

A method for hierarchical comparative analysis of crystal structures

Vladislav A. Blatov

Samara State University, Ac. Pavlov Street 1, Samara 443011, Russia. Correspondence e-mail: blatov@ssu.samara.ru

A geometrical-topological description of crystal structure as a three-dimensional graph with coloured nodes, weighted and coloured edges is used to generate a hierarchical sequence of the structure representations. The solid angles of Voronoi–Dirichlet polyhedra of atoms are used as the edge weights and the nodes and edges are coloured according to chemical reasons. Two operations are defined to derive the representations: contracting an atom to other atoms keeping the local connectivity, and removing an atom together with all its bonds. The atoms of the crystal structure are called origin, removed, contracted or target according to their roles in the operations. Each structure representation is described as a labelled quotient graph and determined by (i) colours of the graph nodes and edges, (ii) some level for edge weights, and (iii) an arrangement of atoms according to their roles. The computer enumeration and topological comparative analysis of all representations for crystal structures of any composition and complexity are implemented into the *TOPOS* program package. The advantages of the method are shown by the analysis of typical inorganic compounds and a molecular packing.

© 2006 International Union of Crystallography
Printed in Great Britain – all rights reserved

1. Introduction

The problem of appropriate classification of crystal structures is one of the principal problems of crystal chemistry. Starting from the first structure determination, crystal chemists tried to find the most valid characteristics that could form an unambiguous system of classification taxons for crystal patterns. Owing to the nature of typical crystallographic data, such characteristics naturally fall into two categories: *geometrical* and *topological*. The geometrical structure parameters (unit-cell dimensions, space group, atomic coordinates, interatomic distances, coordination polyhedra *etc.*) concern the Euclidean metric of crystal space and are obtained directly from experiment. At the same time, they do not contain direct information on interatomic bonding. One can derive such ‘chemical’ information from geometrical features, taking into account the nature of interacting atoms, for instance, by comparing sums of atomic radii with interatomic distances. As a result, some pattern of connectivity can be found that reflects topological features of the structure and is usually represented by a three-dimensional graph. Despite the fact that topological characteristics are closer to the problems of crystal chemistry than geometrical ones, they were rarely considered until the late 1970s. For a long time, coordination number (CN) was the only topological parameter to be exploited in crystallochemical analysis. Thanks to the summary reviews of Wells (1977, 1979), the problem of detailed description of the three-dimensional structure graph came to be considered as an important problem of crystal

chemistry. This approach was further developed by O’Keeffe and co-workers (*e.g.* O’Keeffe & Hyde, 1985, 1996; O’Keeffe *et al.*, 2000; Delgado-Friedrichs *et al.*, 2005), who systematically used new topological structural descriptors (Schläfli and vertex symbols, coordination sequences, tilings) to find and characterize the topologies of ‘ideal’ nets in crystal structures. At present, topological description has become a routine procedure in structure investigations. The theoretical basis of the graph representation has also been intensively developed in the last two decades (Chung *et al.*, 1984; Klee, 1987, 2004; Eon, 1998; Winkler *et al.*, 2001); the authors proposed to reduce the three-dimensional graph to a finite *labelled quotient* graph suitable for computer storage (Blatov, 2000; Delgado-Friedrichs & O’Keeffe, 2003). An intermediate approach that considers both geometrical and topological structure features was first developed for zeolites (Meier, 1968). It introduces the notion of a *secondary building unit* (SBU) as a stable atomic group with a certain topology and well defined geometry. Being an extension of the geometrical notion of a coordination polyhedron (*primary* building unit), the SBU takes into account topological features of a finite region in the three-dimensional graph; its generalized form, *suprapolyhedral invariant*, can serve as a geometrical representation of coordination sequence (Ilyushin & Blatov, 2002).

Crystallochemical classification schemes were developed along with systems of structure descriptors. The oldest geometrical schemes use mainly geometrical structure characteristics and are based on the notion of *structure type* as a primary classification taxon. Although this notion has many

variants, it first considers spatial symmetry and Wyckoff sequences, *i.e.* geometrical parameters, and then atomic coordination as a local topological characteristic (Lima-de-Faria *et al.*, 1990). The structure-type schemes have been widely applied in crystal chemistry as printed [multivolume manuals *Structure Reports* (1931–1990) and *Crystal Structures* (1963–1968); Parthé *et al.* (1993/1994)] and electronic (Bergerhoff *et al.*, 1999) handbooks. Since the 1920s, early topological classification schemes were based on local characteristics of the structure graph, in fact, on the first several members of coordination sequences, although the notion ‘coordination sequence’ itself was introduced much later by Brunner & Laves (1971). Many taxonomists from Machatschki (1928) and Bragg (1930) to Liebau (1985) used the notion of primary building unit to classify inorganic compounds, mainly silicates and phosphates. In the last 20 years, the concept of the SBU [or its chemically stable analog, *molecular building block*, MBB (O’Keeffe *et al.*, 2000)] became more popular to consider various mineral species (Ferraris *et al.*, 1986; Hawthorne *et al.*, 1996; Ferey, 2000) and open-framework structures (Cheetham *et al.*, 1999). This tendency reflects the extension of the analysis of local structure in modern crystal chemistry. The works of Wells (1977, 1979) and O’Keeffe & Hyde (1985, 1996) were the first large-scale topological classifications of basic structure motifs considering net topology *as a whole*. In the last decade, the search for topological relations between simple or more complicated nets has become one of the most important problems of theoretical crystal chemistry (Schindler *et al.*, 1999; Vegas, 2000; Vegas & Jansen, 2002; Delgado-Friedrichs *et al.*, 2005; Hyde *et al.*, 2006). At present, the topological multilevel analysis (focused on SBU, MBB or net topology) is one of the main crystallochemical tools used in the design of new materials (Yaghi *et al.*, 2003; Carlucci *et al.*, 2003; Blatov *et al.*, 2004; Baburin *et al.*, 2005; Öhrström & Larsson, 2005).

Taking into account the large variety even of simple topological motifs found in crystal structures and different chemical ways to separate SBUs in a compound, we need a distinct algorithmic procedure to consider a crystal structure at all allowable levels of its organization. In previous work (Blatov, 2000), we proposed a computer method and developed a program *IsoTest* to enumerate crystal structure representations and to analyse corresponding topologies. These tools were successfully applied to process various classes of inorganic synthetic compounds and minerals (Blatov, 2001; Ilyushin & Blatov, 2002; Ilyushin *et al.*, 2004; Blatov & Peskov, 2006). Below, we present a further development of the method.

