

3. CIF DATA DEFINITION AND CLASSIFICATION

Example 3.2.3.3. *Description of subsets of the reflection list.*

```

loop_
  _reflns_class_number_gt
  _reflns_class_code
  _reflns_class_description
  584 'Main' 'm=0; main reflections'
  226 'Sat1' 'm=1; first-order satellites'
  50 'Sat2' 'm=2; second-order satellites'

```

The REFLNS_CLASS category is used to summarize the properties of subsets of the reflection list. The data names are analogous to several in the REFLNS and REFIN categories, but are applied to individual classes of reflections labelled by `_reflns_class_code` and described by `_reflns_class_description` (see Example 3.2.3.3).

Individual reflections in the structure-factor listing can be recognized through the matching value of `_refln_class_code` as belonging to a particular class labelled by `_reflns_class_code`.

Although classes can be assigned according to arbitrary criteria, the specific case for which the REFLNS_CLASS category was designed was the partitioning of the reflection list into contributions from different components in incommensurately modulated structures. However, the formalism is general and other binning strategies can be described. Note, however, that the specific case of processing of reflections by shells of resolution (in macromolecular crystallography, for example) is handled explicitly by the REFLNS_SHELL category.

The category REFLNS_SCALE provides a listing of the scale factors applied to individual reflections sharing a common value of `_refln_scale_group_code`. Each value is indexed by the matching identifier `_reflns_scale_group_code` of this category.

The REFLNS_SHELL category describes the properties of separate resolution shells of reflections and is a special case of the binning of reflections into classes (compare REFLNS_CLASS above).

Each shell is defined by an upper and lower resolution limit (`_reflns_shell_d_res_high` and `*_low`), and for each shell there are data names for the number of reflections measured and exceeding a threshold of significance, for the percentage of geometrically possible reflections collected, and for the ratios of the mean intensities to their standard uncertainties.

R_{merge} values are also defined for each shell of resolution (both for all measured reflections and for significantly intense ones).

This category also contains a number of deprecated data names reflecting older terminology and notation. Such data names should not be used in creating new CIFs, but will need to be recognized by CIF-reading software in order to process old CIFs.

3.2.4. Atomicity, chemistry and structure

The core CIF dictionary provides many data names for describing the structural model.

The categories describing the atom sites handle these in a general way as sites of significant electron density which might be contributed to by more than one element species. The chemical identification of the compound under study, and where appropriate a model of the molecular connectivity and bonding, are handled separately by the chemistry-related categories. The geometry-related categories are purely derivative, given knowledge of the positions of the atom sites and the crystallographic symmetry; but as with other examples of derived data, they are given their own data names to provide convenient listings and to check the consistency of information provided by other categories. The symmetry-related data names in the core dictionary are restricted to those essential for the construction of a geometric model; Chapter 3.8

describes a symmetry extension dictionary suitable for a more complete description of crystal symmetry.

3.2.4.1. Atom sites

The categories describing atom sites are as follows:

ATOM group

Individual atom sites (§3.2.4.1.1)

ATOM_SITE

Collections of atom sites (§3.2.4.1.2)

ATOM_SITES

Atom types (§3.2.4.1.3)

ATOM_TYPE

These categories permit the traditional interpretation of regular concentrations of electron density in a crystalline lattice as atom sites containing one or more chemical elements, with complete or partial occupancy, and with a spatial distribution affected by thermal displacement or disorder.

Lists of atom-site coordinates and anisotropic displacement factors are covered by data items in the ATOM_SITE category. Identification of the chemical species occupying each site is handled by data items in the ATOM_TYPE category and data items in the ATOM_SITES category record collective information common to all sites.

While the ATOM_SITE category formally contains the data items describing both positions and atomic displacements, the anisotropic displacement parameters are often given in a separate looped list. In the version of the core dictionary embedded in the macromolecular CIF dictionary, which uses the DDL2 formalism, this is recognized by the creation of a separate, but overlapping, ATOM_SITE_ANISOTROP category.

3.2.4.1.1. Individual atom sites

The data items in this category are as follows:

ATOM_SITE

- `_atom_site_label`
- `_atom_site_adp_type`
- `_atom_site_aniso_B_11`
- `_atom_site_aniso_B_12`
- `_atom_site_aniso_B_13`
- `_atom_site_aniso_B_22`
- `_atom_site_aniso_B_23`
- `_atom_site_aniso_B_33`
- `_atom_site_aniso_label`
→ `_atom_site_label`
- `_atom_site_aniso_ratio`
- `_atom_site_aniso_type_symbol`
→ `_atom_site_type_symbol`
- `_atom_site_aniso_U_11`
- `_atom_site_aniso_U_12`
- `_atom_site_aniso_U_13`
- `_atom_site_aniso_U_22`
- `_atom_site_aniso_U_23`
- `_atom_site_aniso_U_33`
- `_atom_site_attached_hydrogens`
- `_atom_site_B_equiv_geom_mean`
- `_atom_site_B_iso_or_equiv`
- `_atom_site_calc_attached_atom`
- `_atom_site_calc_flag`
- `_atom_site_Cartn_x`
- `_atom_site_Cartn_y`
- `_atom_site_Cartn_z`
- `_atom_site_chemical_conn_number`
→ `_chemical_conn_atom_number`
- `_atom_site_constraints`
- `_atom_site_description`
- `_atom_site_disorder_assembly`
- `_atom_site_disorder_group`
- `_atom_site_fract_x`
- `_atom_site_fract_y`
- `_atom_site_fract_z`
- `_atom_site_label_component_0`
- `_atom_site_label_component_1`

3.2. CLASSIFICATION AND USE OF CORE DATA

```

_atom_site_label_component_2
_atom_site_label_component_3
_atom_site_label_component_4
_atom_site_label_component_5
_atom_site_label_component_6
_atom_site_occupancy
† _atom_site_refinement_flags
_atom_site_refinement_flags_adp
_atom_site_refinement_flags_occupancy
_atom_site_refinement_flags_posn
_atom_site_restraints
_atom_site_symmetry_multiplicity
† _atom_site_thermal_displace_type
_atom_site_type_symbol
  → _atom_type_symbol
_atom_site_U_equiv_geom_mean
_atom_site_U_iso_or_equiv
_atom_site_Wyckoff_symbol

```

The bullet (●) indicates a category key. For this category an alternative category key can be formed by taking all the `_atom_site_label_component_*` items together. Anisotropic displacement parameters may also be listed in a separate loop, for which `_atom_site_aniso_label` forms the key. The arrow (→) is a reference to a parent data item. The dagger (†) indicates a deprecated item, which should not be used in the creation of new CIFs.

Data items in the ATOM_SITE category represent the positions of atom sites identified in the structural model, their spatial distribution defined by isotropic or anisotropic displacement parameters, details of restraints or constraints applied during the refinement, and the interpretation of their occupancy due to structural or compositional disorder.

Example 3.2.4.1 is a typical extract from a list of atom-site coordinates, with equivalent isotropic displacement values and refinement conditions. Each site is identified by `_atom_site_label`.

The coordinates are specified as fractional x , y , z values along the unit-cell axes. Coordinates may also be specified in ångström units along orthogonal Cartesian axes using the data names `_atom_site_Cartn_x`, `_atom_site_Cartn_y` and `_atom_site_Cartn_z`. The transformation matrix between Cartesian and fractional coordinates can be given in the ATOM_SITES category.

(Note that occupancy values are unaffected by symmetry. This is discussed later in connection with site multiplicity.)

`_atom_site_U_iso_or_equiv` records the isotropic atomic displacement value U_{iso} in the case of isotropic refinement. In the case of anisotropic refinement, `_atom_site_U_iso_or_equiv` records the equivalent isotropic value U_{eq} , defined as

$$U_{\text{eq}} = (1/3) \sum_i \left[\sum_j (U^{ij} a_i^* a_j^* a_i a_j) \right],$$

where a_i are the real-space cell lengths, a_j^* are the reciprocal-space cell lengths and U^{ij} are the anisotropic displacement parameters.

