CIF applications. XV. *enCIFer*: a program for viewing, editing and visualizing CIFs

Frank H. Allen,* Owen Johnson, Gregory P. Shields, Barry R. Smith‡ and Matthew Towler

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK. Correspondence e-mail: allen@ccdc.cam.ac.uk

The *enCIFer* program permits the location, reporting and correction of syntax and format violations in single- or multi-block crystallographic information files (CIFs). The program also permits the editing of existing individual or looped data items and the addition of new data in these categories, and provides data-entry wizards for the addition of two types of standard information for small-molecule structural studies, namely publication data and chemical and physical property information. Facilities for the graphical visualization and manipulation of structure(s) in a CIF are also provided.

1. Introduction

The crystallographic information file (CIF; Hall *et al.*, 1991; Brown & McMahon, 2002; http://www.iucr.org/iucr-top/cif/) is the international standard for the transfer of crystallographic information among individuals and laboratories and, most importantly, is increasingly being adopted as the required format for submission of crystal structure data to journals and databases. Although the CIF was specifically designed to be human-readable, the syntax requirements make it rather unsuitable for direct editing or enhancement using standard text editors. The core data items of a CIF, recording the results of structure solution and refinement, are normally generated automatically by crystallographic software packages, and the principal need to edit a CIF arises when the data are being prepared for publication in a journal or for transmission to a database. This requires the addition of, *inter alia*, information concerning the authors' names and addresses, a chemical description of the substance, and various chemical and physical properties, for example crystal colour, melting point, *etc*. Even the core data must, at times, be changed or updated, for example to indicate those geometric parameters that should be published in the paper. It is in these editing processes that the CIF syntax conventions can easily be violated unless special care (or software) is employed to check the resulting file.

Close to 95% of new crystal structure data for the Cambridge Structural Database (CSD; Allen, 2002) now arrive at the Cambridge Crystallographic Data Centre (CCDC) in electronic form, and the vast majority of these data are in CIF format. In nearly half of the incoming CIFs, the syntax rules have been violated, and even though many of these violations are relatively minor, they prevent the CIF from being correctly parsed, for example by *Mercury*, the CCDC's structure visualizer (Taylor & Macrae, 2001; Bruno *et al.*, 2002), or by CCDC in-house software systems that underpin the value-added conversion of raw CIFs to entries in the distributed CSD. For this reason, we have now developed the *enCIFer* program as a general-purpose CIF editor. The software incorporates locally written C++ classes, together with the C++ *Qt* library (Trolltech AS, 1995) for building the graphical user interface (GUI). This article describes the facilities available in Version 1.1 of the program.

2. Capabilities of *enCIFer*

2.1. Overview of principal features

The *enCIFer* program operates on single- or multi-block CIFs to permit:

(*a*) Choice of CIF dictionaries for file validation.

(*b*) Location, reporting and correction of syntax violations.

(*c*) Editing of existing individual or looped data items.

(*d*) Addition of new individual or looped data items.

(*e*) Addition of certain standard information *via* two data-entry wizards:

(i) Publication wizard: prompts for the basic bibliographic information required by most journals and databases which accept CIF deposition documents.

(ii) Data wizard: prompts for chemical and physical property information, which enhances a raw CIF for deposition with a journal or database.

(*f*) Visualization of structure(s) in the CIF.

enCIFer may be used to check the syntax integrity of the amended file in all cases where data are edited or added.

2.2. Overview of the interface

The main *enCIFer* window is depicted in Fig. 1. This contains eight segments:

(*a*) *Top-Level Menu* (*File*, *Edit*, *Search*, *Tools*, *Help*).

(*b*) *Toolbar* containing many common program options.

(*c*) *Browser Box* permitting CIF dictionary navigation (top left-hand pane).

(*d*) *Text Editor* (top right-hand pane).

(*e*) *Visualizer* button for displaying crystal structure(s) in the CIF.

(*f*) *Error List* for displaying and navigating error, warning and remark messages generated by *enCIFer* (bottom left-hand pane).

