The allowed values for the data item **_struct_conf_type.id** cover most types of protein and nucleic acid secondary structure (Example 3.6.7.9). The criteria that define the secondary structure may be given using the data item **_struct_conf_type.criteria**. **_struct_conf_type.reference** can be used to specify a reference to the literature in which the criteria are explained in more detail.

The residues that define the beginning and end of each region of secondary structure are identified with the appropriate ***_asym**, ***_comp** and ***_seq** identifiers. The standard labelling system or the author's alternative labelling system may be used. The identification of the residues assigned to each region of secondary structure is linked to the labelling information in the ATOM_SITE category. Unusual features of a conformation may be described using **_struct_conf.details**.

3.6.7.5.3. *Structural interactions*

The data items in these categories are as follows:

(*a*) STRUCT_CONN_TYPE

- **_struct_conn_type.id**
  _struct_conn_type.criteria
  _struct_conn_type.reference

(*b*) STRUCT_CONN

- **_struct_conn.id**
  _struct_conn.conn_type_id
  → **_struct_conn_type.id**
  _struct_conn.details
  _struct_conn.ptnr1_label_alt_id
  → **_atom_sites_alt.id**
  _struct_conn.ptnr1_label_asym_id
  → **_atom_site.label_asym_id**
  _struct_conn.ptnr1_label_atom_id
  → **_chem_comp_atom.atom_id**
  _struct_conn.ptnr1_label_comp_id
  → **_atom_site.label_comp_id**
  _struct_conn.ptnr1_label_seq_id
  → **_atom_site.label_seq_id**
  _struct_conn.ptnr1_auth_asym_id
  → **_atom_site.auth_asym_id**
  _struct_conn.ptnr1_auth_atom_id
  → **_atom_site.auth_atom_id**
  _struct_conn.ptnr1_auth_comp_id
  → **_atom_site.auth_comp_id**
  _struct_conn.ptnr1_auth_seq_id
  → **_atom_site.auth_seq_id**
  _struct_conn.ptnr1_role
  _struct_conn.ptnr1_symmetry
  _struct_conn.ptnr2_label_alt_id
  → **_atom_sites_alt.id**
  _struct_conn.ptnr2_label_asym_id
  → **_atom_site.label_asym_id**
  _struct_conn.ptnr2_label_atom_id
  → **_chem_comp_atom.atom_id**
  _struct_conn.ptnr2_label_comp_id
  → **_atom_site.label_comp_id**
  _struct_conn.ptnr2_label_seq_id
  → **_atom_site.label_seq_id**
  _struct_conn.ptnr2_auth_asym_id
  → **_atom_site.auth_asym_id**
  _struct_conn.ptnr2_auth_atom_id
  → **_atom_site.auth_atom_id**
  _struct_conn.ptnr2_auth_comp_id
  → **_atom_site.auth_comp_id**
  _struct_conn.ptnr2_auth_seq_id
  → **_atom_site.auth_seq_id**
  _struct_conn.ptnr2_role
  _struct_conn.ptnr2_symmetry

The bullet (•) indicates a category key. The arrow (→) is a reference to a parent data item.

The structural interactions that are described with data items in the STRUCT_CONN family of categories are the tertiary result of a structure determination, not the chemical connectivity of the components of the structure. In general, the interactions described using the STRUCT_CONN data items are noncovalent, such as hydrogen bonds, salt bridges and metal coordination.

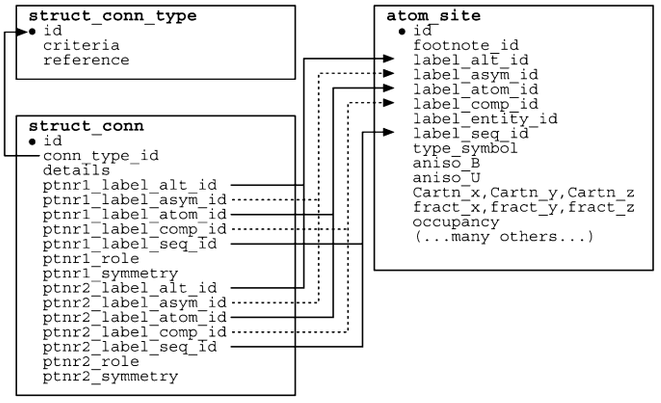

Fig. 3.6.7.10. The family of categories used to describe structural interactions such as hydrogen bonding, salt bridges and disulfide bridges. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

It is useful to think of the structure interactions given in CHEM_COMP_BOND, CHEM_LINK and ENTITY_LINK as the covalent interactions that are known in advance of the structure determination because the chemistry of the components is well defined. Literature or calculated values for these interactions are often used as restraints during the refinement. In contrast, the structural interactions described in the STRUCT_CONN family of categories are not known in advance and are part of the results of the structure determination.

