

3.6. CLASSIFICATION AND USE OF MACROMOLECULAR DATA

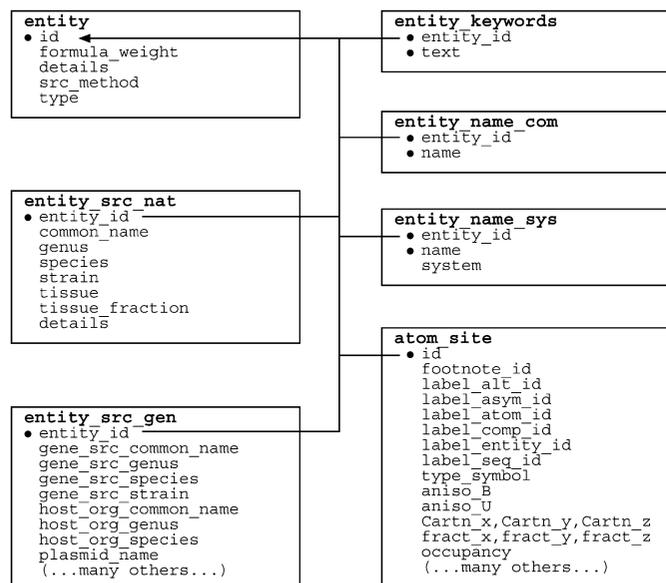


Fig. 3.6.7.5. The family of categories used to describe chemical entities. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (●). Lines show relationships between linked data items in different categories with arrows pointing at the parent data item.

a protease, but it should not be described as an $\alpha\beta$ -barrel; a number of categories within the STRUCT family allow keywords specific to the structure of the macromolecule to be given.

Data items in the ENTITY_NAME_COM category may be used to give any common names for an entity. Several different names can be recorded for each entity if appropriate.

Similarly, data items in the ENTITY_NAME_SYS category may be used to give systematic names for each entity. Again, several

Example 3.6.7.6. An example of the description of the entities in an HIV-1 protease structure (PDB 5HVP), described using data items in the ENTITY, ENTITY_NAME_COM, ENTITY_NAME_SYS and ENTITY_SRC_GEN categories.

```

loop_
  _entity.id
  _entity.type
  _entity.formula_weight
  _entity.details
  1 polymer 10916
; The enzymatically competent form of HIV protease is
a dimer. This entity corresponds to one monomer of
an active dimer.
;
  2 non-polymer 647.2 .
  3 water 18 .

loop_
  _entity_name_com.entity_id
  _entity_name_com.name
  1 'HIV-1 protease monomer'
  1 'HIV-1 PR monomer'
  2 'acetyl-pepstatin'
  2 'acetyl-Ile-Val-Asp-Statine-Ala-Ile-Statine'
  3 'water'

_entity_name_sys.entity_id 1
_entity_name_sys.name 'EC 2.1.1.1'
_entity_name_sys.system 'Enzyme convention'

loop_
  _entity_src_gen.entity_id
  _entity_src_gen.gene_src_common_name
  _entity_src_gen.gene_src_strain
  _entity_src_gen.host_org_common_name
  _entity_src_gen.host_org_genus
  _entity_src_gen.host_org_species
  _entity_src_gen.plasmid_name
  1 'HIV-1' 'NY-5' 'bacteria' 'Escherichia' 'coli'
  'pB322'
  
```

different names can be recorded for each entity if appropriate. The data item `_entity_name_sys.system` can be used to record the system according to which the systematic name was generated.

The ENTITY_SRC_GEN category allows a description of the source of entities produced by genetic manipulation to be given. There are data items for describing the tissue from which the gene was obtained, the plasmid into which it was incorporated for expression, and the host organism in which the macromolecule was expressed (Example 3.6.7.6).

The ENTITY_SRC_NAT category allows a description of the source of entities obtained from a natural tissue to be given. Data items are provided for the common and systematic name (by genus, species and, where relevant, strain) of the organism from which the material was obtained. Other data items can be used to describe the tissue (and if necessary the subcellular fraction of the tissue) from which the entity was isolated.

3.6.7.3.2. Polymer entities

The data items in these categories are as follows:

(a) ENTITY_POLY

- `_entity_poly.entity_id`
→ `_entity.id`
- `_entity_poly.nstd_chirality`
- `_entity_poly.nstd_linkage`
- `_entity_poly.nstd_monomer`
- `_entity_poly.number_of_monomers`
- `_entity_poly.type`
- `_entity_poly.type_details`

(b) ENTITY_POLY_SEQ

- `_entity_poly_seq.entity_id`
→ `_entity.id`
- `_entity_poly_seq.mon_id`
→ `_chem_comp.id`
- `_entity_poly_seq.num`
- `_entity_poly_seq.hetero`

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 polymer type, sequence length and information about any nonstandard features of the polymer may be specified using data items in the ENTITY_POLY category. The sequence of monomers in each polymer entity is given using data items in the ENTITY_POLY_SEQ category. The relationships between categories describing polymer entities are shown in Fig. 3.6.7.6, which also shows how the information describing the polymer is linked to the coordinate list in the ATOM_SITE category and to the full chemical description of each monomer or nonstandard monomer in the CHEM_COMP category.

Non-polymer entities are treated as individual chemical components, in the same way in which monomers within a polymer are treated as individual chemical components. They may be fully described in the CHEM_COMP group of categories (Example 3.6.7.7).

Data items in the ENTITY_POLY category can be used to give the number of monomers in the polymer and to assign the type of the polymer as one of the set of types polypeptide (D), polypeptide (L), polydeoxyribonucleotide, polyribonucleotide, polysaccharide (D), polysaccharide (L) or other. Details of deviations from a standard type may be given in `_entity_poly.type_details`.

In some cases, the polymer is best described as one of the standard types even if it contains some nonstandard features. Flags are provided to indicate the presence of three types of nonstandard features. The presence of chiral centres other than those implied