International Tables for Crystallography (2006). Vol. G, Section 3.6.6, pp. 152–164.

Example 3.6.5.3. *The crystal used in the determination of an HIV-1 protease structure (PDB 5HVP) described using data items in the* EXPTL *and* EXPTL_CRYSTAL *categories.*

```
_exptl.entry_id                   '5HVP'
_exptl.crystals_number            1
_exptl.method     'single-crystal x-ray diffraction'
_exptl.method_details
; graphite monochromatized Cu K(alpha) fixed tube
  and Siemens multiwire detector used
;
_exptl_crystal.id                 1
_exptl_crystal.colour             'colorless'
_exptl_crystal.density_percent_sol0.57
_exptl_crystal.description       'rectangular plate'
_exptl_crystal.size_max           0.30
_exptl_crystal.size_mid           0.20
_exptl_crystal.size_min           0.05
```

3.6.5.3.2. *Crystal growth*

The data items in these categories are as follows:

```
(a) EXPTL_CRYSTAL_GROW
• _exptl_crystal_grow.crystal_id
       → _exptl_crystal.id
  _exptl_crystal_grow.apparatus
  _exptl_crystal_grow.atmosphere
  _exptl_crystal_grow.details
  _exptl_crystal_grow.method
  _exptl_crystal_grow.method_ref
  _exptl_crystal_grow.pH
+ _exptl_crystal_grow.pressure
  _exptl_crystal_grow.seeding
  _exptl_crystal_grow.seeding_ref
+ _exptl_crystal_grow.temp
  _exptl_crystal_grow.temp_details
  _exptl_crystal_grow.time

(b) EXPTL_CRYSTAL_GROW_COMP
• _exptl_crystal_grow_comp.crystal_id
       → _exptl_crystal.id
• _exptl_crystal_grow_comp.id
  _exptl_crystal_grow_comp.conc
  _exptl_crystal_grow_comp.details
  _exptl_crystal_grow_comp.name
  _exptl_crystal_grow_comp.sol_id
  _exptl_crystal_grow_comp.volume
```

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. 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.*

Crystallization strategies and protocols are very varied and may not lend themselves to a formal tabulation. Common or well defined techniques may be indicated using the data item **`_exptl_crystal_grow.method`**, and a literature reference, where appropriate, may be given using **`_exptl_crystal_grow.method_ref`**. Frequently, however, a detailed description of methodology is required; this can be given in **`_exptl_crystal_grow.details`**. Example 3.6.5.4 shows how information about strategies that were attempted and proved unsuccessful can be recorded. In circumstances such as this, the data item **`_exptl_crystal_grow.pH`** would record the final pH.

Where the crystallization protocol is well defined, it is useful to list the individual components of the solution in the category EXPTL_CRYSTAL_GROW_COMP. Example 3.6.5.4 labels the solutions used as 1 and 2, in accordance with the convention that solution 1 contains the molecule to be crystallized and solution 2 (and if necessary additional solutions) contains the precipitant. However, it is permissible and may be preferable to use more explicit labels such as 'well solution' in the **`_exptl_crystal_grow_comp.sol_id`** field.

Example 3.6.5.4. *The growth of HIV-1 protease crystals (PDB 5HVP) described with data items in the* EXPTL_CRYST_GROW *and* EXPTL_CRYSTAL_GROW_COMP *categories.*

```
_exptl_crystal_grow.crystal_id     1
_exptl_crystal_grow.method         'hanging drop'
_exptl_crystal_grow.apparatus      'Linbro plates'
_exptl_crystal_grow.atmosphere     'room air'
_exptl_crystal_grow.pH             4.7
_exptl_crystal_grow.temp           18(3)
_exptl_crystal_grow.time    'approximately 2 days'
_exptl_crystal_grow.details
; The dependence on pH for successful crystal growth
  is very sharp. At pH 7.4 only showers of tiny
  crystals grew, at pH 7.5 well formed single
  crystals grew, at pH 7.6 no crystallization
  occurred at all.
;
loop_
  _exptl_crystal_grow_comp.crystal_id
  _exptl_crystal_grow_comp.id
  _exptl_crystal_grow_comp.sol_id
  _exptl_crystal_grow_comp.name
  _exptl_crystal_grow_comp.volume
  _exptl_crystal_grow_comp.conc
  _exptl_crystal_grow_comp.details
  1  1  1  'HIV-1 protease'  '0.002 ml'  '6 mg/ml'
; The protein solution was in a buffer containing
  25 mM NaCl, 100 mM NaMES/MES buffer, pH 7.5,
  3 mM NaAzide
;
  1  2  2  'NaCl'  '0.200 ml'  '4 M'
  'in 3 mM NaAzide'
  1  3  2  'Acetic Acid'  '0.047 ml'  '100 mM'
  'in 3 mM NaAzide'
  1  4  2  'Na Acetate'  '0.053 ml'  '100 mM'
; in 3 mM NaAzide. Buffer components were mixed
  to produce a pH of 4.7 according to a ratio
  calculated from the pKa. The actual pH of
  solution 2 was not measured.
;
  1  5  2  'water'  '0.700 ml'  'neat'
 'in 3 mM NaAzide'
```

3.6.6. Analysis

The mmCIF dictionary contributes several new categories and data items to the REFINE and REFLN category groups. These reflect common practices in macromolecular crystallography in refinement and in the handling of experimental observations.