This distinction only holds approximately, as there are clearly bonds, such as disulfide links, that are covalent and usually restrained during the refinement but that are also a result of the folding of the protein revealed by the structure determination, and thus should be described using STRUCT_CONN data items.

In general, the STRUCT_CONN data items would not be used to list all the structure interactions. Instead, the author of the mmCIF would use the STRUCT_CONN data items to identify and annotate only the structural interactions worthy of discussion. The relationships between categories used to describe structural interactions are shown in Fig. 3.6.7.10.

Structural interactions such as hydrogen bonds, salt bridges and disulfide bridges can be described in the STRUCT_CONN category. The type of each interaction and the criteria used to identify the interaction can be specified in the STRUCT_CONN_TYPE category (Example 3.6.7.10).

The atoms participating in each interaction are arbitrarily labelled as 'partner 1' and 'partner 2'. Each is identified by the ***_alt**, ***_asym**, ***_atom**, ***_comp** and ***_seq** constituents of the corresponding atom-site label. The role of each partner in the interaction (*e.g.* donor, acceptor) may be specified, and any crystallographic symmetry operation needed to transform the atom from the position given in the ATOM_SITE list to the position where the interaction occurs can be given. The atoms participating in the interaction may also be identified using an alternative labelling scheme if the author has supplied one.

Unusual aspects of the interaction may be discussed in **_struct_conn.details**. The general type of an interaction can be indicated using **_struct_conn.conn_type_id**, which references one of the standard types described using data items in the STRUCT_CONN_TYPE category.

The specific types of structural connection that may be recorded are those allowed for **_struct_conn_type.id**, namely covalent and hydrogen bonds, ionic (salt-bridge) interactions, disulfide

Example 3.6.7.10. *A hypothetical salt bridge and hydrogen bond described with data items in the* STRUCT_CONN_TYPE *and* STRUCT_CONN *categories.*

```
loop_
_struct_conn_type.id
_struct_conn_type.criteria
   saltbr
; negative to positive distance > 2.5 Angstroms,
  < 3.2 Angstroms
;
   hydrog
; N-O distance > 2.5 Angstroms, < 3.5 Angstroms,
  N-O-C  angle < 120 degrees
;

loop_
_struct_conn.id
_struct_conn.conn_type_id
_struct_conn.ptnr1_label_comp_id
_struct_conn.ptnr1_label_asym_id
_struct_conn.ptnr1_label_seq_id
_struct_conn.ptnr1_label_atom_id
_struct_conn.ptnr1_role
_struct_conn.ptnr1_symmetry
_struct_conn.ptnr2_label_comp_id
_struct_conn.ptnr2_label_asym_id
_struct_conn.ptnr2_label_seq_id
_struct_conn.ptnr2_label_atom_id
_struct_conn.ptnr2_role
_struct_conn.ptnr2_symmetry
  C1 saltbr ARG  A  87 NZ1 positive 1_555
            GLU  A  92 OE1 negative 1_555
  C2 hydrog ARG  B 287 N   donor    1_555
            GLY  B 292 O   acceptor 1_555
```

links, metal coordination, mismatched base pairs, covalent residue modifications and covalent modifications of nucleotide bases, sugars or phosphates. The criteria used to define each interaction may be described in detail using `_struct_conn_type.criteria` or a literature reference to the criteria can be given in `_struct_conn_type.reference`.

