```
_exptl_crystal_description
_exptl_crystal_F_000
_exptl_crystal_preparation
_exptl_crystal_pressure_history
_exptl_crystal_recrystallization_method
_exptl_crystal_size_length
_exptl_crystal_size_max
_exptl_crystal_size_mid
_exptl_crystal_size_min
_exptl_crystal_size_rad
_exptl_crystal_thermal_history
```

(*c*) EXPTL_CRYSTAL_FACE

```
• _exptl_crystal_face_index_h
• _exptl_crystal_face_index_k
• _exptl_crystal_face_index_l
  _exptl_crystal_face_diffr_chi
  _exptl_crystal_face_diffr_kappa
  _exptl_crystal_face_diffr_phi
  _exptl_crystal_face_diffr_psi
  _exptl_crystal_face_perp_dist
```

*The bullet* (•) *indicates a category key. Where multiple items within a category are marked by a bullet, they must be taken together to form a compound key.*

The EXPTL category is rather broadly named, but in practice is used to record details about any absorption correction applied and, using `_exptl_special_details`, any other details of the experimental work prior to intensity measurement not specifically described by other data items (*e.g.* `_exptl_crystal_preparation`).

The data items in the EXPTL_CRYSTAL category are designed to record details of experimental measurements on the crystal or crystals used. Since it is usually the case that just one crystal is used throughout the experiment, the category is presented as if it comprises non-looped data names. However, details of a number of crystals may be looped together, in which case `_exptl_crystal_id` is used to identify the different crystals and acts as the category key.

When different crystals are used to collect diffraction intensities, it is likely that the intensities collected from each crystal would need to be scaled by different factors, as recorded by the DIFFRN_SCALE_GROUP category and the `_diffrn_refln_scale_group_code` used for each individual reflection. In these circumstances, it would be good practice to use matching values of `_diffrn_refln_scale_group_code` and `_exptl_crystal_id`, although this is not mandatory.

Note that the $F(000)$ value, which is often calculated as the integer number of electrons in the crystal unit cell, may contain dispersion contributions and is more properly calculated as

$$F(000) = \left[\left(\sum f_r\right)^2 + \left(\sum f_i\right)^2\right]^{1/2},$$

where $f_r$ and $f_i$ are, respectively, the real and imaginary parts of the scattering factors at $\theta = 0$ and the sum is taken over each atom in the unit cell.

The crystal colour may be given as free text using the data item `_exptl_crystal_colour`. Alternatively, the standardized names developed by the International Centre for Diffraction Data to classify specimen colours may be constructed from the items `_exptl_crystal_colour_lustre`, `*_modifier` and `*_primary`, each of which has a restricted set of specific values.

The EXPTL_CRYSTAL_FACE category records details of the crystal faces. The faces are defined by Miller indices and their perpendicular distances from the centre of rotation of the crystal may be recorded in millimetres. Absolute orientations with respect to the goniometer angle settings may also be recorded. The category is currently constructed in a way that cannot distinguish between multiple crystals.

## 3.2.3. Analysis

The categories relevant to the structural analysis are as follows:

*Refinement techniques and results* (§3.2.3.1)

REFINE group

- REFINE
- REFINE_LS_CLASS

*The reflections used in the refinement* (§3.2.3.2)

REFLN group

- REFLN
- REFLNS
- REFLNS_CLASS
- REFLNS_SCALE
- REFLNS_SHELL

In the small-molecule and inorganic studies for which the core dictionary was designed, phasing and structure solution are almost routine, and the dictionary provides few specific fields for recording the details of the structure solution process: `_atom_sites_solution_primary`, `_atom_sites_solution_secondary` and `_atom_sites_solution_hydrogens` (Section 3.2.4.1.2); `_computing_structure_solution` (Section 3.2.5.2); and `_publ_section_exptl_solution` (Section 3.2.5.5). (In contrast, the macromolecular CIF includes extensive details of phasing.) Refinement, however, still allows for a wide range of techniques, practices and interpretation, and there are a large number of data names to allow a full account of the refinement strategy to be given. To complement this, several categories exist to provide a detailed listing and annotation of the structure factors and their treatment according to shells of resolution or other sorting criteria.