(*g*) *Message Log*, a scrolling log of all *enCIFer* messages (bottom right-hand pane).
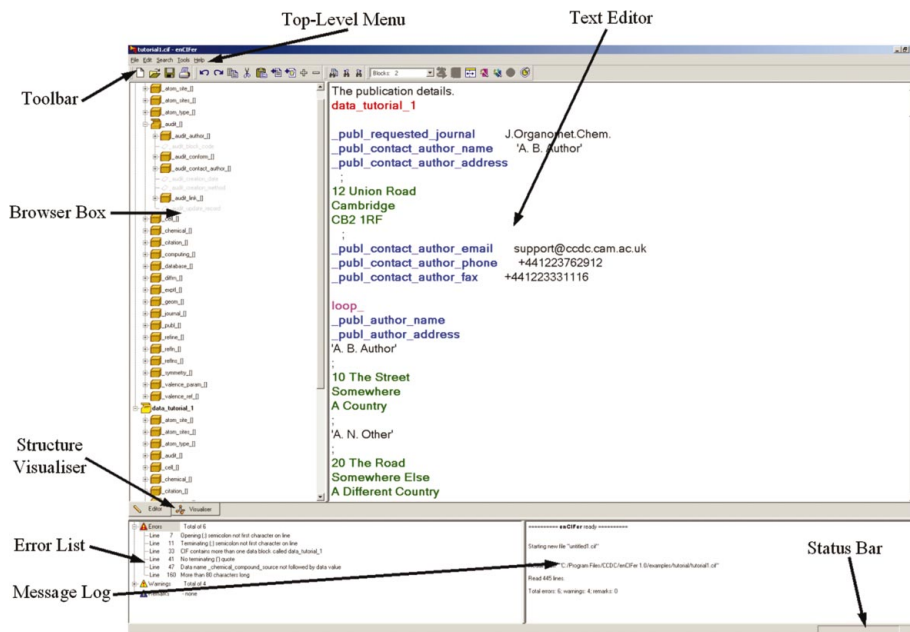
Figure 1
Main control and display window of the *enCIFer* graphical user interface.

(*h*) *Status Bar* at the bottom of the window, which displays help messages and line and column numbers.

The interface can also access:

(*a*) *Loop Editor* (Fig. 2), which provides a spreadsheet view of looped data items.

(*b*) *Wizards* (*e.g.* Fig. 3) for entering crystal, chemical and publication data.

(*c*) *Visualizer* windows (*e.g.* Fig. 4), for graphical display of crystal and molecular structures.

2.3. CIF dictionaries supported

Valid CIF data names and the permitted data value type(s) for each name are expressed in computer-readable dictionaries, where the dictionary syntax is defined in a separate dictionary definition language (DDL; Hall & Cook, 1995). *enCIFer* is able to load
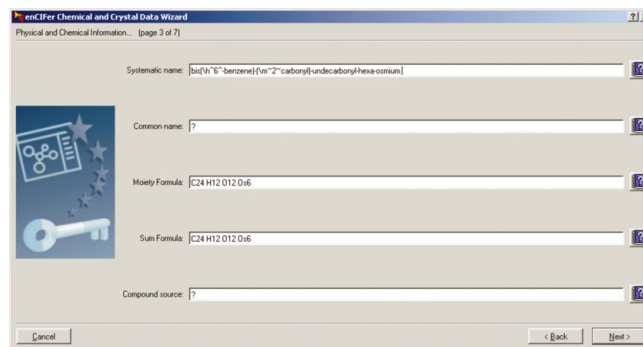

Figure 3
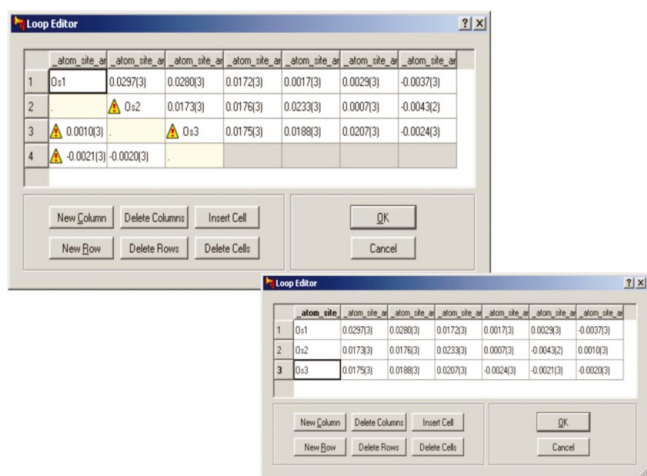One of the data-entry panes of the *enCIFer Crystal Data Wizard*.

Figure 2
The *enCIFer Loop Editor* window showing a loop containing errors before (upper pane) and after correction (lower pane).
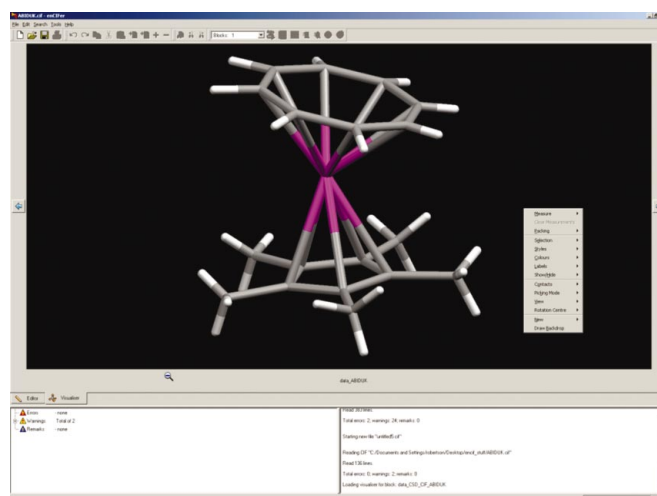

