3. CIF DATA DEFINITION AND CLASSIFICATION

Example 3.6.6.11. *An example of one cycle of refinement described with data items in the* REFINE_HIST *category.*

```
_refine_hist.cycle_id              C134
_refine_hist.d_res_high            1.85
_refine_hist.d_res_low             20.0
_refine_hist.number_atoms_solvent   217
_refine_hist.number_atoms_total     808
_refine_hist.number_reflns_all     6174
_refine_hist.number_reflns_obs     4886
_refine_hist.number_reflns_R_free   476
_refine_hist.number_reflns_R_work  4410
_refine_hist.R_factor_all          .265
_refine_hist.R_factor_obs          .195
_refine_hist.R_factor_R_free       .274
_refine_hist.R_factor_R_work       .160
_refine_hist.details
; Add majority of solvent molecules. B factors
  refined by group. Continued to remove
  misplaced water molecules.
;
```

Data items in the REFINE_B_ISO category can be used to record details of the treatment of isotropic B (displacement) factors during refinement. There is no formal link between the classes identified by **_refine_B_iso.class** and individual atom sites, although relationships may be inferred if the class names are carefully chosen. The category allows the treatment of the atoms in each class (isotropic, anisotropic or fixed) and the value assigned for fixed isotropic B factors to be recorded. Any special details can be given in a free-text field.

Data items in the REFINE_OCCUPANCY category can be used to record details of the treatment of occupancies of groups of atom sites during refinement. As with the treatment of displacement factors in the REFINE_B_ISO category, the classes itemized by **_refine_occupancy.class** are not formally linked to the individual atom sites, but the relationships may be deduced if the class names are chosen carefully.

3.6.6.2.5. *History of the refinement*

The data items in this category are as follows:

REFINE_HIST
● _refine_hist.cycle_id
  _refine_hist.details
  _refine_hist.d_res_high
  _refine_hist.d_res_low
  _refine_hist.number_atoms_solvent
  _refine_hist.number_atoms_total
  _refine_hist.number_reflns_all
  _refine_hist.number_reflns_obs
  _refine_hist.number_reflns_R_free
  _refine_hist.number_reflns_R_work
  _refine_hist.R_factor_all
  _refine_hist.R_factor_obs
  _refine_hist.R_factor_R_free
  _refine_hist.R_factor_R_work

The bullet (●) indicates a category key.

Data items in the REFINE_HIST category can be used to record details about the various steps in the refinement of the structure. They do not provide as thorough a description of the refinement as can be given in other categories for the final model, but instead allow a summary of the progress of the refinement to be given and supported by a small set of representative statistics.

The category is sufficiently compact that a large number of cycles could be summarized, but it is not expected that every cycle of refinement would be routinely reported. Example 3.6.6.11 shows an entry for a single cycle of refinement. It is likely that an author would present a representative sequence of entries in a looped list.

3.6.6.3. Reflection measurements

The categories describing the reflections used in the refinement are as follows:
  REFLN group
  Individual reflections (§3.6.6.3.1)
    REFLN
    REFLN_SYS_ABS
  Groups of reflections (§3.6.6.3.2)
    REFLNS
    REFLNS_SCALE
    REFLNS_SHELL
    REFLNS_CLASS

Data items in the REFLN category can be used to give information about the individual reflections that were used to derive the final model. The related category REFLN_SYS_ABS allows the reflections that should be systematically absent for the space group in which the structure was refined to be tabulated. Data items in the REFLNS category can be used to record information that applies to all of the reflections. Scale factors can be listed in the REFLNS_SCALE category, while the data items in REFLNS_SHELL can be used to record information about the reflection set by shells of resolution. The core CIF dictionary category REFLNS_CLASS, which can be used for a general classification of reflection groups according to criteria other than resolution shell, is not expected to be used in mmCIF applications.

3.6.6.3.1. *Individual reflections*

The data items in these categories are as follows:

