

Daria A. Banaru\*, Sergey M. Aksenov, Alexander M. Banaru and Artem R. Oganov

# Mutual correlations of complexity indices of the crystal structure for the series of mercury-containing minerals

<https://doi.org/10.1515/zkri-2024-0062>

Received January 17, 2024; accepted March 2, 2024;

published online March 29, 2024

**Abstract:** The correlations of values of the crystal structure complexity indices were calculated using the major modern approaches (10 indices in total), in particular, those, proposed by Krivovichev, Oganov, the concept of implicit hierarchical depth, as well as, for the comparison, the measure of regularity of  $(r, R)$ -system introduced by Blatov. To find the correlations, a series of 33 mercury-containing mineral structures with a non-repeating system of crystallographic positions occupied by atoms without any partial occupancy was used. It was shown that almost all pairs of complexity indices are statistically significantly ( $p < 0.05$ ) positively or negatively correlated. The discriminating power of the indices for the set of structures was calculated. It is discussed that all the indices can be used to assess the complexity for any series of the crystal structures, but each of them has its particular advantages and limitations.

**Keywords:** mercury minerals; combinatorial complexity; hierarchical depth; degree of order; quasi-entropy

## 1 Introduction

The interest in studying the crystal structure complexity is primarily related to Pauling's fifth rule which is also called the rule of parsimony and which can be formulated as follows: "the number of essentially different kinds of constituents in a crystal tends to be small" [1]. The complexity of crystal structures can be easily characterized using a Shannon functional [2] which has been retaining popularity in mathematical chemistry for many decades [3–6]. In this approach, the crystal structure represents a "message" in which the symbols are atoms, and the positions that the symbols occupy correspond to the crystallographic orbits (occupied by the atoms). The calculation is carried out within one reduced (primitive) cell of the crystal structure because the translations do not actually increase the amount of information, since they repeat the same structural fragment in all directions. Such approach for the calculation of the crystal structure complexity can be called *combinatorial*, or *Krivovichev complexity*. The concept of combinatorial complexity was further significantly extended by Hornfeck [7, 8] and applied to crystal structures with any positional disorder by Kaußler and Kieslich [9, 10].

Another approach to estimate the complexity of crystal structures was proposed and developed by Oganov and Valle [11–13]. As a measure of the simplicity of the crystal structure, this approach uses two special functionals, viz. the degree of order and quasi-entropy, both calculated from the integral characteristics of the positions of atoms. An important advantage of these functionals is that, unlike combinatorial complexity, a small change in the positions of atoms leads to a small, rather than abrupt, change in the functional. The structural degree of order carries information about how narrow the peaks of the pair correlation function over the distances are and how symmetric their local environments are. There is a relationship between the degree of order and the structure amplitude of the crystal [11]. Another measure of complexity, quasi-entropy, indicates the presence of different building blocks in the structure. The quasi-entropy of the structure grows with increasing diversity of the coordination environment of atoms of the same element. According to the results of the study [11], it was concluded that "the ground state normally

---

In memory of Professor Stanislav V. Borisov (1930–2023).

---

**\*Corresponding author: Daria A. Banaru**, Vernadsky Institute of Geochemistry and Analytical Chemistry RAS, Moscow 119991, Russia, E-mail: banaru@geokhi.ru

**Sergey M. Aksenov**, Kola Science Centre of RAS, Laboratory of Arctic Mineralogy and Material Sciences, 14 Fersman str., Apatity 184209, Russia; and Kola Science Centre of RAS, Geological Institute, 14 Fersman str., Apatity 184209, Russia

**Alexander M. Banaru**, Kola Science Centre of RAS, Laboratory of Arctic Mineralogy and Material Sciences, 14 Fersman str., Apatity 184209, Russia; and Department of Chemistry, Lomonosov Moscow State University, 1/3 Leninskie Hills, Moscow 119991, Russia

**Artem R. Oganov**, Vernadsky Institute of Geochemistry and Analytical Chemistry RAS, Moscow 119991, Russia; and Skolkovo Institute of Science and Technology, 20 Bolshoy Blvd., bld. 1, Moscow 121205, Russia

adopts one of the simplest structures compatible with the chemistry of the compound. Such structures tend to have lower [free] energies” and that “in the ground state and low-energy structures, atoms of each species tend to occupy similar crystallographic sites”. Thus, for a thermodynamically controlled crystallization, simpler structures have an advantage over more complex ones. Let us call the degree of order and quasi-entropy the *Oganov complexity* of crystal structures.

An alternative approach to estimate the complexity of a crystal structure is to calculate the implicit hierarchical depth (IHD) with respect to the initial structural units, which should not be confused with explicit hierarchical depth (see [14]). The value of IHD is equal to the smallest number of equivalence classes of edges linking structural units into a simply connected net. The calculation of implicit hierarchical depth is especially useful for molecular crystals with weak van der Waals interactions, for which other types of complexity reflect the complexity of molecules themselves rather than the packing of these molecules. Since the hierarchical organization of molecular crystals at the supramolecular level is often not obvious, their hierarchical depth is inherently implicit. The parsimony of the structures of molecular crystals is manifested in the fact that the number of intermolecular contacts ranked in descending order of interaction energy and sufficient to form the observed crystal structure, is either equal to IHD, or (less often) slightly exceeds it [15].

Any crystal structure can be considered as a multi-regular (multi-orbit) ( $r$ ,  $R$ )-system (Delone point system), in which the  $r$ -neighborhood of each point of the system does not contain other points of the system, and the  $R$ -neighborhood of each point of space contains at least one more point of the system [16]. The uniformity of this system can be characterized by the combined parameter  $G_3$ , which is the second moment of inertia of the Voronoi–Dirichlet polyhedron (VDP) of a point averaged over all points of the system [17] and reflects both the geometric parameters of a Delone system,  $r$  and  $R$ . As the uniformity of the point system decreases, the value of  $G_3$  increases. Among the periodic lattices, the smallest value  $G_3 = 0.07854$  corresponds to the base-centered cubic (b.c.c.) lattice. According to the results of the calculation of  $G_3$  for the crystal structures of elements, Blatov et al. [17] concluded that in a thermodynamically stable structure, atoms and groups of atoms, between which just non-directional interaction forces act, tend to arrange themselves in space so that their centers of mass reach the lowest value of  $G_3$ .

