International Tables for Crystallography (2006). Vol. G, Section 2.4.5, pp. 47–48.

```
data_cyclohexane

_molecule_name_common           cyclohexane

    loop_
        _atom_id
        _atom_type
        _atom_attach_h          1  C  2   2  C  2   3
                    C  2   4  C  2   5  C  2   6  C  2
loop_
    _bond_id_1
    _bond_id_2
    _bond_type_mif
          1 2 S  2 3 S  3 4 S  4 5 S  5 6 S  6 1 S

loop_
    _reference_conformation
                        $chair   $boat   $twisted_boat

save_chair
    loop_
        _atom_id
        _atom_coord_x
        _atom_coord_y
        _atom_coord_z    1   1.579  0.159  0.263
                         2   0.756  0.507 -0.986
                         3   0.825  0.493  1.541
                         4  -0.549 -0.131  1.590
                         5  -1.377  0.222  0.347
                         6  -0.626 -0.158 -0.937
save_

save_boat
    loop_
        _atom_id
        _atom_coord_x
        _atom_coord_y
        _atom_coord_z    1   1.657 -0.426  0.356
                         2   1.031  0.133 -0.927
                         3   0.960  0.133  1.602
                         4  -0.568 -0.040  1.558
                         5  -1.051 -0.738  0.279
                         6  -0.499 -0.028 -0.964
save_

save_twisted_boat
    loop_
        _atom_id
        _atom_coord_x
        _atom_coord_y
        _atom_coord_z    1   0.933  0.922  0.971
                         2   1.186  0.220 -0.368
                         3  -0.119  0.161  1.796
                         4  -1.135 -0.581  0.911
                         5  -1.371  0.181 -0.397
                         6  -0.083  0.236 -1.238
save_
```

Fig. 2.4.4.3. Atom and bond properties for cyclohexane, together with 3D coordinate representations of three alternative conformations: chair, boat and twisted boat.

2.4.4.5. Global blocks

A global block is similar to a data block except that it is opened with a **global_** statement and contains data that are common or 'default' to all subsequent data blocks in a file. Global data items remain active until re-specified in a subsequent data block or global block.

In some applications it may be efficient to place data that are common to all data blocks within a global block. In particular, save frames may be defined within global blocks and then referenced in subsequent data blocks [this statement corrects an error in Hall & Spadaccini (1994)]. Examples of global data are shown in Figs. 2.4.7.1 and 2.4.7.2, in which a variety of frequently referenced structural units are encapsulated within save frames specified in global blocks.

2.4.5. Atoms, bonds and molecular representations

The MIF dictionary (see Chapter 4.8) contains definitions of the principal data items needed to specify molecular connectivity and spatial representations. These definitions are grouped according to purpose or, as referred to in the DDL dictionary language (Hall & Cook, 1995), by category. Categories are formally specified in the MIF dictionary using the data attribute **_category** but they may also be identified from the data-name construction '***_<category>_<subcategory>_<descriptor>***'. Note that data items appearing in the same looped list must belong to the same category.

The values of some data items are restricted, by definition in the MIF dictionary, to standard codes or states. For example, the item **_bond_type_mif** can only have values S, D, T or O as in its dictionary definitions:

S: single (two-electron) bond;
D: double (four-electron) bond;
T: triple (six-electron) bond;
O: other (*e.g.* coordination) bond.

The MIF dictionary plays the important additional role of validating and standardizing data values. This is illustrated with the data item **_display_colour**, which identifies the colours of 'atom' and 'bond' graphical objects. The colour codes or states for this item are specified in its dictionary definitions as a set of permitted red/green/blue (RGB) ratios, and no other colours may be used in a MIF. This has the technical advantage of making colour states searchable for chemical applications.

Fig. 2.4.4.2 shows MIF data for the molecule (+)-3-bromocamphor. The 'atom' list contains the items **_atom_id**, **_atom_type** and **_atom_attach_h**, which identify the chemical properties of the atoms, plus the items **_atom_coord_x**, ***_y** and ***_z**, which specify the 3D molecular structure in Cartesian coordinates [these are taken from diffraction results (Allen & Rogers, 1970)]. The item **_atom_label** is also used with any graphical depiction of the 3D model. The 'bond' loop in this example uses the simple **_bond_type_mif** conventions described above. The data names needed to depict stereochemistry are discussed with examples (Figs. 2.4.8.1, 2.4.8.2 and 2.4.8.3) in Section 2.4.8.

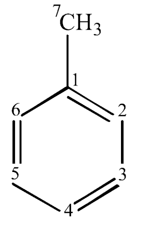
The MIF approach to representing 2D chemical structure separates the specification of chemical atom and bond properties. This provides additional flexibility in the description of the graphical objects, such as atomic nodes and bonded connections. The MIF data required to generate a 2D chemical diagram are shown in Fig. 2.4.4.2. The diagram generated from this data will be in a display area of 500 × 500 coordinate units at a scale of 50 units per cm (the 2D chemical diagram shown in Fig. 2.4.4.2 is not to this scale). The default origin (the bottom left corner of the display area) can be specified with the item **_display_define_origin**. The data used to depict a 2D structure form a two-level loop with the 'atomic' graphical objects at level 1 and the 'bond' graphical objects at level 2. The item **_display_object** has the values '.' (null or no object), 'text' (an element or number string) or 'icon'. The size and colour of the atom site are specified with **_display_size** and **_display_colour**. The bonds connected to each atom site are specified as a sequence of **_display_conn_id** numbers (in loop level 2). These numbers must match one of the `_display_id` numbers at level 1. The connection object is specified with a `_display_conn_symbol` code, which must be a standard value in the dictionary definition, as is the colour of the icon if specified by `_display_conn_colour`.