The data item `_atom_site_adp_type` identifies which value is given. An alternative equivalent isotropic displacement parameter `_atom_site_U_equiv_geom_mean` may be calculated as the geometric mean of the anisotropic parameters,

$$U_{\text{eq}} = (U_i U_j U_k)^{1/3},$$

where the U_i are the principal components of the orthogonalized U^{ij} .

Data names also exist for the corresponding quantities calculated from B values, although the use of B values is discouraged by the IUCr Commission on Crystallographic Nomenclature.

For each site, `_atom_site_calc_flag` takes one of the following values: `d`, to indicate that the atom-site coordinates were determined from the diffraction intensities; `c` or `calc` to indicate that they were calculated from molecular geometry considerations; or `dum`, for a dummy site.

Example 3.2.4.1. List of atom-site coordinates, equivalent isotropic U values and refinement conditions.

```

loop_
  _atom_site_label
  _atom_site_fract_x
  _atom_site_fract_y
  _atom_site_fract_z
  _atom_site_U_iso_or_equiv
  _atom_site_adp_type
  _atom_site_calc_flag
  _atom_site_refinement_flags_posn
  _atom_site_occupancy
  _atom_site_disorder_assembly
  _atom_site_disorder_group
  _atom_site_type_symbol
O1 0.5000 1.0000 0.8011 (2) 0.0259 (6)
  Uni d S 1 . . O
O2 0.33569 (10) 0.98239 (10) 0.88892 (17) 0.0321 (5)
  Uni d . 1 . . O
O3 0.20150 (13) 0.92560 (11) 0.8817 (2) 0.0458 (5)
  Uni d . 1 . . O
O4 0.35539 (11) 0.81530 (10) 0.96958 (17) 0.0333 (5)
  Uni d . 1 . . O
C1 0.43883 (16) 0.95672 (14) 0.7293 (3) 0.0275 (6)
  Uni d . 1 . . C
H1 0.4064 0.9930 0.6730 0.0493 (19)
  Uiso calc R 1 . . H
C2 0.37292 (16) 0.92010 (14) 0.8175 (3) 0.0266 (6)
  Uni d . 1 . . C
H2 0.3246 0.8945 0.7691 0.0493 (19)
  Uiso calc R 1 . . H
C3 0.41827 (16) 0.85968 (14) 0.8983 (3) 0.0280 (6)
  Uni d . 1 . . C
H3 0.4629 0.8850 0.9536 0.0493 (19)
  Uiso calc R 1 . . H

```

Specific restraints or constraints applied to a site may be indicated by one or more of the `_atom_site_refinement_flags_*` items.

The data item `_atom_site_occupancy` defines the fraction of the atom type present at the site. Note that the same site may occur more than once in the list, identified by separate values of `_atom_site_label`. Such an arrangement would represent contributions from separate atom types (perhaps in modelling compositional disorder). The sum of occupancies of all atom types present at a single site may not significantly exceed 1.0 (unless it is a dummy site with no physical significance). Note that an atom of a given chemical species positioned on a special position (e.g. on a twofold axis) will in general be assigned a full occupancy value of 1.0. However, it will occur less often in the unit cell than an atom on a general position (in this example by a factor of 2). To account for this in structure-factor calculations it may be given a population value of 0.5 within the refinement program. A population adjustment of this kind is *not* implied in the assignment of a value to `_atom_site_occupancy`. The multiplicity of the site owing to the space-group symmetry can be recorded in `_atom_site_symmetry_multiplicity`.

The disorder-related data names in this example will be discussed below.

`_atom_site_type_symbol` is a code which must match an entry in the ATOM_TYPE category that supplies information about the elemental composition and scattering factors of the atom or atoms occupying the site. Note that it is quite legitimate to have an atom-type symbol such as 'Fe3+Ni2+', referring to a mixed-composition atom site. The effective physical properties of such a pseudo-atom should be given in full in the ATOM_TYPE category.

Example 3.2.4.2 demonstrates how the anisotropic displacement parameters are conventionally broken out into a separate list. When this is done, each atom site is identified by `_atom_site_aniso_label`, and this must of course match the value of `_atom_site_label` specifying the position of the site.

3. CIF DATA DEFINITION AND CLASSIFICATION

Example 3.2.4.2. *Separate list of anisotropic U values with `_atom_site_aniso_label` acting as the key that uniquely identifies table rows in this listing.*

```
loop_
  _atom_site_aniso_label
  _atom_site_aniso_U_11
  _atom_site_aniso_U_22
  _atom_site_aniso_U_33
  _atom_site_aniso_U_12
  _atom_site_aniso_U_13
  _atom_site_aniso_U_23
O1 0.0256(9) 0.0275(9) 0.0433(12) 0.0020(8)
    0.0038(8) -0.0062(9)
O2 0.0306(10) 0.0418(11) 0.0651(15) -0.0037(10)
    0.0079(11) -0.0049(11)
O3 0.0374(10) 0.0309(10) 0.0315(11) -0.0030(9)
    0.0051(9) 0.0005(9)
C1 0.0279(13) 0.0245(13) 0.0300(15) 0.0004(11)
    -0.0020(12) -0.0034(12)
C2 0.0258(13) 0.0226(13) 0.0314(15) 0.0024(11)
    -0.0024(12) -0.0033(12)
C3 0.0292(13) 0.0246(13) 0.0302(16) -0.0028(11)
    0.0031(12) 0.0005(12)
```

The data item `_atom_site_label` is normally used as the identifier of each individual atom site in a list of coordinates and atomic displacement factors. Historically, the labels given to atom sites have been chosen to summarize useful information about the atom located at the site. Almost invariably the label contains the symbol of the chemical element or elements occupying the site; there may also be indicators of charge, valence, chemical connectivity, disorder, occupation of a site of crystallographic symmetry or grouping within a component of secondary structure within large molecules. In a CIF, it is formally sufficient that atom-site labels are unique, as all the information about composition, valence, connectivity and so on can be extracted from the data items designed specifically to record this information. However, it is preferable that an atom-site label should summarize the relevant features of the site. Many styles and conventions for labelling atoms are in use in crystallography, so to enable interchange with other crystallographic data file formats, the core dictionary contains a detailed but highly flexible set of rules for constructing and parsing atom-site labels.

Labelling atom sites in crystallography usually serves two distinct purposes: (a) to identify the site in the molecule and crystal, and (b) to identify the chemical element that occupies that site. The core dictionary makes this distinction clear by defining `ATOM_SITE` and `ATOM_TYPE` as separate data categories. The connection between the two categories is made through the equivalence of the data items `_atom_site_type_symbol` (in the `ATOM_SITE` list) and `_atom_type_symbol` (in the `ATOM_TYPE` list). Often, however, crystallographers use a single label `_atom_site_label` to define both the site and the chemical species occupying it.

The `_atom_site_label` may be composed of as many as eight separate components; the recommended convention for construction of the string is as follows.

Component 0 [optionally identical to a value of `_atom_type_symbol`] (*mandatory*): A character string containing any character except a blank or an underline, with the proviso that each digit '0'–'9' is used only to designate an oxidation state and, as such, must be followed by a plus '+' or a minus '-' character. It is recommended that the element symbols be used when applicable. Examples of permissible codes are: Cu, Cu²⁺, dummy, Fe³⁺Ni²⁺, S⁻, H*, H(SDS).

Component 1 [atom number code] (*optional*): This string may contain any alphanumeric character except a blank or an underline, but the first character *must* be a digit '0'–'9' and the second character may not be a plus '+' or a minus '-'. Component 1 is

intended primarily to differentiate sites containing the same atom type, but it can be used for any purpose. Examples of combined component 0 and 1 codes are: C1, C103g28, **Fe3+17b**, **H*251**, **boron2a**, **Ni2+2**, **Fe2+Ni2+2**, where component 0 is in bold to indicate how these labels are parsed.