### 3.2.3.1. Structure refinement

The data items in these categories are as follows:

(*a*) REFINE

```
  _refine_diff_density_max
  _refine_diff_density_min
  _refine_diff_density_rms
  _refine_ls_abs_structure_details
  _refine_ls_abs_structure_Flack
  _refine_ls_abs_structure_Rogers
  _refine_ls_d_res_high
  _refine_ls_d_res_low
  _refine_ls_extinction_coef
  _refine_ls_extinction_expression
  _refine_ls_extinction_method
  _refine_ls_goodness_of_fit_all
  _refine_ls_goodness_of_fit_gt
† _refine_ls_goodness_of_fit_obs
  _refine_ls_goodness_of_fit_ref
  _refine_ls_hydrogen_treatment
  _refine_ls_matrix_type
  _refine_ls_number_constraints
  _refine_ls_number_parameters
  _refine_ls_number_reflns
  _refine_ls_number_restraints
  _refine_ls_R_factor_all
  _refine_ls_R_factor_gt
† _refine_ls_R_factor_obs
  _refine_ls_R_Fsqd_factor
  _refine_ls_R_I_factor
  _refine_ls_restrained_S_all
  _refine_ls_restrained_S_gt
† _refine_ls_restrained_S_obs
† _refine_ls_shift/esd_max
† _refine_ls_shift/esd_mean
  _refine_ls_shift/su_max
  _refine_ls_shift/su_max_lt
  _refine_ls_shift/su_mean
  _refine_ls_shift/su_mean_lt
  _refine_ls_structure_factor_coef
  _refine_ls_weighting_details
```

```
  _refine_ls_weighting_scheme
  _refine_ls_wR_factor_all
  _refine_ls_wR_factor_gt
† _refine_ls_wR_factor_obs
  _refine_ls_wR_factor_ref
  _refine_special_details
```

(*b*) REFINE_LS_CLASS

```
• _refine_ls_class_code
        → _reflns_class_code
  _refine_ls_class_d_res_high
  _refine_ls_class_d_res_low
  _refine_ls_class_R_factor_all
  _refine_ls_class_R_factor_gt
  _refine_ls_class_R_Fsqd_factor
  _refine_ls_class_R_I_factor
  _refine_ls_class_wR_factor_all
```

*The bullet (•) indicates a category 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.*

Example 3.2.3.1 shows how the data names in the REFINE category are used. Most of the dictionary entries are detailed and fully explanatory, so only a few points that might require special care are mentioned here.

Two groups of older data names have been superseded by new names that are functionally equivalent, but represent a more correct terminology. One group is of names that include the component '**_obs**' used to indicate 'observed' reflections; this has been replaced by the component '**_gt**' indicating that the measured values are greater than a threshold recorded elsewhere (as the value of **_reflns_threshold_expression**). The other group replaces the component '**_esd**' (for estimated standard deviation) with '**_su**' (for standard uncertainty).

A number of data names describe the extinction coefficient and the method used to determine it. Note that a default value (Zachariasen) is given in the dictionary for the method (**_refine_ls_extinction_method**); this *only* makes sense if this data item is missing from the data block but a value of **_refine_ls_extinction_coef** is present. This can complicate the design of software to read CIFs, which might assign to any missing data name a default value given by the dictionary.

Care is also needed with **_refine_ls_hydrogen_treatment**, which describes the treatment of hydrogen atoms in the refinement. Clearly, the data item only has meaning if there were hydrogen atoms in the model (although, since in this case the default value is undef for 'undefined', it could be argued that the default is appropriate even when hydrogen atoms were not included in the model).

