

Topology CIF dictionary

BY VLADISLAV A. BLATOV AND DAVIDE M. PROSERPIO

The topology CIF dictionary provides data names for describing some topological characteristics of different lattices and their relation to crystal structures. As this is a field in which resolution has not been reached for more complex cases (*e.g.* interpenetrating lattices) the dictionary is likely to expand over time.

TOPOLOGY

This category is the parent of all categories in the dictionary.

TOPOL

The TOPOL category covers data on connectivity between atoms and structural groups and the related structural properties as calculated from the ATOM, CELL and SYMMETRY data.

Example 1 – Connectivity of the diamond crystal structure. All atoms coincide with the nodes and all bonds coincide with the edges, so the atomic network coincides with the underlying net.

```
loop_
_space_group_symop.id
_space_group_symop.operation_xyz
1 x,y,z
2 1/4-x,1/4-y,z
# Other symmetry elements skipped
13 -y,-x,-z
# Other symmetry elements skipped
192 3/4-z,1/2+y,1/4-x

loop_
_atom_site.label
_atom_site.type_symbol
_atom_site.site_symmetry_multiplicity
_atom_site.fract_x
_atom_site.fract_y
_atom_site.fract_z
_atom_site.occupancy
C1 C 6 0.00000 0.00000 0.25000 1.0000

loop_
_topol_repres_node.label
_topol_repres_node.atom_label
C1 C1

loop_
_topol_link.node_label_1
_topol_link.node_label_2
_topol_link.site_symmetry_symop_1
_topol_link.site_symmetry_translation_1
_topol_link.site_symmetry_symop_2
_topol_link.site_symmetry_translation_2
_topol_link.distance
_topol_link.type
_topol_link.multiplicity
Z1 Z1 1 [0 0 0] 20 [-1 -1 0] 3.2122 v 36

_topol_repres.overall_topology_RCSR pcu-b
```

Example 2 – Connectivity of an underlying net of the calcite (CaCO_3) crystal structure. The nodes of the underlying net correspond to Ca atoms and carbonate (CO_3) groups. The underlying net has the NaCl (pcu-b in the RCSR nomenclature) topology.

```
loop_
_space_group_symop.id
_space_group_symop.operation_xyz
1 x,y,z
2 -y,x-y,z
```

```
# Other symmetry elements elided
20 1/3+x-y,2/3-y,1/6-z
# Other symmetry elements elided
36 1/3-y,2/3-x,1/6+z

loop_
_atom_site.label
_atom_site.type_symbol
_atom_site.site_symmetry_multiplicity
_atom_site.fract_x
_atom_site.fract_y
_atom_site.fract_z
_atom_site.occupancy
C1 C 6 0.00000 0.00000 0.25000 1.0000
O1 O 18 0.25930 0.00000 0.25000 1.0000
Ca1 Ca 6 0 0.00000 0.00000 0.00000 1.0000

loop_
_topol_repres_node.label
_topol_repres_node.chemical_formula_sum
_topol_repres_node.fract_x
_topol_repres_node.fract_y
_topol_repres_node.fract_z
ZA1 Ca 0.00000 0.00000 0.25000
ZB1 CO3 0.00000 0.00000 0.00000

loop_
_topol_link.node_label_1
_topol_link.node_label_2
_topol_link.site_symmetry_symop_1
_topol_link.site_symmetry_translation_1
_topol_link.site_symmetry_symop_2
_topol_link.site_symmetry_translation_2
_topol_link.distance
_topol_link.type
_topol_link.multiplicity
ZA1 ZB1 1 [0 0 0] 20 [-1 -1 0] 3.2122 v 36

_topol_repres.overall_topology_RCSR pcu-b

Example 3 – Connectivity of an underlying net of the cuprite ( $\text{Cu}_2\text{O}$ ) crystal structure. Oxygen atoms coincide with the nodes, while copper atoms represent the edges. There are two interpenetrating networks of the diamond topology.

loop_
_space_group_symop.id
_space_group_symop.operation_xyz
1 x,y,z
2 1/2-x,1/2-y,z
# Symmetry elements elided
13 -y,-x,-z
# Symmetry elements elided
48 1/2-z,y,1/2-x

loop_
_atom_site.label
_atom_site.type_symbol
_atom_site.site_symmetry_multiplicity
_atom_site.fract_x
_atom_site.fract_y
_atom_site.fract_z
_atom_site.occupancy
O1 O 2 0.25000 0.25000 0.25000 1.0000
Cu1 Cu 4 0.00000 0.00000 0.00000 1.0000

loop_
_topol_repres_node.label
_topol_repres_node.atom_label
Node1 O1
```

```

loop_
_topol_link.id
_topol_link.node_label_1
_topol_link.node_label_2
_topol_link.site_symmetry_symop_1
_topol_link.site_symmetry_translation_1
_topol_link.site_symmetry_symop_2
_topol_link.site_symmetry_translation_2
_topol_link.type
_topol_link.multiplicity
1 Node1 Node1 1 [0 0 0] 13 [0 0 0] v 4
loop_
_topol_repres_edge.id
_topol_repres_edge.chemical_formula_sum
1 Cu

_topol_repres.overall_topology_RCSR dia