2. An approach to hierarchical representation of crystal structure

Hereafter, we assume that some topology is established for the crystal structure to be analysed. This topology can be represented by a three-dimensional graph or by the corresponding labelled quotient graph. Thus, we consider geometrical features of the crystal structure only when assigning topology;

the successive crystal structure analysis, classification and comparison are exclusively based on a topological representation that unambiguously corresponds to a graph.

Strictly speaking, any crystal structure has an infinite number of representations because an infinite number of topologies (nets) can be defined for an infinite (three-dimensional) lattice. However, only a small (and finite) number of them are crystallochemically significant, *i.e.* have crystallochemical meaning. Ordinarily, such significant representations should be chemically proved, *i.e.* should have a topology corresponding to a system of chemical bonds. Obviously, several representations may be considered for a heterodesmic compound; for instance, any organic molecular crystal may be represented as a packing of non-bonded molecules or as a net of molecular centroids connected by intermolecular contacts. However, even a homodesmic compound may be considered in different ways. For instance, the crystal structure of a binary ionic compound *AB* (like NaCl) has two distinct representations: as a packing of *A* and *B* ions (with primitive cubic topology, **pcu**,¹ in the case of NaCl) or as a packing of larger ions, normally anions [*B*], containing interstitial smaller ions (in NaCl, Cl anions form face-centred cubic packing, **fcu**, with Na cations occupying octahedral voids). Nevertheless, this list of representations is not comprehensive. Taking into account the important role of cation arrays in crystallization (O’Keeffe & Hyde, 1985; Vegas, 2000; Vegas & Jansen, 2002), one can be curious about the topology of the cation sublattice [*A*], which has the same (**fcu**) topology in NaCl. In more complicated structures, the number of significant representations increases. For instance, a silicate crystal structure may be represented as a graph of connected O atoms ignoring Si–O bonds; this representation is significant in relation to oxygen packing. However, in zeolites, it is important to consider the topology of the silicate framework with O spacers, or also the sublattice of the framework intrachannels to examine ion-exchange properties. Therefore, to provide a comprehensive classification and to find implicit relations between crystal structures, we need to consider *all* their representations. The main problem is to enumerate them independently of chemical and crystallographic complexity of a substance.

Let us consider a crystal structure containing *Z* atoms in the asymmetric unit; some of them may be chemically equivalent. Our analysis will be based on the complete representation, where all contacts between the atoms are taken into account; some of these contacts (or even all of them, as in crystals of noble gases) may not be chemical bonds. To determine interatomic contacts, we use a Voronoi–Dirichlet partition of crystal space; as has been shown (Blatov, 2004, and references therein), Voronoi–Dirichlet polyhedra can be successfully used for crystallochemical analysis even of complicated chemical substances (Fig. 1). A Voronoi–Dirichlet polyhedron is constructed for each of the *Z* atoms, and all faces of the Voronoi–Dirichlet polyhedra are considered as interatomic contacts.

¹ The bold three-letter codes indicate the net topology according to the RCSR nomenclature (<http://okeeffe-ws1.la.asu.edu/RCSR/home.htm>).

All the contacts should be classified, depending on the peculiarities of the crystallochemical task. Using parameters of Voronoi–Dirichlet polyhedra, one can separate strong (valence or ionic), specific (H bond, agostic or secondary) or van der Waals (non-valence) bonds (Blatov, 2004). In addition, indirect interatomic contacts (O’Keeffe, 1979), whose lines do not intersect corresponding faces of Voronoi–Dirichlet polyhedra, are considered as non-bonding and form no graph edges. The strength of an atomic interaction of a given type is assumed to be proportional to the solid angle of the corresponding face (Blatov, 2004). Thus, the Voronoi–Dirichlet partition maps onto a three-dimensional graph, whose edges are coloured according to the type of atomic interaction, and have weights equal to the solid angles (Fig. 1).

We assume that the complete representation contains all information on the crystal structure topology; all significant representations may be deduced as its subsets by the following three-step procedure.

(i) The graph edges of the same colour are taken into account, depending on the task conditions. Other edges are either ignored or considered irrespective of their weights. In most cases, the chemical interactions of only one type are of interest; as a rule those are strong bonds. If two or more types of bonds are to be analysed, the bonds of only one type are to be considered at a given pass of the procedure. Then an array of the weights (solid angles) is formed for all the edges with the same colour. For instance, when analysing H-bonded molecular packings, the solid angles for H bonds should be arranged at this step; valence bonds should be treated regardless of their weights, and all van der Waals contacts should be forgotten.

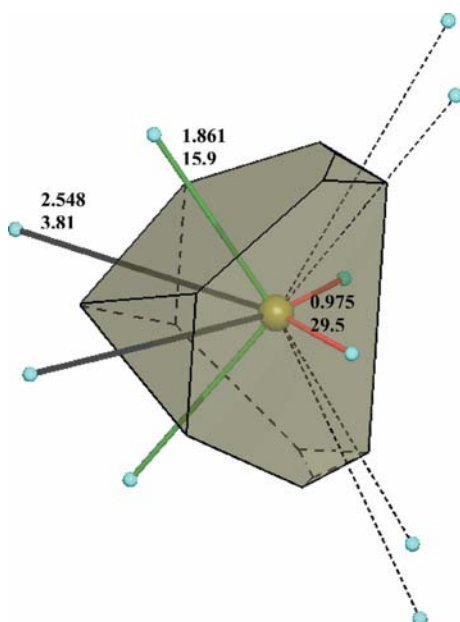


Figure 1
The Voronoi–Dirichlet polyhedron and environment of an O atom in the crystal structure of ice VIII (Besson *et al.*, 1994). Valence, H-bond and non-valence interatomic contacts are coloured red, green and black, respectively. Indirect contacts, forming no graph edges, are dashed. Interatomic distances (Å) and solid angles (in percentage of 4π sr) are given for each non-equivalent graph edge.

