Example 3.6.6.13. *The data set used in the refinement of an HIV-1 protease structure (PDB 5HVP) described using data items in the* REFLNS *and* REFLNS_SHELL *categories.*

```
_reflns.entry_id                         '5HVP'
_reflns.data_reduction_method
; Xengen program scalei. Anomalous pairs were merged.
  Scaling proceeded in several passes, beginning with
  1-parameter fit and ending with 3-parameter fit.
;
_reflns.data_reduction_details
; Merging and scaling based on only those reflections
  with I > sigma(I).
;

_reflns.d_resolution_high                2.00
_reflns.d_resolution_low                 8.00

_reflns.limit_h_max                      22
_reflns.limit_h_min                      0
_reflns.limit_k_max                      46
_reflns.limit_k_min                      0
_reflns.limit_l_max                      57
_reflns.limit_l_min                      0

_reflns.number_obs                       7228
_reflns.observed_criterion_sigma_I       1.0
_reflns.details                          none

loop_
_reflns_shell.d_res_high
_reflns_shell.d_res_low
_reflns_shell.meanI_over_sigI_obs
_reflns_shell.number_measured_obs
_reflns_shell.number_unique_obs
_reflns_shell.percent_possible_obs
_reflns_shell.Rmerge_F_obs
   31.38   3.82   69.8   9024   2540   96.8    1.98
    3.82   3.03   26.1   7413   2364   95.1    3.85
    3.03   2.65   10.5   5640   2123   86.2    6.37
    2.65   2.41    6.4   4322   1882   76.8    8.01
    2.41   2.23    4.3   3247   1714   70.4    9.86
    2.23   2.10    3.1   1140    812   33.3   13.99
```

a reflection as being observed are given using the data item `_reflns.observed_criterion`. This is a free-text field so is not automatically parsable. Therefore it is supplemented in the mmCIF dictionary by data items that can be used to stipulate the criterion in terms of the values of F, I or the uncertainties in these quantities (Example 3.6.6.13). The percentage of the total number of reflections that meet the criterion can be recorded.

Data items are also provided for describing the selection of the reflections used to calculate the free R factor, and for giving the $R_{\rm merge}$ values for all reflections and for the subset of 'observed' reflections. Data items in the REFLNS_SCALE and REFLNS_SHELL categories are used in the same way in the mmCIF and core CIF dictionaries, and Section 3.2.3.2.2 can be consulted for details.

As with the related categories DIFFRN_REFLNS_CLASS and REFINE_LS_CLASS, the core dictionary category REFLNS_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.

3.6.7. Atomicity, chemistry and structure

The basic concepts of the mmCIF model for describing a macromolecular structure were outlined in Section 3.6.3. The present section describes the components of the model in more detail. The category groups used to describe the molecular chemistry and structure are: the ATOM group describing atom positions (Section 3.6.7.1); the CHEMICAL, CHEM_COMP and CHEM_LINK groups describing molecular chemistry (Section 3.6.7.2); the ENTITY group describing distinct chemical species (Section 3.6.7.3); the GEOM group describing molecular or packing geometry (Section 3.6.7.4); the STRUCT group describing the large-scale features of molecular structure (Section 3.6.7.5); and the SYMMETRY group describing the symmetry and space group (Section 3.6.7.6).

The CHEMICAL category group itself is not generally used in an mmCIF. The purpose of this category group in the core CIF dictionary is to specify the chemical identity and connectivity of the relatively simple molecular or ionic species in a small-molecule or inorganic crystal. In principle, a macromolecular structure determined to atomic resolution could be represented as a coherent chemical entity with a complete connectivity graph. However, in practice, biological macromolecules are built from units from a library of models of standard amino acids, nucleotides and sugars. Data items in the CHEM_COMP and CHEM_LINK category groups of the mmCIF dictionary describe the internal connectivity and standard bonding processes between these units.

Molecular or packing geometry is also rarely tabulated for large macromolecular complexes, so the GEOM category group is rarely used in an mmCIF.

3.6.7.1. Atom sites

The categories describing atom sites are as follows:

ATOM group

Individual atom sites (§3.6.7.1.1)
- ATOM_SITE
- ATOM_SITE_ANISOTROP

Collections of atom sites (§3.6.7.1.2)
- ATOM_SITES
- ATOM_SITES_FOOTNOTE

Atom types (§3.6.7.1.3)
- ATOM_TYPE

Alternative conformations (§3.6.7.1.4)
- ATOM_SITES_ALT
- ATOM_SITES_ALT_ENS
- ATOM_SITES_ALT_GEN

The ATOM category group represents a compromise between the representation of a small-molecule structure as an annotated list of atomic coordinates and the need in macromolecular crystallography to present a more structured view organized around residues, chains, sheets, turns, helices *etc.* The locations of individual atoms and other information about the atom sites are given using data items in this category group. The categories within the group may be classified as shown in the summary above.

