
Using the GSAS2CIF program to export [GSAS](#) results



[The crystallographic information file, \(CIF\)](#) was developed by the [IUCr](#) in the early 1990's as a standardized format to document single-crystal structure determinations and to exchange the results between laboratories. In the late 1990's, the [pdCIF](#) dictionary was developed to allow CIF to document powder diffraction measurements, as well as Rietveld results. A very short introduction to CIF is included at the [end of this document](#).

The GSAS2CIF program is used to create CIFs from GSAS results. This web page documents concepts within the program as well as how the program is used.

Overview of the steps to create a CIF in GSAS

1. Complete (*ha!*) your refinement, being sure to run GENLES after making any changes to the .EXP file with EXPEDT/EXPGUI.
2. Run DISAGL to compute interatomic distances and angles.
3. (Optional) Edit the "[publication flags](#)" in the distance and angle listings.
4. (Optional) Use EXPGUI program [FillTemplate](#) to customize copies of CIF template files in the current directory, or manually copy these files to the current directory and use a text editor.
5. Run GSAS2CIF (this creates experiment-specific copies of the template files).
6. Edit the template files (for example, using [FillTemplate](#)) to include information.
7. Run GSAS2CIF again, to incorporate the revised versions of the template files into the complete .CIF file.
... and if you are like me...
8. try out a few more ideas in GENLES.
9. Run DISAGL.
10. Run GSAS2CIF
... repeat steps 8-10 until really finished... ([see quote](#)).

Structure of the GSAS2CIF program

The GSAS2CIF program is used to prepare CIF files containing results from GSAS. It should be noted many important types of information that are needed to describe the sample and the measurement are not supplied as input to GSAS. A few examples of information that should likely be included in a CIF, but are not defined within GSAS are: the data measurement temperature, sample preparation conditions, publication information, and so on. If the CIF will be used as supplementary material, to accompany a publication, this sort of documentary information certainly should be supplied, so that the CIF has value for archival and database purposes. If a CIF is being prepared for submission to an *Acta Crystallographia* journal, the IUCr has [template with a recommended list of CIF data items](#). This template has been utilized for creation of the template files distributed with GSAS/EXPGUI.

Template Files

Since the GSAS2CIF program cannot obtain much of the information needed to create a CIF that documents a structural determination, this information must be supplied by the person running the program in the form of template files. The GSAS2CIF program simply copies this information from template files into the CIF created by GSAS2CIF. Three separate template files are used with GSAS2CIF:

1. one with publication and other overall information,
2. one with information about the sample & specimen and
3. one with information about diffraction instrumentation and data collection parameters.

Note that as distributed, these files do not contain any information, but rather define a set of suggested fields where the user can provide this information. Since much of the information in these fields will be the same for all CIF's prepared by a researcher or using a specific instrument, it will be a good idea to create customized versions of these template files and thus avoid inputting the same information multiple times.

When GSAS2CIF is used to create a CIF file for an experiment named *exnam* (e.g. from file *exnam*.EXP), the GSAS2CIF program creates CIF file *exnam.cif* containing GSAS results. Information from a series of template files is copied directly into the CIF. There will be $N+M+1$ template files, where N is the number of phases and M is the number of data histograms. The files are named as follows:

1. *exnam_publ.cif* for the publication/overall information template file;
2. *exnam_phasen.cif* for the N sample/specimen template file(s);
3. *exnam_INSTmm.cif* for the the M instrument/data sample/specimen template file(s), where *INST* is the instrument name ([see below](#)).

If these files do not exist, they are created and filled with the contents of master versions of the template files. In the case of the *exnam_publ.cif* and *exnam_phasen.cif* files, template files named *template_publ.cif* and *template_phase.cif* are read, if present from the same directory where the *exnam*.EXP file is found, and if not present there, from the GSAS data directory. In the case of the *exnam_INSTmm.cif* file(s), the program first looks for files named *template_INST.cif* in the current directory and the GSAS data directory and if that file is not found, file *template_instrument.cif* is read from the current directory and if not found, the GSAS data directory. This somewhat complex series of template files allows for the creation default template files for commonly-used instruments as well as the potential for reuse of the other template files, by copying these files as needed and also allows the GSAS2CIF program to be reused to update the CIF as needed without loss of input information.

