

## 5.7. Small-molecule crystal structure publication using CIF

BY P. R. STRICKLAND, M. A. HOYLAND AND B. MCMAHON

### 5.7.1. Introduction

The International Union of Crystallography (IUCr) has always understood the importance of the accurate reporting of numerical results, and as far back as its early sponsorship of the Standard Crystallographic File Structure (Brown, 1983, 1988) the IUCr has explored the use of exchange files in publishing (see Chapter 1.1). In 1991, when the first draft of the CIF standard was nearing completion, the main journal of the IUCr for reporting crystal structures, *Acta Crystallographica Section C: Crystal Structure Communications* (hereafter *Acta Cryst. C*), consisted of a collection of concise reports of crystal and molecular structures presented in a standard format that would lend itself well to computerized markup and typesetting from an appropriate input file format. It seemed natural, therefore, to use this journal to test the new draft CIF standard and to develop techniques for machine-based checking of structural data along with the new methods for submitting, typesetting and distributing a crystal-structure report in electronic format. Although adopting a novel data-exchange format for the submission and handling of research papers might have seemed a radical and audacious development, the potential benefits in terms of accuracy and speed of publication were clear.

In parallel with the publication of the CIF standard (Hall *et al.*, 1991), an *Editorial* and revised *Notes for Authors* in *Acta Cryst. C* described the new route to publication using CIFs and invited the crystallographic community to cooperate in this innovative practice. The same issue of the journal contained the first paper to be published by this route (Willis *et al.*, 1991).

This first paper was the outcome of a testing phase which involved considerable interaction with the authors. The first unsolicited article to be submitted in CIF format appeared in the February 1992 issue of *Acta Cryst. C*. A few more were submitted during 1992, the number gradually increasing through the following year. Authors quickly adapted to the compartmentalized style of text entries and by the beginning of 1994 the level of CIF submissions allowed the journal to introduce a production stream that promised faster publication times for articles submitted electronically as CIFs. By the beginning of 1996, it became journal policy to accept *only* electronic submissions in CIF format.

The IUCr was not the only publisher to introduce the submission of structure reports in machine-readable form. In 1990, *Zeitschrift für Kristallographie*, published by R. Oldenbourg Verlag, introduced a new section for the publication of short inorganic and small-molecule structural papers with minimal commentary. To submit a report to this section, the author would use the output file from the refinement program *SHELX76* (Sheldrick, 1976) (at that time a *de facto* exchange standard on account of its widespread distribution), which was processed by a specially developed program *CASTOR* to create a self-contained file for use in publication. When CIF was introduced, it was also accepted as a submission format for this section of *Zeitschrift*. The section flourished and in 1997 it became a separate journal, *Zeitschrift*

*für Kristallographie – New Crystal Structures*. CIF is now the standard submission format for this journal as well as for *Acta Cryst. C*.

In an era dominated by information retrieval *via* the world wide web, it is easy to forget that these innovations in crystallographic publishing predated the http protocol and the universal availability of graphical browsers. However, the independently developed but well defined CIF exchange standard proved easy to integrate with the publication procedures developed for electronic journals. The current delivery formats available to journals like *Acta Crystallographica* and *Zeitschrift für Kristallographie* are HTML and PDF. Nevertheless, the original CIF data are still accessible, and allow readers to visualize structures interactively in three dimensions or perform their own analyses of structural models.

The highly automated submission, checking and publication procedures of *Acta Cryst. C* and the online-only journal *Acta Crystallographica Section E: Structure Reports Online* (hereafter *Acta Cryst. E*) are described in detail in Section 5.7.2 as a case study for the publication of structure reports that are highly ordered in format. However, there are only a few journals that report detailed crystal structures and they represent a very specialized field of publishing. Section 5.7.3 discusses publications in which the reporting of structural data is only a minor or supplementary element of the article. It will become apparent that many of the considerations behind the design of a workflow for handling data-rich papers are also relevant to maximizing the value of data presented in or referenced by any scientific publication.

### 5.7.2. Case study: the fully automated reporting of small-unit-cell crystal structures

This section describes the route to publication of a small-molecule or inorganic single-crystal structure in *Acta Cryst. C* or *E* from the perspective of an author.

#### 5.7.2.1. Assembling the complete article

For many authors the generation of a CIF suitable for publication is quite straightforward, since diffractometer software and structure solution and refinement packages have all been capable of writing or reading the CIF format for some time. In some highly integrated systems, the entire experimental, analysis and report-generating pathway may be controlled through a common user interface.

In other cases, different components must be collected from different sources and merged together, either by software utilities or, in the worst case, by hand-editing. It is a useful feature of the text-based CIF format that it can be modified by text editors or in certain word-processing modes; indeed, this was the only way in which the earliest CIF-based papers could be constructed. However, significant expertise and understanding of the technical details of the file format are needed to produce hand-edited files that are totally free from error. Authors are now encouraged to use software designed to help them create complete and error-free files (*e.g.* the *enCIFer* and *CIFEDIT* editors described in Chapter 5.3).

A complete structure communication comprises the following components.

Affiliations: PETER R. STRICKLAND, MICHAEL A. HOYLAND and BRIAN MCMAHON, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

## 5. APPLICATIONS

- (a) Material common to the article as a whole:
  - (i) title and authors;
  - (ii) synopsis and/or abstract;
  - (iii) comment section;
  - (iv) acknowledgements;
  - (v) references.
- (b) Material relevant to each structure:
  - (i) description of the experimental apparatus;
  - (ii) description of the settings and environmental conditions for the experiment;
  - (iii) experimental data, typically a list of measured and calculated structure factors for a single-crystal X-ray structure determination, or powder diffraction data with measured and calculated powder diffraction profiles;
  - (iv) information about the compound, including source, preparation and formula;
  - (v) summary of structure solution and refinement;
  - (vi) coordinates of atomic sites, their elemental composition, occupancy, anisotropic displacement parameters, whether they are in part of the structure affected by positional disorder, and information about their refinement restraints;
  - (vii) selected geometrical data.
- (c) Graphical illustrations:
  - (i) chemical structural diagrams;
  - (ii) chemical diagrams of reaction pathways, tautomerism, bond properties *etc.*;
  - (iii) crystallographic displacement-ellipsoid diagrams;
  - (iv) crystallographic packing diagrams;
  - (v) other graphs, plots or images.

Different journals will have different requirements for the arrangement of these items. For example, at the time of publication (2005), *Acta Crystallographica* requires that diffraction data (structure factors or Rietveld refinement profiles) are provided as supplementary information in separate files from that containing the body of the paper. This policy originated in the early days of network file transfer where relatively large files of experimental data could be transferred only with difficulty. This is less of a practical constraint now, and a case could be made for including the experimental results as an integral part of a single submission file, especially since there is still no formal mechanism in the core CIF dictionary to enforce an unambiguous connection between separate data blocks containing related data.

There is also not at present a standard way to include graphics within a CIF. The mechanisms of the imgCIF dictionary (Chapter 3.7) offer a possible approach to this problem. It is also possible to envisage the automated generation of views of the structure directly from the numerical data in the CIF. Three-dimensional ellipsoid plots are routinely generated from CIFs submitted to *Acta Crystallographica* for use in the review process and incomplete categories of data names exist in the core dictionary for the representation of two-dimensional diagrams of chemical connectivity. At present, however, neither of these is sufficiently well developed to generate publication-quality graphics in different orientations and styles as preferred by an author.

A journal may provide a *request list* of the data items that it considers recommended or mandatory. The request list for *Acta Cryst. C* and *E* is given in Appendix 5.7.1. An author can test a file intended for publication against a request list with a general-purpose CIF parsing tool such as *cif2cif* (Bernstein, 1998) or *QUASAR* (Hall & Sievers, 1993) (Chapter 5.3). Different request lists may be provided for different kinds of experiments, such as

for powder diffraction experiments or for single-crystal studies using area detectors.

Note that an author always has the freedom to include additional data items in a CIF; the journal will exercise its own policy for the handling of data items not specified in its public request lists. The PUBL\_MANUSCRIPT\_INCL category available in the CIF core dictionary provides a mechanism for requesting the publication of data items that are not normally published by the journal (see Sections 5.7.2.3 and 3.2.5.5).

### 5.7.2.2. Reporting multiple structures and using templates

In CIF format, a data name cannot be repeated within a data block. Therefore, each structure reported in a CIF must occupy a separate data block. A journal might request a separate file for each structure; in the case of *Acta Cryst. C*, however, a single file for the entire submission is required. This file therefore contains several data blocks if the article reports several structures. The data-block codes (*i.e.* the changeable *label* part of a data-block header `data_label`) have no particular significance and are usually chosen by the authors as meaningful identifiers within their own collection of structures. However, each block code may be used *once only* in any individual file.

If an article reports only one structure, the author can include the general text of the article in the same data block that records the structure or in a separate data block. If the file already contains several data blocks (because it reports multiple structures), using a distinct data block for the text of the article is the most natural way of organizing the contents of the file. Fig. 5.7.2.1 shows the structure of a CIF that describes several structures.

Authors often have one or more local template data blocks that already include standard information about their contact details and details of the experiment. These templates may then be added or merged into the data blocks reporting the structures. Several standard crystallographic software packages include programs for merging CIF templates; one of the best known and most widespread is *SHELX97* (Sheldrick, 1997).

Some authors also use programmable macro facilities within commercial word-processing packages to achieve the same purpose. The IUCr application *printCIF for Word* (Westrip, 2004) extends this approach by creating a custom editing and formatting environment within Microsoft *Word*. These are very helpful utilities for authors who are not CIF experts. However, they are restricted to particular operating systems or software environments and are thus not universally available.

The program *enCIFer* (Allen *et al.*, 2004) provides facilities for importing templates and external files, and for adding and maintaining standard information about the authors of a CIF. It provides alternative representations of a CIF as a text file and as a collection of containers and object fields, and provides a great deal of support for authors who are not familiar with the technical details of the CIF format. *enCIFer* and other useful text-editing programs are described in Chapter 5.3.

### 5.7.2.3. Adding extra information to an article

An article for publication in *Acta Cryst. C* or *E* is built from a standard request list of CIF data items. Among the items included in this list are ones that describe molecular geometry: bond and contact distances, bond angles and torsion angles. In most cases, unexceptional values of these are not worth displaying (particularly as *Acta Cryst. C* and *E* make the original CIF data available as supplementary material). Authors can choose which values are to be displayed using a 'publication flag'. For example, the category

```
# ----- Material relevant to the whole paper -----
data_paper
  _publ_section_title
  ; Two dimeric nickel(II) carboxylates
  ;
loop_
  _publ_author_name
  _publ_author_address
  'Morooka, Mitsuo'
  ; Department of Chemistry
  ...
  ;
  'Ohba, Shigeru'
  ; Department of Chemistry
  ...
  ;
  _publ_section_abstract
  ; The structures of two dimeric NiII carboxylates
  and one dimeric NiII silanecarboxylate have been
  determined...
  ;

# -----Material relevant to the first structure ----
data_I
  _chemical_formula_sum      'C34 H54 N2 Ni2 O8'
  _chemical_formula_moiety   'C34 H54 N2 Ni2 O8'
  _chemical_formula_weight    736.2
  _symmetry_cell_setting     monoclinic
loop_
  _symmetry_equiv_pos_as_xyz
+x, +y, +z      -x, +y, 1/2-z
-x, -y, -z      +x, -y, 1/2+z
1/2+x, 1/2+y, +z  1/2-x, 1/2+y, 1/2-z
1/2-x, 1/2-y, -z  1/2+x, 1/2-y, 1/2+z
  _symmetry_space_group_name_H-M      'C 1 2/c 1'
  _symmetry_space_group_name_Hall     '-C 2yc'
  _cell_length_a                      20.520(2)
  _cell_length_b                      10.647(1)
  _cell_length_c                      18.260(2)
  _cell_angle_alpha                   90.0
  _cell_angle_beta                    91.015(8)
  _cell_angle_gamma                   90.0

# --- Material relevant to the second structure ----
data_II
  _chemical_formula_sum      'C80 H68 Cl6 N2 Ni2 O8'
  _chemical_formula_moiety   'C78 H66 N2 Ni2 O8, 2(C H Cl3)'
  _chemical_formula_weight    1515.5
  _symmetry_cell_setting     triclinic
loop_
  _symmetry_equiv_pos_as_xyz
+x, +y, +z      -x, -y, -z
  _symmetry_space_group_name_H-M      'P -1'
  _symmetry_space_group_name_Hall     '-P 1'
  _cell_length_a                      13.231(1)
  _cell_length_b                      13.857(1)
  _cell_length_c                      11.425(1)
  _cell_angle_alpha                   99.48(1)
  _cell_angle_beta                    104.63(1)
  _cell_angle_gamma                   109.68(1)
```

Fig. 5.7.2.1. Structure of a CIF describing several crystal structures.

of data items that describes bond lengths includes the data name `_geom_bond_publ_flag`, which may be assigned the value 'yes' or 'no' for any particular bond length depending on whether it should or should not be displayed.