The ATOM_SITE, ATOM_SITES and ATOM_TYPE categories have many data items that are aliases of equivalent data items in the same categories in the core CIF dictionary, but the conventions for the labelling of the atom sites are different.

The ATOM_SITE_ANISOTROP and ATOM_SITES_FOOTNOTE categories are new to the mmCIF dictionary, as are the categories related to alternative conformations: ATOM_SITES_ALT, ATOM_SITES_ALT_ENS and ATOM_SITES_ALT_GEN.

3.6.7.1.1. *Individual atom sites*

The data items in these categories are as follows:

(*a*) ATOM_SITE

• **`_atom_site.id`** (∼ ***`_atom_site_label`***)
  `_atom_site.adp_type`
+ **`_atom_site.aniso_B[1][1]`**
  ⇌ ***`_atom_site_anisotrop.B[1][1]`***

+ _atom_site.aniso_B[1][2]
  ⇌ _atom_site_anisotrop.B[1][2]
+ _atom_site.aniso_B[1][3]
  ⇌ _atom_site_anisotrop.B[1][3]
+ _atom_site.aniso_B[2][2]
  ⇌ _atom_site_anisotrop.B[2][2]
+ _atom_site.aniso_B[2][3]
  ⇌ _atom_site_anisotrop.B[2][3]
+ _atom_site.aniso_B[3][3]
  ⇌ _atom_site_anisotrop.B[3][3]
  _atom_site.aniso_ratio
  ⇌ _atom_site_anisotrop.ratio
+ _atom_site.aniso_U[1][1]
  ⇌ _atom_site_anisotrop.U[1][1]
+ _atom_site.aniso_U[1][2]
  ⇌ _atom_site_anisotrop.U[1][2]
+ _atom_site.aniso_U[1][3]
  ⇌ _atom_site_anisotrop.U[1][3]
+ _atom_site.aniso_U[2][2]
  ⇌ _atom_site_anisotrop.U[2][2]
+ _atom_site.aniso_U[2][3]
  ⇌ _atom_site_anisotrop.U[2][3]
+ _atom_site.aniso_U[3][3]
  ⇌ _atom_site_anisotrop.U[3][3]
  _atom_site.attached_hydrogens
  _atom_site.auth_asym_id
  _atom_site.auth_atom_id
  _atom_site.auth_comp_id
  _atom_site.auth_seq_id
+ _atom_site.B_equiv_geom_mean
+ _atom_site.B_iso_or_equiv
  _atom_site.calc_attached_atom
  _atom_site.calc_flag
+ _atom_site.Cartn_x
+ _atom_site.Cartn_y
+ _atom_site.Cartn_z
  _atom_site.chemical_conn_number
  → _chemical_conn_atom.number
  _atom_site.constraints
  _atom_site.details (∼ _atom_site_description)
  _atom_site.disorder_assembly
  _atom_site.disorder_group
  _atom_site.footnote_id
+ _atom_site.fract_x
+ _atom_site.fract_y
+ _atom_site.fract_z
  _atom_site.group_PDB
  _atom_site.label_alt_id
  → _atom_sites_alt.id
  _atom_site.label_asym_id
  → _struct_asym.id
  _atom_site.label_atom_id
  → _chem_comp_atom.atom_id
  _atom_site.label_comp_id
  → _chem_comp.id
  _atom_site.label_entity_id
  → _entity.id
  _atom_site.label_seq_id
  → _entity_poly_seq.num
+ _atom_site.occupancy
  _atom_site.refinement_flags
  _atom_site.refinement_flags_adp
  _atom_site.refinement_flags_occupancy
  _atom_site.refinement_flags_posn
  _atom_site.restraints
  _atom_site.symmetry_multiplicity
  _atom_site.thermal_displace_type
  _atom_site.type_symbol
  → _atom_type.symbol
+ _atom_site.U_equiv_geom_mean
+ _atom_site.U_iso_or_equiv
  _atom_site.Wyckoff_symbol