Note, that users should avoid editing the final *exnam.cif* file, but rather should edit the *exnam_*.cif* template files and then

rerun GSAS2CIF to incorporate the revised template files into a new version of the *exptom.cif* file. In this way, if GSAS2CIF is rerun at a later time, the crystallographic results in the .CIF are updated and the template information is retained automatically. This editing can be done with any ASCII text editor, but an application has been included within EXPGUI, [FillTemplate](#), for this purpose.

Other Input Files

In addition to the reading the GSAS experiment file (file *exptom.EXP*), GSAS2CIF also reads the variance-covariance matrix created in GENLES (from file *exptom.CMT*) and a table of interatomic distances and angles created from program DISAGL (file *exptom.DISAGL*). If these files cannot be read, GSAS2CIF produces a warning message, since the CIF will be incomplete without this information. Note that the .CMT and .DISAGL are both derived from GENLES output and thus the .EXP file must have been used to run GENLES and then DISAGL just prior to running GSAS2CIF. If you edit the .EXP file or go back to an archived experiment file and do not rerun GENLES and then DISAGL, (*exptom.Onn*), the most recent *exptom.CMT* and *exptom.DISAGL* will be out-of-sync with the .EXP contents. GSAS2CIF will try to spot this and warn you, but some errors may be hard to catch.

Instrument Name

An instrument name is needed for every GSAS histogram. It is best if this name is unique to a specific instrument, so for commercial instruments, it is best if this name contains part of the instrument serial number or the institution name, etc. The instrument name may be defined in the instrument parameter file, by inclusion of a record of type "INS nn INAME *Instrument name*". If this name is not defined in the original instrument parameter file, when GSAS2CIF is run, it will request an instrument name for each histogram, and this information will be added to the GSAS experiment file. Note that the vertical bar character, (|), should not be used in instrument names.

Publication/Non-Publish Flag for Distances and Angles

The DISAGL program will tabulate all interatomic distances within specified interatomic radii. These radii may be modified using EXPEDT (but not at present EXPGUI). All distances and angles listed by DISAGL in the *exptom.LST* file will be recorded in a more compact format in file named *exptom.DISAGL*. All these distances and angles will then be included in the CIF file when GSAS2CIF is run.

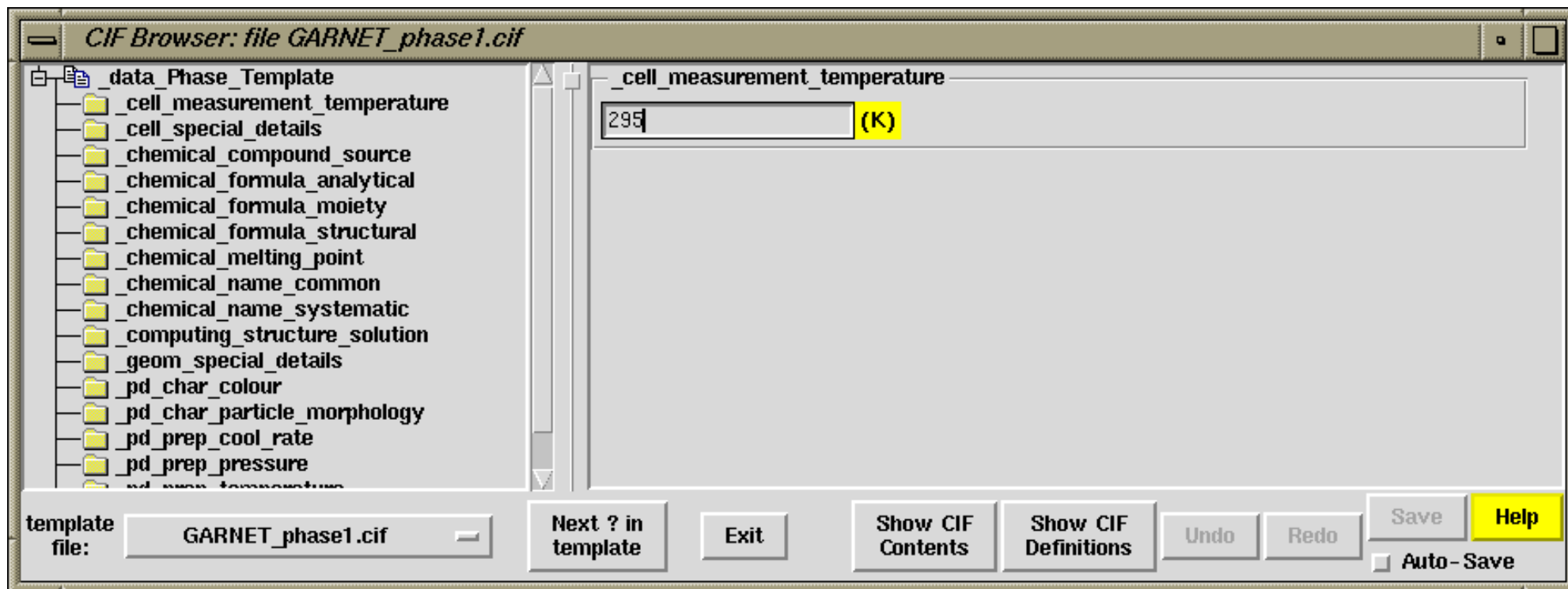
Note that when an IUCr journal publication is prepared from a CIF, these distances and angles are abstracted directly into the publication tables from the CIF, provided that a special CIF flag for each distance/angle (`_geom_bond_publ_flag` and `_geom_angle_publ_flag`) indicates that these values should be published. The character in the first column of the *exptom.DISAGL* indicates the value for this flag, when the distances/angles are tabulated in the CIF. If this character is anything other than "Y" or "y", the distance/angle is flagged as "do not publish". A value of "Y" or "y" indicates this flag should be set to "publish." At present, the only way to change this flag is manually, with a text editing program, such as wordpad or emacs. An EXPGUI program, [CIFSelect](#), is available for this chore.

