

## 3.6. CLASSIFICATION AND USE OF MACROMOLECULAR DATA

STRUCT\_NCS\_ENS\_GEN

STRUCT\_NCS\_OPER

*External databases* (§3.6.7.5.6)

STRUCT\_REF

STRUCT\_REF\_SEQ

STRUCT\_REF\_SEQ\_DIF

 *$\beta$ -sheets* (§3.6.7.5.7)

STRUCT\_SHEET

STRUCT\_SHEET\_TOPOLOGY

STRUCT\_SHEET\_ORDER

STRUCT\_SHEET\_RANGE

STRUCT\_SHEET\_HBOND

*Molecular sites* (§3.6.7.5.8)

STRUCT\_SITE\_GEN

STRUCT\_SITE\_KEYWORDS

STRUCT\_SITE\_VIEW

The results of the determination of a structure can be described in mmCIF using data items in the categories contained in the STRUCT category group. This is a very large group of categories and it has been divided into eight groups of related categories for the discussions that follow: (1) those that describe the structure at the level of biologically relevant assemblies; (2) those that describe the secondary structure of the macromolecules present; (3) those that describe the structural interactions that determine the conformation of the macromolecules; (4) those that describe properties of the structure at the monomer level; (5) those that describe ensembles of identical domains related by noncrystallographic symmetry; (6) those that provide references to related entities in external databases; (7) those that describe the  $\beta$ -sheets present in the structure; and (8) those that provide detailed descriptions of the structure of biologically interesting molecular sites.

## 3.6.7.5.1. Higher-level macromolecular structure

The data items in these categories are as follows:

## (a) STRUCT

- `_struct.entry_id`  
→ `_entry.id`
- `_struct.title`

## (b) STRUCT\_ASYM

- `_struct_asym.id`
- `_struct_asym.details`
- `_struct_asym.entity_id`  
→ `_entity.id`

## (c) STRUCT\_BIOL

- `_struct_biol.id`
- `_struct_biol.details`

## (d) STRUCT\_BIOL\_GEN

- `_struct_biol_gen.asym_id`  
→ `_struct_asym.id`
- `_struct_biol_gen.biol_id`  
→ `_struct_biol.id`
- `_struct_biol_gen.symmetry`
- `_struct_biol_gen.details`

## (e) STRUCT\_BIOL\_KEYWORDS

- `_struct_biol_keywords.biol_id`  
→ `_struct_biol.id`
- `_struct_biol_keywords.text`

## (f) STRUCT\_BIOL\_VIEW

- `_struct_biol_view.biol_id`  
→ `_struct_biol.id`
- `_struct_biol_view.id`
- `_struct_biol_view.details`
- `_struct_biol_view.rot_matrix[1][1]`
- `_struct_biol_view.rot_matrix[1][2]`
- `_struct_biol_view.rot_matrix[1][3]`

```
_struct_biol_view.rot_matrix[2][1]
_struct_biol_view.rot_matrix[2][2]
_struct_biol_view.rot_matrix[2][3]
_struct_biol_view.rot_matrix[3][1]
_struct_biol_view.rot_matrix[3][2]
_struct_biol_view.rot_matrix[3][3]
```

## (g) STRUCT\_KEYWORDS

- `_struct_keywords.entry_id`  
→ `_entry.id`
- `_struct_keywords.text`

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.

The data items in these categories serve two related but distinct purposes.

The first purpose is to label each of the entities in the asymmetric unit, using data items in the STRUCT\_ASYM category. These labels become part of the category key that identifies each coordinate record and they are used extensively throughout the STRUCT family of categories, so care must be taken to select a labelling scheme that is concise and informative.

The second function is descriptive. The categories descending from STRUCT\_BIOL allow the author of the mmCIF to identify and annotate the biologically relevant structural units found by the structure determination. What constitutes a biological unit can depend on the context. Take the case of a structure with two polymers related by noncrystallographic symmetry, each of which binds a small-molecule cofactor. If the author wishes to describe the dimer interface, the biological unit could be taken to be the two protein molecules. If the author wishes to highlight the cofactor binding mode, the biological unit could be taken to be one protein molecule and its bound cofactor. In this second case, there could be an additional biological unit of the second protein molecule and its bound cofactor, which may or may not be identical in conformation to the first.

The relationships between categories used to describe higher-level structure are illustrated in Fig. 3.6.7.7.

The STRUCT category serves to link the structure to the overall identifier for the data block, using `_struct.entry_id`, and to supply a title that describes the entire structure. The importance of this title as a succinct description of the structure should not be underestimated, and the author should express concisely but clearly in `_struct.title` the components of interest and the importance of this particular study. It is useful to think of this title as describing

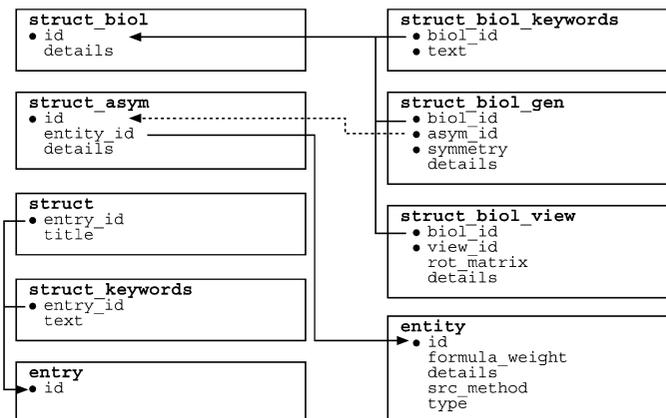


Fig. 3.6.7.7. The family of categories used to describe the higher-level macromolecular structure. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.