The purpose of this work is to compare the results obtained using three approaches to complexity assessment (Krivovichev, Oganov and IHD) for the series of structures of

mercury-containing minerals. The parameter  $G_3$ , despite the fact that simplicity and uniformity are not exactly the same thing, was also added to the list of comparable complexity indices. The series of structures used is valuable for the variety of both chemical composition (from 2 to 5 or more elements included in the composition) and types of chemical bonds, as well as structure dimensionalities (heterodesmic structures with any dimension of the structural motif and homodesmic structures are present in the series; homo-/heterodesmic is usually understood as those with possible/impossible bypassing all the atoms over the shortest interatomic bonds). Some of the structural data used in the work were obtained by the group of Prof. Stanislav V. Borisov<sup>1</sup> [18, 19], in particular, synthetic analogues of aktashite [20], grechishchevite [21], kelyanite [22], poyarkovite [23] and radtkite [24].

## 2 Methods

### 2.1 Structural data

A total of 218 records containing mercury atoms were available in AMCSDB [25] at the time of structural data selection. From these data, the selection was made in accordance with the following criteria: (1) the mineral has its nomenclature name; (2) the atomic coordinates for all the elements of the compound are known; (3) the occupation of all positions occupied by atoms is full. In the presence of several structural entries, single crystal X-ray diffraction data were preferred to neutronography and powder diffraction data. A total of 37 entries were selected. Some of the entries occurred to correspond to structures with the same system of points (isopointal structures), which have the same type of space group and in which the same Wyckoff positions are occupied [26]. In particular, these are the mercury halides calomel, kuzminite and moshelite belonging to the same structural type, as well as the mercury chalcogenides metacinnabar, timannite and coloradoite belonging to the structural type of sphalerite. Of these sets, only one structure (calomel and metacinnabar) was left for further analysis. Also, the same systems of points are occupied in the crystal structures of chursinite and tvalchrelidzeite (Table 1), despite the different setting of the same space-group type ( $P2_1/c$  and  $P2_1/n$ , respectively). However, compositions of chursinite and tvalchrelidzeite have a different number of elements (three and four, respectively), so both of these structures were left for analysis. In total, there were  $n = 33$  structural records left.

<sup>1</sup> Prof. Stanislav V. Borisov (07.07.1930–18.02.2023) was a principal researcher at Nikolaev Institute of Inorganic Chemistry, Siberian Branch of Russian Academy of Sciences, and was one of the pioneers of applying the systematical analysis of the Patterson function into the X-ray analysis to solve the phase problem [41, 42]. He made a significant contribution to the crystal chemistry of natural and synthetic compounds with heavy elements and proposed and developed the original approach for the analysis of the crystal structures based on the stable cationic nets [43, 44].

**Table 1:** Mercury-containing minerals with full occupation of crystallographic positions.

Space group	Wyckoff letter sequence	Name	Formula	Refcode in AMCS
$P\bar{1}$	$i^2ea$	Deanesmithite	Hg <sub>5</sub> CrS <sub>2</sub> O <sub>5</sub>	0005526
$Aa$	$d^6$	Laffittite	HgAgAsS <sub>3</sub>	0000889
$C2/m$	$f^2\bar{3}b$	Edgarbaileyite	Hg <sub>6</sub> Si <sub>2</sub> O <sub>7</sub>	0001331
	$f^6ca$	Grumplucite	HgBi <sub>2</sub> S <sub>4</sub>	0009710
$P2_1/c^*$	$f^3\bar{6}hf$	Radtkeite	Hg <sub>3</sub> S <sub>2</sub> ICl	0005926
	$e^2a$	Imiterite	HgAg <sub>2</sub> S <sub>2</sub>	0012085
	$e^6$	Christite	HgTlAsS <sub>3</sub>	0010787
	$e^8$	Chursinite	Hg <sub>3</sub> AsO <sub>4</sub>	0009483
		Tvalchrelidzeite	Hg <sub>3</sub> SbAsS <sub>3</sub>	0006159
		Edoylerite	Hg <sub>3</sub> CrS <sub>2</sub> O <sub>4</sub>	0005595
$C2/c^*$	$e^{11}$	Simonite	HgTlAs <sub>3</sub> S <sub>6</sub>	0010851
	$f^4$	Aurivilliusite	Hg <sub>2</sub> OI	0019050
	$f^3ec$	Terlinguaite	Hg <sub>2</sub> OCl	0011911
	$f^5ea$	Wattersite	Hg <sub>5</sub> CrO <sub>6</sub>	0005429
	$f^2eb$	Livingstonite	HgSb <sub>4</sub> S <sub>8</sub>	0010580
	$f^{15}$	Poyarkovite	Hg <sub>3</sub> OCl	0005596
$Pbm2$	$d^2c^2$	Magnolite	Hg <sub>2</sub> TeO <sub>3</sub>	0005221
$Aba2$	$b^{18}a$	Vrbaite	Hg <sub>3</sub> Tl <sub>4</sub> Sb <sub>2</sub> As <sub>8</sub> S <sub>20</sub>	0010725
$Pbma$	$e^3d^4ca$	Hanawaltite	Hg <sub>7</sub> O <sub>3</sub> Cl <sub>2</sub>	0005606
$Pnma^*$	$c^2$	Montroydite	HgO	0011897
	$d^4c^6$	Grechishchevite	Hg <sub>6</sub> S <sub>4</sub> IBr <sub>2</sub> Cl	0005919
$Ibam$	$kj^2fa$	Pinchite	Hg <sub>5</sub> O <sub>4</sub> Cl <sub>2</sub>	0001692
$P4j$	$da$	Potarite	HgPd	0005246
$mmm$				
$P4_2/nmc$	$da$	Coccinite	HgI <sub>2</sub>	0019553
$I4/mmm$	$e^2$	Calomel	HgCl	0010993
		Kuzminite	HgBr	0011657
		Moschelite	HgI	0018645
		Kelyanite	Hg <sub>12</sub> SbO <sub>6</sub> BrCl <sub>2</sub>	0004681
$P3$	$d^{20}c^2b^2a^2$			
$R3$	$b^8a$	Aktashite	Hg <sub>3</sub> Cu <sub>6</sub> As <sub>4</sub> S <sub>12</sub>	0018863
$P3_221$	$ba$	Cinnabar	HgS	0000007
$P\bar{3}m$	$ieda$	Jacutingaite	HgPt <sub>2</sub> Se <sub>3</sub>	0019200
$P2_13$	$b^2a^3$	Kuznetsovite	Hg <sub>3</sub> AsO <sub>4</sub> Cl	0015847
$I2_13$	$ba^2$	Corderoite	Hg <sub>3</sub> S <sub>2</sub> Cl <sub>2</sub>	0009337
$F\bar{4}3m$	$ca$	Metacinnabar	HgS	0018202
		Tiemannite	HgSe	0000050
		Coloradoite	HgTe	0017666
$Ia\bar{3}d$	$hge$	Eglestonite	Hg <sub>6</sub> O <sub>2</sub> Cl <sub>3</sub>	0015660

