Table 3.4.2.1. *Category groups defined in the msCIF dictionary*

The groups are listed in the order in which they are described in this chapter.

Section	Category group	Subject covered
(*a*) Experimental measurements		
3.4.3.1.1	CELL	The unit cell, especially cell subsystems for composite structures
3.4.3.1.2	DIFFRN	Multi-dimensional diffraction pattern
3.4.3.1.3	EXPTL	Details of the experiment specific to modulated and composite structures
(*b*) Analysis		
3.4.3.2.1	REFINE	Refinement procedures
3.4.3.2.2	REFLN	Reflection measurements indexed in higher-dimensional space
(*c*) Atomicity, chemistry and structure		
3.4.3.3.1 to 3.4.3.3.3	ATOM	Atom sites in a modulated structure
3.4.3.3.4	GEOM	Geometry of a structure in superspace
3.4.3.3.5	SYMMETRY	Symmetry information
(*d*) File metadata		
3.4.3.4	AUDIT	The structure of the CIF

(Chapter 4.1). This means that the dictionary defines items that are basically related to single-crystal data. The close relationship between the msCIF and core dictionaries has led to synergies between and benefits for both dictionaries.

The design of the msCIF dictionary had two objectives: (i) it should be as functional as possible, *i.e.* as little information in an msCIF as possible should be given as unstructured text; (ii) it should be possible to include even the oldest modulated and composite structures in an msCIF, even if the way they were reported did not follow the guidelines used now.

There were two major difficulties in implementing the msCIF dictionary. Firstly, the number of additional wave vectors used to index a diffraction diagram is theoretically not limited. Secondly, a CIF containing information about a modulated or composite structure should, in general, be composed of several (related) data blocks. As CIF definitions do not at present include vectors or matrices as distinct types of data values, an arbitrary upper limit of 11 was assigned to the dimension of superspace to limit the number of new data names. Linking between data blocks is handled by using recommended values for items in the AUDIT and AUDIT_LINK categories, like those used in the powder CIF (pdCIF) dictionary (Chapters 3.3 and 4.2).

An additional problem arises when special (ideal) modulation functions are considered. Although periodic modulations are normally parameterized by Fourier series, in certain cases it is convenient to use discontinuous functions which lead to a severe reduction in the number of structural parameters. The shape of these functions is not restricted and new materials could require new functions. Given that it is not possible at this moment to define logical or mathematical relations between data values within a CIF [although an initiative for including algorithms in the definitions of CIF dictionaries has been proposed by Spadaccini *et al.* (2000)], general functions cannot be defined and therefore the type of special functions included in the msCIF dictionary are those implemented in the most widely used program, *JANA2000* (Petříček & Dušek, 2000). They only apply to one-dimensional modulations and are sawtooth displacive functions and occupational crenel functions. Both functions define discontinuous occupational atomic domains and are normally combined with (smoother) atomic modulation functions (involving atom positions and/or thermal parameters) that are expressed by Fourier series. Because of the discreteness of the atomic domains, members of the set of harmonic functions used to expand these series are no longer mutually orthogonal as they are only defined within each atomic domain and not in the (internal space) interval $[0, 1]$. As a consequence, severe correlation effects among the coefficients of the Fourier series are expected. A solution for this problem lies in the selection and orthogonalization of a set of basic functions (Petříček *et al.*, 1995). The atomic modulation functions are then expressed as linear combinations of an orthogonal basis whose elements are specific combinations of harmonic functions. Discontinuous atomic domains are being increasingly used in composite materials, in some cases revealing that considering such materials as composites or modulated structures is a matter of convenience (Elcoro *et al.*, 2003; Pérez-Mato *et al.*, 2003).

CIFs that conform to the msCIF dictionary are highly itemized for the human reader, but have a strong relational structure even though the dictionary itself is written in DDL1.

The major drawbacks of the dictionary are:

(i) Items describing the superspace symmetry should be reconsidered and perhaps included within the symmetry CIF (symCIF) dictionary (Chapters 3.8 and 4.7).

(ii) There is still some information in an msCIF that cannot be interpreted by a computer [*e.g.* rigid rotations and translations around (along) noncrystallographic axes cannot be parsed, since the description of the axes is textual].

(iii) A full description of the modulation in terms of orthogonalized functions (used when the atomic domains are discrete) is not supported yet.