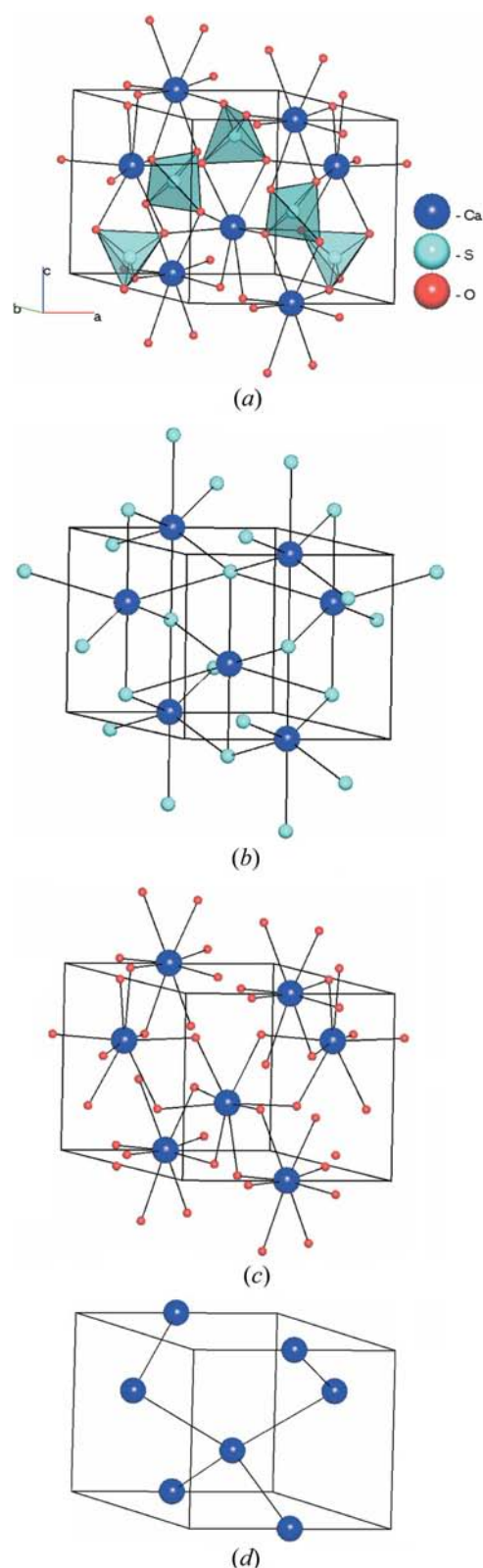


Figure 2
 γ -CaSO₄ crystal structure (Lager *et al.*, 1984, space group $P6_222$): (a) complete representation ($\{\text{Ca}, \text{S}, \text{O}\}, \emptyset, \emptyset, \emptyset$), and its subrepresentations (b) ($\{\text{Ca}, \text{S}\}, \emptyset, \{\text{O}\}, \{\text{S}\}$) with origin Ca and S atoms, contracted O atoms and target S atoms (the **sma** topology); (c) ($\{\text{Ca}, \text{O}\}, \{\text{S}\}, \emptyset, \emptyset$) with origin Ca and O atoms, and removed S atoms; (d) ($\{\text{Ca}\}, \{\text{S}\}, \{\text{O}\}, \{\text{Ca}\}$) with origin and target Ca atoms, removed S atoms and contracted O atoms (the **qtz** topology).

(ii) The entire array of solid angles is divided into several groups by a clustering algorithm. We have used a simple approach where two solid angles belong to the same group if their difference is smaller than a given value [a typical solid angle estimation error, 1.5% of 4π sr, should normally be applied (Blatov, 2004)]. Thus, n distinct coordination spheres are separated in the atomic environment. Then different topologies are generated by successively rejecting the farthest coordination sphere. As a result, $n - 1$ additional representations of the crystal structure are produced from the complete one. It is important that no 'best' representations are chosen at this step but all levels of interatomic interaction are clearly distinguished for further analysis, depending on the matter in hand.

(iii) Each of the n representations is used to generate a set of subrepresentations according to an improved scheme initially proposed by Blatov (2000). Every subrepresentation is unambiguously determined by an arrangement of the set $\{Z\}$ of all Z atoms into the four subsets: origin $\{OA\}$, removed $\{RA\}$, contracted $\{CA\}$ and target $\{TA\}$ atoms. Two operations are defined on the subsets to derive a graph of the subrepresentation from the graph of an initial i th representation: contracting an atom to other atoms keeping the local connectivity, when the atom is suppressed, but all graph paths passing through it are retained (Figs. 2*a, b*), and removing an atom together with all its bonds (Figs. 2*c, d*). The four-subset arrangement is determined by the role of atoms in those operations. Namely, origin atoms form a new net that characterizes the subrepresentation topology; removed atoms are eliminated from the initial net by the removing operation; contracted atoms merge with target atoms, passing the bonds to them.

All the sets $\{OA\}$, $\{RA\}$, $\{CA\}$ and $\{TA\}$ form a collection ($\{OA\}$, $\{RA\}$, $\{CA\}$, $\{TA\}$) that, together with the initial representation, unambiguously determines the subrepresentation topology (Figs. 2*a–d*). As a result, all crystal-structure representations are generated with the coloured edges selected at step (i); therefore, the representations should be called valence (H bond, van der Waals *etc.*) according to the chosen bond type. If other types of atomic interactions are to be taken into account, the procedure is repeated for each of the found representations with another colour filter. For instance, considering H-bonded complex groups, we may first enumerate all representations of the complex groups depending on the strength of intermolecular H bonds, and then analyse all types of intramolecular valence bonding for every H-bond representation.

With the concept of collection, the successful enumeration of the significant subrepresentations becomes easily formalizable as a computer algorithm. Firstly, any collection has a number of properties reflecting the crystal-structure relations that can be formulated in terms of set theory.

(i) $\{OA\} \cap \{RA\} = \emptyset$; $\{OA\} \cap \{CA\} = \emptyset$, $\{RA\} \cap \{CA\} = \emptyset$ because an atom cannot play more than one role in the crystal structure.

(ii) $\{OA\} \cup \{RA\} \cup \{CA\} = \{Z\}$, *i.e.* every atom must have a crystallochemical role.

(iii) $\{OA\} \neq \emptyset$, other sets may be empty. This property arises because only the origin atoms are nodes in the graph of the crystal structure subrepresentation; other atoms determine the graph topology. Obviously, the collection ($\{OA\}, \emptyset, \emptyset, \emptyset$) means that $\{OA\} = \{Z\}$; it describes the initial representation.

(iv) $\{TA\} \subseteq \{OA\}$, because the target atoms are always selected from the origin atoms; unlike other origin atoms they are the centres of complex structural groups.

(v) $\{TA\} \neq \emptyset \Leftrightarrow \{CA\} \neq \emptyset$ because the target and contracted atoms together form the structural groups.

Secondly, the collections, together with the topological operations, map onto *all* the crystal structure transformations applied in crystallochemical analysis. Namely, origin atoms correspond to the centres of structural groups in a given structure consideration. If a structural group has no distinct central atom, a pseudoatom (PA) coinciding with the group's centroid should be added to the $\{Z\}$ set; this case is typical in the analysis of molecular packings. Removed atoms are atoms to be ignored in the current crystal-structure representation, like atoms of interstitial ions and molecules in porous substances or, say, alkali metals in framework coordination compounds. Contracted atoms, together with target atoms, form complex structural groups but the contracted atoms are not directly considered; they merely provide the structure connectivity whereas the target atoms coincide with the groups' centroids. The difference between origin and target atoms is that the target atoms always correspond to *polyatomic* structural groups whereas the origin atoms symbolize all structural units, both mono- and polyatomic.