The other items in the request list comprise the complete set of items that are by default extracted for publication from a CIF if they are present. An author may of course add more detail to an article within standard free-text fields (such as `_publ_section_`

`comment`). However, if the additional information is present as a data item that is not in the standard request list, the typesetting software can be told to add this item dynamically to the request list, thus including the extra information in the published article. The way to do this is to list the additional data name or names as values of `'_publ_manuscript_incl_extra_item'`. The example below shows how to request that atom-site multiplicities and Wyckoff symbols are included in the table of atomic positions. These are data names defined in the core dictionary; this is indicated by the value 'yes' of `_publ_manuscript_incl_extra_defn`.

```
loop_
  _publ_manuscript_incl_extra_item
  _publ_manuscript_incl_extra_defn
  '_atom_site_symmetry_multiplicity'    yes
  '_atom_site_Wyckoff_symbol'          yes
  '_Smith_crystal_magnetic_perm'       no
```

In this example, the author has also requested the publication of the value of the magnetic permeability of the crystal, which does not have a standard dictionary definition, but which has been recorded under a local data name, `_Smith_crystal_magnetic_perm`. Note that for this item, `_publ_manuscript_incl_extra_defn` takes the value 'no'. The journal typesetting software has no procedure for handling arbitrary additional content, but it may be configured to recognize such a data name and typeset it in the desired style. Once the software is aware of this new item, it will automatically extract and format it in future submissions, as long as the author continues to list it under `_publ_manuscript_incl_extra_item`. It is best if the informal data name includes a registered reserved prefix (see Section 3.1.2.2), especially if machine-readable definitions are also provided in an appropriate DDL dictionary format and accessible through the IUCr register of CIF dictionaries (Section 3.1.8.2).

Care is needed when using `_publ_manuscript_incl_extra_item`:

(i) The extra items requested must be surrounded by quote marks, otherwise CIF software will try to interpret them as active data names.

(ii) The list is cumulative: if several `_publ_manuscript_incl_extra_item` loops appear in the file (one per data block), the request list that is generated will include all the extra items that appear in all of these loops, and that request list will be applied in full to all the data blocks in the file. It is therefore not possible to ask for an extra item from one data block but not another.

(iii) Not all possible terms in the official dictionaries may be recognized and handled appropriately by the journal software. To check this, the author can generate a preview of the formatted paper by using the *printcif* service, described in Section 5.7.2.4.

Two examples of this approach are shown in Fig. 5.7.2.2. Atom-site positions and displacement parameters are often displayed without the associated Wyckoff symbols or multiplicities (to save space). In the first example, the author indicates that the Wyckoff symbols should be displayed.

In the second example, the author wishes to publish a table of a set of items not defined in the core CIF dictionary (in this example, contact distances with associated charge density and Laplacian functions). Here, utility data names are used to extract regularly tabulated data of arbitrary content from the CIF to create a table in the published article.

#### 5.7.2.4. Previewing the article

The appearance of the plain-text ordered arrangement of content in a CIF differs a great deal from its typeset representation in

```

loop_
  _publ_manuscript_incl_extra_item
  _publ_manuscript_incl_extra_defn
  ' _atom_site_Wyckoff_symbol'          yes

loop_
  _atom_site_label
  _atom_site_type_symbol
  _atom_site_Wyckoff_symbol
  _atom_site_fract_x
  _atom_site_fract_y
  _atom_site_fract_z
  _atom_site_U_iso_or_equiv
Sn1 Sn 8g 0.20466(6) 0.28795(8) 0.2500 0.0104(2)
Ga1 Ga 8g 0.20466(6) 0.28795(8) 0.2500 0.0104(2)
Sn2 Sn 8f 0.0000 0.31416(11) 0.04613(8) 0.0120(3)
Ga2 Ga 8f 0.0000 0.31416(11) 0.04613(8) 0.0120(3)
Sm1 Sm 8e 0.20025(4) 0.0000 0.0000 0.0092(1)
Sn3 Sn 4c 0.0000 0.02592(13) 0.2500 0.0091(4)
Ga3 Ga 4c 0.0000 0.02592(13) 0.2500 0.0091(4)
Sm2 Sm 4c 0.0000 0.63989(8) 0.2500 0.0097(1)
          (a)

loop_
  _publ_manuscript_incl_extra_item
  _publ_manuscript_incl_extra_defn
  ' _geom_extra_tableA_col_1'          no
  ' _geom_extra_tableA_col_2'          no
  ' _geom_extra_tableA_col_3'          no
  ' _geom_extra_tableA_col_4'          no
  ' _geom_extra_table_head_A'          no
  ' _geom_table_footnote_A'           no

  _geom_extra_table_head_A
; Distances (l, \%A) and values for the charge
density [\r(r), e \%A^-3] and the Laplacian
[\\nabla^2\r(r), e/\%A^-5] at the (3,-1) critical
points for attractive shortened intramolecular
contacts in (I).
;
loop_
  _geom_extra_tableA_col_1
  _geom_extra_tableA_col_2
  _geom_extra_tableA_col_3
  _geom_extra_tableA_col_4
Contact      l          \r(r)  \\nabla^2\r(r)
N5...C8      2.813(5)  0.092  1.19
O6...N1      2.913(5)  0.079  0.95
N1...H16B    2.58      0.070  0.82
O1...H3      2.26      0.111  1.60
O5...H15     2.24      0.112  1.50
O7...H30B    2.34      0.095  1.28
O7...H31C    2.36      0.093  1.25
O7...H34B    2.36      0.092  1.24
O7...H35C    2.37      0.090  1.24

  _geom_table_footnote_A      # maps to \tablefn
; The sums of the van der Waals radii (Zefirov &
Zorky, 1989) are N and C = 3.21 \%A, N and
O = 2.79 \%A, N and H = 2.66 \%A, and
O and H = 2.45 \%A.
;
          (b)

```

Fig. 5.7.2.2. Examples of authors' request-list extensions for items not normally printed in a paper. (a) Printing additional standard data items. The data are listed as normal in the ATOM\_SITE loop. (b) A complete table of non-standard quantities associated with contact distances is generated, complete with table caption and footnote.

a journal article. It can help authors, therefore, if they can see how their article will appear in print (or as an online article) before they formally submit their article to a journal. *Acta Cryst. C* and *E* provide an online web service for this called *printcif* (<http://journals.iucr.org/services/cif/printcif.html>).

When an author uploads a CIF to the service, the data within it are extracted (using a dynamically enhanced request list if the publication of extra items has been requested) and translated through a sequence of software filters to T<sub>E</sub>X (Knuth, 1986). The T<sub>E</sub>X file is processed and a final document representation (a 'preprint') in PostScript or Portable Document Format (Adobe Systems Incorporated, 1999, 2004) is generated. The preprint is then downloaded to the author. The primary translation engine is the program *cif-tex* (Section 5.3.5.3). However, *printcif* has additional content filters which are not distributed with *cif-tex*; these are modified frequently to make additional pattern-based text substitutions or to make changes to the typographic style of the preprint to match any changes in the style of *Acta Cryst. C* or *E*.

A new approach to document formatting is being explored in the development of *printCIF for Word* (Westrip, 2004), an embedded Visual Basic application suitable for CIF editing and formatting within *Word* (Section 5.3.3.4.2). This allows users to preview their article as they work on it. However, *printCIF for Word* does not have access to the constantly updated translation filters used by *printcif*.

### 5.7.2.5. Data validation

The highly structured format of a CIF allows automated validation of the self-consistency and integrity of the structural data reported in it. What was traditionally a part of the referee's task in checking crystal structure papers can now be handled by software. *Acta Cryst. C* and *E* require authors to check their structures before submitting them for publication. The same checks are run on each CIF after submission and a report of the results is made available to the referees for use during the peer-review process.

The routine checking of submissions for errors was introduced by the IUCr journals in the early 1990s, initially as a manual procedure. When CIF was introduced, the new format was readily adopted as a standard interchange format from which the input files for different checking programs could be generated automatically. The development of a workflow based on CIF proved worthwhile, as CIF increasingly became the format for submission in the first place. Over time, too, much of the checking software became capable of reading CIFs directly, so that the intermediate data-conversion processes could be avoided.

Over several years, a great deal of experience was gained in the types of error that could most easily be detected using checking software. A major component of the checking suite was *UNIMOL*, which had been developed by the Cambridge Crystallographic Data Centre for checking the molecular geometry of database entries (Allen *et al.*, 1974). Other types of checks could be performed by running other general-purpose crystallographic packages under the direction of pre-defined scripts designed to exploit their particular strengths. Among the programs used in this way were *NRCVAX* (Gabe *et al.*, 1989), which incorporated the powerful *MISSYM* algorithm of Le Page (1988), *PARST* (Nardelli, 1983), an early version of *PLATON* (Spek, 1990) and the *BUNYIP* routine for detecting additional symmetry (Hester & Hall, 1996) within the *Xtal* program system (Hall *et al.*, 2000).

As experience grew in running these processes in increasingly automated ways, and in collecting, parsing and reformatting the most relevant diagnostic output, it became apparent that a modular system could be designed to perform most of the data checking entirely automatically. Preliminary work on the set of tests developed for the *PREPUB* component of the *Xtal* system (du Boulay & Hall, 1996) led, through close cooperation with the IUCr editorial office and Ton Spek, the author of *PLATON* (Spek, 2003), to the

implementation of *checkcif*, which is described in Section 5.7.2.6 below.

### 5.7.2.6. Automated data validation: *checkcif*

The current service for checking structural data submitted to IUCr journals is known as *checkcif* and is available at <http://journals.iucr.org/services/cif/checkcif.html>. Versions of this service have been made available to other publishers for some time. In 2003, a general service was introduced at <http://checkcif.iucr.org> to provide structural checks on CIF data sets destined for publication in non-IUCr journals or database deposition, or indeed to allow authors to assess the quality of their structure determinations whether they wish to publish them or not.

The tests carried out by *checkcif* include:

- (i) a simple file syntax check: essential in the early days of manual CIF construction, but of less importance now as syntax-preserving editing programs have become more widespread;
- (ii) tests for the self-consistency of mutually dependent data items present in the CIF;
- (iii) a large collection of analytic tests on structural chemistry and molecular geometry based on the program *PLATON* (Spek, 2003).

The checks carried out at the time of publication (2005) are listed in Appendix 5.7.2 and on the CD-ROM accompanying this volume. The current list is available from <http://journals.iucr.org/services/cif/datavalidation.html>.