\*in different settings

## 2.2 Krivovichev complexity

The calculation of combinatorial complexity was carried out according to the ordinary Shannon formula [2]:

$$I_G = - \sum_{i=1}^k p_i \log_2 p_i \text{ (bit/atom)}, \quad (1)$$

where  $p_i = v_i/v$  is the probability of finding some atom in the  $i$ th orbit,  $v_i$  is the multiplicity of this orbit,  $v$  is the sum of the multiplicities of all occupied orbits. The total complexity of the entire reduced cell was calculated as

$$I_{G,\text{tot}} = vI_G. \quad (2)$$

The normalized complexity was calculated with respect to the maximum possible value of  $I_G$  for a given  $v$ :

$$I_{G,\text{norm}} = I_G/I_{G,\text{max}} = I_G/\log_2 v. \quad (3)$$

For instance, for one of the simplest structures, calomel (Hg<sub>2</sub>Cl<sub>2</sub>, or HgCl, see Table 1), the reduced cell is twice smaller than the conventional one, as the latter is body-centered ( $I4/mmm$ ). Thus, for the two occupied orbits with the same Wyckoff letter  $e$  and equal  $v_i = 2$ , one obtains  $I_G = -2 \cdot (2/4) \cdot \log_2(2/4) = 1.0$  bit/atom,  $I_{G,\text{tot}} = 4 \cdot 1.0 = 4.0$  bit/cell, and  $I_{G,\text{norm}} = 1.0/\log_2 4 = 0.5$ .

## 2.3 Oganov complexity

The method proposed by Oganov and Valle utilizes a matrix  $F$  called “fingerprints” of a crystal structure, with the matrix elements being the values of  $F_{AB}$  distance function:

$$F_{AB}(R) = \sum_{A_i B_j} \frac{\delta(R - R_{ij})}{4\pi R_{ij}^2 \left( \frac{N_A N_B}{V_{\text{u.c.}}} \right) \Delta} - 1, \quad (4)$$

where  $R_{ij}$  is the distance between atoms,  $\delta(R - R_{ij})$  is the Dirac  $\delta$ -function,  $i$  runs through all  $N_A$  atoms of element  $A$  within the unit cell of volume  $V_{\text{u.c.}}$ ,  $j$  runs through  $N_B$  atoms of element  $B$  within  $R_{\text{max}}$ , the value of which should be chosen to be large enough to yield converging results. The  $R$  parameter in formula (4) should not be confused with that of ( $r, R$ )-system (Delone point system). Before summation, each peak of the  $\delta$ -function is approximated by a normal distribution with a user-adjustable standard deviation value  $\sigma$  (usually  $\sigma = 0.02$  Å), and then transformed into a histogram with an also adjustable bar width  $\Delta$  (usually  $\Delta = 0.05$  Å) [12] (Figure 1). To assess the similarity of two “fingerprints”, the value of the  $F_{AB}$  function is represented as a vector  $F_{AB}$  in  $(R_{\text{max}}/\Delta)$ -dimensional space. The degree of order is calculated using the function

$$P_{AB} = \frac{N_B}{V_{\text{u.c.}}} \int_0^{R_{\text{max}}} F_{AB}^2(R) R^2 dR. \quad (5)$$

For the whole structure

$$P = \sum_{A,B} w_{AB} P_{AB}, \quad (6)$$

$$w_{AB} = \frac{N_A N_B}{\sum_{\text{u.c.}} N_A N_B}. \quad (7)$$

The quasi-entropy is calculated using the formula [11]:

$$S_{\text{str}} = - \sum_A \frac{N_A}{N} (1 - D_{A,A_j}) \ln(1 - D_{A,A_j}), \quad (8)$$

where  $N$  is the total number of atoms in an elementary cell,  $D_{A,A_j}$  is the cosine distance between the vectors  $F_{A,A_j}$ , calculated separately for the  $i$ th and  $j$ th positions occupied by the atoms of element  $A$ .

Later Lyakhov et al. [13] introduced the value of the local degree of order for the atom  $A_i$ :

$$\Pi_i = \sqrt{\sum_B \frac{N_B}{N} \frac{\Delta}{(V_{\text{u.c.}}/N)^{1/3}} |F_{A_i,B}|^2}, \quad (9)$$

which is then used to calculate the average value of  $\langle \Pi \rangle$  for all atoms in the crystal structure.