(*b*) ATOM_SITE_ANISOTROP

• _atom_site_anisotrop.id
+ _atom_site_anisotrop.B[1][1] (∼ _atom_site_aniso_B_11)
+ _atom_site_anisotrop.B[1][2] (∼ _atom_site_aniso_B_12)
+ _atom_site_anisotrop.B[1][3] (∼ _atom_site_aniso_B_13)
+ _atom_site_anisotrop.B[2][2] (∼ _atom_site_aniso_B_22)
+ _atom_site_anisotrop.B[2][3] (∼ _atom_site_aniso_B_23)
+ _atom_site_anisotrop.B[3][3] (∼ _atom_site_aniso_B_33)
  _atom_site_anisotrop.ratio (∼ _atom_site_aniso_ratio)
  → _atom_site.id
  _atom_site_anisotrop.type_symbol
  (∼ _atom_site_aniso_type_symbol)
  → _atom_type.symbol
+ _atom_site_anisotrop.U[1][1] (∼ _atom_site_aniso_U_11)
+ _atom_site_anisotrop.U[1][2] (∼ _atom_site_aniso_U_12)
+ _atom_site_anisotrop.U[1][3] (∼ _atom_site_aniso_U_13)
+ _atom_site_anisotrop.U[2][2] (∼ _atom_site_aniso_U_22)
+ _atom_site_anisotrop.U[2][3] (∼ _atom_site_aniso_U_23)
+ _atom_site_anisotrop.U[3][3] (∼ _atom_site_aniso_U_33)

The bullet (•) indicates a category 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. The double arrow (⇌) indicates alternative names in a distinct category.

The refined coordinates of the atoms in the crystallographic asymmetric unit are stored in the ATOM_SITE category. Atom positions and their associated uncertainties may be given using either Cartesian or fractional coordinates, and anisotropic displacement factors and occupancies may be given for each position.

The relationships between categories describing atom sites are shown in Fig. 3.6.7.1.

Several of the mmCIF data names arise from the need to associate atom sites with residues and chains. As in the core CIF dictionary, the identifier for the atom site is the data item **_atom_site_label**. To accommodate standard practice in macromolecular crystallography, the mmCIF atom identifier is the aggregate of **_atom_site.label_alt_id**, ***.label_asym_id**, ***.label_atom_id**, ***.label_comp_id** and ***.label_seq_id**. For the two types of files to be compatible, the data item **_atom_site.id**, which is independent of the different modes of identifying atoms (discussed below), was introduced. The mmCIF identifier **_atom_site.id** is aliased to the core CIF identifier **_atom_site_label**.

Since the identifier does not need to be a number, it is quite possible (although it is not recommended) to use a complex label with an internal structure corresponding to the label components that the mmCIF dictionary provides as separate data items. This scheme is described in Section 3.2.4.1.1. However, normal practice in mmCIFs should be to label sites with the functional components available and to assign a simple numeric sequence to the values of **_atom_site.id** (see Example 3.6.7.1).

In addition to labelling information, each entry in the ATOM_SITE list must contain a value for the data item **_atom_site.type_symbol**, which is a pointer to the table of element symbols in the ATOM_TYPE category. All other data items in the ATOM_SITE category are optional, but it is normal practice to

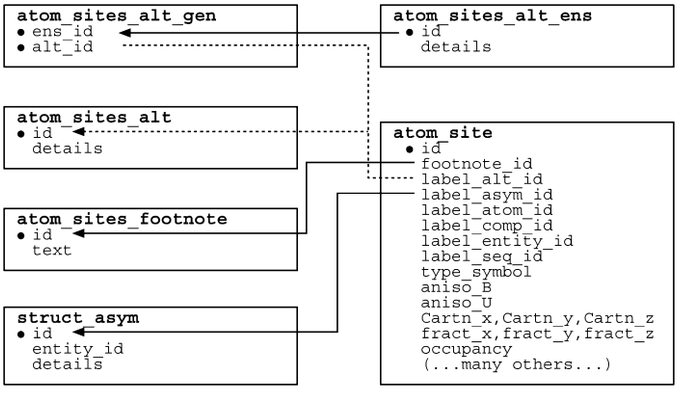

Fig. 3.6.7.1. The family of categories used to describe atom sites. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

Example 3.6.7.1. *Part of the coordinate list for an HIV-1 protease structure (PDB 5HVP) described with data items in the* ATOM_SITE *category. Atoms are given for both polymer and non-polymer regions of the structure, and atoms in the side chain of residue 12 adopt alternative conformations.*

```
loop_
_atom_site.group_PDB
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_seq_id
_atom_site.label_alt_id
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.footnote_id
_atom_site.auth_seq_id
_atom_site.id
 ATOM N N   THR  A  12  .  26.095  32.930  14.590
      1.00  18.97  4  12 8
 ATOM C CA  THR  A  12  .  25.734  32.995  16.032
      1.00  19.80  4  12 9
 ATOM C C   THR  A  12  .  24.695  34.106  16.113
      1.00  20.92  4  12 10
 ATOM O O   THR  A  12  .  24.869  35.118  15.421
      1.00  21.84  4  12 11
 ATOM C CB  THR  A  12  .  26.911  33.346  17.018
      1.00  20.51  4  12 12
 ATOM O OG1 THR  A  12  3  27.946  33.921  16.183
      0.50  20.29  4  12 13
 ATOM O OG1 THR  A  12  4  27.769  32.142  17.103
      0.50  20.59  4  12 14
 ATOM C CG2 THR  A  12  3  27.418  32.181  17.878
      0.50  20.47  4  12 15
 ATOM C CG2 THR  A  12  4  26.489  33.778  18.426
      0.50  20.00  4  12 16
# - - - abbreviated - - -
 HETATM C C1 APS  C  .  1  4.171  29.012   7.116
      0.58  17.27  1 300 101
 HETATM C C2 APS  C  .  1  4.949  27.758   6.793
      0.58  16.95  1 300 102
 HETATM O O3 APS  C  .  1  4.800  26.678   7.393
      0.58  16.85  1 300 103
 HETATM N N4 APS  C  .  1  5.930  27.841   5.869
      0.58  16.43  1 300 104
# - - - abbreviated - - -
```

