Example 3.2.4.6. *Different representations of a chemical formula.*

```
_chemical_formula_iupac    '[Mo (C O)4 (C18 H33 P)2]'
_chemical_formula_moiety    'Mo,4(C O),2(C18 H33 P)'
_chemical_formula_structural
                          '((C O)4 (P (C6 H11)3)2)Mo'
_chemical_formula_sum              'C40 H66 Mo O4 P2'
```

Example 3.2.4.7. *Representation of a two-dimensional chemical diagram.*

```
loop_
     _chemical_conn_atom_number
     _chemical_conn_atom_type_symbol
     _chemical_conn_atom_display_x
     _chemical_conn_atom_display_y
     _chemical_conn_atom_NCA
     _chemical_conn_atom_NH
        1   S    .39  .81   1   0
        2   S    .39  .96   2   0
        3   N    .14  .88   3   0
        4   C    .33  .88   3   0
        5   C    .11  .96   2   2
        6   C    .03  .96   2   2
        7   C    .03  .80   2   2
        8   C    .11  .80   2   2
        9   S    .54  .81   1   0
        10  S    .54  .96   2   0
        11  N    .80  .88   3   0
        12  C    .60  .88   3   0
        13  C    .84  .96   2   2
        14  C    .91  .96   2   2
        15  C    .91  .80   2   2
        16  C    .84  .80   2   2
```

is connected. Together with the CHEMICAL_CONN_BOND category, the data items in the CHEMICAL_CONN_ATOM category provide a basic description of the chemical structure. Although the description of the chemical structure provided in these two categories is not as extensive as the information that may be conveyed in a molecular information file (Chapter 2.4), it should allow a substructure to be searched for in a suitable database.

The CHEMICAL_CONN_BOND category lists pairs of atoms that contribute to chemical bonds and describes the nature of the bond between them (Example 3.2.4.8). Taken with data items in the CHEMICAL_CONN_ATOM category, data items in this category complete the basic description of a molecular entity.

Bond types are assigned from a list that specifies single, double, triple, quadruple, aromatic, polymeric, delocalized double and π bonds. These are not intended to cover all possible cases, but to characterize a molecular model suitable for database substructure searching.

3.2.4.3. Molecular or packing geometry

The categories describing geometry are as follows:

GEOM group

- GEOM
- GEOM_ANGLE
- GEOM_BOND
- GEOM_CONTACT
- GEOM_HBOND
- GEOM_TORSION

The molecular and packing geometry can be calculated fully given the unit-cell parameters, the space group and a list of atom sites. Therefore, all the information about geometry in the GEOM category group is derivative. However, it is useful to record it within the file both as a check on the primary information stored in other categories and as a method for flagging values to be published.

Example 3.2.4.8. *Bond types in a chemical connectivity table.*

```
loop_
     _chemical_conn_bond_atom_1
     _chemical_conn_bond_atom_2
     _chemical_conn_bond_type
       4     1     doub     4     3     sing
       4     2     sing     5     3     sing
       6     5     sing     7     6     sing
       8     7     sing     8     3     sing
       10    2     sing     12    9     doub
       12    11    sing     12    10    sing
       13    11    sing     14    13    sing
       15    14    sing     16    15    sing
       16    11    sing     17    5     sing
       18    5     sing     19    6     sing
       20    6     sing     21    7     sing
       22    7     sing     23    8     sing
       24    8     sing     25    13    sing
       26    13    sing     27    14    sing
       28    14    sing     29    15    sing
       30    15    sing     31    16    sing
       32    16    sing
```

3.2.4.3.1. *Contents of the geometry-related categories*

The data items in these categories are as follows:

(*a*) GEOM

```
_geom_special_details
```

(*b*) GEOM_ANGLE

- _geom_angle_atom_site_label_1 → _atom_site_label
- _geom_angle_atom_site_label_2 → _atom_site_label
- _geom_angle_atom_site_label_3 → _atom_site_label
- _geom_angle_site_symmetry_1
- _geom_angle_site_symmetry_2
- _geom_angle_site_symmetry_3

```
_geom_angle
_geom_angle_publ_flag
```

(*c*) GEOM_BOND

- _geom_bond_atom_site_label_1 → _atom_site_label
- _geom_bond_atom_site_label_2 → _atom_site_label
- _geom_bond_site_symmetry_1
- _geom_bond_site_symmetry_2

```
_geom_bond_distance
_geom_bond_publ_flag
_geom_bond_valence
```