Component 2 [residue code] (*optional*): This string may contain any character except a blank or underline. It is intended primarily to give specific structural information such as the molecular fragment or amino-acid type, e.g. C1_gly, O1_SO4. If component 2 is present, it is separated from the concatenated components 0 and 1 with an underline character.

Components 3–7 [sequence, remoteness, chain order, alternate, footnote codes] (*optional*): These strings may contain any character except a blank or an underline. The underline character is used to separate the individual components. The names associated with the separate components suggest their roles in constructing composite labels that match the conventions of site labelling in the PDB format for macromolecular structure files. However, they are not restricted to these functions and may be used in other ways.

Component 0 is normally identical to an `_atom_type_symbol` code in the `ATOM_TYPE` list. However, if it is not, an `_atom_site_type_symbol` code must appear in the `ATOM_TYPE` list in order to identify the atom type. In these cases, component 0 may contain any code consistent with the rules given in the dictionary. Thus, component 0 could be `ca` to identify an alpha carbon, provided that the `_atom_site_type_symbol` is encoded as `c` to indicate that the atom type is carbon.

Multiple occupation of a single atom site by different atom species (compositional disorder) may be handled simply by having multiple values of `_atom_site_label` referring to the same site in the crystal structure. Alternatively, multiple occupancy of an atom site may be denoted by a unique character or characters in component 0 of the atom label, with the `ATOM_TYPE` list containing the equivalent pseudo element label entry with values that are weighted averages of those for the constituent elements. The proportions of the atom types should then be defined using `_atom_type_description`.

This `_atom_site_label` construction is flexible, visually decipherable and well suited to computer applications. The components can be easily identified and stripped with a single pass, from left to right, along the label string. Note that the underline separators are only used if higher-order components exist. If intermediate components are not used they may be omitted provided the underline separators are retained. For example, the label `C233_ggg` is acceptable and contains the components 0: `c`, 1: `233`, 2: `null` and 3: `ggg`. There is no requirement that the same number of components should be used in each label.

The `_atom_site_label` may be replaced by separate data items specifying the individual components of an atom label; this may be useful for large lists of site coordinates, for example in a macromolecular structure, where site-labelling components follow a systematic convention and where subsets of the atom sites need to be searched for or extracted using individual label components. Such uses are not common in files built with core CIF data names; the mmCIF dictionary identifies substructural components in biological macromolecules by alternative techniques (Section 3.6.7).

There is no comparable fragmentation of the components of `_atom_site_aniso_label`. Where separate lists of anisotropic displacement parameters use complex atom-site labels, either the coordinate list should use `_atom_site_label` alone or the processing software needs to be able to construct a value for `_atom_site_label` from the separate components `_atom_site_label_component_*` in order to test the equivalence between the

3.2. CLASSIFICATION AND USE OF CORE DATA

Example 3.2.4.3. *Chemical connectivity table; atoms are linked back to atom-site positions through matching values of `_atom_site_chemical_conn_number` and `_chemical_conn_atom_number`.*

```
loop_
  _atom_site_label
  _atom_site_chemical_conn_number
  _atom_site_fract_x
  _atom_site_fract_y
  _atom_site_fract_z
  _atom_site_U_iso_or_equiv
S1 1 0.74799(9) -0.12482(11) 0.27574(9) 0.0742(3)
S2 2 1.08535(10) 0.16131(9) 0.34061(9) 0.0741(3)
N1 3 1.0650(2) -0.1390(2) 0.2918(2) 0.0500(5)
Cl 4 0.9619(3) -0.0522(3) 0.3009(2) 0.0509(6)
# - - - data truncated for brevity - - -

loop_
  _chemical_conn_atom_number
  _chemical_conn_atom_type_symbol
  _chemical_conn_atom_display_x
  _chemical_conn_atom_display_y
  _chemical_conn_atom_NCA
  _chemical_conn_atom_NH
1 S .39 .81 1 0 2 S .39 .96 2 0
3 N .14 .88 3 0 4 C .33 .88 3 0
5 C .11 .96 2 2 6 C .03 .96 2 2
# - - - data truncated for brevity - - -
```

Example 3.2.4.4. *Handling of occupational disorder of atom sites.*

```
loop_
  _atom_site_label
  _atom_site_fract_x
  _atom_site_fract_y
  _atom_site_fract_z
  _atom_site_occupancy
  _atom_site_disorder_assembly
  _atom_site_disorder_group
B2 0.9639(7) 0.6536(5) 0.4464(5) 1 . .
F21 0.9411(18) 0.7083(11) 0.5388(9) 0.50 A 1
F22 1.008(2) 0.5331(6) 0.4747(10) 0.50 A 1
F23 0.8364(14) 0.6845(17) 0.3951(15) 0.50 A 1
F24 1.0718(17) 0.6896(14) 0.3764(12) 0.50 A 1
F21A 0.9727(18) 0.7141(15) 0.5293(13) 0.50 A 2
F22A 1.0540(15) 0.5401(8) 0.4635(15) 0.50 A 2
F23A 0.8216(9) 0.6479(15) 0.4461(14) 0.50 A 2
F24A 1.007(2) 0.7096(17) 0.3476(11) 0.50 A 2
```

labels in the coordinates and anisotropic displacement parameters lists.

While either atom-labelling technique is permitted, it is recommended that the individual label components are *not* used unless there is an overwhelming argument to do so.

Information about the molecular model is sometimes embedded in a labelling convention. In CIF, this information is usually expressed through other data items.

The connectivity of a molecule is described by the CHEMICAL group of categories, and more specifically through the CHEMICAL_CONN_ATOM and CHEMICAL_CONN_BOND categories.

The link between atom sites in the coordinate list and the corresponding atoms in the molecular model is established using the data item `_chemical_conn_atom_number` in the CHEM_CONN_ATOM category, and the data items `_chemical_conn_bond_atom_1` and `_chemical_conn_bond_atom_2` in the CHEMICAL_CONN_BOND category. The values of these data items must match values for the data item `_atom_site_chemical_conn_number` in the ATOM_SITE list. Example 3.2.4.3 shows an extract from a connectivity table; a more complete version of this table is given in the relevant category descriptions in the dictionary.

Note that there is no guarantee that the refined atom-site coordinates that characterize the asymmetric unit will correspond to loca-

tions within a single connected molecular species. Crystal symmetry transformations may need to be applied to individual sites in order to map the contents of a connected molecular residue to real space in the unit cell. There is no provision in the CHEMICAL_CONN categories for the specification of these symmetry transformations; thus these higher-order molecular geometries are best described using data items in the GEOM categories, which do allow for the specification of symmetry transformations.

It may also be the case that not all atom positions have been located; this is particularly true for hydrogen atoms, and the data item `_atom_site_attached_hydrogens` is provided for book-keeping purposes to indicate hydrogen atoms known to be bonded to an atom but whose positions have not been refined (or calculated).

Example 3.2.4.4 shows how the disorder of a group of bonded atoms over a set of atom sites (occupational disorder) is described. In this example of a disordered tetrafluoroborate anion, the data item `_atom_site_disorder_assembly` takes the value A, and the data item `_atom_site_disorder_group` takes the values 1 and 2, indicating the two alternative positions of the disordered group.

The remaining items in this category are clearly described in their individual dictionary entries. However, the now-deprecated data item `_atom_site_refinement_flags` should be mentioned. This was allowed to take values obtained by concatenating one or more of the single-letter flags:

- . no refinement constraints;
- S special-position constraint on site;
- G rigid-group refinement of site;
- R riding-atom site attached to non-riding atom;
- D distance or angle restraint on site;
- T thermal displacement constraints;
- U U_{iso} or U^{ij} restraint (rigid bond);
- P partial occupancy constraint.

These individual flags are listed in the dictionary using the DDL field `_enumeration`, which denotes a list of mutually exclusive permitted values. As concatenation of values is allowed here, dictionary-based software must be modified to handle this data item as a special case. To avoid the need for this in future, the data item was marked as deprecated from version 2.3 of the dictionary, and is replaced by the three separate items `_atom_site_refinement_flags_posn`, `*_adp` and `*_occupancy`. For each of these, the relevant combinations of refinement flags are fully enumerated (for example `_atom_site_refinement_flags_adp` may take any one of the values T, U or TU). This logically separates the different types of refinement constraints or restraints that an author might want to record and allows software to parse the data item.