Although the results from *checkcif* provide valuable indications of possible inconsistencies or data errors, an article for publication is not accepted or rejected on the basis of the *checkcif* report alone. The report is always read by a reviewer as part of a considered critical appraisal of the article.

Sometimes, particular data values are so far from the expected values that some response is required from the author to explain them. The unusual values may be a consequence of poor experimental conditions that the author was unable to improve, or of poor crystal quality; they may indicate an uncertainty in part of the structure determination that the author considers acceptable, particularly if the purpose of the study is to concentrate on a different part of the structure; or they may genuinely indicate novel chemical features. Whatever the case, anomalous values usually need to be discussed by the author and the reviewer or editor, and often need to be commented on in the article. For *Acta Cryst. C* and *E*, *checkcif* generates in CIF format a list of the tests that have highlighted unusual values in the author's CIF (called 'A alerts'), together with a text field for each of these tests in which the author may justify or discuss the apparently anomalous results (see Fig. 5.7.2.3). Together these comprise a 'validation reply form'. The author can complete this form and paste it into the final version of the CIF submitted for publication. The editor handling the paper can then read the comments in the validation reply form and decide whether to accept the paper for publication. The submission system will automatically return to the author any CIF which generates an A alert but does not contain a completed validation reply form.

Every article published in *Acta Cryst. E* has as part of its supplementary material a summary of the *checkcif* report for the structure described in it. This summary includes any validation reply that the author has supplied. It also includes selected numerical data items identified by the journal editors as characterizing the overall quality and completeness of the structure determination.

The characterization of the 'quality' of a structure is a contentious issue. For journals, where there is active selection of

```
The following ALERTS were generated. Each ALERT has the format
test-name ALERT_alert-type_alert-level.
Click on the hyperlinks for more details of the test.
Alert level A
PLAT035_ALERT_1_A No _chemical_absolute_configuration info given .      ?
PLAT761_ALERT_1_A CIF Contains no X-H Bonds .....                      ?
PLAT762_ALERT_1_A CIF Contains no X-Y-H or H-Y-H Angles .....          ?
Alert level C
PLAT125_ALERT_4_C No _symmetry_space_group_name_Hall Given .....      ?
PLAT152_ALERT_1_C Supplied and Calc Volume s.u. Inconsistent .....     ?
PLAT707_ALERT_1_C D...A Calc 3.324(2), Rep 3.321(2), Dev... 1.50 Sigma
C13 -04 1.555 2.556
Alert level G
REFLT03_ALERT_4_G Please check that the estimate of the number of Friedel pairs is
correct. If it is not, please give the correct count in the
_publ_section_exptl_refinement section of the submitted CIF.
From the CIF: _diffrn_reflns_theta_max 27.00
From the CIF: _reflns_number_total 2393
Count of symmetry unique reflns 1395
Completeness (_total/calc) 171.54 TEST3: Check
Friedels for noncentro structure
Estimate of Friedel pairs measured 998
Fraction of Friedel pairs measured 0.715
Are heavy atom types Z>Si present yes
3 ALERT level A = In general: serious problem
0 ALERT level B = Potentially serious problem
3 ALERT level C = Check and explain
1 ALERT level G = General alerts; check

5 ALERT type 1 CIF construction/syntax error, inconsistent or missing data
0 ALERT type 2 Indicator that the structure model may be wrong or deficient
0 ALERT type 3 Indicator that the structure quality may be low
2 ALERT type 4 Improvement, methodology, query or suggestion

(a)
```

#### Publication of your CIF

You should always attempt to resolve as many as possible of the alerts in all categories. Often the minor alerts point to easily fixed oversights, errors and omissions in your CIF or refinement strategy, so attention to these fine details can be worthwhile. In order to resolve some of the more serious problems it may be necessary to carry out additional measurements or structure refinements. However, the nature of your study may justify the reported deviations from the submission requirements of the journal and these should be commented upon in the discussion or experimental section of a paper - after all, they might represent an interesting feature.

If level A alerts remain, which you believe to be justified deviations, and you intend to submit this CIF for publication in *Acta Crystallographica* Section C or Section E, you should additionally insert an explanation in your CIF using the Validation Reply Form (VRF) below. Your explanation will be considered as part of the review process.

If you intend to submit to another section of *Acta Crystallographica* or *Journal of Applied Crystallography* or *Journal of Synchrotron Radiation*, you should make sure that at least a basic structural check is run on the final version of your CIF prior to submission.

```
# start Validation Reply Form
_vrf_PLAT035_99107abs
;
PROBLEM: No _chemical_absolute_configuration info given .      ?
RESPONSE: ...
;
_vrf_PLAT761_99107abs
;
PROBLEM: CIF Contains no X-H Bonds .....                      ?
RESPONSE: ...
;
_vrf_PLAT762_99107abs
;
PROBLEM: CIF Contains no X-Y-H or H-Y-H Angles .....          ?
RESPONSE: ...
;
# end Validation Reply Form
```

If you wish to submit your CIF for publication in *Acta Crystallographica* Section C or E, you should upload your CIF via the web. If your CIF is to form part of a submission to another IUCr journal, you will be asked, either during electronic submission or by the Co-editor handling your paper, to upload your CIF via our web site.

(b)

Fig. 5.7.2.3. Extracts from a *checkcif* report for a 'publication check' on a CIF to be submitted to an IUCr journal. (a) Alerts of various levels of severity are listed. (b) The journal policy on the handling of alerts is summarized and a validation reply form listing the A alerts is supplied for the author to fill in.

articles for publication, it can be difficult to assign criteria for assessing the quality of the structure determination without these being seen as judging the quality or worth of the scientific work giving rise to the result. Thus journals rely upon the experience and discernment of referees to identify structures 'worth' publish-

ing. However, in a comprehensive collection of structural data sets, such as in a public structural database, it might be possible to identify particular data items that could be used for weighting individual data sets when the database is being ‘mined’ for particular patterns or characteristic values. It will be interesting to see whether a consensus emerges on what items would be suitable. It is clear that reliance on a single indicator will not be appropriate for sophisticated studies. The old idea that a structure could be classed as ‘good’ or ‘bad’ on the basis of its final residual  $R$  factor alone has long been abandoned, but it may be possible to stipulate criteria for a set of interrelated data items and use these to filter specific information from a database.

#### 5.7.2.7. Submission and review

When an author has previewed and checked the contents of the CIF and has made the changes suggested by a careful study of the preprint and the *checkcif* report, the article may finally be submitted to *Acta Cryst. C* or *E* by file upload over the web. Other files completing or supporting the submission are also transferred to the editorial office at this time. These include structure-factor or powder profile listings for each structure, figures and chemical diagrams, and sometimes other supplementary documents. Structure-factor listings are supplied in CIF format. Figures may be in one of a number of standard graphics file formats, and at the moment have to be uploaded as separate files. Future extensions to CIF, perhaps following the imgCIF approach, may allow all the items needed to submit an article, including figures, to be prepared as a single file.

When all the files have arrived at the editorial office, a review document is generated that can be sent to the referees. This document contains: the text and tables of the article that will appear in the final publication, but laid out in a more open style suitable for annotation by hand; tables of atomic positions and geometry (containing all the data in the CIF, not just the subset that has been selected for displaying in the published article); certain fields from the CIF that are not normally printed but which may contain details of the way in which the experiment was carried out (these fields might have been completed manually or by the software controlling the experiment); the figures and other supplementary documents; and a print-out of the report from a final *checkcif* cycle, including a displacement-ellipsoid plot of the molecule in a minimal-overlap least-squares plane view. This composite document provides the information that a referee will typically want to consider in a compact and convenient form. Because the CIF is so highly structured, producing this review document is in most cases entirely automatic. The complete CIF as submitted by the author and the experimental data are also made available to the reviewer.

If revisions are requested, authors may upload modified files. The generation of revised versions of an article is also largely automatic.

#### 5.7.2.8. Publication

When the final version of a CIF for *Acta Cryst. C* or *E* is approved, the article is ready for publication. Once more, the data fields required for the published article are extracted from the CIF and sorted. If the author has asked for additional items to be printed by using `_publ_manuscript_incl_extra_item`, these also are extracted. The result is transformed to a file suitable for processing by typesetting software. For *Acta Cryst. C* this was originally a  $\text{\TeX}$  file; now a further transformation generates an SGML file that conforms to the document type definition (DTD) common to all IUCr journals. This allows not only typesetting and printing, but also the generation of the HTML for the navigable online

version of the article, and the extraction of metadata for building online tables of contents and for supplying to bibliographic databases.

The conventional published article then appears in a monthly issue. Each article is still similar in style to the type of structure report published in journals for decades, although tables of atomic positions and geometric data are not usually displayed now, since these data are so readily available from the online article.

The online version of the journal, however, presents a much more information-rich version of the article. Each article is generally available in the form of a PDF file, suitable for downloading and offline printing. There is also an HTML version of the same text, and this version has rich internal links that make it easy to scroll back and forth through the article, jump to specific sections and see figures in low-resolution thumbnail or high-resolution views. The reference list contains links to the articles that are cited. There may also be links to related records in chemical or crystal structure databases. The reader may also download the experimental data and any supplementary documents associated with the article. As mentioned above, for *Acta Cryst. E* a summary of the check report is also available.

Finally, the structural data may be downloaded directly in CIF format. The CIF is presented in two ways. If a reader follows one link in a web browser, the file is interpreted simply as a text file and appears as a simple listing in the browser window, from which it may be printed or saved to disk. However, if the reader follows the other link, the CIF is transmitted to the browser with a header declaring its MIME type (Freed & Borenstein, 1996) as ‘chemical/x-cif’. This is one of several MIME types registered for particular presentations of chemistry-related content by Rzepa & Murray-Rust (1998). The reader may then configure a web browser to respond in a specific way to content tagged with this MIME type; typically a helper application such as a molecular visualizer [*e.g. Mercury* (Bruno *et al.*, 2002)] will be launched that allows three-dimensional visualization and manipulation of the molecular or crystal structure.

When an article has been published in *Acta Cryst. C* or *E*, the CIF is transferred to the relevant public structural databases. Thus, the transcription errors that used to cause so many problems for data harvesters are completely avoided and one of the initial goals of the CIF project is achieved: uncorrupted data transfer from diffractometer, through publication, to a final repository.

Because *Acta Cryst. C* and *E* handle almost exclusively the publication of structure reports, the editorial workflow based on CIF lends itself to a very high level of automation and the journals are produced efficiently and on short timescales. Routine refereeing of structures is made very easy by the provision of checking reports, and the universal use of e-mail and web file transfer means that production times can be very fast.

#### 5.7.3. CIF and other journals

Not every journal will be able to benefit to the same extent from the handling of CIFs. For many journals, structure reports will be secondary to the main purpose of most articles, and CIF data will more usually be deposited as supplementary or supporting documents, while only a summary (if anything) of the structure will be reported in the article body.

Nevertheless, the ability to extract data from CIFs automatically and the ability of much crystallographic software to read CIFs mean that even journals that do not specialize in crystallography can provide a production stream that includes careful checking of crystal structure data. The IUCr continues to develop *checkcif* as

a service which can be used by other publishers to enhance their checking of crystal structures, and there is considerable interest in this approach.

All journals publishing the results of crystal structure determinations may easily collect the supporting data in CIF format and transfer the files to public databases, improving the accuracy and efficiency of the database-building procedures.

### 5.7.3.1. Including CIF data in an article

For journals other than those specializing in full-scale structure reports, including CIF data in tables or reports of structures within general articles is rather more problematic. The translation of CIF data into XML seems to be a promising route to explore, as journals and reference volumes are increasingly being typeset from XML files. Traditionally, publishing has emphasized content markup that leads to a particular typographic representation. Modern trends are towards markup that tags the content by purpose, with the representation directed by external ‘style files’. Consider Fig. 5.7.3.1, which shows the typeset representation of a set of data items in a CIF for a structural paper.