A new category group, the PHASING group, has been introduced to provide a structured description of phasing strategies, as macromolecular crystallography differs strongly from small-molecule crystallography in how phases are determined. The data model for phasing in the current version of the mmCIF dictionary cannot describe all approaches to phasing yet. Additions and revisions to the data items in the PHASING group of categories are anticipated in future versions of the dictionary.

3.6.6.1. Phasing

The categories describing phasing are as follows:

PHASING group

Overall description of phasing (§3.6.6.1.1)
  PHASING

Phasing via molecular averaging (§3.6.6.1.2)
  PHASING_AVERAGING

Phasing via isomorphous replacement (§3.6.6.1.3)
  PHASING_ISOMORPHOUS

Phasing via multiple-wavelength anomalous dispersion (§3.6.6.1.4)
  PHASING_MAD
  PHASING_MAD_CLUST

PHASING_MAD_EXPT
PHASING_MAD_RATIO
PHASING_MAD_SET

Phasing via multiple isomorphous replacement (§3.6.6.1.5)

PHASING_MIR
PHASING_MIR_DER
PHASING_MIR_DER_REFLN
PHASING_MIR_DER_SHELL
PHASING_MIR_DER_SITE
PHASING_MIR_DER_SHELL

Phasing data sets (§3.6.6.1.6)

PHASING_SET
PHASING_SET_REFLN

The data items in the PHASING category group can be used to record details about the phasing of the structure and cover the various methods used in the phasing process. Many data items are provided for multiple isomorphous replacement (MIR) and multiple-wavelength anomalous dispersion (MAD). More limited sets of data items are provided for phasing using molecular averaging and phasing *via* using a structure that is isomorphous to the present structure. The current version of the mmCIF dictionary does not provide specific data items for recording the details of phasing *via* molecular replacement.

3.6.6.1.1. *Overall description of phasing*

The single data item in this category is as follows:

```
PHASING
• _phasing.method
```

The bullet (•) indicates a category key.

Phasing of macromolecular structures often involves the application of more than one of the methods described in the PHASING section of the mmCIF dictionary, such as when phases generated from a multiple isomorphous replacement experiment are improved by molecular averaging. The PHASING category is used to list the methods that were used.

At present, the category contains a single data item, the purpose of which is to specify the method employed in the structure determination. It may have one or more of the values listed in the dictionary (Example 3.6.6.1).

Example 3.6.6.1. *The methods used to generate the phases for a hypothetical structure described with the data item in the* PHASING *category.*

```
loop_
_phasing.method
   'mir'
   'averaging'
```

3.6.6.1.2. *Phasing via molecular averaging*

The data items in this category are as follows:

```
PHASING_AVERAGING
• _phasing_averaging.entry_id
        → _entry.id
  _phasing_averaging.details
  _phasing_averaging.method
```

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

When more than one copy of a molecule is present in the asymmetric unit, phases can be improved by averaging an electron-density map over the multiple images of the molecule. In some special cases with very high noncrystallographic symmetry, *de novo* phases have been derived by iterative application of molecular averaging, but more often averaging is used to improve phases determined by another method.

There are many protocols used for phasing with averaging and they are very varied. It was not thought to be appropriate to specify data items for any one approach in the current version of the mmCIF dictionary. The data items that are provided allow a text-based description of the protocol to be given; a formalism for recording a fully parsable description of molecular averaging needs to be developed for future revisions of the dictionary.

Example 3.6.6.2. *Phase improvement with molecular averaging for a hypothetical structure described with data items in the* PHASING_AVERAGING *category.*

```
_phasing_averaging.entry_id    'EXAMHYPO'
_phasing_averaging.method
; Iterative threefold averaging alternating with
  phase extensions by 0.5 reciprocal lattice units
  per cycle.
;

_phasing_averaging.details
; The position of the threefold axis was redetermined
  every five cycles.
;
```

Data items in the PHASING_AVERAGING category allow free-text descriptions to be given of the method used for structure determination or phase improvement using averaging over multiple observations of the molecule in the asymmetric unit and of any specific details of the application of the method to the current structure determination (Example 3.6.6.2). Note that the reference to the method is to be used to describe the method itself, and not as a reference to a software package; references to software packages would be made using data items in the SOFTWARE category.

3.6.6.1.3. *Phasing via isomorphous replacement*

The data items in this category are as follows:

```
PHASING_ISOMORPHOUS
• _phasing_isomorphous.entry_id
        → _entry.id
  _phasing_isomorphous.details
  _phasing_isomorphous.method
  _phasing_isomorphous.parent
```

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

Phases for many macromolecular structures are obtained from a previous determination of the same structure in the same crystal lattice. Examples of this are the determination of the structure of a point mutant or the determination of a structure in which a ligand is bound to an active site that was empty in the previous structure determination. In these cases, the new structure is essentially isomorphous with the parent structure, hence this method of phasing is termed ‘isomorphous phasing’ in the mmCIF dictionary. It is not to be confused with multiple isomorphous phasing (MIR), a phasing technique that involves the use of heavy-atom derivatives. MIR phasing is discussed in Section 3.6.6.1.5.