The weighting scheme used in the refinement is described by the two data names **_refine_ls_weighting_scheme** and **_refine_ls_weighting_details**. The first of the two can take only one of the three values sigma (weights assigned based on measured standard uncertainties), unit (unit or no weights applied) or calc (calculated weights applied). The actual mathematical expression used in the weighting scheme should be stated in **_refine_ls_weighting_details**.

A wide variety of 'residual structure-factor difference measures', referred to as $R$ factors, are used in crystallography as indicators of refinement quality. The core CIF dictionary contains definitions for the three most commonly used $R$ factors. The 'conventional $R$ factor' is defined as

$$R = \frac{\sum |F(\text{meas.}) - F(\text{calc.})|}{\sum |F(\text{meas.})|},$$

Example 3.2.3.1. *Summary of refinement results.*

```
_refine_special_details
                sfls:_F_calc_weight_full_matrix

_refine_ls_structure_factor_coef  F
_refine_ls_matrix_type            full
_refine_ls_weighting_scheme       calc
_refine_ls_weighting_details
                  'w=1/(u^2^(F)+0.0004F^2^)'
_refine_ls_hydrogen_treatment     refxyz
_refine_ls_extinction_method      Zachariasen
_refine_ls_extinction_coef        3514(42)
_refine_ls_extinction_expression
;Larson, A. C. (1970). "Crystallographic Computing",
 edited by F. R. Ahmed. Eq. (22) p. 292. Copenhagen:
 Munksgaard.
;
_refine_ls_abs_structure_details
;The absolute configuration was assigned to agree
 with that of its precursor l-leucine at the
 chiral centre C3.
;
_refine_ls_number_reflns          1408
_refine_ls_number_parameters      272
_refine_ls_number_restraints      0
_refine_ls_number_constraints     0
_refine_ls_R_factor_all           .038
_refine_ls_R_factor_gt            .034
_refine_ls_wR_factor_all          .044
_refine_ls_wR_factor_gt           .042
_refine_ls_goodness_of_fit_all   1.462
_refine_ls_goodness_of_fit_gt    1.515
_refine_ls_shift/su_max           .535
_refine_ls_shift/su_mean          .044
_refine_diff_density_min         -.108
_refine_diff_density_max          .131
```

where $F(\text{meas.})$ and $F(\text{calc.})$ are the measured and calculated structure factors, respectively. In the data item **_refine_ls_R_factor_all**, the sum used in the calculation is taken over all the reflections collected, whereas in the data item **_refine_ls_R_factor_gt**, the sum is taken over reflections with a value greater than the limit specified by **_refine_threshold_expression**. In both cases, the reflections included in the calculation may be limited to those between specified resolution limits.

This $R$ factor is calculated from the $F$ values, regardless of whether the structure-factor coefficient $|F|$, $|F|^2$ or $I$ was actually used in the refinement, and is often taken as a convenient indicator of the relative quality of a structure determination. As most structure refinements used to be performed on $|F|$, it allows a structure determined today to be compared with an older study.

Many refinements are now carried out on $|F|^2$, although some may still use the absolute value of the structure factor $|F|$ or the net intensity $I$. The weighted residual factor $wR$ and goodness of fit $S$ for a refinement should be reported according to the coefficients actually used in the refinement. For example, the weighted residual over all reflections, **_refine_ls_wR_factor_all**, is defined as

$$wR = \left( \frac{\sum w[Y(\text{meas.}) - Y(\text{calc.})]^2}{\sum wY(\text{meas.})^2} \right)^{1/2},$$

where $w$ represents the weights and $Y$ represents the structure-factor coefficient, either $|F|$, $|F|^2$ or $I$ as specified by **_refine_ls_structure_factor_coef**.

This distinction between the conventional $R$ factor, which is invariably calculated using $F$ values, and the $wR$ and $S$ factors also holds for similar expressions defined on subsets of the reflections, *e.g.* **_reflns_class_wR_factor_all**.

