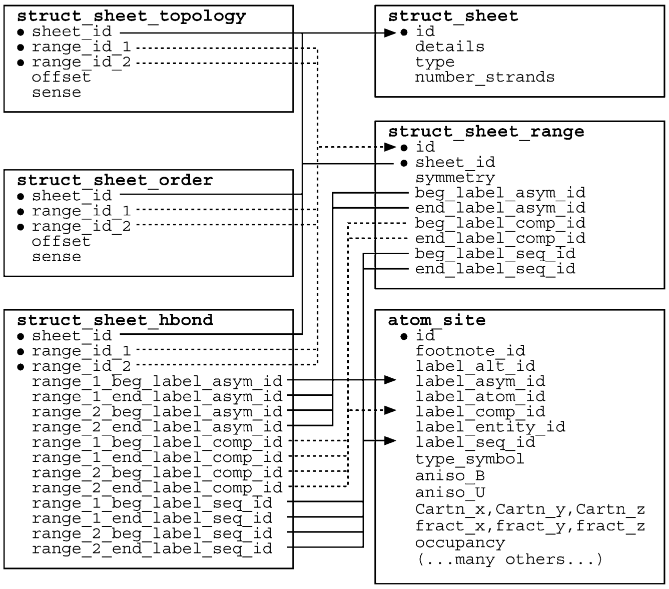

Fig. 3.6.7.14. The family of categories used to describe β-sheets. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (●). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

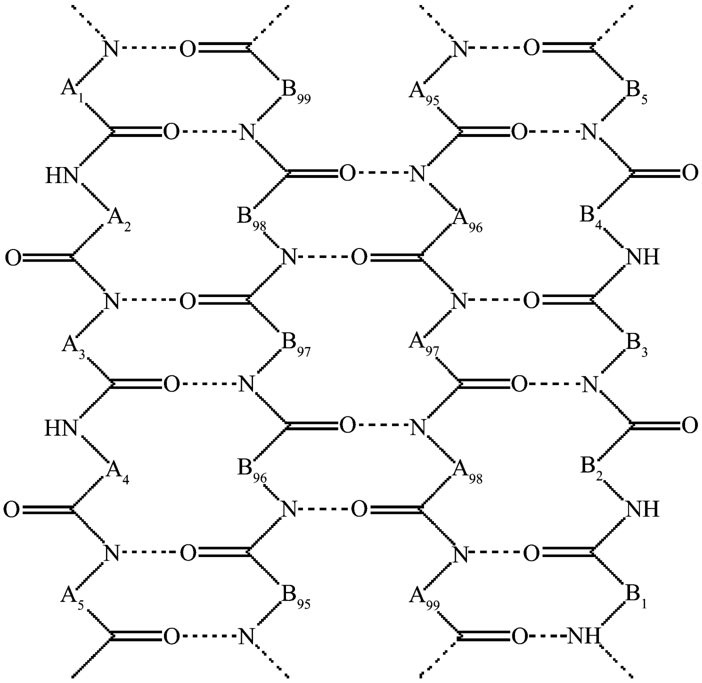

Fig. 3.6.7.15. A hypothetical β-sheet to be described with data items in the STRUCT_SHEET, STRUCT_SHEET_ORDER, STRUCT_SHEET_RANGE and STRUCT_SHEET_HBOND categories. Note that the strands come from two different polypeptides, labelled A and B.

Example 3.6.7.14. *A hypothetical β-sheet described with data items in the STRUCT_SHEET, STRUCT_SHEET_ORDER, STRUCT_SHEET_RANGE and STRUCT_SHEET_HBOND categories.*

```
loop_
  _struct_sheet.id
  _struct_sheet.number_strands
    S1    4

loop_
  _struct_sheet_order.sheet_id
  _struct_sheet_order.range_id_1
  _struct_sheet_order.range_id_2
  _struct_sheet_order.sense
    S1 1 2 anti-parallel
    S1 2 3 anti-parallel
    S1 3 4 anti-parallel
    S2 1 2 anti-parallel

loop_
  _struct_sheet_range.sheet_id
  _struct_sheet_range.id
  _struct_sheet_range.beg_label_comp_id
  _struct_sheet_range.beg_label_asym_id
  _struct_sheet_range.beg_label_seq_id
  _struct_sheet_range.end_label_comp_id
  _struct_sheet_range.end_label_asym_id
  _struct_sheet_range.end_label_seq_id
    S1 1 PRO A 1  LEU A 5
    S1 2 CYS B 95 PHE B 99
    S1 3 CYS A 95 PHE A 99
    S1 4 PRO B 1  LEU B 5

loop_
  _struct_sheet_hbond.sheet_id
  _struct_sheet_hbond.range_id_1
  _struct_sheet_hbond.range_id_2
  _struct_sheet_hbond.range_1_beg_label_atom_id
  _struct_sheet_hbond.range_1_beg_label_seq_id
  _struct_sheet_hbond.range_2_beg_label_atom_id
  _struct_sheet_hbond.range_2_beg_label_seq_id
    S1 1 2 A 3  O 97
    S1 2 3 B 98 O 96
    S1 3 4 A 97 O 3
```