3.6.7.5.4. *Structural features of monomers*

The data items in these categories are as follows:

(*a*) STRUCT_MON_DETAILS

- `_struct_mon_details.entry_id`
  → `_entry.id`

`_struct_mon_details.prot_cis`
`_struct_mon_details.RSCC`
`_struct_mon_details.RSR`

(*b*) STRUCT_MON_NUCL

- `_struct_mon_nucl.label_alt_id`
  → `_atom_sites_alt.id`
- `_struct_mon_nucl.label_asym_id`
  → `_atom_site.label_asym_id`
- `_struct_mon_nucl.label_comp_id`
  → `_atom_site.label_comp_id`
- `_struct_mon_nucl.label_seq_id`
  → `_atom_site.label_seq_id`

`_struct_mon_nucl.alpha`
`_struct_mon_nucl.auth_asym_id`
  → `_atom_site.auth_asym_id`
`_struct_mon_nucl.auth_comp_id`
  → `_atom_site.auth_comp_id`
`_struct_mon_nucl.auth_seq_id`
  → `_atom_site.auth_seq_id`
`_struct_mon_nucl.beta`
`_struct_mon_nucl.chi1`
`_struct_mon_nucl.chi2`
`_struct_mon_nucl.delta`
`_struct_mon_nucl.details`
`_struct_mon_nucl.epsilon`
`_struct_mon_nucl.gamma`
`_struct_mon_nucl.mean_B_all`
`_struct_mon_nucl.mean_B_base`
`_struct_mon_nucl.mean_B_phos`
`_struct_mon_nucl.mean_B_sugar`
`_struct_mon_nucl.nu0`
`_struct_mon_nucl.nu1`
`_struct_mon_nucl.nu2`
`_struct_mon_nucl.nu3`
`_struct_mon_nucl.nu4`
`_struct_mon_nucl.P`
`_struct_mon_nucl.RSCC_all`
`_struct_mon_nucl.RSCC_base`
`_struct_mon_nucl.RSCC_phos`
`_struct_mon_nucl.RSCC_sugar`
`_struct_mon_nucl.RSR_all`
`_struct_mon_nucl.RSR_base`
`_struct_mon_nucl.RSR_phos`
`_struct_mon_nucl.RSR_sugar`
`_struct_mon_nucl.tau0`
`_struct_mon_nucl.tau1`
`_struct_mon_nucl.tau2`
`_struct_mon_nucl.tau3`
`_struct_mon_nucl.tau4`
`_struct_mon_nucl.taum`
`_struct_mon_nucl.zeta`

(*c*) STRUCT_MON_PROT

- `_struct_mon_prot.label_alt_id`
  → `_atom_sites_alt.id`
- `_struct_mon_prot.label_asym_id`
  → `_atom_site.label_asym_id`
- `_struct_mon_prot.label_comp_id`
  → `_atom_site.label_comp_id`
- `_struct_mon_prot.label_seq_id`
  → `_atom_site.label_seq_id`

`_struct_mon_prot.auth_asym_id`
  → `_atom_site.auth_asym_id`
`_struct_mon_prot.auth_comp_id`
  → `_atom_site.auth_comp_id`
`_struct_mon_prot.auth_seq_id`
  → `_atom_site.auth_seq_id`
`_struct_mon_prot.chi1`
`_struct_mon_prot.chi2`
`_struct_mon_prot.chi3`
`_struct_mon_prot.chi4`
`_struct_mon_prot.chi5`
`_struct_mon_prot.details`
`_struct_mon_prot.RSCC_all`
`_struct_mon_prot.RSCC_main`
`_struct_mon_prot.RSCC_side`
`_struct_mon_prot.RSR_all`
`_struct_mon_prot.RSR_main`
`_struct_mon_prot.RSR_side`
`_struct_mon_prot.mean_B_all`
`_struct_mon_prot.mean_B_main`
`_struct_mon_prot.mean_B_side`
`_struct_mon_prot.omega`
`_struct_mon_prot.phi`
`_struct_mon_prot.psi`

(*d*) STRUCT_MON_PROT_CIS

- `_struct_mon_prot_cis.label_alt_id`
  → `_atom_sites_alt.id`
- `_struct_mon_prot_cis.label_asym_id`
  → `_atom_site.label_asym_id`
- `_struct_mon_prot_cis.label_comp_id`
  → `_atom_site.label_comp_id`
- `_struct_mon_prot_cis.label_seq_id`
  → `_atom_site.label_seq_id`

`_struct_mon_prot_cis.auth_asym_id`
  → `_atom_site.auth_asym_id`
`_struct_mon_prot_cis.auth_comp_id`
  → `_atom_site.auth_comp_id`
`_struct_mon_prot_cis.auth_seq_id`
  → `_atom_site.auth_seq_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.

Most macromolecules have complex structures which contain regions of well defined structure and flexible regions that are difficult to model accurately. Overall measures of the quality of a model, such as the standard crystallographic *R* factors, do not represent the local quality of the model. During the development of