2.4.6. Bonding conventions

Chemical information systems use a variety of conventions to specify attributes such as aromaticity, bond-order alternation, tautomerism *etc.* These system-dependent conventions decide the values that are permitted for quantities such as bond order, electronic charge and hydrogen-atom count. Most systems also provide for redundancy between chemical attributes. For example, the valency, the number of connected non-hydrogen atoms, the number of terminal hydrogen atoms and the bond types associated with a given atom are clearly related. Operational systems make use of these relationships to perform internal checks and to provide flexibility in substructure search processes.

The MIF data definitions provide for three bonding conventions. These are the data items `_bond_type_mif`, `_bond_type_casreg3` and `_bond_type_ccdc`. The 'mif' convention defines only single, double, triple and other bonds, while the 'casreg3' convention (Mockus & Stobaugh, 1980) extends these to include aromaticity in terms of 'ring alternating normalized bonds' and tautomerism *via* a 'tautomer normalized bond'. The 'ccdc' convention is that employed in the Cambridge Structural Database System (Allen *et al.*, 1991; Allen, 2002) to categorize bond types encountered in both organic and metal-organic molecules.

An important advantage of the MIF approach is that a molecule can be represented using all three bonding conventions within the same data block. An example of alternative bonding conventions encoded for toluene is shown in Fig. 2.4.6.1.



```
data_toluene

_molecule_name_common         toluene

loop_
    _atom_id
    _atom_type
    _atom_attach_h
          1  C  0    2  C  1    3  C  1    4  C  1
          5  C  1    6  C  1    7  C  3

loop_
    _bond_id_1
    _bond_id_2
    _bond_type_mif
    _bond_type_casreg3
    _bond_type_ccdc
          1  2  D  A  A    2  3  S  A  A
          3  4  D  A  A    4  5  S  A  A
          5  6  D  A  A    1  6  S  A  A
          1  7  S  S  S
```

Fig. 2.4.6.1. Three alternative bonding conventions for toluene stored in the same MIF data block.

2.4.7. Structural templates

In many chemical information systems, it is standard practice to build complete 2D molecular representations through the use of a library of commonly referenced structural templates, *e.g.* ligands, functional groups, amino-acid units *etc.*

In a MIF, molecular templates can be encapsulated as save frames, either within a data block for a specific molecule, or within a global block that is accessible to many data blocks. A simple application of a MIF template is shown in Fig. 2.4.7.1, where a 4-methylcyclohexyl ligand is used to encode the molecule tris(methylcyclohexyl)phosphine. In this example a molecular fragment is constructed in the save frame mechex, where the 'atom' sites and 'bond' connections appear in `_atom_*` and `_bond_*` loops. The molecule (2-methylcyclohexyl)(3-methylcyclohexyl)(4-methylcyclohexyl)phosphine is encoded by referencing the template fragment as the save frame $mechex. In the 'atom' loop, the item `_atom_environment` identifies the components of the target molecule as an 'atom' or 'frag' (fragment). If the component is a fragment, the items `_atom_frag_key` and `_atom_frag_id` are used to specify the frame code and the ID of the attached atom in the fragment, respectively. In the 'bond' loop, the connections from the atom P(1) to the template are encoded simply in terms of the `_atom_id` values. The necessary redefinition of the hydrogen and non-hydrogen counts of the template atoms is accomplished using the `_atom_attach_h` and `_atom_attach_nh` items, respectively. The external values override any values that are contained in, or derived from, the data in the template.

The same approach is used to construct the dipeptide alanylserine in Fig. 2.4.7.2. This employs the template peptide units described by the atoms and bonds in the save frames $alanyl and $seryl. The complete dipeptide is specified in its 'atom' list as the template peptides (identified by their save-frame names) and an additional carboxylate O atom. Note that only the atom sites affected by molecule formation are identified explicitly in this list, which gives the values of `_atom_attach_nh`, `_atom_attach_h` and `_atom_charge` for the modified sites in the zwitterionic form of alanylserine.

2.4.8. Stereochemistry and geometry at stereogenic centres

The Cahn–Ingold–Prelog (CIP) notation (Cahn *et al.*, 1966; Prelog & Helmchen, 1982) is available in the MIF definitions to specify the stereochemistry of a molecule. The CIP notation is restricted to tetrahedral atomic centres and to olefinic type stereogenic bonds, and the CIF approach is unsuitable for describing molecules with partially known stereochemistry, molecules containing more complex geometries or substructural queries. The MIF data items representing stereochemical quantities are as follows:

```
_define_stereo_relationship
_atom_cip
_bond_cip
_stereo_atom_id
_stereo_bond_id_1
_stereo_bond_id_2
_stereo_geometry
_stereo_vertex_id
```

The CIP stereochemical designators (R, S, E, Z, r, s, e, z *etc.*) are specified with the MIF data items `_atom_cip` and `_bond_cip`. The MIF atom-property data for the molecule (+)-3-bromocamphor are shown in Fig. 2.4.8.1. In this the absolute configuration is expressed as the atom CIP values R, R and S for nodes 1, 3 and 4. The period in this example is used to indicate a null field.

references