Chapter 10.2

CIFTAB

Sheldrick, G. M. (2008). *Acta Cryst*. **A64, 112-122.**

George W. Sheldrick
Dept. of Structural Chemistry
Tammannstraße 4
37077 Göttingen, Germany
email: gsheldr@shelx.uni-ac.gwdg.de

10.2 CIF, CIFTAB and Electronic Publication

10.2.1 CIF archive format

The *CIF* format represents a major step forward in the archiving, publication and communication of crystallographic data. At last it is possible to publish crystal structures and incorporate structural data into the crystallographic databases without the expensive and error-prone retyping of tables by hand. CIF format also provides a convenient method of transferring data from one program system to another. The ACTA instruction instructs SHELXL to write two CIF-format files: *name.fcf* contains the reflection data and 'name.cif' all other data. These files contain all the items needed for archiving the structure; those answers not known to SHELXL (e.g. the color of the crystal) are left as a question mark. In general the final 'name.cif' file should be edited using CIFTAB or any text editor to replace most of these question marks. The file is then suitable for deposition in the CSD (organic) and ICSD (inorganic crystal structure) databases.

For publication of a routine structure determination via electronic mail it will normally be necessary to add the authors' names, title, text etc., which may also be done in CIF-format; this is followed by the edited contents of one or more *.cif* files each describing one structure (or possibly the same structure at different temperatures etc.). In general SHELXL provides all the CIF identifiers required by Acta Cryst. except those that begin with '_publ'. Further details are given below, and an example of a paper submitted to Acta Cryst. in this way may be found in the file *example.cif* (it has been brought up to date for the 1997 requirements for authors; whether it would pass the new stricter quality controls is another matter!). SHELXL users are strongly recommended to familiarize themselves with the definitive paper by the I.U.Cr. Commission on Crystallographic Data by Hall, Allen & Brown (1991), and with the current Acta Crystallographica Instructions for Authors.

Since the archiving of macromolecular data in CIF format is still being debated, SHELXL only creates a standard 'small molecule' CIF file, suitable for Acta Cryst. etc.; a macromolecular CIF file is likely to contain much more information. However the LIST 6 instruction in the new version of SHELXL does produce a CIF format reflection data file suitable for archiving with the PDB. This file also contains all the information necessary for the calculation of electron density maps, though as yet it appears that no standard macromolecular graphics package is able to read CIF format. Macromolecular coordinates etc. should be deposited in PDB format; SHELXPRO provides the necessary facilities for extending the *.pdb* file produced by SHELXL so that it can be used as a template for deposition.

10.2.2 The auxiliary program CIFTAB

CIFTAB is a simple program that reads CIF files and convert them into tables. It may prove useful for padding out Ph.D. theses and for submission of table to old-fashioned journals.

CIFTAB in *WinGX* is started from the menu item Publish-CIF Tools-CIFTAB :
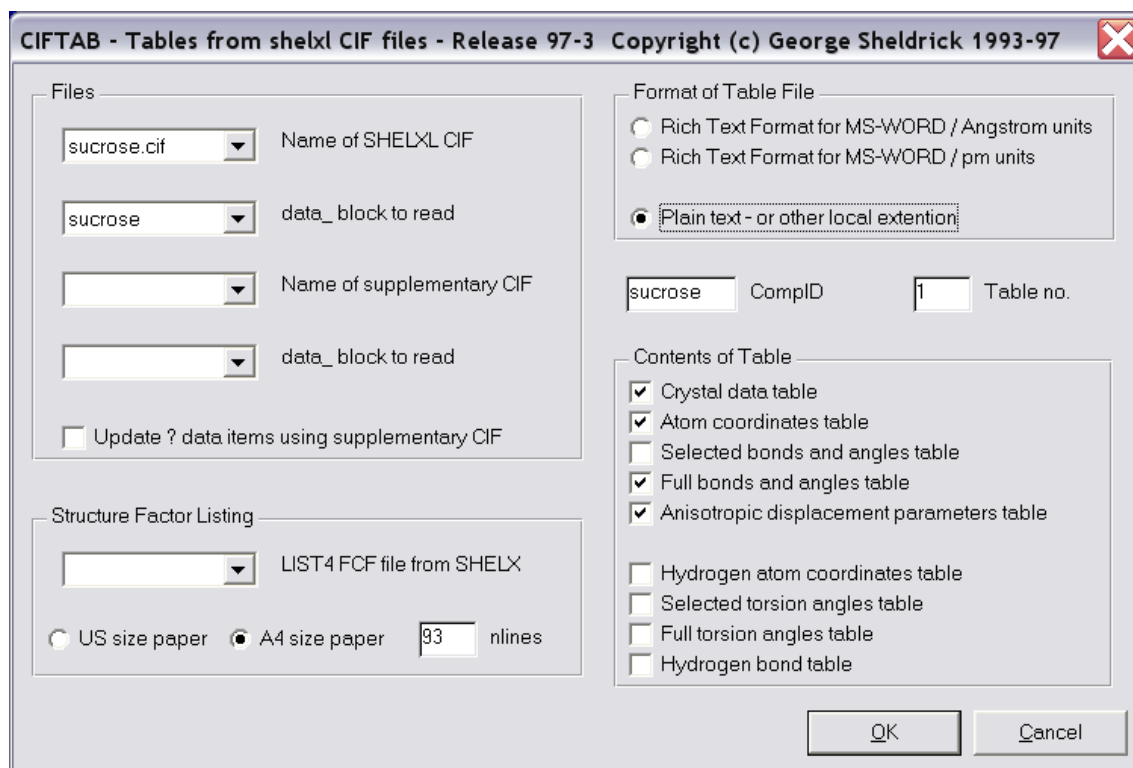
CIFTAB enables tables to be produced from the *.cif* or *.fcf* files written by SHELXL and provides the following facilities, which may be selected from a simple menu.