3.4.3. Arrangement of the dictionary

The msCIF dictionary detailed in Chapter 4.3 includes 19 new categories. Another 18 already exist in the core CIF dictionary, but include new items (16) or revised definitions (2). The category structure of the msCIF dictionary is summarized in Table 3.4.2.1 and is listed in full in Appendix 3.4.1. The appendix also lists for each category the section of this chapter in which the category is described.

Many of the modifications to categories that already exist in the core CIF dictionary result from the need to use more than three integer indices to label the diffracted intensities (in the cases of CELL, DIFFRN_REFLN, DIFFRN_REFLNS, DIFFRN_STANDARD_REFLN and EXPTL_CRYSTAL_FACE) or the need to use superspace symmetry (in the cases of GEOM_ANGLE, GEOM_BOND, GEOM_CONTACT, GEOM_TORSION, SPACE_GROUP and SPACE_GROUP_SYMOP). Apart from the categories that describe the atomic modulation functions, there are two that are specific to composite structures (CELL_SUBSYSTEM and CELL_SUBSYSTEMS).

The rest of this section summarizes the contents of the dictionary, organized by categories within the functional groups outlined in Table 3.1.10.1. As in the other chapters in this part of the volume, the classification is under the headings *Experimental measurements* (Section 3.4.3.1), *Analysis* (Section 3.4.3.2), *Atomicity, chemistry and structure* (Section 3.4.3.3) and *File metadata* (Section 3.4.3.4). The msCIF dictionary adds no new data items concerned with the publication or reporting of structures to those already present in the core CIF dictionary.

The data items within each category are listed in the detailed commentary below. Where relevant, data items that represent a unique identifier for a looped list ('category keys') are listed first and are marked by a bullet (•). The remaining data items in each category are listed alphabetically.

3.4.3.1. Experimental measurements

The categories relevant to a structure determination experiment are in the CELL, DIFFRN and EXPTL groups. Most of the items extend the existing core CIF categories, but the categories describing cell subsystems are new.

3.4.3.1.1. *Cell and modulation wave vectors*

The categories describing the unit cell (or cells for composite structures), the wave vectors of the modulations and, for composites, the cell subsystems are as follows:

CELL group
- CELL ¶
- CELL_SUBSYSTEM
- CELL_SUBSYSTEMS
- CELL_WAVE_VECTOR
- CELL_WAVE_VECTORS

Categories marked with ¶ are already defined in the core CIF dictionary.

The data items in these categories are as follows:

(*a*) CELL
```
_cell_modulation_dimension
_cell_reciprocal_basis_description
```

(*b*) CELL_SUBSYSTEM
```
_cell_subsystem_code
_cell_subsystem_description
_cell_subsystem_matrix_W_1_1
   ⋮
```
(and `_cell_subsystem_matrix_W_`*m*_*n* for all combinations $1 \leq m,n \leq 11$)
```
   ⋮
_cell_subsystem_matrix_W_11_11
```

(*c*) CELL_SUBSYSTEMS
```
_cell_subsystems_number
```

(*d*) CELL_WAVE_VECTOR
```
_cell_wave_vector_seq_id
_cell_wave_vector_x
_cell_wave_vector_y
_cell_wave_vector_z
```

(*e*) CELL_WAVE_VECTORS
```
_cell_wave_vectors_meas_details
_cell_wave_vectors_pressure_max
_cell_wave_vectors_pressure_min
_cell_wave_vectors_temp_max
_cell_wave_vectors_temp_min
_cell_wave_vectors_variation
```

As explained in Section 3.4.2, the msCIF dictionary arbitrarily allows an upper limit of 11 for the dimension of superspace for which data names are defined. `_cell_modulation_dimension` specifies the number of additional reciprocal vectors needed to index the whole diffraction pattern and has values d ranging from 1 to 8 to express the dimensionality $(3 + d)$ of the superspace. `_cell_reciprocal_basis_description` is a text field allowing a free description of the higher-dimensional basis chosen.

For a composite structure, different cell subsystems may be specified. Each such subsystem is identified and characterized by the data items in the CELL_SUBSYSTEM category (see Section 3.4.4.1). `_cell_subsystems_number` gives the number of such subsystems as an independent check of the completeness of the description.

Data items in the CELL_WAVE_VECTOR category specify the wave vectors of the modulation (see Section 3.4.4.1). In accordance with the limits on dimensionality of the current version of the msCIF dictionary, no more than eight additional modulation wave vectors may be specified. The number used must agree with the value of `_cell_modulation_dimension`.

The data items in the CELL_WAVE_VECTORS category describe the experimental conditions during the determination of the independent modulation wave vectors.