```

_topol.repres_occurrence_total (Count)

The total number of occurrences in literature and databases of the underlying net topology at the time the data file was prepared.

_topol.special_details (Text)

A description of topological information not covered by the existing data names in the topology categories.

TOPOL_LINK

The TOPOL_LINK category describes the crystal structure connectivity and encodes the weighted colored symmetry-labeled quotient graph, from which the whole periodic net describing the overall topology of the crystal structure can be restored. The definition of symmetry-labeled quotient graph was given by Klein (1996) and examples of weights and colors for the graph edges and vertices are provided by Blatov (2006). The connections described in TOPOL_LINK may correspond to any vectors in the structure, not just bonds or contacts. The nodes that are linked are listed in TOPOL_REPRES_NODE. In order to properly describe the connectivity,

`_topol_link.node_label_1, _topol_link.node_label_2,`
`_topol_link.site_symmetry_symop_1,`
`_topol_link.site_symmetry_translation_1,`
`_topol_link.site_symmetry_symop_2, and`
`_topol_link.site_symmetry_translation_2` must be given for each link, which is identified by `_topol_link.id`. Other items in this category are optional.

References: Klein, H.-J. (1996). Systematic generation of models for crystal structures. *Math. Model. Sci. Comput.* **6**, 325–330; Blatov, V. A. (2006). A method for hierarchical comparative analysis of crystal structures. *Acta Cryst. A* **62**, 356–364.

Category key(s): `_topol_link.id`

_topol_link.node_label_1 (Code)

The labels of two nodes that form a link. These must match nodes specified in TOPOL_REPRES_NODE.

Values must match those for the following item(s):

`_topol_repres_node.label`

_topol_link.node_label_2 (Code)

The labels of two nodes that form a link. These must match nodes specified in TOPOL_REPRES_NODE.

Values must match those for the following item(s):

`_topol_repres_node.label`

_topol_link.site_symmetry_symop_1 (Code)

The symmetry operation that is applied to the coordinates of the node given by `_topol_link.node_label_1` before addition of the translations in `_topol_link.site_symmetry_translation_1`. The value must match a value of `_space_group_symop.id`. No normalization of the resulting coordinates into the interval [0, 1] is carried out. For example, $(x + \frac{1}{2}, y + \frac{1}{2}, z)$ is not the same as $(x - \frac{1}{2}, y + \frac{1}{2}, z)$ for these purposes.

Values must match those for the following item(s):

`_space_group_symop.id`

_topol_link.site_symmetry_symop_2 (Code)

The symmetry operation that is applied to the coordinates of the node given by `_topol_link.node_label_2` before addition of the translations in `_topol_link.site_symmetry_translation_2`. The value must match a value of `_space_group_symop.id`. No normalization of the resulting coordinates into the interval [0, 1] is carried out. For example, $(x + \frac{1}{2}, y + \frac{1}{2}, z)$ is not the same as $(x - \frac{1}{2}, y + \frac{1}{2}, z)$ for these purposes.

Values must match those for the following item(s):

`_space_group_symop.id`

_topol_link.site_symmetry_translation_1 (Integer[3])

A vector of lattice translations that are added to the coordinates after application of the symmetry operation given by `_topol_link.site_symmetry_symop_1` to generate the node used in calculating the link. For example, if the symmetry operation referred to by `_topol_link.site_symmetry_symop_1` is $(x - \frac{1}{2}, y + \frac{1}{2}, z)$, the translation vector is $[0, -1, 0]$ and the original position is $(0.2, 0.7, 1.0)$ in fractional coordinates, then the resultant position is $(-0.3, 0.2, 1.0)$.