3.2.4.1.2. Collections of atom sites

The data items in this category are as follows:

```
ATOM_SITES
  _atom_sites_Cartn_tran_matrix_11
  _atom_sites_Cartn_tran_matrix_12
  _atom_sites_Cartn_tran_matrix_13
  _atom_sites_Cartn_tran_matrix_21
  _atom_sites_Cartn_tran_matrix_22
  _atom_sites_Cartn_tran_matrix_23
  _atom_sites_Cartn_tran_matrix_31
  _atom_sites_Cartn_tran_matrix_32
  _atom_sites_Cartn_tran_matrix_33
  _atom_sites_Cartn_tran_vector_1
  _atom_sites_Cartn_tran_vector_2
  _atom_sites_Cartn_tran_vector_3
  _atom_sites_Cartn_transform_axes
  _atom_sites_fract_tran_matrix_11
  _atom_sites_fract_tran_matrix_12
  _atom_sites_fract_tran_matrix_13
```

3. CIF DATA DEFINITION AND CLASSIFICATION

```
_atom_sites_fract_tran_matrix_21
_atom_sites_fract_tran_matrix_22
_atom_sites_fract_tran_matrix_23
_atom_sites_fract_tran_matrix_31
_atom_sites_fract_tran_matrix_32
_atom_sites_fract_tran_matrix_33
_atom_sites_fract_tran_vector_1
_atom_sites_fract_tran_vector_2
_atom_sites_fract_tran_vector_3
_atom_sites_solution_hydrogens
_atom_sites_solution_primary
_atom_sites_solution_secondary
_atom_sites_special_details
```

This category records information that applies collectively to the atom sites of the structural model. At present, the topics covered are the transformation matrix between Cartesian and cell fractional coordinates, and the methods used to locate the initial atom sites. `_atom_sites_solution_primary` describes how the first atom sites were determined, `_atom_sites_solution_secondary` describes how the remaining non-hydrogen sites were located and `_atom_sites_solution_hydrogens` describes how hydrogen atoms were located. The codes that are allowed for each of these refer to distinct solution methods, and at present only the seven formal values listed below are provided (although other values might be added in the future):

```
difmap  difference-electron-density map;
vecmap  real-space vector search;
heavy   heavy-atom method;
direct  structure-invariant direct methods;
geom    inferred from neighbouring sites;
disper  anomalous-dispersion techniques;
isomor  isomorphous structure methods.
```

3.2.4.1.3. Atom types

The data items in this category are as follows:

ATOM_TYPE

- `_atom_type_symbol`
- `_atom_type_analytical_mass_%`
- `_atom_type_description`
- `_atom_type_number_in_cell`
- `_atom_type_oxidation_number`
- `_atom_type_radius_bond`
- `_atom_type_radius_contact`
- `_atom_type_scatter_Cromer_Mann_a1`
- `_atom_type_scatter_Cromer_Mann_a2`
- `_atom_type_scatter_Cromer_Mann_a3`
- `_atom_type_scatter_Cromer_Mann_a4`
- `_atom_type_scatter_Cromer_Mann_b1`
- `_atom_type_scatter_Cromer_Mann_b2`
- `_atom_type_scatter_Cromer_Mann_b3`
- `_atom_type_scatter_Cromer_Mann_b4`
- `_atom_type_scatter_Cromer_Mann_c`
- `_atom_type_scatter_dispersion_imag`
- `_atom_type_scatter_dispersion_real`
- `_atom_type_scatter_dispersion_source`
- `_atom_type_scatter_length_neutron`
- `_atom_type_scatter_source`
- `_atom_type_scatter_versus_stol_list`

The bullet (•) indicates a category key.

The data items in this category record details about the atomic species associated with each occupied atom site in the structural model. While these will typically be standard properties of the naturally occurring chemical elements, they may also be synthetic atom types, for example in cases where a single atom site may be occupied with partial occupancies by atoms of different elements.

As mentioned in Section 3.2.4.1.1, there are two ways of dealing with such a case: the same location in the coordinate list may be populated by multiple entries, each for an atom of a particular element with an associated occupancy fraction; or a single entry

Example 3.2.4.5. Reference to atomic scattering factors.

```
loop_
_atom_type_symbol
_atom_type_oxidation_number
_atom_type_number_in_cell
_atom_type_scatter_dispersion_real
_atom_type_scatter_dispersion_imag
_atom_type_scatter_source
C 0 72 .017 .009
International Tables Vol IV Table 2.2B
H 0 100 0 0
International Tables Vol IV Table 2.2B
O 0 12 .047 .032
International Tables Vol IV Table 2.2B
N 0 4 .029 .018
International Tables Vol IV Table 2.2B
```

may be made for the synthetic atom type, the properties of which are described fully in the ATOM_TYPE list.

Each different atom type has a unique `_atom_type_symbol` identifier. In principle, this could be any string of characters, but the dictionary recommends certain conventions to encourage compatibility with the atom-site labelling rules. It is recommended that the identifier be the normal chemical element symbol when the atom type is a pure element. If some other labelling is used, the identifier may be composed of any character except an underline, with the additional proviso that digits designate an oxidation state and must be followed by a '+' or '-' character.

The data item `_atom_type_scatter_versus_stol_list` can be used to give a table of scattering factors as a function of $(\sin \theta)/\lambda$. This is a text field with no specified internal structure, except the suggestion that it is well commented and the lists should be regularly formatted. However, it is generally enough to list the atomic scattering factors of each element and to provide a reference to the source of the values, as in Example 3.2.4.5.

3.2.4.2. Chemical identification and connectivity information

The categories describing chemical identity and connectivity are as follows:

CHEMICAL group

Chemical identification (§3.2.4.2.1)

CHEMICAL

CHEMICAL_FORMULA

Chemical connectivity (§3.2.4.2.2)

CHEMICAL_CONN_ATOM

CHEMICAL_CONN_BOND

As indicated in Section 3.2.4.1.1, the chemical interpretation of the coordinate list of regions of significant electron density is not always easy. Occupational and compositional disorder, symmetry-equivalent locations, and unrefined atom sites all contribute to the difficulties, but it is usually possible in modern studies to construct a sensible chemical model. The CHEMICAL category group provides the data names needed to describe the chemical identity and properties of the material characterized in the structural study.

3.2.4.2.1. Chemical identification

The data items in these categories are as follows:

(a) CHEMICAL

```
_chemical_absolute_configuration
_chemical_compound_source
_chemical_melting_point
_chemical_melting_point_gt
_chemical_melting_point_lt
_chemical_name_common
_chemical_name_mineral
_chemical_name_structure_type
_chemical_name_systematic
```

3.2. CLASSIFICATION AND USE OF CORE DATA

```
_chemical_optical_rotation
_chemical_properties_biological
_chemical_properties_physical
_chemical_temperature_decomposition
_chemical_temperature_decomposition_gt
_chemical_temperature_decomposition_lt
_chemical_temperature_sublimation
_chemical_temperature_sublimation_gt
_chemical_temperature_sublimation_lt
```

(b) CHEMICAL_FORMULA

```
_chemical_formula_analytical
_chemical_formula_iupac
_chemical_formula_moiety
_chemical_formula_structural
_chemical_formula_sum
_chemical_formula_weight
_chemical_formula_weight_meas
```

The CHEMICAL category itself deals with the large-scale chemical properties of the compound from which the crystal under study was formed: its various formal and common names, its source, melting point, decomposition and sublimation temperatures (as experimentally determined values, or as upper or lower possible values if not measured directly), its biological or physical properties, and where applicable the absolute configuration and optical rotation.