The number of subrepresentations is determined only by the number of *colours* of the graph nodes (atoms) in the initial representation. The colours distinguish topologically different atoms in a current structure description. Atoms of the same colour must belong to the same $\{Z\}$ subset. One can consider all the Z atoms as of different colours; this is the most common approach. However, in most cases, it is reasonable to unite the atoms of the same chemical element into the same colour group to reduce the total number of subrepresentations and to simplify topological analysis. Sometimes, chemically equivalent atoms should be distinguished if they play different crystallochemical roles. For instance, in magnetite Fe_3O_4 , the Fe atoms with different coordination numbers should be distinguished in the $\{Z\}$ subsets. If, say, there are two atoms of different colours, A and B , $\{A, B\} = \{Z\}$, the following four subrepresentations are possible for the initial representation ($\{A, B\}, \emptyset, \emptyset, \emptyset$):

(i) ($\{A\}, \{B\}, \emptyset, \emptyset$), *i.e.* the subnet of A atoms;

(ii) ($\{A\}, \emptyset, \{B\}, \{A\}$), *i.e.* the net of A atoms with the $A-B-A$ bridges (B atoms are spacers);

(iii) ($\{B\}, \{A\}, \emptyset, \emptyset$);

(iv) ($\{B\}, \emptyset, \{A\}, \{B\}$).

Nets of (iii) and (iv) are obtained by interchanging A and B in (i) and (ii).

Using the collection properties, it is easily seen that no other subrepresentations apart from these four may be constructed. As is shown below, some of them may have trivial topology

Table 1

Interatomic contacts corresponding to faces of Voronoi–Dirichlet polyhedra in the crystal structure of α -SrSi₂.

Atom	R (Å)	Ω (%)	Atom	R (Å)	Ω (%)
Si environment (CN = 10)			Sr environment (CN = 20)		
Si	2.394×3	19.58	Si	3.251×6	8.95
Sr	3.251×3	8.95	Si	3.373×2	8.66
Sr	3.373	8.66	Si	3.830×6	1.92
Sr	3.830×3	1.92	Sr	4.002×6	2.91

(disconnected graph) or infinite-connected nodes but the remaining subrepresentations can be crystallochemically significant.

This approach has been implemented into the crystallochemical software package *TOPOS* (Blatov, 2004), which is available at <http://www.topos.ssu.samara.ru>. With the program *AutoCN* of the package, one can classify interatomic contacts and then can generate all crystal structure representations as labelled quotient graphs using both solid angles and interatomic distances as weights of the graph edges. Subsequent enumeration of the subrepresentations is done with the improved *IsoTest* program. Since every representation is described as a net graph, a number of conventional topological

descriptors (Delgado-Friedrichs & O’Keeffe, 2005; Öhrström & Larsson, 2005) may be used to characterize the graph topology for further comparison and storage in electronic databases. The following descriptors are applied in *TOPOS*:

Coordination sequence (CS) $\{N_k\}$ is a set of sequential numbers N_1, N_2, \dots of atoms in first, second *etc.* coordination spheres of an atom in the net. The first ten coordination spheres are usually considered at the topological classification. The coordination number is equal to N_1 and the graph node is called N_1 -connected or N_1 -coordinated.

Extended Schläfli symbol (ES) contains a detailed description of all shortest circuits (closed chains of connected atoms) for each angle at each non-equivalent atom. The *total Schläfli symbol* summarizes all the Schläfli symbols for the non-equivalent atoms with stoichiometric coefficients.

Vertex symbol (VS) gives information similar to ES but for rings (circuits without shortcuts).

The CS + ES + VS combination allows one to distinguish any topologically non-equivalent nets (no counterexamples are known). If several atoms in the asymmetric unit have the same CS, ES and VS, they are assumed to be topologically equivalent (a topological supersymmetry exists). Thus, the net topology is determined by a set of CS + ES + VS combinations for all topologically non-equivalent atoms. Two nets are assumed to be isomorphic if an isomorphism exists between their sets of CS + ES + VS combinations. We call the corresponding crystal structures *isotypic* (at a given representation), or belonging to the same *topological type*. A procedure of automatic arrangement of all crystal-structure representations into topological types irrespective of the structure composition and complexity is realized in the *IsoTest* program. Then the topological types can be related to the archetypes collected in the *TOPOS* topological databases. At present, the databases contain more than 18000 two- and three-dimensional net topologies, including information from *RCSR* (<http://okeeffe-ws1.la.asu.edu/RCSR/home.htm>), framework zeolites (<http://www.iza-structure.org/databases/>) and *EPINET* (<http://epinet.anu.edu.au>) databases, and on sphere packings (Koch *et al.*, 2006, and references therein).

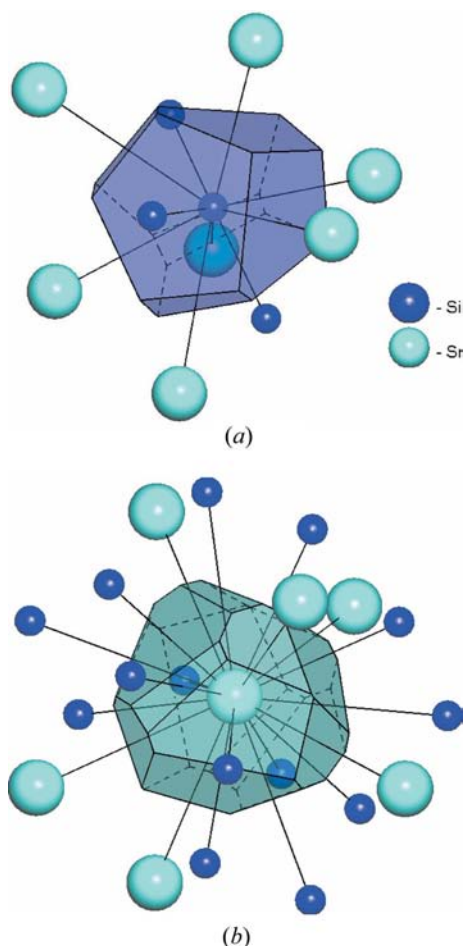


Figure 3 Voronoi–Dirichlet polyhedra and environment of (a) the Si atom and (b) the Sr atom in the crystal structure of α -SrSi₂.

