Topological representations of crystal structures: generation, analysis and implementation in the TopCryst system

Alexander P. Shevchenko, Aleksandr A. Shabalin, Igor Yu. Karpukhin and Vladislav A. Blatov

Samara Center for Theoretical Materials Science (SCTMS), Samara State Technical University, Samara, Russian Federation; Laboratory of Coherent Optics (LCO), Samara Branch of P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Samara, Russian Federation; Samara Center for Theoretical Materials Science (SCTMS), Samara University, Samara, Russian Federation

ABSTRACT
Main modern approaches to the topological representation of crystal structures of different chemical classes are overviewed. The problem of automated generation and analysis of such representations is discussed, and a new free web service is presented, which enables the user to describe topological features of crystal structures of any complexity and chemical composition in a fully automated mode. The service requires only an input file with crystallographic information in the standard CIF format and generates all reasonable representations of the structure by selecting structural units with rigorous algorithms. Then the representations are assigned to known topological types and relations to other crystal structures, which have to same architectures, are established. The service is interfaced to a set of topological databases, which in turn have gateways to world-wide crystallographic databases. A number of examples of topological analysis of different classes of chemical structures are presented, and an outlook is given for further development and applications of the service for big data analysis and data mining in crystal chemistry and materials science.

A new free web service TopCryst is presented to describe crystal structures of any complexity and chemical composition in a fully automated mode.

1. Introduction
The atomic structures of crystalline solids were determined by diffraction methods during more than a century. Although the diffraction experiment provides the data on the distribution of electron density, thus keeping the crystal space continuity, this information is usually lost in public access. The retained structural information, which is now collected in a number of electronic world-wide databases, such as Cambridge Structural Database (CSD) [1], Inorganic Crystal Structure Database (ICSD) [2], Crystallography Open Database (COD) [3] and Pearson’s Crystal Data (PCD) [4], describes only positions of maxima of electron density (atoms) and the structure symmetry, thus bearing only geometrical properties of the structure. When a chemist analyzes this information, he/she should again restore the structure connectivity, i.e. the bonds between atoms. Since besides the atom names, only geometrical information is available at this stage, the criteria for determining the bonds can also be only geometrical. One can distinguish two groups of such criteria: (i) distance criteria, which use interatomic distances or other parameters derived from them, such as atomic radii [5] or bond strengths [6], and (ii) polyhedron criteria, which rest upon Voronoi polyhedra [7]; the criteria from these groups can be combined. However, for a long time, there was no universal method for automated determination of atomic coordination numbers, and
2. What is topological representation of crystal structure?

As was mentioned above, the initial crystallographic information is purely geometrical, and it should be supplied with the information on the structure connectivity for further crystallochemical analysis. Thus, one comes to the notion of topological representation of crystal structure.

2.1. General concept

Any model of a crystal structure where connections between atoms and complex structural groups are established can be treated as a topological representation. The set of atoms or structural groups considered as a whole forms a topological space $T$, on which a topology is defined as a family of pairwise sets (links) from $T$ as well as all their unions and intersections. Such topological space together with the defined topology can be visualized as a graph, which possesses translational symmetry to be equal to or higher than the symmetry of $T$, which is described by one of the space groups. This periodic graph is called a crystallographic or non-crystallographic net [14] depending on whether its symmetry group is isomorphic or non-isomorphic to a space group, and it describes a particular topological representation of the crystal structure. The most general representation includes all atoms of the crystal structure and all connections between them; we call it complete representation since all other (partial) representations can be derived from it. Certainly, the rigorous notion of complete representation is abstract; it is hardly possible to determine all interatomic links as in general their number can be large and depends on a particular task. For example, one could be interested not only in direct interatomic interactions, even the weakest, but also in relations between distant atoms, which is important when analyzing atomic sublattices (e.g., cation arrays [15]). There are three basic topological operations for generating partial representations: (i) removing an atom, (ii) removing a link and (iii) contracting an atomic group to its centroid, which is equivalent to separating structural units and representing them by their centroids. All these operations result in a simpler representation, thus they are called simplifications. The net, which is constructed for a particular representation and defines its topology, is called underlying net [16]. The structures that have the underlying nets of the same topology belong to the same isoreticular series.

2.2. How to determine structure connectivity?

To build the complete representation one should determine all links between atoms and classify the links in accordance with chemical reasons. As was mentioned above, interatomic distance is the primary geometrical descriptor for the analysis of chemical interactions in crystals. However, crystal chemists also use his/her intuition and experience to distinguish different kinds of bond, to accept or ignore weak interactions, and to select structural units. It would be extremely useful to develop a computer procedure that mimics this human’s reasoning and provides a structure model to be close to an ordinary crystallochemical representation. Recently [17], we proposed such procedure, which was implemented into ToposPro as the Domains algorithm. This algorithm uses parameters of atomic Voronoi polyhedra in addition to interatomic distances and atomic radii to account for the whole environment of the atom when analyzing a particular interatomic contact. The whole set of the contacts determined by the Voronoi polyhedron is then clusterized to separate bonds of different kinds such as valence, H bond, specific or van der Waals interaction. As a result, atomic connectivity can be analyzed in any kind of crystal structure with the same set of options, which is important when processing big sets of diverse structural data. For example, Voronoi polyhedron of a copper atom in the crystal structure of $[\text{Cu(acac)}_2]$ (acac =...
acetylacetonato) (ACACCU41)\(^1\) [18] is confined by 16 faces of different sizes (Table 1 and Figure 1); as a result, the Domains algorithm distinguishes valence bonds, van der Waals contacts, and very weak ‘indirect’ contacts, which do not correspond to any bonding.

