

## 3.6. CLASSIFICATION AND USE OF MACROMOLECULAR DATA

## (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.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

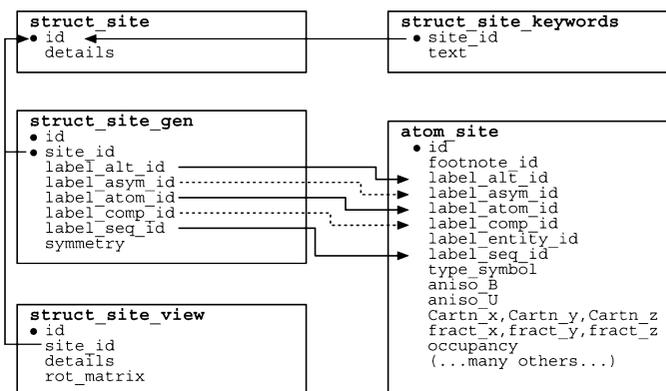


Fig. 3.6.7.16. The family of categories used to describe molecular sites. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent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

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.