angles associated with each phasing set, the wavelength of the radiation used in the experiment, the source of the radiation, the detector type, and the ambient temperature.

Data items in the PHASING_SET_REFLN category are used to record the values of the measured structure factors and their uncertainties. Several distinct data sets may be present in this list, with reflections in each set identified by the appropriate value of **_phasing_set_refln.set_id**.

3.6.6.2. Refinement

The categories describing refinement are as follows:

REFINE group

Overall description of the refinement (§3.6.6.2.1)

REFINE

REFINE_FUNCT_MINIMIZED

Analysis of the refined structure (§3.6.6.2.2)

REFINE_ANALYZE

Restraints and refinement by shells of resolution (§3.6.6.2.3)

REFINE_LS_RESTR

REFINE_LS_RESTR_NCS

REFINE_LS_RESTR_TYPE

REFINE_LS_SHELL

REFINE_LS_CLASS

Equivalent atoms in the refinement (§3.6.6.2.4)

REFINE_B_ISO

REFINE_OCCUPANCY

History of the refinement (§3.6.6.2.5)

REFINE_HIST

The macromolecular CIF dictionary contains many more data items for describing the refinement process than the core CIF dictionary does. In addition to new items in the REFINE category itself, additional categories have been introduced to describe in great detail the function minimized and the restraints applied, and the history of the refinement process, which often has many cycles. The REFINE_ANALYZE category can be used to give details of many of the quantities that may be used to assess the quality of the refinement. The REFINE_LS_SHELL category allows results to be reported by shells of resolution, and in effect replaces the more general core CIF category REFINE_LS_CLASS.

3.6.6.2.1. *Overall description of the refinement*

The data items in these categories are as follows:

(*a*) REFINE

- **_refine.entry_id**
  (→ **_entry.id**)
  _refine.aniso_B[1][1]
  _refine.aniso_B[1][2]
  _refine.aniso_B[1][3]
  _refine.aniso_B[2][2]
  _refine.aniso_B[2][3]
  _refine.aniso_B[3][3]
  _refine.B_iso_max
  _refine.B_iso_mean
  _refine.B_iso_min
  _refine.correlation_coeff_Fo_to_Fc
  _refine.correlation_coeff_Fo_to_Fc_free
  _refine.details (∼ ***_refine_special_details***)
+ ***_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_all
  _refine.ls_number_reflns_obs
  (∼ ***_refine_ls_number_reflns***)
  _refine.ls_number_reflns_R_free
  _refine.ls_number_reflns_R_work
  _refine.ls_number_restraints
  _refine.ls_percent_reflns_obs
  _refine.ls_percent_reflns_R_free
  _refine.ls_R_factor_all
  _refine.ls_R_factor_gt
  _refine.ls_R_factor_obs
  _refine.ls_R_factor_R_free
  _refine.ls_R_factor_R_free_error
  _refine.ls_R_factor_R_free_error_details
  _refine.ls_R_factor_R_work
  _refine.ls_R_Fsqd_factor_obs
  (∼ ***_refine_ls_R_Fsqd_factor***)
  _refine.ls_R_I_factor_obs (∼ ***_refine_ls_R_I_factor***)
  _refine.ls_redundancy_reflns_all
  _refine.ls_redundancy_reflns_obs
  _refine.ls_restrained_S_all
  _refine.ls_restrained_S_obs
  _refine.ls_shift_over_esd_max
  (∼ ***_refine_ls_shift/esd_max***)
  _refine.ls_shift_over_esd_mean
  (∼ ***_refine_ls_shift/esd_mean***)
  _refine.ls_shift_over_su_max
  (∼ ***_refine_ls_shift/su_max***)
  _refine.ls_shift_over_su_max_lt
  (∼ ***_refine_ls_shift/su_max_lt***)
  _refine.ls_shift_over_su_mean
  (∼ ***_refine_ls_shift/su_mean***)
  _refine.ls_shift_over_su_mean_lt
  (∼ ***_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_obs
  _refine.ls_wR_factor_R_free
  _refine.ls_wR_factor_R_work
  _refine.occupancy_max
  _refine.occupancy_min
  _refine.overall_FOM_free_R_set
  _refine.overall_FOM_work_R_set
  _refine.overall_SU_B
  _refine.overall_SU_ML
  _refine.overall_SU_R_Cruickshank_DPI
  _refine.overall_SU_R_free
  _refine.solvent_model_details
  _refine.solvent_model_param_bsol
  _refine.solvent_model_param_ksol