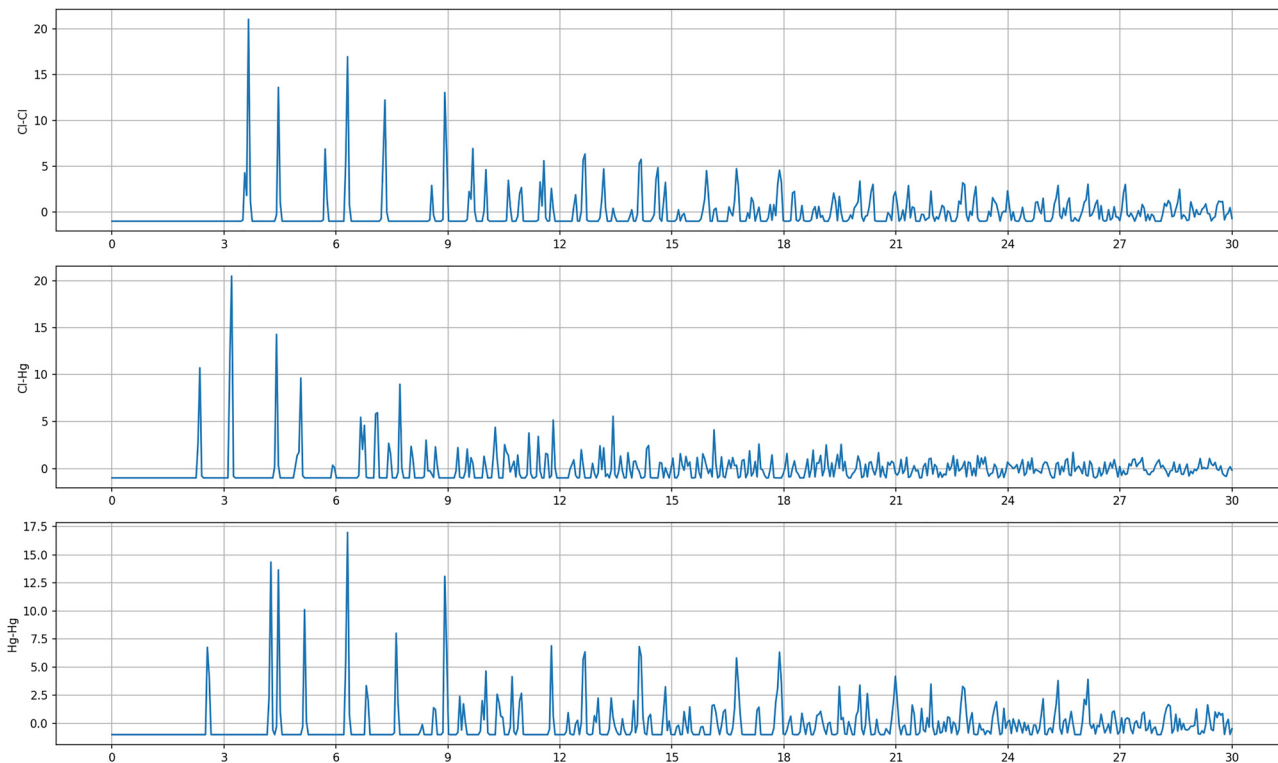


Figure 1:  $F_{AB}(R)$  for the crystal structure of calomel at  $\sigma = 0.02 \text{ \AA}$ ,  $\Delta = 0.05 \text{ \AA}$ ,  $R_{\max} = 30 \text{ \AA}$ .

For instance, for calomel, in which each of the two elements (Hg and S) occupies exactly one orbit,  $S_{\text{str}} = 0$ . Fortunately, the degrees of order for this structure are non-zero:  $P = \langle \Pi \rangle = 5.77$ .

## 2.4 Hierarchical depth

The initial adjacency matrix of atoms not needed for calculating Krivovichev and Oganov complexities (Figure 2a) and different from that conventional for chemists (Figure 2b) was constructed by the method of solid angles for all  $\Omega_{\text{at}} > 0$ . The net obtained in this way (Figure 2c) is the Delone graph [27]. The alternative domain method widely used in constructing models of crystal structures [17], also allows one to build a Delone graph, but takes into account the chemical characteristics of atoms of different elements and attributes the different types of chemical bonds to the edges of the graph, which does not matter for the current problem. While considering the midpoints of the edges of the Delone graph as “atoms” (Figure 2d), its edge combinatorial complexity  $H_{\text{edge}}$  was calculated using the formula similar to (1) and those recently used for molecular crystals [28, 29].

Next, a critical net [30] was found in the Delone graph, containing the minimal necessary number of edges for the net to be simply connected, and all edges were subsequently removed from it. Then the edges were gradually added back, building all possible graphs first with one class of edges equivalent with respect to the space group, then with two equivalence classes, three equivalence classes etc., until a simply connected graph was obtained (Figure 3a). Usually, aside from such a subgraph of the critical net, there are many other ways to connect structural units into a simply connected graph (Figure 3b–f), however, all of these graphs must have the least number (IHD) of equivalence

classes of edges. Unlike an ordinary crystallographic net with all vertices being not less than 3-coordinated, a simply connected subgraph of the Delone graph may have 2-coordinated (Figure 3d) or even 1-coordinated (Figure 3c) vertices. According to [31], the value of IHD depends on the space group and the list of Wyckoff positions (WP) occupied by the structural units, viz.:

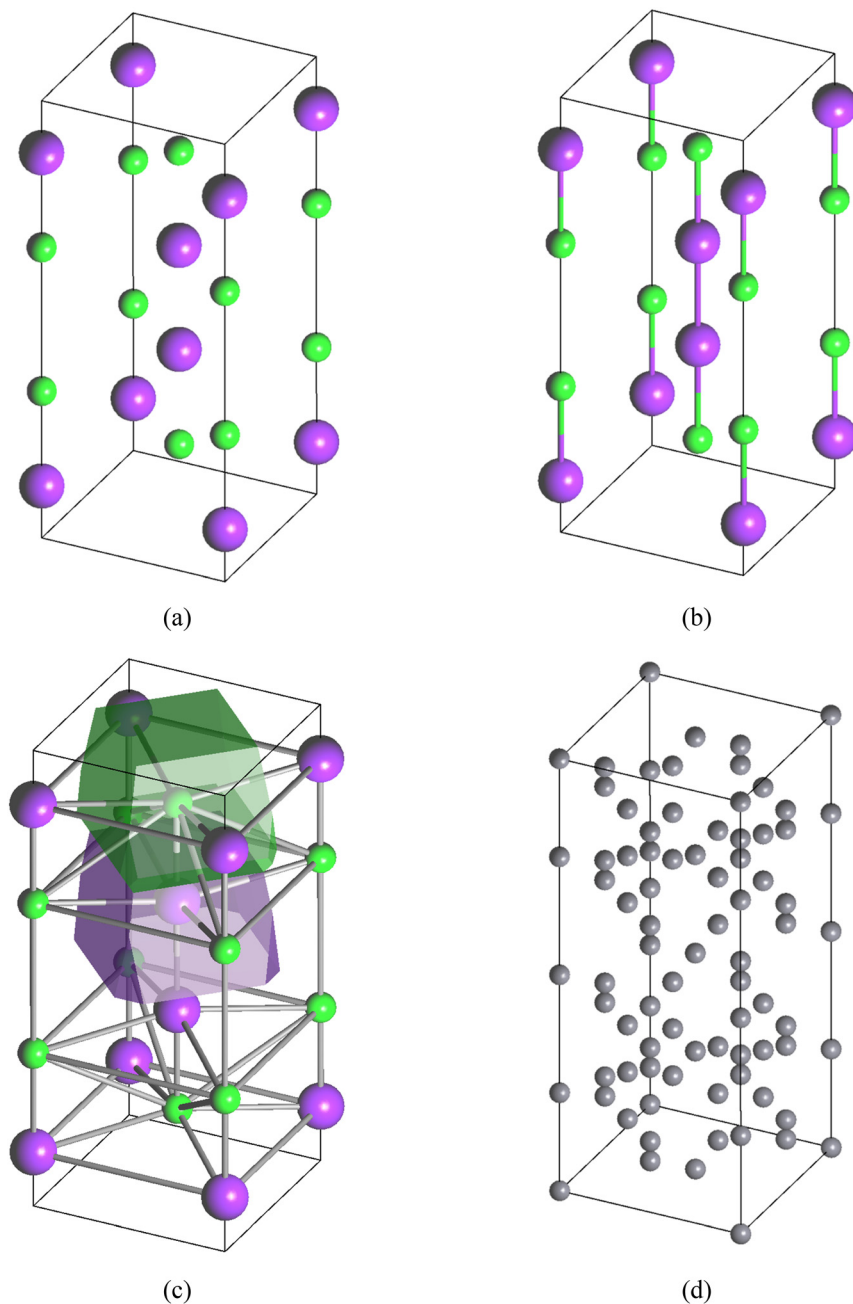
$$\text{IHD} \equiv \inf(e^{\vee}) = |U_{\text{SG}}| + Z' - 1 - f(\text{WP}_1, \text{WP}_2, \dots, \text{WP}_Z), \quad (10)$$

