

3.6. CLASSIFICATION AND USE OF MACROMOLECULAR DATA

- `_struct_ncs_dom_lim.beg_label_seq_id`
→ `_atom_site.label_seq_id`
- `_struct_ncs_dom_lim.dom_id`
- `_struct_ncs_dom_lim.end_label_alt_id`
→ `_atom_sites.alt.id`
- `_struct_ncs_dom_lim.end_label_asym_id`
→ `_atom_site.label_asym_id`
- `_struct_ncs_dom_lim.end_label_comp_id`
→ `_atom_site.label_comp_id`
- `_struct_ncs_dom_lim.end_label_seq_id`
→ `_atom_site.label_seq_id`
- `_struct_ncs_dom_lim.beg_auth_asym_id`
→ `_atom_site.auth_asym_id`
- `_struct_ncs_dom_lim.beg_auth_comp_id`
→ `_atom_site.auth_comp_id`
- `_struct_ncs_dom_lim.beg_auth_seq_id`
→ `_atom_site.auth_seq_id`
- `_struct_ncs_dom_lim.end_auth_asym_id`
→ `_atom_site.auth_asym_id`
- `_struct_ncs_dom_lim.end_auth_comp_id`
→ `_atom_site.auth_comp_id`
- `_struct_ncs_dom_lim.end_auth_seq_id`
→ `_atom_site.auth_seq_id`

(e) STRUCT_NCS_OPER

- `_struct_ncs_oper.id`
- `_struct_ncs_oper.code`
- `_struct_ncs_oper.details`
- `_struct_ncs_oper.matrix[1][1]`
- `_struct_ncs_oper.matrix[1][2]`
- `_struct_ncs_oper.matrix[1][3]`
- `_struct_ncs_oper.matrix[2][1]`
- `_struct_ncs_oper.matrix[2][2]`
- `_struct_ncs_oper.matrix[2][3]`
- `_struct_ncs_oper.matrix[3][1]`
- `_struct_ncs_oper.matrix[3][2]`
- `_struct_ncs_oper.matrix[3][3]`
- `_struct_ncs_oper.vector[1]`
- `_struct_ncs_oper.vector[2]`
- `_struct_ncs_oper.vector[3]`

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.

Biological macromolecular complexes may be built from domains related by symmetry transformations other than those arising from the crystal lattice symmetry. These domains are not necessarily discrete molecular entities: they may be composed of one or more segments of a single polypeptide or nucleic acid chain, of segments from more than one chain, or of small-molecule components of the structure. The categories above allow the distinct domains that participate in ensembles of structural elements related by noncrystallographic symmetry to be listed and described in detail. The relationships between categories used to describe noncrystallographic symmetry are shown in Fig. 3.6.7.12.

In the mmCIF model of noncrystallographic symmetry, the highest level of organization is the ensemble, which corresponds to the complete symmetry-related aggregate (e.g. tetramer, icosahedron). An identifier is given to the ensemble using the data item `_struct_ncs_ens.id`.

The symmetry-related elements within the ensemble are referred to as domains. The elements of structure that are to be considered part of the domain are specified using the data items in the `STRUCT_NCS_DOM` and `STRUCT_NCS_DOM_LIM` categories. By using the `STRUCT_NCS_DOM_LIM` data items appropriately, domains can be defined to include ranges of polypeptide chain or nucleic acid strand, bound ligands or cofactors, or even bound solvent molecules. Note that the category keys for `STRUCT_NCS_DOM_LIM` include the domain ID and the range specifiers. Thus a single domain may be composed of any number of ranges of elements.

Finally, the ensemble is generated from the domains using the rotation matrix and translation vector specified by data items in

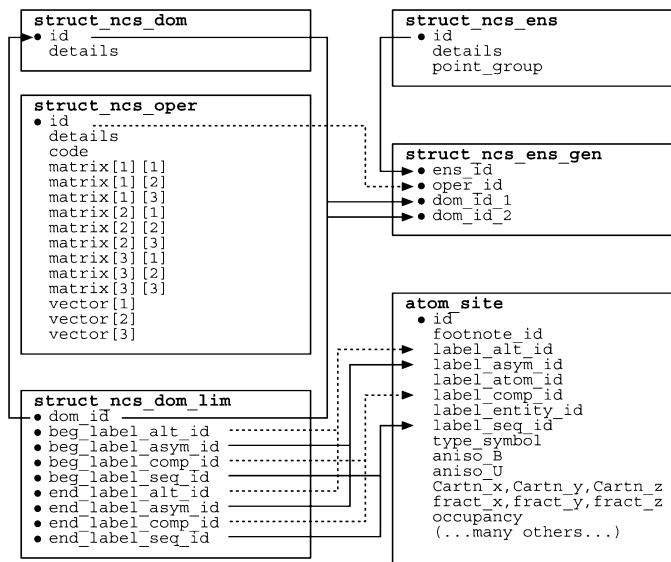


Fig. 3.6.7.12. The family of categories used to describe noncrystallographic symmetry. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

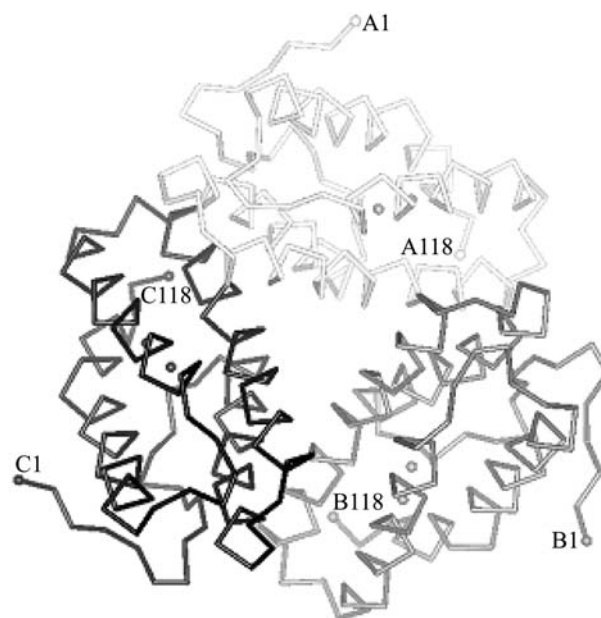


Fig. 3.6.7.13. Noncrystallographic symmetry in the structure of trimeric haemerythrin (PDB 1HR3) to be described with data items in the `STRUCT_NCS_ENS`, `STRUCT_NCS_ENS_GEN`, `STRUCT_NCS_DOM` and `STRUCT_NCS_DOM_LIM` categories.

the `STRUCT_NCS_OPER` category, which are referenced by the data items in the `STRUCT_NCS_ENS_GEN` category. There are data items appropriate for two common methods of describing noncrystallographic symmetry:

(1) In the first method, the coordinate list includes all copies of domains related by noncrystallographic symmetry and the aim is to describe the relationships between domains in the ensemble; in this case the data items in `STRUCT_NCS_ENS_GEN` specify a pair of domains and reference the appropriate operator in `STRUCT_NCS_OPER`. This method is indicated by giving the data item `_struct_ncs_oper.code` the value given.

(2) In the second method, the coordinate list contains only one copy of the domain and the aim is to generate the entire ensemble; in this case the data items in `STRUCT_NCS_ENS_GEN`