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.

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
```

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.