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

Copyright © 2006 International Union of Crystallography

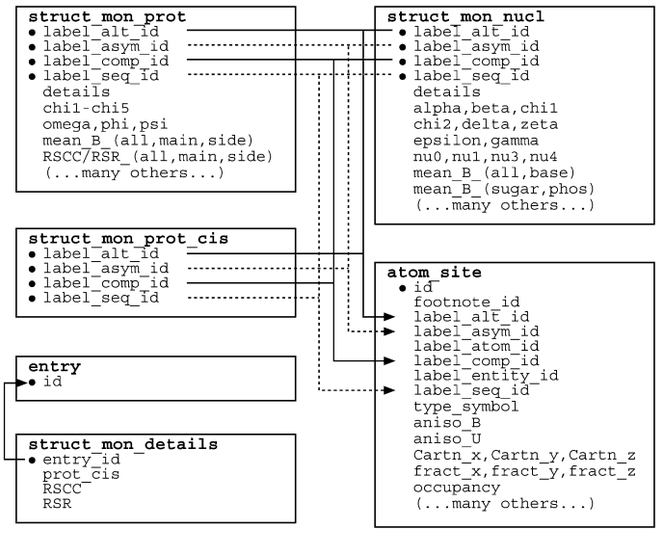

Fig. 3.6.7.11. The family of categories used to describe the structural features of monomers. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

Example 3.6.7.11. *A hypothetical example of the structural features of a single protein residue described with data items in the* STRUCT_MON_PROT *category.*

```
_struct_mon_prot.label_comp_id          ARG
_struct_mon_prot.label_seq_id           35
_struct_mon_prot.label_asym_id          A
_struct_mon_prot.label_alt_id           .
_struct_mon_prot.chi1                 -67.9
_struct_mon_prot.chi2                -174.7
_struct_mon_prot.chi3                 -67.7
_struct_mon_prot.chi4                 -86.3
_struct_mon_prot.chi5                   4.2
_struct_mon_prot.RSCC_all               0.90
_struct_mon_prot.RSR_all                0.18
_struct_mon_prot.mean_B_all            30.0
_struct_mon_prot.mean_B_main           25.0
_struct_mon_prot.mean_B_side           35.1
_struct_mon_prot.omega                180.1
_struct_mon_prot.phi                  -60.3
_struct_mon_prot.psi                  -46.0
```

the mmCIF dictionary, it was found that the biological crystallography community felt that mmCIF should contain data items that allowed the local quality of the model to be recorded: these data items are found in the categories STRUCT_MON_DETAILS, STRUCT_MON_NUCL (for nucleotides), and STRUCT_MON_PROT and STRUCT_MON_PROT_CIS (for proteins). Using these categories, quantities that reflect the local quality of the structure, such as isotropic displacement factors, real-space *R* factors and real-space correlation coefficients, can be given at the monomer and submonomer levels.

In addition, these categories can be used to record the conformation of the structure at the monomer level by listing side-chain torsion angles. These values can be derived from the atom coordinate list, so it would not be common practice to include them in an mmCIF for archiving a structure unless it was to highlight conformations that deviate significantly from expected values (Engh & Huber, 1991). However, there are applications, such as comparative studies across a number of independent determinations of the same structure, where it would be useful to store torsion-angle information without having to recalculate it each time it is needed.

The relationships between the categories used to describe the structural features of monomers are shown in Fig. 3.6.7.11.

Three indicators of the quality of a structure at the local level are included in this version of the dictionary: the mean displacement (*B*) factor, the real-space correlation coefficient (Jones *et al.*, 1991) and the real-space *R* factor (Brändén & Jones, 1990). Other indicators are likely to be added as they become available. In the current version of the dictionary, these metrics can be given at the monomer level, or at the levels of main- and side-chain for proteins, or base, phosphate and sugar for nucleic acids (Altona & Sundaralingam, 1972).

The variables used when calculating real-space correlation coefficients and real-space *R* factors, such as the coefficients used to calculate the map being evaluated or the radii used for including points in a calculation, can be recorded using the data items `_struct_mon_details.RSC` and `_struct_mon_details.RSR`.

These data items are also provided for recording the full conformation of the macromolecule, using a full set of data items for the torsion angles of both proteins and nucleic acids. Although one could use these data items to describe the whole macromolecule, it is more likely that they would be used to highlight regions of the structure that deviate from expected values (Example 3.6.7.11). Deviations from expected values could imply inaccuracies in the model in poorly defined parts of the structure, but in some cases nonstandard torsion angles are found in very well defined regions and are essential to the proper configurations of active sites or ligand binding pockets.

A special case of nonstandard conformation is the occurrence of *cis* peptides in proteins. As the *cis* conformation occurs quite often, the category STRUCT_MON_PROT_CIS is provided so that an explicit list can be made of *cis* peptides. The related data item `_struct_mon_details.prot_cis` allows an author to specify how far a peptide torsion angle can deviate from the expected value of 0.0 and still be considered to be *cis*.

In these categories, properties are listed by residue rather than by individual atom. The only label components needed to identify the residue are `*_alt`, `*_asym`, `*_comp` and `*_seq`. If the author has provided an alternative labelling system, this can also be used. Since the analysis is by individual residue, there is no need to specify symmetry operations that might be needed to move one residue so that it is next to another.

3.6.7.5.5. *Noncrystallographic symmetry*

Data items in these categories are as follows:

(*a*) STRUCT_NCS_ENS

- `_struct_ncs_ens.id`

  `_struct_ncs_ens.details`

  `_struct_ncs_ens.point_group`

(*b*) STRUCT_NCS_ENS_GEN

- `_struct_ncs_ens_gen.dom_id_1`
  → `_struct_ncs_dom.id`
- `_struct_ncs_ens_gen.dom_id_2`
  → `_struct_ncs_dom.id`
- `_struct_ncs_ens_gen.ens_id`
  → `_struct_ncs_ens.id`
- `_struct_ncs_ens_gen.oper_id`
  → `_struct_ncs_oper.id`

(*c*) STRUCT_NCS_DOM

- `_struct_ncs_dom.id`

  `_struct_ncs_dom.details`

(*d*) STRUCT_NCS_DOM_LIM

- `_struct_ncs_dom_lim.beg_label_alt_id`
  → `_atom_sites_alt.id`
- `_struct_ncs_dom_lim.beg_label_asym_id`
  → `_atom_site.label_asym_id`
- `_struct_ncs_dom_lim.beg_label_comp_id`
  → `_atom_site.label_comp_id`