(*a*) REFLN
● _refln.index_h
● _refln.index_k
● _refln.index_l
  _refln.A_calc
  _refln.A_calc_au
  _refln.A_meas
  _refln.A_meas_au
  _refln.B_calc
  _refln.B_calc_au
  _refln.B_meas
  _refln.B_meas_au
  _refln.class_code
  _refln.crystal_id
        → _exptl_crystal.id
  _refln.d_spacing
  _refln.F_calc
  _refln.F_calc_au
  _refln.F_meas
  _refln.F_meas_au
  _refln.F_meas_sigma (∼ _refln_F_sigma)
  _refln.F_meas_sigma_au
  _refln.F_squared_calc
  _refln.F_squared_meas
  _refln.F_squared_sigma
  _refln.fom
  _refln.include_status
  _refln.intensity_calc
  _refln.intensity_meas
  _refln.intensity_sigma
  _refln.mean_path_length_tbar
  _refln.phase_calc
  _refln.phase_meas
  _refln.refinement_status
  _refln.scale_group_code
        → _reflns_scale.group_code
  _refln.sint_over_lambda (∼ _refln_sint/lambda)
  _refln.status (∼ _refln_observed_status)
  _refln.symmetry_epsilon
  _refln.symmetry_multiplicity
  _refln.wavelength

Example 3.6.6.12. *Part of the reflection list for an HIV-1 protease structure (PDB 5HVP) described with data items in the* REFLN *category.*

```
loop_
_refln.index_h
_refln.index_k
_refln.index_l
_refln.F_squared_calc
_refln.F_squared_meas
_refln.F_squared_sigma
_refln.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
```

```
  _refln.wavelength_id
        →  _diffrn_radiation.wavelength_id
```

(*b*) REFLN_SYS_ABS
- `_refln_sys_abs.index_h`
- `_refln_sys_abs.index_k`
- `_refln_sys_abs.index_l`
  `_refln_sys_abs.I`
  `_refln_sys_abs.I_over_sigmaI`
  `_refln_sys_abs.sigmaI`

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. Items in italics have aliases in the core CIF dictionary formed by changing the full stop (.) to an underscore (_) except where indicated by the ∼ symbol.

Data items in the REFLN category are used in the same way in the mmCIF and core CIF dictionaries, and Section 3.2.3.2.1 can be consulted for details. However, in macromolecular crystallography it is not usual for reflection intensities to be given in units of electrons (the units specified by the core CIF dictionary). Thus it was necessary to introduce in the mmCIF dictionary data items for the magnitudes of structure factors and their *A* and *B* components in arbitrary units (Example 3.6.6.12). A figure of merit (`_refln.fom`) can also be included for reflections that were phased using experimental methods.

The REFLN_SYS_ABS category allows the intensities of the reflections that should be systematically absent to be tabulated. The ratio of the intensity to its standard uncertainty, given in the data item `_refln_sys_abs.I_over_sigmaI`, can be used to assess whether the reflection is indeed absent. The decision as to whether it is absent is left to the user of the mmCIF and is not recorded in the mmCIF.

3.6.6.3.2. *Groups of reflections*

The data items in these categories are as follows:

(*a*) REFLNS
- `_reflns.entry_id`
     → `_entry.id`
  `_reflns.B_iso_Wilson_estimate`
  `_reflns.data_reduction_details`
  `_reflns.data_reduction_method`
  `_reflns.d_resolution_high`
  `_reflns.d_resolution_low`
  `_reflns.details` (∼ `_reflns_special_details`)
  `_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_all` (∼ `_reflns_number_total`)
`_reflns.number_gt`
`_reflns.number_obs` (∼ `_reflns_number_observed`)
`_reflns.observed_criterion`
`_reflns.observed_criterion_F_max`
`_reflns.observed_criterion_F_min`
`_reflns.observed_criterion_I_max`
`_reflns.observed_criterion_I_min`
`_reflns.observed_criterion_sigma_F`
`_reflns.observed_criterion_sigma_I`
`_reflns.percent_possible_obs`
`_reflns.R_free_details`
`_reflns.Rmerge_F_all`
`_reflns.Rmerge_F_obs`
`_reflns.threshold_expression`

(*b*) REFLNS_SCALE
- `_reflns_scale.group_code`
  `_reflns_scale.meas_F`
  `_reflns_scale.meas_F_squared`
  `_reflns_scale.meas_intensity`

(*c*) 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`

(*d*) 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`

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. Items in italics have aliases in the core CIF dictionary formed by changing the full stop (.) to an underscore (_) except where indicated by the ∼ symbol.

Data items in the REFLNS category of the core CIF dictionary can be used to summarize the properties or attributes of the complete set of reflections used in refinement (Section 3.2.3.2.2). The mmCIF dictionary adds a number of data items to this category, including the formal category key required by the DDL2 data model. There are also data items for describing the data-reduction method and recording any relevant details about data reduction, and for giving an estimate of the overall Wilson *B* factor for the data set.

A number of the new data items relate to the issue of how reflections are flagged as being observed and are thus used in the refinement. In the core CIF dictionary, the criteria used to consider

references