### 2.3. Representations of crystal structures

Any crystallochemical consideration of a crystal structure is a simplification because it is impossible and unreasonable to account for all interatomic interactions and sometimes even for all atoms in the structure. The corresponding topological representation depends on the nature of bonding in the crystal and on the crystallochemical task to be solved within this representation. Below, we consider typical topological representations for different classes of crystal structures.

#### 2.3.1. Covalent crystals

In 3D covalent crystals, the structure framework is formed by strong valence bonds; all other (e.g. van der Waals) interactions are much weaker and usually should be ignored. All atoms of the framework are included into the representation, and the net coincides with the framework of valence bonded atoms. Thus, a variety of natural and hypothetical 3D carbon allotropes can be perfectly discriminated by topology of their atomic nets. These topologies are collected in the database SACADA [19] as well as in the ToposPro TTD Collection. If the crystal consists of low-dimensional (0D, 1D or 2D) structural units, two representations are possible: (i) representation of the low-dimensional structural unit, when only valence bonds are considered, and (ii) representation of the whole structure, when the contacts between the structural units are also taken into account. For example, the structure of the C8 carbon polymorph [20] can be described as the whole net with the pcb topology, but also as a network of 0D cubic units whose centers form a body-centered cubic (bcc) net (Figure 2). Another example is the selenium structure [21], which consists of simple chains 2C1, but if one takes into account both intrachain valence bonds and interchain van der Waals contacts, the resulting topology is primitive cubic (pcc), which is realized in many structures with strong bonds like α-Po or NaCl (Figure 3).

#### 2.3.2. intermetallic compounds

Metals, metal alloys and intermetallic compounds are similar to covalent crystals since metal atoms being formally uncharged are connected by bonds of one kind, metallic bonding. Thus, the main topological representation includes all atoms and all direct contacts between them. However, to establish correlations between intermetallic compounds of different stoichiometric composition and structure, one can select polyatomic structural units using the nanocluster approach [22]. In this approach, intermetallic structure is represented as an assembly of multishell onion-like nanoclusters, whose centers are allocated in the most symmetrical positions of the structure. The nanoclusters have no common internal atoms but can share the atoms of their external shells. The underlying net consists of the nanocluster centers (atoms or centers of voids) and links between them, which correspond to the contacts between the outer-shell or shared atoms of the nanoclusters. This approach enables one to reveal simple topological motifs in

![Figure 1. Voronoi polyhedron of a copper atom in the crystal structure of \(\text{Cu(acac)}_2\) (ACACCU41). The valence bonds are shown by solid lines; other contacts are depicted by dotted lines. The ‘indirect’ contacts correspond to the faces, which are not crossed by the contact lines. Hereafter the TopCryst outputs are given in the ESI for all structures shown in the figures.](image)

**Table 1.** Parameters of the Voronoi polyhedron of a copper atom and descriptors of the contacts of this atom with neighboring atoms in the crystal structure of \(\text{Cu(acac)}_2\): interatomic distance (Dist.), area of a face of the Voronoi polyhedron (SSeg), volume of the pyramid based on the face with the top in the copper atom (VSeg), solid angle of the face (SAng), and the number of the face vertices (NV).

| Number of contact | Atom | Dist, Å | SSeg, % | VSeg, % | SAng, % | NV | Type |
|-------------------|------|---------|---------|---------|---------|-----|------|
| 1, 2              | 2×O1 | 1.912   | 18.29   | 15.71   | 19.66   | 7   | Val  |
| 3, 4              | 2×O2 | 1.916   | 18.16   | 15.64   | 19.55   | 7   | Val  |
| 5, 6              | 2×H1 | 2.996   | 4.99    | 6.72    | 4.18    | 5   | vdW  |
| 7, 8              | 2×C3 | 3.002   | 6.48    | 8.74    | 5.09    | 7   | vdW  |
| 9, 10             | 2×H5 | 3.373   | 0.59    | 0.90    | 0.43    | 3   | No bond |
| 11, 12            | 2×C2 | 3.389   | 0.65    | 0.99    | 0.48    | 3   | No bond |
| 13, 14            | 2×C1 | 3.420   | 0.42    | 0.64    | 0.30    | 3   | No bond |
| 15, 16            | 2×H4 | 3.434   | 0.42    | 0.65    | 0.30    | 3   | No bond |

The SSeg, VSeg and SAng values are given in percentage of the total area, volume of the Voronoi polyhedron, and total solid angle (4π steradian). The valence (Val) bonds have much larger SSeg, VSeg and SAng values and smaller Dist values than the van der Waals (vdW) contacts.
complex crystal structures. For example, the complete representation of Nd$_2$Fe$_{17}$ (230,594) \cite{23} has the complex 1212,12,14,20T1 topology, where the numbers 12, 12, 12, 14 and 20 designate the coordination of five non-equivalent metal atoms. However, the nanocluster representation indicates that the 40-atom two-shell nanoclusters are packed over the nodes of an underlying net with the topologically simple uninodal face-centered cubic (fcu) topology (Figure 4). Many examples of the decomposition of extremely complex intermetallic structures into nanoclusters are given in \cite{24}. The topological types of nanoclusters in all known intermetallic

Figure 2. The C8 structure represented as the entire carbon network with the pcb topology (left) or as a network of the centers of 0D cubic units, which follows the bcu topology (right).