give either the Cartesian or fractional coordinates. Most macromolecular structures use Cartesian coordinates. Isotropic displacement factors are normally placed directly in the ATOM_SITE category, using **_atom_site.B_iso_or_equiv**. Anisotropic displacement factors may be placed directly in the ATOM_SITE category *or* in the ATOM_SITE_ANISOTROP category. *U*'s may be used instead of *B*'s. It is not acceptable to use both *U*'s and *B*'s, nor is it acceptable to have anisotropic displacement factors in both the ATOM_SITE category and the ATOM_SITE_ANISOTROP category.

Each atom within each chemical component is uniquely identified using the data item **_atom_site.label_atom_id**, which is a reference to the data item **_chem_comp_atom.atom_id** in the CHEM_COMP_ATOM category.

The specific object in the asymmetric unit to which the atom belongs is indicated using the data item **_atom_site.label_asym_id**, which is a reference to the data item **_struct_asym.id** in the STRUCT_ASYM category. For macromolecules, it is useful to think of this identifier as a chain ID.

The chemical component to which the atom belongs is indicated using the data item **_atom_site.label_comp_id**, which is a reference to the data item **_chem_comp.id** in the CHEM_COMP category. The chemical component that is referenced in this way may be either a non-polymer or a monomer in a polymer; if it is a monomer in a polymer, it is useful to think of this identifier as the residue name.

The correspondence between the sequence of an entity in a polymer and the sequence information in the coordinate list (and in the STRUCT categories) is established using the data item **_atom_site.label_seq_id**, which is a reference to the data item **_entity_poly_seq.num** in the ENTITY_POLY_SEQ category. This identifier has no meaning for entities that are not part of a polymer; in a polymer it is useful to think of this identifier as the residue number. Note that this is strictly a number. If the combination of a number with an insertion code is needed, **_atom_site.auth_seq_id** should be used (see below).

An alternative set of identifiers can be used for the ***_asym_id**, ***_atom_id**, ***_comp_id** and ***_seq_id** identifiers, but not for ***_alt_id**. The **_atom_site.label_*** data names are standard; there are rules for these identifiers such as the requirement that residue numbers are sequential integers. Different databases may also have their own rules. However, the author of an mmCIF may wish to use a nonstandard labelling scheme, *e.g.* to reflect the residue numbering scheme of a structure to which the present structure is homologous, apart from insertions and gaps. Another situation in which a nonstandard labelling scheme might be used is to follow a local convention for atom names in a non-polymer, such as a haem, that conflicts with the scheme required by a database in which the structure is to be deposited. In these situations, alternative identifiers can be given using the data names (**_atom_site.auth_***).

In regions of the structure with alternative conformations, the specific conformation to which an atom belongs can be indicated using the data item **_atom_site.label_alt_id**, which is a reference to the data item **_atom_sites_alt.id** in the ATOM_SITES_ALT category.

The chemically distinct part of the structure (*e.g.* polymer chain, ligand, solvent) to which an atom belongs can be indicated using the data item **_atom_site.label_entity_id**, which is a reference to the data item **_entity.id** in the ENTITY category.

Most of the information that needs to be associated with an atom site is conveyed by the values of specific data names in mmCIF. However, for historical reasons, a pointer to additional free-text information about an atom site or about a group of atom sites can be given using the data item **_atom_site.footnote_id**, which is a reference to the data item **_atom_sites_footnote.id** in the ATOM_SITES_FOOTNOTE category.

The data item **_atom_site.group_PDB** is a place holder for the tags used by the PDB to identify types of coordinate records. It allows interconversion between mmCIFs and PDB format files. The only permitted values are ATOM and HETATM.

