

3. CIF DATA DEFINITION AND CLASSIFICATION

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_scst

(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

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_scst
  _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
```

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.

3.6. CLASSIFICATION AND USE OF MACROMOLECULAR DATA

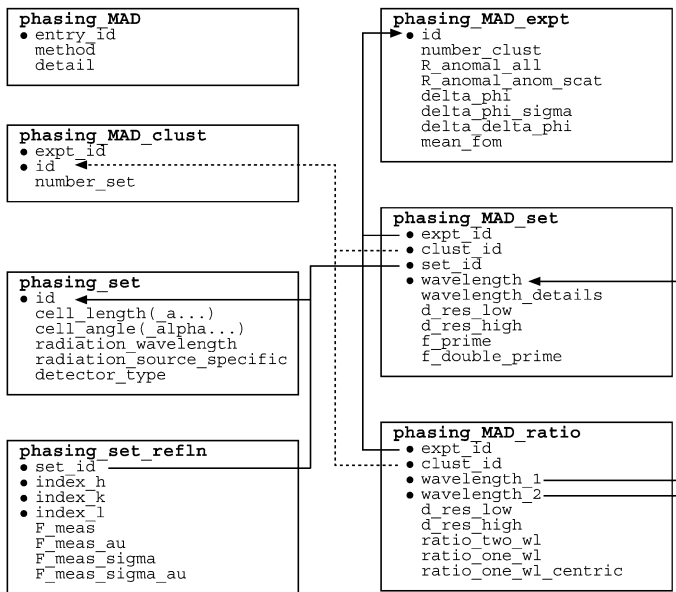


Fig. 3.6.6.1. The family of categories used to describe MAD phasing. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data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