Figure 3. Representations of the structure of elementary selenium: (left) 2C1 chains of valence bonds and (right) pcu framework of valence and van der Waals contacts.

Figure 4. Nanocluster representation of Nd$_2$Fe$_{17}$: the contents of the primitive cell of the whole structure with an unusual 12,12,12,14,29T1 topology (left), a 40-atom two-shell nanocluster unit (middle) and a fragment of the underlying net with the fcu topology, the nodes of which correspond to the centers of the nanoclusters (right).
compounds are gathered in the *ToposPro* Topological Types of Nanoclusters (TTN) collection [25].

### 2.3.3. Ionic inorganic compounds

In ionic crystals, there are two oppositely charged structural components, and this feature extends the set of possible topological representations. The most general representation is similar to the representation used for covalent crystals: all atoms are included into the net, and only strongest (ionic or ion-covalent) bonds are considered as the net edges. If cations or anions are complex, a simplified representation can be built where the complex ions are represented by their centroids [26] (Figure 5).

However, at least two more representations are viable: (i) the anionic packing since many ionic inorganic structures are based on such a packing, and (ii) the cation array, which can support the general structural motif in many cases [27]. In these representations, the atoms of one component (cations or anions) are removed, and the underlying net is determined by establishing direct links between the atoms of the other component using the distance criterion or Voronoi partition. Thus, the Na₃PS₄ crystal structure can be considered as a packing of sodium cations or as a packing of centers of the PS₄³⁻ anions, which coincide with the phosphorous atoms; in both cases, the topologies are well known and regular (Figure 5).

### 2.3.4. Coordination compounds

Structural units of coordination compounds naturally include complexing metal atoms and mono- or polyatomic molecular ligands. Such representation is widespread and applicable to any coordination compound that is why it is called ‘standard’ in *ToposPro*. However, if there are polynuclear complex groups, an alternative ‘cluster’ representation is viable, in which the nodes of the underlying net coincide with the centers of these groups, while edges mimic the links between them. To automatically recognize such groups and construct the ‘cluster’ representation, a rigorous topological criterion was proposed and implemented into *ToposPro* [28]. To apply this criterion, all the shortest atomic cycles are determined, which meet at each non-equivalent atom of the

---

**Figure 5.** Representations of the Na₃PS₄ crystal structure (121,569): (top left) complete network of the 4,6,6,8T21 topology, (top right) the underlying net wzz of sodium atoms and centroids of the PS₄³⁻ anions, (bottom left) packing net sqc489 of Na⁺ cations, and (bottom right) packing net bcu-x of PS₄³⁻ anions.
crystal structure, and the bonds belonging to short or long cycles are classified as intra- or intercluster links, respectively. The cluster representation can be realized in two ways [28,29]: (i) the ‘all nodes’ representation, when complex ligands are represented by several nodes of the underlying net, and (ii) the ‘single node’ representation, in which the ligands are represented by their centroids irrespective of their complexity. For example, the crystal structure of [Cd$_2$(DDPP)(H$_2$O)$_2$] • DMA•H$_2$O (H$_4$DDPP = 3,5-di(2',5'-dicarboxyphenyl)pyridine; DMA = dimethylacetamide) (HOHHIY, 101,303) [30] admits four representations (Figure 6): standard representation with cadmium atoms and DDPP ligands as structural units (top left) and three cluster representations. Two of them are obtained for the intercluster bonds belonging to the cycles of size higher than 6 and consist of tetranuclear clusters [C$_8$H$_8$O$_{20}$Cd$_4$], and the ligand either divided into rings in the ‘all nodes’ representation or considered as a whole in the ‘single node’ representation. An alternative and topologically the simplest cluster representation (top right) is obtained for with intercluster bonds belonging to the cycles of size higher than 12 and formed by one type of node, where the tetranuclear cluster is extended with the dicarboxyldiphenylpyridine fragments of two DDPP ligands.

The local coordination of ligands can be described with the nomenclature proposed in [31]; the coordination modes of the ligands in coordination compounds are gathered in the ToposPro Topological Types of Ligands (TTL) collection [32].

### 2.3.5. Molecular crystals

The most typical (i.e. ‘standard’ in the ToposPro terminology) representation of a molecular crystal consists of molecules connected by intermolecular interactions. Thus, the underlying net is formed by the molecular centroids linked by the edges, each of which mimics the whole variety of the interactions between a particular pair of molecules. However, if the intermolecular interactions can be differentiated, one can consider additional representations. For example, the crystals with hydrogen bonds can be represented as both molecular packings and networks of H-bonded molecular ensembles [33] (Figure 7).

The method of connection of molecules can be formalized in a nomenclature to be similar to that for ligands [34]. With this nomenclature, the connection modes of molecules in molecular crystals were gathered in the ToposPro Topological Types of Molecules (TTM) collection [32].

