

5. APPLICATIONS

system, although the *enCIFer* program runs on a variety of different operating systems and machine hardware platforms. Nevertheless, the use of a common menu style makes the initial use of the program much easier for novice users and allows the program to be effectively used without detailed study of its documentation.

5.3.3.1.3. *The task bar*

The task bar allows rapid one-click access to the standard operations of creating a new document, opening, saving or printing the contents of the current file, copying, cutting and pasting text, searching for specific text within the document, and undoing or redoing previous edits.

Two buttons allow insertion of complete text files. One allows the user to select any file from local or network-mounted file systems. The other imports a specific file (the location of which may be specified by the user through the 'Preferences...' selection of the main 'Edit' menu). While this specific file may contain anything, it is intended to be a template CIF that a user will tailor to meet their own requirements. The default provided with the software is a standard template distributed by the IUCr for use in submitting articles to *Acta Crystallographica*. In either case, the file is imported at the current editing location and is not subject to validation upon input; the user must manually revalidate the file after import.

An icon on the task bar allows the user to run a validation procedure. This icon will be dimmed (indicating that the validation procedure may *not* be run) unless the user has modified the contents of the CIF. Other icons on the task bar behave in the same manner, allowing the procedures with which they are associated to be executed only under appropriate circumstances. Thus, for example, the looped list editor is not invoked unless the user clicks within the reserved word `loop_` in a list header.

Similarly, the 'help' icon in the task bar is dimmed unless the user has selected a data name in the CIF; when this is done, the icon is activated and clicking on it launches a help window containing the CIF dictionary definition of the data item.

The task bar also contains a drop-down menu listing all the data-block names in the current file. When the user selects one of the data-block names, the edit cursor is positioned at the head of the matching data block in the edit window. This is a rapid and efficient way of navigating within large and complex files.

The other buttons provided on the task bar allow the user to: reduce or increase the font size in the editing window; create a new looped list within the loop-editing window; invoke the publication and data-entry wizards; and hide or reveal the dictionary browse window pane.

Users may modify the appearance of the task bar to retain or conceal subsets of these icons, depending on which they find most useful.

5.3.3.1.4. *The main edit pane*

The main edit pane is a text-editing area where the user may directly modify the content of a CIF. Colours and font styles are used to indicate different syntactic elements. The details of the colours and styles may be modified to suit the user.

For the novice user, this is perhaps the most immediately helpful feature offered by this program. When a trailing semicolon is inadvertently lost from an extended text field, typical sequential parsers may interpret succeeding tokens as part of the quoted text and produce misleading error reports. Within the *enCIFer* edit window, all such text is marked up in a specific colour (green by default) so that the fault is much more obvious to the human eye and its source much easier to locate.

Two other typographic cues are used to help the user to trace errors, or to ensure that certain text has been input correctly. Subscripts and superscripts are represented in a smaller typeface (and in a different colour) so that missing delimiter characters are again obvious to the eye. Secondly, some special characters in the conventional CIF encoding (such as Greek letters) are displayed in an appropriate symbol font when the file is first loaded, so that for example the input string `\a` is rendered as α . Note that the backslash character is retained, and that the symbol character is not generated as new text is input or edited. This scheme therefore has some potential for confusion, but is nevertheless helpful in checking that less obvious special codes have been entered correctly.

The user is free to enter arbitrary text in this pane, possibly breaking CIF syntax rules in the process. Only when the revalidation process is manually invoked will the file be rescanned and any errors reported.

5.3.3.1.5. *The dictionary browse pane*

The upper left-hand pane in Fig. 5.3.3.1 illustrates the dictionary browser, an optional graphical view of the contents of the CIF dictionary against which the file is being validated. (The presence or absence of this pane is toggled from an icon in the task bar.) Box icons represent the contents of categories, and the tree of category containers may be expanded or collapsed as desired to show individual items within categories.

