Example 3.6.3.7. *Separate chemical components forming the crambin polypeptide.*

```
loop_
_chem_comp.id
_chem_comp.mon_nstd_flag
_chem_comp.formula
_chem_comp.name
  ethanol .   'C2 H6 O1'         "ethanol"
  ALA     yes 'C3 H7 N1 O2'      "alanine"
  ARG     yes 'C6 H14 N4 O2'     "arginine"
  ASN     yes 'C4 H8 N2 O3'      "asparagine"
  ASP     yes 'C4 H7 N1 O4'      "aspartic acid"
  CYS     yes 'C3 H7 N1 O2 S1'   "cysteine"
  GLU     yes 'C5 H9 N1 O4'      "glutamic acid"
  GLY     yes 'C2 H5 N1 O2'      "glycine"
  ILE     yes 'C6 H13 N1 O2'     "isoleucine"
  LEU     yes 'C6 H13 N1 O2'     "leucine"
  PHE     yes 'C9 H11 N1 O2'     "phenylalanine"
  PRO     yes 'C5 H9 N1 O2'      "proline"
  SER     yes 'C3 H7 N1 O3'      "serine"
  THR     yes 'C4 H9 N1 O3'      "threonine"
  TYR     yes 'C9 H11 N1 O3'     "tyrosine"
  VAL     yes 'C5 H11 N1 O2'     "valine"
```

example should be noted. The composite labelling of each site includes a pointer to the description of the parent molecule as a specific object in the asymmetric unit (`_atom_site.label_asym_id`) and to the relevant monomeric building block of which the atom is a member (`_atom_site.label_comp_id`). The label component `_atom_site.label_alt_id` indicates alternative conformations in which an atom site may be found. For example, the atom sites numbered 3 and 4 are alternative locations for the α-carbon of the terminal residue. It may be deduced from the occupancies that the alternative conformations A and B are modelled with 80% and 20% occupancy, respectively, but this can be stated explicitly using the ATOM_SITES_ALT category. The sequence heterogeneity at residue 22 is shown by the presence of pointers to proline and serine, and the occupancy factors show that proline and serine are present in the ratio 60 to 40. There is also an alternative conformation within the serine at residue 22, split equally across two sites.

3.6.4. Content of the macromolecular CIF dictionary

Because it is derived from the core CIF dictionary, the mmCIF dictionary shares the same general structure as outlined in Chapter 3.2. However, DDL2 permits the formal assignment of categories to *category groups*. Table 3.6.4.1 lists the major category groups in the mmCIF dictionary (a full list is given in Appendix 3.6.1 and at the beginning of Chapter 4.5).

Small capitals are used for the names of category groups and individual categories in this volume, but the identifiers in the dictionary are actually lower-case strings.

The ordering of category groups in the remainder of this chapter follows the thematic scheme of Table 3.1.10.1. The discussion proceeds under the headings *Experimental measurements* (Section 3.6.5), *Analysis* (Section 3.6.6), *Atomicity, chemistry and structure* (Section 3.6.7), *Publication* (Section 3.6.8) and *File metadata* (Section 3.6.9).

Certain conventions of style and layout have been followed to summarize the large amount of information in the mmCIF dictionary and to help the reader navigate their way through this chapter. Appendix 3.6.1 is an overview of the mmCIF dictionary structure by category and lists all the categories with the number of the section in which they are discussed. This acts as an index between the alphabetical ordering within the dictionary and the thematic ordering of this chapter. Each thematic section lists the