![Figure 6. Representations of the crystal structure of [Cd$_2$(ddpp)(H$_2$O)$_2$] • DMA•H$_2$O (top center) and the corresponding underlying nets: standard representation (top left), cluster representations: ‘all nodes’ (bottom left, the 3,3,10T49 topology), ‘single node’ (bottom right, the flu topology) and alternative (top right, the pcu topology).](image)


2.3.6. Porous structures

A special kind of topological representation is possible for the structures that contain voids, pores and channels. The free space in such structures can be described by a natural tiling whose elements, natural tiles, are polyhedral units, which fill in the crystal space without gaps or intersections (Figure 8). The procedure for constructing natural tilings for nets of any complexity was formalized in a rigorous algorithm [35] and implemented into ToposPro. In particular, natural tilings were constructed for all the known zeolite frameworks [36], presented at the International Zeolite Association (IZA) website (http://www.iza-structure.org/data bases/) and all natural tiles that occurred in these frameworks were included into the ToposPro Topological Types of Tiles (TTT) collection [36]. The underlying net in the tiling representation describes the system of cages and channels thus elucidating the free space topology and the method of assembling porous frameworks from polyhedral units [37,38].

3. Automation of topological analysis with TopCryst

The concept of topological representation admits a mathematical formalization of the crystal structure description that in turn enables one to automate crystallochemical analysis. ToposPro can essentially help in such analysis, but it assumes that the user has a good background in topological methods since the ToposPro applied programs contain a lot of options that should be correctly specified to obtain the required results. This is the reverse side of the ToposPro flexibility: the software renders many methods for the crystal structure analysis but requires strictness from the user. Moreover, if the analysis consists of several steps, the user has to run the procedure at each step manually. This also requires perfect understanding of all steps from the user and does not enable him/her to make the whole analysis fully automated. The TopCryst service described below was designed to overcome this ToposPro shortcoming, to make the ToposPro tools available for a broad crystallochemical community and to take one more step toward automation of the analysis of crystal structures resting upon the information from a CIF file (Crystallographic Information File) (Figure 9 [39]). At present, the ToposPro TTD, TTO, TTL and TTT collections have interfaces with TopCryst.

3.1. TopCryst algorithms

TopCryst uses algorithms, which we earlier implemented into ToposPro, but now they are united in the same analytical procedure, which requires no initial information from the user besides standard CIF file. The entire procedure of topological analysis includes the following steps (Figure 10):

1. Reading the information from a CIF file provided by the user. The standard CIF parser from the Python Anaconda library is used, which is supplemented with a special procedure of BONDS class that parses the topological data encoded with the new Topology CIF dictionary (https://www.iucr.org/resources/cif/dictionaries/cif_topology). If the CIF file contains information on the structure connectivity encoded as a labeled quotient graph in accordance with the rules of the Topology CIF dictionary, the next step is passed. This option is useful if the structure connectivity should be calculated with another algorithm or the topology of an already simplified underlying net is analyzed.

2. Determination of the structure connectivity with the Domains algorithm. The default options of the ToposPro AutoCN program are
used, which provide a crystallochemically correct assignment of interatomic bonds and interactions of various types.

(3) Assigning the structure to a specific chemical class or classes depending on the chemical composition of the compound and the dimensionality of the structural units. The service recognizes covalent, ionic and coordination compounds, as well as molecular structures where the structural groups are connected by hydrogen, specific, Coulomb or van der Waals interactions.

(4) Simplifying the structure to an underlying net containing the associated structural unit centroids. The service chooses the simplification options depending on the chemical class of the crystal structure.

(5) Calculating topological indices [40] (coordination sequences and point symbols for nodes), which characterize the topological type of the underlying net.

(6) Determining the underlying net topology by matching the calculated topological indices and the topological indices of the topological types from the TTD collection.

(7) Outputting the results to a file (see examples in the ESI).

All steps (1)–(7) are performed in an automated mode and require no user intervention. With the obtained information, the user can then look for other structures that have the same structural units or the underlying net topology in the TTL, TTO or TTT collections. Note that the current TopCryst
version covers not all ToposPro features, but only those that can be generated in a fully automated mode and hence can be used by an amateur in topological analysis. An extended TopCryst version will include more procedures for building representations, in particular, the nanocluster and tiling approaches as well as recognizing interpenetration.

3.2. TopCryst tools and interfaces

The TopCryst user interface (front-end) is programmed in HTML, PHP, CSS and JavaScript in the form of a webpage available for free at https://topcryst.com/. Conventionally, it can be divided into two main parts: a service for determining the topology of a structure from a CIF file, and a search engine for underlying topologies (Search topology) and topological objects (Search topological objects and Search structure) in the TTD, TTL, TTO and TTT collections.

The interface of the service for determining the topology from a CIF file is designed as a window for uploading the file to be sent to the server through an AJAX request for further processing. At the same time, this window displays the current status of the file processing: 'file transfer,' 'position in the queue,' 'analysis' and 'result.' After the analysis is completed, the result is displayed on the web page in the form of a PDF report, a list of the resulting CIF files with the underlying nets to be downloaded by the user, and a list of the underlying net topologies with the links to the webpages, which contain the information on these topologies.