In the more detailed and more general method for describing β-sheets, data items in the STRUCT_SHEET_RANGE category specify the range of residues that form strands in the sheet, data items in the STRUCT_SHEET_ORDER category specify the relative pairwise orientation of strands and data items in the STRUCT_SHEET_HBOND category provide details of specific hydrogen-bonding interactions between strands (see Fig. 3.6.7.15 and Example 3.6.7.14). Note that the specifiers for the strand ranges include the amino acid (**`*_comp_id`** and **`*_seq_id`**), the chain (**`*_asym_id`**) and a symmetry code (**`_struct_sheet_range.symmetry`**). Thus sheets that are composed of strands from more than one polypeptide chain or from polypeptides in more than one asymmetric unit can be described.

It is conventional to assign the number 1 to an outermost strand. The choice of which outermost strand to number as 1 is arbitrary, but would usually be the strand encountered first in the amino-acid sequence. The remaining strands are then numbered sequentially across the sheet.

In some simple cases, the complete hydrogen bonding of the sheet could be inferred from the strand-range pairings and the relationship between the strands (parallel or antiparallel). However, in most cases it is necessary to specify at least one hydrogen bond between adjacent strands in order to establish the registration. The data items in the STRUCT_SHEET_HBOND category can be used to do this. Hydrogen bonds also need to be specified precisely when a sheet contains a nonstandard feature such as a β-bulge. This is a case where it is sufficient to specify a single hydrogen-bonding interaction to establish the registration; here only the **`*_beg_*`** or **`*_end_*`** data items need to be used to reference the atom-label components. However, it is preferable, wherever possible, to specify the initial and final atoms of the two ranges participating in the hydrogen bonding.

3.6.7.5.8. *Molecular sites*

The data items in these categories are as follows:

(*a*) STRUCT_SITE
● **`_struct_site.id`**
`_struct_site.details`

(*b*) STRUCT_SITE_KEYWORDS

```
● _struct_site_keywords.site_id
        → _struct_site.id
● _struct_site_keywords.text
```

(*c*) STRUCT_SITE_GEN

```
● _struct_site_gen.id
● _struct_site_gen.site_id
        → _struct_site.id
  _struct_site_gen.details
  _struct_site_gen.label_alt_id
        → _atom_sites_alt.id
  _struct_site_gen.label_asym_id
        → _atom_site.label_asym_id
  _struct_site_gen.label_atom_id
        → _chem_comp_atom.atom_id
  _struct_site_gen.label_comp_id
        → _atom_site.label_atom_id
  _struct_site_gen.label_seq_id
        → _atom_site.label_seq_id
  _struct_site_gen.auth_asym_id
        → _atom_site.auth_asym_id
  _struct_site_gen.auth_atom_id
        → _atom_site.auth_atom_id
  _struct_site_gen.auth_comp_id
        → _atom_site.auth_comp_id
  _struct_site_gen.auth_seq_id
        → _atom_site.auth_seq_id
  _struct_site_gen.symmetry
```