First, it can be seen that several CIF data items are omitted from the printed representation, such as the *International Tables* space-group number and the Hall symbol for the space group. For compactness, the printed data value does not have a legend or annotation if the meaning of an item is clear from the context; thus, the crystal system and Hermann–Mauguin space-group symbol are printed without any accompanying text. The journal may also omit information that is implicit given other data; thus the cell angles are not printed for an orthorhombic cell. On the other hand, units, which are implicit in the definition of a CIF data item, are printed. Related items are grouped together in a single expression, as in the case of the  $\theta$  range or the crystal dimensions. In some cases, numerical values have been rounded to meet the journal’s policy.

All of these transformations are matters of style, but it can be seen that they are not always trivial mappings to single data names. The style files determining the transformation from a detailed explicit data tabulation in the initial CIF may need to implement complex logical tests to suit the requirements of the journal.

Fig. 5.7.3.2 shows the same extract in  $\text{\TeX}$ , the markup and typesetting language that was used for several years to produce *Acta Cryst. C*. It can be seen from this extract that the actual markup maps very closely to the initial CIF. All the cell parameters, including the cell angles, are present in the source file. The expansion of the macros (e.g.  $\backslash\text{cell}\alpha$ ) executes the logic required to determine whether the value is to be printed and generates the additional text surrounding the value. Each data name is mapped to a distinct macro (even if the macros themselves have identical or near-identical internal structure), which preserves the semantic labelling of the original CIF. These macros are maintained in a separate file referenced and executed by every invocation of the typesetting program.

In contrast, Fig. 5.7.3.3 shows part of the SGML now used to typeset *Acta Cryst. C* and to generate HTML versions of the articles online. It is immediately seen that the markup emphasizes typographic style and positioning, and there is no explicit labelling by semantic element. Additional labelling is now found in the document structure; the individual items are marked up as ‘list items’ ( $\langle\text{li}\rangle$ ), but the arrangement of this list into a tabular form is a feature of the typesetting engine, not the SGML.

It is clear that the  $\text{\TeX}$  macros provide a representation of the contents of the CIF that could easily be converted back to the initial

<b>data_I</b>	
<code>_symmetry_cell_setting</code>	orthorhombic
<code>_symmetry_space_group_name_H-M</code>	'P 21 21 21'
<code>_symmetry_space_group_name_Hall</code>	'P 2ac 2ab'
<code>_symmetry_Int_Tables_number</code>	19
<b>loop_</b>	
<code>_symmetry_equiv_pos_as_xyz</code>	
'x, y, z'	
'-x+1/2, -y, z+1/2'	
'-x, y+1/2, -z+1/2'	
'x+1/2, -y+1/2, -z'	
<code>_chemical_formula_moiety</code>	'C10 H8 Br N S'
<code>_diffrn_radiation_type</code>	MoK $\alpha$
<code>_exptl_absorpt_coefficient_mu</code>	4.280
<code>_cell_length_a</code>	5.7339 (7)
<code>_cell_length_b</code>	14.8229 (15)
<code>_cell_length_c</code>	23.469 (2)
<code>_cell_angle_alpha</code>	90.0
<code>_cell_angle_beta</code>	90.0
<code>_cell_angle_gamma</code>	90.0
<code>_cell_volume</code>	1994.7 (4)
<code>_cell_formula_units_Z</code>	8
<code>_cell_measurement_temperature</code>	296 (2)
<code>_cell_measurement_reflns_used</code>	58
<code>_cell_measurement_theta_min</code>	4.806
<code>_cell_measurement_theta_max</code>	11.635
<code>_exptl_crystal_description</code>	plate
<code>_exptl_crystal_colour</code>	colourless
<code>_exptl_crystal_size_max</code>	0.40
<code>_exptl_crystal_size_mid</code>	0.32
<code>_exptl_crystal_size_min</code>	0.04
<code>_exptl_crystal_density_meas</code>	?
<code>_exptl_crystal_density_diffrn</code>	1.693
<code>_exptl_crystal_density_method</code>	'not measured'
<code>_exptl_crystal_F_000</code>	1008
	(a)
<b>Crystal data</b>	
C <sub>10</sub> H <sub>8</sub> BrNS	Mo K $\alpha$ radiation
$M_r = 254.14$	Cell parameters from 58
Orthorhombic, $P2_12_12_1$	reflections
$a = 5.7339 (7) \text{ \AA}$	$\theta = 4.8\text{--}11.6^\circ$
$b = 14.8229 (15) \text{ \AA}$	$\mu = 4.28 \text{ mm}^{-1}$
$c = 23.469 (2) \text{ \AA}$	$T = 296 (2) \text{ K}$
$V = 1994.7 (4) \text{ \AA}^3$	Plate, colourless
$Z = 8$	$0.40 \times 0.32 \times 0.04 \text{ mm}$
$D_x = 1.693 \text{ Mg m}^{-3}$	
	(b)

Fig. 5.7.3.1. Typesetting of structural data. The contents of the CIF (a) are transformed into a typeset representation (b) that omits, annotates or reorders the incoming data according to context and the style rules of the journal.

input CIF. At present, such bidirectional translation is not possible from the SGML file.

Clearly, therefore, a mapping to SGML that preserved semantic markup would be preferable. It is most likely that suitable bidirectional translations would be based on XML.

### 5.7.3.2. CIF and XML

XML is a specific concrete implementation of SGML suitable for generation of online browsable content. Mature style transformation mechanisms for XML exist and others are under active development.

```

\structno
\noindent{\nineit Crystal data}\nobreak\par
\vskip2pt\begindoublecolumns\twocoltrue\defaultfont
\raggedright
\everypar={\global\parindent=0pt\hangindent=1em
\hangafter=1 }\noindent

\chemformiupac{C$_{10}$H$_{8}$BrNS}
\chemform{C$_{10}$H$_{8}$BrNS}
\chemformsum{C$_{10}$H$_{8}$BrNS}
\molwt{254.14}
\system{Orthorhombic}\def\sgsetno{2}
\defaultfont
\sgHM{P2}_{1}2_{1}2_{1}2_{1}
\cella{5.7339 (7)}
\cellb{14.8229 (15)}
\cellc{23.469 (2)}
\cellalpha{90.00}
\cellbeta{90.00}
\cellgamma{90.00}
\cellvol{1994.7 (4)}
\cellz{8}
\dx{1.693}
\dm{missing}
\densmetha{not measured}
\radiationtype{Mo {it K}\alpha}
\wavelength{0.71073}
\cellrefl{58}
\cellthetamin{4.8}
\cellthetamax{11.6}
\absorpmu{4.28}
\celltemp{296 (2)}
\shape{Plate}
\colour{Colourless}
\sizemax{0.40}
\sizemid{0.32}
\sizemin{0.04}
\sizead{missing}
\origin{see text}

```

Fig. 5.7.3.2. Part of a T<sub>E</sub>X file used to print the article shown in Fig. 5.7.3.1(b).

Section 5.3.8.2.1 describes one transformation to XML in the biological structures field, designed primarily for database interchange rather than publication. This transformation preserves the underlying data model of an mmCIF very closely, and one might anticipate similar XML transformations for small-molecule CIF applications and for publications. It is even possible that the XML transformations referred to in Chapter 5.3 could be used for publishing articles if suitable style transformations are developed, but this has not been tested yet.

One difficulty with a simple CIF-to-XML transformation is that it could be easily adapted to the publication of structure reports in dedicated journals, but would not necessarily be compatible with other XML implementations developed by an unspecialized publishing house. This could be avoided by the registration of an XML name space covering transformed CIF data and the production of portable stylesheet transformations that could be adopted and modified to meet the requirements of different publishing houses. As yet, we know of no initiatives in this direction.

XML name spaces have been registered to safeguard the development of subject-specific methods of representation as part of a project by the International Union of Pure and Applied Chemistry (Becker, 2001). One markup language that falls within the scope of this project is Chemical Markup Language (CML) (Murray-Rust & Rzepa, 1999, 2001).

Further discussions of the relationship between CIF and XML representations and a proposal for extensions to certain CIF data values to accommodate the wider range of data structures permitted in XML are given by Bernstein (2000).

```

<sec id="sec2.1">
  <sec id="sec2.1.1">
    <st><?print tpct=0pt>Crystal data</st>
    <p>
      <l id="l1" type="unord">
        <li><p>C<inf arrange="stagger">10</inf>H<inf
          arrange="stagger">8</inf>BrNS</p></li>
        <li><p><it>M</it><inf
          arrange="stagger"><it>r</it></inf> =
          254.14</p></li>
        <li><p>Orthorhombic, &nbsp;
          <fi type="tex" print-info="tth" img.type="tth"
          img.data="teximages/ga1014fi3.tth">
          P2_{1}2_{1}2_{1}</fi></p></li>
        <li><p><it>a</it> =
          5.7339&emsp14; (7) &emsp14; &Aring;</p></li>
        <li><p><it>b</it> =
          14.8229&emsp14; (15) &emsp14; &Aring;</p></li>
        <li><p><it>c</it> =
          23.469&emsp14; (2) &emsp14; &Aring;</p></li>
        <li><p><it>V</it> =
          1994.7&emsp14; (4) &emsp14; &Aring;<sup
          arrange="stagger">3</sup></p></li>
        <li><p><it>Z</it> =
          8</p></li>
        <li><p><it>D</it><inf
          arrange="stagger"><it>x</it></inf> =
          1.693&emsp14; Mg&emsp14; m<sup
          arrange="stagger">&minus;3</sup></p></li>
        <li><p>Mo <it>K</it>&alpha; radiation</p></li>
        <li><p>Cell parameters from 58 <?print show
          &softreturn;>reflections</p></li>
        <li><p>&thetas; = 4.8&ndash;11.6&deg;</p></li>
        <li><p>&mu; = 4.28&emsp14; mm<sup
          arrange="stagger">&minus;1</sup></p></li>
        <li><p><it>T</it> =
          296&emsp14; (2) &emsp14; K</p></li>
        <li><p>Plate, colourless</p></li>
        <li><p>0.40 &times; 0.32 &times;
          0.04&emsp14; mm</p></li>
      </l>
    </p>
  </sec>

```

Fig. 5.7.3.3. Part of the SGML file used to print the article shown in Fig. 5.7.3.1(b).

We acknowledge the guidance, enthusiasm and dedication of past and present members of the editorial boards of *Acta Crystallographica Sections C* and *E* in developing the journals along the path described in this chapter. Particular tribute must be paid to Syd Hall, George Ferguson, Bill Clegg, David Watson and Tony Linden. We are very grateful to Ton Spek for his close involvement with the development of checking software, and also wish to acknowledge George Sheldrick, Mario Nardelli, Eric Gabe, Peter White, Yvon Le Page, Alan Mighell, Vicky Karen, Doug du Boulay, Mike Dacombe and Charlie Bugg for their help in the early days of automated structure checking. We wish also to pay tribute to the dedication and effort of our colleagues in the IUCr editorial office: Gillian Holmes, Sean Conway, Amanda Berry, Sarah Froggatt and Lisa Stephenson; and we thank the many authors who have been willing to test new approaches through the years.

### Appendix 5.7.1

#### Request list for *Acta Crystallographica Section C*

Table A5.7.1.1 contains the request list for *Acta Crystallographica Section C* as given in the 2005 *Notes for Authors*. This list is appropriate for a single-crystal X-ray diffraction study and gives all the data items that are displayed in an article *if they are present in the CIF*. In principle, a smaller set of mandatory data items could be supplied as a separate request list. However, certain items may be

## 5.7. SMALL-MOLECULE CRYSTAL STRUCTURE PUBLICATION USING CIF

considered mandatory or not depending on the nature of the study and on the presence of other data items in the CIF, so checking for mandatory items is performed through higher-level algorithmic checks during the pre-submission validation stage.