(*b*) REFINE_FUNCT_MINIMIZED

- **_refine_funct_minimized.type**
  _refine_funct_minimized.number_terms
  _refine_funct_minimized.residual
  _refine_funct_minimized.weight

The bullet (•) *indicates a category 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 marked with a plus* (+) *have companion data names for the standard uncertainty in the reported value, formed by appending the string* **_esd** *to the data name listed.*

There is already an extensive set of data names in the REFINE category of the core dictionary, and Section 3.2.3.1 should be read with the present section. The only data items discussed in this section are entries in the mmCIF dictionary that do not have a counterpart in the core CIF dictionary. Analogues of a number of *R* factors in the core CIF dictionary have been added to the mmCIF dictionary to express these same *R* factors indepen-

dently for the free and working sets of reflections. The remaining new data items have more specialized roles, which are discussed below.

The data item **_refine.entry_id** has been added to the REFINE category to provide the formal category key required by the DDL2 data model.

Many macromolecular structure refinements now use the statistical cross-validation technique of monitoring a 'free' *R* factor (Brünger, 1997). $R_{\rm free}$ is calculated the same way as the conventional least-squares *R* factor, but using a small subset of reflections that are not used in the refinement of the structural model. Thus $R_{\rm free}$ tests how well the model predicts experimental observations that are not themselves used to fit the model.

The mmCIF dictionary provides data names for $R_{\rm free}$ and for the complementary $R_{\rm work}$ values for the 'working' set of reflections, which are the reflections that are used in the refinement. Separate data items are provided for unweighted and weighted versions of each *R* factor. A fixed percentage of the total number of reflections is usually assigned to the free group, and this percentage can be specified. Further details about the method used for selecting the free reflections can be given using **_reflns.R_free_details**. The estimated error in the $R_{\rm free}$ value may also be given, along with the method used for determining its value.

The purposes of having a set of reflections that are not used in the refinement are to monitor the progress of the refinement and to ensure that the *R* factor is not being artificially reduced by the introduction of too many parameters. However, as the refinement converges, the working and free *R* factors both approach stable values. It is common practice, particularly in structures at high resolution, to stop monitoring $R_{\rm free}$ at this point and to include all the reflections in the final rounds of refinement. It is thus worth noting a distinction between **_refine.ls_R_factor_obs** and **_refine.ls_R_factor_R_work**: **_refine.ls_R_factor_obs** relates to a refinement in which all reflections more intense than a specified threshold were used, while **_refine.ls_R_factor_R_work** relates to a refinement in which a subset of the observed reflections were excluded from the refinement and were used to calculate the free *R* factor. The dictionary allows the use of both values if a free *R* factor were calculated for most of the refinement, but all of the observed reflections were used in the final rounds of refinement; the protocol for this may be explained in **_refine.details**. When a full history of the refinement is provided using data items in the REFINE_HIST category, it is preferable to specify a change in protocol using data items in this category.

Other data items help to provide an assessment of the quality of the refinement. The scale-independent correlation coefficient between the observed and calculated structure factors may be recorded for the reflections included in the refinement using the data item **_refine.correlation_coeff_Fo_to_Fc**. There is a similar data item for the reflections that were not included in the refinement.

