Example 3.6.7.4. *The description of a component (adriamycin) of a macromolecule with data items in the* CHEM_COMP, CHEM_COMP_ATOM, CHEM_COMP_BOND, CHEM_COMP_TOR *and* CHEM_COMP_TOR_VALUE *categories (Leonard et al., 1993).*

```
_chem_comp.id                 'DM2'
_chem_comp.name               'adriamycin'
_chem_comp.type                non-polymer
_chem_comp.formula            'C27 H29 N1 O11'
_chem_comp.number_atoms_all  68
_chem_comp.number_atoms_nh   39
_chem_comp.formula_weight    543.51

loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.model_Cartn_x
_chem_comp_atom.model_Cartn_y
_chem_comp_atom.model_Cartn_z
   DM2 'C1'   C  12.996  0.476  12.694
   DM2 'C2'   C  13.982 -0.225  13.183
   DM2 'C3'   C  12.482  0.165  11.515
# - - - abbreviated - - -

loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.value_order
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
   DM2 'C1' 'C2' sing  1.517  0.0210
   DM2 'C2' 'C3' sing  1.445  0.0040
# - - - abbreviated - - -

loop_
  _chem_comp_tor.comp_id
  _chem_comp_tor.id
  _chem_comp_tor.atom_id_1
  _chem_comp_tor.atom_id_2
  _chem_comp_tor.atom_id_3
  _chem_comp_tor.atom_id_4
    phe  phe_chi1   N    CA   CB   CG
    phe  phe_chi2   CA   CB   CG   CD1
    phe  phe_ring1  CB   CG   CD1  CE1
    phe  phe_ring2  CB   CG   CD2  CE2
    phe  phe_ring3  CG   CD1  CE1  CZ
    phe  phe_ring4  CD1  CE1  CZ   CE2
    phe  phe_ring5  CE1  CZ   CE2  CD2

loop_
  _chem_comp_tor_value.tor_id
  _chem_comp_tor_value.comp_id
  _chem_comp_tor_value.angle
  _chem_comp_tor_value.dist
    phe_chi1   phe  -60.0  2.88
    phe_chi1   phe  180.0  3.72
    phe_chi1   phe   60.0  2.88
    phe_chi2   phe   90.0  3.34
    phe_chi2   phe  -90.0  3.34
    phe_ring1  phe  180.0  3.75
    phe_ring2  phe  180.0  3.75
    phe_ring3  phe    0.0  2.80
    phe_ring4  phe    0.0  2.80
    phe_ring5  phe    0.0  2.80
```

restraint is used in refinement, where the value of the angle is assumed to be close to the target value.) As torsion angles can have more than one target value, the target values are specified in the CHEM_COMP_TOR_VALUE category.

Data items in the CHEM_COMP_LINK category can be used to provide a table of links between the components of the structure. Each link is assigned an identifier (`_chem_comp_link.link_id`) and the types of monomer at each end of the link are stated. The types are those allowed for the parent data item `_chem_comp.type`.

The use of many of these data items to describe a typical component is shown in Example 3.6.7.4.

3.6.7.2.3. *Chemical links*

The data items in these categories are as follows:

(*a*) CHEM_LINK

- `_chem_link.id`

  `_chem_link.details`

(*b*) CHEM_LINK_ANGLE

- `_chem_link_angle.atom_id_1`
- `_chem_link_angle.atom_id_2`
- `_chem_link_angle.atom_id_3`
- `_chem_link_angle.link_id`
  → `_chem_link.id`

  `_chem_link_angle.atom_1_comp_id`

  `_chem_link_angle.atom_2_comp_id`

  `_chem_link_angle.atom_3_comp_id`

+ `_chem_link_angle.value_angle`
+ `_chem_link_angle.value_dist`

(*c*) CHEM_LINK_BOND

- `_chem_link_bond.atom_id_1`
- `_chem_link_bond.atom_id_2`
- `_chem_link_bond.link_id`
  → `_chem_link.id`

  `_chem_link_bond.atom_1_comp_id`

  `_chem_link_bond.atom_2_comp_id`

+ `_chem_link_bond.value_dist`

  `_chem_link_bond.value_order`

(*d*) CHEM_LINK_CHIR

- `_chem_link_chir.id`
- `_chem_link_chir.link_id`
  → `_chem_link.id`

  `_chem_link_chir.atom_comp_id`

  `_chem_link_chir.atom_id`

  `_chem_link_chir.atom_config`

  `_chem_link_chir.number_atoms_all`

  `_chem_link_chir.number_atoms_nh`

  `_chem_link_chir.volume_flag`

+ `_chem_link_chir.volume_three`

(*e*) CHEM_LINK_CHIR_ATOM

- `_chem_link_chir_atom.atom_id`
- `_chem_link_chir_atom.chir_id`
  → `_chem_link_chir.id`