(*d*) STRUCT_SITE_VIEW

```
● _struct_site_view.id
  _struct_site_view.details
  _struct_site_view.rot_matrix[1][1]
  _struct_site_view.rot_matrix[1][2]
  _struct_site_view.rot_matrix[1][3]
  _struct_site_view.rot_matrix[2][1]
  _struct_site_view.rot_matrix[2][2]
  _struct_site_view.rot_matrix[2][3]
  _struct_site_view.rot_matrix[3][1]
  _struct_site_view.rot_matrix[3][2]
  _struct_site_view.rot_matrix[3][3]
  _struct_site_view.site_id
        → _struct_site.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.

Substrate-binding sites, active sites, metal coordination sites and any other sites of interest may be described using data items in a collection of categories descending from STRUCT_SITE. These categories are intended to enable the author to generate views of molecular sites that could be used as figures in a report describing the structure or to enable a database to store standard views of common molecular sites (*e.g.* ATP-binding sites or the coordination of a calcium atom). The relationships between categories used to describe structural sites are shown in Fig. 3.6.7.16.

An identifier for each site that an author wishes to describe is given using `_struct_site.id` and the site can be described using `_struct_site.details`.

Keywords can be given for each site using data items in the STRUCT_SITE_KEYWORD category. Because keywords can be given at many levels of the mmCIF description of a structure, it may be worth duplicating the most significant higher-level keywords at this level to ensure that the site is detected in all search strategies.

The structural elements that generate each molecular site can be specified using data items in the STRUCT_SITE_GEN category. 'Structural elements' in this sense may be at any level of detail in the structure: single atoms, complete amino acids or nucleotides, or elements of secondary, tertiary or quaternary structure. Therefore the labels for each element may include, as required, the relevant `*_alt`, `*_asym`, `*_atom`, `*_comp` or `*_seq` parts of atom or residue identifiers. If the author has used an alternative labelling scheme, this can also be used. Noteworthy features of a structural element that forms part of the site can be described using the data item `_struct_site_gen.details`. Any crystallographic symmetry operations that are needed to form the site can be given using `_struct_site_gen.symmetry`.

Data items in the STRUCT_SITE_VIEW category allow the author to specify an orientation of the molecular site that gives a useful view of the components. The comments given in `_struct_site_view.details` could be used as a figure caption if the view is intended for use as a figure in a report.

Example 3.6.7.15 illustrates the use of these categories for describing a DNA binding site.

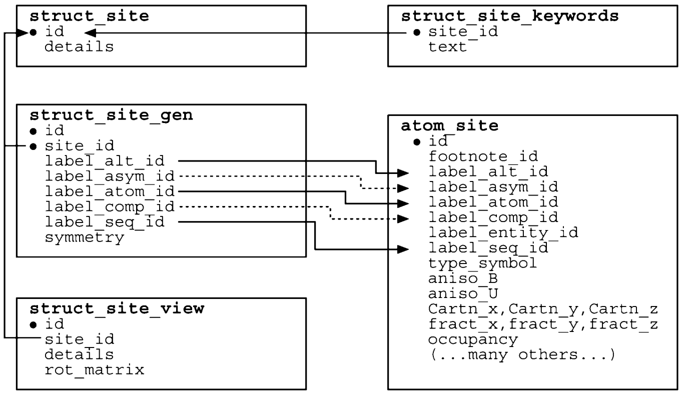

Fig. 3.6.7.16. The family of categories used to describe molecular sites. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (●). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

Example 3.6.7.15. *A DNA binding site with an intercalated drug (NDB DDF040) described with data items in the STRUCT_SITE, STRUCT_SITE_KEYWORDS, STRUCT_SITE_GEN and STRUCT_SITE_VIEW categories.*

```
loop_
_struct_site.id
_struct_site.details
   B1  'Binding at TG/AC Step 1'

loop_
_struct_site_keywords.site_id
_struct_site_keywords.text
   B1  'Intercalation complex'

loop_
_struct_site_gen.id
_struct_site_gen.site_id
_struct_site_gen.label_asym_id
_struct_site_gen.label_comp_id
_struct_site_gen.label_seq_id
_struct_site_gen.symmetry
   1  B1  A  T    1  1_555
   2  B1  A  G    2  1_555
   3  B1  A  C    5  8_555
   4  B1  A  A    6  8_555
   5  B1  D  DM2  .  8_555

loop_
_struct_site_view.id
_struct_site_view.site_id
_struct_site_view.details
_struct_site_view.rot_matrix[1][1]
_struct_site_view.rot_matrix[1][2]
# - - - abbreviated - - -
_struct_site_view.rot_matrix[3][3]
   View1  B1
  'View along the base-pair plane'
   0.133  0.922  . . . . . . -0.172
```