3.4.3.1.2. *Data collection*

The categories describing data collection are as follows:

DIFFRN group
- DIFFRN_REFLN ¶
- DIFFRN_REFLNS ¶
- DIFFRN_STANDARD_REFLN ¶

Categories marked with ¶ are already defined in the core CIF dictionary.

New data items in these categories are as follows:

(*a*) DIFFRN_REFLN
```
_diffrn_refln_index_m_1
_diffrn_refln_index_m_2
_diffrn_refln_index_m_3
_diffrn_refln_index_m_4
_diffrn_refln_index_m_5
_diffrn_refln_index_m_6
_diffrn_refln_index_m_7
_diffrn_refln_index_m_8
```

(*b*) DIFFRN_REFLNS
```
_diffrn_reflns_limit_index_m_1_max
_diffrn_reflns_limit_index_m_1_min
_diffrn_reflns_limit_index_m_2_max
_diffrn_reflns_limit_index_m_2_min
_diffrn_reflns_limit_index_m_3_max
_diffrn_reflns_limit_index_m_3_min
_diffrn_reflns_limit_index_m_4_max
_diffrn_reflns_limit_index_m_4_min
_diffrn_reflns_limit_index_m_5_max
_diffrn_reflns_limit_index_m_5_min
_diffrn_reflns_limit_index_m_6_max
_diffrn_reflns_limit_index_m_6_min
_diffrn_reflns_limit_index_m_7_max
_diffrn_reflns_limit_index_m_7_min
_diffrn_reflns_limit_index_m_8_max
_diffrn_reflns_limit_index_m_8_min
_diffrn_reflns_satellite_order_max
```

(*c*) DIFFRN_STANDARD_REFLN
```
_diffrn_standard_refln_index_m_1
_diffrn_standard_refln_index_m_2
_diffrn_standard_refln_index_m_3
_diffrn_standard_refln_index_m_4
_diffrn_standard_refln_index_m_5
_diffrn_standard_refln_index_m_6
_diffrn_standard_refln_index_m_7
_diffrn_standard_refln_index_m_8
```

The data items in these categories are straightforward extensions of the core CIF dictionary definitions to the indexing of diffraction intensities by higher-dimensional components. The `_diffrn_refln_index_m_*` items are the additional Miller indices m_i indexing the modulation wave vectors $\mathbf{q}_i$ when the diffraction wave vector is written as $\mathbf{H} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* + \sum_i m_i\mathbf{q}_i$.

The upper limit of 8 was chosen arbitrarily to limit the number of data items defined in the msCIF dictionary.

`_diffrn_reflns_limit_index_m_*` items provide independent checks on the range of values recorded for each of the additional Miller indices m_i. `_diffrn_standard_refln_index_m_*` items allow the higher-dimensional Miller indices of standard reflections to be recorded.

3.4.3.1.3. *Experimental measurements on the crystal*

The categories describing measurements on the crystal or crystals used in the experiment are as follows:

EXPTL group
EXPTL_CRYSTAL ¶
EXPTL_CRYSTAL_FACE ¶

Categories marked with ¶ are already defined in the core CIF dictionary.

New data items in these categories are as follows:

(*a*) EXPTL_CRYSTAL
_exptl_crystal_type_of_structure

(*b*) EXPTL_CRYSTAL_FACE
_exptl_crystal_face_index_m_1
_exptl_crystal_face_index_m_2
_exptl_crystal_face_index_m_3
_exptl_crystal_face_index_m_4
_exptl_crystal_face_index_m_5
_exptl_crystal_face_index_m_6
_exptl_crystal_face_index_m_7
_exptl_crystal_face_index_m_8

_exptl_crystal_type_of_structure specifies the structure type as `cryst` (crystalline), `mod` (modulated) or `comp` (composite). These are the only three types of structure handled at present by the msCIF dictionary.

The extensions to the EXPTL_CRYSTAL_FACE category permit the indexing of crystal faces using the higher-dimensional Miller indices introduced for aperiodic structures.

3.4.3.2. Analysis

The categories related to refinement that have been extended in this dictionary are as follows:

Refinement techniques and results (§3.4.3.2.1)
REFINE ¶
The reflections used in the refinement (§3.4.3.2.2)
REFLN ¶
REFLNS ¶

Categories marked with ¶ are already defined in the core CIF dictionary.