The optical rotation in solution may be reported using the data name `_chemical_optical_rotation` by an expression of the form

$$[\alpha]_W^T = \pm \frac{100\alpha}{lc} \quad (c = \text{CONC}, \text{SOLV}),$$

where $[\alpha]_W^T$ is the signed optical rotation in degrees at temperature T and wavelength labelled by code W , l is the length of the optical cell, CONC is the concentration of the solution (given as the mass of the substance in g in a standard 100 ml of solution), and SOLV is the chemical formula of the solvent. This can be marked up within the constraints of the ASCII character set to which CIF is restricted as `[\a]^25^~D~ = +108 (c = 3.42, CHCl~3~)`, where the measurement is taken using the D line of the atomic spectrum of sodium.

Data items in the CHEMICAL_FORMULA category describe the chemical formula and formula mass of the compound under study. The quoted formula must reflect the overall stoichiometry of the crystal under study, and must, when multiplied by the Z value `_cell_formula_units_z`, account for the total contents of the unit cell.

A number of data names are provided to account for different conventions in the presentation of chemical formulae. `_chemical_formula_analytical` is appropriate for a gross formula determined by standard chemical analysis, including all trace elements identified in the sample. Standard uncertainties on the proportions of elements present are acceptable, e.g.

```
_chemical_formula_analytical 'Fe2.45(2) Ni1.60(3) S4'
```

`_chemical_formula_sum` is another aggregate formula, in which all discrete bonded residues and ions are summed over the constituent elements. Where appropriate, the formulae of separate residues of a complex may be described by `_chemical_formula_moiety`, in which the formula for each moiety is supplied as a sum of the individual elements within the moiety, or by `_chemical_formula_structural`, in which sub-components within individual moieties are further identified, so that the overall expression permits the identification of particular bonded groups. Within these formula expressions, certain rules must be observed to allow parsing by software. The final data item relating to the chemical formula, `_chemical_formula_iupac`, is for formulae that

are constructed according to the rules of the International Union for Pure and Applied Chemistry.

The ordering and notation rules are explained in detail in the dictionary, but are repeated here for convenience. Within each group of atoms for which a formula is present:

- (i) only recognized element symbols may be used;
- (ii) each element symbol is followed by a 'count' number ('1' is implicit and may be omitted);
- (iii) a space or parenthesis must separate each cluster of (element symbol + count);
- (iv) where a group of elements is enclosed in parentheses, the multiplier for the group must follow the closing parentheses. That is, all element and group multipliers are assumed to be printed as subscripted numbers. (An exception to this rule exists for `_chemical_formula_moiety`, where pre- and post-multipliers are permitted for molecular units.)
- (v) Unless the elements are ordered in a manner that corresponds to their chemical structure, as in `_chemical_formula_structural`, the order of the elements within any group or moiety depends on whether or not carbon is present. If carbon is present, the order should be: C, then H, then the other elements in alphabetical order of their symbol. If carbon is not present, the elements are listed purely in alphabetic order of their symbol. This is the 'Hill' system used by *Chemical Abstracts*. This ordering is used in `_chemical_formula_moiety` and `_chemical_formula_sum`.

For `_chemical_formula_moiety` some additional rules apply:

- (i) Moieties are separated by commas, ','.
- (ii) The order of elements within a moiety follows the general rules outlined above as the 'Hill' system.
- (iii) Parentheses are *not* used within moieties but may surround a moiety. Parentheses may not be nested.
- (iv) Charges should be placed at the end of the moiety. The charge '+' or '-' may be preceded by a numerical multiplier and should be separated from the last (element symbol + count) by a space. Pre- or post-multipliers may be used for individual moieties.

Example 3.2.4.6 illustrates the differences between some of these data items.

3.2.4.2.2. Chemical connectivity

The data items in these categories are as follows:

(a) CHEMICAL_CONN_ATOM

- `_chemical_conn_atom_number`
- `_chemical_conn_atom_charge`
- `_chemical_conn_atom_display_x`
- `_chemical_conn_atom_display_y`
- `_chemical_conn_atom_NCA`
- `_chemical_conn_atom_NH`
- `_chemical_conn_atom_type_symbol`

(b) CHEMICAL_CONN_BOND

- `_chemical_conn_bond_atom_1`
→ `_chemical_conn_atom_number`
- `_chemical_conn_bond_atom_2`
→ `_chemical_conn_atom_number`
- `_chemical_conn_bond_type`

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. The arrow (→) is a reference to a parent data item.

The CHEMICAL_CONN_ATOM category labels the chemical atoms in a connected representation of the molecular species and can also give the coordinates for the atoms in a two-dimensional chemical diagram (Example 3.2.4.7). Each atom may also carry an indication of the number of connected non-hydrogen atoms (`*_NCA`) and the number of hydrogen atoms (`*_NH`) to which it

3. CIF DATA DEFINITION AND CLASSIFICATION

Example 3.2.4.6. *Different representations of a chemical formula.*

```
_chemical_formula_iupac  '[Mo (C O)4 (C18 H33 P)2]'
```

```
_chemical_formula_moiety  'Mo,4(C O),2(C18 H33 P)'
```

```
_chemical_formula_structural
```

```
'((C O)4 (P (C6 H11)3)2)Mo'
```

```
_chemical_formula_sum      'C40 H66 Mo O4 P2'
```

Example 3.2.4.7. *Representation of a two-dimensional chemical diagram.*

```
loop_
  _chemical_conn_atom_number
  _chemical_conn_atom_type_symbol
  _chemical_conn_atom_display_x
  _chemical_conn_atom_display_y
  _chemical_conn_atom_NCA
  _chemical_conn_atom_NH
  1  S   .39 .81  1  0
  2  S   .39 .96  2  0
  3  N   .14 .88  3  0
  4  C   .33 .88  3  0
  5  C   .11 .96  2  2
  6  C   .03 .96  2  2
  7  C   .03 .80  2  2
  8  C   .11 .80  2  2
  9  S   .54 .81  1  0
 10  S   .54 .96  2  0
 11  N   .80 .88  3  0
 12  C   .60 .88  3  0
 13  C   .84 .96  2  2
 14  C   .91 .96  2  2
 15  C   .91 .80  2  2
 16  C   .84 .80  2  2
```

is connected. Together with the CHEMICAL_CONN_BOND category, the data items in the CHEMICAL_CONN_ATOM category provide a basic description of the chemical structure. Although the description of the chemical structure provided in these two categories is not as extensive as the information that may be conveyed in a molecular information file (Chapter 2.4), it should allow a substructure to be searched for in a suitable database.

The CHEMICAL_CONN_BOND category lists pairs of atoms that contribute to chemical bonds and describes the nature of the bond between them (Example 3.2.4.8). Taken with data items in the CHEMICAL_CONN_ATOM category, data items in this category complete the basic description of a molecular entity.

Bond types are assigned from a list that specifies single, double, triple, quadruple, aromatic, polymeric, delocalized double and π bonds. These are not intended to cover all possible cases, but to characterize a molecular model suitable for database substructure searching.

3.2.4.3. Molecular or packing geometry

The categories describing geometry are as follows:

GEOM group

GEOM

GEOM_ANGLE

GEOM_BOND

GEOM_CONTACT

GEOM_HBOND

GEOM_TORSION

The molecular and packing geometry can be calculated fully given the unit-cell parameters, the space group and a list of atom sites. Therefore, all the information about geometry in the GEOM category group is derivative. However, it is useful to record it within the file both as a check on the primary information stored in other categories and as a method for flagging values to be published.

