

3.6. CLASSIFICATION AND USE OF MACROMOLECULAR DATA

words for each biological unit can be given using data items in the STRUCT_BIOL_KEYWORD category.

The entities that comprise the biological unit are specified using data items in the STRUCT_BIOL_GEN category by reference to the appropriate values of `_struct_asym.id` and by specifying any symmetry transformation that must be applied to the entities to generate the biological unit.

Data items in the STRUCT_BIOL_VIEW category allow the author to specify an orientation of the biological unit that provides a useful view of the structure. The comments given in `_struct_biol_view.details` may be used as a figure caption if the view is intended to be a figure in a report describing the structure.

The example of crambin in Section 3.6.3 shows the relations between the categories defining higher-level structure for the straightforward case of a single protein molecule (with a small co-crystallization molecule and solvent) in the asymmetric unit. The structure of HIV-1 protease with a bound inhibitor (PDB 5HVP), shown in Example 3.6.7.8, is considerably more complex. There are two entities: the monomeric form of the enzyme and the small-molecule inhibitor. The asymmetric unit contains two copies of the enzyme monomer (both fully occupied) and two copies of the inhibitor (each of which is partially occupied) (Fig. 3.6.7.8). Three biological assemblies are constructed for this system. One biological unit contains only the dimeric enzyme (Fig. 3.6.7.8*b*), the second contains the dimeric enzyme with one partially occupied conformation of the inhibitor (Fig. 3.6.7.8*c*) and the third contains the dimeric enzyme with the second partially occupied conformation of the inhibitor (Fig. 3.6.7.8*d*). There are alternative conformations of the side chains in the enzyme that correlate with the binding mode of the inhibitor.

3.6.7.5.2. Secondary structure

The data items in these categories are as follows:

(a) STRUCT_CONF_TYPE

- `_struct_conf_type.id`
- `_struct_conf_type.criteria`
- `_struct_conf_type.reference`

(b) STRUCT_CONF

- `_struct_conf.id`
- `_struct_conf.beg_label_asym_id`
→ `_atom_site.label_asym_id`
- `_struct_conf.beg_label_comp_id`
→ `_atom_site.label_comp_id`
- `_struct_conf.beg_label_seq_id`
→ `_atom_site.label_seq_id`
- `_struct_conf.beg_auth_asym_id`
→ `_atom_site.auth_asym_id`
- `_struct_conf.beg_auth_comp_id`
→ `_atom_site.auth_comp_id`
- `_struct_conf.beg_auth_seq_id`
→ `_atom_site.auth_seq_id`
- `_struct_conf.conf_type_id`
→ `_struct_conf_type.id`
- `_struct_conf.details`
- `_struct_conf.end_label_asym_id`
→ `_atom_site.label_asym_id`
- `_struct_conf.end_label_comp_id`
→ `_atom_site.label_comp_id`
- `_struct_conf.end_label_seq_id`
→ `_atom_site.label_seq_id`
- `_struct_conf.end_auth_asym_id`
→ `_atom_site.auth_asym_id`
- `_struct_conf.end_auth_comp_id`
→ `_atom_site.auth_comp_id`
- `_struct_conf.end_auth_seq_id`
→ `_atom_site.auth_seq_id`

The bullet (•) indicates a category key. The arrow (→) is a reference to a parent data item.

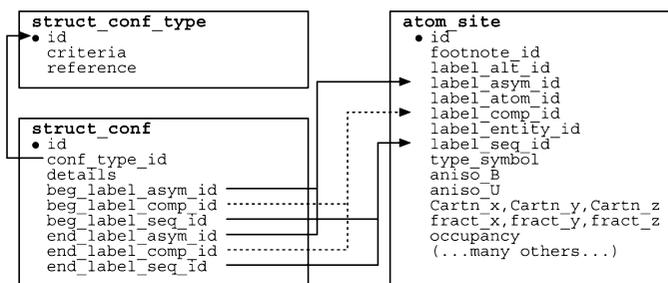


Fig. 3.6.7.9. The family of categories used to describe secondary structure. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

Example 3.6.7.9. Secondary structure in an HIV-1 protease structure (PDB 5HVP) described with data items in the STRUCT_CONF_TYPE and STRUCT_CONF categories.

```

loop_
  _struct_conf_type.id
  _struct_conf_type.criteria
  HELX_RH_AL_P 'author judgement'
  STRN          'author judgement'
  TURN_TY1_P   'author judgement'
  TURN_TY1P_P  'author judgement'
  TURN_TY2_P   'author judgement'
  TURN_TY2P_P  'author judgement'

loop_
  _struct_conf.id
  _struct_conf.conf_type_id
  _struct_conf.beg_label_comp_id
  _struct_conf.beg_label_asym_id
  _struct_conf.beg_label_seq_id
  _struct_conf.end_label_comp_id
  _struct_conf.end_label_asym_id
  _struct_conf.end_label_seq_id
  HELX1  HELX_RH_AL_P  ARG  A   87  GLN  A   92
  HELX2  HELX_RH_AL_P  ARG  B  287  GLN  B  292
  STRN1  STRN          PRO  A    1  LEU  A    5
  STRN2  STRN          CYS  B  295  PHE  B  299
  STRN3  STRN          CYS  A   95  PHE  A  299
  STRN4  STRN          PRO  B  201  LEU  B  205
  TURN1  TURN_TY1P_P  ILE  A   15  GLN  A   18
  TURN2  TURN_TY2_P   GLY  A   49  GLY  A   52
  TURN3  TURN_TY1P_P  ILE  A   55  HIS  A   69
  TURN4  TURN_TY1_P   THR  A   91  GLY  A   94

```

The primary structure of a macromolecule is defined by the sequence of the components (amino acids, nucleic acids or sugars) in the polymer chain. The polymer chains assume conformations based on the torsion angles adopted by the rotatable bonds in the polymer backbone; the resulting conformations are referred to as the secondary structure of the polymer. Several patterns of values of backbone torsion angles have been described and given names, such as α -helix, β -strand, turn and coil for proteins, and A-, B- and Z-helix for nucleic acids.

In the mmCIF dictionary, these secondary structures are described in the STRUCT_CONF and STRUCT_CONF_TYPE categories. Note that the data items in these categories describe only the secondary structure; the tertiary organization of β -strands into β -sheets is described in the STRUCT_SHEET_* categories. There are no data items for describing the tertiary organization of α -helices or nucleic acids in the current version of the mmCIF dictionary.

The relationships between categories used to describe secondary structure are shown in Fig. 3.6.7.9.

The type of the secondary structure is specified in the STRUCT_CONF_TYPE category, along with the criteria used to identify it. The range of monomers assigned to each secondary-structure element is given in the STRUCT_CONF category.