3.4.3.2.1. *Refinement techniques and results*

New data items in this category are as follows:

REFINE
_refine_ls_mod_func_description
_refine_ls_mod_hydrogen_treatment
_refine_ls_mod_overall_phason_coeff
_refine_ls_mod_overall_phason_formula

During the early stages of the development of the msCIF dictionary, several sets of data items were defined to accommodate the need to specify residual *R* factors for the different sets of main reflections and satellite reflections. It was then recognized that the binning of reflection classes had more general application, and these new data items were transferred to the core CIF dictionary, where, of course, they are still available for use in an msCIF.

The new items in the REFINE category in the msCIF dictionary are specific to the refinement of modulated structures. **_refine_ls_mod_func_description** allows a free-text description of the types of modulation present in the structural model and how they are handled. The treatment of hydrogen-atom modulation parameters is specified by **_refine_ls_mod_hydrogen_treatment**. Information on an overall phason correction (the use of which should in general be discouraged) may be given using the **_refine_ls_mod_overall_phason_*** items.

3.4.3.2.2. *The reflections used in the refinement*

New data items in these categories are as follows:

(*a*) REFLN
_refln_index_m_1
_refln_index_m_2
_refln_index_m_3
_refln_index_m_4
_refln_index_m_5
_refln_index_m_6
_refln_index_m_7
_refln_index_m_8

(*b*) REFLNS
_reflns_limit_index_m_1_max
_reflns_limit_index_m_1_min
_reflns_limit_index_m_2_max
_reflns_limit_index_m_2_min
_reflns_limit_index_m_3_max
_reflns_limit_index_m_3_min
_reflns_limit_index_m_4_max
_reflns_limit_index_m_4_min
_reflns_limit_index_m_5_max
_reflns_limit_index_m_5_min
_reflns_limit_index_m_6_max
_reflns_limit_index_m_6_min
_reflns_limit_index_m_7_max
_reflns_limit_index_m_7_min
_reflns_limit_index_m_8_max
_reflns_limit_index_m_8_min

As with the **_diffrn_refln_*** and **_diffrn_reflns_*** items (Section 3.4.3.1.2), these data names extend the corresponding core data items into the higher-dimensional space used in the treatment of modulated structures and composites. They apply to the list of reflections used in the refinement, as distinct from the experimentally collected set of intensities described by the **_diffrn_*** data items.

3.4.3.3. Atomicity, chemistry and structure

The categories relevant to the description of the structural model are as follows:

ATOM group
Atom sites (§3.4.3.3.1)
ATOM_SITE ¶
ATOM_SITE_PHASON
Modulation functions as Fourier series (§3.4.3.3.2)
ATOM_SITE_DISPLACE_FOURIER
ATOM_SITE_DISPLACE_FOURIER_PARAM
ATOM_SITE_FOURIER_WAVE_VECTOR
ATOM_SITE_OCC_FOURIER
ATOM_SITE_OCC_FOURIER_PARAM
ATOM_SITE_ROT_FOURIER
ATOM_SITE_ROT_FOURIER_PARAM
ATOM_SITE_U_FOURIER
ATOM_SITE_U_FOURIER_PARAM
ATOM_SITES_DISPLACE_FOURIER
ATOM_SITES_MODULATION
ATOM_SITES_ROT_FOURIER
Special modulation functions (§3.4.3.3.3)
ATOM_SITE_DISPLACE_SPECIAL_FUNC
ATOM_SITE_OCC_SPECIAL_FUNC
Molecular or packing geometry (§3.4.3.3.4)
GEOM group
GEOM_ANGLE ¶
GEOM_BOND ¶
GEOM_CONTACT ¶
GEOM_TORSION ¶
Symmetry information (§3.4.3.3.5)
SYMMETRY group
SPACE_GROUP ¶
SPACE_GROUP_SYMOP ¶

Categories marked with ¶ are already defined in the core CIF dictionary.

Most of the new categories introduced to the msCIF dictionary appear here, since their function is to describe in great detail the

modulation of the atom-site properties. They fall naturally into families describing the modulation of atomic displacement, of site occupation or of thermal parameters.

New data items are added to several categories in the core CIF dictionary that describe molecular or packing geometry. There are also new data items to describe superspace-group symmetry.

3.4.3.3.1. *Atom sites*

Data items in these categories are as follows:

(*a*) ATOM_SITE

```
  _atom_site_displace_modulation_flag
  _atom_site_occ_modulation_flag
  _atom_site_subsystem_code
        → _cell_subsystem_code
  _atom_site_U_modulation_flag
```

(*b*) ATOM_SITE_PHASON

```
• _atom_site_phason_atom_site_label
        → _atom_site_label
  _atom_site_phason_coeff
  _atom_site_phason_formula
```

The bullet (•) *indicates a category key. The arrow* (→) *is a reference to a parent data item.*

The ATOM_SITE category is extended in the msCIF dictionary by the addition of a small number of items that may appear in the main looped list of atom-site information (see Section 3.2.4.1.1). The `*_flag` items indicate whether each individual atom site has been modelled through modulation of atomic displacement, site occupation or thermal parameters. In each case, the default value of the item is `no`, so that any or all of the flags may be omitted when that particular type of modulation has not been applied to the structural model.