Example 3.6.3.8. *Partial listing of the atomic coordinates of crambin.*

```
loop_
_atom_site.label_seq_id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_alt_id
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.footnote_id
_atom_site.label_entity_id
_atom_site.id
1  N  N    THR chain_a A  16.864  14.059   3.442
   0.80  6.22  .   A    1
1  N  N    THR chain_a B  17.633  14.126   4.146
   0.20  8.40  .   A    2
1  C  CA   THR chain_a A  16.868  12.814   4.233
   0.80  4.45  .   A    3
1  C  CA   THR chain_a B  17.282  12.671   4.355
   0.20  7.82  .   A    4
1  C  C    THR chain_a .  15.583  12.775   4.990
   1.00  4.39  .   A    5
1  O  O    THR chain_a .  15.112  13.824   5.431
   1.00  7.04  .   A    6
1  C  CB   THR chain_a A  18.060  12.807   5.200
   0.80  5.42  .   A    7
1  C  CB   THR chain_a B  18.202  11.709   5.108
   0.20 11.07  .   A    8
1  O  OG1  THR chain_a A  19.233  12.892   4.380
   0.80  7.87  .   A    9
1  O  OG1  THR chain_a B  17.662  10.381   4.831
   0.20 14.39  .   A   10
1  C  CG2  THR chain_a A  18.117  11.578   6.092
   0.80  6.88  .   A   11
1  C  CG2  THR chain_a B  17.973  11.955   6.599
   0.20 19.74  .   A   12
# - - abbreviated - - -
22 N  N    PRO chain_a .   4.909  12.659  -3.127
   0.60  3.03  .   A  352
22 C  CA   PRO chain_a .   6.035  13.459  -2.622
   0.60  3.04  .   A  353
22 C  C    PRO chain_a .   6.362  13.139  -1.174
   0.60  3.08  .   A  354
22 O  O    PRO chain_a .   5.473  12.959  -0.323
   0.60  3.67  .   A  355
22 C  CB   PRO chain_a .   5.528  14.895  -2.825
   0.60  4.19  .   A  356
22 C  CG   PRO chain_a .   4.614  14.846  -4.059
   0.60  3.91  .   A  357
22 C  CD   PRO chain_a .   3.904  13.493  -3.885
   0.60  3.25  .   A  358
22 N  N    SER chain_a .   4.909  12.659  -3.127
   0.40  3.03  .   A  366
22 C  CA   SER chain_a .   6.035  13.459  -2.622
   0.40  3.04  .   A  367
22 C  C    SER chain_a .   6.362  13.139  -1.174
   0.40  3.08  .   A  368
22 O  O    SER chain_a .   5.473  12.959  -0.323
   0.40  3.67  .   A  369
22 C  CB   SER chain_a .   5.644  14.934  -2.679
   0.40  3.96  .   A  370
22 O  OG   SER chain_a C   4.712  15.250  -1.677
   0.20  3.53  .   A  371
22 O  OG   SER chain_a D   6.688  15.800  -2.315
   0.20  7.09  .   A  372
```

categories discussed in that section. Within each subsection, the data names within the relevant categories are listed. Category keys, pointers to parent data items and aliases to data items in the core CIF dictionary are indicated. For each category, the data item (or set of data items that must be considered together) that forms the category key is marked by a bullet (•) and listed first; the other data names follow in alphabetical order.

For measured or derived numerical quantities that should be specified with a standard uncertainty (in older terminology, an estimated standard deviation), the core dictionary uses the DDL1

Table 3.6.4.1. *Major category groups defined in the mmCIF dictionary*

The groups are listed in the order in which they are described in this chapter. There is also an INCLUSIVE category group, which serves as a formal higher-order container group to which all other category groups belong.

Section	Category group	Subject covered
(*a*) Experimental measurements		
3.6.5.1	CELL	Unit cell
3.6.5.2	DIFFRN	Diffraction experiment
3.6.5.3	EXPTL	Experimental conditions
(*b*) Analysis		
3.6.6.1	PHASING	Phasing techniques
3.6.6.2	REFINE	Refinement procedures
3.6.6.3	REFLN	Reflection measurements
(*c*) Atomicity, chemistry and structure		
3.6.7.1	ATOM	Atom sites
3.6.7.2	CHEMICAL	Chemical properties and nomenclature
3.6.7.3	ENTITY	Chemical entities
3.6.7.4	GEOM	Geometry of atom sites
3.6.7.5	STRUCT	Crystallographic structure
3.6.7.6	SYMMETRY	Symmetry information
3.6.7.7	VALENCE	Bond-valence information
(*d*) Publication		
3.6.8.1	CITATION	Bibliographic references
3.6.8.2	COMPUTING	Computational details of the experiment
3.6.8.3	DATABASE	Database information
3.6.8.4	IUCR	Journal housekeeping and the contents of a published article
(*e*) File metadata		
3.6.9.1	AUDIT	Dictionary maintenance and identification
3.6.9.2	ENTRY	Links between data blocks
3.6.9.3	COMPLIANCE	Compliance with previous dictionaries

attribute `_type_conditions_esd` and allows the standard uncertainty of the value to be placed in parentheses after the numerical value, as in

```
_cell_length_a      58.39(5)
```

