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.

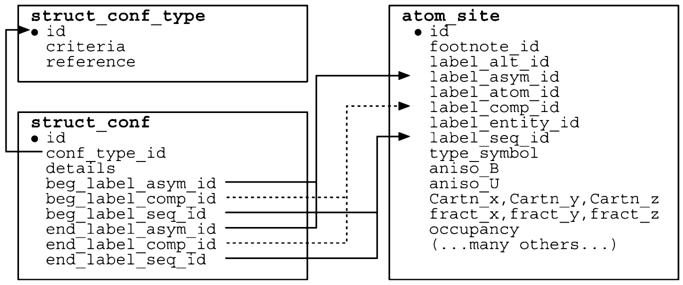

Fig. 3.6.7.9. The family of categories used to describe secondary structure. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (●). Lines show relationships between linked data items in different categories with arrows pointing at the parent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

The allowed values for the data item **_struct_conf_type.id** cover most types of protein and nucleic acid secondary structure (Example 3.6.7.9). The criteria that define the secondary structure may be given using the data item **_struct_conf_type.criteria**. **_struct_conf_type.reference** can be used to specify a reference to the literature in which the criteria are explained in more detail.

The residues that define the beginning and end of each region of secondary structure are identified with the appropriate ***_asym**, ***_comp** and ***_seq** identifiers. The standard labelling system or the author's alternative labelling system may be used. The identification of the residues assigned to each region of secondary structure is linked to the labelling information in the ATOM_SITE category. Unusual features of a conformation may be described using **_struct_conf.details**.

3.6.7.5.3. *Structural interactions*

The data items in these categories are as follows:

```
(a) STRUCT_CONN_TYPE
• _struct_conn_type.id
  _struct_conn_type.criteria
  _struct_conn_type.reference

(b) STRUCT_CONN
• _struct_conn.id
  _struct_conn.conn_type_id
        → _struct_conn_type.id
  _struct_conn.details
  _struct_conn.ptnr1_label_alt_id
        → _atom_sites_alt.id
  _struct_conn.ptnr1_label_asym_id
        → _atom_site.label_asym_id
  _struct_conn.ptnr1_label_atom_id
        → _chem_comp_atom.atom_id
  _struct_conn.ptnr1_label_comp_id
        → _atom_site.label_comp_id
  _struct_conn.ptnr1_label_seq_id
        → _atom_site.label_seq_id
  _struct_conn.ptnr1_auth_asym_id
        → _atom_site.auth_asym_id
  _struct_conn.ptnr1_auth_atom_id
        → _atom_site.auth_atom_id
  _struct_conn.ptnr1_auth_comp_id
        → _atom_site.auth_comp_id
  _struct_conn.ptnr1_auth_seq_id
        → _atom_site.auth_seq_id
  _struct_conn.ptnr1_role
  _struct_conn.ptnr1_symmetry
  _struct_conn.ptnr2_label_alt_id
        → _atom_sites_alt.id
  _struct_conn.ptnr2_label_asym_id
        → _atom_site.label_asym_id
  _struct_conn.ptnr2_label_atom_id
        → _chem_comp_atom.atom_id
  _struct_conn.ptnr2_label_comp_id
        → _atom_site.label_comp_id
  _struct_conn.ptnr2_label_seq_id
        → _atom_site.label_seq_id
  _struct_conn.ptnr2_auth_asym_id
        → _atom_site.auth_asym_id
  _struct_conn.ptnr2_auth_atom_id
        → _atom_site.auth_atom_id
  _struct_conn.ptnr2_auth_comp_id
        → _atom_site.auth_comp_id
  _struct_conn.ptnr2_auth_seq_id
        → _atom_site.auth_seq_id
  _struct_conn.ptnr2_role
  _struct_conn.ptnr2_symmetry
```

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

The structural interactions that are described with data items in the STRUCT_CONN family of categories are the tertiary result of a structure determination, not the chemical connectivity of the components of the structure. In general, the interactions described using the STRUCT_CONN data items are noncovalent, such as hydrogen bonds, salt bridges and metal coordination.

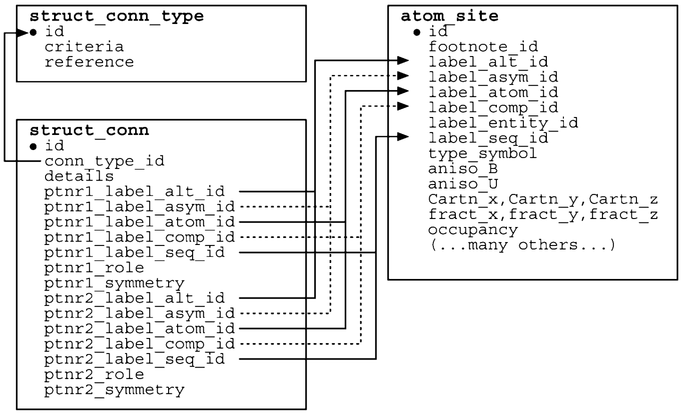

Fig. 3.6.7.10. The family of categories used to describe structural interactions such as hydrogen bonding, salt bridges and disulfide bridges. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

It is useful to think of the structure interactions given in CHEM_COMP_BOND, CHEM_LINK and ENTITY_LINK as the covalent interactions that are known in advance of the structure determination because the chemistry of the components is well defined. Literature or calculated values for these interactions are often used as restraints during the refinement. In contrast, the structural interactions described in the STRUCT_CONN family of categories are not known in advance and are part of the results of the structure determination.

This distinction only holds approximately, as there are clearly bonds, such as disulfide links, that are covalent and usually restrained during the refinement but that are also a result of the folding of the protein revealed by the structure determination, and thus should be described using STRUCT_CONN data items.

In general, the STRUCT_CONN data items would not be used to list all the structure interactions. Instead, the author of the mmCIF would use the STRUCT_CONN data items to identify and annotate only the structural interactions worthy of discussion. The relationships between categories used to describe structural interactions are shown in Fig. 3.6.7.10.

Structural interactions such as hydrogen bonds, salt bridges and disulfide bridges can be described in the STRUCT_CONN category. The type of each interaction and the criteria used to identify the interaction can be specified in the STRUCT_CONN_TYPE category (Example 3.6.7.10).

The atoms participating in each interaction are arbitrarily labelled as 'partner 1' and 'partner 2'. Each is identified by the ***_alt**, ***_asym**, ***_atom**, ***_comp** and ***_seq** constituents of the corresponding atom-site label. The role of each partner in the interaction (*e.g.* donor, acceptor) may be specified, and any crystallographic symmetry operation needed to transform the atom from the position given in the ATOM_SITE list to the position where the interaction occurs can be given. The atoms participating in the interaction may also be identified using an alternative labelling scheme if the author has supplied one.

Unusual aspects of the interaction may be discussed in **_struct_conn.details**. The general type of an interaction can be indicated using **_struct_conn.conn_type_id**, which references one of the standard types described using data items in the STRUCT_CONN_TYPE category.

The specific types of structural connection that may be recorded are those allowed for **_struct_conn_type.id**, namely covalent and hydrogen bonds, ionic (salt-bridge) interactions, disulfide