Not much information is needed to characterize isomorphous phasing. The ‘parent’ structure (the structure used to generate the initial phases for the present structure) is described in a free-text field and a second free-text field can be used to give details of the application of the method to the determination of the present structure (for instance, the removal of solvent or a bound ligand). In Example 3.6.6.3, the parent structure is the PDB entry 5HVP and the structure that is the subject of the present data block is identified as ‘HVP+CmpdA’. `_phasing_isomorphous.method` allows

Example 3.6.6.3. *Isomorphous replacement phasing of an HIV-1 protease structure described using data items in the* PHASING_ISOMORPHOUS *category.*

```
_phasing_isomorphous.entry_id    'HVP+CmpdA'
_phasing_isomorphous.parent      'PDB entry 5HVP'
_phasing_isomorphous.details
; The inhibitor and all solvent atoms were removed
  from the parent structure before beginning
  refinement. All static disorder present in the
  parent structure was also removed.
;
```

any formal techniques that were used in the application of the method to the present structure determination to be described, for example rigid-body refinement. Note that this data item is not to be used to reference a software package; this would be done using data items in the SOFTWARE category.

3.6.6.1.4. *Phasing via multiple-wavelength anomalous dispersion*

The data items in these categories are as follows:

(*a*) PHASING_MAD

```
• _phasing_MAD.entry_id
       → _entry.id
  _phasing_MAD.details
  _phasing_MAD.method
```

(*b*) PHASING_MAD_CLUST

```
• _phasing_MAD_clust.expt_id
       → _phasing_MAD_clust.expt_id
• _phasing_MAD_clust.id
  _phasing_MAD_clust.number_set
```

(*c*) PHASING_MAD_EXPT

```
• _phasing_MAD_expt.id
  _phasing_MAD_expt.delta_delta_phi
  _phasing_MAD_expt.delta_phi
  _phasing_MAD_expt.delta_phi_sigma
  _phasing_MAD_expt.mean_fom
  _phasing_MAD_expt.number_clust
  _phasing_MAD_expt.R_normal_all
  _phasing_MAD_expt.R_normal_anom_scat
```

(*d*) PHASING_MAD_RATIO

```
• _phasing_MAD_ratio.expt_id
       → _phasing_MAD_expt.id
• _phasing_MAD_ratio.clust_id
       → _phasing_MAD_clust.id
• _phasing_MAD_ratio.wavelength_1
       → _phasing_MAD_set.wavelength
• _phasing_MAD_ratio.wavelength_2
       → _phasing_MAD_set.wavelength
  _phasing_MAD_ratio.d_res_high
  _phasing_MAD_ratio.d_res_low
  _phasing_MAD_ratio.ratio_one_wl
  _phasing_MAD_ratio.ratio_one_wl_centric
  _phasing_MAD_ratio.ratio_two_wl
```

(*e*) PHASING_MAD_SET

```
• _phasing_MAD_set.clust_id
       → _phasing_MAD_clust.id
• _phasing_MAD_set.expt_id
       → _phasing_MAD_expt.id
• _phasing_MAD_set.set_id
       → _phasing_set.id
• _phasing_MAD_set.wavelength
  _phasing_MAD_set.d_res_high
  _phasing_MAD_set.d_res_low
  _phasing_MAD_set.f_double_prime
  _phasing_MAD_set.f_prime
  _phasing_MAD_set.wavelength_details
```

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.

Example 3.6.6.4. *MAD phasing of the structure of N-cadherin (Shapiro et al., 1995) described using data items in the* PHASING_MAD *and related categories.*

```
_phasing_MAD.entry_id       'NCAD'

loop_
_phasing_MAD_expt.id
_phasing_MAD_expt.number_clust
_phasing_MAD_expt.R_normal_all
_phasing_MAD_expt.R_normal_anom_scat
_phasing_MAD_expt.delta_delta_phi
_phasing_MAD_expt.delta_phi_sigma
_phasing_MAD_expt.mean_fom
   1  2  0.063  0.451  58.5  20.3  0.88
   2  1  0.051  0.419  36.8  18.2  0.93

loop_
_phasing_MAD_clust.id
_phasing_MAD_clust.expt_id
_phasing_MAD_clust.number_set
  'four wavelength'  1  4
  'five wavelength'  1  5
  'five wavelength'  2  5

loop_
_phasing_MAD_ratio.expt_id
_phasing_MAD_ratio.clust_id
_phasing_MAD_ratio.wavelength_1
_phasing_MAD_ratio.wavelength_2
_phasing_MAD_ratio.d_res_low
_phasing_MAD_ratio.d_res_high
_phasing_MAD_ratio.ratio_two_wl
_phasing_MAD_ratio.ratio_one_wl
_phasing_MAD_ratio.ratio_one_wl_centric
   1  'four wavelength'  1.4013 1.4013  20.00  4.00
        .  0.084 0.076
   1  'four wavelength'  1.4013 1.3857  20.00  4.00
     0.067      .     .
   1  'four wavelength'  1.4013 1.3852  20.00  4.00
     0.051      .     .
   1  'four wavelength'  1.4013 1.3847  20.00  4.00
     0.044      .     .
   1  'four wavelength'  1.3857 1.3857  20.00  4.00
        .  0.110 0.049
   1  'four wavelength'  1.3857 1.3852  20.00  4.00
     0.049      .     .
# - - - abbreviated - - -

loop_
_phasing_MAD_set.expt_id
_phasing_MAD_set.clust_id
_phasing_MAD_set.set_id
_phasing_MAD_set.wavelength
_phasing_MAD_set.wavelength_details
_phasing_MAD_set.d_res_low
_phasing_MAD_set.d_res_high
_phasing_MAD_set.f_prime
_phasing_MAD_set.f_double_prime
   1 'four wavelength' aa 1.4013 'pre-edge'  20.00
      3.00   -12.48   3.80
   1 'four wavelength' bb 1.3857 'peak'      20.00
      3.00.  -31.22  17.20
   1 'four wavelength' cc 1.3852 'edge'      20.00
      3.00   -13.97  29.17
```