Note that data names are also provided for reporting *unweighted* residuals on $|F|^2$ or $I$, but these are rarely used in practice, with

Example 3.2.3.2. *Structure-factor listing.*

```
loop_
     _refln_index_h
     _refln_index_k
     _refln_index_l
     _refln_F_squared_calc
     _refln_F_squared_meas
     _refln_F_squared_sigma
     _refln_include_status
    2   0   0       85.57       58.90      1.45 o
    3   0   0    15718.18    15631.06     30.40 o
    4   0   0    55613.11    49840.09     61.86 o
    5   0   0      246.85      241.86     10.02 o
    6   0   0       82.16       69.97      1.93 o
    7   0   0     1133.62      947.79     11.78 o
    8   0   0     2558.04     2453.33     20.44 o
    9   0   0      283.88      393.66      7.79 o
    10  0   0      283.70      171.98      4.26 o
```

the exception of $R(I)$ in Rietveld refinements against powder data, where it is generally called the Bragg $R$ factor, $R_{\rm Bragg}$ or $R_B$.

The data items in the REFINE_LS_CLASS category are similar to several in the general REFINE category, but correspond to values for separate reflection classes as described in the REFLNS_CLASS category. The data name **_refine_ls_class_code** identifies the individual classes through a direct match with a corresponding value of **_reflns_class_code**.

### 3.2.3.2. Reflection measurements

The categories describing the reflections used in the refinement are as follows:

REFLN group

*Individual reflections* (§3.2.3.2.1)

REFLN

*Groups of reflections* (§3.2.3.2.2)

REFLNS

REFLNS_CLASS

REFLNS_SCALE

REFLNS_SHELL

The main category in this group is REFLN, which stores the list of reflections used in the structure refinement process, their associated structure factors and information about how each reflection was handled. The distinction between the REFLN (singular) category and the REFLNS (plural) category parallels the distinction between the categories DIFFRN_REFLN and DIFFRN_REFLNS: data items in the REFLN category store information about individual reflections, while data items in the REFLNS category store information about the complete set of reflections, or about subsets of reflections selected by shells of resolution, scaling factors or other criteria.

#### 3.2.3.2.1. *Individual reflections*

The data items in this category are as follows:

REFLN

- **_refln_index_h**
- **_refln_index_k**
- **_refln_index_l**

**_refln_A_calc**
**_refln_A_meas**
**_refln_B_calc**
**_refln_B_meas**
**_refln_class_code**
→ **_reflns_class_code**
**_refln_crystal_id**
→ **_exptl_crystal_id**
**_refln_d_spacing**
**_refln_F_calc**
**_refln_F_meas**
**_refln_F_sigma**
**_refln_F_squared_calc**
**_refln_F_squared_meas**
**_refln_F_squared_sigma**
**_refln_include_status**
**_refln_intensity_calc**
**_refln_intensity_meas**
**_refln_intensity_sigma**
**_refln_mean_path_length_tbar**
† **_refln_observed_status**
**_refln_phase_calc**
**_refln_phase_meas**
**_refln_refinement_status**
**_refln_scale_group_code**
→ **_reflns_scale_group_code**
**_refln_sint/lambda**
**_refln_symmetry_epsilon**
**_refln_symmetry_multiplicity**
**_refln_wavelength**
**_refln_wavelength_id**
→ **_diffrn_radiation_wavelength_id**

*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 dagger (†) indicates a deprecated item, which should not be used in the creation of new CIFs.*

Example 3.2.3.2 shows a typical structure-factor listing produced by a refinement program. This kind of structure-factor listing is suitable for deposition with a journal or a database. The Miller indices for each reflection are accompanied by the calculated and measured values of the quantity used in the refinement, and the standard uncertainty derived from the measurement. There is also an indication of whether each reflection was included in the refinement and in the calculation of $R$ factors.