Table A5.7.1.1. Request list for Acta Crystallographica Section C

(a) Data names relating to the text of an article	
<code>_publ_contact_author_name</code>	Contact author's name
<code>_publ_contact_author_address</code>	Contact author's address
<code>_publ_contact_author_email</code>	E-mail address to be published
<code>_publ_contact_author_fax</code>	For editorial communications
<code>_publ_contact_author_phone</code>	For editorial communications
<code>_publ_contact_letter</code>	Letter of submission, with date
<code>_publ_requested_journal</code>	'Acta Crystallographica Section C'
<code>_publ_requested_category</code>	Publication choice (FI, FM, FO, AD)
<code>_publ_section_title</code>	Title of paper
<code>_publ_section_title_footnote</code>	Footnote to title of paper
<code>_publ_author_name</code>	List of author(s) name(s)
<code>_publ_author_footnote</code>	Footnote(s) to author(s) name(s)
<code>_publ_author_address</code>	Author(s) address(es)
<code>_publ_section_synopsis</code>	Synopsis for compounds that cannot be shown as a chemical diagram
<code>_publ_section_abstract</code>	Abstract of paper in English
<code>_publ_section_comment</code>	Discussion of study
<code>_publ_section_acknowledgements</code>	Acknowledgements
<code>_publ_section_references</code>	References
<code>_publ_section_figure_captions</code>	Legends to figures
(b) Data names relating to the experimental data	
<code>_publ_section_exptl_prep</code>	Compound preparation details
<code>_chemical_formula_sum</code>	Chemical formula as sum of elements
<code>_chemical_formula_moiety</code>	Chemical formula in moieties
<code>_chemical_formula_weight</code>	Chemical formula mass (Da)
<code>_chemical_melting_point</code>	Melting point (K)
<code>_symmetry_cell_setting</code>	Code for cell setting
<code>_symmetry_space_group_name_H-M</code>	Space-group symbol, including unique axis
<code>_symmetry_equiv_pos_as_xyz</code>	Equivalent positions in order used by <code>_geom</code>
<code>_cell_length_a</code>	Unit-cell lengths (Å)
<code>_cell_length_b</code>	
<code>_cell_length_c</code>	
<code>_cell_angle_alpha</code>	Unit-cell angles (°)
<code>_cell_angle_beta</code>	
<code>_cell_angle_gamma</code>	
<code>_cell_volume</code>	Unit-cell volume (Å <sup>3</sup> )
<code>_cell_formula_units_Z</code>	Number of formulae per unit cell
<code>_exptl_crystal_density_diffn</code>	Density calculated from unit cell and contents (Mg m <sup>-3</sup> )
<code>_exptl_crystal_density_meas</code>	Density measured experimentally (Mg m <sup>-3</sup> )
<code>_exptl_crystal_density_method</code>	Method used to measure density experimentally
<code>_diffn_radiation_type</code>	Radiation type (e.g. neutron or Mo Kα)
<code>_diffn_radiation_wavelength</code>	Radiation wavelength (Å)
<code>_cell_measurement_reflns_used</code>	Number of reflections used to measure unit cell
<code>_cell_measurement_theta_min</code>	Minimum $\theta$ of reflections used to measure unit cell (°)
<code>_cell_measurement_theta_max</code>	Maximum $\theta$ of reflections used to measure unit cell (°)
<code>_cell_measurement_temperature</code>	Measurement temperature (K)
<code>_exptl_absorpt_coefficient_mu</code>	Linear absorption coefficient (mm <sup>-1</sup> )
<code>_exptl_crystal_description</code>	Crystal habit description
<code>_exptl_crystal_size_max</code>	Maximum dimension of crystal (mm)
<code>_exptl_crystal_size_mid</code>	Medial dimension of crystal (mm)
<code>_exptl_crystal_size_min</code>	Minimum dimension of crystal (mm)
<code>_exptl_crystal_size_rad</code>	Radius of spherical or cylindrical crystal (mm)
<code>_exptl_crystal_colour</code>	Crystal colour
<code>_diffn_measurement_device_type</code>	Diffraction make and type
<code>_diffn_measurement_method</code>	Mode of intensity measurement and scan
<code>_diffn_detector_area_resol_mean</code>	Resolution of area detector (pixels mm <sup>-1</sup> )
<code>_exptl_absorpt_correction_type</code>	Code for absorption correction
<code>_exptl_absorpt_process_details</code>	Literature reference for absorption correction [e.g. '(North et al., 1968)']
<code>_exptl_absorpt_correction_T_min</code>	Minimum transmission factor from corrections
<code>_exptl_absorpt_correction_T_max</code>	Maximum transmission factor from corrections
<code>_diffn_reflns_number</code>	Total number of reflections measured
<code>_reflns_number_total</code>	Number of symmetry-independent reflections
<code>_reflns_number_gt</code>	Number of reflections > $\sigma$ threshold
<code>_reflns_threshold_expression</code>	$\sigma$ expression for $F$ , $F^2$ or $I$ threshold
<code>_diffn_reflns_theta_max</code>	Maximum $\theta$ of measured reflections (°)
<code>_diffn_reflns_theta_full</code>	$\theta$ to which available reflections are 'complete' (°)
<code>_diffn_measured_fraction_theta_max</code>	Fraction of unique reflections measured to $\theta_{\max}$

Table A5.7.1.1 (cont.).

<code>_diffn_measured_fraction_theta_full</code>	Fraction of unique reflections measured to $\theta_{\text{full}}$
<code>_diffn_reflns_av_R_equivalents</code>	$R$ factor for symmetry-equivalent intensities
<code>_diffn_reflns_limit_h_min</code>	Minimum/maximum $h$ index of measured data
<code>_diffn_reflns_limit_h_max</code>	
<code>_diffn_reflns_limit_k_min</code>	Minimum/maximum $k$ index of measured data
<code>_diffn_reflns_limit_k_max</code>	
<code>_diffn_reflns_limit_l_min</code>	Minimum/maximum $l$ index of measured data
<code>_diffn_reflns_limit_l_max</code>	
<code>_diffn_standards_number</code>	Number of standards used in measurement
<code>_diffn_standards_interval_count</code>	Number of measurements between standards
<code>_diffn_standards_interval_time *</code>	Time (min) between standards
<code>_diffn_standards_decay_%</code>	Percentage decrease in standards intensity
<code>_refine_ls_structure_factor_coef</code>	Code for $F$ , $F^2$ or $I$ used in least-squares refinement
<code>_refine_ls_R_factor_gt</code>	$R$ factor of $F$ for reflections > threshold
<code>_refine_ls_wR_factor_ref</code>	$R$ factor of coefficient for refinement reflections
<code>_refine_ls_goodness_of_fit_ref</code>	Goodness of fit $S$ for refinement reflections
<code>_refine_ls_number_reflns</code>	Number of reflections used in refinement
<code>_refine_ls_number_parameters</code>	Number of parameters refined
<code>_refine_ls_weighting_scheme</code>	Code for weight type
<code>_refine_ls_weighting_details</code>	Weighting expression
<code>_refine_ls_hydrogen_treatment</code>	Code for H-atom treatment
<code>_refine_ls_shift/su_max</code>	Maximum shift/s.u. ratio after final refinement cycle
<code>_refine_diff_density_max</code>	Maximum/minimum values of final difference map (e Å <sup>-3</sup> )
<code>_refine_diff_density_min</code>	
<code>_refine_ls_extinction_method</code>	Description of extinction methods applied
<code>_refine_ls_extinction_coef</code>	Extinction coefficient applied in corrections
<code>_refine_ls_abs_structure_details</code>	Absolute structure method and Friedel-pair number
<code>_refine_ls_abs_structure_Flack</code>	Measure of absolute structure
<code>_refine_ls_abs_structure_Rogers†</code>	Measure of absolute structure
<code>_publ_section_exptl_refinement</code>	Special details of the refinement
<code>_computing_data_collection</code>	Reference to data-collection software
<code>_computing_cell_refinement</code>	Reference to cell-refinement software
<code>_computing_data_reduction</code>	Reference to data-reduction software
<code>_computing_structure_solution</code>	Reference to structure-solution software
<code>_computing_structure_refinement</code>	Reference to structure-refinement software
<code>_computing_molecular_graphics</code>	Reference to visualization software
<code>_computing_publication_material</code>	Reference to publication preparation software
<code>loop_</code>	
<code>_atom_type_symbol</code>	Atom type symbol (usually element symbol)
<code>_atom_type_description</code>	Description of atom type
<code>_atom_type_scatter_source</code>	Reference to scattering factors applied
<code>_atom_type_scatter_dispersion_real</code>	Real anomalous-dispersion value applied
<code>_atom_type_scatter_dispersion_imag</code>	Imaginary anomalous-dispersion value applied
<code>loop_</code>	
<code>_atom_site_label</code>	Unique label identifying the atom site
<code>_atom_site_fract_x</code>	Fractional coordinates of atom site
<code>_atom_site_fract_y</code>	
<code>_atom_site_fract_z</code>	
<code>_atom_site_U_iso_or_equiv</code>	Isotropic atomic displacement parameter, or equivalent from anisotropic atomic displacement parameters
<code>_atom_site_occupancy</code>	Occupancy fraction for site (default is 1.0)
<code>_atom_site_disorder_assembly</code>	Code that identifies functional group suffering disorder
<code>_atom_site_disorder_group</code>	Code that identifies disorder group
<code>_atom_site_adp_type</code>	Atomic displacement parameter type
<code>loop_</code>	
<code>_geom_site_aniso_label</code>	Unique label identifying the atom site
<code>_geom_site_aniso_U_11</code>	Elements of anisotropic atomic displacement parameter tensor
<code>_geom_site_aniso_U_22</code>	
<code>_geom_site_aniso_U_33</code>	
<code>_geom_site_aniso_U_12</code>	
<code>_geom_site_aniso_U_13</code>	
<code>_geom_site_aniso_U_23</code>	
<code>loop_</code>	
<code>_geom_bond_atom_site_label_1</code>	Labels identifying the atom sites 1 and 2
<code>_geom_bond_atom_site_label_2</code>	
<code>_geom_bond_site_symmetry_1</code>	Symmetry codes (e.g. 2,554) for atom sites 1 and 2
<code>_geom_bond_site_symmetry_2</code>	
<code>_geom_bond_distance</code>	Distance between atom sites 1 and 2 (Å)
<code>_geom_bond_publ_flag</code>	Flag for print request (yes or no)

\* Alternative to `_diffn_standards_interval_count`. † Alternative to `_refine_ls_abs_structure_Flack`.