Example 3.2.4.8. *Bond types in a chemical connectivity table.*

```
loop_
  _chemical_conn_bond_atom_1
  _chemical_conn_bond_atom_2
  _chemical_conn_bond_type
  4  1  doub  4  3  sing
  4  2  sing  5  3  sing
  6  5  sing  7  6  sing
  8  7  sing  8  3  sing
 10  2  sing 12  9  doub
 12 11  sing 12 10  sing
 13 11  sing 14 13  sing
 15 14  sing 16 15  sing
 16 11  sing 17  5  sing
 18  5  sing 19  6  sing
 20  6  sing 21  7  sing
 22  7  sing 23  8  sing
 24  8  sing 25 13  sing
 26 13  sing 27 14  sing
 28 14  sing 29 15  sing
 30 15  sing 31 16  sing
 32 16  sing
```

3.2.4.3.1. Contents of the geometry-related categories

The data items in these categories are as follows:

(a) GEOM

_geom_special_details

(b) GEOM_ANGLE

- _geom_angle_atom_site_label_1
→ _atom_site_label
- _geom_angle_atom_site_label_2
→ _atom_site_label
- _geom_angle_atom_site_label_3
→ _atom_site_label
- _geom_angle_site_symmetry_1
- _geom_angle_site_symmetry_2
- _geom_angle_site_symmetry_3
- _geom_angle
- _geom_angle_publ_flag

(c) GEOM_BOND

- _geom_bond_atom_site_label_1
→ _atom_site_label
- _geom_bond_atom_site_label_2
→ _atom_site_label
- _geom_bond_site_symmetry_1
- _geom_bond_site_symmetry_2
- _geom_bond_distance
- _geom_bond_publ_flag
- _geom_bond_valence

(d) GEOM_CONTACT

- _geom_contact_atom_site_label_1
→ _atom_site_label
- _geom_contact_atom_site_label_2
→ _atom_site_label
- _geom_contact_site_symmetry_1
- _geom_contact_site_symmetry_2
- _geom_contact_distance
- _geom_contact_publ_flag

(e) GEOM_HBOND

- _geom_hbond_atom_site_label_A
→ _atom_site_label
- _geom_hbond_atom_site_label_D
→ _atom_site_label
- _geom_hbond_atom_site_label_H
→ _atom_site_label
- _geom_hbond_site_symmetry_A
- _geom_hbond_site_symmetry_D
- _geom_hbond_site_symmetry_H
- _geom_hbond_angle_DHA
- _geom_hbond_distance_DA
- _geom_hbond_distance_DH
- _geom_hbond_distance_HA
- _geom_hbond_publ_flag

```
(f) GEOM_TORSION
• _geom_torsion_atom_site_label_1
  → _atom_site_label
• _geom_torsion_atom_site_label_2
  → _atom_site_label
• _geom_torsion_atom_site_label_3
  → _atom_site_label
• _geom_torsion_atom_site_label_4
  → _atom_site_label
• _geom_torsion_site_symmetry_1
• _geom_torsion_site_symmetry_2
• _geom_torsion_site_symmetry_3
• _geom_torsion_site_symmetry_4
  _geom_torsion
  _geom_torsion_publ_flag
```

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. **_symmetry_** items have a default value and may be omitted from the list. The arrow (→) is a reference to a parent data item.

Most categories within this group record distances or angles specified by atom-site labels and are well characterized. The GEOM category currently provides the single data name `_geom_special_details` in which any other details of the geometry that an author considers noteworthy may be stored. Examples of information that might be stored in this data item are least-squares equations of planes, out-of-plane distances, dihedral angles between planes and general comments about the calculation of standard uncertainties.

A subtlety in the geometry-related categories arises from the need to record geometric relationships that involve atoms that are not listed in the ATOM_SITE coordinate list, but that can be derived from the coordinates in this list by the application of a crystallographic symmetry transformation. Thus atom sites in the geometry lists are identified both by their atom-site labels (which must identically match one of the entries in the ATOM_SITE list) and by the code for the symmetry transformation that has been applied to the initial location. Since the atom-site labels may refer to atoms in their original location as well as to atoms in symmetry-related locations, the formal key for these categories involves the site labels as well as the symmetry codes. However, in many cases (as discussed further below) the symmetry codes may be absent from a list, and a parser must supply suitable default or null values for the missing components when constructing or checking a complete key.

In many cases, interest is focused on intramolecular distances and angles, and on intramolecular contacts within a single asymmetric unit. In such cases, the geometry lists would contain only atoms listed explicitly in the ATOM_SITE list and the symmetry codes all refer trivially to the identity transformation.

The examples in this section demonstrate various ways of handling geometry lists with trivial or non-trivial symmetry transformations. In Example 3.2.4.9, showing treatment of bond angles, the relevant data items (`_geom_angle_site_symmetry_*`) are absent, which is one method for indicating the identity transformation. Dictionary validation software must therefore be able to handle both the presence and absence of these components of the formal category key.

The symmetry transformations in this and related categories take the form of codes '*n klm*' or *n_klm*, where *n* refers to the symmetry operation that is applied to the coordinates stored in `_atom_site_fract_x`, `_atom_site_fract_y` and `_atom_site_fract_z`. The value of *n* must match a number given in `_symmetry_equiv_pos_site_id`. *k*, *l* and *m* refer to the translations that are subsequently applied to the symmetry-transformed coordinates to generate the atom used in calculating the contact. These translations (*x*, *y*, *z*) are related to (*k*, *l*, *m*) by

Example 3.2.4.9. List of bond angles.

```
loop_
  _geom_angle_atom_site_label_1
  _geom_angle_atom_site_label_2
  _geom_angle_atom_site_label_3
  _geom_angle
  _geom_angle_publ_flag
C2 O1 C5 111.6(2) yes
O1 C2 C3 110.9(2) yes
O1 C2 O21 122.2(3) yes
C3 C2 O21 127.0(3) yes
C2 C3 N4 101.3(2) yes
C2 C3 C31 111.3(2) yes
C2 C3 H3 107(1) no
N4 C3 C31 116.7(2) yes
```

Example 3.2.4.10. List of bonds.

```
loop_
  _geom_bond_atom_site_label_1
  _geom_bond_atom_site_label_2
  _geom_bond_distance
  _geom_bond_site_symmetry_1
  _geom_bond_site_symmetry_2
  _geom_bond_publ_flag
O1 C2 1.342(4) 1_555 1_555 yes
O1 C5 1.439(3) 1_555 1_555 yes
C2 C3 1.512(4) 1_555 1_555 yes
C2 O21 1.199(4) 1_555 1_555 yes
C3 N4 1.465(3) 1_555 1_555 yes
C3 C31 1.537(4) 1_555 1_555 yes
C3 H3 1.00(3) 1_555 1_555 no
N4 C5 1.472(3) 1_555 1_555 yes
```

$$k = 5 + x, \quad l = 5 + y, \quad m = 5 + z.$$

By adding 5 to the translations, the use of negative numbers is avoided. As an example, the symmetry code 7.645 means that the symmetry operation with label '7' in the `_symmetry_equiv_pos_site_id` list is applied and the resulting position is translated $+1.0 \times a$ along the *x* axis, $-1.0 \times b$ along the *y* axis and $0.0 \times c$ along the *z* axis, where *a*, *b* and *c* are the unit-cell edges.

List entries with a `_geom_angle_publ_flag` value of `yes` are those that should be published.

The GEOM_BOND category records intramolecular bond distances. In Example 3.2.4.10, all the atoms are untransformed and are at the positions given in the ATOM_SITE list. The symmetry code is 1_555, where the trivial symmetry operation *x*, *y*, *z* is numbered '1' by `_symmetry_equiv_pos_site_id`.

The GEOM_CONTACT category records nonbonded interatomic contacts. In Example 3.2.4.11, all the atoms are untransformed and are at the positions given in the ATOM_SITE list, and therefore the symmetry codes all have the value '.' (meaning 'inapplicable'). This is another method for indicating the identity transformation.

The GEOM_HBOND category records details about hydrogen bonds. Unlike other categories in the GEOM group, the GEOM_HBOND category records information about both distances and angles, including donor–acceptor, donor–hydrogen and acceptor–hydrogen distances and the included angle at the hydrogen-atom site (see Example 3.2.4.12). The comments above about the interpretation of symmetry codes and their relevance in the formal assignment of the category key also apply to this category.