When DISAGL is first run, this flag character is set to N (do not publish) for all distances and angles. However, if any of these flags are changed in the .DISAGL file, these flag settings will be retained if DISAGL is rerun.

The FillTemplate program

The EXPGUI FillTemplate program is used to edit the CIF template files used within GSAS2CIF. The *expnam_*.cif* files are edited by this program, if they exist. If these files are not found (because GSAS2CIF has not been run), the *template_*.cif* in the current directory are edited. If these files are not found as well, the *template_*.cif* in the GSAS data directory are copied to the current directory and are then made available for editing.

The FillTemplate program is accessed in EXPGUI from the "CIF Export" submenu of the "Import/Export" menu. The FillTemplate program opens a window, as is shown below:



The box on the left side of the window displays the data names in the CIF as a hierarchical tree. Note that subentries can be displayed or hidden by clicking on the +/- sign to the right of each grouping. For example, by default the data names contained in each loop are not shown. However, clicking on the + sign to the right of the "loop_0" listing causes the data names in the loop to be shown (as in the [example below](#).)

Clicking on an individual data name causes the associated value to be displayed on the right. The value associated with the data name can be edited by clicking on the entry box. The information you enter is copied into the in-memory copy of the CIF when you click on another data name, etc.

In the example shown above, the data name `_cell_measurement_temperature` has been selected. This is defined in the CIF dictionary as a number between 0 and infinity with units of Kelvins. These units (K) are displayed adjacent to the entry box. When appropriate input is validated to require that valid numbers in the allowed range are input or that standard uncertainties (esd's) are entered only where allowed. Likewise, if the CIF dictionary defines a enumerated list of values for a

data name, a menu button is offered in place of an entry box. In this way, only a valid entry from the list can be selected.

The controls on the bottom of the window have the following effects:

Template file

The menu button labeled "Template file" offers a list of template files. This button can be used to select a file to be edited. If changes have been made to the current CIF, that have not been saved to disk, the user is offered a chance to save or discard these changes.

Next ? in template

If the button labeled "Next ? in template" is pressed, the program scans forward in the current file, and if needed through subsequent files, for a data item where the value is a single question mark (?), since this value indicates a item where a value has not been specified. Note that the `_journal_` data items, which are intended to be used by *Acta Crystallographia* editors, are ignored by this button.