_topol_link.site_symmetry_translation_2 (Integer[3])

A vector of lattice translations that are added to the coordinates after application of the symmetry operation given by `_topol_link.site_symmetry_symop_2` to generate the node used in calculating the link. For example, if the symmetry operation referred to by `_topol_link.site_symmetry_symop_2` is $(x - \frac{1}{2}, y + \frac{1}{2}, z)$, the translation vector is $[0, -1, 0]$ and the original position is $(0.2, 0.7, 1.0)$ in fractional coordinates, then the resultant position is $(-0.3, 0.2, 1.0)$.

_topol_link.distance (Real; ångströms)

The link length in ångströms.

_topol_link.id (Code)

The identifier of the link.

_topol_link.multiplicity (Integer)

The number of these links in the unit cell.

_topol_link.order (Real)

The number of electron pairs participating in the bond described by `_topol_link.type`.

_topol_link.special_details (Text)

Information about the link that is not expressed using other data names, for example, bond subtypes and explanations of links with `_topol_link.type` of 'sb'.

TOPOLOGY_CIF

_topol_link.type

(*Code*)
The chemical bond type associated with the connection between the two sites. If no bond exists, use an undelimited period character. If the bond type is unknown, use ? or leave out the data name. The number of electron pairs participating in the bond can be indicated using _topol_link.order.

The data value must be one of the following:

ar	aromatic bond
v	valence bond
pi	π bond
hb	hydrogen bond
vw	van der Waals contact
sb	special bond type described in _topol_link.special_details
.	no bond

_topol_link.Voronoi_solidangle

(*Real; none*)

The solid angle fraction of the interatomic contact A—X, which is the percentage of the sphere of unit radius cut by the pyramid with the basal face of the Voronoi polyhedron of A or X, the two atoms defining the contact. The total solid angle (the whole sphere) is equal to 100. The face used is that corresponding to the A—X interatomic contact.

TOPOL_REPRES

The TOPOL_REPRES category describes a particular crystal structure representation, which corresponds to the periodic (underlying) net topology specified in the TOPOL_BOND category. The underlying net is the net of centroids of structural units. The edges of this net represent the links between the units.

_topol_repres.genus

(*Index*)

The genus of the underlying net, defined as the cyclomatic number of its own quotient graph: $g = 1 + e - v$, where e and v are the number of edges and vertices in the quotient graph. The quotient graph is a finite graph that contains all of the information of the periodic net: the vertices of the graph are the vertices of a translational repeat unit and the edges are all the edges of the repeat unit.

Reference: Delgado-Friedrichs, O. & O'Keeffe, M. (2005). *J. Solid State Chem.* **178**, 2480–2485.

_topol_repres.overall_topology

(*Text*)

The overall topology symbol in an arbitrary form.

Example: 'face-centered cubic topology'

_topol_repres.overall_topology_EPINET

(*Text*)

The identifier for the overall topology as listed in the EPINET database at <http://epinet.anu.edu.au>.

Example: 'sqcc6'

_topol_repres.overall_topology_RCSR

(*Text*)

The overall topology symbol according to the RCSR nomenclature described by O'Keeffe *et al.* (2008).

Reference: O'Keeffe, M., Peskov, M. A., Ramsden, S. J. & Yaghi, O. M. (2008). *Acc. Chem. Res.* **41**, 1782–1789.

Example: 'dia'

_topol_repres.overall_topology_SP

(*Text*)

The overall topology symbol according to the nomenclature of Fischer for sphere packings described by Koch *et al.* (2006).

Reference: Koch, E., Fischer, W. & Sowa, H. (2006). *Acta Cryst. A* **62**, 152–167.

Example: '4/6/c1'

TOPOLOGY CIF DICTIONARY

TOPOL_REPRES_EDGE

_topol_repres.overall_topology_TOPOS

(*Text*)

The overall topology symbol according to the TOPOS nomenclature. TOPOS symbols NDn are interpreted as follows: N is a sequence of degrees (coordination numbers) of all independent nodes; D is one of the letters C (chain), L (layer) or T (three-periodic) designating the dimensionality of the net; and n enumerates non-isomorphic nets with a given ND sequence. For finite (molecular) graphs the symbols $NMK-n$ are used, where k is the number of vertices (atoms) in the graph.