Note that, strictly speaking, this category should only be populated if coordinates for the hydrogen atom are available (because the mandatory component of the category key `_geom_hbond_`

3. CIF DATA DEFINITION AND CLASSIFICATION

Example 3.2.4.11. *List of nonbonded interatomic contacts.*

```
loop_
  _geom_contact_atom_site_label_1
  _geom_contact_atom_site_label_2
  _geom_contact_distance
  _geom_contact_site_symmetry_1
  _geom_contact_site_symmetry_2
  _geom_contact_publ_flag
O(1) O(2) 2.735(3) . . yes
H(O1) O(2) 1.82 . . no
```

Example 3.2.4.12. *List of hydrogen-bond distances and angles.*

```
loop_
  _geom_hbond_atom_site_label_D
  _geom_hbond_atom_site_label_H
  _geom_hbond_atom_site_label_A
  _geom_hbond_distance_DH
  _geom_hbond_distance_HA
  _geom_hbond_distance_DA
  _geom_hbond_angle_DHA
  _geom_hbond_publ_flag
N6 HN6 OW 0.888(8) 1.921(12) 2.801(8) 169.6(8) yes
OW HO2 O7 0.917(6) 1.923(12) 2.793(8) 153.5(8) yes
OW HO1 N10 0.894(8) 1.886(11) 2.842(8) 179.7(9) yes
```

Example 3.2.4.13. *List of torsion angles.*

```
loop_
  _geom_torsion_atom_site_label_1
  _geom_torsion_atom_site_label_2
  _geom_torsion_atom_site_label_3
  _geom_torsion_atom_site_label_4
  _geom_torsion
  _geom_torsion_site_symmetry_1
  _geom_torsion_site_symmetry_2
  _geom_torsion_site_symmetry_3
  _geom_torsion_site_symmetry_4
  _geom_torsion_publ_flag
C(9) O(2) C(7) C(2) 71.8(2) . . . . yes
C(7) O(2) C(9) C(10) -168.0(3) . . . . 2_666 yes
C(10) O(3) C(8) C(6) -167.7(3) . . . . yes
C(8) O(3) C(10) C(9) -69.7(2) . . . . 2_666 yes
O(1) C(1) C(2) C(3) -179.5(4) . . . . no
O(1) C(1) C(2) C(7) -0.6(1) . . . . no
```

`_atom_site_label_H` needs a parent label in the atom-site list). In practice, hydrogen bonds can be assumed between donor atoms and acceptors even if the hydrogen atom is not specifically located.

The items in the GEOM_TORSION category describe the torsion angle in degrees generated for the bonded sequence of four atom sites identified by the `_geom_torsion_atom_site_label_*` codes. As with other geometry-specific site labels, these must match labels specified as `_atom_site_label` in the atom list. The torsion angle definition is that of Klyne & Prelog (1960).

Example 3.2.4.13 includes two sites that have been generated by crystallographic symmetry operations and lattice translations from the parent sites in the atom list.

3.2.4.4. Symmetry and space-group information

The categories describing symmetry are as follows:

SYMMETRY group

Original symmetry categories (§3.2.4.4.1)

SYMMETRY

SYMMETRY_EQUIV

Replacement symmetry categories (§3.2.4.4.2)

SPACE_GROUP

SPACE_GROUP_SYMOP

The SPACE_GROUP and older SYMMETRY categories contain information about the symmetry of the crystal; specifically the

space group and the symmetry-equivalent positions for that space group. More information about the symmetry is available in the symCIF dictionary described in Chapter 3.8 and presented in Chapter 4.7. The categories SPACE_GROUP and SPACE_GROUP_SYMOP were imported from symCIF in version 2.3 of the core dictionary, and are intended to replace the SYMMETRY and SYMMETRY_EQUIV categories. In most cases, there are strict equivalences between data items in the two sets. The new categories have been adopted for greater compatibility with future expansions to the symmetry CIF dictionary, and to correct some potentially misleading practices in the original categories. Although all the data items in SYMMETRY and SYMMETRY_EQUIV_POS are now formally marked as deprecated, it is likely that the older data items will remain in circulation for some time.

3.2.4.4.1. Data items in SYMMETRY and related categories

The data items in these categories are as follows:

(a) SYMMETRY

† `_symmetry_cell_setting`

† `_symmetry_Int_Tables_number`

† `_symmetry_space_group_name_H-M`

† `_symmetry_space_group_name_Hall`

(b) SYMMETRY_EQUIV

†• `_symmetry_equiv_pos_site_id`

† `_symmetry_equiv_pos_as_xyz`

The bullet (•) indicates a category key. In practice `_symmetry_equiv_pos_site_id` is often absent from older CIFs. The dagger (†) indicates a deprecated item, which should not be used in the creation of new CIFs.

The data items in the SYMMETRY category (now superseded by SPACE_GROUP) were used to record the space group. The Hermann–Mauguin (H-M) symbol was given by `_symmetry_space_group_name_H-M`. The dictionary definition recommended the use of the ‘full’ H-M symbol as listed in *International Tables for Crystallography* Volume A, but was not explicit about the meaning of ‘full’. The dictionary examples showed short-form symbols expanded to a complete representation of individual symmetry elements; thus *Pnnn* would be given as ‘P 2/n 2/n 2/n’, and the monoclinic space group *P2₁/m* would be given as ‘P 1 2₁/m 1’ for the *b*-axis unique setting or ‘P 1 1 2₁/m’ for the *c*-axis unique setting.

In practice, abbreviated symbols were often used, following conventions established over many years; thus ‘P 2₁/m’ was often given as the Hermann–Mauguin symbol when the ‘usual’ *b* setting of a monoclinic cell had been chosen. It is recommended that these conventions should continue to be followed when the new data item `_space_group_name_H-M_alt` is used instead.

The dictionary examples also suggested concise ways of indicating the origin choice within the `_symmetry_space_group_name_H-M` field; since there is no formal description of how to do this, different authors used different wording. Hence, `_symmetry_space_group_name_H-M` was always best considered as a container for the representation of the space group that would appear in a published article, and not as a machine-readable source of information about the crystallographic symmetry.

The two mechanisms for conveying the symmetry transformations in a fully machine-readable form were the Hall symbol `_symmetry_space_group_name_Hall` (Hall, 1981; Hall & Grosse-Kunstleve, 2001) and a complete listing of the symmetry operations using data items in the SYMMETRY_EQUIV category.

The data item `_symmetry_cell_setting` indicates the crystal system, not (as suggested by its name) the setting used.

Example 3.2.4.14. A list of symmetry-equivalent positions.

```
loop_
  _symmetry_equiv_pos_site_id
  _symmetry_equiv_pos_as_xyz
  1      x,y,z
  2      1/2-x,-y,1/2+z
  3      1/2+x,1/2-y,-z
  4      -x,1/2+y,1/2-z
```

The SYMMETRY_EQUIV category, now superseded by SPACE_GROUP_SYMOP, provided a list of symmetry-equivalent positions in algebraic notation. Formally, `_symmetry_equiv_pos_site_id` acted as a category key, with any arbitrary numeric value that uniquely identifies each operator. Historically, the earliest versions of the core dictionary did not have such an identifier at all and the separate equivalent positions were indexed by their position in the `_symmetry_equiv_pos_as_xyz` list. This interpretation was vulnerable to inadvertent re-ordering of the list of equivalent positions, and for this reason, as well as to satisfy the formal need for a category key, `_symmetry_equiv_pos_site_id` was added (Example 3.2.4.14). For compatibility with software that was written to handle the earlier arrangement, it is recommended that `_symmetry_equiv_pos_site_id` gives sequential integer labels, starting with 1, to the equivalent positions in the sequence in which they appear in the CIF.

Note that the `_symmetry_equiv_pos_as_xyz` list must contain *all* symmetry-equivalent positions of the space group, including those generated by lattice centring and a centre of symmetry, if present.

3.2.4.4.2. Data items in SPACE_GROUP and related categories

Data items in these categories are as follows:

- (a) SPACE_GROUP
- `_space_group_crystal_system`
 - `_space_group_id`
 - `_space_group_IT_number`
 - `_space_group_name_Hall`
 - `_space_group_name_H-M_alt`
- (b) SPACE_GROUP_SYMOP
- `_space_group_symop_id`
 - `_space_group_symop_operation_xyz`
 - `_space_group_symop_sg_id`

The bullet (•) indicates a category key.