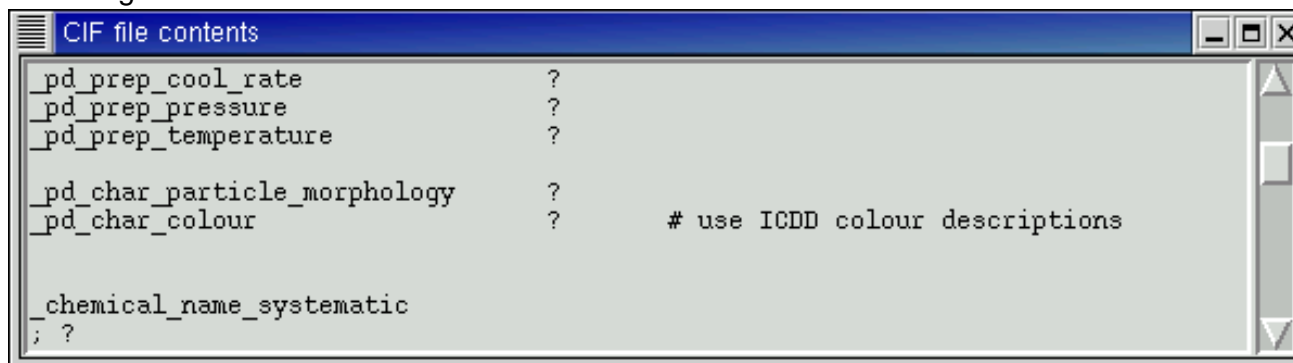
Exit

The "Exit" button causes the program to exit. If any changes have been made to a data item or the CIF has been changed, but have not been saved, a chance is offered to save these items.

Show (Hide) CIF contents

The "Show CIF contents" button causes a window to be displayed that shows the text of the CIF, as shown below. As CIF data items are selected by clicking on data names or through use of the other buttons, the window is scrolled forward or backward to show the appropriate section. Note that it is possible to make editing changes directly to the CIF using this window, but the program cannot verify that these edits have the correct syntax. Further, the program will not note that changes have been made to the CIF so another section of the file must be changed, before the edits can be saved to disk (see the [Save](#) button).

After the "Show CIF contents" button is pressed, the label changes to "Hide CIF contents"; pressing the button again causes the window to be hidden.



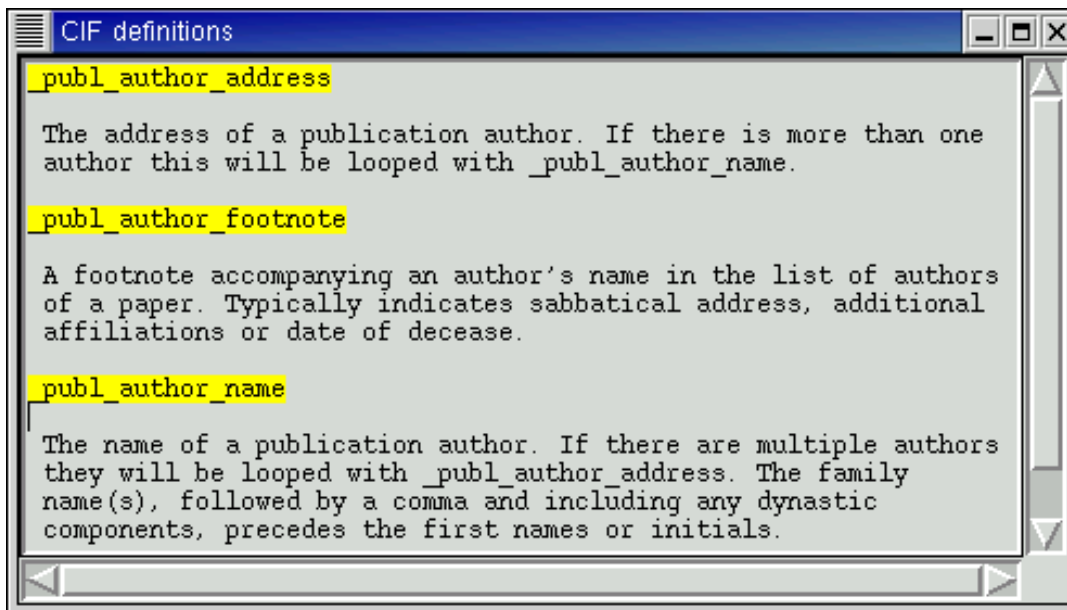
```
CIF file contents
_pd_prep_cool_rate      ?
_pd_prep_pressure      ?
_pd_prep_temperature    ?

_pd_char_particle_morphology  ?
_pd_char_colour         ?      # use ICDD colour descriptions

_chemical_name_systematic
; ?
```