Table A5.7.1.1 (cont.).

loop_	
_geom_angle_atom_site_label_1	Labels identifying the atom sites 1, 2 and 3
_geom_angle_atom_site_label_2	
_geom_angle_atom_site_label_3	
_geom_angle_site_symmetry_1	Symmetry codes for atom sites 1, 2 and 3
_geom_angle_site_symmetry_2	
_geom_angle_site_symmetry_3	
_geom_angle	Angle between atom sites 1, 2 and 3 (°)
_geom_angle_publ_flag	Flag for print request (yes or no)
loop_	
_geom_torsion_atom_site_label_1	Labels identifying the atom sites 1, 2, 3 and 4
_geom_torsion_atom_site_label_2	
_geom_torsion_atom_site_label_3	
_geom_torsion_atom_site_label_4	
_geom_torsion_site_symmetry_1	Symmetry codes for atom sites 1, 2, 3 and 4
_geom_torsion_site_symmetry_2	
_geom_torsion_site_symmetry_3	
_geom_torsion_site_symmetry_4	
_geom_torsion	Torsion angle between atom sites 1, 2, 3 and 4 (°)
_geom_torsion_publ_flag	Flag for print request (yes or no)
loop_	
_geom_hbond_atom_site_label_D	Donor-atom label in hydrogen bond
_geom_hbond_atom_site_label_H	H-atom label in hydrogen bond
_geom_hbond_atom_site_label_A	Acceptor-atom label in hydrogen bond
_geom_hbond_site_symmetry_D	Symmetry code for donor site
_geom_hbond_site_symmetry_H	Symmetry code for H-atom site
_geom_hbond_site_symmetry_A	Symmetry code for acceptor site
_geom_hbond_distance_DH	Donor atom-to-H-atom distance (Å)
_geom_hbond_distance_HA	H-atom-to-acceptor atom distance (Å)
_geom_hbond_distance_DA	Donor atom-to-acceptor atom distance (Å)
_geom_hbond_angle_DHA	Donor to H to acceptor angle (°)
_geom_hbond_publ_flag	Flag for print request (yes or no)

(c) Data names for adding items to the standard request list

loop_	
_publ_manuscript_incl_extra_item	Additional CIF item submitted for publication
_publ_manuscript_incl_extra_defn	Is item defined in core dictionary? (yes or no)

(d) Data names for structure-factor lists

loop_	
_refln_index_h	Miller indices <i>h</i> , <i>k</i> and <i>l</i>
_refln_index_k	
_refln_index_l	
_refln_F_meas	Measured <i>F</i>
_refln_F_squared_meas‡	Measured <i>F</i> <sup>2</sup>
_refln_F_sigma	Standard uncertainty of <i>F</i>
_refln_F_squared_sigma‡	Standard uncertainty of <i>F</i> <sup>2</sup>
_refln_F_calc	Calculated <i>F</i>
_refln_F_squared_calc‡	Calculated <i>F</i> <sup>2</sup>

‡ Alternative to the corresponding data name without 'squared'.

## Appendix 5.7.2

### Data validation using *checkcif*

Table A5.7.2.1 lists the *checkcif* tests concerned primarily with the completeness and self-consistency of individual or closely related data items. These tests were developed from the routines of *PREPUB* (du Boulay & Hall, 1996) and in the IUCr Editorial Office. Table A5.7.2.2 lists the tests applied specifically by the program *PLATON* (Spek, 2003), which performs a more detailed crystallographic analysis of the structure itself.

Each entry in each table has an identifying code and a numeric type. The type is used to categorize the alert messages generated when the tested values deviate from assigned norms. Type 1 refers to syntactic or other errors of construction in the CIF, or to inconsistent or missing data. Type 2 alerts indicate that the structure model may be wrong or deficient. Type 3 alerts indicate that the quality of the structure may be low, owing to limited or incomplete data coverage. Alerts of type 4 are indicative of deviations from style or suggested good practice, or may offer suggestions for improvement in presentation. The alerts within each category may be of varying levels of severity.

Full details of the tests and algorithms applied for the *checkcif* tests may be found at <http://journals.iucr.org/services/cif/>

datavalidation.html or on the CD-ROM accompanying this volume. These include comments which provide help in interpreting the results of the tests and suggest ways in which the author can improve the data. The comments were provided by A. Linden and other members of the IUCr journal editorial boards.

The tests listed in Tables A5.7.2.1 and A5.7.2.2 are appropriate for small-unit-cell single-crystal structure determinations. More discriminating tests are being introduced for powder diffraction studies and for modulated structures.

Table A5.7.2.1. List of data-validation tests applied by *checkcif*

Test name	Type	Purpose
ABSMU01	1	Check that $\mu$ is consistent with the cell contents
ABSTM01	1	Check that $T_{\min}$ is less than $T_{\max}$
ABSTM02	3	Check that $T_{\min}$ and $T_{\max}$ are appropriate to the crystal size and $\mu$
ABSTY01	1	Check that <code>_expt1_absorpt_correction_type</code> is a recognized keyword
ABSTY02	1	Check that <code>_expt1_absorpt_correction_type</code> contains some reference text
CELLK01	1	Check that temperature is in Kelvin
CELLT01	1	Check that $\theta_{\min}$ is less than $\theta_{\max}$
CELLV01	1	Check that the <code>_cell_volume</code> matches <code>_cell_length</code> and <code>_cell_angle</code> values
CELLV02	1	Check that the <code>_cell_volume</code> s.u. matches <code>_cell_length</code> and <code>_cell_angle</code> s.u. values
CELLZ01	1	Check consistency between formula, Z, atom list and symmetry
CHEMS01	1	Check that the <code>_chemical_formula_sum</code> is properly constructed
CHEMS02	1	Check that the stated category is consistent with the formula of the compound
CHEMW01	1	Check consistency between <code>_chemical_formula_weight</code> and <code>_chemical_formula_sum</code>
CHEMW03	2	Check consistency between weight, Z, symmetry and atom list
CRYSC01	1	Check that colour of crystal is consistent with expected colour code combinations
CRYSR01	1	Check that the radius of the crystal is given for a spherical or cylindrical crystal
CRYSS01	1	Check consistency of crystal dimensions
CRYSS02	3	Check that the values of <code>_expt1_crystal_size_*</code> are not larger than expected
DENSD01	1	Check consistency of density, cell volume and weight
DENSX01	1	Check that <code>_expt1_crystal_density_meas</code> matches <code>_expt1_crystal_density_diffn</code>
DIFMN01	1	Check that <code>_refine_diff_density_min</code> is less than <code>_refine_diff_density_max</code>
DIFMN02	2	Check that <code>_refine_diff_density_min</code> is within expected limits
DIFMN03	1	Check for adjacent site information if <code>_refine_diff_density_min</code> is outside expected limits
DIFMX01	2	Check that <code>_refine_diff_density_max</code> is within expected limits
DIFMX02	1	Check for adjacent site information if <code>_refine_diff_density_max</code> is outside expected limits
FCOEF01	1	Check that the value of <code>_refine_ls_structure_factor_coef</code> is recognized
FORMU01	1/2	Check consistency between formulae and atom site data
GOODF01	1/2	Check that <code>_refine_ls_goodness_of_fit_ref</code> is within expected limits
HYDTR01	1	Check that the value of <code>_refine_ls_hydrogen_treatment</code> is recognized
RADNT01	1	Check that the radiation type is recognized
RADNW01	1	Check that <code>_diffrn_radiation_wavelength</code> matches <code>_diffrn_radiation_type</code>
REFLE01	3	Check that <code>_reflns_threshold_expression</code> contains a multiplier which is below the limit
REFLG01	1	Check that <code>_reflns_number_gt</code> is less than or equal to <code>_diffrn_reflns_number</code>
REFLL01	1	Check that <code>_diffrn_reflns_limit</code> values are in the correct order
REFLT01	1	Check that <code>_reflns_number_total</code> is less than or equal to <code>_diffrn_reflns_number</code>
REFLT02	1	Check that <code>_reflns_number_total</code> is greater than or equal to <code>_reflns_number_gt</code>
REFLT03	1/3/4	Check consistency of <code>_reflns_number_total</code> with cell volume, symmetry and $\theta_{\max}$
REFNR01	3	Check the ratio of <code>_refine_ls_number_reflns</code> and <code>_refine_ls_number_parameters</code>
RFACG01	3	Check that <code>_refine_ls_R_factor_gt</code> is within expected limits
RFACR01	3	Check that <code>_refine_ls_wR_factor_ref</code> is within expected limits
RINTA01	3	Check that <code>_diffrn_reflns_av_R_equivalents</code> is within expected limits
SHFSU01	2	Check that <code>_refine_ls_shift/su_max</code> is within expected limits
STRDE01	1	Check that <code>_refine_ls_abs_structure_details</code> is present if necessary
STRVA01	2/4	Check that <code>_refine_ls_abs_structure_Flack</code> is within expected limits
STRVA02	2/3/4	Check that <code>_refine_ls_abs_structure_Rogers</code> is within expected limits
SYMMG01	1	Check that the <code>_symmetry_space_group_name_H-M</code> value is recognized

## 5.7. SMALL-MOLECULE CRYSTAL STRUCTURE PUBLICATION USING CIF

Table A5.7.2.1 (cont.).

SYMMG02	1	Check consistency between space group name and symmetry positions
SYMMS01	1	Check that the <code>_symmetry_cell_setting</code> matches one of the keywords
SYMMS02	1	Check consistency between cell setting and cell parameters
THETM01	3	Check that <code>_diffrn_reflms_theta_max</code> is greater than expected limits
WEIGH01	1	Check that the value of <code>_refine_ls_weighting_scheme</code> is recognized

Table A5.7.2.2. List of data-validation tests applied by PLATON

Test name	Type	Purpose
PLAT020	3	Check $R(\text{int})$
PLAT021	1	Check expected number of reflections (max = 1 centro, 2 non-centro)
PLAT022	3	Check expected number of reflections
PLAT023	3	Check $\theta_{\text{max}}$
PLAT024	4	Check for required Friedel pair averaging $Z \leq \text{Si}$
PLAT025	1	Check for $h_{\text{min}} \dots l_{\text{max}}$
PLAT026	3	Check for weak data
PLAT027	3	Check <code>_diffrn_reflms_theta_full</code>
PLAT028	3	Check <code>_diffrn_measured_fraction_theta_max</code>
PLAT029	3	Check <code>_diffrn_measured_fraction_theta_full</code>
PLAT030	1	Check <code>_diffrn_reflms_number &gt; _reflns_number_total</code>
PLAT031	4	Check need for extinction correction parameter
PLAT032	4	Check s.u. Flack parameter
PLAT033	2	Check Flack parameter value
PLAT034	1	Check for Flack parameter value specified $Z > \text{Si}$ , non-centro
PLAT035	1	Check for <code>_chemical_absolute_configuration</code>
PLAT036	1	Check for missing Flack parameter s.u.
PLAT037	1	Check <code>_diffrn_reflms_theta_full</code>
PLAT038	1	Check <code>_diffrn_measured_fraction_theta_max</code>
PLAT039	1	Check <code>_diffrn_measured_fraction_theta_full</code>
PLAT040	1	Test for H atoms [0, 1]
PLAT041	1	Test sum formula
PLAT042	1	Test moiety formula
PLAT043	1	Test for molecular weight
PLAT044	1	Check reported with calculated density
PLAT045	1	Check reported and calculated Z
PLAT046	1	Check reported density with calculated density from $Z * MW$
PLAT047	1	Test sum formula given
PLAT048	1	Test moiety formula given
PLAT049	1	Check calculated density $> 1.0$
PLAT050	1	Test for $\mu$ given [0, 1]
PLAT051	1	Test for difference $\mu(\text{cif})$ with $\mu(\text{calc})$ [%]
PLAT052	1	Test for specification of absorption correction method [0, 1]
PLAT053	1	Test for specification crystal dimension min [0, 1]
PLAT054	1	Test for specification crystal dimension mid [0, 1]
PLAT055	1	Test for specification crystal dimension max [0, 1]
PLAT056	1	Test for specification crystal radius [0, 1]
PLAT057	3	Test for correction for absorption needed
PLAT058	1	Test for specification $T_{\text{max}}$ [0, 1]
PLAT059	1	Test for specification $T_{\text{min}}$ [0, 1]
PLAT060	3	$RR$ test
PLAT061	3	$RR'$ test
PLAT062	4	Rescale $T_{\text{min}}$ and $T_{\text{max}}$
PLAT063	3	Test for crystal size
PLAT064	1	Test for $T_{\text{max}} \geq T_{\text{min}}$
PLAT065	3	Test for applicability of (semi-)empirical absorption correction [0, 1]
PLAT066	1	Test whether predicted and reported transmission ranges are identical
PLAT067	1	Ensure that minimum dimension $<$ max dimension
PLAT068	1	Test for $F(000)$ calc/reported difference
PLAT070	1	Test for duplicate labels
PLAT071	1	Test for uninterpretable labels
PLAT074	1	Test for occupancy = 0.0
PLAT075	1	Test for occupancy greater than 1.0
PLAT076	1	Test for occupancy less than 1.0 for atom on special position
PLAT077	4	Test for non-integral number of atoms in unit cell
PLAT080	2	Test maximum shift/error
PLAT081	1	Test for maximum shift/error given
PLAT082	2	Test for reasonable $R1$
PLAT083	2	Test for extreme second weighting parameter ( <i>SHELXL</i> )
PLAT084	2	Test for reasonable $wR2$
PLAT085	2	Test for default <i>SHELXL</i> weighting scheme
PLAT086	2	Test for reasonable $S$ (too low)
PLAT087	2	Test for reasonable $S$ (too high)
PLAT088	3	Test for reasonable data/parameter ratio (centro)
PLAT089	3	Test for reasonable data/parameter ratio (non-centro) ( $Z_{\text{max}} < 18$ )
PLAT090	3	Test for reasonable data/parameter ratio (non-centro) ( $Z_{\text{max}} > 18$ )
PLAT091	1	Test for 'No wavelength given'
PLAT092	4	Test for wavelength type [Cu, Mo, Ag] [0, 1]
PLAT093	1	Test for inconsistency 'mixed' versus 'no refined H'
PLAT094	2	Test for maximum/minimum residual density ratio
PLAT095	1	Test for residual density maximum given [0, 1]
PLAT096	1	Test for residual density minimum given [0, 1]
PLAT097	2	Test maximum residual density
PLAT098	2	Test for minimum residual density

Table A5.7.2.2 (cont.).