Figure 4
The *enCIFer* visualizer window.

dictionaries which conform to the DDL1.4 format (http://www.iucr.org/iucr-top/cif/), including the small-molecule core dictionary and the powder diffraction dictionary. The current DDL1 dictionaries are included in the *enCIFer* distribution with the permission of the IUCr as copyright holder. *enCIFer* does not support DDL2 dictionaries, *e.g.* mmCIF, the macromolecular CIF dictionary.

2.4. File operations

On launching *enCIFer*, the *Text Editor* pane shows an empty CIF. An existing CIF can be loaded using *Open*, or by supplying a file name on the *encifer* command line in UNIX, or by drag and drop in Windows. Multiple CIFs can be loaded and viewed either in separate *enCIFer* windows or in a single reusable window. Optionally, *enCIFer* can be configured to open or insert a template CIF.

2.5. CIF display and syntax or format checking

Once a CIF is loaded into *enCIFer*, its contents are displayed in the *Text Editor* pane using configurable colour highlighting according to CIF syntax, *e.g.* bold red for the data-block header, bold blue for data names in CIF dictionaries, bold magenta for loop keywords, *etc.* This colour coding is particularly useful in tracking content errors where text fields lack a closing semicolon; this type of error is notoriously difficult to locate otherwise. The CIF is parsed to check for dictionary compliance and syntax violations. Optionally, the CIF can be checked for the presence of mandatory data items, listed in configurable files, that may be required by specific journals or databases. Check results are classified into errors, warnings and remarks, and these messages are listed in the *Error List* box, with a summary written to the *Message Log*. The message lists may be expanded or contracted by double-clicking the appropriate icons adjacent to the words *Errors*, *Warnings* or *Remarks*. Double-clicking on a specific *Error* or *Warning* message displays and highlights the relevant line of the CIF in the *Text Editor* pane.

2.6. CIF editing with CIF dictionary assistance

Text may be typed into the *Text Editor* pane as for a standard plain text editor. The *enCIFer* editor supports copy, cut and paste, undo and redo, and find and replace mechanisms. Within extended text fields, limited support exists for special representation of Greek symbols and subscript or superscript text. To assist the editing process, CIF dictionary information can be accessed in two ways, firstly by simply right-clicking on the data item to be edited in the *Text Editor* pane, and secondly by using the *Browser Box*. The browser pane provides a hierarchical view of each CIF in terms of data blocks and their data items. The hierarchy is defined by the arrangement of data categories and data items in the CIF dictionaries, and the full hierarchy can be viewed by clicking on *Expand* (and *Contract*) buttons in the browser. Data items present in the current CIF block are shown as black text, while data items which are not present are in grey. This provides a means of navigating the blocks and data items present in the CIF: right-clicking on a black data name provides dictionary information about that data item and allows the data value to be edited for non-looped data items, while right-clicking on a grey data name allows it to be inserted and/or the corresponding data value set in the current CIF data block displayed in the *Text Editor* pane.

2.7. Editing or inserting CIF loops

CIF loop structures can be inserted or edited using the spreadsheet-style *Loop Editor* (Fig. 2) as an alternative to using the *Text*

Table 1
Information that can be added to a CIF using the *Crystal Data Wizard* in *enCIFer*.

(*a*) Physical and chemical information

Systematic chemical name
Common chemical name
Moiety formula
Sum formula
Source of the chemical compound
Crystallization solvent
Melting point
Crystal habit
Crystal colour
Temperature of diffraction experiment
Pressure of diffraction experiment

(*b*) Diffraction information

Radiation type
Radiation wavelength
Radiation source

Editor pane (which is disabled when the *Loop Editor* is invoked). An existing loop is displayed as a spreadsheet, with the data names shown as column headings and with the loop rows numbered sequentially. Spreadsheet cells are colour coded according to their data content, with grey used for empty cells, yellow for cells containing '.' (placeholder values) and blue for cells containing '?' (unknown values). Cells containing values that are incompatible with the CIF dictionary definitions show a yellow warning triangle. Data item assistance can be obtained by right-clicking on a spreadsheet cell. The regular arrangement of data values in this format allows easy visual detection of out-of-phase errors, where a column (or columns) of data values is omitted yet the total number of included data values is still an integer multiple of the number of declared data items. This is another type of error that can be difficult to detect by other means. Apart from simply altering data values, loop-editing facilities also include the ability to resize or move columns and rows, and to add or delete columns, rows or cells. Changes made using the *Loop Editor* can be reviewed before they are applied to the target CIF. Finally, a completely new loop can be inserted into a CIF data block at the current cursor position in the *Text Editor* pane.

