Example 3.6.6.9. *Results of the refinement of an HIV-1 protease structure (PDB 5HVP) described with data items in the* REFINE_LS_RESTR *and* REFINE_LS_SHELL *categories.*

```
loop_
_refine_ls_restr.type
_refine_ls_restr.dev_ideal_target
_refine_ls_restr.dev_ideal
_refine_ls_restr.number
_refine_ls_restr.criterion
_refine_ls_restr.rejects
'p_bond_d'          0.020  0.018  1654 '>2 sigma' 22
'p_angle_d'         0.030  0.038  2246 '>2 sigma' 139
'p_planar_d'        0.040  0.043  498  '>2 sigma' 21
'p_planar'          0.020  0.015  270  '>2 sigma' 1
'p_chiral'          0.150  0.177  278  '>2 sigma' 2
'p_singtor_nbd'     0.500  0.216  582  '>2 sigma' 0
'p_multtor_nbd'     0.500  0.207  419  '>2 sigma' 0
'p_xyhbond_nbd'     0.500  0.245  149  '>2 sigma' 0
'p_planar_tor'      3.0    2.6    203  '>2 sigma' 9
'p_staggered_tor'   15.0   17.4   298  '>2 sigma' 31
'p_orthonormal_tor' 20.0   18.1   12   '>2 sigma' 1

loop_
_refine_ls_shell.d_res_low
_refine_ls_shell.d_res_high
_refine_ls_shell.number_reflns_obs
_refine_ls_shell.R_factor_obs
   8.00   4.51   1226   0.196
   4.51   3.48   1679   0.146
   3.48   2.94   2014   0.160
   2.94   2.59   2147   0.182
   2.59   2.34   2127   0.193
   2.34   2.15   2061   0.203
   2.15   2.00   1647   0.188
```

parameters that lie too far from a target, as these data are often published as part of a description of the refinement and are often deposited with the coordinates in an archive. However, the types of restraints applied depend strongly on the software package used, and as new refinement packages regularly become available, it was clearly not advisable to provide program-specific data items in the mmCIF dictionary. The approach taken in the mmCIF dictionary has been to allow the value of `_refine_ls_restr.type` to be a free-text field, so that arbitrary labels can be given to restraints that are particular to a software package, but to recommend the use of specific labels for restraints applied by particular programs. The dictionary provides examples for labels specific to the programs *PROTIN/PROLSQ* (Hendrickson & Konnert, 1979) and *RESTRAIN* (Driessen *et al.*, 1989). These program-specific representations have particular prefixes; thus the value `p_bond_d` is a bond-distance restraint as applied by *PROTIN/PROLSQ*. Values for `_refine_ls_restr.type` appropriate for other refinement programs may be suggested in future versions of the mmCIF dictionary.

Data items in the REFINE_LS_RESTR_TYPE category can be used to specify the ranges within which quantities are allowed to vary for each type of restraint. The special value indicated by a full stop (.) represents a restraint unbounded on the high or low side.

Data items in the REFINE_LS_RESTR_NCS category can be used to record details about the restraints applied to atom positions in domains related by noncrystallographic symmetry during least-squares refinement, and also to record the deviation of the restrained atomic parameters at the end of the refinement. The domains related by noncrystallographic symmetry are defined in the STRUCT_NCS_DOM and related categories (see Section 3.6.7.5.5). The quantities that can be recorded for each restrained domain are the root-mean-square deviations of the displacement and positional parameters, and the weighting coefficients used in the noncrystallographic restraint of each type of parameter. Any special aspects of the way the restraints were applied may be described using `_refine_ls_restr_ncs.ncs_model_details`.

Data items in the REFINE_LS_SHELL category are used to summarize details of the results of the least-squares refinement by shells of resolution (Example 3.6.6.9). The resolution range, in ångströms, forms the category key; for each shell the quantities reported, such as the number of reflections above the threshold for counting as significantly intense, are all defined in the same way as the corresponding data items used to describe the results of the overall refinement in the REFINE category.

The core dictionary category REFINE_LS_CLASS was introduced after the release of the first version of the mmCIF dictionary. It provides a more general way of describing the treatment of particular subsets of the observations, but it is not expected to be used in macromolecular structural studies, where partition by shells of resolution is traditional.

### 3.6.6.2.4. *Equivalent atoms in the refinement*

The data items in these categories are as follows:

(*a*) REFINE_B_ISO

- `_refine_B_iso.class`
  `_refine_B_iso.details`
  `_refine_B_iso.treatment`
  `_refine_B_iso.value`

(*b*) REFINE_OCCUPANCY

- `_refine_occupancy.class`
  `_refine_occupancy.details`
  `_refine_occupancy.treatment`
  `_refine_occupancy.value`

*The bullet* (•) *indicates a category key.*

In macromolecular structure refinement, displacement factors or occupancies are often treated as equivalent for groups of atoms. An example would be the case where most of the atoms in the structure are refined with isotropic displacement factors, but a bound metal atom is allowed to refine anisotropically. Another example would be where the occupancies for all of the atoms in the protein part of a macromolecular complex are fixed at 1.0, but the occupancies of atoms in a bound inhibitor are refined. The REFINE_B_ISO and REFINE_OCCUPANCY categories can be used to record this information (Example 3.6.6.10).