PHASING_MAD and related categories are used to provide information about phasing using the multiple-wavelength anomalous dispersion (MAD) technique. The data model used for MAD phasing in the current version of the mmCIF dictionary is that of Hendrickson, as exemplified in the structure determination of N-cadherin (Shapiro *et al.*, 1995; Example 3.6.6.4). In current practice, MAD phasing is often treated as a special case of MIR phasing and the PHASING_MIR categories would be more appropriate to describe the results.

Unlike the PHASING_MIR categories, there is no provision in the current mmCIF model of MAD phasing for analysis of the overall phasing statistics and the contribution to the phasing of each data set by bins of resolution, and no provision for giving a list of the phased reflections. This will need to be addressed in future versions of the mmCIF dictionary.

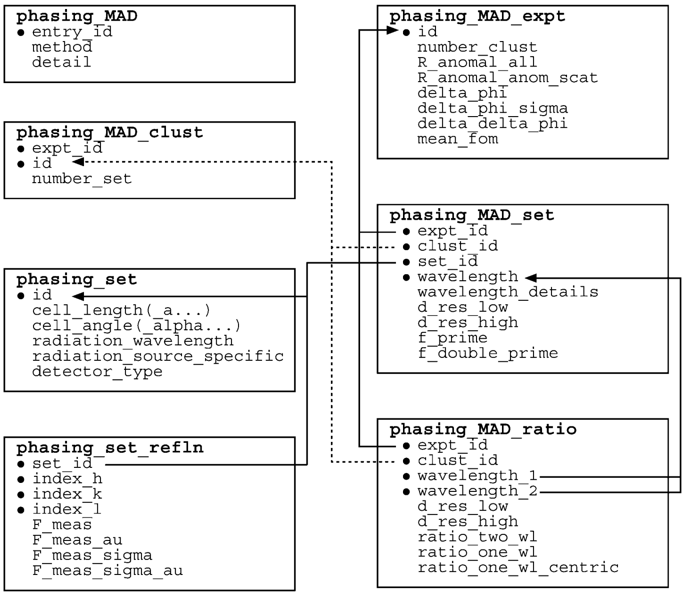

Fig. 3.6.6.1. The family of categories used to describe MAD phasing. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (●). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

The relationships between categories describing MAD phasing are shown in Fig. 3.6.6.1.

Data items in the PHASING_MAD category allow a brief overview of the method that was used to be given and allow special aspects of the phasing strategy to be noted; data items in this category are analogous to the data items in the other overview categories describing phasing techniques.

In the data model for MAD phasing used in the present version of the mmCIF dictionary, a collection of data sets measured at different wavelengths can be used to construct more than one set of phases. These phase sets will produce electron-density maps with different local properties. The model of the structure is often constructed using information from a collection of these maps. The collections of multiple phase sets are referred to as ‘experiments’ and the groups of data sets that contribute to each experiment are referred to as ‘clusters’. Data items in PHASING_MAD_EXPT identify each experiment and give the number of contributing clusters. Additional data items record the phase difference between the structure factors due to normal scattering from all atoms and from only the anomalous scatterers, the standard uncertainty of this quantity, the mean figure of merit, and a number of other indicators of the quality of the phasing.

Data items in the PHASING_MAD_CLUST category can be used to label the clusters of data sets and give the number of data sets allocated to each cluster. In Example 3.6.6.4 two experiments are described. The first experiment contains two clusters, one of which contains four data sets and the second of which contains five data sets. The second experiment contains a single cluster of five data sets. Note that the author has chosen informative labels to identify the clusters (‘four wavelength’, ‘five wavelength’). Carefully chosen labels can help someone reading the mmCIF to trace the complex relationships between the categories.

Data items in the PHASING_MAD_RATIO category can be used to record the ratios of phasing statistics (Bijvoet differences) between pairs of data sets in a MAD phasing experiment, within shells of resolution characterized by **_phasing_MAD_ratio.d_res_high** and ***.d_res_low**.

The data sets used in the MAD phasing experiments are described using data items in the PHASING_MAD_SET category. Each data set is characterized by resolution shell and wavelength, and by the f' and f'' components of the anomalous scattering factor at that wavelength. The actual observations in each data set and the experimental conditions under which they were made are recorded using data items in the PHASING_SET and PHASING_SET_REFLN categories.