The Search topology interface is represented by a field for the input of the topology name in one of available nomenclatures [32], a tool to download a CIF file with the net of this topology, links to RCSR and EPINET databases if they contain such net, a JSmol [41] net visualizer, and a list of representatives grouped by the representation types with the links to the CSD and ICSD records at the Cambridge Crystallographic Data Centre website (https://www.ccdc.cam.ac.uk/). The user can search by topology name, representative CSD Reference Code or ICSD Collection Code as well as the natural tile name.

The Search topological objects and Search structure interfaces enable the user to search for the webpages of natural tiles and ligands by their names, as well as for the webpages of individual structures by their CSD Reference Code or ICSD Collection Code. On these webpages, a JSmol visualization of natural tiles and ligands is available.

The server part of the service (back-end) is written in PHP and includes a queue system for processing user files (CIFs) and a topology search module. The queue system arranges the CIF files uploaded by the user depending on the availability of free processors, sends the files for calculation with the subsequent...
Figure 10. *TopCryst* procedure of the topological analysis of a crystal structure.
transfer to the module, which provides determination of the underlying net topology. The results of the analysis are then displayed on the webpage as a list of underlying topologies for all possible representations of the crystal structure (Figure 9). The output includes the name of the topology if it was found in the TTD collection, otherwise ‘unknown topology’ is indicated with the description of coordination numbers of all non-equivalent nodes (for example, 3^1,5,6-c = 3^1,5,6,c means a net with 11 3-coordinated, one 5-coordinated and one 6-coordinated nodes). The composition of structural units (secondary building units, SBUs) follows the underlying topology name. For the cluster representation, the size of cycles (RINGS) is indicated, which divides intra- or inter-cluster links. The current TopCryst version considers cycles up to size eight, so the representations with larger intercluster cycles are not listed (such large cycle do not correspond to compact cluster groups).

The computational part of the service is written in Python using external libraries os, sys, numpy, scipy, CifFile, itertools and re. It includes the main CrystNet module and 11 additional modules. Reading and parsing input files, as well as preparing output files, is designed using the read_write module. The constants module stores tabular data such as atomic numbers, symbols, radii, etc. The functions for calculating internal coordinates and transforming the coordinate system of a crystal structure are contained in the geometry module. The structure_data module is responsible for checking the completeness and correctness of the structural data received from the CIF file. The Polyhedron and ChemBond modules are used to build the Voronoi partition and to determine the structure connectivity. All symmetry transformations are handled using the symmetry module. The main functionality related to topological analysis of periodic nets is implemented in the structure and functions modules. Two additional modules exception and t1180 ensure uninterrupted operating of the service.

### 3.3. What tasks can be solved with TopCryst?

The current TopCryst version is intended to perform the standard topological analysis, which includes the determination of the local atomic connectivity (coordination numbers of atoms), recognition of structural units and determination of the method of their connection as well as the overall topology of the corresponding underlying net. The results of such analysis can be briefly expressed by the following phrase: ‘The crystal structure X is composed by the structural units Y, which are connected into the Z topological motif,’ and contain the minimal data-set called XYZ block. However, in most cases, additional information is provided depending on the crystal structure features:

1. all representations that can be built based on the chemical and topological structure with their description and classification (see part 2.3);
2. all other structures that have the same underlying net, coordination mode of a ligand, or topological type of a tile.

Besides the ordinary description of a crystal structure, these data enable one to solve typical tasks of crystallochemical analysis, which are considered below.

#### 3.4.1. determination of the underlying topology

In most cases, the determination of the overall topology is not a trivial task. Visual analysis can result to errors because of the structure complexity and/or geometrical distortions of the atomic network. If the degrees of the underlying net nodes are higher than six, the reliable visual determination is possible only in special cases when the underlying net has no more than two or three topologically different nodes, high space-group symmetry, or a well-established isostuctural analog. Thus, the following problematic cases can take place:

1. Different underlying nets have very similar local topologies. Since the visual analysis always rests upon the consideration of a finite part of the crystal structure, such similarity can result in a wrong assignment of the overall topology. For example, the und and unc nets look very similar (Figure 11 [42,43]), but differ from each other starting from the fourth coordination sphere of the nodes.
2. The underlying net is too complicated for the visual analysis. Thus, we will not be able to visually determine the topology of the structure \([\text{Ce}_6(\text{OH})_2\text{O}_4(\text{atph})_6(\text{H}_2\text{O})_6][\text{C}_6\text{H}_7\text{ON})_2(\text{H}_2\text{O})_{10} (\text{H}_2\text{atph} = 2\text{-aminoterephthalic acid})](19,457) [44] (Figure 12).
3. The underlying net is highly distorted. For example, the complex shape of an optically active molecule of myrotheucime A essentially hinders visual analysis of the packing motif, which, however, belongs to a quite common topological type of hexagonal close packing (hcp) [45] (Figure 13).
4. Different underlying nets are geometrically close. In this very rare case, the structures are geometrically similar and hence have the same local topological parameters, but nonetheless their overall topologies are different. For example, \(\beta\)- and \(\gamma\)-HgSeO\(_3\) have the same space group (P2\(_1\)/c), close unit cell dimensions, equal numbers of non-equivalent atoms and the same coordination numbers of the
corresponding atoms that allows one to formally relate these phases to the same structure type. However, the corresponding underlying nets have different topological indices and hence different topologies [46, 47].