  `_chem_link_chir_atom.atom_comp_id`

  `_chem_link_chir_atom.dev`

(*f*) CHEM_LINK_PLANE

- `_chem_link_plane.id`
- `_chem_link_plane.link_id`
  → `_chem_link.id`

  `_chem_link_plane.number_atoms_all`

  `_chem_link_plane.number_atoms_nh`

(*g*) CHEM_LINK_PLANE_ATOM

- `_chem_link_plane_atom.atom_id`
- `_chem_link_plane_atom.plane_id`
  → `_chem_link_plane.id`

  `_chem_link_plane_atom.atom_comp_id`

(*h*) CHEM_LINK_TOR

- `_chem_link_tor.id`
- `_chem_link_tor.link_id`
  → `_chem_link.id`

  `_chem_link_tor.atom_1_comp_id`

  `_chem_link_tor.atom_2_comp_id`

  `_chem_link_tor.atom_3_comp_id`

  `_chem_link_tor.atom_4_comp_id`

  `_chem_link_tor.atom_id_1`

  `_chem_link_tor.atom_id_2`

  `_chem_link_tor.atom_id_3`

  `_chem_link_tor.atom_id_4`

(*i*) CHEM_LINK_TOR_VALUE

- `_chem_link_tor_value.tor_id`
  → `_chem_link_tor.id`

+ `_chem_link_tor_value.angle`
+ `_chem_link_tor_value.dist`

(*j*) ENTITY_LINK

```
• _entity_link.link_id
        → _chem_link.id
  _entity_link.details
  _entity_link.entity_id_1
        → _entity.id
  _entity_link.entity_id_2
        → _entity.id
  _entity_link.entity_seq_num_1
        → _entity_poly_seq.num
  _entity_link.entity_seq_num_2
        → _entity_poly_seq.num
```

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. Data items marked with a plus (+) have companion data names for the standard uncertainty in the reported value, formed by appending the string ***_esd*** *to the data name listed.*

The geometry of the links between chemical components or entities can be described in the CHEM_LINK group of categories. Chemical components may be linked together according to the type of the component; defining the linking according to the type of the component rather than by each component in turn allows a type of polymer link for all the monomers in a polymer to be specified (*e.g.* L-peptide linking). The geometry of the links can be specified in the remaining CHEM_LINK categories. The relationships between categories used to describe links between chemical components are shown in Fig. 3.6.7.4, which also shows how information about the links is passed to the CHEM_COMP and CHEM_LINK categories. For simplicity, the categories CHEM_COMP_PLANE, CHEM_COMP_PLANE_ATOM, CHEM_COMP_CHIR, CHEM_COMP_CHIR_ATOM and ENTITY_LINK are not included in Fig. 3.6.7.4.

Note that this category group can be used to describe the links that connect the monomers within a macromolecular polymer (using the CHEM_LINK categories) and also the intramolecular links between separate molecules in the whole complex (using the ENTITY_LINK category). Intramolecular links, for example a covalent bond formed between a bound ligand and an amino-acid side chain, are usually discovered as a result of the structure determination, and it would therefore seem more appropriate to describe them in the STRUCT_CONN category. However, since one of the roles of the CHEM_LINK category group is to record target values used for restraints or constraints during the refinement of the model of the structure, ideal values for the geometry of any entity-to-entity links should be given here.

Data items in the CHEM_LINK category are used to assign a unique identifier to each link and allow the author to record any unusual aspects of each link. The other categories in the CHEM_LINK category group describe the geometric model of each link, and are closely analogous to the similarly named categories in the CHEM_COMP group.

The relationships among these categories are complex (see Fig. 3.6.7.4). Each atom that participates in an aspect of the link (for example, a bond, an angle, a chiral centre, a torsion angle or a plane) must be identified and it must also be specified whether the atom is in the first or second of the components that form the link.

Data items in the CHEM_LINK_BOND category describe the bonds between atoms participating in an intermolecular link between chemical components. Bond restraints may be described by the distance between the bonded atoms, the bond order or both.