Test name	Type	Purpose
PLAT099	1	Test for minimum residual density less than zero [0, 1]
PLAT110	2	Test for additional translational symmetry [0, 1]
PLAT111	2	Test for additional centre of symmetry [0, 100]
PLAT112	2	Test for additional symmetry [0, 1]
PLAT113	2	Report new space group suggested by <i>ADDSYM</i>
PLAT114	2	Report on <i>ADDSYM</i> problem
PLAT120	1	Test for consistent <code>_symmetry_space_group_name_H-M</code> and symmetry operations
PLAT121	1	Test for valid <code>_symmetry_space_group_name_H-M</code>
PLAT122	1	Test for ? <code>_symmetry_space_group_name_H-M</code>
PLAT123	1	Test for interpretable space-group symmetry
PLAT124	1	Test for <code>_symmetry_equiv_pos_as_xyz</code> present
PLAT125	4	Test for ? <code>_symmetry_space_group_name_Hall</code>
PLAT126	1	Test for <code>_symmetry_space_group_name_Hall</code> error
PLAT127	1	Test for <code>_symmetry_space_group_name_Hall</code> consistency
PLAT128	4	Test for non-standard monoclinic space-group setting
PLAT129	4	Test for unusual non-standard space-group name
PLAT130	1	Test for cubic: $a = b = c$
PLAT131	1	Test for cubic: $\alpha = \beta = \gamma = 90$
PLAT132	1	Test for trigonal/hexagonal: $a = b$
PLAT133	1	Test for trigonal/hexagonal: $\alpha = \beta = 90$
PLAT134	1	Test for trigonal/hexagonal: $\gamma = 120$
PLAT135	1	Test for tetragonal: $a = b$
PLAT136	1	Test for tetragonal: $\alpha = \beta = \gamma = 90$
PLAT137	1	Test for orthorhombic: $\alpha = \beta = \gamma = 90$
PLAT138	1	Test for monoclinic more than 1 angle off 90 degrees
PLAT139	1	Test for rhombohedral $a = b = c$
PLAT140	1	Test for rhombohedral $\alpha = \beta = \gamma$
PLAT141	4	S.u. on $a$ axis small or missing
PLAT142	4	S.u. on $b$ axis small or missing
PLAT143	4	S.u. on $c$ axis small or missing
PLAT144	4	S.u. on $\alpha$ small or missing
PLAT145	4	S.u. on $\beta$ small or missing
PLAT146	4	S.u. on $\gamma$ small or missing
PLAT147	1	S.u. on symmetry restricted cell angle
PLAT150	1	Check volume
PLAT151	1	Check for s.u. on volume
PLAT152	1	Check for consistency of s.u. on volume and cell parameters
PLAT155	4	Check for reduced cell $aP$
PLAT156	4	Check for non-standard axial order
PLAT157	4	Check for non-standard monoclinic $\beta$ angle $< 90$ degrees
PLAT161	4	Missing $x$ -coordinate s.u.
PLAT162	4	Missing $y$ -coordinate s.u.
PLAT163	4	Missing $z$ -coordinate s.u.
PLAT164	4	Check for refined C—H atoms
PLAT165	3	Check for $R$ -flagged non-H atoms
PLAT166	4	Check for calc flagged atoms with s.u.'s on coordinates
PLAT170	4	Check for sufficient data in atom data loop
PLAT199	1	Test for <i>SHELXL</i> room-temperature default (cell)
PLAT200	1	Test for <i>SHELXL</i> room-temperature default (data collection)
PLAT201	2	Test for isotropic non-H atoms in main residue(s)
PLAT202	3	Test for isotropic non-H atoms in anion? or solvent?
PLAT210	3	Test for all-isotropic a.d.p.(s)
PLAT211	2	Test for NPD a.d.p.'s (1.0) in main residue(s)
PLAT212	2	Test for NPD a.d.p.'s in anion? & solvent? [0, 1]
PLAT213	2	Test ratio a.d.p. max/min in main residue(s)
PLAT214	2	Test ratio a.d.p. max/min in anion? or solvent?
PLAT215	3	Test for unusual disordered atom a.d.p. in main residue
PLAT216	3	Test for unusual disordered atom a.d.p. in minor residue
PLAT217	1	Test for incomplete $U_{ij}$ data
PLAT220	2	Test $U_{\text{eq}}(\text{max})/U_{\text{eq}}(\text{min})$ range for non-H atoms in non-solvent
PLAT221	4	Test $U_{\text{eq}}(\text{max})/U_{\text{eq}}(\text{min})$ range for non-H atoms in solvent
PLAT222	3	Test $U_{\text{eq}}(\text{max})/U_{\text{eq}}(\text{min})$ range for H atoms in non-solvent
PLAT223	4	Test $U_{\text{eq}}(\text{max})/U_{\text{eq}}(\text{min})$ range for H atoms in solvent
PLAT230	2	Hirshfeld rigid-bond test [ <i>Acta Cryst.</i> (1976), A32, 239–244]
PLAT231	4	Hirshfeld rigid-bond test [ <i>Acta Cryst.</i> (1976), A32, 239–244]
PLAT232	2	Hirshfeld rigid-bond test (metal-X) [ <i>Acta Cryst.</i> (1976), A32, 239–244]
PLAT233	4	Hirshfeld rigid-bond test (metal-X) [ <i>Acta Cryst.</i> (1976), A32, 239–244]
PLAT241	2	Test for unusually high $U_{\text{eq}}$ as compared with bonded neighbours
PLAT242	2	Test for unusually low $U_{\text{eq}}$ as compared with bonded neighbours
PLAT243	4	Test for unusually high solvent $U_{\text{eq}}$ as compared with bonded neighbours
PLAT244	4	Test for unusually low solvent $U_{\text{eq}}$ as compared with bonded neighbours
PLAT250	2	Test for unusual anisotropic average $U_{ij}$
PLAT301	3	Test for main residue(s) disorder
PLAT302	4	Test for (anion/solvent) disorder
PLAT305	2	Test for isolated hydrogen atoms
PLAT306	2	Test for isolated oxygen atoms
PLAT307	2	Test for isolated metal atoms
PLAT308	2	Test for single-bonded metal atoms
PLAT309	2	Test for single-bonded oxygen atoms
PLAT310	2	Test for 'too close' (symmetry-related) full-weight atoms
PLAT311	2	Test for isolated disordered oxygen atoms
PLAT312	2	Test for C=O—H
PLAT313	2	Test for O with three covalent bonds
PLAT318	2	Hybridization problem on N in main residue(s)
PLAT319	2	Hybridization problem on N in solvent/ion

## 5. APPLICATIONS

Table A5.7.2.2 (cont.).

Test name	Type	Purpose
PLAT320	2	Hybridization problem on C in main residue(s)
PLAT321	2	Hybridization problem on C in solvent/ion
PLAT322	2	Hybridization problem on non-C in main residue(s)
PLAT323	2	Hybridization problem on non-C in solvent/ion
PLAT324	2	Check for possibly missing H on coordinating X—N—X in main residue
PLAT325	2	Check for possibly missing H on coordinating X—N—X in solvent/anion
PLAT326	2	Check for possibly missing hydrogen atom on carbon with $sp^3$ -like geometry in the main residue
PLAT327	2	Check for possibly missing hydrogen atom on carbon with $sp^3$ -like geometry in the solvent/anion
PLAT328	2	Check for possibly missing H on potentially $sp^3$ P
PLAT330	2	Check average phenyl C—C
PLAT331	2	Check average phenyl C—C
PLAT332	2	Check phenyl C—C range
PLAT333	2	Check average in multiply substituted benzene C—C
PLAT334	2	Check average in multiply substituted benzene C—C
PLAT335	2	Check multiply substituted benzene C—C range
PLAT338	2	Check average torsion angle in cyclohexane ring
PLAT340	3	Check bond precision for C—C in light-atom structures [ $Z(\max) < 20$ ]
PLAT341	3	Check bond precision for C—C in structures [ $19 < Z(\max) < 40$ ]
PLAT342	3	Check bond precision for C—C in structures [ $Z(\max) > 39$ ]
PLAT350	3	Test for short C—H (ångstrom difference) XRAY: 0.96
PLAT351	3	Test for long C—H (ångstrom difference) XRAY: 0.96
PLAT352	3	Test for short N—H (ångstrom difference) XRAY: 0.87
PLAT353	3	Test for long N—H (ångstrom difference) XRAY: 0.87
PLAT354	3	Test for short O—H (ångstrom difference) XRAY: 0.82
PLAT355	3	Test for long O—H (ångstrom difference) XRAY: 0.82
PLAT360	2	Test for short C4—C4 (ångstrom difference) XRAY: 1.54
PLAT361	2	Test for long C4—C4 (ångstrom difference) XRAY: 1.54
PLAT362	2	Test for short C4—C3 (ångstrom difference) XRAY: 1.52
PLAT363	2	Test for long C4—C3 (ångstrom difference) XRAY: 1.52
PLAT364	2	Test for short C4—C2 (ångstrom difference) XRAY: 1.46
PLAT365	2	Test for long C4—C2 (ångstrom difference) XRAY: 1.46
PLAT366	2	Test for short C?—C? (ångstrom difference) XRAY: 1.50
PLAT367	2	Test for long C?—C? (ångstrom difference) XRAY: 1.50
PLAT368	2	Test for short C3—C3 (ångstrom difference) XRAY: 1.34
PLAT369	2	Test for long C3—C3 (ångstrom difference) XRAY: 1.34
PLAT370	2	Test for short C3—C2 (ångstrom difference) XRAY: 1.31
PLAT371	2	Test for long C3—C2 (ångstrom difference) XRAY: 1.31
PLAT372	2	Test for short C2—C2 (ångstrom difference) XRAY: 1.25
PLAT373	2	Test for long C2—C2 (ångstrom difference) XRAY: 1.25
PLAT374	2	Test for long N—N bond ( $> 1.45 \text{ \AA}$ )
PLAT380	4	Test for incorrectly oriented methyl moiety
PLAT390	3	Test methyl moiety X—C—H bond angle
PLAT391	3	Test methyl moiety H—C—H bond angle
PLAT395	2	Test X—O—Y angle
PLAT396	2	Test Si—O—Si angle
PLAT410	2	Test for short non-bonding intra H···H contacts
PLAT411	2	Test for short non-bonding inter H···H contacts
PLAT412	2	Test for short non-bonding intra H···H contacts (involving XH3)
PLAT413	2	Test for short non-bonding inter H···H contacts (involving XH3)
PLAT414	2	Test for short non-bonding intra D—H···H—X contacts
PLAT415	2	Test for short non-bonding inter D—H···H—X contacts
PLAT416	2	Test for short non-bonding intra D—H···H—D contacts
PLAT417	2	Test for short non-bonding inter D—H···H—D contacts
PLAT420	2	Test for D—H without acceptor
PLAT430	2	Test for short non-bonding inter D···A contacts
PLAT431	2	Test for short non-bonding inter HL···A contacts (HL = halogen)
PLAT432	2	Test for short non-bonding inter X···Y contacts
PLAT433	4	Test for short non-bonding minor···minor inter X···Y contacts
PLAT480	4	Test for too large H···A
PLAT481	4	Test for too large D···A
PLAT482	4	Test for too small D—H···A angle
PLAT601	2	Test for solvent accessible voids
PLAT602	4	Test for too large solvent accessible voids
PLAT603	4	Test for too large unit cell for void search
PLAT604	4	Test for too many voids
PLAT701	1	Test for consistency of bonds and coordinates in CIF
PLAT702	1	Test for consistency of angles and coordinates in CIF
PLAT703	1	Test for consistency of torsions and coordinates in CIF
PLAT704	1	Test for consistency of contact distances and coordinates in CIF
PLAT705	1	Test for consistency of H-bond D—H distances and coordinates in CIF
PLAT706	1	Test for consistency of H-bond H···A distances and coordinates in CIF
PLAT707	1	Test for consistency of H-bond D···A distances and coordinates in CIF
PLAT708	1	Test for consistency of H-bond D—H···A angles and coordinates in CIF
PLAT710	4	Test for linear torsions in CIF
PLAT711	1	Test for label problem for bonds in CIF
PLAT712	1	Test for label problem for angles in CIF
PLAT713	1	Test for label problem for torsions in CIF
PLAT714	1	Test for label problem for contact distances in CIF
PLAT715	1	Test for label problem for H-bond D—H distances in CIF
PLAT716	1	Test for label problem for H-bond H···A distances in CIF
PLAT717	1	Test for label problem for H-bond D···A distances in CIF