3. Examples of the analysis

In all the cases below, the difference level of 1.5% was used to cluster solid angles corresponding to interatomic contacts. The graph nodes were distinguished chemically: they had the same colour if and only if corresponding atoms were of the same chemical element.

3.1. Metallic and covalent bonding: α -SrSi₂

Data on interatomic distances and solid angles for α -SrSi₂ (Evers, 1978) within the first coordination spheres of Si and Sr atoms (Figs. 3a, b) are given in Table 1. Obviously, three groups of contacts can be selected both by solid angle and by distance criteria. Three *valence* representations can be produced:

Table 2

Interatomic contacts with O atoms corresponding to faces of Voronoi–Dirichlet polyhedra for Ca and S atoms in the crystal structure of γ -CaSO₄.

R (Å)	Ω (%)	R (Å)	Ω (%)
Ca environment (CN = 8)		S environment (CN = 4)	
2.384×4	14.49	1.477×4	24.95
2.543×4	10.51	3.827×4†	0.05

† Non-bonding contacts.

(i) $\Omega > 19.5\%$, $R < 2.4$ Å, with only covalent Si–Si bonds. This net is well known in crystal chemistry: it represents the topology **srs**, or 3/10/*c*1 sphere packing (Koch *et al.*, 2006), being widespread in three-dimensional nets both in inorganic (Blatov *et al.*, 2004) and in metal-organic frameworks (Ockwig *et al.*, 2005; Baburin *et al.*, 2005). No non-trivial subrepresentations are possible in this case.

(ii) $\Omega > 8.5\%$, $R < 3.4$ Å, with strong Si–Sr bonds. This representation characterizes the α -SrSi₂ structure type with CN_{Si} = 7 and CN_{Sr} = 8. The following subrepresentations arise from enumerating collections:

(iia) ({Si}, {Sr}, \emptyset , \emptyset) is equivalent to the representation (i);

(iib) ({Sr}, {Si}, \emptyset , \emptyset) has trivial topology because the Sr atoms do not directly connect with each other; removing the Si links leads to a disconnected graph;

(iic) ({Si}, \emptyset , {Sr}, {Si}) gives a 26-connected uninodal Si net that describes the topology of the Si array taking into account the Si–Sr–Si bridges.

(iid) ({Sr}, \emptyset , {Si}, {Sr}) corresponds to a structure with infinite polymeric ligand (Si)_∞ because the contracted Si atoms form a three-dimensional net. The topology of the Sr array cannot be described as a graph because each Sr atom is connected with all other Sr atoms through the –(Si)_n–bridges.

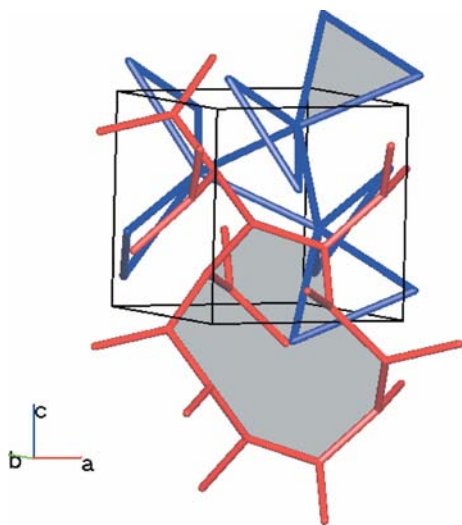


Figure 4

Interpenetrating 3-connected **srs** (red) and 6-connected **lcy** (blue) nets in the crystal structure of α -SrSi₂. The smallest 10- and 3-rings are shaded in the nets.

Table 3

Subrepresentations for the complete representation of the γ -CaSO₄ crystal structure.

No.	Subrepresentation	Crystallochemical description
1†	(({Ca,S},{O}, \emptyset , \emptyset)	Array of isolated Ca and S atoms
2	(({Ca,O},{S}, \emptyset , \emptyset)	Three-dimensional Ca–O subnet
3†	(({S,O},{Ca}, \emptyset , \emptyset)	Isolated sulfate ions
4	(({Ca,S}, \emptyset ,{O},{Ca,S})	Three-dimensional Ca–S net with O spacers
5	(({Ca,S}, \emptyset ,{O},{Ca})	Three-dimensional net of [CaO _n] units and S atoms
6	(({Ca,S}, \emptyset ,{O},{S})	Three-dimensional net of Ca ²⁺ and [SO ₄] ²⁻ ions
7	(({Ca,O}, \emptyset ,{S},{Ca,O})	Three-dimensional Ca–O net with S spacers
8	(({Ca,O}, \emptyset ,{S},{Ca})	Equivalent to No. 2‡
9	(({Ca,O}, \emptyset ,{S},{O})	Equivalent to No. 7‡
10	(({S,O}, \emptyset ,{Ca},{S,O})	Three-dimensional net of [SO ₄] ²⁻ ions with Ca spacers
11†	(({S,O}, \emptyset ,{Ca},{S})	Equivalent to No. 3‡
12	(({S,O}, \emptyset ,{Ca},{O})	Equivalent to No. 10‡
13†	(({Ca},{S,O}, \emptyset , \emptyset)	Isolated Ca ²⁺ ions
14	(({Ca},{S},{O},{Ca})	Three-dimensional net of Ca ²⁺ ions with O spacers
15†	(({Ca},{O},{S},{Ca})	Equivalent to No. 13‡
16	(({Ca}, \emptyset ,{S,O},{Ca})	Three-dimensional net of Ca ²⁺ ions with [SO ₄] ²⁻ spacers
17†	(({S},{Ca,O}, \emptyset , \emptyset)	Isolated S atoms
18†	(({S},{Ca},{O},{S})	Isolated centres of SO ₄ ions
19†	(({S},{O},{Ca},{S})	Equivalent to No. 17‡
20†	(({S}, \emptyset ,{Ca,O},{S})	S atoms connected with (Ca,O) _∞ polymeric ligands
21†	(({O},{Ca,S}, \emptyset , \emptyset)	Isolated O atoms
22†	(({O},{Ca},{S},{O})	Isolated tetrahedral [O ₄] groups with S spacers
23	(({O},{S},{Ca},{O})	Three-dimensional net of O atoms with Ca spacers
24	(({O}, \emptyset ,{Ca,S},{O})	Three-dimensional net of O atoms with Ca and S spacers

† Not a three-dimensional net, or existence of polymeric ligands. ‡ Since Ca and S atoms are not directly connected.