3.4.2. relations between structures of different composition and nature

Underlying topology is an important criterion for establishing relations between geometrically and chemically different crystal structures. TopCryst enables the user to find all topological analogs for a particular underlying topology i.e. to immediately relate the compound A under consideration to the corresponding isoreticular series. As a result, the following questions can be answered:

1. Is this topology unknown? In this case, the architecture of A is unique, and hence it can possess some special properties.
2. How numerous is the isoreticular series if the topology has representatives among other compounds? If the topology is rare, the chemical composition of A is very likely unique that could again result in unusual properties. On the contrary, if the isoreticular series contains many representatives, the user can find and explore chemically similar compounds to uncover similar properties in A.
3. Does this topology occur only in a particular class of compounds or in different classes? The answer to this question determines whether the topology is specific and formed, thanks to a special combination of the structural units, or it is common and does not

Figure 11. Underlying nets in the standard representation of the crystal structures of [(Sn(cns))30.964(sb(cns)2Cl)0.1018] (101,591) (left) and (μ3-aminobenzoato)-silver(I) monohydrate [Ag(abz)(H2O)] (abz = aminobenzoato) (FEVGIZ) (right) with the und and unc topologies and the ‘4 1226 4672 106 148 194 248 308,’ ‘4 1226 4874 104 144 186 234 292’ coordination sequences, respectively. Unlike [(Sn(cns))30.964(sb(cns)2Cl)0.1018], the [Ag(abz)(H2O)] structure has additional cluster representations of the 3,3,3,T90 and gra topologies, which are detected by TopCryst.

Figure 12. A complicated underlying net of the crystal structure [Ce2(oh)3.104(atph)0.4(h2o)7][c2(h2o)3](H2o)20 (H2atph = 2-aminoterephthalic acid) (19457) in the standard representation belonging to the 3,4,8T15 topological type (left). On the contrary, the cluster representation is characterized by an easily recognizable face-centered cubic (fcu) topology (right).
depend on the chemical nature of the compound. For example, currently, the TopCryst database contains 39,794 references to crystal structures with the dia topology, and these compounds belong to all chemical classes distinguished in the database. This means that the dia topology is caused by the factors to be common for quite different compounds and is not determined by their chemical composition.

What structural units and what coordination are needed to build this topology? There are clear relations between the local coordination of structural units and the overall topology of the underlying net [47]. The information on the underlying net topology together with coordination modes of ligands enables one to find relations for a given compound or a class of compounds.
3.4.3. relations between different topologies: multilevel analysis

TopCryst provides different representations for the same crystal structure, and this approach can be considered as the analysis of the crystal structure at different levels of its organization. Since such representations differ by chosen structural units, one can establish relations between different coexisting overall topologies thus elucidating deeper structural correspondences between different compounds or even classes of compounds. For example, the topologies of the CaSO$_4$ and NaCl structures are interrelated, since the complete CaSO$_4$ net transforms into the NaCl-type (pcu) net in the standard representation (Figure 14 [48]).

Another example is presented in Figure 6, where relations between the ant, flu and pcu topologies are seen after choosing different structural units in the same crystal structure. Such relations can be used in the analysis of structure assembling or crystal growth as well as in the reticular design of new compounds and materials [49].

4. Concluding remarks

The web service TopCryst described in this paper presents the first attempt of full automation of topological analysis of crystal structures of any complexity and chemical composition. In fact, it can be interfaced with the automatic diffractometers to obtain a primary crystallochemical description of a just determined crystal structure. Besides the formal description of the structure, the service uses a collection of topological databases for the search for relations with hundreds of thousands of other crystal structures at different levels of their organization. This takes the crystallochemical analysis far beyond the traditional crystallographic description and provides much more possibilities in establishing new laws and regularities in material science. Now the service is in progress; we plan to develop interfaces with other ToposPro topological collections including TTM and TTN, and to implement artificial intelligence tools for automated search for correlations between chemical, geometrical and topological descriptors, which are contained in the TopCryst system of databases. This step will enable the scientists over the world to figure out the huge amount of experimental information accumulated in crystallographic databases and to develop other tools for big data analysis and data mining in crystal chemistry and materials science.

Note

1. Hereafter the CSD six-letter Reference Codes or ICSD numerical Collection Codes are given in parentheses.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Ministry of Education and Science of the Russian Federation (project no. 0778-2020-0005).

Notes on contributors

Alexander P. Shevchenko is a Senior Researcher at the Samara Center for Theoretical Materials Science (SCTMS) since 2013. His research interests are related to the use of geometrical and topological approaches for the analysis of crystal structures, the development and computer implementation of algorithms for calculating structural descriptors, as well as the collection of such data in electronic storages. He is a developer of the program package ToposPro since 1989. Now he adjusts machine learning algorithms to the design of crystal structures and prediction of their properties. He develops and administers all SCTMS knowledge databases.

Aleksandr A. Shabalin is the IT Project Manager at the Samara Center for Theoretical Materials Science (SCTMS). His interests concern the development and implementation of websites and scientific web applications, including those based on Big Data, as well as the creation and management of web development teams. Since 2017, he is the SCTMS team leader for the TopologyFinder, OxidationStatePredictor and PathFinder web application projects. He also supports the SCTMS websites based on these web applications.

Igor Yu. Karpukhin is currently a leading mathematician at Samara University. Since graduating from the Kuibyshev State University with a degree in Mathematics, he worked as an engineer, programmer, system administrator and database administrator. Since 2018, he participates in the development of web applications of the Samara Center for Theoretical Materials Science as a lead programmer.