As in the core CIF dictionary, anisotropic displacement parameters in an mmCIF can be given in the same list as the atom positions and occupancies, or can be given in a separate list. However, DDL2 does not permit the same data names to be used for both constructs. Therefore, in mmCIF, anisotropic displacement parameters presented in a separate list are handled in a separate category with its own key, **_atom_site_anisotrop.id**, which must match a corresponding label in the atom-site list, **_atom_site.id**.

The individual elements of the anisotropic displacement matrix are labelled slightly differently in the mmCIF dictionary than in the core CIF dictionary in order to emphasize their matrix character. However, the definitions of the corresponding data items are identical in the two dictionaries.

3.6.7.1.2. *Collections of atom sites*

The data items in these categories are as follows:

(*a*) ATOM_SITES

- **_atom_sites.entry_id**
  → **_entry.id**
  _atom_sites.Cartn_transf_matrix[1][1]
  (∼ **_atom_sites_Cartn_tran_matrix_11**)
  _atom_sites.Cartn_transf_matrix[1][2]
  (∼ **_atom_sites_Cartn_tran_matrix_12**)
  _atom_sites.Cartn_transf_matrix[1][3]
  (∼ **_atom_sites_Cartn_tran_matrix_13**)
  _atom_sites.Cartn_transf_matrix[2][1]
  (∼ **_atom_sites_Cartn_tran_matrix_21**)
  _atom_sites.Cartn_transf_matrix[2][2]
  (∼ **_atom_sites_Cartn_tran_matrix_22**)
  _atom_sites.Cartn_transf_matrix[2][3]
  (∼ **_atom_sites_Cartn_tran_matrix_23**)
  _atom_sites.Cartn_transf_matrix[3][1]
  (∼ **_atom_sites_Cartn_tran_matrix_31**)
  _atom_sites.Cartn_transf_matrix[3][2]
  (∼ **_atom_sites_Cartn_tran_matrix_32**)
  _atom_sites.Cartn_transf_matrix[3][3]
  (∼ **_atom_sites_Cartn_tran_matrix_33**)
  _atom_sites.Cartn_transf_vector[1]
  (∼ **_atom_sites_Cartn_tran_vector_1**)
  _atom_sites.Cartn_transf_vector[2]
  (∼ **_atom_sites_Cartn_tran_vector_2**)
  _atom_sites.Cartn_transf_vector[3]
  (∼ **_atom_sites_Cartn_tran_vector_3**)
  _atom_sites.Cartn_transform_axes
  _atom_sites.fract_transf_matrix[1][1]
  (∼ **_atom_sites_fract_tran_matrix_11**)
  _atom_sites.fract_transf_matrix[1][2]
  (∼ **_atom_sites_fract_tran_matrix_12**)
  _atom_sites.fract_transf_matrix[1][3]
  (∼ **_atom_sites_fract_tran_matrix_13**)
  _atom_sites.fract_transf_matrix[2][1]
  (∼ **_atom_sites_fract_tran_matrix_21**)
  _atom_sites.fract_transf_matrix[2][2]
  (∼ **_atom_sites_fract_tran_matrix_22**)
  _atom_sites.fract_transf_matrix[2][3]
  (∼ **_atom_sites_fract_tran_matrix_23**)
  _atom_sites.fract_transf_matrix[3][1]
  (∼ **_atom_sites_fract_tran_matrix_31**)
  _atom_sites.fract_transf_matrix[3][2]
  (∼ **_atom_sites_fract_tran_matrix_32**)
  _atom_sites.fract_transf_matrix[3][3]
  (∼ **_atom_sites_fract_tran_matrix_33**)
  _atom_sites.fract_transf_vector[1]
  (∼ **_atom_sites_fract_tran_vector_1**)
  _atom_sites.fract_transf_vector[2]
  (∼ **_atom_sites_fract_tran_vector_2**)
  _atom_sites.fract_transf_vector[3]
  (∼ **_atom_sites_fract_tran_vector_3**)
  _atom_sites.solution_hydrogens
  _atom_sites.solution_primary
  _atom_sites.solution_secondary
  _atom_sites.special_details

(*b*) ATOM_SITES_FOOTNOTE

- **_atom_sites_footnote.id**
  _atom_sites_footnote.text

The bullet (•) *indicates a category 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.*

The ATOM_SITES category of the core dictionary, which is used to record information that applies collectively to all the atom sites in the model of the structure, is incorporated without change into the mmCIF dictionary, and Section 3.2.4.1.2 can be consulted for details.

In practice, the data names in the PHASING categories are preferred to the aliases to the core CIF data items **_atom_sites.solution_primary**, ***_secondary** and ***_hydrogens**. The data items in the mmCIF PHASING categories are designed to allow a much more detailed description of how a macromolecular structure was solved.