`_atom_site_subsystem_code` identifies the cell subsystem to which the atom site must be assigned in the description of composite structures. Each value of `_atom_site_subsystem_code` must match one of the values of `_cell_subsystem_code` in the overall description of the subsystems defined for a composite.

The ATOM_SITE_PHASON category allow details of an atom-dependent phason correction, as implemented in *JANA2000*, to be given. The use of these phason corrections is discouraged.

3.4.3.3.2. *Modulation functions as Fourier series*

Data items in these categories are as follows:

(*a*) ATOM_SITE_DISPLACE_FOURIER

```
• _atom_site_displace_Fourier_id
  _atom_site_displace_Fourier_atom_site_label
        → _atom_site_label
  _atom_site_displace_Fourier_axis
  _atom_site_displace_Fourier_wave_vector_seq_id
        → _atom_site_Fourier_wave_vector_seq_id
```

(*b*) ATOM_SITE_DISPLACE_FOURIER_PARAM

```
• _atom_site_displace_Fourier_param_id
        → _atom_site_displace_Fourier_id
  _atom_site_displace_Fourier_param_cos
  _atom_site_displace_Fourier_param_modulus
  _atom_site_displace_Fourier_param_phase
  _atom_site_displace_Fourier_param_sin
```

(*c*) ATOM_SITE_FOURIER_WAVE_VECTOR

```
  _atom_site_Fourier_wave_vector_description
  _atom_site_Fourier_wave_vector_seq_id
  _atom_site_Fourier_wave_vector_x
  _atom_site_Fourier_wave_vector_y
  _atom_site_Fourier_wave_vector_z
```

(*d*) ATOM_SITE_OCC_FOURIER

```
• _atom_site_occ_Fourier_id
  _atom_site_occ_Fourier_atom_site_label
        → _atom_site_label
  _atom_site_occ_Fourier_wave_vector_seq_id
        → _atom_site_Fourier_wave_vector_seq_id
```

(*e*) ATOM_SITE_OCC_FOURIER_PARAM

```
• _atom_site_occ_Fourier_param_id
        → _atom_site_occ_Fourier_id
  _atom_site_occ_Fourier_param_cos
  _atom_site_occ_Fourier_param_modulus
  _atom_site_occ_Fourier_param_phase
  _atom_site_occ_Fourier_param_sin
```

(*f*) ATOM_SITE_ROT_FOURIER

```
• _atom_site_rot_Fourier_id
  _atom_site_rot_Fourier_atom_site_label
        → _atom_site_label
  _atom_site_rot_Fourier_axis
  _atom_site_rot_Fourier_wave_vector_seq_id
        → _atom_site_Fourier_wave_vector_seq_id
```

(*g*) ATOM_SITE_ROT_FOURIER_PARAM

```
• _atom_site_rot_Fourier_param_id
        → _atom_site_rot_Fourier_id
  _atom_site_rot_Fourier_param_cos
  _atom_site_rot_Fourier_param_modulus
  _atom_site_rot_Fourier_param_phase
  _atom_site_rot_Fourier_param_sin
```

(*h*) ATOM_SITE_U_FOURIER

```
• _atom_site_U_Fourier_id
  _atom_site_U_Fourier_atom_site_label
        → _atom_site_label
  _atom_site_U_Fourier_tens_elem
  _atom_site_U_Fourier_wave_vector_seq_id
        → _atom_site_Fourier_wave_vector_seq_id
```

(*i*) ATOM_SITE_U_FOURIER_PARAM

```
• _atom_site_U_Fourier_param_id
        → _atom_site_U_Fourier_id
  _atom_site_U_Fourier_param_cos
  _atom_site_U_Fourier_param_modulus
  _atom_site_U_Fourier_param_phase
  _atom_site_U_Fourier_param_sin
```

(*j*) ATOM_SITES_DISPLACE_FOURIER

```
  _atom_sites_displace_Fourier_axes_description
```