Overall standard uncertainties for positional and displacement parameters can be recorded according to a number of conventions. A maximum-likelihood residual for the positional parameters can be given using **_refine.overall_SU_ML** and the corresponding value for the displacement parameters can be given using **_refine.overall_SU_B**. Diffraction-component precision indexes for the displacement parameters based on the crystallographic *R* factor (the Cruickshank DPI; Cruickshank, 1999) can be given using **_refine.overall_SU_R_Cruickshank_DPI**. The corresponding value for $R_{\rm free}$ can be given using **_refine.overall_SU_R_free**.

Example 3.6.6.7. *Results of the overall refinement of an HIV-1 protease structure (PDB 5HVP) described using data items in the* REFINE *and* REFINE_FUNCT_MINIMIZED *categories.*

```
_refine.entry_id                          '5HVP'
_refine.ls_number_reflns_obs              12901
_refine.ls_number_restraints              6609
_refine.ls_number_parameters              7032
_refine.ls_R_factor_obs                   0.176
_refine.ls_weighting_scheme               calc
_refine.ls_weighting_details
;  Sigdel model of Konnert-Hendrickson:
   Sigdel: Afsig +  Bfsig*(sin(theta)/lambda-1/6)
   Afsig = 22.0, Bfsig = -150.0 at the beginning
      of refinement.
   Afsig = 15.5, Bfsig =  -50.0 at the end of
      refinement.
;
loop_
  _refine_funct_minimized.type
  _refine_funct_minimized.number_terms
  _refine_funct_minimized.residual
    'sum(W*Delta(Amplitude)^2^'       3009   1621.3
    'sum(W*Delta(Plane+Rigid)^2^'       85     56.68
    'sum(W*Delta(Distance)^2^'        1219    163.59
    'sum(W*Delta(U-tempfactors)^2^'   1192     69.338
```

The quality of a data set used for the refinement of a macromolecular structure is often given not only in terms of the scaling residuals, but also in terms of the data redundancy (the ratio of the number of reflections measured to the number of crystallographically unique reflections). Data items are provided to express the redundancy of all reflections, as well as those that have been marked as 'observed' (*i.e.* exceeding the threshold for inclusion in the refinement). The percentage of the total number of reflections that are considered observed is another metric of the quality of the data set, and a data item is provided for this (**_refine.ls_percent_reflns_obs**).

The limited resolution of many macromolecular data sets makes it inappropriate to refine anisotropic displacement factors for each atom. For these low- to medium-resolution studies, an overall anisotropic displacement model may be refined. The data items **_refine.aniso_B*** are provided for recording the unique elements of the matrix that describes the refined anisotropy.

The two-parameter method for modelling the contribution of the bulk solvent to the scattering proposed by Tronrud is used in several refinement programs. The data items **_refine.solvent_model_*** can be used to record the scale and displacement factors of this model, and any special aspects of its application to the refinement.

The average phasing figure of merit can be given for the working and free reflections. Unusually high or low values of displacement factors or occupancies can be a sign of problems with the refinement, so data items are provided to record the high, low and mean values of each. Further indicators of the quality of the refinement are found in the REFINE_ANALYZE category (Section 3.6.6.2.2).

The data items in the REFINE_FUNCT_MINIMIZED category allow a brief description of the function minimized during refinement to be given (Example 3.6.6.7). It is not possible to reconstruct the functioned minimized during the refinement by automatic parsing of the values of these data items, but the details given in them may still be helpful to someone reading the mmCIF.

3.6.6.2.2. *Analysis of the refined structure*

The data items in this category are as follows:

REFINE_ANALYZE

- **_refine_analyze.entry_id**
  → **_entry.id**

```
_refine_analyze.Luzzati_coordinate_error_free
_refine_analyze.Luzzati_coordinate_error_obs
_refine_analyze.Luzzati_d_res_low_free
_refine_analyze.Luzzati_d_res_low_obs
_refine_analyze.Luzzati_sigma_a_free
_refine_analyze.Luzzati_sigma_a_free_details
_refine_analyze.Luzzati_sigma_a_obs
_refine_analyze.Luzzati_sigma_a_obs_details
_refine_analyze.number_disordered_residues
_refine_analyze.occupancy_sum_hydrogen
_refine_analyze.occupancy_sum_non_hydrogen
_refine_analyze.RG_d_res_high
_refine_analyze.RG_d_res_low
_refine_analyze.RG_free
_refine_analyze.RG_free_work_ratio
_refine_analyze.RG_work
```