Example 3.6.7.2. *Footnotes for particular groups of atom sites in an HIV-1 protease structure (PDB 5HVP) using data items in the ATOM_SITES_FOOTNOTE category.*

```
loop_
  _atom_sites_footnote.id
  _atom_sites_footnote.text
     3
; The positions of these water molecules correlate
  with the alternative orientations of the
  inhibitor. Water molecules with alternative ID
  "1" and occupancy 0.58 correlate with
  inhibitor orientation "1". Water molecules with
  alternative ID "2" and occupancy 0.42 correlate
  with inhibitor orientation "2".
;
     4
; Side chains of these residues adopt alternative
  orientations that do not correlate with the
  alternative orientation of the inhibitor.
;
```

The data item **_atom_sites.entry_id** has been added to the ATOM_SITES category to provide the formal category key required by the DDL2 data model.

The ATOM_SITES_FOOTNOTE category can be used to note something about a group of sites in the ATOM_SITE coordinate list, each of which is flagged with the same value of **_atom_site.footnote_id**. For example, an author may wish to note atoms for which the electron density is very weak, or atoms for which static disorder has been modelled. Example 3.6.7.2 shows how an author has used these data items to describe alternative orientations in part of a structure. However, the very large number of data names describing specific structural characteristics in the mmCIF dictionary mean that these rather general data names are rarely needed.

3.6.7.1.3. *Atom types*

The data items in this category are as follows:

ATOM_TYPE

- **_atom_type.symbol**
  _atom_type.analytical_mass_percent
  (∼ **_atom_type_analytical_mass_%**)
  _atom_type.description
  _atom_type.number_in_cell
  _atom_type.oxidation_number
  _atom_type.radius_bond
  _atom_type.radius_contact
  _atom_type.scat_Cromer_Mann_a1
  _atom_type.scat_Cromer_Mann_a2
  _atom_type.scat_Cromer_Mann_a3
  _atom_type.scat_Cromer_Mann_a4
  _atom_type.scat_Cromer_Mann_b1
  _atom_type.scat_Cromer_Mann_b2
  _atom_type.scat_Cromer_Mann_b3
  _atom_type.scat_Cromer_Mann_b4
  _atom_type.scat_Cromer_Mann_c
  _atom_type.scat_dispersion_imag
  _atom_type.scat_dispersion_real
  _atom_type.scat_dispersion_source
  _atom_type.scat_length_neutron
  _atom_type.scat_source
  _atom_type.scat_versus_stol_list

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

The ATOM_TYPE category, which provides information about the atomic species associated with each atom site in the model of the structure, is used in the same way in the mmCIF dictionary as in the core CIF dictionary. See Section 3.2.4.1.3 for details.

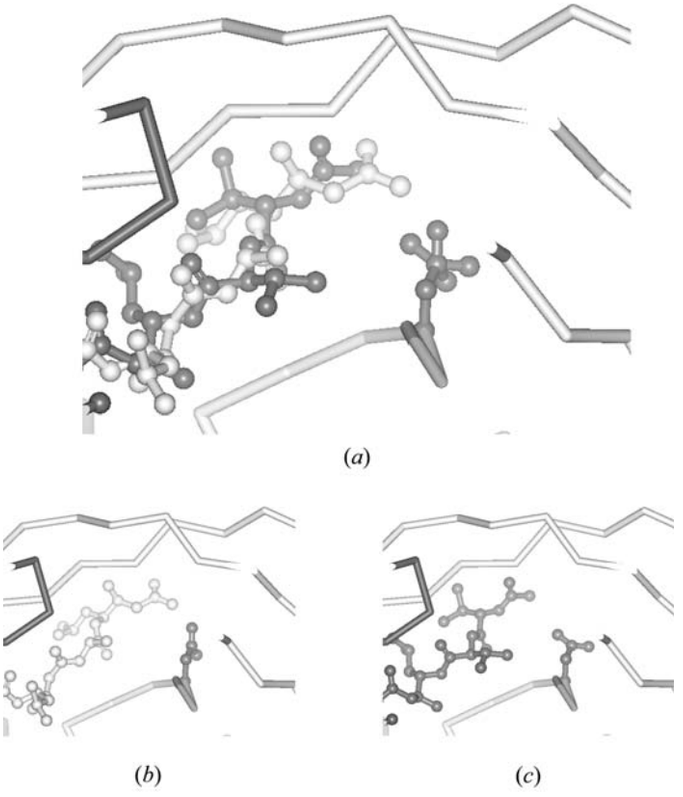

Fig. 3.6.7.2. Alternative conformations in an HIV-1 protease structure (PDB 5HVP) to be described with data items in the ATOM_SITES_ALT, ATOM_SITES_ALT_ENS and ATOM_SITES_ALT_GEN categories. (*a*) Complete structure, (*b*) ensemble 1, (*c*) ensemble 2.

