cif applications

CIF applications. XIII. *CIFEDIT*, a program for viewing and editing CIFs

Brian H. Toby

NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8562, USA. Correspondence e-mail: brian.toby@nist.gov

The *CIFEDIT* program displays the contents of single- or multi-block crystallographic information files (CIF). The program can also be used to edit the data values for CIF items, including adding or deleting sets of entries in a loop. CIF dictionaries are used to display definitions for data names and units. In the case of DDL 1.4 dictionaries, validation information is used to check numerical values. The program is written using the cross-platform scripting language Tcl/Tk and thus is largely platform-independent. It has been tested under Windows, Mac OS X and Unix.

1. Introduction

The crystallographic information file (CIF) is used to communicate results of structure determination studies, as well as associated data and other related information (Hall *et al.*, 1991). This article describes an application, *CIFEDIT*, that can be used for viewing the contents of a CIF as well as editing it. Information about many topics related to CIF, including other software tools, can be found on the IUCr website http://www.iucr.org/iucr-top/cif. Subsequent to the release of *CIFEDIT*, the Cambridge Crystallographic Data Centre have released a program with similar capabilities, *enCIFer* (Smith & Johnson, 2002).

The *CIFEDIT* program has been written using the Tcl/Tk scripting language, a programming practice that encourages portability (Ousterhout, 1994). It has been tested using Windows and multiple Unix implementations. Limited testing has also been performed on a Macintosh under OS X. It is anticipated that *CIFEDIT* will function under all operating systems where Tcl/Tk is available. The program highlights the programming features of Tcl/Tk in that only ~2600 lines of code were written to implement this program. The code was designed from the beginning to be embedded into larger programs, to add CIF reading and editing capabilities. This code has been utilized in the programs *EXPGUI*, *pdCIFplot* and *FillCIF* (Toby, 2001; Toby *et al.*, 2003; Toby, 2003). Other software authors are also encouraged to use this code.

2. Capabilities

The program *CIFEDIT* reads a CIF into memory and then parses it. The data names and loops within the CIF are then displayed in a hierarchical browser, as seen on the left-hand side of Fig. 1. Selecting a CIF data name causes the associated value to be displayed. The program offers two modes: values can be viewed in 'browse mode' and changed in 'edit mode'. The program offers two methods for working with information in loops. If the loop entry is selected, then one set of values for all data names in the loop is displayed, as shown in Fig. 2. Alternatively, if a single data name within a loop is selected, then all values (up to 100) for that data name are shown. This limit can easily be increased or decreased by changing the variable
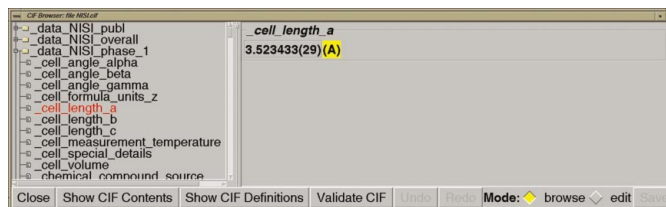

Figure 1
Screen image from the *CIFEDIT* program showing, on the left-hand side, a hierarchical view of the CIF contents and, on the right-hand side, the contents of CIF data item `_cell_length_a`. The program is in 'browse' mode, so the value cannot be changed.
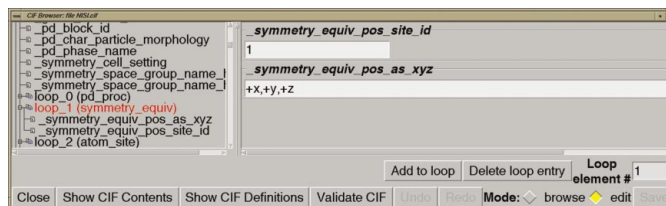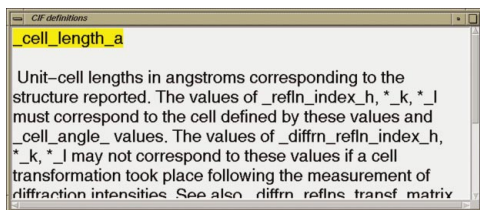

Figure 2
Screen image from the *CIFEDIT* program showing the selection of an entire loop. Values for a selected 'row' in the loop are shown to the right. The program is in 'edit' mode, so the values can be changed.