(iii) The complete representation with all contacts from Table 1, where CN_{Si} = 10 and CN_{Sr} = 20. Because the number of subrepresentations depends on the number of colours of atoms only, the similar (iii*a*), (iii*b*), (iii*c*) and (iii*d*) collections may be considered. Again, the (iii*a*) subrepresentation gives the 3-connected **srs** Si net, but (iii*b*) is not trivial thanks to weaker Sr–Sr metallic bonding, and the corresponding graph has 6-connected nodes and the **lcy** topology (6/3/*c*1 sphere packing). Considering the (iii*a*) and (iii*b*) subrepresentations together, one can mention that α -SrSi₂ can be represented as an interpenetrating array of **srs** Si and **lcy** Sr nets if all the Si–Sr contacts are omitted (Fig. 4). This seemingly sophisticated description for this structure can be quite reasonable in more complicated cases. In particular, according to Baburin *et al.* (2005), the crystal structure of Li₂Pt₃B is an interpenetration of an **srs** Li net and an **lcy** net of Pt₃B structural groups. Lastly, both (iii*c*) and (iii*d*) subrepresentations contain polymeric ligands and are insignificant.

3.2. Ionic and covalent bonding: γ -CaSO₄

The crystal structure of calcium sulfate (Lager *et al.*, 1984, space group *P6₂22*) is chosen as a more complicated example

of a ternary compound containing rather large alkaline-earth cations whose CNs are often a problem to be unambiguously determined. In the case of γ -CaSO₄, the calcium environment can be split into two coordination spheres, and in total there are three distinct groups of interatomic contacts and three *valence* representations (Table 2):

(i) $\Omega > 24.5\%$, $R < 1.45 \text{ \AA}$. This case is insignificant because it corresponds to isolated arrays of disconnected Ca cations and SO₄ anions;

(ii) $\Omega > 14.0\%$, $R < 2.35 \text{ \AA}$;

(iii) $\Omega > 10.5\%$, $R < 2.55 \text{ \AA}$.

These last two representations are similar; they describe three-dimensional CaSO₄ nets with mainly covalent bonds within SO₄ ions and ionic contacts between Ca²⁺ and [SO₄]²⁻ nodes, however the net topologies are different. Below, we analyse in detail the complete (iii) representation. A successive enumeration of collections constructed for $\{Z\} = \{\text{Ca}, \text{S}, \text{O}\}$ results in the subrepresentations given in Table 3. The complete representation and the most significant subrepresentations Nos. 6, 2 and 14 are shown in Figs. 2(a)–(d).

Close inspection of Table 3 shows that, thanks to peculiarities of the structure topology, at least half of the subrepresentations (11 out of 24) describe disconnected atomic arrays and some of the remaining ones are topologically equivalent. Only the ten subrepresentations 2, 4–7, 10, 14, 16, 23 and 24 should be tested for crystallochemical significance. In turn, only two of them (Nos. 6 and 14) have well known crystallochemical topology.

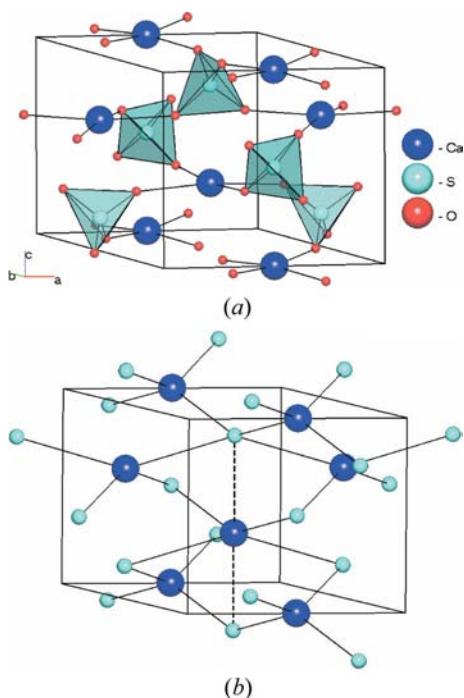


Figure 5
(a) γ -CaSO₄ crystal structure representation at $\Omega > 14.0\%$; (b) its subrepresentation $(\{\text{Ca}, \text{S}\}, \emptyset, \{\text{O}\}, \{\text{S}\})$ with the **qtz** topology. The dashed lines show the edges to be added to transform the **qtz** topology into the **sma** topology (cf. Fig. 2b).

Subrepresentation No. 6 $(\{\text{Ca}, \text{S}\}, \emptyset, \{\text{O}\}, \{\text{S}\})$ (Fig. 2b) describes the CaSO₄ crystal structure as a packing of Ca and SO₄ ions with the **sma** topology, known also as 6-connected 6/4/h3 sphere packing (Sowa *et al.*, 2003). Both Ca and SO₄ nodes have the same topological descriptors so they are topologically equivalent. This description is crystallochemically the most reasonable but not the only possible one.

Subrepresentation No. 14 $(\{\text{Ca}\}, \{\text{S}\}, \{\text{O}\}, \{\text{Ca}\})$ (Fig. 2d) corresponds to the subnet of Ca²⁺ ions connected by O spacers; the S atoms are ignored. Although such a representation of an inorganic salt is unusual, it allows one to keep track of genetic relations between different net topologies. Indeed, this subrepresentation has the topology of a 4-connected Si net in quartz (**qtz**, 4/6/h1 sphere packing); this means that **sma** and **qtz** topologies could be related. Really, we can obtain the **qtz** net from **sma** by breaking two links; this choice is unambiguous in space symmetry *P6₂22*. In this case, the relation is evident due to symmetry because **sma**