3.6.7.1.4. *Alternative conformations*

The data items in these categories are as follows:

(*a*) ATOM_SITES_ALT

- **_atom_sites_alt.id**
  _atom_sites_alt.details

(*b*) ATOM_SITES_ALT_ENS

- **_atom_sites_alt_ens.id**
  _atom_sites_alt_ens.details

(*c*) ATOM_SITES_ALT_GEN

- **_atom_sites_alt_gen.alt_id**
  → **_atom_sites_alt.id**
- **_atom_sites_alt_gen.ens_id**
  → **_atom_sites_alt_gen.ens_id**

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.

Biological macromolecules are often very flexible, and as the resolution of a structure determination increases, it becomes increasingly possible to model reliably the alternative conformations that the structure adopts. Typically, partial occupancies are assigned to atom sites within the alternative conformations to indicate the relative frequency of occurrence of each conformation. It can, however, be difficult to deduce the possible different conformations of the whole structure from inspection of the atom-site occupancies alone. For instance, a segment of protein main chain might adopt one of three slightly different conformations, and within each conformation a particular side chain might adopt one of two possible conformations, one of which sterically distorts an adjacent residue sequence, while the other does not. The data model in the mmCIF dictionary allows these kinds of correlations in positions to be described.

The relationships between the categories used to describe alternative conformations are shown in Fig. 3.6.7.1.

In the core CIF dictionary, alternative conformations are indicated by using the **_atom_site.disorder_assembly** and ***.disorder_group** data items. Aliases to these data items are present in the mmCIF dictionary, but it is not intended that they should be used to describe disorder in a macromolecular structure.

The model for describing alternative conformations in mmCIF uses the ATOM_SITES_ALT family of categories. Ensembles of correlated alternative conformations can be identified using the category ATOM_SITES_ALT_ENS. Each ensemble is generated from one or more of the alternative conformations given in the list of alternative sites in the ATOM_SITES_ALT category. Data items in the

Example 3.6.7.3. *Alternative conformations in an HIV-1 protease structure (PDB 5HVP) described with data items in the ATOM_SITES_ALT, ATOM_SITES_ALT_ENS and ATOM_SITES_ALT_GEN categories.*

```
loop_
_atom_sites_alt.id
_atom_sites_alt.details
  .
; Atom sites with the alternative ID set to null are
  not modelled in alternative conformations
;
  1
; Atom sites with the alternative ID set to 1 have
  been modelled in alternative conformations with
  respect to atom sites marked with alternative
  ID 2. The conformations of amino-acid side chains
  with alternative ID set to 1 correlate with the
  conformation of the inhibitor marked with
  alternative ID 1. Atoms in these side chains have
  been given an occupancy of 0.58 to match the
  occupancy assigned to the inhibitor.
;
  2
; Atom sites with the alternative ID set to 2 have
  been modelled in alternative conformations with
  respect to atom sites marked with alternative
  ID 1. The conformations of amino-acid side chains
  with alternative ID set to 2 correlate with the
  conformation of the inhibitor marked with
  alternative ID 2. Atoms in these side chains have
  been given an occupancy of 0.42 to match the
  occupancy assigned to the inhibitor.
;

loop_
_atom_sites_alt_ens.id
_atom_sites_alt_ens.details
  'Ensemble 1'
; The inhibitor binds to the enzyme in two, roughly
  twofold symmetric, alternative conformations.

  This conformational ensemble includes the more-
  populated conformation of the inhibitor (ID=1) and
  the amino-acid side chains that correlate with this
  inhibitor conformation.
;
  'Ensemble 2'
; The inhibitor binds to the enzyme in two, roughly
  twofold symmetric, alternative conformations.

  This conformational ensemble includes the less-
  populated conformation of the inhibitor (ID=2) and
  the amino-acid side chains that correlate with this
  inhibitor conformation.
;

loop_
_atom_sites_alt_gen.ens_id
_atom_sites_alt_gen.alt_id
   'Ensemble 1'  .
   'Ensemble 1'  1
   'Ensemble 2'  .
   'Ensemble 2'  2
```

ATOM_SITES_ALT_GEN category explicitly tie together the alternative conformations that contribute to each ensemble. Finally, the atoms in each alternative conformation are identified in the ATOM_SITE category by the data item **_atom_site.label_alt_id**.

The current version of the mmCIF dictionary cannot be used to describe an NMR structure determination completely. However, an mmCIF can be used to store the multiple models usually used to describe a structure determined by NMR using the data items in these categories.