3.6.6.1.5. *Phasing via multiple isomorphous replacement*

The data items in these categories are as follows:

(*a*) PHASING_MIR
• **_phasing_MIR.entry_id**
  → **_entry.id**
  _phasing_MIR.details
  _phasing_MIR.d_res_high
  _phasing_MIR.d_res_low
  _phasing_MIR.FOM
  _phasing_MIR.FOM_acentric
  _phasing_MIR.FOM_centric
  _phasing_MIR.method
  _phasing_MIR.reflns
  _phasing_MIR.reflns_acentric
  _phasing_MIR.reflns_centric
  _phasing_MIR.reflns_criterion

(*b*) PHASING_MIR_SHELL
• **_phasing_MIR_shell.d_res_high**
• **_phasing_MIR_shell.d_res_low**
  _phasing_MIR_shell.FOM
  _phasing_MIR_shell.FOM_acentric
  _phasing_MIR_shell.FOM_centric
  _phasing_MIR_shell.loc
  _phasing_MIR_shell.mean_phase
  _phasing_MIR_shell.power
  _phasing_MIR_shell.R_cullis
  _phasing_MIR_shell.R_kraut
  _phasing_MIR_shell.reflns
  _phasing_MIR_shell.reflns_acentric
  _phasing_MIR_shell.reflns_anomalous
  _phasing_MIR_shell.reflns_centric

(*c*) PHASING_MIR_DER
• **_phasing_MIR_der.id**
  _phasing_MIR_der.d_res_high
  _phasing_MIR_der.d_res_low
  _phasing_MIR_der.der_set_id
  → **_phasing_set.id**
  _phasing_MIR_der.details
  _phasing_MIR_der.native_set_id
  → **_phasing_set.id**
  _phasing_MIR_der.number_of_sites
  _phasing_MIR_der.power_acentric
  _phasing_MIR_der.power_centric
  _phasing_MIR_der.R_cullis_acentric
  _phasing_MIR_der.R_cullis_anomalous
  _phasing_MIR_der.R_cullis_centric
  _phasing_MIR_der.reflns_acentric
  _phasing_MIR_der.reflns_anomalous
  _phasing_MIR_der.reflns_centric
  _phasing_MIR_der.reflns_criteria

(*d*) PHASING_MIR_DER_REFLN
• **_phasing_MIR_der_refln.der_id**
  → **_phasing_MIR_der.id**
• **_phasing_MIR_der_refln.index_h**
• **_phasing_MIR_der_refln.index_k**
• **_phasing_MIR_der_refln.index_l**
• **_phasing_MIR_der_refln.set_id**
  → **_phasing_set.id**
  _phasing_MIR_der_refln.F_calc
  _phasing_MIR_der_refln.F_calc_au
  _phasing_MIR_der_refln.F_meas
  _phasing_MIR_der_refln.F_meas_au
  _phasing_MIR_der_refln.F_meas_sigma
  _phasing_MIR_der_refln.F_meas_sigma_au
  _phasing_MIR_der_refln.HL_A_iso
  _phasing_MIR_der_refln.HL_B_iso
  _phasing_MIR_der_refln.HL_C_iso

```
  _phasing_MIR_der_refln.HL_D_iso
  _phasing_MIR_der_refln.phase_calc
```

(*e*) PHASING_MIR_DER_SHELL

```
• _phasing_MIR_der_shell.d_res_high
• _phasing_MIR_der_shell.d_res_low
• _phasing_MIR_der_shell.der_id
        → _phasing_MIR_der.id
  _phasing_MIR_der_shell.fom
  _phasing_MIR_der_shell.ha_ampl
  _phasing_MIR_der_shell.loc
  _phasing_MIR_der_shell.phase
  _phasing_MIR_der_shell.power
  _phasing_MIR_der_shell.R_cullis
  _phasing_MIR_der_shell.R_kraut
  _phasing_MIR_der_shell.reflns
```

(*f*) PHASING_MIR_DER_SITE

```
• _phasing_MIR_der_site.der_id
        → _phasing_MIR_der.id
• _phasing_MIR_der_site.id
  _phasing_MIR_der_site.atom_type_symbol
        → _atom_type.symbol
+ _phasing_MIR_der_site.B_iso
+ _phasing_MIR_der_site.Cartn_x
+ _phasing_MIR_der_site.Cartn_y
+ _phasing_MIR_der_site.Cartn_z
  _phasing_MIR_der_site.details
+ _phasing_MIR_der_site.fract_x
+ _phasing_MIR_der_site.fract_y
+ _phasing_MIR_der_site.fract_z
  _phasing_MIR_der_site.occupancy
  _phasing_MIR_der_site.occupancy_anom
  _phasing_MIR_der_site.occupancy_anom_su
  _phasing_MIR_der_site.occupancy_iso
  _phasing_MIR_der_site.occupancy_iso_su
```

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. 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.*

PHASING_MIR and related categories provide information about phasing by methods involving multiple isomorphous replacement (MIR). These same categories may also be used to describe phasing by related techniques, such as single isomorphous replacement (SIR) and single or multiple isomorphous replacement plus anomalous scattering (SIRAS, MIRAS). The relationships between the categories describing MIR phasing are shown in Fig. 3.6.6.2.

As with the other overview categories described in this section, the PHASING_MIR category contains data items that can be used for text-based descriptions of the method used and any special aspects of its application. There are also items for describing the resolution limit of the reflections that were phased, the figures of merit for all reflections and for the acentric reflections phased in the native data set, and the total numbers of reflections and their inclusion threshold in the native data set. Statistics for the phasing can be given by shells of resolution using data items in the PHASING_MIR_SHELL category.