Example 3.6.6.10. *The handling of displacement factors and occupancies during the refinement of an HIV-1 protease structure (PDB 5HVP) described with data items in the* REFINE_B_ISO *and* REFINE_OCCUPANCY *categories.*

```
loop_
_refine_B_iso.class
_refine_B_iso.treatment
   'protein'    isotropic
   'solvent'    isotropic
   'inhibitor'  isotropic

loop_
_refine_occupancy.class
_refine_occupancy.treatment
_refine_occupancy.value
_refine_occupancy.details
   'protein'                   fix  1.00  .
   'solvent'                   fix  1.00  .
   'inhibitor orientation 1'   fix  0.65  .
   'inhibitor orientation 2'   fix  0.35
; The inhibitor binds to the enzyme in two
  alternative conformations. The occupancy of
  each conformation was adjusted so as to result
  in approximately equal mean thermal factors
  for the atoms in each conformation.
;
```

Example 3.6.6.11. *An example of one cycle of refinement described with data items in the* REFINE_HIST *category.*

```
_refine_hist.cycle_id                 C134
_refine_hist.d_res_high               1.85
_refine_hist.d_res_low                20.0
_refine_hist.number_atoms_solvent      217
_refine_hist.number_atoms_total        808
_refine_hist.number_reflns_all        6174
_refine_hist.number_reflns_obs        4886
_refine_hist.number_reflns_R_free      476
_refine_hist.number_reflns_R_work     4410
_refine_hist.R_factor_all             .265
_refine_hist.R_factor_obs             .195
_refine_hist.R_factor_R_free          .274
_refine_hist.R_factor_R_work          .160
_refine_hist.details
; Add majority of solvent molecules. B factors
  refined by group. Continued to remove
  misplaced water molecules.
;
```

Data items in the REFINE_B_ISO category can be used to record details of the treatment of isotropic *B* (displacement) factors during refinement. There is no formal link between the classes identified by `_refine_B_iso.class` and individual atom sites, although relationships may be inferred if the class names are carefully chosen. The category allows the treatment of the atoms in each class (isotropic, anisotropic or fixed) and the value assigned for fixed isotropic *B* factors to be recorded. Any special details can be given in a free-text field.

Data items in the REFINE_OCCUPANCY category can be used to record details of the treatment of occupancies of groups of atom sites during refinement. As with the treatment of displacement factors in the REFINE_B_ISO category, the classes itemized by `_refine_occupancy.class` are not formally linked to the individual atom sites, but the relationships may be deduced if the class names are chosen carefully.

#### 3.6.6.2.5. *History of the refinement*

The data items in this category are as follows:

REFINE_HIST

- `_refine_hist.cycle_id`

`_refine_hist.details`
`_refine_hist.d_res_high`
`_refine_hist.d_res_low`
`_refine_hist.number_atoms_solvent`
`_refine_hist.number_atoms_total`
`_refine_hist.number_reflns_all`
`_refine_hist.number_reflns_obs`
`_refine_hist.number_reflns_R_free`
`_refine_hist.number_reflns_R_work`
`_refine_hist.R_factor_all`
`_refine_hist.R_factor_obs`
`_refine_hist.R_factor_R_free`
`_refine_hist.R_factor_R_work`

*The bullet* (•) *indicates a category key.*

Data items in the REFINE_HIST category can be used to record details about the various steps in the refinement of the structure. They do not provide as thorough a description of the refinement as can be given in other categories for the final model, but instead allow a summary of the progress of the refinement to be given and supported by a small set of representative statistics.

The category is sufficiently compact that a large number of cycles could be summarized, but it is not expected that every cycle of refinement would be routinely reported. Example 3.6.6.11 shows an entry for a single cycle of refinement. It is likely that an author would present a representative sequence of entries in a looped list.

### 3.6.6.3. Reflection measurements

The categories describing the reflections used in the refinement are as follows:

REFLN group
*Individual reflections* (§3.6.6.3.1)
REFLN
REFLN_SYS_ABS
*Groups of reflections* (§3.6.6.3.2)
REFLNS
REFLNS_SCALE
REFLNS_SHELL
REFLNS_CLASS

Data items in the REFLN category can be used to give information about the individual reflections that were used to derive the final model. The related category REFLN_SYS_ABS allows the reflections that should be systematically absent for the space group in which the structure was refined to be tabulated. Data items in the REFLNS category can be used to record information that applies to all of the reflections. Scale factors can be listed in the REFLNS_SCALE category, while the data items in REFLNS_SHELL can be used to record information about the reflection set by shells of resolution. The core CIF dictionary category REFLNS_CLASS, which can be used for a general classification of reflection groups according to criteria other than resolution shell, is not expected to be used in mmCIF applications.

#### 3.6.6.3.1. *Individual reflections*

The data items in these categories are as follows:

(*a*) REFLN

- `_refln.index_h`
- `_refln.index_k`
- `_refln.index_l`

`_refln.A_calc`
`_refln.A_calc_au`
`_refln.A_meas`
`_refln.A_meas_au`
`_refln.B_calc`
`_refln.B_calc_au`
`_refln.B_meas`
`_refln.B_meas_au`
`_refln.class_code`
`_refln.crystal_id`
→ `_exptl_crystal.id`
`_refln.d_spacing`
`_refln.F_calc`
`_refln.F_calc_au`
`_refln.F_meas`
`_refln.F_meas_au`
`_refln.F_meas_sigma` (∼ `_refln_F_sigma`)
`_refln.F_meas_sigma_au`
`_refln.F_squared_calc`
`_refln.F_squared_meas`
`_refln.F_squared_sigma`
`_refln.fom`
`_refln.include_status`
`_refln.intensity_calc`
`_refln.intensity_meas`
`_refln.intensity_sigma`
`_refln.mean_path_length_tbar`
`_refln.phase_calc`
`_refln.phase_meas`
`_refln.refinement_status`
`_refln.scale_group_code`
→ `_reflns_scale.group_code`
`_refln.sint_over_lambda` (∼ `_refln_sint/lambda`)
`_refln.status` (∼ `_refln_observed_status`)
`_refln.symmetry_epsilon`
`_refln.symmetry_multiplicity`
`_refln.wavelength`