3.6. CLASSIFICATION AND USE OF MACROMOLECULAR DATA

Example 3.6.5.2. *Data collection for an HIV-1 protease crystal (PDB 5HVP) described with data items in the DIFFRN and related categories.*

```
_diffrn.id                            'set1'
_diffrn.crystal_id                    1
_diffrn.ambient_temp                  293(3)
_diffrn.ambient_environment
; Mother liquor from the reservoir of the vapor
  diffusion experiment, mounted in room air
;
_diffrn.crystal_support
;  0.7 mm glass capillary, sealed with dental wax
;
_diffrn.crystal_treatment
; Equilibrated in rotating anode radiation enclosure
  for 18 hours prior to beginning of data collection.
;
_diffrn_detector.diffrn_id            'set1'
_diffrn_detector.detector             'multiwire'
_diffrn_detector.type                 'Siemens'

_diffrn_measurement.diffrn_id         'd1'
_diffrn_measurement.device          '3-circle camera'
_diffrn_measurement.device_type     'Supper model x'
_diffrn_measurement.device_details    'none'
_diffrn_measurement.method            'omega scan'
_diffrn_measurement.details
; 440 frames, 0.20 degrees, 150 sec, detector
  distance 12 cm, detector angle 22.5 degrees
;
_diffrn_radiation.diffrn_id           'set1'
_diffrn_radiation.collimation
                              '0.3 mm double pinhole'
_diffrn_radiation.monochromator       'graphite'
_diffrn_radiation.type                'Cu Kalpha'
_diffrn_radiation.wavelength_id       1
_diffrn_radiation_wavelength.id       1
_diffrn_radiation_wavelength.wavelength 1.54
_diffrn_radiation_wavelength.wt       1.0
_diffrn_source.diffrn_id              'set1'
_diffrn_source.source                 'rotating anode'
_diffrn_source.type                   'Rigaku RU-200'
_diffrn_source.power                  50
_diffrn_source.current                180
_diffrn_source.target       '8mm x 0.4 mm broad-focus'
```

the DIFFRN categories. These differences were introduced in order to accommodate data from more than one experiment in the same table. For example, in the core CIF dictionary, the Miller indices `_diffrn_refln_index_h`, `*_k` and `*_l` play the role of the category key for the DIFFRN_REFLN category. In the mmCIF dictionary, the category key is formed by the data items `_diffrn_refln.id` and `_diffrn_refln.diffrn_id`.

3.6.5.3. Growth, description and analysis of the crystal

The categories describing the crystal properties and growth are as follows:

EXPTL group
Crystal properties (§3.6.5.3.1)
    EXPTL
    EXPTL_CRYSTAL
    EXPTL_CRYSTAL_FACE
Crystal growth (§3.6.5.3.2)
    EXPTL_CRYSTAL_GROW
    EXPTL_CRYSTAL_GROW_COMP

Categories in the EXPTL category group are used to describe experimental measurements on the crystal (*e.g.* of its shape, size and density) and the growth of the crystal. Data items in the EXPTL category are used to describe the gross properties of the crystal or crystals used in the experiment. Data items in the EXPTL_CRYSTAL

category are used to describe the crystal properties in detail and allow for cases where multiple crystals are used. The data items in the EXPTL_CRYSTAL_FACE category are used to describe the crystal faces.

Data items for describing crystal growth are given in two categories that are not found in the current version of the core CIF dictionary. Data items in the EXPTL_CRYSTAL_GROW category are used to describe the conditions and methods used to grow the crystals, and data items in the EXPTL_CRYSTAL_GROW_COMP category can be used to list the components of the solutions in which the crystals were grown.

3.6.5.3.1. *Crystal properties*

The data items in these categories are as follows:

(*a*) EXPTL
- `_exptl.entry_id`
    → `_entry.id`
  `_exptl.absorpt_coefficient_mu`
  `_exptl.absorpt_correction_T_max`
  `_exptl.absorpt_correction_T_min`
  `_exptl.absorpt_correction_type`
  `_exptl.absorpt_process_details`
  `_exptl.crystals_number`
  `_exptl.details` (∼ `_exptl_special_details`)
  `_exptl.method`
  `_exptl.method_details`

(*b*) EXPTL_CRYSTAL
- `_exptl_crystal.id`
  `_exptl_crystal.colour`
  `_exptl_crystal.colour_lustre`
  `_exptl_crystal.colour_modifier`
  `_exptl_crystal.colour_primary`
  `_exptl_crystal.density_diffrn`
  `_exptl_crystal.density_Matthews`
+ `_exptl_crystal.density_meas`
  `_exptl_crystal.density_meas_gt`
  `_exptl_crystal.density_meas_lt`
+ `_exptl_crystal.density_meas_temp`
  `_exptl_crystal.density_meas_temp_gt`
  `_exptl_crystal.density_meas_temp_lt`
  `_exptl_crystal.density_method`
  `_exptl_crystal.density_percent_sol`
  `_exptl_crystal.description`
  `_exptl_crystal.F_000`
  `_exptl_crystal.preparation`
  `_exptl_crystal.size_max`
  `_exptl_crystal.size_mid`
  `_exptl_crystal.size_min`
  `_exptl_crystal.size_rad`

(*c*) EXPTL_CRYSTAL_FACE
- `_exptl_crystal_face.crystal_id`
    → `_exptl_crystal.id`
- `_exptl_crystal_face.index_h`
- `_exptl_crystal_face.index_k`
- `_exptl_crystal_face.index_l`
  `_exptl_crystal_face.diffr_chi`
  `_exptl_crystal_face.diffr_kappa`
  `_exptl_crystal_face.diffr_phi`
  `_exptl_crystal_face.diffr_psi`
  `_exptl_crystal_face.perp_dist`

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 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.

Data items in these categories are used in the same way in the mmCIF and core CIF dictionaries, and Section 3.2.2.3 can be consulted for details (see Example 3.6.5.3). Identifiers have been introduced to the categories to provide the formal category keys required by the DDL2 data model.

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