

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      1      \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_EX (Knuth, 1986). The T_EX 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 \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