where  $e^{\vee}$  is the number of equivalence classes of edges in the Delone graph,  $|U_{\text{SG}}|$  is the number of elements in any minimal generating subset of the space group,  $Z'$  is the total number of orbits occupied by the structural units, and the parameter  $f(\text{WP}_1, \text{WP}_2, \dots, \text{WP}_Z)$  is influenced by the site-symmetry groups of occupied positions. In the simplest case, when structural units occupy only general positions,  $f = 0$ .

For calomel (Figures 2 and 3), in the Delone graph the vertices corresponding to Hg and S atoms are 10- and 14-coordinated, respectively. However, all the edges fall in the  $e^{\vee} = 7$  equivalence classes with respect to the group  $I4/mmm$ . The midpoints of the edges occupy Wyckoff positions  $mg^2feba$  ( $v_i = 16, 8 \times 2, 8, 4, 2$ , and  $2$ , respectively), in accordance with (1) resulting in  $H_{\text{edge}} = 2.502 \text{ bit/edge}$ . Out of the seven equivalence classes, six combinatorially different triples are sufficient to make the graph simply connected (IHD = 3).

## 2.5 Discriminating power

According to [32], a discriminating power of typology, depending on the probability of two unrelated objects to belong to the same type, can be represented as a parameter:



**Figure 2:** Unit cell of calomel, view along [421], represented as: not connected atoms (a), covalently connected atoms (b), the Delone graph (c), edge midpoints of the Delone graph (d).

$$D = 1 - \frac{1}{n(n-1)} \sum_{j=1}^s x_j(x_j-1), \quad (11)$$

## 2.6 Software

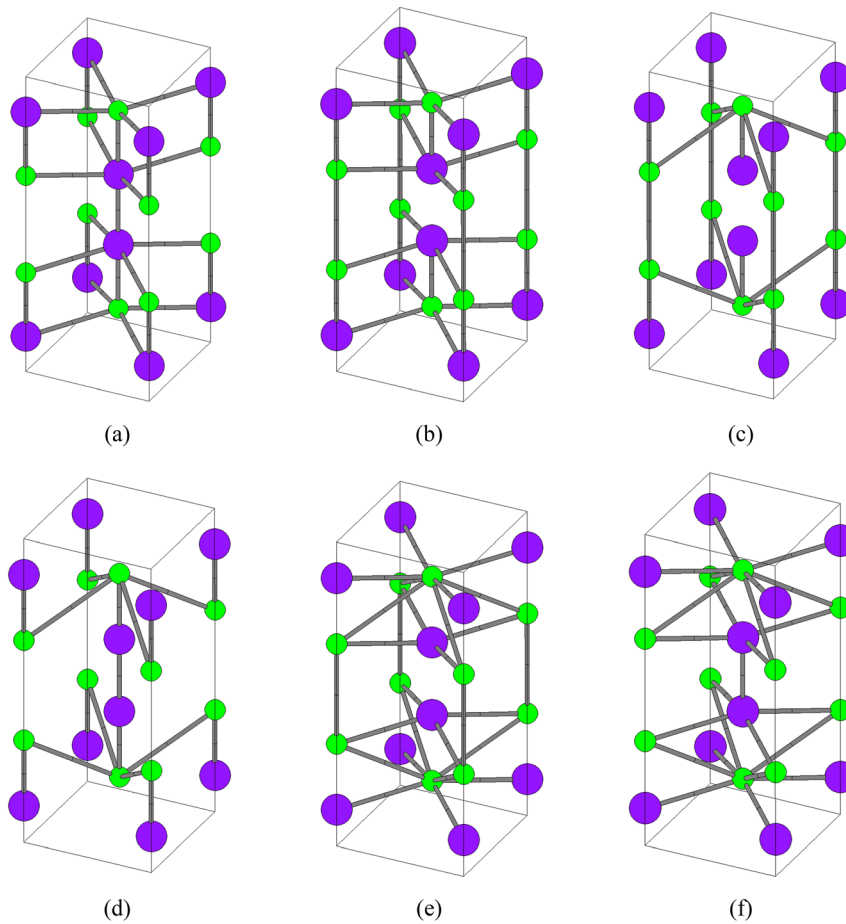
Calculations of all combinatorial complexities, IHD and  $G_3$  were performed in ToposPro ver. 5.5.2.0 software package [33]. Calculations of Oganov complexities were performed using Fingerprints utility [34] with values  $\sigma = 0.02 \text{ \AA}$ ,  $\Delta = 0.05 \text{ \AA}$ ,  $R_{\max} = 30 \text{ \AA}$ .

## 3 Results and discussion

Table 2 shows the paired linear correlation coefficients for the indices used. The level of statistical significance  $p < 0.05$

where  $n$  is the number of unrelated objects (in our case, structures),  $s$  is the number of types in the model,  $x_j$  is the number of objects assigned to the  $j$ th type in the model, and the types should not overlap. Thus,  $D = 1$  shows that the typology is able to distinguish all the objects in the sample. When  $D = 0$ , all objects are classified as of the same type.  $D = 0.5$  means that a randomly selected sample object has 50 % probability to belong to the same type than the next randomly selected object. In [32], the formula (11) was applied to microbiological taxons. In this work two structures are considered to belong to the same type if the values of a certain complexity index for them are equal.





**Figure 3:** Unit cell of calomel, view along [421], represented as the critical net (a) and other subgraphs of the Delone graph (b–f) with the least number (IHD) of equivalence classes of edges.

at  $n = 33$  corresponds to the correlation coefficient of 0.344 and more. Only  $G_3$  does not have a statistically significant correlation with other indices confirming that complexity and uniformity are different properties of a system of points. As it was shown by the values of  $H_{\text{edge}}$  and  $G_3$  calculated via the centers of mass of organic molecules in the structures of organic minerals [35], an increase in the diversity of intermolecular contacts does not necessarily entail an increase in the non-uniformity of the Delone system of centers of mass.

This conclusion is now confirmed for non-molecular structures, as well.