An MIR phasing experiment involves one or more derivatives. The remaining categories in this group are used to describe aspects of each derivative (Example 3.6.6.5). A derivative in this context does not necessarily correspond to a data set; for instance, the same data set could be used to one resolution limit as an isomorphous scatterer and to a different resolution (and with a different sigma cutoff) as an anomalous scatterer. These would be treated as two distinct derivatives, although both derivatives would point to the same data sets *via* **_phasing_MIR_der.der_set_id** and **_phasing_MIR_der.native_set_id** (see Fig. 3.6.6.2).

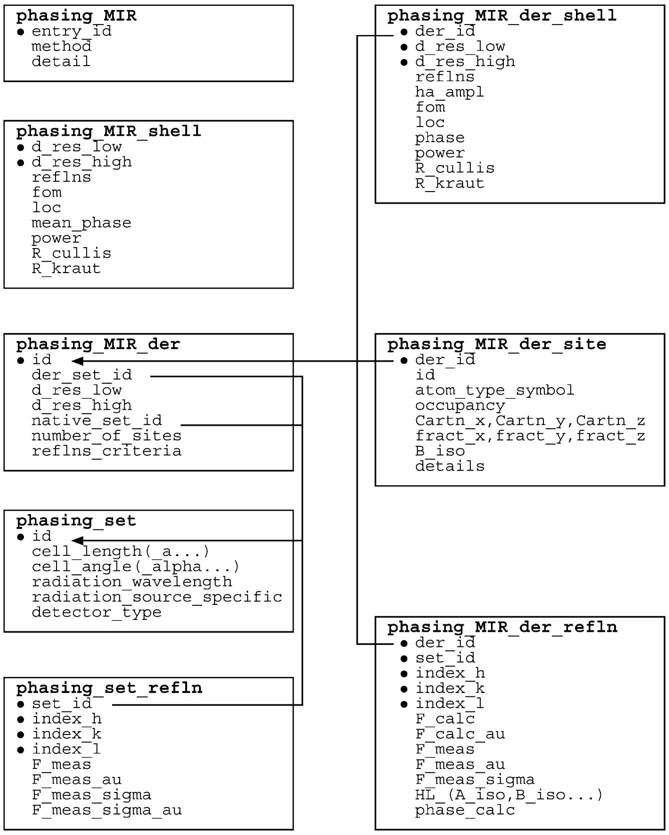

Fig. 3.6.6.2. The family of categories used to describe MIR phasing. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

Data items in the PHASING_MIR_DER category can be used to identify and describe each derivative. The resolution limits for the individual derivatives need not match those of the overall phasing experiment, as the phasing power of each derivative as a function of resolution will vary. Many of the statistical descriptors of phasing given in the PHASING_MIR category are repeated in this category, as derivatives vary in quality and their contribution to the phasing must be assessed individually. These same statistical measures can be given for shells of resolution in the PHASING_MIR_DER_SHELL category.

Data items in the PHASING_MIR_DER_REFLN category can be used to provide details of each reflection used in an MIR phasing experiment. The pointer **_phasing_MIR_der_refln.set_id** links the reflection to a particular set of experimental data and **_phasing_MIR_der_refln.der_id** points to a particular derivative used in the phasing (as mentioned above, derivatives in this context do not equate to data sets). The phase assigned to each reflection and the measured and calculated values of its structure factor can be given. (It is not necessary to include the measured values of the structure factors in this list, since they are accessible in the PHASING_SET_REFLN category, but it may be convenient to present them here). Data items are also provided for the *A*, *B*, *C* and *D* phasing coefficients of Hendrickson & Lattman (1970).

The heavy atoms identified in each derivative can be listed using data items in the PHASING_MIR_DER_SITE category. Most of the data names are clear analogues of similar items in the ATOM_SITE category; an exception is **_phasing_MIR_der_site.occupancy_anom**, which specifies the relative anomalous occupancy of the atom type present at a heavy-atom site in a particular derivative.

Example 3.6.6.5. *Phasing of the structure of bovine plasma retinol-binding protein (Zanotti et al., 1993) described using data items in the* PHASING_MIR *and related categories.*

```
_phasing_MIR.entry_id          '1HBP'
_phasing_MIR.method
;  Standard phase refinement (Blow & Crick, 1959)
;

loop_
_phasing_MIR_shell.d_res_low
_phasing_MIR_shell.d_res_high
_phasing_MIR_shell.reflns
_phasing_MIR_shell.FOM
 15.0  8.3   80  0.69       8.3  6.4  184  0.73
  6.4  5.2  288  0.72       5.2  4.4  406  0.65
  4.4  3.8  554  0.54       3.8  3.4  730  0.53
  3.4  3.0  939  0.50

loop_
_phasing_MIR_der.id
_phasing_MIR_der.number_of_sites
_phasing_MIR_der.details
KAu(CN)2  3
     'major site interpreted in difference Patterson'
K2HgI4    6 'sites found in cross-difference Fourier'
K3IrCl6   2 'sites found in cross-difference Fourier'
All      11 'data for all three derivatives combined'

loop_
_phasing_MIR_der_shell.der_id
_phasing_MIR_der_shell.d_res_low
_phasing_MIR_der_shell.d_res_high
_phasing_MIR_der_shell.ha_ampl
_phasing_MIR_der_shell.loc
   KAu(CN)2  15.0  8.3   54   26
   KAu(CN)2   8.3  6.4   54   20
# - - - abbreviated - - -
   K2HgI4    15.0  8.3  149   87
   K2HgI4     8.3  6.4  121   73
# - - - abbreviated - - -
   K3IrCl6   15.0  8.3   33   27
   K3IrCl6    8.3  6.4   40   23
# - - - abbreviated - - -

loop_
_phasing_MIR_der_site.der_id
_phasing_MIR_der_site.id
_phasing_MIR_der_site.atom_type_symbol
_phasing_MIR_der_site.occupancy
_phasing_MIR_der_site.fract_x
_phasing_MIR_der_site.fract_y
_phasing_MIR_der_site.fract_z
_phasing_MIR_der_site.B_iso
   KAu(CN)2  1  Au  0.40  0.082  0.266  0.615  33.0
   KAu(CN)2  2  Au  0.03  0.607  0.217  0.816  25.9
   K2HgI4    1  Hg  0.63  0.048  0.286  0.636  33.7
   K2HgI4    2  Hg  0.34  0.913  0.768  0.889  36.7
# - - - abbreviated - - -

_phasing_MIR_der_refln.index_h           6
_phasing_MIR_der_refln.index_k           1
_phasing_MIR_der_refln.index_l          25
_phasing_MIR_der_refln.der_id        HGPT1
_phasing_MIR_der_refln.set_id      'NS1-96'
_phasing_MIR_der_refln.F_calc_au    106.66
_phasing_MIR_der_refln.F_meas_au    204.67
_phasing_MIR_der_refln.F_meas_sigma   6.21
_phasing_MIR_der_refln.HL_A_iso      -3.15
_phasing_MIR_der_refln.HL_B_iso      -0.76
_phasing_MIR_der_refln.HL_C_iso       0.65
_phasing_MIR_der_refln.HL_D_iso       0.23
_phasing_MIR_der_refln.phase_calc   194.48
```