Tables of crystal data, atom parameters, bond lengths and angles, anisotropic displacement parameters and hydrogen atom coordinates may be produced in a format specified in a file *ciftab.???* (where *???* is any three letter combination). A standard ASCII file *ciftab.def* is provided; users may use it as a model for preparing standard ASCII tables files for input to word processors etc.

The format file is simply copied to the output file, except that directives (lines beginning with '?' or '$') have a special meaning, '\n\' (where n is a number) is replaced by the ASCII character n (e.g. \12\ starts a new page), and CIF identifiers (which begin with the character '_') are replaced by the appropriate number or string from the CIF file. CIF identifiers may optionally be followed (without an intervening space) by one or more of: '<n', '>n', ':n' and '=n' where n is an integer; the CIF identifier (including qualifier) must be terminated by one space that is not

copied to the output file. '<n' left justifies the CIF item so that it starts in column n, and is usually used for strings. '>n' right justifies a string or justifies a number so that the figure immediately to the left of the decimal point appears in column n; if there is no decimal point then the last digit appears in column n. In either case the standard deviation (if any) extends to the right with brackets but without intervening spaces. If '<n' and '>n' are both absent, the CIF item is inserted at the current position. If ':n' is absent the item is treated as a string (see above), otherwise it is treated as a number; n is the power of 10 with which the CIF item should be multiplied, and is useful for converting Å to pm or printing coordinates as integers; n may be negative, zero or positive. '=n' rounds the CIF item (after application of ':n') so that there are not more than n figures after the decimal point; n must be zero or positive.

A line beginning with 'loop_' is repeated until the corresponding loop in the CIF file is exhausted; all the CIF items in the line must be in the same loop in the CIF input file

A line containing at least 4 consecutive underscores is copied to the output file unchanged, and may be used for drawing a horizontal line. There are also two pseudo-CIF-identifiers: '_tabno' is the number of the table, and '_comno' is a number or text string to identify the compound. Both may be set via the CIFTAB menu. '_tabno' but not '_comno' is incremented each time it is used.

An underscore '_' followed by a space may be used to continue on the next line without creating a new line in the output file. Lines beginning with question marks are output to the console (without the leading question mark) as questions; if the answer to the question is not 'Y' or 'y', everything in the format file is skipped until the next line which begins with a question mark. Lines beginning with a dollar '$' are not interpreted as text, but are scanned for the following strings (upper or lower case, quotes not essential):

 'xtext': output should be formatted for the Siemens SHELXTL XCIF program (which now incorporates XTEXT, which was a separate program in version 4 of SHELXTL).

 'xtext,deutsch': as above, but translated into German.

The above directive, if present, should be the first line of the format file.

The directive $symops:n, where n is an integer, prints the symmetry operations used to generate equivalent atoms, starting each line of text in column n. These operators are referenced by '#m' (where m is an integer) after the atom name. The line beginning '$symops:n' usually follows the tables of selected bond lengths and angles, torsion angles and hydrogen bonds.

The remaining directives may appear at any point in the format file except immediately after a continuation line marker, but always on a line beginning with '$'.

'h=none': leave out all hydrogen atoms.

'h=only': leave out all non-hydrogen atoms.

'h=free': leave out riding or rigid group hydrogens but include the rest.

'h=all': include all hydrogen and all other atoms.

The hydrogen atom directives apply only to tables of coordinates; hydrogen atoms are recognized by the .._type_symbol 'H'. A common user error on writing format files is to forget that 'h=only' etc. applies until it is replaced by another 'h=...' directive!   The publication flags can be used to control which hydrogen atoms appear in tables of bond lengths, angles etc.