The next most weakly correlated value with the rest of the indices is the normalized Krivovichev complexity  $I_{G,\text{norm}}$  statistically significantly ( $p < 0.05$ ) correlated only with  $I_G$  and quasi-entropy. Seven indices, with the exception of  $G_3$ ,  $I_{G,\text{norm}}$  and  $\langle \Pi \rangle$ , are statistically significantly positively or negatively correlated with each other. The degrees of order  $P$  and  $\langle \Pi \rangle$  demonstrate a strong positive correlation between

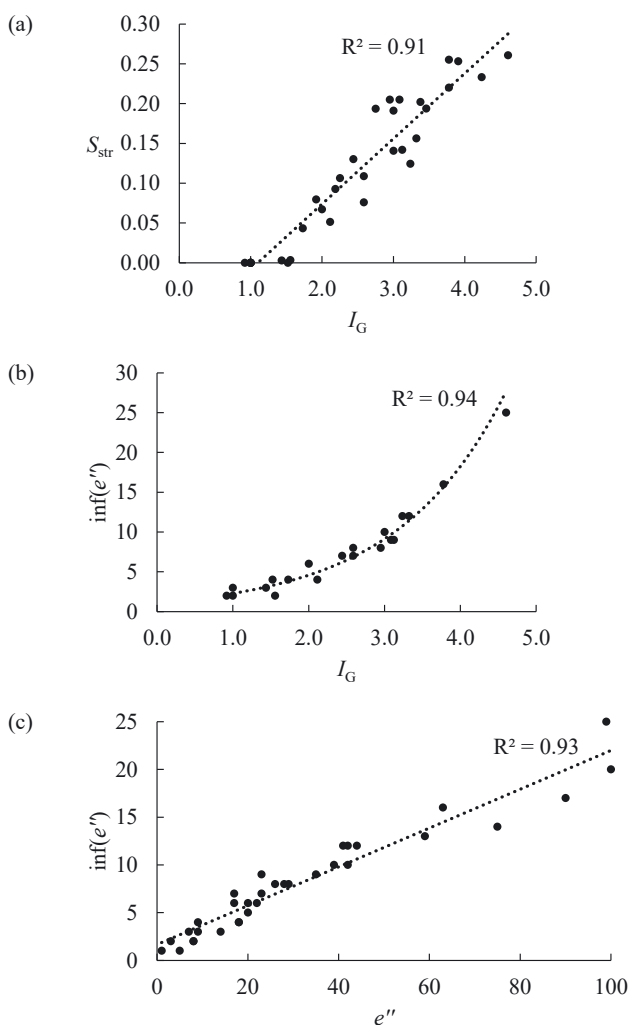
**Table 2:** Pearson correlation coefficients for indices based on the series of mercury-containing mineral structures ( $n = 33$ ).

	$I_G$	$I_{G,\text{tot}}$	$I_{G,\text{norm}}$	$S_{\text{str}}$	$P$	$\langle \Pi \rangle$	$e''$	$\text{inf}(e'')$	$H_{\text{edge}}$	$G_3$
$I_G$	1.000									
$I_{G,\text{tot}}$	0.839	1.000								
$I_{G,\text{norm}}$	0.382	0.183	1.000							
$S_{\text{str}}$	0.953	0.754	0.427	1.000						
$P$	-0.717	-0.544	0.141	-0.728	1.000					
$\langle \Pi \rangle$	-0.488	-0.328	0.350	-0.495	0.953	1.000				
$e''$	0.905	0.929	0.322	0.853	-0.626	-0.399	1.000			
$\text{inf}(e'')$	0.949	0.895	0.361	0.886	-0.632	-0.396	0.965	1.000		
$H_{\text{edge}}$	0.919	0.781	0.102	0.862	-0.874	-0.724	0.857	0.873	1.000	
$G_3$	0.306	0.159	0.069	0.340	-0.242	-0.187	0.173	0.303	0.221	1.000

themselves and a negative correlation with the other five indices from the same seven ones. The  $H_{\text{edge}}$  value demonstrates the best negative correlation with  $P$  and  $\langle \Pi \rangle$ . In general, this index is the only one among the specified seven indices whose all correlation coefficients are greater than 0.75, which can be considered a strong correlation, so  $H_{\text{edge}}$  looks in some sense the most universal index to take into account various aspects of complexity of the crystal structure. The latter index is of a Shannon nature, but unlike the ordinary combinatorial complexity it takes into account not the division of vertices (atoms), but that of edges (bonds) into equivalence classes. As atoms have no knowledge of space-group types, their mutual arrangement is determined by nothing more than a system of chemical bonds between them, and energetic and therefore symmetric diversity of them is ultimately expressed by the value of  $H_{\text{edge}}$ . For molecular crystals in which the translational degrees of freedom of individual atoms are strongly constrained by chemical bonds within the molecule, it is  $H_{\text{edge}}$  that determines the decisive difference in complexity for crystal structures of the same structural class [29].

Logarithmic measures  $I_G$ ,  $S_{\text{str}}$  and  $H_{\text{edge}}$  naturally have a better linear correlation with each other than with other indices. Figure 4a shows a diagram of the correlation between Oganov quasi-entropy and Krivovichev complexity, which is approximated by a linear dependence with a high coefficient of determination  $R^2 = 0.91$ . The IHD depends on Krivovichev complexity exponentially (Figure 4b), but also with a high  $R^2 = 0.94$ . Graphs similar to those, but sometimes with a lower  $R^2$ , can be obtained for other indices of the seven most strongly correlated measures. Surprisingly, the IHD increases clearly linearly with the increase in the number of edge equivalence classes in Delone graph (Figure 4c). For graphs with a different number of classes of equivalent vertices, this correlation is predictable: the more symmetrically unequal vertices there are, the more unequal edges there are in the critical net. For structures with the same number of classes of equivalent vertices, this correlation is less obvious and requires further verification on a larger series of structures.