Vladislav A. Blatov is the head of General and Inorganic Chemistry Department at Samara State Technical University, the organizer and director of the Samara Center for Theoretical
Materials Science. His research interests concern geometrical and topological methods in materials science and crystal chemistry and their computer implementation. He is the main developer of the program package ToposPro since 1989. He invented many original algorithms for analyzing and classifying crystal structures, searching for correlations in crystallographic data and predicting new crystalline materials. Now he works under development of knowledge databases and artificial intelligence systems in materials science.

ORCID

Vladislav A. Blatov https://orcid.org/0000-0002-4048-7218

References

[1] Groom CR, Bruno IJ, Lightfoot MP, et al. The Cambridge structural database. Acta Crystallogr Sect B. 2016;72:171–179.
[2] Hellenbrandt M. The Inorganic Crystal Structure Database (ICSD) - Present and future. Crystallogr Rev. 2004;10:17–22.
[3] Gražulis S, Chateigner D, Downs RT, et al. Crystallography Open Database - an open-access collection of crystal structures. J Appl Crystallogr. 2009;42:726–729.
[4] Villars P, Cenzual K. Pearson’s Crystal Data: Crystal Structure Database for Inorganic Compounds (on DVD). Materials Park: ASM International; 2021.
[5] Slater JC. Atomic radii in crystals. J Chem Phys. 1964;41:3199–3204.
[6] Brown ID. The Chemical Bond in Inorganic Chemistry: the Bond Valence Model. Oxford: Oxford University Press; 2006.
[7] Blatov VA. Voronoi–Dirichlet polyhedra in crystal chemistry: theory and applications. Crystallogr Rev. 2004;10:249–318.
[8] Delgado-Friedrichs O, O’-Keeffe M. Crystal nets as graphs: Terminology and definitions. J Solid State Chem. 2005;178:2480–2485.
[9] Delgado-Friedrichs O, Foster MD, O’-Keeffe M, et al. What do we know about three-periodic nets? J Solid State Chem. 2005;178:2533–2554.
[10] O’-Keeffe M, Peskov MA, Ramsden SJ, et al. The Reticular Chemistry Structure Resource (RCSR) database of, and symbols for, crystal nets. Acc Chem Res. 2008;41:1782–1789.
[11] Ramsden SJ, Robins V, Hyde ST. Three-Dimensional Euclidean nets from two-dimensional hyperbolic tilings: kaleidoscopic examples. Acta Crystallogr Sect a. 2009;65:81–108.
[12] Alexandrov EV, Shevchenko AP, Blatov VA. Topological databases: Why do we need them for design of coordination polymers? Cryst Growth Des. 2019;19:2604–2614.
[13] Blatov VA, Shevchenko AP, Proserpio DM. Applied topological analysis of crystal structures with the program package topospro. Cryst Growth Des. 2014;14:3576–3586.
[14] Klee WE. Crystallographic nets and their quotient graphs. Cryst Res Technol. 2004;39:959–968.
[15] O’-Keeffe M, Hyde BG. An alternative approach to non-molecular crystal structures with emphasis on the arrangements of cations. Cation Ordering Electron Transf. Berlin, Heidelberg: Springer; 1985. pp. 77–144.
[16] Alexandrov EV, Blatov VA, Kochetkov AV, et al. Underlying nets in three-periodic coordination polymers: Topology, taxonomy and prediction from a computer-aided analysis of the Cambridge Structural Database. Cryst Eng Comm. 2011;13:3947–3958.
[17] Blatov VA. A method for topological analysis of rod packings. Struct Chem. 2016;27:1605–1611.
[18] Brock AJ, Whittaker JJ, Powell JA, et al. Elastically flexible crystals have disparate mechanisms of molecular movement induced by strain and heat. Angew Chem Int Ed. 2018;57:11325–11328.
[19] Hoffmann R, Kabanov AA, Golov AA, et al. Homo citans and carbon allotropes: for an ethics of citation. Angew Chem Int Ed. 2016;55:10962–10976.
[20] Pokropivny A, Volz S. C8 phase’: Supercubane, tetrahedral, BC-8 or carbon sodalite? Phys Status Solidi. 2012;249:1704–1708.
[21] Akahama Y, Kobayashi M, Kawamura H. Structural studies of pressure-induced phase transitions in selenium up to 150 GPa. Phys Rev B. 1993;47:20–26.
[22] Blatov VA. Nanocluster analysis of intermetallic structures with the program package TOPOS. Struct Chem. 2012;23:955–963.
[23] Bouchaal N, Jemmali M, Bartoli T, et al. Influence of Fe-substitution on structural, magnetic and magneto-caloric properties of Nd2Fe17-x–x–15–x Nickel-rich solutions. J Solid State Chem. 2018;258:501–509.
[24] Akhmetshina TG, Blatov VA. Topological methods for complex intermetallics. Zeitschrift Fur Krist - Cryst Mater. 2017;232:497–506.
[25] Pankova AA, Akhmetshina TG, Blatov VA, et al. A collection of topological types of nanoclusters and its application to icoshedron-based intermetallics. Inorg Chem. 2015;54:6616–6630.
[26] Famprikis T, Kudo O, Foo S, et al. Under pressure: Mechanochemical effects on structure and ion conduction in the sodium-ion solid electrolyte Na3PS4. J Am Chem Soc. 2020;142:18422–18436.
[27] Blatov VA. Crystal structures of inorganic oxocacids salts perceived as cation arrays: a periodic-graph approach. Struct Bond. 2011;138:31–66.
[28] Shevchenko AP, Blatov VA. Simplify to understand: how to elucidate crystal structures? Struct Chem. 2021;32:507–519.
[29] Bonneau C, O’-Keeffe M, Proserpio DM, et al. Deconstruction of crystalline networks into underlying nets: Relevance for terminology guidelines and crystallographic databases. Cryst Growth Des. 2018;18:3411–3418.
[30] Ting G, Li Z, Zheng G, et al. Fluorescent sensing properties of three coordination polymers constructed by 3,5-Di(2’,5-dicarboxylphenyl)pyridine ligand. Zeitschrift Fur Anorg Und Allg Chemie. 2019;645:934–939.
[31] Vologzhanina A, Sokolov A, Purygin P, et al. Knowledge-based approaches to H-Bonding patterns in Hetero-cycle-1-Carboxyhydrazoneamides. Cryst Growth Des. 16:6354–6362. DOI:10.1021/acs.cgd.6b00990.