Show (Hide) CIF definitions

As CIF data names are selected, their definitions are shown in the CIF Definitions window, as shown below. After the "Show CIF definitions" button is pressed, the label changes to "Hide CIF definitions"; pressing the button again causes the window to be hidden.



Undo

As changes are made to the CIF template, they are recorded and can be reversed using the "Undo" button. There is no limit to the number of changes that are recorded. However, changes cannot be undone after the CIF has been saved to disk.

Redo

If changes have been reversed with the ["Undo"](#) button, the changes can be reapplied using the "Redo" button. The list of changes available for "Redo" is cleared when a new edit is made or when the CIF is saved.

Save

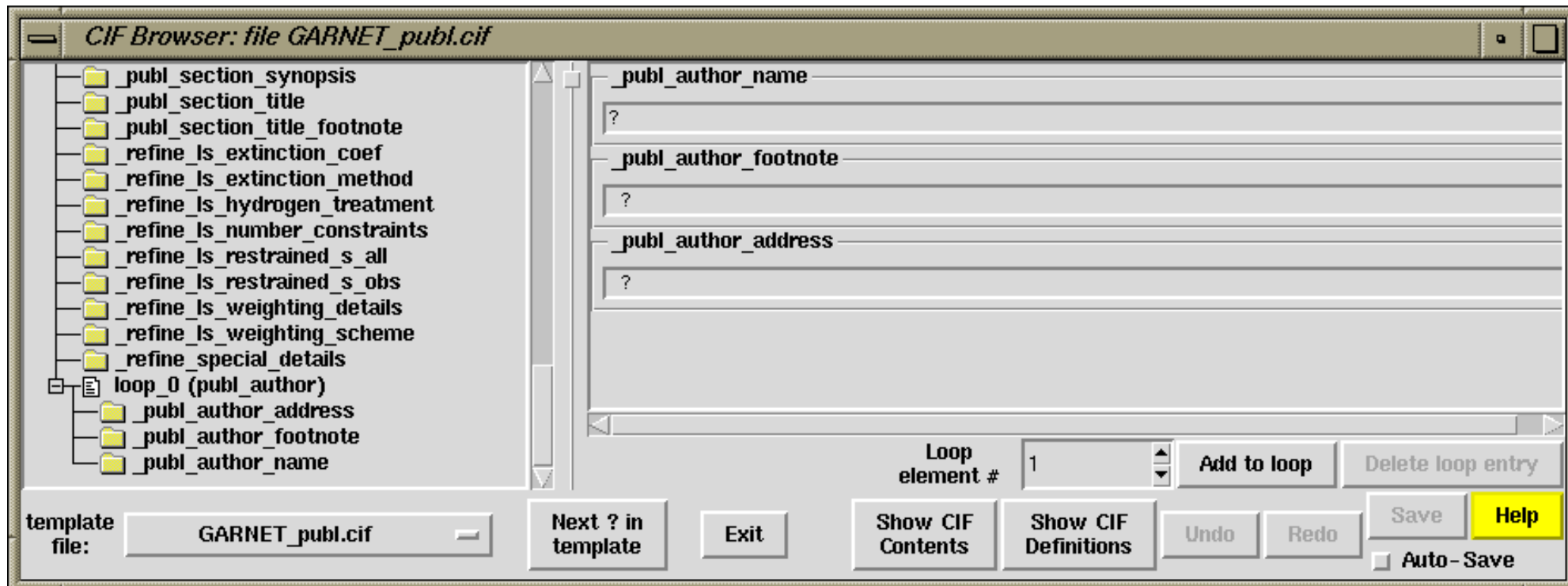
Changes made to the CIF are not saved to the disk file automatically, unless the [Auto-Save](#) checkbox is set. When changes have been made, but not saved to disk, this button is made active.