(*k*) ATOM_SITES_MODULATION

```
  _atom_sites_modulation_global_phase_t_1
  _atom_sites_modulation_global_phase_t_2
  _atom_sites_modulation_global_phase_t_3
  _atom_sites_modulation_global_phase_t_4
  _atom_sites_modulation_global_phase_t_5
  _atom_sites_modulation_global_phase_t_6
  _atom_sites_modulation_global_phase_t_7
  _atom_sites_modulation_global_phase_t_8
```

(*l*) ATOM_SITES_ROT_FOURIER

```
  _atom_sites_rot_Fourier_axes_description
```

The bullet (•) *indicates a category key. The arrow* (→) *is a reference to a parent data item.*

It is common to represent a modulated structure using a reference periodic structure on which are superimposed atomic modulation functions expanded as Fourier series. (A full discussion of this is given in Section 3.4.4.3.) The msCIF dictionary provides separate categories for listing the modulated parameters that apply to atom positions, site occupancies and thermal parameters. The structuring of the data items within each of these categories follows a similar pattern.

For example, consider the modulation of the atomic displacements. The ATOM_SITE_DISPLACE_FOURIER category allows a listing of the axis along which the displacement occurs (`*_axis`) and the wave vectors contributing to that displacement component (`*_wave_vector_seq_id`) for each relevant atom site (labelled by `_atom_site_displace_Fourier_atom_site_label`).

_wave_vector_seq_id** is a pointer to the description of the separate modulation wave vectors and must match one of the identifiers **_atom_site_Fourier_wave_vector_seq_id** listed separately in the ATOM_SITE_FOURIER_WAVE_VECTOR category. Likewise, the ***_atom_site_label** data item must match a value of **_atom_site_label** in the main list of atom positions. This is how the modulation is linked to the atom list. The item **_atom_site_displace_Fourier_id** is the formal key for the ATOM_SITE_DISPLACE_FOURIER category. It is used to locate the matching Fourier coefficients in the ATOM_SITE_DISPLACE_FOURIER_PARAM category. The coefficients may be reported in a sine–cosine (**_atom_site_displace_Fourier_param_sin**, ***_cos**) or modulus–argument (_mod**, ***_phase**) representation.

Where a group of atoms is treated as a rigid group, the categories above describe only the translational part of the positional distortion. ATOM_SITE_ROT_FOURIER and ATOM_SITE_ROT_FOURIER_PARAM are used to describe the rotational components.

ATOM_SITE_OCC_FOURIER and ATOM_SITE_U_FOURIER, and their associated *_PARAM categories, are the analogous categories for the modulation of site occupation and thermal parameters.

All the categories above describe the properties of individual atom sites. Larger-scale descriptions of the displacive modulation or of the rotational component of a rigid group are covered by the categories ATOM_SITES_DISPLACE_FOURIER and ATOM_SITES_ROT_FOURIER, each of which at present contains one descriptive data item.

The ATOM_SITES_MODULATION category contains data items describing the initial phases of the modulation waves, which are essential for determining the space group of the commensurate superstructure. More details are given in the dictionary.

3.4.3.3.3. *Special modulation functions*

Data items in these categories are as follows:

(*a*) ATOM_SITE_DISPLACE_SPECIAL_FUNC

• **_atom_site_displace_special_func_atom_site_label**
  → **_atom_site_label**
  _atom_site_displace_special_func_sawtooth_ax
  _atom_site_displace_special_func_sawtooth_ay
  _atom_site_displace_special_func_sawtooth_az
  _atom_site_displace_special_func_sawtooth_c
  _atom_site_displace_special_func_sawtooth_w

(*b*) ATOM_SITE_OCC_SPECIAL_FUNC

• **_atom_site_occ_special_func_atom_site_label**
  → **_atom_site_label**
  _atom_site_occ_special_func_crenel_c
  _atom_site_occ_special_func_crenel_w

The bullet (•) indicates a category key. The arrow (→) is a reference to a parent data item.

Several data items cover modulation functions that are not expressed as Fourier expansions. The examples in the current msCIF dictionary are restricted to the one-dimensional modulations (sawtooth displacive and occupational crenel functions) implemented in the program *JANA2000* (see Section 3.4.2).