In this example, the squared structure factors $|F|^2$ are listed. When refinement is performed against the structure factors $F$ or the intensities $I$, the data items **_refln_F_calc** or **_refln_intensity_calc** and the corresponding data names for the measured values and uncertainties should be used.

Individual calculated structure-factor components $A = |F|\cos\varphi$ and $B = |F|\sin\varphi$ may also be listed, along with the phase $\varphi$, using the data names **_refln_A_calc**, **_refln_B_calc** and **_refln_phase_calc**. Corresponding measured values have equivalent ***_meas** names.

The **_refln_include_status** flag is used to indicate whether reflections were used in the refinement and in the calculation of $R$ factors, and if they were not used, to give the reason for exclusion of the reflection from the refinement. The flag o, which indicates that a reflection was used in the refinement, was originally chosen to indicate that the value of the reflection was higher than the limit specified by **_reflns_observed_criterion** and that the reflection was thus 'observed'. The data item **_reflns_observed_criterion** is now deprecated in favour of **_reflns_threshold_status**, and the value o is now taken to indicate not only that the reflection has an intensity suitable for inclusion in the refinement, but also that the reflection satisfies all other criteria used to select reflections for inclusion in the refinement.

Various other flags indicate reflections that were not included in the refinement. Reflections outside the range of $d$ spacings bounded by the values **_refine_ls_d_res_high** and **_refine_ls_d_res_low** are flagged with h or l, respectively. Reflections within the resolution limits but below the intensity threshold are flagged with <. Systematically absent reflections are flagged with -. Sometimes a value can be identified as having a systematic error; these reflections can be flagged with x. However, great care must be taken in excluding reflections with apparently 'anomalous' structure factors (*i.e.* where the measured values are substantially different from the calculated ones), so as not to introduce bias into the refinement.

The flag **_refln_refinement_status** is used specifically to indicate whether a reflection was included in or excluded from

the refinement. Use of **_refln_include_status** to provide more information about each reflection is greatly preferred.

Other data names in this category allow the recording of specific information about each reflection, such as the symmetry reinforcement factor $\varepsilon$, the number of reflections symmetry-equivalent under the Laue symmetry, the $d$ spacing, the mean path length through the crystal $\bar{t}$, the $(\sin\theta)/\lambda$ value and, in the case of Laue experiments, the mean wavelength of the radiation. (For polychromatic radiation, the wavelength information might instead be given by **_refln_wavelength_id**, which is a code identifying a matching entry in the DIFFRN_RADIATION category.)

Other codes provide links to identifiers in other categories. The **_refln_class_code** identifies a set of reflections binned as described by entries in the REFLNS_CLASS category. **_refln_scale_group_code** identifies groups of reflections to which the same structure-factor scaling has been applied.

Note that the values of the Miller indices in this list must correspond to the cell defined by the lengths and angles recorded in the CELL category; they may, however, be different from the Miller indices in the DIFFRN_REFLN list if a transformation of the original cell has taken place. In this case, the transformation matrix is given using the **_diffrn_reflns_transf_matrix_*** items.

The usual use of a CIF as an archive of a completed structure determination implies that the values given in the REFLN list are derived from the final cycle of refinement, but this is not a formal requirement. Care should be taken when preparing a CIF for archiving that the structural model corresponds to the refinement cycle summarized in the accompanying REFLN table, especially if the file is constructed from fragments output from different programs.

3.2.3.2.2. *Groups of reflections*

The data items in these categories are as follows:

(*a*) REFLNS

```
  _reflns_d_resolution_high
  _reflns_d_resolution_low
  _reflns_Friedel_coverage
  _reflns_limit_h_max
  _reflns_limit_h_min
  _reflns_limit_k_max
  _reflns_limit_k_min
  _reflns_limit_l_max
  _reflns_limit_l_min
  _reflns_number_gt
† _reflns_number_observed
  _reflns_number_total
† _reflns_observed_criterion
  _reflns_special_details
  _reflns_threshold_expression
```

