1 mmCIF Syntax

The syntax used in mmCIF data files and dictionaries is derived from the STAR (Self-defining Text Archive and Retrieval) grammar and is simlar in most respects to the syntax which is used for the core (*small molecule*) CIF.

In its simplest form an mmCIF file looks like a paired collection of data item names and values. For instance, in the following example of assigning values to cell constants the interpretation of the syntax is obvious.

```
_cell.entry_id                          T100A
_cell.length_a                          68.39
_cell.length_a_esd                       0.05
_cell.length_b                          88.70
_cell.length_b_esd                       0.12
_cell.length_c                          76.27
_cell.length_c_esd                       0.06
```

mmCIF data item names can be identified by the leading underscore character. The underscore is followed by a text string which is interpreted in mmCIF as containing both a category name and a keyword name separated by a period. The keyword portion of the name is the unique identifier of the data item within the category. In the example above, all of the data items belong to the `CELL` category. The above example also illustrates the one-to-one correspondence required between item names and item values. Data category and data item names are not case sensitive. Data names and data values should be arranged within a file such that no line length exceeds 80 characters.

The next example illustrates the how text strings are expressed in mmCIF. Short text strings may be enclosed in single or double quotation marks. Text strings which span multiple lines are enclosed by semi-colons placed the first character position of the line. There are two special characters used as placeholders for mmCIF item values which for some reason cannot be explicitly assigned. The question mark (`?`) is used to mark an item value as missing. A period (`.`) may be used to identify that there is no appropriate value for the item or that a value has been intentionally omitted.

```
_diffrn_measurement.diffrn_id          'Data set 1'
_diffrn_measurement.device             '3-circle camera'
_diffrn_measurement.method             'omega scan'
_diffrn_measurement.details
; 440 frames, 0.20 degrees, 150 sec, detector distance 12 cm, detector
  angle 22.5 degrees
;
_diffrn_measurement.specimen_support    ?
```

Vectors and tables of data may be encoded in mmCIF using a loop_ directive. To build a table, the data item names corresponding to the table columns are preceded by the loop_ directive, and followed by the corresponding rows of data. The following example builds a table of author names.

```
loop_
_citation_author.citation_id
_citation_author.ordinal
_citation_author.name
  primary  1  'Fitzgerald, P.M.D.'
```

```
  primary  2  'McKeever, B.M.'
  primary  3  'Van Middlesworth, J.F.'
  primary  4  'Springer, J.P.'
  primary  5  'Heimbach, J.C.'
  primary  6  'Leu, C.-T.'
  primary  7  'Herber, W.K.'
  primary  8  'Dixon, R.A.F.'
  primary  9  'Darke, P.L.'
  2        1  'Navia, M.A.'
  2        2  'Fitzgerald, P.M.D.'
  2        3  'McKeever, B.M.'
  2        4  'Leu, C.-T.'
  2        5  'Heimbach, J.C.'
  2        6  'Herber, W.K.'
  2        7  'Sigal, I.S.'
  2        8  'Darke, P.L.'
  2        9  'Springer, J.P.'
```

The use of the `loop_` directive in mmCIF has a few restrictions. First, it is required that all of the data items within the loop belong to the same mmCIF category. Second, the number of data values following the loop must be an exact multiple of the number of data item names. Finally, mmCIF prohibits the nesting of `loop_` directives.

mmCIF uses data blocks to organize related information and data. A data block is a logical partition of a data file or dictionary created using a `data_` directive. A data block may be named by appending a text string after the `data_` directive and a data block is terminated by either another `data_` directive or by the end of the file. The following example shows a very simple example of a pair of abbreviated data blocks.

```
#
# --- Lines beginning with # are treated as comments
#
data_X987A
_entry.id                              X987A
_exptl_crystal.id                   'Crystal A'
_exptl_crystal.colour               'pale yellow'
_exptl_crystal.density_diffrn       1.113
_exptl_crystal.density_Matthews     1.01

_cell.entry_id                         X987A
_cell.length_a                         95.39
_cell.length_a_esd                      0.05
_cell.length_b                         48.80
_cell.length_b_esd                      0.12
_cell.length_c                         56.27
_cell.length_c_esd                      0.06

# Second data block
data_T100A

_entry.id                              T100A
_exptl_crystal.id                   'Crystal B'
_exptl_crystal.colour               'orange'
_exptl_crystal.density_diffrn       1.156
_exptl_crystal.density_Matthews     1.06
```

```
_cell.entry_id                          T100A
_cell.length_a                          68.39
_cell.length_a_esd                       0.05
_cell.length_b                          88.70
_cell.length_b_esd                       0.12
_cell.length_c                          76.27
_cell.length_c_esd                       0.06
```

The above example illustrates how data blocks can be used to separate similar information pertaining to different structures. This separation is required because the mmCIF syntax prohibits the repetition of the same category at multiple places within the same data block. As a result, the simple concatentation of the contents of the above two data blocks into a single data block would be syntactically incorrect.

- Merging the data blocks in the above example raises some additional issues associated with the mmCIF data model and the structure of these specific categories. In above example, it would be possible to merge the information in the EXPTL_CRYSTAL category into a single data block by reorganizing this category using a loop_ directive. However, certain mmCIF categories like CELL and ENTRY may contain only a single value within the data block and therefore cannot be looped. The single-valued property of the data items in these categories is a consequence of the definition of the key items in these two categories. The category key for the CELL category, _cell.entry_id, is a defined as a child definition of _entry.id. This item is defined as the data block identifier and may therefore assume only a single value.

Definitions in the mmCIF dictionary are encapsulated in named save frames. A save frame begins with the save_ directive and is terminated by another save_ directive. Save frames are named by appending a text string to the save_ token. In mmCIF dictionaries save frames are used to encapsulate item and category definitions. The mmCIF dictionary is composed of a data block containing thousands of save frames, where each save frame contains a different definition. Save frames may only appear in mmCIF dictionaries and they may not be nested. The following example shows the save frame containing the definition of the data item _exptl.details.

```
save__exptl.details
    _item_description.description
;               Any special information about the experimental work prior to the
                intensity measurement. See also _exptl_crystal.preparation.
;
    _item.name                   '_exptl.details'
    _item.category_id              exptl
    _item.mandatory_code           no
    _item_aliases.alias_name     '_exptl_special_details'
    _item_aliases.dictionary       cif_core.dic
    _item_aliases.version          2.0.1
    _item_type.code                text
     save_
```

- Save frames play a much more important role in STAR than in mmCIF. In a STAR file, a save frame acts as a reuseable cell of information which can be referenced and expanded within the file. Save frames are referenced in a STAR file by preceding the save frame name with a dollar sign. The use of save frames in mmCIF has been limited to the organization and scoping features that they provide. mmCIF does not support references to save frames or the use of save frames for purposes other than

for encapsulating dictionary definitions.

mmCIF Syntax

Table of Contents

1 mmCIF Syntax 1