3.6.6.1.6. *Phasing data sets*

The data items in these categories are as follows:

(*a*) PHASING_SET

```
• _phasing_set.id
  _phasing_set.cell_angle_alpha
  _phasing_set.cell_angle_beta
  _phasing_set.cell_angle_gamma
  _phasing_set.cell_length_a
  _phasing_set.cell_length_b
  _phasing_set.cell_length_c
  _phasing_set.detector_specific
  _phasing_set.detector_type
  _phasing_set.radiation_source_specific
  _phasing_set.radiation_wavelength
  _phasing_set.temp
```

(*b*) PHASING_SET_REFLN

```
• _phasing_set_refln.index_h
• _phasing_set_refln.index_k
• _phasing_set_refln.index_l
• _phasing_set_refln.set_id
         → _phasing_set.id
  _phasing_set_refln.F_meas
  _phasing_set_refln.F_meas_au
  _phasing_set_refln.F_meas_sigma
  _phasing_set_refln.F_meas_sigma_au
```

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.

Data items in the PHASING_SET family of categories are homologous to items with related names in the CELL and DIFFRN families of categories. The PHASING_SET categories were added to the mmCIF data model so that intensity and phase information for the data sets used in phasing could be stored in the same data block as the information for the refined structure. It is not necessary to store all the experimental information for each data set (*e.g.* the raw data sets or crystal growth conditions); it is assumed that the full experimental description of each phasing set would be recorded in a separate data block (see Example 3.6.6.6).

Data items in the PHASING_SET category identify each set of diffraction data used in a phasing experiment and can be used to summarize relevant experimental conditions. Because a given data set may be used in a number of different ways (for example, as an isomorphous derivative and as a component of a multiple-wavelength calculation), it is appropriate to store the reflections in a category distinct from either the PHASING_MAD or PHASING_MIR family of categories, but accessible to both these families (and any similar categories that might be introduced later to describe new phasing methods). Figs. 3.6.6.1 and 3.6.6.2 show how reference is made to the relevant sets from within the PHASING_MAD and PHASING_MIR categories.

Each phasing set is given a unique value of `_phasing_set.id`. The other PHASING_SET data items record the cell dimensions and

Example 3.6.6.6. *The phasing sets used in the structure determination of bovine plasma retinol-binding protein (Zanotti et al., 1993) described with data items in the* PHASING_SET *and* PHASING_SET_REFLN *categories.*

```
_phasing_set.id                     'NS1-96'
_phasing_set.cell_angle_alpha       90.0
_phasing_set.cell_angle_beta        90.0
_phasing_set.cell_angle_gamma       90.0
_phasing_set.cell_length_a          38.63
_phasing_set.cell_length_b          38.63
_phasing_set.cell_length_c          82.88
_phasing_set.radiation_wavelength   1.5145
_phasing_set.detector_type          'image plate'
_phasing_set.detector_specific      'RXII'

_loop
_phasing_set_refln.set_id
_phasing_set_refln.index_h
_phasing_set_refln.index_k
_phasing_set_refln.index_l
_phasing_set_refln.F_meas_au
_phasing_set_refln.F_meas_sigma_au
  'NS1-96'  15  15  32  181.79  3.72
  'NS1-96'  15  15  33   34.23  1.62
# - - - abbreviated - - -
```

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 independently 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_{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_{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_{free} and for the complementary R_{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_{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_{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_{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