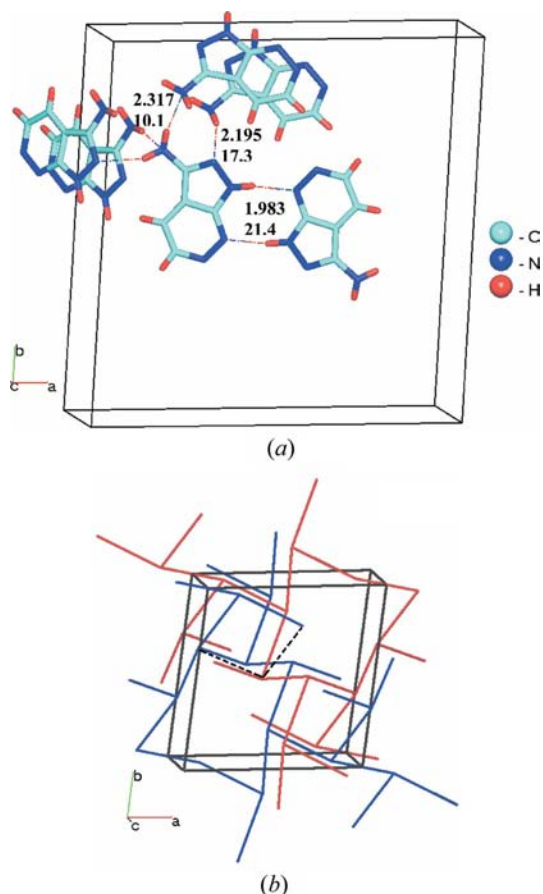


Figure 6
(a) Complete representation of the crystal structure of 3-amino-1H-pyrazolo(3,4-c)pyridazine with H bonds (dot-and-dash lines). Interatomic distances N...H (Å) and corresponding solid angles (in percentage of $4\pi \text{ sr}$) are given for each non-equivalent H bond. (b) The subrepresentation $(\{\text{PA}\}, \emptyset, \{\text{Z}\}-\{\text{PA}\}, \{\text{PA}\})$ for H-bond representation at $\Omega(\text{N}\cdots\text{H}) > 17\%$. The two interpenetrating molecular arrays of the **lig** topology are differently coloured. Some additional weak H bonds with $\Omega(\text{N}\cdots\text{H}) = 10.1\%$ leading to the 5/4/t5 topology are shown by dashed lines.

corresponds to the Q (quartz) lattice complex (Fischer & Koch, 2002). Finally, the **qtz** topology describes also the $(\{\text{Ca}, \text{S}\}, \emptyset, \{\text{O}\}, \{\text{S}\})$ subrepresentation of representation (ii) with 4-coordinated Ca atoms (Figs. 5*a*, *b*).

3.3. Molecular packing with H bonds: 3-amino-1*H*-pyrazolo(3,4-*c*)pyridazine

While analysing packing of molecules, one should choose molecular centroids (pseudoatoms, PA) as target atoms and all the atoms of the molecules as contracted atoms. In this case, the set $\{Z\}$ should be extended with the $\{\text{PA}\}$ set. A typical collection corresponding to a molecular packing is $(\{\text{PA}\}, \emptyset, \{Z\}-\{\text{PA}\}, \{\text{PA}\})$. Naturally, the initial representation should include at least two clusters of interatomic contacts corresponding to valence (intramolecular) and non-valence (H-bonded or van der Waals) contacts. As a result, the molecules will be contracted to their centroids and such a subrepresentation will be described by a net of centroids connected by intermolecular contacts.

As an example, let us consider the crystal structure of 3-amino-1*H*-pyrazolo(3,4-*c*)pyridazine (Cacciari *et al.*, 2003). There are three groups of H bonds distinguished by solid angles and lengths of $\text{N}\cdots\text{H}$ contacts (Fig. 6*a*). Obviously, it is not reasonable to examine different valence representations in this case (molecules should be considered as a whole), therefore only *H-bond* representations, and only their molecular packing subrepresentations $(\{\text{PA}\}, \emptyset, \{Z\}-\{\text{PA}\}, \{\text{PA}\})$, are studied below.

(i) $\Omega(\text{N}\cdots\text{H}) > 21\%$, $R(\text{N}\cdots\text{H}) < 2.0 \text{ \AA}$. In this case, only the strongest H bonds are taken into account and the packing consists of dimers (Fig. 6*a*).

(ii) $\Omega(\text{N}\cdots\text{H}) > 17\%$, $R(\text{N}\cdots\text{H}) < 2.2 \text{ \AA}$. The next group of H bonds connects the dimers into two interpenetrating three-dimensional arrays where every molecule is 3-connected (Fig. 6*b*). Either of the arrays has the LiGe (**lig**) topology that is also known as $3/8/t1$ sphere packing (Koch *et al.*, 2006). Two interpenetrating **lig** arrays were recently found in valence-bonded metal-organic frameworks (Blatov *et al.*, 2004).

(iii) $\Omega(\text{N}\cdots\text{H}) > 10\%$, $R(\text{N}\cdots\text{H}) < 2.35 \text{ \AA}$. Considering all H bonds, one can unite the two 3-connected **lig** arrays into a 5-connected single framework of the $5/4/t5$ topology (Fig. 6*b*). Thus, the two sphere packings, $5/4/t5$ and $3/8/t1$, can be closely related in the packings of molecules.

4. Concluding remarks

Crystallochemical description of a crystal structure traditionally rests upon selecting structural groups according to chemical reasoning. These reasons always have a local character and arise from chemists' long experience of dealing with finite molecular groups that are stable in gaseous or liquid phases. So the last part of the term *crystallochemical* still dominates over the first. To find *crystallochemical* regularities, *i.e.* common properties of extended atomic arrays on a global level of their organization, we need to search for typical infinite periodic motifs. In this way, geometrical descriptors (bond

distances and angles, types of coordination polyhedra *etc.*), being efficient for small molecular units, become inappropriate to identify rather flexible polymeric structural groups. Thus, the topology of periodic nets is the fundamental characteristic to identify the types of crystal structure organization. As was stated above, the mathematical tools of the topological approach have been highly developed during the last few years, and now the problem is arising of systematically searching for typical periodic motifs. Although this problem has been solved by analysing both natural (Wells, 1977, 1979; O'Keeffe & Hyde, 1985, 1996; Vegas, 2000; Vegas & Jansen, 2002) and model nets (Koch *et al.*, 2006) for the last few decades, the results obtained embrace a small part of the data sets accumulated in crystallographic databases worldwide. This work pursued two main objectives to change this situation:

(i) an algorithm and a computer program should be developed to enumerate successively all possible representations of a crystal structure as periodic (usually three-dimensional) array(s);

(ii) all the representations should topologically be studied irrespective of their *chemical* meaning. Only in this way may *crystallochemical* regularities be found that can differ from familiar chemical regularities. The examples given above show that typical crystal-structure topologies usually correspond to chemically reasonable representations.

An important application of the proposed method is the possibility to establish genetic relations between nets with different and with the same connectivity. Thus, the net with a higher connectivity can be transformed to a simpler net by breaking some edges, as in the **sma-qtz** example. The relation between two nets with equal coordination numbers of atoms (for instance, between 6-connected **sma** and **ley** nets) should be established through an intermediate common net with higher or lower connectivity. Systematic investigation of such topological relationships is in progress.