Calculating IHD without using a critical net is much time-consuming computational task, since, in this case, in order to test the connectivity of a net at each step of the calculation algorithm it is required to construct  $C_e^i$  (a combination, i.e. a selection of  $i$  edges from the set of  $e''$  edges) subnets of Delone graph for  $i = 1, 2, \dots, \text{inf}(e'')$ . If the value of  $e''$  is very large, then such a task becomes practically impossible to solve in a reasonable time even using a supercomputer. For example, for the *vrbaite* structure ( $e'' = 100$ ), at the last step it would be necessary to test the connectivity of  $C_{100}^{20} = 535983370403809682970$  subnets. Cutting off most of



**Figure 4:** Dependencies of quasi-entropy (a) and implicit hierarchical complexity (b) on combinatorial complexity (bit/atom), and the dependence of IHD on the number of equivalence classes of edges in the reduced cell of the Delone graph (c) for the structures of mercury minerals ( $n = 33$ ).

the edges with small solid angles of atomic VDP greatly simplifies the task. Thus, even a rough construction of a critical net for a *vrbaite* with a step of  $\Delta\Omega_{\text{at}} = 1.5\%$  lowers  $e''$  from 100 to 25 and reduces  $C_e^i$  by about 16 orders of magnitude. The idea of a critical net was recently developed into the concept of a skeletal net [36], which is equally conveniently applicable to both molecular and non-molecular crystals, and the hierarchical analysis of structures in newer versions of ToposPro software package [33] has become even more routine. Compared to the critical net, the skeletal net has a weaker restriction (3D-periodicity instead of 1-connectivity), which for some structures leads to a mismatch of these nets, however, the skeletal net contains more information about the nature of the structure. In

**Table 3:** Discriminating power of indices for a series of mercury-containing mineral structures ( $n = 33$ ).

Index	$I_G$	$I_{G,tot}$	$I_{G,norm}$	$S_{str}$	$P$
$s$	26	31	28	27	33
$D$	0.9754	0.9962	0.9886	0.9602	1.0000
Index	$\langle \Pi \rangle$	$e''$	$\inf(e'')$	$H_{edge}$	$G_3$
$s$	33	26	17	32	33
$D$	1.0000	0.9867	0.9583	0.9981	1.0000

particular, the skeletal net is able to distinguish a framework, inside the cavities of which there is a weakly bound particle, as well as to distinguish some interpenetrating nets [36]. For the instance of calomel mentioned above, the critical net and the skeletal net are identical (Figure 3a), in RCSR [37] its topologic type having refcode **fsx**.

The degrees of order and  $G_3$  have the greatest discriminating power among the considered indices (see formula (11)) equal to 1, or 100 %, and IHD has the least one (~96 %) (Table 3). Quasi-entropy has approximately the same low discriminating power, mainly due to the fact that seven structures have  $S_{str} = 0$  (calomel, cinnabar, coccinite, imiterite, metacinnabar, montroydite and potarite). In these structures, the atoms of each element occupy only one regular point system, which also means that the combinatorial complexity of these structures coincides with the chemical complexity (the Shannon complexity of the chemical formula [38]).

## 4 Conclusions

The aim of this paper was not to cover all the possible indices, which can be used for characterizing the complexity of a crystal structure, but we looked only at indices that can be obtained during automated (or semi-automated, in case of hierarchical complexity) procedures performed over a crystal structure using powerful and freely distributed software. For example, the Baur distortion index of a coordination polyhedron [39], the calculation of which is integrated into the popular program (also with a free license) VESTA [40], in the current version of the program cannot be automatically calculated as the average value for all coordination polyhedra in the structure, which makes the routine use of this index as a measure of complexity almost impossible. As the summary, the Table 4 shows the advantages and limitations of utilizing the considered indices.

**Table 4:** Advantages and limitations of complexity indices of various types.

Indices	«+»	«-»
Combinatorial complexities (including edge complexity)	Simple calculations, strong correlations with other indexes	High sensitivity to small changes in structure
Degrees of order	High discriminating power, low sensitivity to small changes in structure	Weak correlations with other indices
Quasi-entropy	High discriminating power, low sensitivity to small changes in structure	Indistinguishability of structures with the minimal allowable combinatorial complexity
Hierarchical depth	Accounting for the actual system of contacts in the structure	Low discriminating power

**Research ethics:** Not applicable.

**Author contributions:** The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** The authors state no conflict of interest.

**Research funding:** The work with databases and Oganov complexity calculations were carried out by D.A.B. and A.R.O. in accordance with the state task of the GEOKHI RAS. Krivovichev and hierarchical complexity calculations were performed by S.M.A. and A.M.B. in accordance with the Russian Science Foundation (Grant No. 20-77-10065).

**Data availability:** Not applicable.

## References

- Pauling L. The principles determining the structure of complex ionic crystals. *J. Am. Chem. Soc.* 1929, 51, 1010–1026.
- Krivovichev S. V. Topological complexity of crystal structures: quantitative approach. *Acta Crystallogr., Sect. A* 2012, 68, 393–398.
- Sabirov D. S., Shepelevich I. S. Information entropy in chemistry: an overview. *Entropy* 2021, 23, 1240.
- Aksenov S. M., Charkin D. O., Banaru A. M., Banaru D. A., Volkov S. N., Deineko D. V., Kuznetsov A. N., Rastsvetaeva R. K., Chukanov N. V., Shkurskii B. B., Yamnova N. A. Modularity, polytypism, topology, and complexity of crystal structures of inorganic compounds (Review). *J. Struct. Chem.* 2023, 64, 1797–2028.
- Krivovichev S. V., Krivovichev V. G., Hazen R. M., Aksenov S. M., Avdontceva M. S., Banaru A. M., Gorelova L. A., Ismagilova R. M., Korniyakov I. V., Kuporev I. V., Morrison S. M., Panikorovskii T. L., Starova G. L. Structural and chemical complexity of minerals: an update. *Mineral. Mag.* 2022, 86, 183–204.



