighting schemes, should be deleted, as it is essentially meaningless; trailing zeros should be removed from the decimal fraction parts of numerical terms in weighting-scheme formulae and anywhere else in the absence of s.u. values requiring a certain number of decimal places to define the precision.

The use of CONF and appropriately chosen MPLA instructions allows s.u. values to be obtained direct from the refinement for torsion angles and interplanar dihedral angles, when it is desired to quote these in the publication. These s.u. values are more reliable than any obtained by subsequent calculations in other programs, because covariance effects are taken into account; once the refinement program has terminated, the covariance information is lost.

4. Refinement of non-centrosymmetric structures

There has been much debate about the correct procedures for refining non-centrosymmetric structures, interpreting the results in terms of absolute configuration, polarity or other terminology, and the statistical significance of such results, and these will not be repeated here (Flack & Bernardinelli, 1999, 2000). *SHELXL* and *SHELXTL* adopt the method of refining the Flack (1983) parameter, which is obtained together with an associated s.u., allowing its significance to be assessed. A number of problems and misunderstandings can arise from inexpert use of this method.

It has been demonstrated that the most reliable results are obtained by explicit refinement of the Flack parameter through combined TWIN and BASF instructions, rather than allowing the program to calculate this separately from the other refined parameters

(the default action in the absence of these instructions). However, the inclusion of TWIN *without* BASF leads to a completely false result, with a Flack parameter of exactly zero and no s.u., giving the impression that the structure has been refined with the correct hand. *checkCIF/PLATON* now detects this error, raising a serious alert.

In cases where the Flack parameter can not be reliably determined, because of insufficient anomalous scattering effects (*e.g.* when a non-centrosymmetric structure containing only C, H, N and O atoms is refined from data collected with Mo $K\alpha$ radiation, a common occurrence), Friedel pairs in the data set should be merged and not used as independent data (MERG 3 or 4 instruction), and the meaningless Flack parameter should not be reported. It is recognized and accepted that the data/parameter ratio will be lower in such cases. When the Flack parameter is refined, its value and s.u. should be reported (even if the correct absolute configuration is known from other information, such as a non-inversion synthetic route from an enantiomerically pure starting material of known configuration), and the number of Friedel pairs in the data set should be reported. The preferred format for this is, for example, 'Flack (1983); 1945 Friedel pairs' as the entry for the CIF data name `_refine_ls_abs_structure_details`, with the full reference given in the reference list (not here, as is done by the refinement program). A simple method of determining the number of Friedel pairs is either to read the data set successively into *SHELXS* and *SHELXL* without MERG instructions, or to read the data set twice into *SHELXL* (once without MERG and once with MERG 3), and to take the difference in the number of unique reflections reported in each case.

5. Other points of notation and references

The notation used by the *SHELX* family of programs for space groups is not in accord with standard IUCr CIF practice. Each part (the Bravais lattice and then the symbols for axes and/or planes) should be separated by spaces, and parentheses should not be used for subscripts; because of the embedded spaces, surrounding quotation marks are needed. This is best illustrated by common examples. Thus, P-1 should be 'P -1'; P2(1)/c should be 'P 21/c'; P2(1)2(1)2(1) should be 'P 21 21 21'; and (less common!) I4(1)/amd should be 'I 41/a m d'.

The intensity threshold for 'observed' reflections should be written 'I>2\s(I)' instead of '>2sigma(I)'.

The most commonly used references for *SHELXL*97 (the most recent version of the public domain refinement program) and *SHELXTL* are as given below (Sheldrick,

1997, 2000). Where structure solution has used the public domain *SHELXS*97 software, this may be referenced as (Sheldrick, 1990) or it may be combined with *SHELXL*97 (Sheldrick, 1997), as indicated here, in preference to having two separate and almost identical 1997 references. The year cited for *SHELXTL* tends to vary, depending on which particular upgrade has been used, and some authors cite the software as, for example, '(Bruker, 2001)', with further variations in the name of the firm (including Siemens and Bruker–Nonius), because of a history of mergers and commercial take-overs. Consistency is desirable, *e.g.* in not having a reference to Bruker with a date of 1993, or Siemens with 2001.

Versions of *SHELXL* earlier than 1997 are considered obsolete by their author, and should no longer be in use. *SHELXL93* contained a number of known faults, and these particularly affect the generation of CIF output. The journal has a policy of recommending authors to update to the current version and hence obtain a more acceptable CIF. For further information, see Sheldrick (2003).

References

Blessing, R. H. (1995). *Acta Cryst.* A**51**, 33–58.
Blessing, R. H. (1997). *J. Appl. Cryst.* **30**, 421–426.
Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.
Flack, H. D. & Bernardinelli, G. (1999). *Acta Cryst.* A**55**, 908–915.
Flack, H. D. & Bernardinelli, G. (2000). *J. Appl. Cryst.* **33**, 1143–1148.
Otwinowski, Z. & Minor, W. (1997). *Methods in Enzymology*, Vol. 276, *Macromolecular Crystallography*, Part A, edited by C. W. Carter Jr and R. M. Sweet, pp. 307–326. New York: Academic Press.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. (1997). *SHELXS*97 and *SHELXL*97. University of Göttingen, Germany.
Sheldrick, G. M. (2000). *SHELXTL*. Version 6.10. Bruker AXS Inc., Madison, Wisconsin, USA.
Sheldrick, G. M. (2002). *SADABS*. Version 2.03. University of Göttingen, Germany.
Sheldrick, G. M. (2003). http://SHELX.uni-ac.gwdg.de/SHELX/index.html
Spek, A. L. (2002). *PLATON*. University of Utrecht, The Netherlands.