Reference: Aman, F., Asiri, A. M., Siddiqui, W. A., Arshad, M. N., Ashraf, A., Zakharov, N. S. & Blatov, V. A. (2014). *Cryst. Eng. Comm.* **16**, 1963–1970.

Example: '3, 3, 4T3' (The third three-periodic trinodal net with two 3-coordinated and one 4-coordinated independent nodes)

_topol_repres.period

(*Count*)

Periodicity of the underlying net. The allowed data values have the following meaning:

- 0 0-periodic (finite)
- 1 1-periodic (chain)
- 2 2-periodic (layer)
- 3 3-periodic (framework)

_topol_repres.td10

(*Index*)

The topological density TD10 of the underlying net. This is the cumulative sum of the first ten shells of topological neighbours including the central atom. For structures with more than one kind of vertex in the asymmetric unit the value given is a weighted average over the vertices.

_topol_repres.total_point_symbol

(*Text*)

The total point symbol of the underlying net. This value summarizes all the point symbols for the non-equivalent nodes with their stoichiometric coefficients.

Examples: '6^6' (Point symbol for diamond), '4.6^2.24^2.6^10.8^3' (3,6-coordinated underlying net of TiO₂)

TOPOL_REPRES_EDGE

The TOPOL_REPRES_EDGE category describes the chemical composition of the edges of the underlying net.

Category key(s): _topol_repres_edge.id

_topol_repres_edge.chemical_formula_iupac

(*Text*)

Formula of the residue or ion which corresponds to the edge, expressed in conformance with IUPAC rules.

_topol_repres_edge.chemical_formula_moiety

(*Text*)

Formula of the residue or ion which corresponds to the edge. The formula is written in accordance with the rules of the _chemical_formula.moiety tag.

_topol_repres_edge.chemical_formula_sum

(*Text*)

Formula of the residue or ion which corresponds to the edge. The formula is written in accordance with the rules of the _chemical_formula.sum tag.

_topol_repres_edge.id

(*Code*)

The label of the edge. These must match labels specified as _topol_link.id in the TOPOL_LINK list.

Values must match those for the following item(s):

_topol_link.id

TOPOL_REPRES_EDGE**DATA DICTIONARIES****TOPOLOGY_CIF****TOPOL_REPRES_ENTANGL**

The TOPOL_REPRES_ENTANGL category describes entanglements in the underlying net. This category is a placeholder for future development of descriptions of entanglement.

TOPOL_REPRES_NODE

The TOPOL_REPRES_NODE category describes the chemical composition, structure and topological properties of the nodes of the underlying net. Nodes may be specified by reference to atom sites, or by explicitly giving their coordinates.

Reference: Blatov, V. A., O'Keeffe, M. & Proserpio, D. M. (2010). *Cryst. Eng. Comm.* 12, 44-48.

Category key(s): `_topol_repres_node.label`

_topol_repres_node.atom_label *(Code)*

The atom label corresponding to this node. Not all nodes have to coincide with atom sites.

Values must match those for the following item(s):

`_atom_site.label`

_topol_repres_node.chemical_formula_iupac *(Text)*

Formula of the residue or ion which corresponds to the node, expressed in conformance with IUPAC rules.

_topol_repres_node.chemical_formula_moiety *(Text)*

Formula of the residue or ion which corresponds to the node. The formula is written in accordance with the rules of the `_chemical_formula.moiety` tag.

_topol_repres_node.chemical_formula_sum *(Text)*

Formula of the residue or ion which corresponds to the node. The formula is written in accordance with the rules of the `_chemical_formula.sum` tag.

_topol_repres_node.coordination_sequence *(Integer)*

The coordination sequence is a sequence of numbers counting the atoms in the 1st, 2nd, 3rd etc. coordination shells of any given node in the net. In other words, the k th entry in the list is the number of vertices linked to the node by a path of exactly k steps. It is usually listed up to $k = 10$.