6. Banaru D., Hornfeck W., Aksenov S., Banaru A. On the origin of combinatorial complexity of the crystal structures with 0D, 1D, or 2D primary motifs. *CrystEngComm* 2023, 25, 2144–2158.
7. Hornfeck W. On an extension of Krivovichev's complexity measures. *Acta Crystallogr., Sect. A* 2020, 76, 534–548.
8. Hornfeck W. Crystallographic complexity partition analysis. *Z. Kristallogr.* 2022, 237, 127–134.
9. Kaußler C., Kieslich G. crystIT: complexity and configurational entropy of crystal structures via information theory. *J. Appl. Crystallogr.* 2021, 54, 306–316.
10. Hallweger S. A., Kaußler C., Kieslich G. The structural complexity of perovskites. *Phys. Chem. Chem. Phys.* 2022, 24, 9196–9202.
11. Oganov A. R., Valle M. How to quantify energy landscapes of solids. *J. Chem. Phys.* 2009, 130, 104504.
12. Valle M., Oganov A. R. Crystal fingerprint space – a novel paradigm for studying crystal-structure sets. *Acta Crystallogr. Sect. A* 2010, 66, 507–517.
13. Lyakhov A. O., Oganov A. R., Valle M. How to predict very large and complex crystal structures. *Comput. Phys. Commun.* 2010, 181, 1623–1632.
14. Krivovichev S. V. Structure description, interpretation and classification in mineralogical crystallography. *Crystallogr. Rev.* 2017, 23, 2–71.
15. Banaru A. M., Aksenov S. M., Banaru D. A. Critical molecular coordination numbers in the structural class  $P2_1/c$ ,  $Z = 4(1)$ . *Moscow Univ. Chem. Bull.* 2021, 78, 325–333.
16. Galiulin R. V. Delone systems as a basis of geometry of the discrete world. *Comp. Math. Math. Phys.* 2003, 43, 754–765.
17. Blatov V. A., Shevchenko A. P., Serenzhkin V. N. Crystal space analysis by means of Voronoi–Dirichlet polyhedra. *Acta Crystallogr., Sect. A* 1995, 51, 909–916.
18. Magarill S. A., Pervukhina N. V., Borisov S. V., Pal'chik N. A. Crystal chemistry and features of the structure formation of mercury oxo- and chalcogenides. *Russ. Chem. Rev.* 2007, 76, 101–131.
19. Borisov S. V., Magarill S. A., Pervukhina N. V., Peresypkina E. V. Crystal chemistry of mercury oxo- and chalcogenides. *Crystallogr. Rev.* 2005, 11, 87–123.
20. Vasil'ev V. I., Pervukhina N. V., Borisov S. V., Magarill S. A., Naumov D. Y., Kurat'eva N. V. Aktashite  $Cu_6Hg_3As_4S_{12}$  from the Aktash deposit, Altai, Russia: refinement and crystal chemical analysis of the structure. *Geol. Ore Depos.* 2010, 52, 656–661.
21. Pervukhina N. V., Vasil'ev V. I., Borisov S. V., Magarill S. A., Naumov D. Y. The crystal structure of a polymorph of  $Hg^{2+}_3S_2Br_{1.0}Cl_{0.5}I_{0.5}$ . *Can. Mineral.* 2003, 41, 1445–1453.
22. Pervukhina N. V., Borisov S. V., Magarill S. A., Naumov D. Y., Vasil'ev V. I. The crystal structure of kelyanite,  $(Hg_2)_6(SbO_6)BrCl_2$ . *Am. Mineral.* 2008, 93, 1666–1669.
23. Vasil'ev V. I., Pervukhina N. V., Romanenko G. V., Magarill S. A., Borisov S. V. New data on the mercury oxide-chloride mineral poyarkovite; the second find, and crystal-structure determination. *Can. Mineral.* 1999, 37, 119–126.
24. Pervukhina N. V., Borisov S. V., Magarill S. A., Naumov D. Y., Vasiliev V. I., Nenashev B. G. Crystal structure of the synthetic analog of radtkeite  $Hg_3S_2Cl_{1.00}I_{1.00}$ . *J. Struct. Chem.* 2004, 45, 720–723.
25. Downs R. T., Hall-Wallace M. The American mineralogist crystal structure database. *Am. Mineral.* 2003, 88, 247–250.
26. Lima-de-Faria J., Hellner E., Liebau F., Makovicky E., Parthé E. Nomenclature of inorganic structure types. Report of the international union of crystallography commission on crystallographic nomenclature subcommittee on the nomenclature of inorganic structure types. *Acta Crystallogr., Sect. A* 1990, 46, 1–11.
27. Danzer L., Dolbilin N. Delone graphs; some species and local rules. In *The mathematics of long-range aperiodic order (Waterloo, ON, 1995)*; Kluwer Acad. Publ.: Dordrecht, Waterloo, ON, Canada, 1997; pp. 85–114.
28. Banaru A. M., Aksenov S. M., Krivovichev S. V. Complexity parameters for molecular solids. *Symmetry* 2021, 13, 1399.
29. Banaru A. M., Aksenov S. M. Complexity of molecular nets: topological approach and descriptive statistics. *Symmetry* 2022, 14, 220.
30. Banaru A. M., Banaru D. A. Zorkii structural classes and critical topology of molecular crystals. *J. Struct. Chem.* 2020, 61, 1485–1502.
31. Lord E. A., Banaru A. M. Number of generating elements in space group of a crystal. *Moscow Univ. Chem. Bull.* 2012, 67, 50–58.
32. Hunter P. R., Gaston M. A. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J. Clin. Microbiol.* 1988, 26, 2465–2466.
33. Blatov V. A., Shevchenko A. P., Proserpio D. M. Applied topological analysis of crystal structures with the program package ToposPro. *Cryst. Growth Des.* 2014, 14, 3576–3586.
34. Davari M., Rakitin M. Fingerprints. 2015. Freely: [https://uspex-team.org/online\\_utilities/fingerprints/](https://uspex-team.org/online_utilities/fingerprints/).
35. Banaru A. M., Banaru D. A., Aksenov S. M. On the subset of intermolecular contacts generating a molecular crystal: topological features of organic minerals. *Crystallogr. Rep.* 2022, 67, 1133–1145.
36. Blatova O. A., Blatov V. A. Hierarchical topological analysis of crystal structures: the skeletal net concept. *Acta Crystallogr. Sect. A* 2024, 80, 65–71.
37. O'Keeffe M., Peskov M. A., Ramsden S. J., Yaghi O. M. The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets. *Acc. Chem. Res.* 2008, 41, 1782–1789.
38. Siidra O. I., Zenko D. S., Krivovichev S. V. Structural complexity of lead silicates: crystal structure of  $Pb_{21}[Si_7O_{22}]_2[Si_4O_{13}]$  and its comparison to hyttsoite. *Am. Mineral.* 2014, 99, 817–823.
39. Baur W. H. The geometry of polyhedral distortions. Predictive relationships for the phosphate group. *Acta Crystallogr. Sect. B* 1974, 30, 1195–1215.
40. Momma K., Izumi F. VESTA: a three-dimensional visualization system for electronic and structural analysis. *J. Appl. Crystallogr.* 2008, 41, 653–658.
41. Ilyukhin V. V., Borisov S. V. Quantitative evaluations of maxima in the three-dimensional patterson function. *J. Struct. Chem.* 1961, 1, 70–75.
42. Borisov S. V., Golovachev V. P., Ilyukhin V. V., Kuz'min E. A., Belov N. V. Systematic analysis of the Patterson function. *J. Struct. Chem.* 1972, 13, 166–182.
43. Borisov S. V., Magarill S. A., Pervukhina N. V. Crystallographic analysis of a series of inorganic compounds. *Russ. Chem. Rev.* 2015, 84, 393–421.
44. Borisov S. V., Pervukhina N. V., Magarill S. A. Structural analysis in modern crystallography. *Crystallogr. Rep.* 2022, 67, 860–870.