2.8. Data-entry wizards

Two wizards are provided in *enCIFer*, namely a *Publication Wizard* for entering bibliographic data, and a *Crystal Data Wizard* for entering additional crystallographic and chemical information for small-molecule structural studies. Both wizards operate on the current CIF data block, as determined by the cursor position in the *Text Editor* pane. Both wizards will display any relevant data already present in the CIF, so they also provide facilities for editing or updating that information. All additions made with each wizard may be reviewed before they are incorporated into the target CIF. The act of data incorporation takes care of the necessary CIF syntax and format rules.

The *Publication Wizard* permits entry of contact author details, which can also be entered automatically *via* appropriate preference settings, and of journal and author information. The journal information for which it prompts will vary depending on whether the CIF is being submitted for publication, has already been published, or is being submitted directly to a database as a private communication. A scrolling list of journals that are already represented in the Cambridge Structural Database is provided to assist data entry.

The *Crystal Data Wizard* prompts for entry of any or all of the physical and chemical data and diffraction information summarized

Table 2
Summary of the structure visualization features available in *enCIFer*.

Rotate, translate and scale the three-dimensional crystal structure display
View down cell axes, reciprocal cell axes and normals to planes
Range of visualization options, *e.g.* different display styles, colouring and labelling
    options, ability to hide and then redisplay atoms, molecules, *etc.*
Measure distances, angles and torsion angles
Create and display centroids and least-squares mean planes
Draw crystal-packing diagrams for a single unit cell

in Table 1. This wizard also permits the crystal system to be selected from a pull-down list containing the allowed CIF values, and for the space-group number (*International Tables for Crystallography*, 1995) to be entered if it is not already present in the CIF. The Hermann–Mauguin space-group symbol may already be specified in the CIF, or will be given an initial value by *enCIFer* using any symmetry-equivalent positions already present in the CIF, or may be selected directly in the wizard from a pull-down list of common space-group settings that correspond to the given space-group number. The wizard will generate warnings if there are inconsistencies between the crystal system, the space-group number and the Hermann–Mauguin symbol, so that these can be resolved before any information is incorporated into the target CIF.

2.9. Structure visualization

Crystal structure visualizer windows may be displayed by clicking the *Visualizer* button just below the *Text Editor* pane (Fig. 1). By default, a 2×3 grid of visualizer windows is shown, with one window for each data block containing crystal structure data in the CIF. A zoom facility permits isolation of an individual structure in a single visualizer window (Fig. 4). Right-clicking in the visualizer background or on a specific object (atom, bond, plane, *etc.*) will generate menus which access the display options summarized in Table 2. The visualization facilities in *enCIFer* use much of the underlying C++ code that is used in the CCDC's *Mercury* program (Taylor & Macrae, 2001, Bruno *et al.*, 2002). *Mercury* can itself read CIFs and provides more extensive structure visualization facilities, particularly for generating and exploring networks of intermolecular contacts. *Mercury* is freely downloadable from http://www.ccdc.cam.ac.uk/ for *bona fide* research purposes.

3. Program availability and documentation

The *enCIFer* program is available as a free download from http://www.ccdc.cam.ac.uk/ for *bona fide* research use. Program executables are available for a number of operating systems, including Windows, Linux (Intel), Solaris (SPARC and Intel) and SGI (IRIX). Version 1.0 was released in April 2003, and Version 1.1 was made available in April 2004. Note that *enCIFer* is not currently supported for Macintosh computers; however, a Mac OS X port is under consideration.

Full documentation (73 pages) and three *enCIFer* tutorials are provided with the download. Documentation may be freely accessed and viewed in HTML format or as a PDF file *via* the CCDC website noted above. Complete installation instructions are provided with the downloaded files. User support for *enCIFer* is provided by the CCDC and queries about the program and its operation may be e-mailed to support@ccdc.cam.ac.uk, or as otherwise directed from time to time on the CCDC website.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Brown, I. D. & McMahon, B. (2002). *Acta Cryst.* B**58**, 317–324.
Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P. M., Pearson, J. & Taylor, R. (2002). *Acta Cryst.* B**58**, 389–397.
Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Cryst.* A**47**, 655–685.
Hall, S. R. & Cook, A. P. F. (1995). *J. Chem. Inf. Comput. Sci.* **35**, 819–825.
International Tables for Crystallography (1995). Vol. A. Dordrecht: Kluwer Academic Publishers.
Taylor, R. & Macrae, C. F. (2001). *Acta Cryst.* B**57**, 815–827.
Trolltech AS (1995). *Qt.* Trolltech AS, Oslo, Norway.