'brack': Atom names should include brackets (if present in the CIF file).

'nobrack': Brackets are deleted from the atom names.

'flag': Only output items for which the publication flag is 'Y' or 'y'.

'noflag': Output all items, ignoring the publication flag.

The default settings are '$h=none,brack,flag'. The standard tables file *ciftab.def* illustrates the use of most of these facilities. CIFTAB extends some of the standard CIF codes to make them more suitable for tables, and also takes special action when items such as _refine_ls_extinction_coef are missing or undefined.

The above description refers to the version of CIFTAB distributed with SHELXL. The simplest method of altering the contents and format of results tables is to create a different ciftab.??? format file (or a collection of such files for various purposes), using the standard file ciftab.def as a starting model. Thus the output can be tailored to different journals, doctoral theses, reports, etc.

The more ambitious user may wish to make some changes in the CIFTAB program itself, to incorporate additional options not provided by the program as distributed. The flexibility of the format file, however, provides most of the facilities that are likely to be needed, and the standard CIFTAB does include a procedure for replacing undefined data items by values taken from one or more other files conforming to CIF rules. Thus items such as diffractometer or area detector operating parameters, details of absorption corrections, and crystal color,

which are unknown to SHELXL, can be incorporated from separate files. This is more reliable than using a text editor.

10.2.3 Using SHELXL CIF files for publication in Acta Cryst

The process of converting a virgin SHELXL CIF output file into an electronic manuscript submission for Acta Cryst. Section C may seem at first rather complex and daunting, but the journal's Instructions for Authors are very detailed, and much of the conversion is routine and can be semi-automated; it can soon become an accustomed habit!

The important first step is to be properly informed of what is involved. The I.U.Cr. makes a variety of useful information available, and it can conveniently be accessed in its most up-to-date form at the World Wide Web location http://www.iucr.ac.uk/welcome.html by any standard Web browser. Printed Instructions for Authors can be found each year in the journal itself, and copies are available on request from The Managing Editor, I.U.Cr., 5 Abbey Square, Chester CH1 2HU, England. The Chester office can also supply copies of a technical account of how a CIF becomes a printed paper (reprinted from McMahon, 1993), and of 'A Guide to CIF for Authors' (published in 1995).

For a manuscript describing a single structure, the SHELXL CIF output needs only the addition of a well-defined set of publication information (the items that begin with '_publ'), itself in correct CIF format. A template for this can be obtained by ftp from I.U.Cr., and the SHELXL CIF output is attached to the end of it. Into the template are inserted (by any standard text editor) items such as manuscript title, authors' names and addresses, descriptive text, some extra experimental details as necessary (such as chemical synthesis and crystallization details, and a description of hydrogen atom refinement procedures), literature references, acknowledgments, and figure captions. There is also a place for inserting a formal submission letter. Some parts of the SHELXL output need changing; in particular, bond lengths and angles to be printed in the journal must be identified by changing their publication flag from '.?' to 'yes'.

When the CIF appears to be ready for submission, its completeness and validity can be checked anonymously by e-mailing it to the address checkcif@iucr.ac.uk; a report will be automatically generated and returned by e-mail listing and CIF syntax errors and any unrecognized data items. If there are no errors, the file is also checked for completeness and for some aspects of self-consistency (geometry is checked against coordinates, and possible higher symmetry is searched for). Any errors or omissions should be corrected and the checkcif procedure repeated, until everything is correct.

Beware of adding anything to e-mailed CIF submissions which does not accord with the syntax rules. In particular, there must be no non-CIF lines at the beginning or

end of the message, and this includes automatically appended e-mail signatures! These should be disabled or, safer, set up such that every line begins with the # character, which signals a CIF comment line to be ignored.

There is also a facility for previewing a manuscript in the form which will be produced from the CIF. Sending the CIF by e-mail to printcif@iucr.ac.uk will produce, as a reply message, a PostScript file of the manuscript; this can be printed or viewed by appropriate software. A useful feature is the highlighting (in bold) of any items which may subsequently be queried by editorial staff, and it may be possible to deal with these potential problems now, before final submission.

When everything is ready and checked, the CIF is e-mailed to med@iucr.ac.uk; after automatic checking is complete, a reply will list any problems requiring attention, will give a Co-Editor reference, and will ask for further material to be sent. This includes structure factor data, figures (diagrams), a copyright transfer form, and a formal signed letter of submission. The last two must still be sent by normal mail, but the others can be transferred electronically (ftp), using the method specified in the Instructions for Authors and the submission acknowledgment. None of these items should be sent until the acknowledgment and reference code arrive.

If these instructions are followed carefully, the editorial process should proceed smoothly! I am grateful to Bill Clegg for writing much of this chapter.