Table A5.7.2.2 (cont.).

Test name	Type	Purpose
PLAT718	1	Test for label problem for H-bond D—H···A angles in CIF
PLAT720	4	Test for unusual labels
PLAT721	1	Test for consistency of bonds and coordinates in CIF
PLAT722	1	Test for consistency of angles and coordinates in CIF
PLAT723	1	Test for consistency of torsions and coordinates in CIF
PLAT724	1	Test for consistency of contact distances and coordinates in CIF
PLAT725	1	Test for consistency of H-bond D—H distances and coordinates in CIF
PLAT726	1	Test for consistency of H-bond H···A distances and coordinates in CIF
PLAT727	1	Test for consistency of H-bond D···A distances and coordinates in CIF
PLAT728	1	Test for consistency of H-bond D—H···A angles and coordinates in CIF
PLAT731	1	Test for consistency of bond s.u.'s and coordinate s.u.'s in CIF
PLAT732	1	Test for consistency of angle s.u.'s and coordinate s.u.'s in CIF
PLAT733	1	Test for consistency of torsion s.u.'s and coordinate s.u.'s in CIF
PLAT734	1	Test for consistency of contact distance s.u.'s and coordinate s.u.'s in CIF
PLAT735	1	Test for consistency of H-bond D—H distance s.u.'s and coordinate s.u.'s in CIF
PLAT736	1	Test for consistency of H-bond H···A distance s.u.'s and coordinate s.u.'s in CIF
PLAT737	1	Test for consistency of H-bond D···A distance s.u.'s and coordinate s.u.'s in CIF
PLAT738	1	Test for consistency of H-bond D—H···A angle s.u.'s and coordinate s.u.'s in CIF
PLAT741	1	Test for missing bond s.u. in CIF
PLAT742	1	Test for missing angle s.u. in CIF
PLAT743	1	Test for missing torsion s.u. in CIF
PLAT744	1	Test for missing contact distance s.u. in CIF
PLAT745	1	Test for missing H-bond D—H distance s.u. in CIF
PLAT746	1	Test for missing H-bond H···A distance s.u. in CIF
PLAT747	1	Test for missing H-bond D···A distance s.u. in CIF
PLAT748	1	Test for missing H-bond D—H···A angle s.u. in CIF
PLAT751	4	Test for senseless bond s.u. in CIF
PLAT752	4	Test for senseless angle s.u. in CIF
PLAT753	4	Test for senseless torsion s.u. in CIF
PLAT754	4	Test for senseless contact distance s.u. in CIF
PLAT755	4	Test for senseless H-bond D—H distance s.u. in CIF
PLAT756	4	Test for senseless H-bond H···A distance s.u. in CIF
PLAT757	4	Test for senseless H-bond D···A distance s.u. in CIF
PLAT758	4	Test for senseless H-bond D—H···A angle s.u. in CIF
PLAT761	1	Test for the presence of at least one X—H in the CIF
PLAT762	1	Test for at least one X—Y—H or H—Y—H entry in the CIF
PLAT763	1	Test for missing bonds in CIF
PLAT764	4	Test for overcomplete bonds in CIF
PLAT770	2	Test for suspect C—H bonds in CIF (not caught otherwise)
PLAT771	2	Test for suspect N—H bonds in CIF (not caught otherwise)
PLAT772	2	Test for suspect O—H bonds in CIF (not caught otherwise)
PLAT773	2	Test for suspect C—C bonds in CIF (not caught otherwise)
PLAT779	2	Test for suspect angle in CIF (not caught otherwise)
PLAT780	2	Test whether coordinates form a connected set
PLAT790	4	Test whether c.g. residue in unit-cell box
PLAT798	4	Test for alphanumeric label on coordinate record
PLAT799	4	Test for alphanumeric label on displacement-parameter record
PLAT801	4	Test for missing, incomplete or out-of-order cell data
PLAT802	1	Test for input lines longer than 80 characters
PLAT803	1	Test for loop problem in CIF-read
PLAT804	4	Test for ARU-pack problem in PLATON
PLAT805	4	Test for insufficient coordinate data
PLAT806	4	Test for insufficient $U_{ij}$ data
PLAT810	4	Test for out-of-memory problem
PLAT850	4	Test for BASF/TWIN problem in SHELXL

## References

- Adobe Systems Incorporated (1999). *PostScript language reference*. 3rd ed. Reading, MA: Addison-Wesley Longman.
- Adobe Systems Incorporated (2004). *PDF reference*. 5th ed. *Adobe Portable Document Format*. Version 1.6. <http://partners.adobe.com/public/developer/en/pdf/PDFReference16.pdf>.
- Allen, F. H., Johnson, O., Shields, G. P., Smith, B. R. & Towler, M. (2004). *CIF applications. XV. enCIFer: a program for viewing, editing and visualizing CIFs*. *J. Appl. Cryst.* **37**, 335–338.
- Allen, F. H., Kennard, O., Motherwell, W. D. S., Town, W. G., Watson, D. G., Scott, T. J. & Larson, A. C. (1974). *The Cambridge Crystallographic Data Centre, part 3. The unique molecule program*. *J. Appl. Cryst.* **7**, 73–78.
- Becker, E. D. (2001). *Secretary General's Report*. *Chem. Int.* **23**, 135.

## 5.7. SMALL-MOLECULE CRYSTAL STRUCTURE PUBLICATION USING CIF

- Bernstein, H. J. (1998). *cif2cif*. CIF copy program. <http://www.iucr.org/iucr-top/cif/software/cif2cif/cif2cif.scr/>.
- Bernstein, H. J. (2000). *xmlCIF: a proposal for faithful representation of Extensible Markup Language (XML) documents within Crystallographic Information File (CIF) data sets*. <http://www.bernstein-plus-sons.com/software/xmlCIF/>.
- Boulay, D. J. du & Hall, S. R. (1996). *PREPUB*. Pre-publication tests on CIF structural data. <http://xtal.sourceforge.net/man/prepub-desc.html>.
- Brown, I. D. (1983). *The standard crystallographic file structure*. *Acta Cryst.* **A39**, 216–224.
- Brown, I. D. (1988). *Standard Crystallographic File Structure-87*. *Acta Cryst.* **A44**, 232.
- Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). *New software for searching the Cambridge Structural Database and visualizing crystal structures*. *Acta Cryst.* **B58**, 389–397.
- Freed, N. & Borenstein, N. (1996). *Multipurpose Internet Mail Extensions (MIME) part two: media types*. Internet Engineering Task Force. Request for comment 2046. <http://www.ietf.org/rfc/rfc2046.txt>
- Gabe, E. J., Le Page, Y., Charland, J.-P., Lee, F. L. & White, P. S. (1989). *NRCVAX – an interactive program system for structure analysis*. *J. Appl. Cryst.* **22**, 384–387.
- Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *The Crystallographic Information File (CIF): a new standard archive file for crystallography*. *Acta Cryst.* **A47**, 655–685.
- Hall, S. R., du Boulay, D. J. & Olthof-Hazekamp, R. (2000). *Xtal crystallographic software*. <http://xtal.sourceforge.net>.
- Hall, S. R. & Sievers, R. (1993). *CIF applications. I. QUASAR: for extracting data from a CIF*. *J. Appl. Cryst.* **26**, 469–473.
- Hester, J. R. & Hall, S. R. (1996). *BUNYIP: in search of errant symmetry*. *J. Appl. Cryst.* **29**, 474–478.
- Knuth, D. E. (1986). *The TeXbook*. Computers and typesetting, Vol. A. Reading, MA: Addison-Wesley.
- Le Page, Y. (1988). *MISSYMI.1 – a flexible new release*. *J. Appl. Cryst.* **21**, 983–984.
- Murray-Rust, P. & Rzepa, H. S. (1999). *Chemical markup, XML and the Worldwide Web. 1. Basic principles*. *J. Chem. Inf. Comput. Sci.* **39**, 928–942.
- Murray-Rust, P. & Rzepa, H. S. (2001). *Chemical markup, XML and the Worldwide Web. 2. Information objects and the CMLDOM*. *J. Chem. Inf. Comput. Sci.* **41**, 1113–1123.
- Nardelli, M. (1983). *PARST. A system of FORTRAN routines for calculating molecular structure parameters from results of crystal structure analyses*. *Comput. Chem.* **7**, 95–98.
- Rzepa, H. S., Murray-Rust, P. & Whitaker, B. J. (1998). *The application of chemical Multipurpose Internet Mail Extensions (chemical MIME) internet standards to electronic mail and world-wide web information exchange*. *J. Chem. Inf. Comput. Sci.* **38**, 976–982.
- Sheldrick, G. M. (1976). *SHELX76. Program for crystal structure determination*. University of Cambridge, England.
- Sheldrick, G. M. (1997). *SHELX97. Program for the refinement of crystal structures*. University of Göttingen, Germany. <http://shelx.uni-ac.gwdg.de/SHELX/>.
- Spek, A. L. (1990). *PLATON, an integrated tool for the analysis of the results of a single crystal structure determination*. *Acta Cryst.* **A46** (Suppl.), C34.
- Spek, A. L. (2003). *Single-crystal structure validation with the program PLATON*. *J. Appl. Cryst.* **36**, 7–13.
- Westrip, S. P. (2004). *printCIF for Word*. <http://www.iucr.org/iucr-top/cif/software/printCIFforWord/index.html>.
- Willis, A. C., Beckwith, A. L. J. & Tozer, M. J. (1991). *trans-3-Benzoyl-2-tert-butyl-4-isobutyl-1,3-oxazolidin-5-one*. *Acta Cryst.* **C47**, 2276–2277.