(*d*) GEOM_CONTACT

- _geom_contact_atom_site_label_1 → _atom_site_label
- _geom_contact_atom_site_label_2 → _atom_site_label
- _geom_contact_site_symmetry_1
- _geom_contact_site_symmetry_2

```
_geom_contact_distance
_geom_contact_publ_flag
```

(*e*) GEOM_HBOND

- _geom_hbond_atom_site_label_A → _atom_site_label
- _geom_hbond_atom_site_label_D → _atom_site_label
- _geom_hbond_atom_site_label_H → _atom_site_label
- _geom_hbond_site_symmetry_A
- _geom_hbond_site_symmetry_D
- _geom_hbond_site_symmetry_H

```
_geom_hbond_angle_DHA
_geom_hbond_distance_DA
_geom_hbond_distance_DH
_geom_hbond_distance_HA
_geom_hbond_publ_flag
```

(*f*) GEOM_TORSION

- `_geom_torsion_atom_site_label_1`
  → `_atom_site_label`
- `_geom_torsion_atom_site_label_2`
  → `_atom_site_label`
- `_geom_torsion_atom_site_label_3`
  → `_atom_site_label`
- `_geom_torsion_atom_site_label_4`
  → `_atom_site_label`
- `_geom_torsion_site_symmetry_1`
- `_geom_torsion_site_symmetry_2`
- `_geom_torsion_site_symmetry_3`
- `_geom_torsion_site_symmetry_4`

  `_geom_torsion`
  `_geom_torsion_publ_flag`

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. ***_symmetry_**** *items have a default value and may be omitted from the list. The arrow (→) is a reference to a parent data item.*

Most categories within this group record distances or angles specified by atom-site labels and are well characterized. The GEOM category currently provides the single data name **_geom_special_details** in which any other details of the geometry that an author considers noteworthy may be stored. Examples of information that might be stored in this data item are least-squares equations of planes, out-of-plane distances, dihedral angles between planes and general comments about the calculation of standard uncertainties.

A subtlety in the geometry-related categories arises from the need to record geometric relationships that involve atoms that are not listed in the ATOM_SITE coordinate list, but that can be derived from the coordinates in this list by the application of a crystallographic symmetry transformation. Thus atom sites in the geometry lists are identified both by their atom-site labels (which must identically match one of the entries in the ATOM_SITE list) *and* by the code for the symmetry transformation that has been applied to the initial location. Since the atom-site labels may refer to atoms in their original location as well as to atoms in symmetry-related locations, the formal key for these categories involves the site labels as well as the symmetry codes. However, in many cases (as discussed further below) the symmetry codes may be absent from a list, and a parser must supply suitable default or null values for the missing components when constructing or checking a complete key.

In many cases, interest is focused on intramolecular distances and angles, and on intramolecular contacts within a single asymmetric unit. In such cases, the geometry lists would contain only atoms listed explicitly in the ATOM_SITE list and the symmetry codes all refer trivially to the identity transformation.

The examples in this section demonstrate various ways of handling geometry lists with trivial or non-trivial symmetry transformations. In Example 3.2.4.9, showing treatment of bond angles, the relevant data items (**_geom_angle_site_symmetry_***) are absent, which is one method for indicating the identity transformation. Dictionary validation software must therefore be able to handle both the presence and absence of these components of the formal category key.

The symmetry transformations in this and related categories take the form of codes *'n klm'* or *n_klm*, where *n* refers to the symmetry operation that is applied to the coordinates stored in **_atom_site_fract_x**, **_atom_site_fract_y** and **_atom_site_fract_z**. The value of *n* must match a number given in **_symmetry_equiv_pos_site_id**. *k*, *l* and *m* refer to the translations that are subsequently applied to the symmetry-transformed coordinates to generate the atom used in calculating the contact. These translations (x, y, z) are related to (k, l, m) by

Example 3.2.4.9. *List of bond angles.*

```
loop_
     _geom_angle_atom_site_label_1
     _geom_angle_atom_site_label_2
     _geom_angle_atom_site_label_3
     _geom_angle
     _geom_angle_publ_flag
     C2  O1  C5   111.6(2)  yes
     O1  C2  C3   110.9(2)  yes
     O1  C2  O21  122.2(3)  yes
     C3  C2  O21  127.0(3)  yes
     C2  C3  N4   101.3(2)  yes
     C2  C3  C31  111.3(2)  yes
     C2  C3  H3   107(1)    no
     N4  C3  C31  116.7(2)  yes
```