3.4.3.3.4. *Molecular or packing geometry*

New data items in these categories are as follows:

(*a*) GEOM_ANGLE
_geom_angle_av
_geom_angle_max
_geom_angle_min
_geom_angle_site_ssg_symmetry_1
_geom_angle_site_ssg_symmetry_2
_geom_angle_site_ssg_symmetry_3

(*b*) GEOM_BOND
_geom_bond_distance_av
_geom_bond_distance_max
_geom_bond_distance_min
_geom_bond_site_ssg_symmetry_1
_geom_bond_site_ssg_symmetry_2

(*c*) GEOM_CONTACT
_geom_contact_distance_av
_geom_contact_distance_max
_geom_contact_distance_min
_geom_contact_site_ssg_symmetry_1
_geom_contact_site_ssg_symmetry_2

(*d*) GEOM_TORSION
_geom_torsion_av
_geom_torsion_max
_geom_torsion_min
_geom_torsion_site_ssg_symmetry_1
_geom_torsion_site_ssg_symmetry_2
_geom_torsion_site_ssg_symmetry_3
_geom_torsion_site_ssg_symmetry_4

For each of the geometry categories, there are two groups of extensions. One set covers maximum, minimum and average values of bonds, contact distances, angles and torsion angles. The other extends the symmetry-operation code used in geometry listings in the core CIF dictionary (see Section 3.2.4.3.2) to the higher-dimensional superspace form.

3.4.3.3.5. *Symmetry information*

New data items in these categories are as follows:

(*a*) SPACE_GROUP
_space_group_ssg_IT_number
_space_group_ssg_name
_space_group_ssg_name_IT
_space_group_ssg_name_WJJ
_space_group_ssg_WJJ_code

(*b*) SPACE_GROUP_SYMOP
_space_group_symop_ssg_id
_space_group_symop_ssg_operation_algebraic

At present, the msCIF dictionary extends the core CIF dictionary symmetry categories to describe superspace groups for one-dimensional modulated structures in four ways: as the superspace-group number in Janssen *et al.* (2004) (**_space_group_ssg_IT_number**), as the *International Tables* superspace-group symbol (***_ssg_name_IT**), as one of the notations from de Wolff *et al.* (1981) (***_ssg_name_WJJ**, ***_ssg_WJJ_code**), or in some other formalism (***_ssg_name**). At present, superspace-group names for higher dimensions can only be indicated using **_space_group_ssg_name**.

Symmetry operations in the superspace group are specified in the SPACE_GROUP_SYMOP category by an obvious extension to the method used in the core dictionary. These items must always be present in a CIF corresponding to a modulated or composite structure.

3.4.3.4. File metadata

The categories modified in the msCIF dictionary to formalize the construction of a multi-block description of modulated or composite structures are as follows:

AUDIT group
  AUDIT ¶
  AUDIT_LINK ¶

Categories marked with ¶ are already defined in the core CIF dictionary.

Data items revised in these categories are as follows:

(*a*) AUDIT
`_audit_block_code`
(*b*) AUDIT_LINK
`_audit_link_block_code`

The core dictionary definitions of these items are revised in order to formalize the relationships between multiple data blocks representing reference and modulated structures. Guidance is provided in the msCIF dictionary on how to label data blocks in a way that makes their mutual relationships clear.

3.4.4. Use of the msCIF dictionary

In this section, some of the capabilities of the dictionary will be demonstrated using simple examples. More detailed examples can be found at http://www.iucr.org/iucr-top/cif/ms and on the CD-ROM accompanying this volume.

3.4.4.1. Description of reciprocal space

Modulated and composite structures need more than three reciprocal vectors in order to index the whole set of reflections with integer numbers. Hence a diffraction vector is written as

$$\mathbf{H} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* + m_1\mathbf{q}_1 + \ldots + m_d\mathbf{q}_d, \qquad (3.4.4.1)$$

where the notation has been chosen according to the core CIF dictionary. In the case of a modulated structure, $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$ are the reciprocal vectors of the reference structure (and therefore h, k and l index the main reflections). $\mathbf{q}_1, \ldots, \mathbf{q}_d$ are the modulation wave vectors. They are three-dimensional vectors with some irrational component (if the modulated structure is incommensurate) in the lattice spanned by $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$. d is the dimension of the modulation. In the case of composite structures, the diffraction pattern can be indexed using $3 + d$ (arbitrarily selected) vectors $\mathbf{a}_k^*$ ($k = 1, \ldots, 3+d$). $\mathbf{a}_1^*$ ($\equiv \mathbf{a}^*$), $\mathbf{a}_2^*$ ($\equiv \mathbf{b}^*$) and $\mathbf{a}_3^*$ ($\equiv \mathbf{c}^*$) normally span the reciprocal lattice of the main reflections of one of the substructures (notice that this is only one particular, but highly intuitive, choice). The remaining d vectors with $k = 4, \ldots, d$ are the wave vectors of the modulation [$\mathbf{q}_1, \ldots, \mathbf{q}_d$ in equation (3.4.4.1)].