Auto-Save

When the Auto-Save mode is not selected, changes made to the in-memory copy of the CIF that is displayed in the [CIF Contents window](#) are not written to the disk file until the ["Save"](#) button is pressed. If the Auto-Save mode is selected, these changes are saved to disk automatically.

CIF loops

CIF loops allow multiple values to be associated with one or more data items, in effect defining a table of data. Clicking on a loop, causes all the data names in the loop to be displayed, as is shown below.



When a loop is displayed, extra controls appear, as are defined below.

Loop element

The Loop element #" spinbox" is used to select which "row" from the loop is displayed. The up arrow advances to the next row, while the down arrow reverses by one entry. Numbers can also be typed into the entry box; the number is accepted when Enter is pressed. The keyboard up and down arrows can also be used to advance between entries. Other keys such as Page Up, Home, etc. advance in large increments.

Add to loop

A new row can be added to the end of a loop using the "Add to loop" button. The value for each new entry is initialized as "?" (meaning value unknown or unspecified.)

Delete loop entry

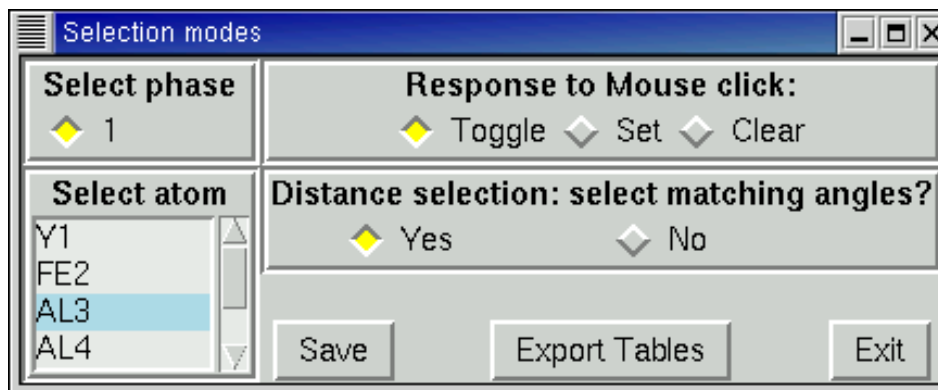
This deletes the current row from the loop. First select the row to delete with the "[Loop element #](#)" spinbox. Note that the values are displayed for confirmation before the delete operation is performed. It is not possible to delete all entries from a loop, so this button is disabled when a loop has only a single row defined.

It is also possible to click on the data name for a item inside a loop, in this case, all entries for that data item in the loop are displayed (a column).

The CIFSelect utility

The CIFselect utility is available within EXPGUI to view interatomic distances and angles generated by DISAGL and to set the CIF publication flags for these values. CIFselect is accessed in EXPGUI from the "CIF Export" submenu of the "Import/Export" menu. CIFselect opens two windows, one for controlling the program and one to display the distances.

The control window is shown to the right. The buttons on this window are described below.



Select Phase

This provides a list of phases.

Select Atom

This provides a list of atoms for the current phase.

Response to Mouse Click

When the mouse is clicked on a distance or angle (see below), the publication flag for the may be changed. The effect is dictated by the mode selected here. In **"Toggle"** mode the publication flag is set to the opposite of the previous value. In **"Set"** mode, the publication flag will be set to "Y" (publish), regardless of the previous state of the flag. In **"Clear"** mode, the publication flag will be set to "N" (do not publish), regardless of the previous state of the flag.

Distance selection: select matching angles?

When this option is set to Yes, the publication flag will be set to "Yes" for all angles where both the publication flags for the distances are also set to "Yes". This is illustrated in the window shown below. Note that the distance AL3-O6_a is set to be not published. The five angles that involve atom O6_a were automatically set to be not published. If this option is set to Yes, turning on the publication flag for distance AL3-O6_a would cause the publication flag to be set on as well. If this option is set to "No", angle flags must be changed individually.

Save

This button causes the current flag settings to be saved in the *expnam.DISAGL* file.

Export Tables

This button creates a file with all bond distances and angles for the current phase, formatted in the arrangement used in the window below. Each entry is separated by a comma, suitable for reading into spreadsheet programs.