An angle at a link may be described in the CHEM_LINK_ANGLE category as either an angle at the vertex atom or as a distance between the atoms attached to the vertex. For data items in both the CHEM_LINK_BOND and CHEM_LINK_ANGLE categories, a target value and its associated standard uncertainty may be specified (Example 3.6.7.5).

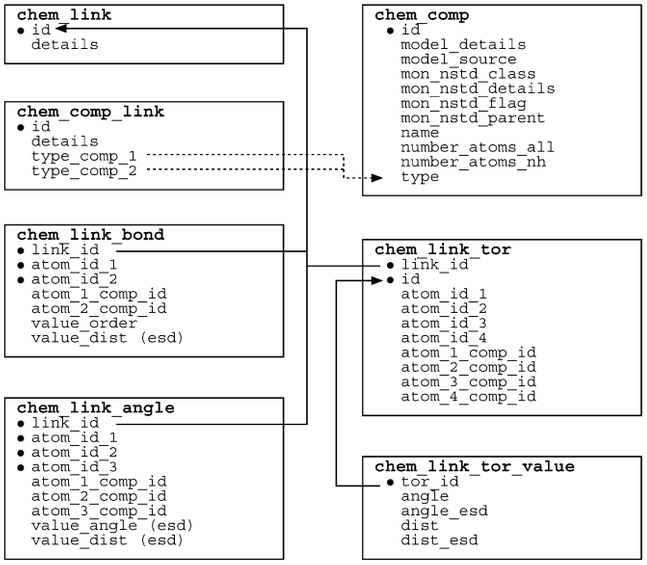

Fig. 3.6.7.4. The family of categories used to describe the links between chemical components. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

Data items in the CHEM_LINK_CHIR category can be used to describe the conformation of chiral centres in a link between two chemical components. The absolute configuration and the chiral volume may be specified, as well as the total number of atoms and the number of non-hydrogen atoms bonded to the chiral centre. There is also a flag to indicate whether a restrained chiral volume should match the target value in sign as well as in magnitude. Because chiral centres can involve a variable number of atoms, a separate list of the atoms should be given in CHEM_LINK_CHIR_ATOM.

Data items in the CHEM_LINK_PLANE category can be used to list planes defined across a link between two chemical components. Because planes can involve a variable number of atoms, a separate list of the atoms should be given in CHEM_LINK_PLANE_ATOM.

Data items in the CHEM_LINK_TOR category can be used to give details of the torsion angles across a link between two chemical

Example 3.6.7.5. *A peptide bond described with data items in the CHEM_LINK_BOND and CHEM_LINK_ATOM categories.*

```
loop_
  _chem_link_bond.link_id
  _chem_link_bond.value_dist
  _chem_link_bond.value_dist_esd
  _chem_link_bond.atom_id_1
  _chem_link_bond.atom_1_comp_id
  _chem_link_bond.atom_id_2
  _chem_link_bond.atom_2_comp_id
    PEPTIDE  1.329  0.014  C  1  N  2

loop_
  _chem_link_angle.link_id
  _chem_link_angle.value_angle
  _chem_link_angle.value_angle_esd
  _chem_link_angle.atom_id_1
  _chem_link_angle.atom_1_comp_id
  _chem_link_angle.atom_id_2
  _chem_link_angle.atom_2_comp_id
  _chem_link_angle.atom_id_3
  _chem_link_angle.atom_3_comp_id
    PEPTIDE  116.2  2.0  CA 1  C  1  N  2
    PEPTIDE  123.0  1.6  O  1  C  1  N  2
    PEPTIDE  121.7  1.8  C  1  N  2  CA 2
```

components. The torsion angle may be described either as an angle or as a distance between the first and last atoms. As torsion angles can have more than one target value, the target values are specified in the CHEM_LINK_TOR_VALUE category.

The ENTITY_LINK category is used to identify the participants in links between distinct molecular entities. A pointer to the details of the link is given in **_entity_link.link_id**, which matches a value of **_chem_link.id** in the CHEM_LINK category.