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

Example 3.6.6.13. *The data set used in the refinement of an HIV-1 protease structure (PDB 5HVP) described using data items in the* REFLNS *and* REFLNS_SHELL *categories.*

```
_reflns.entry_id                        '5HVP'
_reflns.data_reduction_method
; Xengen program scalei. Anomalous pairs were merged.
  Scaling proceeded in several passes, beginning with
  1-parameter fit and ending with 3-parameter fit.
;
_reflns.data_reduction_details
; Merging and scaling based on only those reflections
  with I > sigma(I).
;

_reflns.d_resolution_high               2.00
_reflns.d_resolution_low                8.00

_reflns.limit_h_max                     22
_reflns.limit_h_min                     0
_reflns.limit_k_max                     46
_reflns.limit_k_min                     0
_reflns.limit_l_max                     57
_reflns.limit_l_min                     0

_reflns.number_obs                      7228
_reflns.observed_criterion_sigma_I      1.0
_reflns.details                         none

loop_
_reflns_shell.d_res_high
_reflns_shell.d_res_low
_reflns_shell.meanI_over_sigI_obs
_reflns_shell.number_measured_obs
_reflns_shell.number_unique_obs
_reflns_shell.percent_possible_obs
_reflns_shell.Rmerge_F_obs
   31.38  3.82  69.8  9024  2540  96.8   1.98
    3.82  3.03  26.1  7413  2364  95.1   3.85
    3.03  2.65  10.5  5640  2123  86.2   6.37
    2.65  2.41   6.4  4322  1882  76.8   8.01
    2.41  2.23   4.3  3247  1714  70.4   9.86
    2.23  2.10   3.1  1140   812  33.3  13.99
```

a reflection as being observed are given using the data item `_reflns.observed_criterion`. This is a free-text field so is not automatically parsable. Therefore it is supplemented in the mmCIF dictionary by data items that can be used to stipulate the criterion in terms of the values of F, I or the uncertainties in these quantities (Example 3.6.6.13). The percentage of the total number of reflections that meet the criterion can be recorded.

Data items are also provided for describing the selection of the reflections used to calculate the free R factor, and for giving the $R_{\rm merge}$ values for all reflections and for the subset of 'observed' reflections. Data items in the REFLNS_SCALE and REFLNS_SHELL categories are used in the same way in the mmCIF and core CIF dictionaries, and Section 3.2.3.2.2 can be consulted for details.

As with the related categories DIFFRN_REFLNS_CLASS and REFINE_LS_CLASS, the core dictionary category REFLNS_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.

3.6.7. Atomicity, chemistry and structure

The basic concepts of the mmCIF model for describing a macromolecular structure were outlined in Section 3.6.3. The present section describes the components of the model in more detail. The category groups used to describe the molecular chemistry and structure are: the ATOM group describing atom positions (Section 3.6.7.1); the CHEMICAL, CHEM_COMP and CHEM_LINK groups describing molecular chemistry (Section 3.6.7.2); the ENTITY group describing distinct chemical species (Section 3.6.7.3); the GEOM group describing molecular or packing geometry (Section 3.6.7.4); the STRUCT group describing the large-scale features of molecular structure (Section 3.6.7.5); and the SYMMETRY group describing the symmetry and space group (Section 3.6.7.6).

The CHEMICAL category group itself is not generally used in an mmCIF. The purpose of this category group in the core CIF dictionary is to specify the chemical identity and connectivity of the relatively simple molecular or ionic species in a small-molecule or inorganic crystal. In principle, a macromolecular structure determined to atomic resolution could be represented as a coherent chemical entity with a complete connectivity graph. However, in practice, biological macromolecules are built from units from a library of models of standard amino acids, nucleotides and sugars. Data items in the CHEM_COMP and CHEM_LINK category groups of the mmCIF dictionary describe the internal connectivity and standard bonding processes between these units.

Molecular or packing geometry is also rarely tabulated for large macromolecular complexes, so the GEOM category group is rarely used in an mmCIF.

3.6.7.1. Atom sites

The categories describing atom sites are as follows:

ATOM group
Individual atom sites (§3.6.7.1.1)
ATOM_SITE
ATOM_SITE_ANISOTROP
Collections of atom sites (§3.6.7.1.2)
ATOM_SITES
ATOM_SITES_FOOTNOTE
Atom types (§3.6.7.1.3)
ATOM_TYPE
Alternative conformations (§3.6.7.1.4)
ATOM_SITES_ALT
ATOM_SITES_ALT_ENS
ATOM_SITES_ALT_GEN

The ATOM category group represents a compromise between the representation of a small-molecule structure as an annotated list of atomic coordinates and the need in macromolecular crystallography to present a more structured view organized around residues, chains, sheets, turns, helices *etc.* The locations of individual atoms and other information about the atom sites are given using data items in this category group. The categories within the group may be classified as shown in the summary above.

The ATOM_SITE, ATOM_SITES and ATOM_TYPE categories have many data items that are aliases of equivalent data items in the same categories in the core CIF dictionary, but the conventions for the labelling of the atom sites are different.

The ATOM_SITE_ANISOTROP and ATOM_SITES_FOOTNOTE categories are new to the mmCIF dictionary, as are the categories related to alternative conformations: ATOM_SITES_ALT, ATOM_SITES_ALT_ENS and ATOM_SITES_ALT_GEN.

3.6.7.1.1. *Individual atom sites*

The data items in these categories are as follows:

(*a*) ATOM_SITE
• **`_atom_site.id`** (∼ ***`_atom_site_label`***)
  `_atom_site.adp_type`
+ `_atom_site.aniso_B[1][1]`
        ⇌ ***`_atom_site_anisotrop.B[1][1]`***