This is also permitted in mmCIF, but it is preferable to use a separate data item to record the standard uncertainty, as in

```
_cell.length_a        58.39
_cell.length_a_esd     0.05
```

There are many of these kinds of data names in the mmCIF dictionary. The name of each is derived by adding `_esd` to the data name for the value. They are indicated by a + symbol in the category summaries in this chapter.

3.6.5. Experimental measurements

The CELL, DIFFRN and EXPTL category groups are used to describe the crystallographic experiment. The data items used for this purpose in mmCIF are for the most part identical to those in the core CIF dictionary. A complete discussion of the data names in each category may be found in Section 3.2.2.

mmCIF also contains the new categories EXPTL_CRYSTAL_GROW and EXPTL_CRYSTAL_GROW_COMP (Section 3.6.5.3.2), which are used to provide a more structured description of crystallization than is available in the core CIF dictionary.

3.6.5.1. Crystal cell parameters and measurement conditions

The categories describing the crystal unit cell and its determination are as follows:

CELL group
- CELL
- CELL_MEASUREMENT
- CELL_MEASUREMENT_REFLN

The mmCIF dictionary differs from the core CIF dictionary in assigning separate categories to data names that define the crystal unit-cell parameters and to data names relating to the experimental determination of the unit cell. Details of the unit-cell parameters are given in the CELL category and data items in the distinct CELL_MEASUREMENT category are used to describe how the unit-cell parameters were measured. The category CELL_MEASUREMENT_REFLN, which is used to list the reflections used in the unit-cell determination, is common to the core and mmCIF dictionaries.

The data items in these categories are as follows:

(*a*) CELL

```
• _cell.entry_id
        → _entry.id
+ _cell.angle_alpha
+ _cell.angle_beta
+ _cell.angle_gamma
  _cell.details (~ _cell_special_details)
  _cell.formula_units_Z
+ _cell.length_a
+ _cell.length_b
+ _cell.length_c
+ _cell.reciprocal_angle_alpha
+ _cell.reciprocal_angle_beta
+ _cell.reciprocal_angle_gamma
+ _cell.reciprocal_length_a
+ _cell.reciprocal_length_b
+ _cell.reciprocal_length_c
+ _cell.volume
  _cell.Z_PDB
```

(*b*) CELL_MEASUREMENT

```
• _cell_measurement.entry_id
        → _entry.id
+ _cell_measurement.pressure
  _cell_measurement.radiation
  _cell_measurement.reflns_used
+ _cell_measurement.temp
        (~ _cell_measurement_temperature)
  _cell_measurement.theta_max
  _cell_measurement.theta_min
  _cell_measurement.wavelength
```

(*c*) CELL_MEASUREMENT_REFLN

```
• _cell_measurement_refln.index_h
• _cell_measurement_refln.index_k
• _cell_measurement_refln.index_l
  _cell_measurement_refln.theta
```

The bullet (•) indicates a category key. Where multiple items within a category are marked with a bullet, they must be taken together to form a compound key. Items in italics have aliases in the core CIF dictionary formed by changing the full stop (.) to an underscore (_) except where indicated by the ~ symbol. Data items marked with a plus (+) have companion data names for the standard uncertainty in the reported value, formed by appending the string `_esd` to the data name listed.

The summary above includes the formal category keys that have been introduced in mmCIF because the corresponding core categories do not expect looped data, and therefore do not require the specification of a unique identifier. In the relational model of DDL2, all categories are considered to be tables and therefore each category must have a unique identifier. Where core CIF categories have one or more data names that fulfil the role of table-row identifiers, these have generally been carried over as category keys in the mmCIF dictionary (for example, the data items that correspond to the *h*, *k*, and *l* Miller indices of a reflection in the CELL_MEASUREMENT_REFLN category).