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

In small-molecule crystallography, there is general agreement on the metrics that should be used to assess the quality of a structure determination, and data items in the REFINE category of the core CIF dictionary can be used to record them. For macromolecular structure determinations, no such agreement has been achieved yet and new metrics are frequently suggested as the field evolves. The REFINE_ANALYZE category can be used to record the metrics that were in common use at the time that the mmCIF dictionary was constructed; it is anticipated that new metrics will be added in future versions of the dictionary, and that some of the current metrics may fall into disuse.

Luzzati (1952) devised a method for estimating the average positional shift that would be needed in an idealized refinement to reach an R factor of zero by using a plot of R factors against resolution. For some time, macromolecular crystallographers have used a modification of this approach to assess the average positional error. Recent practice has used Luzzati plots based on the free R values to yield a cross-validated error estimate. Data items are provided for recording these coordinate-error estimates and the range of resolution included in the plot (Example 3.6.6.8). Related data names allow the specification of the value of σ_a used in constructing the Luzzati plot.

A general feature of introducing more parameters in the model of the structure is a reduction in the R factor, but the statistical significance of this is often obscured by the simultaneous reduction in the ratio of observations to parameters. Attempts to extend Hamilton's (1965) test to macromolecular structures are usually confounded by the use of restraints. Tickle *et al.* (1998) proposed the use of a Hamilton generalized R factor analyzed separately for reflections in the working set (those used in the refinement) and for reflections in the free set (those set aside for cross validation), and these metrics are often reported in the literature. Data items are provided for recording the Hamilton generalized R factor for the working and free set of reflections, and for the ratio of the two.

Other indicators of a successful refinement involve the relative order of the model. Data items are provided for recording the sum of the occupancies of the hydrogen and non-hydrogen atoms in the model. The number of disordered residues may also be recorded.

3.6.6.2.3. *Restraints and refinement by shells of resolution*

The data items in these categories are as follows:

(*a*) REFINE_LS_RESTR

```
• _refine_ls_restr.type
  _refine_ls_restr.criterion
  _refine_ls_restr.dev_ideal
  _refine_ls_restr.dev_ideal_target
  _refine_ls_restr.number
  _refine_ls_restr.rejects
  _refine_ls_restr.weight
```

Example 3.6.6.8. *Aspects of the refinement of an HIV-1 protease structure (PDB 5HVP) described with data items in the REFINE_ANALYZE category.*

```
loop_
_refine_analyze.entry_id                          '5HVP'
_refine_analyze.Luzzati_coordinate_error_obs       0.32
_refine_analyze.Luzzati_d_res_low_obs              5.0
```

(*b*) REFINE_LS_RESTR_NCS

```
• _refine_ls_restr_ncs.dom_id
        → _struct_ncs_dom.id
  _refine_ls_restr_ncs.ncs_model_details
  _refine_ls_restr_ncs.rms_dev_B_iso
  _refine_ls_restr_ncs.rms_dev_position
  _refine_ls_restr_ncs.weight_B_iso
  _refine_ls_restr_ncs.weight_position
```

(*c*) REFINE_LS_RESTR_TYPE

```
• _refine_ls_restr_type.type
        → _refine_ls_restr.type
  _refine_ls_restr_type.distance_cutoff_high
  _refine_ls_restr_type.distance_cutoff_low
```