In a composite structure, the $(3 + d)$-dimensional reciprocal basis of the subsystem ν is determined by a $(3 + d) \times (3 + d)$ matrix W^ν [see van Smaalen (1995) and references therein]:

$$\mathbf{a}_i^{*\nu} = \sum_{k=1}^{3+d} W_{ik}^\nu \mathbf{a}_k^*, \quad i = 1, \ldots, 3 + d, \qquad (3.4.4.2)$$

where the subscripts $i = 1$, 2 and 3 label the reciprocal vectors $\mathbf{a}^{*\nu}$, $\mathbf{b}^{*\nu}$ and $\mathbf{c}^{*\nu}$, and $i = 4, \ldots, d$ label the wave vectors of the modulation expressed as linear combinations of $\mathbf{a}^{*\nu}$, $\mathbf{b}^{*\nu}$ and $\mathbf{c}^{*\nu}$.

The simplest case corresponds to a one-dimensional ($d = 1$) modulated structure. Consider for example the incommensurate phase of K_2SeO_4. The wave vector of the modulation can be chosen to be $\mathbf{q}_1 = \alpha\mathbf{a}^*$. Relevant information about the diffraction pattern of this compound is expressed using both the core CIF and msCIF dictionaries as shown in Example 3.4.4.1.

A more complicated example is the composite structure $(LaS)_{1.14}NbS_2$. The two mutually incommensurate subsystems (along the a axis) are (van Smaalen, 1991) NbS_2 ($\nu = 1$) and LaS ($\nu = 2$). The reciprocal basis can be chosen to be $\mathbf{a}_1^* = \mathbf{a}^{*1}$, $\mathbf{a}_2^* = \mathbf{b}^{*1}$, $\mathbf{a}_3^* = \mathbf{c}^{*1}$ and $\mathbf{a}_4^* = \mathbf{a}^{*2}$. For this particular choice, the two W matrices [see equation (3.4.4.2)] are

Example 3.4.4.1. *msCIF description of the diffraction pattern of a one-dimensional modulated structure.*

```
_exptl_crystal_type_of_structure         'mod'
_cell_reciprocal_basis_description
; a*,b*,c* (reciprocal basis spanning the lattice of
  main reflections), q modulation wave vector.
;

_diffrn_symmetry_description
; The whole diffraction pattern shows orthorhombic
  symmetry. The following extinction rules were
  detected:
     0kl0    k+l=odd
     h0lm    h+m=odd
     hk0m    m=odd
     h00m    h,m=odd
  Superspace group: P:Pnam:-1ss
;

_diffrn_reflns_satellite_order_max            1

_diffrn_reflns_theta_max                  40.14
_diffrn_reflns_theta_min                   3.32
_diffrn_reflns_limit_h_max                    8
_diffrn_reflns_limit_k_max                   18
_diffrn_reflns_limit_l_max                   10
_diffrn_reflns_limit_index_m_1_max            1
_diffrn_reflns_limit_h_min                    0
_diffrn_reflns_limit_k_min                    0
_diffrn_reflns_limit_l_min                    0
_diffrn_reflns_limit_index_m_1_min           -1

# Modulation wave vector
  loop_
     _cell_wave_vector_seq_id
     _cell_wave_vector_x
         1        0.318(5)
```

$$W^1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad W^2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

This information is transcribed to CIF format as shown in Example 3.4.4.2. (Note that the default values for the wave vector components and the elements of W are 0.)

3.4.4.2. Description of symmetry

The symmetry of a modulated or composite structure is described by a superspace group which leaves the $(3 + d)$-dimensional embedding of the structure invariant. Superspace is built of two orthogonal subspaces and both of them are kept invariant separately by the superspace symmetry operations. (In reciprocal space this means that, for these structures, main reflections and satellite reflections are never transformed into one another by superspace symmetry operations.) Consequently, superspace groups are not general $(3 + d)$-dimensional space groups. The standard notation for superspace groups only covers the one-dimensional superspace groups, which are listed in Janssen *et al.* (2004). As a consequence, msCIFs must include a list of all the symmetry operations in an (x, y, z) format (using as symbols $x_1 \ldots x_{3+d}$) similar to that used in the core CIF dictionary. Superspace-group names for one-dimensional structures can be expressed either according to Janssen *et al.* (2004) or according to the original notation of de Wolff *et al.* (1981). Alternative names or higher-dimensional superspace groups can also be included, but not parsed.

references