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).

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 3.6.6.1. *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'
```

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.

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.

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
```

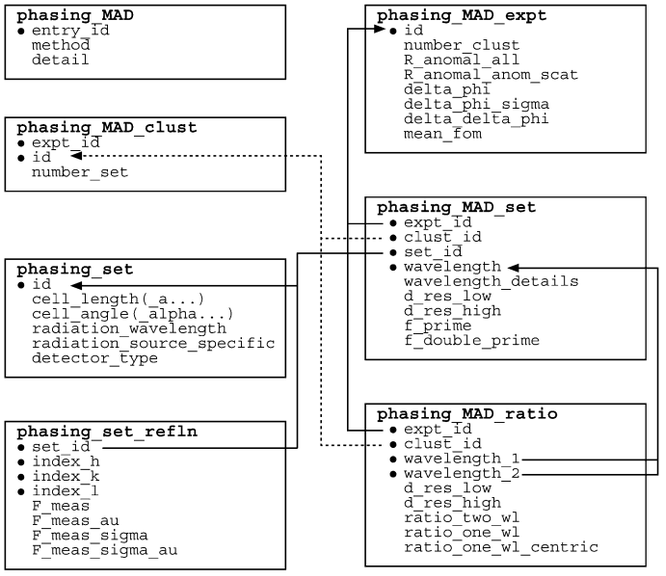

Fig. 3.6.6.1. The family of categories used to describe MAD phasing. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data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

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.

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).

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 indepen-