Example: '[4 12 24 42 64 92 124 162 204 252]' (The diamond coordination sequence)

_topol_repres_node.extended_point_symbol *(Text)*

The extended point symbol of the node lists all shortest circuits for each angle for each non-equivalent atom. $A(b).B(c) \dots$ there are b A -rings and c B -rings for all the $N(N - 1)$ circuits per node. It is sorted so shortest rings came first. For 4-coordinated nodes only, the angles are grouped in opposite pairs; ab, cd and ac, bd and ad, bc (written in lexicographic order smallest numbers first).

Examples: '6(2).6(2).6(2).6(2).6(2)' (ES for a vertex in the diamond structure), '4.6(2).4.8(3).6(2).6(2)' (ES for one vertex of feldspar net), '7(2).9(2).7(3).7(3).7(3)' (ES for the vertex of qzd net), '4.4.4.4.6(3).6(3).6(5).6(5).6(5)' (ES for the vertex of 5-c sqp net)

_topol_repres_node.fract_x *(Real; none)*

Atom site coordinates as fractions of the cell length values.

_topol_repres_node.fract_y *(Real; none)*

Atom site coordinates as fractions of the cell length values.

_topol_repres_node.fract_z

(Real; none)

Atom site coordinates as fractions of the cell length values.

_topol_repres_node.label

(Text)

The label of the node, which corresponds to a particular node of the crystal structure representation.

_topol_repres_node.point_symbol

(Text)

The (short) point symbol of the node. This lists the number and size of the shortest closed chains of connected nodes (circuits) starting from any non-equivalent node in the net. For an N -coordinated node there are $N(N - 1)$ circuits

Examples: '6^6' (Point symbol for a diamond vertex), '4^2.6^3.8' (Point symbol for a feldspar 4-coordinated vertex), '7^5.9' (Point symbol for the vertex of 4-c qzd net), '4^4.6^6' (Point symbol for the vertex of 5-c sqp net)

_topol_repres_node.structural_formula_InChI

(Text)

Formula of the residue or ion, which corresponds to the node. The formula is written in accordance with the rules for IUPAC international chemical identifiers (InChI) as described by Heller *et al.* (2015).

Reference: Heller, S. R., McNaught, A., Pletnev, I. & Tchekhovskoi, D. (2015). *J. Cheminform.* 7:23.

_topol_repres_node.structural_formula_SMILES

(Text)

Formula of the residue or ion which corresponds to the node. The formula is written in SMILES notation for describing chemical structure as formalised by the OpenSMILES group (<https://www.opensmiles.org>).

_topol_repres_node.symmetry_multiplicity *(Index)*

The number of different sites that are generated by the application of the space-group symmetry to the coordinates given for this site. It is equal to the multiplicity given for this Wyckoff site in *International Tables for Crystallography* Vol. A (2002).

_topol_repres_node.vertex_symbol

(Text)

The vertex symbol of a node provides similar information to the extended point symbol, but only for rings, which are circuits that contain no shortcuts, that is, are not the sum of two smaller circuits. There may be circuits that cannot be rings. If there are no rings meeting at a particular angle of the node, the symbol '*' is used instead of the $\mathbf{a}^\wedge \mathbf{a}$ symbol. It is sorted so shortest rings came first. For 4-coordinated nodes only, the angles are grouped in opposite pairs; ab, cd and ac, bd and ad, bc (written in lexicographic order smallest numbers first). In the ordering the symbol '*' is equivalent to zero.

Examples: '6(2).6(2).6(2).6(2).6(2)' (Vertex symbol for diamond), '4.6(2).4.8.6.6(2)' (VS for one vertex of feldspar net), '7(2).*.7(3).7(3).7(3)' (VS for the vertex of qzd net), '4.4.4.4.6.6(5).6(5).6(5)' (VS for the vertex of 5-c sqp net)

_topol_repres_node.Wyckoff_symbol *(Code)*

The Wyckoff symbol (letter) as listed in the space-group section of *International Tables for Crystallography*, Vol. A (1987).

TOPOL_REPRES_OCCURRENCE

The TOPOL_REPRES_OCCURRENCE category lists the appearances of the underlying net topology in crystal structures.

Category key(s): `_topol_repres_occurrence.id`

TOPOLOGY_CIF**TOPOLOGY CIF DICTIONARY****TOPOL_REPRES_TILING_TILE****_topol_repres_occurrence.citation_id** *(Code)*

Reference to a publication, where a crystal structure with the underlying net topology was characterized. This item is a pointer to an item described in the core CITATION category. If a publication and database entry are not directly related, they should be listed in separate rows.