Example 3.6.7.3 is a simplified version of the example given in the mmCIF dictionary (see Fig. 3.6.7.2).

3.6.7.2. Molecular chemistry

The categories describing molecular chemistry are as follows:

Molecular chemistry in the core CIF dictionary (§3.6.7.2.1)

CHEMICAL group
- CHEMICAL
- CHEMICAL_CONN_ATOM
- CHEMICAL_CONN_BOND
- CHEMICAL_FORMULA

Chemical components (§3.6.7.2.2)

CHEM_COMP group
- CHEM_COMP
- CHEM_COMP_ANGLE
- CHEM_COMP_ATOM
- CHEM_COMP_BOND
- CHEM_COMP_CHIR
- CHEM_COMP_CHIR_ATOM
- CHEM_COMP_PLANE
- CHEM_COMP_PLANE_ATOM
- CHEM_COMP_TOR
- CHEM_COMP_TOR_VALUE

Chemical links (§3.6.7.2.3)

CHEM_LINK group
- CHEM_COMP_LINK
- CHEM_LINK
- CHEM_LINK_ANGLE
- CHEM_LINK_BOND
- CHEM_LINK_CHIR
- CHEM_LINK_CHIR_ATOM
- CHEM_LINK_PLANE
- CHEM_LINK_PLANE_ATOM
- CHEM_LINK_TOR
- CHEM_LINK_TOR_VALUE
- ENTITY_LINK

The detailed chemistry of the components of a macromolecular structure can be described using data items in the CHEM_COMP and CHEM_LINK category groups. These mmCIF categories are used in preference to those in the CHEMICAL category group in the core CIF dictionary, as macromolecules are in most cases linked assemblies of a limited number of monomers and so they are most efficiently described by defining the monomers and the links between them, rather than by a formal definition of every bond and angle.

All the categories relevant to molecular chemistry are listed in the summary above; note in particular the presence of the category ENTITY_LINK within the formal CHEM_LINK category group.

3.6.7.2.1. *Molecular chemistry in the core CIF dictionary*

The data items in these categories are as follows:

(*a*) CHEMICAL

```
• _chemical.entry_id
        → _entry.id
  _chemical.absolute_configuration
  _chemical.compound_source
  _chemical.melting_point
  _chemical.melting_point_gt
  _chemical.melting_point_lt
  _chemical.name_common
  _chemical.name_mineral
  _chemical.name_structure_type
  _chemical.name_systematic
  _chemical.optical_rotation
  _chemical.properties_biological
  _chemical.properties_physical
+ _chemical.temperature_decomposition
  _chemical.temperature_decomposition_gt
  _chemical.temperature_decomposition_lt
+ _chemical.temperature_sublimation
  _chemical.temperature_sublimation_gt
  _chemical.temperature_sublimation_lt
```

(*b*) CHEMICAL_CONN_ATOM

```
• _chemical_conn_atom.number
  _chemical_conn_atom.charge
  _chemical_conn_atom.display_x
  _chemical_conn_atom.display_y
  _chemical_conn_atom.NCA
  _chemical_conn_atom.NH
  _chemical_conn_atom.type_symbol
```

(*c*) CHEMICAL_CONN_BOND

```
• _chemical_conn_bond.atom_1
• _chemical_conn_bond.atom_2
  _chemical_conn_bond.type
```

(*d*) CHEMICAL_FORMULA

```
• _chemical_formula.entry_id
        → _entry.id
  _chemical_formula.analytical
  _chemical_formula.iupac
  _chemical_formula.moiety
  _chemical_formula.structural
  _chemical_formula.sum
  _chemical_formula.weight
  _chemical_formula.weight_meas
```

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 (_). 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.*

Descriptions of molecular chemistry in an mmCIF are normally made using data items in the CHEM_COMP and CHEM_LINK category groups. The CHEMICAL category group is retained in the mmCIF dictionary solely for consistency with the core CIF dictionary and Section 3.2.4.2 may be consulted for details.

Two of the categories in this group, CHEMICAL_CONN_ATOM and CHEMICAL_CONN_BOND, have existing category keys in the core dictionary. The formal keys **_chemical.entry_id** and **_chemical_formula.entry_id** have been added to CHEMICAL and CHEMICAL_FORMULA, respectively, to provide the category keys required by the DDL2 data model.

It is emphasized that these items will not appear in the description of a macromolecular structure, but they are retained to allow the representation of small-molecule or inorganic structures in the DDL2 formalism of mmCIF.

3.6.7.2.2. *Chemical components*

Data items in these categories are as follows:

(*a*) CHEM_COMP

```
• _chem_comp.id
  _chem_comp.formula
  _chem_comp.formula_weight
  _chem_comp.model_details
  _chem_comp.model_erf
```