(*d*) REFINE_LS_SHELL

```
• _refine_ls_shell.d_res_high
• _refine_ls_shell.d_res_low
  _refine_ls_shell.number_reflns_all
  _refine_ls_shell.number_reflns_obs
  _refine_ls_shell.number_reflns_R_free
  _refine_ls_shell.number_reflns_R_work
  _refine_ls_shell.percent_reflns_obs
  _refine_ls_shell.percent_reflns_R_free
  _refine_ls_shell.R_factor_all
  _refine_ls_shell.R_factor_obs
  _refine_ls_shell.R_factor_R_free
  _refine_ls_shell.R_factor_R_free_error
  _refine_ls_shell.R_factor_R_work
  _refine_ls_shell.redundancy_reflns_all
  _refine_ls_shell.redundancy_reflns_obs
  _refine_ls_shell.wR_factor_all
  _refine_ls_shell.wR_factor_obs
  _refine_ls_shell.wR_factor_R_free
  _refine_ls_shell.wR_factor_R_work
```

(*e*) REFINE_LS_CLASS

```
• _refine_ls_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. 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.

These categories were introduced in the mmCIF dictionary to allow a detailed description of several aspects of structure refinement to be given. Data items in the REFINE_LS_RESTR category allow geometric restraints to be specified and the deviations of restrained parameters from ideal values in the final model to be given. The type of the geometric restraints can be described in more detail using data items in the REFINE_LS_RESTR_TYPE category. Data items in the REFINE_LS_RESTR_NCS category can be used to give information about any restraints on noncrystallographic symmetry used in the refinement and the category REFINE_LS_SHELL contains data items that allow the results of refinement to be given by shells of resolution.

Data items in the REFINE_LS_RESTR category can be used to record details about the restraints applied to various classes of parameters during least-squares refinement (Example 3.6.6.9). It is clearly useful to tabulate the various classes of restraint, their deviation from ideal target values and the criteria used to reject

Example 3.6.6.9. *Results of the refinement of an HIV-1 protease structure (PDB 5HVP) described with data items in the* REFINE_LS_RESTR *and* REFINE_LS_SHELL *categories.*

```
loop_
_refine_ls_restr.type
_refine_ls_restr.dev_ideal_target
_refine_ls_restr.dev_ideal
_refine_ls_restr.number
_refine_ls_restr.criterion
_refine_ls_restr.rejects
'p_bond_d'          0.020  0.018  1654 '>2 sigma' 22
'p_angle_d'         0.030  0.038  2246 '>2 sigma' 139
'p_planar_d'        0.040  0.043  498  '>2 sigma' 21
'p_planar'          0.020  0.015  270  '>2 sigma' 1
'p_chiral'          0.150  0.177  278  '>2 sigma' 2
'p_singtor_nbd'     0.500  0.216  582  '>2 sigma' 0
'p_multtor_nbd'     0.500  0.207  419  '>2 sigma' 0
'p_xyhbond_nbd'     0.500  0.245  149  '>2 sigma' 0
'p_planar_tor'      3.0    2.6    203  '>2 sigma' 9
'p_staggered_tor'   15.0   17.4   298  '>2 sigma' 31
'p_orthonormal_tor' 20.0   18.1   12   '>2 sigma' 1