CIF(maxRows). However, display of very large loops can overtax the computer operating system.

The program utilizes information from CIF dictionaries to display units or enumerated values for a data item. When a data item has a list of enumerated values, those values, along with the CIF special values of '?' and '.', are the only options offered to the user. For DDL version 1.4 dictionaries, numerical values are checked when entered against the allowed range specified for the data item.

As shown in Fig. 3, definitions from CIF dictionaries are displayed in a separate window. Dictionaries are accessed *via* an index file, which provides a list of defined data names, locations for the definitions and validation information. *CIFEDIT* is distributed with the copies of six CIF dictionaries and an index file that references these dictionaries. If a user wishes to update a dictionary or select dictionaries to be accessed by the program, the index file must be

Figure 3
Screen image from the *CIFEDIT* program showing a CIF dictionary definition. This definition is displayed automatically when the `_cell_length_a` data name is selected in the hierarchical CIF viewer. Definitions are read from a CIF dictionary (in this case the core CIF dictionary). Users may select which dictionaries will be referenced by the program.

rebuilt. This is accomplished by executing a separate Tcl/Tk script, indexCIFdict.tcl, which is also supplied with *CIFEDIT*. The index file is read from the subdirectory where the *CIFEDIT* script is found. This means that, by duplicating this single subdirectory, it becomes possible to have multiple versions of *CIFEDIT*, where each copy validates CIFs against differing sets of dictionaries.

A third window can be used to view or edit the contents of the CIF directly. Editing the CIF from this window bypasses the normal dictionary-based validation procedures, so it is not recommended for routine use. However, it can be useful for more advanced purposes, such as correction of syntax errors.

The in-memory processing used in *CIFEDIT* can be expected to pose limitations on the maximum size of CIF that can be read, but the limit will depend on the operating system and the amount of available memory. Since many computer systems behave poorly when all resources are consumed, a limit of 100 000 data values has been coded into *CIFEDIT*. This limits *CIFEDIT* to a *circa* 25 Mb task size (in Linux) and is sufficient for CIFs of the order of 0.5 Mb in length. If resources permit, this limit can easily be raised by editing the cifedit.tcl file to change the value of the variable CIF(maxvalues). On larger computers, CIFs as large as 3 Mb have been read successfully.

3. Distribution and documentation

The *CIFEDIT* program is not subject to copyright and thus is freely available for use, modification and redistribution. The program, including source code, can be downloaded from a NIST website as part of the NCNR CIFtools distributions (ftp://ftp.ncnr.nist.gov/pub/cryst/cif). A separate WWW page describes how to install and use the program (http://www.ncnr.nist.gov/xtal/software/cif/cifedit.html). WWW pages have also been provided to describe the internal operation of the program by documenting the computer code (http://www.ncnr.nist.gov/xtal/software/cif/cifedit/cifedit_doc.html).

References

Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Cryst.* A**47**, 655–685.
Ousterhout, J. K. (1994). *Tcl and the Tk Toolkit.* London: Addison-Wesley Professional Computing.
Smith, B. & Johnson, O. (2002). *enCIFer: a Program for Checking and Editing CIF Files.* Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England. (http://www.ccdc.cam.ac.uk/prods/encifer/.)
Toby, B. H. (2001). *J. Appl. Cryst.* **34**, 210–213.
Toby, B. H. (2003). *J. Appl. Cryst.* **36**, 1285–1287.
Toby, B. H., Von Dreele, R. B. & Larson, A. C. (2003). *J. Appl. Cryst.* **36**, 1290–1294.