3.6.7.3. Distinct chemical species

The categories describing distinct chemical entities are as follows:

ENTITY group
Entities (§3.6.7.3.1)
ENTITY
ENTITY_KEYWORDS
ENTITY_NAME_COM
ENTITY_NAME_SYS
ENTITY_SRC_GEN
ENTITY_SRC_NAT
Polymer entities (§3.6.7.3.2)
ENTITY_POLY
ENTITY_POLY_SEQ

The ENTITY categories of the mmCIF dictionary should be used in preference to the CHEMICAL categories of the core CIF dictionary. In a typical small-molecule structure determination, for which the core CIF dictionary was designed, the substance being studied can be thought of as a single chemical species, even if it contains distinct ions or ligands. In a macromolecular structure, it is more often the case that separate descriptions are appropriate for each of the distinct chemical species that comprise the structural complex. The ENTITY categories allow the species present and their basic chemical properties to be specified. Their structures and connectivity are described in other categories.

It is important, therefore, to remember that the ENTITY data do not represent the result of the crystallographic experiment; those results are given using the ATOM_SITE data items and are discussed and described using data items in the STRUCT family of categories. The ENTITY categories describe the chemistry of the molecules under investigation and are most usefully considered as the ideal groups to which the structure is restrained or constrained during refinement.

It is also important to remember that entities do not correspond directly to the total contents of the asymmetric unit. Entities are described only once, even in structures in which the entity occurs several times. The STRUCT_ASYM data items, which reference the list of entities, describe and label the contents of the asymmetric unit.

The following discussion treats the data items used for entities in general (Section 3.6.7.3.1) and those used more specifically to describe polymeric entities (Section 3.6.7.3.2) separately.

3.6.7.3.1. *Description of entities*

The data items in these categories are as follows:

(*a*) ENTITY
• **_entity.id**
_entity.details
_entity.formula_weight
_entity.src_method
_entity.type

(*b*) ENTITY_KEYWORDS
• **_entity_keywords.entity_id**
→ **_entity.id**
• **_entity_keywords.text**

(*c*) ENTITY_NAME_COM
• **_entity_name_com.entity_id**
→ **_entity.id**
• **_entity_name_com.name**

(*d*) ENTITY_NAME_SYS
• **_entity_name_sys.entity_id**
→ **_entity.id**
• **_entity_name_sys.name**
_entity_name_sys.system

(*e*) ENTITY_SRC_GEN
• **_entity_src_gen.entity_id**
→ **_entity.id**
_entity_src_gen.gene_src_common_name
_entity_src_gen.gene_src_details
_entity_src_gen.gene_src_genus
_entity_src_gen.gene_src_species
_entity_src_gen.gene_src_strain
_entity_src_gen.gene_src_tissue
_entity_src_gen.gene_src_tissue_fraction
_entity_src_gen.host_org_common_name
_entity_src_gen.host_org_details
_entity_src_gen.host_org_genus
_entity_src_gen.host_org_species
_entity_src_gen.host_org_strain
_entity_src_gen.plasmid_details
_entity_src_gen.plasmid_name

(*f*) ENTITY_SRC_NAT
• **_entity_src_nat.entity_id**
→ **_entity.id**
_entity_src_nat.common_name
_entity_src_nat.details
_entity_src_nat.genus
_entity_src_nat.species
_entity_src_nat.strain
_entity_src_nat.tissue
_entity_src_nat.tissue_fraction

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.

An entity in mmCIF is a chemically distinct molecular component of the structural complex described in the mmCIF. The three possible types of molecular entities are polymer, non-polymer and water. Note that the 'water' entity is water, and only water. Any other well ordered solvent molecules or ions should be treated as non-polymer entities. The relationships between categories used to describe the features of entities are shown in Fig. 3.6.7.5, which also shows how the information describing the entity is linked to the coordinate list in the ATOM_SITE category.

Data items in the ENTITY category are used to label each distinct chemical molecule with a reference code (**_entity.id**), to give the formula weight in daltons (if available) and to define the type of the entity as one of polymer, non-polymer or water. The method by which the entity was produced may be indicated using the item **_entity.src_method**, whose allowed values are `nat` (indicating that the sample was isolated from a natural source), `man` (indicating a genetically manipulated source) or `syn` (indicating a chemical synthesis). A value of `nat` indicates that additional details should be given in the ENTITY_SRC_NAT category and a value of `man` indicates that additional details should be given in the ENTITY_SRC_GEN category. As these flags are only relevant to the macromolecular entities of a structural complex, a value of '.', indicating 'inapplicable', should be given to **_entity.src_method** for solvent or water molecules. The **_entity.details** field can be used for a free-text description of any special features of the entity.

Keywords characterizing the individual molecular species may be given using data items in the ENTITY_KEYWORD category. These keywords should only be used to record information that does not depend on knowledge of the molecular structure. Thus a polypeptide could be described as a polypeptide, or an enzyme, or