A dictionary view is generated for each separate data block in the CIF. Within the dictionary view of an individual data block, those data items present in the data block are shown in bold; other items defined in the dictionary but absent from the current data block appear in a lighter colour.

Within the dictionary browse pane, a user may select (with a click of the appropriate mouse button) a menu of three options which depend on whether the data name is present or absent in the data block. If present, one option positions the cursor in the editing window at the location of the selected data item. If the item is absent from the data block, the user is given the option to paste the data name into the editing window at the current insertion point. The other options (in both cases) are to copy the data name to the clipboard or to open the help window with the CIF dictionary definition of the selected item.

5.3.3.1.6. *The error notification pane and logging area*

The lower left-hand pane of Fig. 5.3.3.1 illustrates typical error notices generated by the parser when the validation process is invoked. At present, the classification of the severity of errors is guided by the editorial requirements of databases and journals, and does not necessarily match the formal errors dictated by the CIF specification. It is likely that this will change in future releases as validation is driven increasingly by the dictionaries rather than by hard-coded subroutines.

A convenient feature is that double-clicking on the line number in the error report relocates the cursor to that line in the editing pane. At present, error messages are listed by line only – they are not grouped by data block.

The user has a small number of options to control error notification. The choice of the maximum number of consecutive error lines to permit before error checking is abandoned is a useful way, especially for novices, to reduce the amount of output generated by severe syntax errors and to focus on repairing individual errors. The user may also specify a file that contains a set of CIF data names which are considered *mandatory* components of a particular file. Absence of any of these items from the current data

5.3. SYNTACTIC UTILITIES FOR CIF

	geom_bond_atom_ste_label_1	geom_bond_atom_ste_label_2	geom_bond_distance	geom_bond_publ fl
1	C11	C151	1.327(13)	yes
2	C11	C122	1.36(2)	yes
3	C11	C22	1.469(13)	yes
4	C11	S121	1.710(8)	yes
5	C11	S152	1.708(17)	yes
6	S121	C131	1.724(11)	yes
7	C131	C141	1.342(14)	no
8	C141	C151	1.372(13)	no
9	C122	C132	1.35(2)	no
10	C132	C142	1.35(2)	?
11	C142	S152	1.72(2)	?

Fig. 5.3.3.2. The *enCIFer* loop editor.

block is flagged as an error. The program log in the lower right-hand part of the program window records the history of the user's interactions with the file during the current editing session.

Information is written to the status bar (the lower margin of the window) to indicate the location by line and column number of the editing cursor.

5.3.3.1.7. The loop editor

The program has a useful spreadsheet-style editor for looped lists (Fig. 5.3.3.2). A particular benefit of this style of display is that the spreadsheet cells are arranged in a rectangular grid, so that visual scans can often detect deviations from a pattern of values within a column, thus making it easy to identify placement errors where values have been omitted or inadvertently conjoined. Such errors are not always obvious by direct visual inspection of a CIF, where the layout of a looped list need not follow any regular pattern.

The buttons to add or delete columns allow for the straightforward addition or deletion of data items from the loop. If the user selects the 'New Column' button, a small pop-up window helpfully provides a view of the associated dictionary (in the same hierarchical category-based tree view of the dictionary browser pane) to help the user select the required new data name. The 'Insert Cell' and 'Delete Cells' buttons are convenient tools for the realignment of rows and columns where values have been omitted or misplaced.

The loop editor is invoked from one of two buttons in the task bar, allowing either the creation of a new looped list or the modification of an existing one. As with the application as a whole, there is no dynamic validation of input; the new list must be saved and the entire CIF then manually revalidated.