Exit

This button causes CIFselect to exit. If there are unsaved changes to publication flags, you will be given an opportunity to save or discard these changes.

The distance and angle display window is shown below. The items that are selected for publication are highlighted in yellow.

Distances and angles around atom AL3						
	distance	O6_a	O6_b	O6_i	O6_j	O6_k
O6_a	1.9832(7)					
O6_b	1.9832(7)	85.47(3)				
O6_i	1.9832(7)	85.47(3)	85.47(3)			
O6_j	1.9832(7)	179.966	94.53(3)	94.53(3)		
O6_k	1.9832(7)	94.53(3)	179.966	94.53(3)	85.47(3)	
O6_l	1.9832(7)	94.53(3)	94.53(3)	179.966	85.47(3)	85.47(3)
Y1_a	3.40697(5)					
Y1_b	3.40697(4)					
Y1_e	3.40697(4)					
Y1_f	3.40697(4)					
V1_a	3.40697(5)					

Acknowledgments

The GSAS2CIF program was written by Brian H. Toby, Robert B. Von Dreele and Allen C. Larson. The distance/angle display format for CIFselect was suggested by Lachlan Cranswick, based on SHELX.

Richard L. Harlow first got me interested in the problem of a universal file format for powder diffraction data, leading eventually to my involvement with CIF and then this programming effort. For all this, I may someday forgive him.

Quote:

"A Rietveld refinement is never perfected, merely abandoned." Peter Stephens

Appendix: A quick & incomplete introduction to CIF

A CIF file consists of logical groups of information that are called data blocks, since each block is initiated by a label of form *data_label*. In the simple case, where a single crystallographic model is determined from a single diffraction dataset (histogram), the CIF can be a single block. In the case where either multiple datasets (histograms) or where multiple phases are used in the refinement, a CIF will require several data blocks to describe the data and results.

CIF consists of a series of tags, called data names, and values associated with these data names. Together the data name and value are called a data item. A separate document, the CIF (or pdCIF) dictionary defines the meaning of each data name. If a value does not contain any spaces, may be specified without quotes, but either single or double quotes may be used to delimit strings. A data value of one or more lines is quoted with semicolon characters (;). The two semicolons must be the first character on each line and the final semicolon is expected to be followed by a white-space character such as a space or then end-of-line. With the exception of semicolon delimiters, which must be the first character of a line to be

recognized, CIF treats multiple spaces, blank lines and other "white-space" characters as a single space.

The following lines give examples of a few CIF data items.

```
_pd_calc_method    'Rietveld Refinement'

_cell_volume      1811.00(5)

_symmetry_space_group_name_H-M  "I a 3 d"

_pd_proc_ls_background_function
;  GSAS Background function number 1 with 4 terms.
  Shifted Chebyshev function of 1st kind
    1:    139.025      2:   -11.5408      3:    9.75652      4:    3.90497
;

```

CIF also allows multiple values to be associated with a CIF data item. This is done by preceding the data name with the keyword `loop_`. If two or more data names follow the `loop_` keyword, a table can be constructed, as is shown in the following examples.

```
loop_  _symmetry_equiv_pos_as_xyz
      +x,+y,+z      +z,+x,+y
      +y,+z,+x      +x+1/2,+y,-z+1/2

loop_
      _atom_site_label
      _atom_site_fract_x
      _atom_site_fract_y
      _atom_site_fract_z
Y1      0.125      0.0      0.25
FE2      0.0      0.0      0.0
AL3      0.375      0.0      0.25
O4      -0.02946(5)  0.05385(5)  0.15068(6)

```

Finally, it should be pointed out that two values in CIF have special meanings. If a value is supplied as a single period (`.`), the meaning is the value is not defined or is inappropriate. If the value is a question mark (`?`), this means that the value is unknown or not specified.

For more information, see: [The CIF home page](#); short intros to [pdCIF](#) & [CIF syntax](#)

[EXPGUI top](#) [Previous page](#)

[Comments, corrections or questions: crystal@NIST.gov](mailto:crystal@NIST.gov)

Last modified 19-October-2002 \$Revision: 1.4 \$ \$Date: 2002/09/05 21:18:57 \$