Example 3.2.4.10. *List of bonds.*

```
loop_
     _geom_bond_atom_site_label_1
     _geom_bond_atom_site_label_2
     _geom_bond_distance
     _geom_bond_site_symmetry_1
     _geom_bond_site_symmetry_2
     _geom_bond_publ_flag
      O1  C2   1.342(4)  1_555  1_555  yes
      O1  C5   1.439(3)  1_555  1_555  yes
      C2  C3   1.512(4)  1_555  1_555  yes
      C2  O21  1.199(4)  1_555  1_555  yes
      C3  N4   1.465(3)  1_555  1_555  yes
      C3  C31  1.537(4)  1_555  1_555  yes
      C3  H3   1.00(3)   1_555  1_555  no
      N4  C5   1.472(3)  1_555  1_555  yes
```

$$k = 5 + x, \qquad l = 5 + y, \qquad m = 5 + z.$$

By adding 5 to the translations, the use of negative numbers is avoided. As an example, the symmetry code 7_645 means that the symmetry operation with label '7' in the **_symmetry_equiv_pos_site_id** list is applied and the resulting position is translated $+1.0 \times a$ along the x axis, $-1.0 \times b$ along the y axis and $0.0 \times c$ along the z axis, where a, b and c are the unit-cell edges.

List entries with a **_geom_angle_publ_flag** value of `yes` are those that should be published.

The GEOM_BOND category records intramolecular bond distances. In Example 3.2.4.10, all the atoms are untransformed and are at the positions given in the ATOM_SITE list. The symmetry code is 1_555, where the trivial symmetry operation x, y, z is numbered '1' by **_symmetry_equiv_pos_site_id**.

The GEOM_CONTACT category records nonbonded interatomic contacts. In Example 3.2.4.11, all the atoms are untransformed and are at the positions given in the ATOM_SITE list, and therefore the symmetry codes all have the value '.' (meaning 'inapplicable'). This is another method for indicating the identity transformation.

The GEOM_HBOND category records details about hydrogen bonds. Unlike other categories in the GEOM group, the GEOM_HBOND category records information about both distances and angles, including donor–acceptor, donor–hydrogen and acceptor–hydrogen distances and the included angle at the hydrogen-atom site (see Example 3.2.4.12). The comments above about the interpretation of symmetry codes and their relevance in the formal assignment of the category key also apply to this category.

Note that, strictly speaking, this category should only be populated if coordinates for the hydrogen atom are available (because the mandatory component of the category key **_geom_hbond_**

Example 3.2.4.11. *List of nonbonded interatomic contacts.*

```
loop_
     _geom_contact_atom_site_label_1
     _geom_contact_atom_site_label_2
     _geom_contact_distance
     _geom_contact_site_symmetry_1
     _geom_contact_site_symmetry_2
     _geom_contact_publ_flag
     O(1)  O(2)     2.735(3)  .  .  yes
     H(O1) O(2)     1.82      .  .  no
```

Example 3.2.4.12. *List of hydrogen-bond distances and angles.*

```
loop_
     _geom_hbond_atom_site_label_D
     _geom_hbond_atom_site_label_H
     _geom_hbond_atom_site_label_A
     _geom_hbond_distance_DH
     _geom_hbond_distance_HA
     _geom_hbond_distance_DA
     _geom_hbond_angle_DHA
     _geom_hbond_publ_flag
N6  HN6 OW  0.888(8) 1.921(12) 2.801(8) 169.6(8) yes
OW  HO2 O7  0.917(6) 1.923(12) 2.793(8) 153.5(8) yes
OW  HO1 N10 0.894(8) 1.886(11) 2.842(8) 179.7(9) yes
```

Example 3.2.4.13. *List of torsion angles.*

```
    loop_
     _geom_torsion_atom_site_label_1
     _geom_torsion_atom_site_label_2
     _geom_torsion_atom_site_label_3
     _geom_torsion_atom_site_label_4
     _geom_torsion
     _geom_torsion_site_symmetry_1
     _geom_torsion_site_symmetry_2
     _geom_torsion_site_symmetry_3
     _geom_torsion_site_symmetry_4
     _geom_torsion_publ_flag
C(9)  O(2) C(7)  C(2)    71.8(2)  .  .  .  .      yes
C(7)  O(2) C(9)  C(10) -168.0(3)  .  .  .  2_666  yes
C(10) O(3) C(8)  C(6)  -167.7(3)  .  .  .  .      yes
C(8)  O(3) C(10) C(9)   -69.7(2)  .  .  .  2_666  yes
O(1)  C(1) C(2)  C(3)  -179.5(4)  .  .  .  .      no
O(1)  C(1) C(2)  C(7)    -0.6(1)  .  .  .  .      no
```

atom_site_label_H needs a parent label in the atom-site list). In practice, hydrogen bonds can be assumed between donor atoms and acceptors even if the hydrogen atom is not specifically located.