The data items in the SPACE_GROUP category record the space group and crystal system. They recognize the common practice of supplying the space group in Hermann–Mauguin notation, though the H-M symbol does not contain complete information about the symmetry and the space-group origin. `_space_group_name_H-M_alt` allows the use of any legitimate H-M symbol as listed in *International Tables for Crystallography* Volume A or derived by similar principles. It does not give rigorous direction on how the symbols should be presented. It is recommended that the use of this symbol in CIFs containing articles for publication should follow the guidelines for `_symmetry_space_group_name_H-M` (Section 3.2.4.4.1).

Because a given space-group type may be described by more than one Hermann–Mauguin symbol, the space-group type should be specified by the use of `_space_group_IT_number`.

Two mechanisms exist for conveying fully machine-readable descriptions of the symmetry transformations relevant to the space group and setting. The first is the Hall symbol (Hall, 1981; Hall & Grosse-Kunstleve, 2001), which uniquely defines the space group

Example 3.2.4.15. A list of symmetry operators using data items from the SPACE_GROUP_SYMOP category.

```
loop_
  _space_group_symop_id
  _space_group_symop_operation_xyz
  1      x,y,z
  2      1/2-x,-y,1/2+z
  3      1/2+x,1/2-y,-z
  4      -x,1/2+y,1/2-z
```

and its reference to a particular coordinate system; it is specified in the data item `_space_group_name_Hall`. Alternatively, the symmetry operations may be listed in full using data items in the SYMMETRY_EQUIV category.

The SPACE_GROUP_SYMOP category provides a list of the symmetry operators for a space group in algebraic notation. It replaces the category SYMMETRY_EQUIV_POS. Unlike the older category, where in practice the category key could be omitted from listings (and must therefore be generated implicitly by parsing software), the category key `_space_group_symop_id` *must* be given. See Example 3.2.4.15, which may be compared with Example 3.2.4.14.

3.2.4.5. Bond-valence information

Categories describing bond valences are as follows:

```
VALENCE group
  VALENCE_PARAM
  VALENCE_REF
```

Data items in these categories are as follows:

- (a) VALENCE_PARAM
- `_valence_param_atom_1`
 - `_valence_param_atom_1_valence`
 - `_valence_param_atom_2`
 - `_valence_param_atom_2_valence`
 - `_valence_param_B`
 - `_valence_param_details`
 - `_valence_param_id`
 - `_valence_param_ref_id`
 - `_valence_ref_id`
 - `_valence_param_Ro`
- (b) VALENCE_REF
- `_valence_ref_id`
 - `_valence_ref_reference`

The arrow (→) is a reference to a parent data item.

The data items in this category group relate to bond valences, which are widely used in inorganic crystallography to confirm and analyse the results of crystal structure determinations. Bond valences are determined from the bond lengths and have the useful property that their sum around any atom is equal to the atom valence (formal charge). They are increasingly being published with bond lengths. The data item `_geom_bond_valence` in the GEOM_BOND category allows the bond valence to be associated with the bond length.

The two categories discussed here list the parameters used to calculate the bond valences and their literature sources. These items might also be published, particularly where there is some uncertainty about the appropriate parameters to use.

The data items in the VALENCE_PARAM category define the parameters used for calculating bond valences from bond lengths. In addition to the parameters, a pointer to the reference for the source of the parameters (in VALENCE_REF) is given (Example 3.2.4.16).

Example 3.2.4.16. A list of bond-valence parameters.

```
loop_
  _valence_param_atom_1
  _valence_param_atom_1_valence
  _valence_param_atom_2
  _valence_param_atom_2_valence
  _valence_param_Ro
  _valence_param_B
  _valence_param_ref_id
  _valence_param_details
  Cu 2 O -2 1.679 0.37 a .
  Cu 2 O -2 1.649 0.37 j .
  Cu 2 N -3 1.64 0.37 m '2-coordinate N'
  Cu 2 N -3 1.76 0.37 m '3-coordinate N'
loop_
  _valence_ref_id
  _valence_ref_reference
a
'Brown & Altermatt (1985), Acta Cryst. B41, 244-247'
j
'Liu & Thorp (1993), Inorg. Chem. 32, 4102-4205'
m
; See, Krause & Strub (1998), Inorg. Chem.
  37, 5369-5375'
;
```

3.2.5. Publication

As an archival file format, CIF is well suited to the complete documentation of a structural study and the categories described in this section provide data items suitable for the generation of a fully documented report, either as an informal laboratory notebook document or as a formal published article.

3.2.5.1. Literature citations

The categories describing literature citations are as follows:

```
CITATION group
  CITATION
  CITATION_AUTHOR
  CITATION_EDITOR
```

The entries in the CITATION category group provide a set of data items suitable for the structured recording of references to the literature. At present, they are designed for the storage and retrieval of information about journal articles and individual chapters in books. They do not currently cover conference proceedings, pamphlets, preprints, theses or other kinds of publication. Reference lists are usually requested by journals that accept articles in CIF format as a single text field in `_publ_section_references`, but the categories in the CITATION group may become more useful for storing citation lists in the future, especially if converters become available to and from other bibliographic formats such as EndNote and BibTeX.

Data items in these categories are as follows:

(a) CITATION

```
• _citation_id
  _citation_abstract
  _citation_abstract_id_CAS
  _citation_book_id_ISBN
  _citation_book_publisher
  _citation_book_publisher_city
  _citation_book_title
  _citation_coordinate_linkage
  _citation_country
  _citation_database_id_CSD
  _citation_database_id_Medline
  _citation_journal_abbrev
  _citation_journal_full
  _citation_journal_id_ASTM
  _citation_journal_id_CSD
  _citation_journal_id_ISSN
  _citation_journal_issue
```

```
_citation_journal_volume
_citation_language
_citation_page_first
_citation_page_last
_citation_special_details
_citation_title
_citation_year
```

(b) CITATION_AUTHOR

```
• _citation_author_citation_id
  → _citation_id
  _citation_author_name
  _citation_author_ordinal
```

(c) CITATION_EDITOR

```
• _citation_editor_citation_id
  → _citation_id
  _citation_editor_name
  _citation_editor_ordinal
```

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

The CITATION category provides the bulk of the information about individual citations. `_citation_id` provides a link to the CITATION_AUTHOR and CITATION_EDITOR categories, where multiple authors, and, if appropriate, multiple editors are listed.

Example 3.2.5.1 shows how a fully populated citation list is structured across these categories.

The authors of a cited reference are listed using items from the CITATION_AUTHOR category. The value of `_citation_author_citation_id` must match a value of `_citation_id` in the CITATION category, and this data item forms the link between the authors and the citations. `_citation_author_ordinal` is used to record the order in which the authors are listed.

The editors of a cited reference are listed using items from the CITATION_EDITOR category. The value of `_citation_editor_citation_id` must match a value of `_citation_id` in the CITATION category, and this data item forms the link between the editors and the citations. `_citation_editor_ordinal` is used to record the order in which the editors are listed.

Example 3.2.5.1. A structured bibliographic reference list.

```
loop_
  _citation_id
  _citation_title
  _citation_page_first
  _citation_page_last
  _citation_year
  _citation_journal_abbrev
  _citation_journal_volume
  _citation_journal_id_ISSN
  1
; Angle calculations for 3- and 4-circle X-ray
  and neutron diffractometers
;
  457 464 1967 'Acta Cryst.' 22 0365-110X
  2
'Space-group notation with an explicit origin'
  517 525 1981 'Acta Cryst. Section A' 37
  0108-7673
  3 ? 521 523 1960 'Experientia' 16 ?
loop_
  _citation_author_citation_id
  _citation_author_name
  1 'Busing, W. R.'
  1 'Levy, H. A.'
  2 'Hall, S. R.'
  3 'Klyne, W.'
  3 'Prelog, V.'
```