[32] Blatov VA, Alexandrov EV, Shevchenko AP Topology: ToposPro. In: Comprehensive Coordination Chemistry III. Elsevier Inc.; 2021. pp. 389–412.

[33] Grosjean A, Spackman PR, Edwards AJ, et al. Insights into Host–Guest Binding in hydroquinone clathrates: Single-Crystal X-ray and neutron diffraction, and complementary computational studies on the Hydroquinone-CO2 clathrate. Cryst Growth Des. 2021;21:3477–3486.

[34] Aman F, Asiri AM, Siddiqui WA, et al. Multilevel topological description of molecular packings in 1,2-benzothiazines. Cryst Eng Comm. 2014;16:1963–1970.

[35] Blatov VA, Delgado-Friedrichs O, O’Keeffe M, et al. Three-Periodic nets and tilings: Natural tilings for nets. Acta Crystallogr Sect A Found Crystallogr. 2007;63:418–425. doi:10.1107/S1087436107003287.

[36] Blatova OA, Golov AA, Blatov VA. Natural tilings and free space in zeolites: Models, statistics, correlations, prediction. Zeitschrift Fur Krist - Cryst Mater. 2019;234:421–436.

[37] Blatov VA, Ilyushin GD, Proserpio DM. The zeolite conundrum: Why are there so many hypothetical zeolites and so few observed? A possible answer from the zeolite-type frameworks perceived as packings of tiles. Chem Mater. 2013;25:412–424.

[38] Kuznetsova ED, Blatova OA, Blatov VA. Predicting new zeolites: a combination of thermodynamic and kinetic factors. Chem Mater. 2018;30:2829–2837.

[39] Jin Y-W, Zhu H. oly[[pentaqua(μ-4-pyridine-2,4,6-tricarboxylato)(μ-4-pyridine-2,4,6-tricarboxylato)]disamarium(iii)] monohydrate]. Acta Crystallogr Sect E. 2012;68:m287.

[40] Blatov VA, O’Keeffe M, Proserpio DM. Vertex-, face-, point-, Schlüfi-, and Delaney-symbols in nets, polyhedra and tilings: Recommended terminology. Cryst Eng Comm. 2010;12:44–48.

[41] Jmol/jsmol: an open-source Java viewer for chemical structures in 3D. 2022. https://jmol.sourceforge.net.

[42] Arlt S, Harloff J, Schulz A, et al. Heavy neutral and anionic pnictogen thiocyanates. Inorg Chem. 2019;58:5305–5313.

[43] Du M-X, Wang M-S, Guo G-C. Solvent-Dependent chirality transformation: Construction of a helical homochiral silver(i) MOF using the V-shaped ligand 3-Aminobenzoate. Inorg Chem Commun. 2018;92:22–26.

[44] Zhang Y-F, Wang Q, Xue D-X, et al. Single-Crystal synthesis and diverse topologies of Hexanuclear CeIV-based metal–organic frameworks. Inorg Chem. 2020;59:11233–11237.

[45] Begum S, Chakraborty TK. Cpticl-mediated reductive cyclization: Total synthesis of Pestaloliolactone A, Myrotheciumone A, and Scabrol a. J Org Chem. 2021;86:11812–11821.

[46] Blatov VA, Peskov MV. A comparative crystallochemical analysis of binary compounds and simple anhydrous salts containing pyramidal anions LO3 (L = S, Se, Te, Cl,Br, I). Acta Crystallogr Sect B Struct Sci. 2006;62:457–466.

[47] Alexandrov EV, Shevchenko AP, Asiri AA, et al. New knowledge and tools for crystal design: Local coordination versus overall network topology and much more. Cryst EngComm. 2015;17:2913–2924.

[48] Cheng GC, Zussman J. The crystal structure of anhydrite (CaSO4). Acta Crystallogr. 1963;16:767–769.

[49] Yaghi OM, Kalmutzi MK, Diercks CS. Introduction to reticulocystalogy: Metal-organic frameworks and covalent organic frameworks. Wiley Blackwell; 2019.