The items in the GEOM_TORSION category describe the torsion angle in degrees generated for the bonded sequence of four atom sites identified by the **_geom_torsion_atom_site_label_*** codes. As with other geometry-specific site labels, these must match labels specified as **_atom_site_label** in the atom list. The torsion angle definition is that of Klyne & Prelog (1960).

Example 3.2.4.13 includes two sites that have been generated by crystallographic symmetry operations and lattice translations from the parent sites in the atom list.

3.2.4.4. Symmetry and space-group information

The categories describing symmetry are as follows:

SYMMETRY group

Original symmetry categories (§3.2.4.4.1)

SYMMETRY

SYMMETRY_EQUIV

Replacement symmetry categories (§3.2.4.4.2)

SPACE_GROUP

SPACE_GROUP_SYMOP

The SPACE_GROUP and older SYMMETRY categories contain information about the symmetry of the crystal; specifically the space group and the symmetry-equivalent positions for that space group. More information about the symmetry is available in the symCIF dictionary described in Chapter 3.8 and presented in Chapter 4.7. The categories SPACE_GROUP and SPACE_GROUP_SYMOP were imported from symCIF in version 2.3 of the core dictionary, and are intended to replace the SYMMETRY and SYMMETRY_EQUIV categories. In most cases, there are strict equivalences between data items in the two sets. The new categories have been adopted for greater compatibility with future expansions to the symmetry CIF dictionary, and to correct some potentially misleading practices in the original categories. Although all the data items in SYMMETRY and SYMMETRY_EQUIV_POS are now formally marked as deprecated, it is likely that the older data items will remain in circulation for some time.

3.2.4.4.1. *Data items in SYMMETRY and related categories*

The data items in these categories are as follows:

(*a*) SYMMETRY

† **_symmetry_cell_setting**
† **_symmetry_Int_Tables_number**
† **_symmetry_space_group_name_H-M**
† **_symmetry_space_group_name_Hall**

(*b*) SYMMETRY_EQUIV

†• **_symmetry_equiv_pos_site_id**
† **_symmetry_equiv_pos_as_xyz**

The bullet (•) *indicates a category key. In practice* ***_symmetry_equiv_pos_site_id*** *is often absent from older CIFs. The dagger* (†) *indicates a deprecated item, which should not be used in the creation of new CIFs.*

The data items in the SYMMETRY category (now superseded by SPACE_GROUP) were used to record the space group. The Hermann–Mauguin (H-M) symbol was given by **_symmetry_space_group_name_H-M**. The dictionary definition recommended the use of the 'full' H-M symbol as listed in *International Tables for Crystallography* Volume A, but was not explicit about the meaning of 'full'. The dictionary examples showed short-form symbols expanded to a complete representation of individual symmetry elements; thus *Pnnn* would be given as `'P 2/n 2/n 2/n'`, and the monoclinic space group $P2_1/m$ would be given as `'P 1 21/m 1'` for the *b*-axis unique setting or `'P 1 1 21/m'` for the *c*-axis unique setting.

In practice, abbreviated symbols were often used, following conventions established over many years; thus `'P 21/m'` was often given as the Hermann–Mauguin symbol when the 'usual' *b* setting of a monoclinic cell had been chosen. It is recommended that these conventions should continue to be followed when the new data item **_space_group_name_H-M_alt** is used instead.

The dictionary examples also suggested concise ways of indicating the origin choice within the **_symmetry_space_group_name_H-M** field; since there is no formal description of how to do this, different authors used different wording. Hence, **_symmetry_space_group_name_H-M** was always best considered as a container for the representation of the space group that would appear in a published article, and not as a machine-readable source of information about the crystallographic symmetry.

The two mechanisms for conveying the symmetry transformations in a fully machine-readable form were the Hall symbol **_symmetry_space_group_name_Hall** (Hall, 1981; Hall & Grosse-Kunstleve, 2001) and a complete listing of the symmetry operations using data items in the SYMMETRY_EQUIV category.

The data item **_symmetry_cell_setting** indicates the crystal system, not (as suggested by its name) the setting used.