(*b*) REFLNS_CLASS

```
• _reflns_class_code
  _reflns_class_d_res_high
  _reflns_class_d_res_low
  _reflns_class_description
  _reflns_class_number_gt
  _reflns_class_number_total
  _reflns_class_R_factor_all
  _reflns_class_R_factor_gt
  _reflns_class_R_Fsqd_factor
  _reflns_class_R_I_factor
  _reflns_class_wR_factor_all
```

(*c*) REFLNS_SCALE

```
• _reflns_scale_group_code
  _reflns_scale_meas_F
  _reflns_scale_meas_F_squared
  _reflns_scale_meas_intensity
```

(*d*) REFLNS_SHELL

```
• _reflns_shell_d_res_high
• _reflns_shell_d_res_low
† _reflns_shell_meanI_over_sigI_all
† _reflns_shell_meanI_over_sigI_gt
† _reflns_shell_meanI_over_sigI_obs
  _reflns_shell_meanI_over_uI_all
  _reflns_shell_meanI_over_uI_gt
  _reflns_shell_number_measured_all
  _reflns_shell_number_measured_gt
† _reflns_shell_number_measured_obs
  _reflns_shell_number_possible
  _reflns_shell_number_unique_all
  _reflns_shell_number_unique_gt
† _reflns_shell_number_unique_obs
  _reflns_shell_percent_possible_all
  _reflns_shell_percent_possible_gt
† _reflns_shell_percent_possible_obs
  _reflns_shell_Rmerge_F_all
  _reflns_shell_Rmerge_F_gt
† _reflns_shell_Rmerge_F_obs
  _reflns_shell_Rmerge_I_all
  _reflns_shell_Rmerge_I_gt
† _reflns_shell_Rmerge_I_obs
```

*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 dagger (†) indicates a deprecated item, which should not be used in the creation of new CIFs.*

The data items in the REFLNS category describe properties or attributes of the complete set of reflections used in the structure refinement. Several are derivative and may be obtained from the information in the reflections list, but it is convenient to present them separately so that they do not need to be calculated again. They can also be used to check the consistency of the reflections list.

The **_reflns_limit_*** data items define the upper and lower bounds on the Miller indices and on the interplanar $d$ spacings.

The **_reflns_threshold_expression** is a text field describing the criterion applied to mark individual reflections as 'significantly intense' (*i.e.* distinct from the background level). This is typically expressed as a multiple of the standard uncertainty on the quantity used in refinement, *e.g.* `I>2u(I)`.

The number of reflections with values higher than the threshold is reported in **_reflns_number_gt**. The total number of reflections measured is given by **_reflns_number_total**. Although the use of these data names appears to be obvious, different practices have been used in the past to report total numbers (*e.g.* by neglecting symmetry-equivalent reflections) and the definitions in the dictionary should be consulted. Both numbers may contain Friedel-equivalent reflections (those which are symmetry-equivalent under the Laue symmetry but inequivalent under the crystal class).

The proportion of Friedel-related reflections present is reported separately by **_reflns_Friedel_coverage**, defined as $(N_C - N_L)/N_L$, where $N_C$ is the number of reflections obtained on averaging under the symmetry of the crystal class and $N_L$ is the number obtained on averaging under the Laue class. The definition in the dictionary provides examples of how the value of this data name may be used as an indicator of the fraction of the available reciprocal space sampled in the diffraction experiment.

The deprecated data names **_reflns_observed_criterion** and **_reflns_number_observed** reflect the old use of 'observed' as a term describing significantly intense reflections. They should not be used in the creation of new CIFs, but are retained to ensure that the information can be extracted from old CIFs.

The free-text field **_reflns_special_details** can be used to discuss any aspects of the reflections list not covered by other data names. It is recommended that information about the averaging of symmetry-equivalent reflections (including Friedel pairs) should be given here.

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 REFINE 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.

$R_{\rm merge}$ 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
```