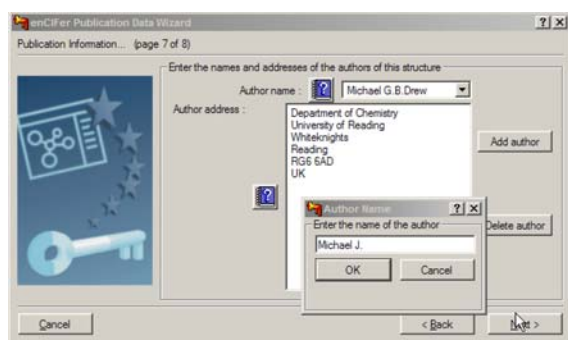


Fig. 5.3.3.3. The *enCIFer* publication data wizard. Information about the title and authors of an article to be submitted for publication is requested through a sequence of linked dialogue boxes.

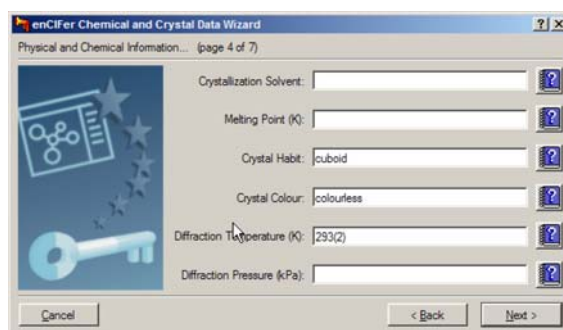


Fig. 5.3.3.4. The *enCIFer* chemical and crystal data wizard.

5.3.3.1.8. The publication and chemical and crystal data wizards

The user may invoke data-entry 'wizards', subordinate programs that prompt for particular data items useful for the publication of a crystal structure report or for the deposition of a crystal structure in a database. This is the kind of information that might be requested in the *Notes for authors* for a journal, and it is helpful if the information is routinely requested from inexperienced authors during normal use of the software. The data-entry tools are known as 'wizards' because they will utilize information already in the file.

Hence, as shown in Fig. 5.3.3.3, details of an article's contact author are retrieved from the CIF and used to seed a list of contributing authors. As the address for each author is entered, the program makes each new address available as a stored record for easier input of additional information.

Fig. 5.3.3.4 demonstrates the same approach to encouraging authors to supplement information already in the CIF with related chemical (or crystal) data not usually provided by the CIF generators embedded in crystallographic structure determination programs.

5.3.3.1.9. The visualization window

A final useful feature of *enCIFer* is its ability to visualize the three-dimensional structure of molecules described in the data blocks of a CIF. Fig. 5.3.3.5 demonstrates crystal packing with

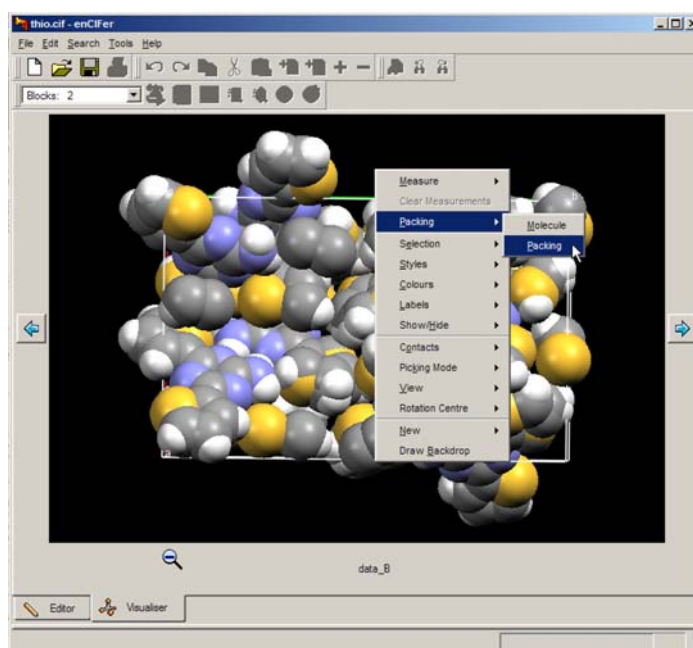


Fig. 5.3.3.5. Visualization of a molecular and crystal three-dimensional structure with *enCIFer*.