References

- Baburin, I. A., Blatov, V. A., Carlucci, L., Ciani, G. & Proserpio, D. M. (2005). *J. Solid State Chem.* **178**, 2452–2474.
- Bergerhoff, G., Berndt, M., Brandenburg, K. & Degen, T. (1999). *Acta Cryst.* **B55**, 147–156.
- Besson, J. M., Pruzan, P., Klotz, S., Hamel, G., Silvi, B., Nelmes, R. J., Loveday, J. S., Wilson, R. M. & Hull, S. (1994). *Phys. Rev. B*, **49**, 12540–12550.
- Blatov, V. A. (2000). *Acta Cryst.* **A56**, 178–188.
- Blatov, V. A. (2001). *Z. Kristallogr.* **216**, 165–171.
- Blatov, V. A. (2004). *Crystallogr. Rev.* **10**, 249–318.
- Blatov, V. A., Carlucci, L., Ciani, G. & Proserpio, D. M. (2004). *CrystEngComm*, **6**, 377–395.
- Blatov, V. A. & Peskov, M. V. (2006). *Acta Cryst.* **B62**, 457–466.
- Bragg, W. L. (1930). *Z. Kristallogr.* **74**, 237–305.
- Brunner, G. O. & Laves, F. (1971). *Wiss. Z. Tech. Univ. Dresden*, **20**, 387–390.
- Cacciari, B., Spalluto, G. & Ferretti, V. (2003). *J. Heterocycl. Chem.* **40**, 1065–1069.
- Carlucci, L., Ciani, G. & Proserpio, D. M. (2003). *Coord. Chem. Rev.* **246**, 247–289.

- Cheetham, A. K., Ferey, G. & Loiseau, T. (1999). *Angew. Chem. Int. Ed. Engl.* **38**, 3268–3292.
- Chung, S. J., Hahn, Th. & Klee, W. E. (1984). *Acta Cryst.* **A40**, 42–50.
- Delgado-Friedrichs, O., Foster, M. D., O’Keeffe, M., Proserpio, D. M., Treacy, M. M. J. & Yaghi, O. M. (2005). *J. Solid State Chem.* **178**, 2533–2554.
- Delgado-Friedrichs, O. & O’Keeffe, M. (2003). *Acta Cryst.* **A59**, 351–360.
- Delgado-Friedrichs, O. & O’Keeffe, M. (2005). *J. Solid State Chem.* **178**, 2480–2485.
- Eon, J.-G. (1998). *J. Solid State Chem.* **138**, 55–65.
- Evers, J. (1978). *J. Solid State Chem.* **24**, 199–207.
- Ferey, G. (2000). *J. Solid State Chem.* **152**, 37–48.
- Ferraris, G., Mellini, M. & Merlino, S. (1986). *Rend. Soc. Ital. Mineral. Petrol.* **41**, 181–192.
- Fischer, W. & Koch, E. (2002). *International Tables for Crystallography*, Vol. A, 5th ed., edited by Th. Hahn, Section 14.1. Dordrecht/Boston/London: Kluwer Academic Publishers.
- Hawthorne, F. C., Burns, P. C. & Grice, J. D. (1996). *Rev. Mineral.* **33**, 41–115.
- Hyde, S. T., Delgado-Friedrichs, O., Ramsden, S. J. & Robins, V. (2006). *Solid State Sci.* **8**, 740–752.
- Ilyushin, G. D. & Blatov, V. A. (2002). *Acta Cryst.* **B58**, 198–218.
- Ilyushin, G. D., Blatov, V. A. & Zakutkin, Yu. A. (2004). *Z. Kristallogr.* **219**, 468–478.
- Klee, W. E. (1987). *Z. Kristallogr.* **179**, 67–76.
- Klee, W. E. (2004). *Cryst. Res. Technol.* **39**, 959–968.
- Koch, E., Fischer, W. & Sowa, H. (2006). *Acta Cryst.* **A62**, 152–167.
- Lager, G. A., Armbruster, T., Rotella, F. J., Jorgensen, J. D. & Hinks, D. G. (1984). *Am. Mineral.* **69**, 910–918.
- Liebau, F. (1985). *Structural Chemistry of Silicates: Structure, Bonding and Classification*. Berlin: Springer.
- Lima-de-Faria, J., Hellner, E., Liebau, F., Makovicky, E. & Parthé, E. (1990). *Acta Cryst.* **A46**, 1–11.
- Machatschki, F. C. (1928). *Zentralbl. Mineral. Geol. Palaeontol. Abt. A*, pp. 97–104.
- Meier, W. M. (1968). *Molecular Sieves*, pp. 10–27. London: Society of Chemistry and Industry.
- Ockwig, N. W., Delgado-Friedrichs, O., O’Keeffe, M. & Yaghi, O. M. (2005). *Acc. Chem. Res.* **38**, 176–182.
- Öhrström, L. & Larsson, K. (2005). *Molecule-Based Materials. The Structural Network Approach*. Amsterdam: Elsevier.
- O’Keeffe, M. (1979). *Acta Cryst.* **A35**, 772–775.
- O’Keeffe, M., Eddaoudi, M., Li, H., Reineke, T. & Yaghi, O. M. (2000). *J. Solid State Chem.* **152**, 3–20.
- O’Keeffe, M. & Hyde, B. G. (1985). *Struct. Bond.* **61**, 77–144.
- O’Keeffe, M. & Hyde, B. G. (1996). *Crystal Structures. I. Patterns and Symmetry*. Washington, DC: Mineralogical Society of America.
- Parthé, E., Gelato, L., Chabot, B., Penzo, M., Cenxual, K. & Gladyshevskii, R. (1993/1994). *TYPIX. Standardized Data & Crystal Chemical Characterization of Inorganic Structure Types. Gmelin Handbook of Inorganic & Organometallic Chemistry*, Vols. 1–4. Berlin: Springer.
- Schindler, M., Hawthorne, F. C. & Baur, W. H. (1999). *Acta Cryst.* **B55**, 811–829.
- Sowa, H., Koch, E. & Fischer, W. (2003). *Acta Cryst.* **A59**, 317–326.
- Vegas, A. (2000). *Crystallogr. Rev.* **7**, 189–283.
- Vegas, A. & Jansen, M. (2002). *Acta Cryst.* **B58**, 38–51.
- Wells, A. F. (1977). *Three-Dimensional Nets and Polyhedra*. New York: Interscience.
- Wells, A. F. (1979). *Further Studies of Three-Dimensional Nets*. Am. Crystallogr. Assoc. Monograph No. 8.
- Winkler, B., Pickard, C. J., Milman, V. & Thimm, G. (2001). *Chem. Phys. Lett.* **337**, 36–42.
- Yaghi, O. M., O’Keeffe, M., Ockwig, N. W., Chae, H. K., Eddaoudi, M. & Kim, J. (2003). *Nature (London)*, **423**, 705–714.