

3.6. CLASSIFICATION AND USE OF MACROMOLECULAR DATA

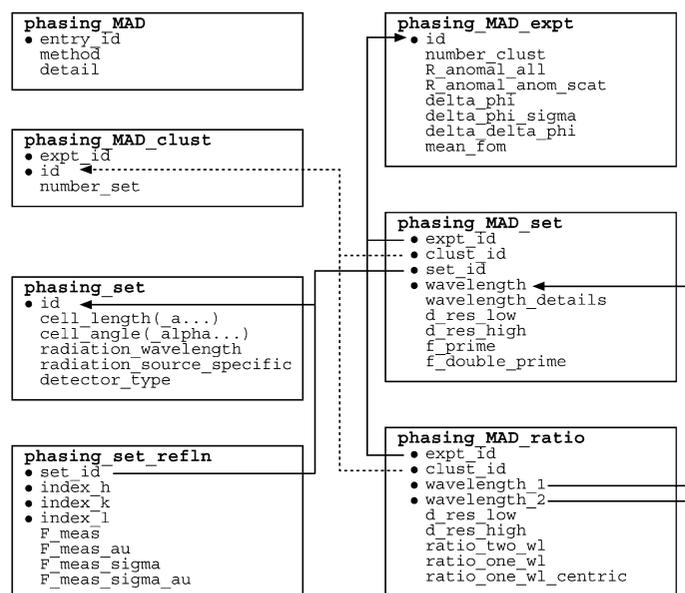


Fig. 3.6.6.1. The family of categories used to describe MAD phasing. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data 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_refl.der_id`
→ `_phasing_MIR_der.id`
- `_phasing_MIR_der_refl.index_h`
- `_phasing_MIR_der_refl.index_k`
- `_phasing_MIR_der_refl.index_l`
- `_phasing_MIR_der_refl.set_id`
→ `_phasing_set.id`
- `_phasing_MIR_der_refl.F_calc`
- `_phasing_MIR_der_refl.F_calc_au`
- `_phasing_MIR_der_refl.F_meas`
- `_phasing_MIR_der_refl.F_meas_au`
- `_phasing_MIR_der_refl.F_meas_sigma`
- `_phasing_MIR_der_refl.F_meas_sigma_au`
- `_phasing_MIR_der_refl.HL_A_iso`
- `_phasing_MIR_der_refl.HL_B_iso`
- `_phasing_MIR_der_refl.HL_C_iso`

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_refl.index_h 6
  _phasing_MIR_der_refl.index_k 1
  _phasing_MIR_der_refl.index_l 25
  _phasing_MIR_der_refl.der_id HGPT1
  _phasing_MIR_der_refl.set_id 'NS1-96'
  _phasing_MIR_der_refl.F_calc_au 106.66
  _phasing_MIR_der_refl.F_meas_au 204.67
  _phasing_MIR_der_refl.F_meas_sigma 6.21
  _phasing_MIR_der_refl.HL_A iso -3.15
  _phasing_MIR_der_refl.HL_B iso -0.76
  _phasing_MIR_der_refl.HL_C iso 0.65
  _phasing_MIR_der_refl.HL_D iso 0.23
  _phasing_MIR_der_refl.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_refl.index_h`
- `_phasing_set_refl.index_k`
- `_phasing_set_refl.index_l`
- `_phasing_set_refl.set_id`
→ `_phasing_set.id`
- `_phasing_set_refl.F_meas`
- `_phasing_set_refl.F_meas_au`
- `_phasing_set_refl.F_meas_sigma`
- `_phasing_set_refl.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_refl.set_id
  _phasing_set_refl.index_h
  _phasing_set_refl.index_k
  _phasing_set_refl.index_l
  _phasing_set_refl.F_meas_au
  _phasing_set_refl.F_meas_sigma_au
  'NS1-96' 15 15 32 181.79 3.72
  'NS1-96' 15 15 33 34.23 1.62
# - - - abbreviated - - -

```