loop_
_refine_ls_shell.d_res_low
_refine_ls_shell.d_res_high
_refine_ls_shell.number_reflns_obs
_refine_ls_shell.R_factor_obs
   8.00   4.51   1226   0.196
   4.51   3.48   1679   0.146
   3.48   2.94   2014   0.160
   2.94   2.59   2147   0.182
   2.59   2.34   2127   0.193
   2.34   2.15   2061   0.203
   2.15   2.00   1647   0.188
```

parameters that lie too far from a target, as these data are often published as part of a description of the refinement and are often deposited with the coordinates in an archive. However, the types of restraints applied depend strongly on the software package used, and as new refinement packages regularly become available, it was clearly not advisable to provide program-specific data items in the mmCIF dictionary. The approach taken in the mmCIF dictionary has been to allow the value of **_refine_ls_restr.type** to be a free-text field, so that arbitrary labels can be given to restraints that are particular to a software package, but to recommend the use of specific labels for restraints applied by particular programs. The dictionary provides examples for labels specific to the programs *PROTIN/PROLSQ* (Hendrickson & Konnert, 1979) and *RESTRAIN* (Driessen *et al.*, 1989). These program-specific representations have particular prefixes; thus the value p_bond_d is a bond-distance restraint as applied by *PROTIN/PROLSQ*. Values for **_refine_ls_restr.type** appropriate for other refinement programs may be suggested in future versions of the mmCIF dictionary.

Data items in the REFINE_LS_RESTR_TYPE category can be used to specify the ranges within which quantities are allowed to vary for each type of restraint. The special value indicated by a full stop (.) represents a restraint unbounded on the high or low side.

Data items in the REFINE_LS_RESTR_NCS category can be used to record details about the restraints applied to atom positions in domains related by noncrystallographic symmetry during least-squares refinement, and also to record the deviation of the restrained atomic parameters at the end of the refinement. The domains related by noncrystallographic symmetry are defined in the STRUCT_NCS_DOM and related categories (see Section 3.6.7.5.5). The quantities that can be recorded for each restrained domain are the root-mean-square deviations of the displacement and positional parameters, and the weighting coefficients used in the noncrystallographic restraint of each type of parameter. Any special aspects of the way the restraints were applied may be described using **_refine_ls_restr_ncs.ncs_model_details**.

Data items in the REFINE_LS_SHELL category are used to summarize details of the results of the least-squares refinement by shells of resolution (Example 3.6.6.9). The resolution range, in ångströms, forms the category key; for each shell the quantities reported, such as the number of reflections above the threshold for counting as significantly intense, are all defined in the same way as the corresponding data items used to describe the results of the overall refinement in the REFINE category.

The core dictionary category REFINE_LS_CLASS was introduced after the release of the first version of the mmCIF dictionary. It provides a more general way of describing the treatment of particular subsets of the observations, but it is not expected to be used in macromolecular structural studies, where partition by shells of resolution is traditional.

3.6.6.2.4. *Equivalent atoms in the refinement*

The data items in these categories are as follows:

(*a*) REFINE_B_ISO

- **_refine_B_iso.class**
  _refine_B_iso.details
  _refine_B_iso.treatment
  _refine_B_iso.value

(*b*) REFINE_OCCUPANCY

- **_refine_occupancy.class**
  _refine_occupancy.details
  _refine_occupancy.treatment
  _refine_occupancy.value

The bullet (•) *indicates a category key.*

In macromolecular structure refinement, displacement factors or occupancies are often treated as equivalent for groups of atoms. An example would be the case where most of the atoms in the structure are refined with isotropic displacement factors, but a bound metal atom is allowed to refine anisotropically. Another example would be where the occupancies for all of the atoms in the protein part of a macromolecular complex are fixed at 1.0, but the occupancies of atoms in a bound inhibitor are refined. The REFINE_B_ISO and REFINE_OCCUPANCY categories can be used to record this information (Example 3.6.6.10).

Example 3.6.6.10. *The handling of displacement factors and occupancies during the refinement of an HIV-1 protease structure (PDB 5HVP) described with data items in the* REFINE_B_ISO *and* REFINE_OCCUPANCY *categories.*

```
loop_
_refine_B_iso.class
_refine_B_iso.treatment
   'protein'    isotropic
   'solvent'    isotropic
   'inhibitor'  isotropic

loop_
_refine_occupancy.class
_refine_occupancy.treatment
_refine_occupancy.value
_refine_occupancy.details
   'protein'                   fix  1.00  .
   'solvent'                   fix  1.00  .
   'inhibitor orientation 1'   fix  0.65  .
   'inhibitor orientation 2'   fix  0.35
; The inhibitor binds to the enzyme in two
  alternative conformations. The occupancy of
  each conformation was adjusted so as to result
  in approximately equal mean thermal factors
  for the atoms in each conformation.
;
```

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`