Values must match those for the following item(s):

_citation.id**_topol_repres_occurrence.id** *(Text)*

A unique identifier for the literature or database reference.

TOPOL_REPRES_TILING

The TOPOL_REPRES_TILING category describes the natural tiling corresponding to the underlying net. A tiling is a partition of crystal space using generalised polyhedra, and a natural tiling is one for which tiles are the smallest possible that conserve the full symmetry of the net and for which the faces are all locally strong rings. This means that there is no single largest face (face with the largest number of vertices) as such a face will be the sum of the other smaller faces. The tile signature contains the sizes of the tile faces and the number of faces of a given size in the tile.

Reference: Blatov, V. A., Delgado-Friedrichs, O., O'Keeffe M. & Proserpio D. M. (2007). *Acta Cryst. A63*, 418-425.

_topol_repres_tiling.Dsize *(Index)*

The number of distinct (not symmetry-related) chambers in the tiling.

_topol_repres_tiling.dual *(Text)*

The overall topology symbol of the dual net, which corresponds to the net of the dual of the natural tiling.

_topol_repres_tiling.edges *(Index)*

Number of independent tile edges in the natural tiling.

_topol_repres_tiling.faces *(Index)*

Number of independent tile faces in the natural tiling.

_topol_repres_tiling.signature *(Text)*

The tiling signature, written in the form $\alpha[A^a.B^b\dots] + \beta[C^c.D^d\dots] + \dots$, where square brackets envelop tile symbols, α, β, \dots are stoichiometric coefficients, A, B, C, D, \dots are sizes of tile faces, a, b, c, d, \dots are numbers of the faces of a given size in the tile. The signature is written in a lexicographic order, smallest numbers first: $5[6^4] + [6^3] = 56463$ better than $[6^3] + 5[6^4] = 63564$.

Examples: ' $[6^4]$ ' (Natural tiling for diamond), ' $3[4^6]+[4^6.6^8]+[4^{12}.6^8.8^6]$ ' (Natural tiling for zeolite LTA)

_topol_repres_tiling.tiles *(Index)*

Number of independent tiles in the natural tiling.

_topol_repres_tiling.vertices*(Index)*

Number of independent tile vertices in the natural tiling.

TOPOL_REPRES_TILING_FACES

The TOPOL_REPRES_TILING_FACES category tabulates the faces belonging to each tile in the tiling. Together with the TOPOL_REPRES_TILING_TILES category it tabulates the information contained in **_topol_repres_tiling.signature**. See the TOPOL_REPRES_TILING category for further information.

Category key(s): **_topol_repres_tiling_faces.tile_id**
_topol_repres_tiling_faces.size

Expanded description of $3[4^6]+[4^6.6^8]+[4^{12}.6^8.8^6]$ tiling

```
loop_
  _topol_repres_tiling_tile.id
  _topol_repres_tiling_tile.count
  a      3
  b      1
  c      1

loop_
  _topol_repres_tiling_faces.tile_id
  _topol_repres_tiling_faces.size
  _topol_repres_tiling_faces.count
  a      4    6
  b      4    6
  b      6    8
  c      4    12
  c      6    8
  c      8    6
```

_topol_repres_tiling_faces.count*(Index)*

The number of faces of this size in the tile.

_topol_repres_tiling_faces.size*(Count)*

The size of the tile face.

_topol_repres_tiling_faces.tile_id*(Code)*

The tile to which this face belongs. It must be one of the values provided in **_topol_repres_tiling_tile.id**.

Values must match those for the following item(s):

_topol_repres_tiling_tile.id**TOPOL_REPRES_TILING_TILE**

The TOPOL_REPRES_TILING_TILE category provides information on each of the tiles in the tiling. Together with the TOPOL_REPRES_TILING_FACES category it tabulates the information contained in **_topol_repres_tiling.signature**. See the TOPOL_REPRES_TILING category for further information.

Category key(s): **_topol_repres_tiling_tile.id**

_topol_repres_tiling_tile.count*(Index)*

The number of this kind of tile in the tiling.

_topol_repres_tiling_tile.id*(Code)*

An arbitrary, unique identifier for this tile type.