

# Where Are Genes in Paulingite? Mathematical Principles of Formation of Inorganic Materials on the Atomic Level<sup>a</sup>

V.Ya. Shevchenko<sup>b</sup> and S.V. Krivovichev<sup>b,c</sup>

## Abstract

By means of the concept of universal optimum and general principles of inorganic gene, structures of paulingite-related zeolites and minerals have been constructed. The structures of zeolite Rho and paulingite are considered as members of zeolite family generated by the work of 4-colored cellular automaton (CA). The ideal symmetry of the members of the family is cubic, space group  $Im\bar{3}m$ ,  $a = 15 + 10n$  [Å], where  $n$  is a number of the CA cycle ( $n=0$  for zeolite Rho and 2 for paulingite). A new hypothetical zeolite of the family with  $n=1$  is predicted and named ISC-1 (Institute of Silicate Chemistry-1); atomic coordinates and theoretical X-ray powder diffraction pattern have been calculated. It appears to be very probable that universal optimum contains all necessary information for its material realization (sharp configurations) and construction of materials with certain type of interaction potential (even with restrictions existing in theory for this potential).

## Keywords

Universal optimum • Inorganic gene • Paulingite • Cellular automaton • Zeolite

Nobody enters here, if he is not a geometrician.  
Platon  
Le chimiste est le seul scientifique, qui cree l'objet de ses etudes.  
Marcelline Berthelot

## Introduction

Over the last 20–30 years, we have been witnessing (and participating in) the development of new approaches to the problem of the structure of matter and considerable expansion of the borders of classic crystallography.

First of all, this expansion was induced by the advances in experimental techniques (in particular, high-resolution microscopy) applied to small natural objects: biological, organic, inorganic, and their combinations. To explain the experimental data obtained, there is a need to understand the very basic principles of construction of matter as a part of the material world around us. Indeed, how many atoms are needed to form a substance of macro-world as we know it? For instance, metallic iron. We know the series of atomic Fe clusters consisting of  $N$  atoms forming a “magic” sequence of numbers. What are the properties of these clusters? Do their structure correspond to the structure of macro-iron? These problems are under active debate and investigation at the moment. Perhaps, at the most advanced degree, development of these ideas has been determined by Mackay who first published on this issue in 1967 [1]. The most important parts of this idea are atomistics (i.e., everything starts from atoms) and hierarchy (i.e., matter is organized on the levels of increasing complexity). At the core of this idea is the concept of information that determines an inorganic structure in

<sup>a</sup>*Structural Chemistry* 2008, 19(4):571–577.

<sup>b</sup>Institute of Silicate Chemistry, Russian Academy of Sciences, Nab. Makarova, 2, 199034 St. Petersburg, Russia  
e-mail: exchem@isc.nw.ru

<sup>c</sup>Department of Crystallography, St. Petersburg State University, University Emb. 7/9, 199034 St. Petersburg, Russia

the same way as genetic code determines biological structures. The ideas of Mackay were in correspondence with the works of W. Pearson who in 1972 suggested describing structures of intermetallics by means of codes that characterize stacking sequences of atomic layers (for the sake of justice, we note that analogous approaches were developed by the Soviet scientist G. S. Zhdanov back in 1946). Mackay further developed the idea of description of structures of inorganic materials on the basis of another important concept of ‘curvature’ in the sense of ‘grammar’ of inorganic and biological structures. In this grammar, letters are atoms, and words are clusters of connected atoms (building blocks, structural complexes, etc.). What determines the structures of these clusters and the structure of matter comprising these clusters? Today we are able to provide some partial answers to this question using mathematics (geometry) and information theory. It happens that mutual arrangement of  $N$  points (atoms) interacting with each other in  $n$ -dimensional space is uniquely determined and, in turn, determines so-called spherical code that can be used to save and to transmit information. Authors of Ref. [2] named these configurations “universal optima”. A special condition is the requirement that interaction potential has to be continuous and differentiable. All known universal optima are listed in Table 1. It is worthy to note that the sharp configurations—independently from the type of interaction potential—are the most closely packed

arrangements. Thus, the principle of dense filling of space suggested by Laves in 1967 for metallic alloys has now been proved mathematically, rigorously.

Experimental proof of the results of [2] had been performed in [3] using advanced computer modeling. Two more universal optima have been found (for 40 points in 10-dimensional space and for 64 points in 14-dimensional space). First five configurations in Table 1 are vertices of certain regular polytopes (these are polytypes with all faces being simplices). Next seven cases were obtained from the root lattices  $E_8$  in  $R^8$  and Leech lattice in  $R^{24}$ . 240-point configurations and 196550 points have minimal vectors in these lattices. In terms of sphere packings, these are contact configurations with kissing points in corresponding packings. Each configuration has a spherical cap in each point with its radius being as maximal as possible but without overlaps. Kissing points of this cap provide spherical code in the space of corresponding dimensionality.

Mathematical problem of contact numbers of spheres is a partial case of the problem of a spherical code. For each sphere, there is a corresponding spherical cap on the central sphere and a kissing point. These points are called an alphabet. The problem is to transfer kissing points from certain caps to the other sphere. The set of caps from which signals are transmitted form a word. If transmission is somehow distorted, the caps should be relatively large. In that case, the point should first appear in the cap where it has previously been, so it can be reconstructed since the caps do not intersect. The word can be reconstructed accordingly. The more spherical caps there are on the certain sphere, the more different information can be encoded.

In the framework of information theory, instead of continuous wave in 1-dimensional space, one can speak about vector set in  $n$ -dimensional space.

If connection channel has no memory and with the increase of the word length  $n$ , the number of decoding steps do not increase exponentially ( $2^n$ ) but as  $n^2$ . Using hierarchical principles of coding-decoding, one may decrease computational complexity till  $\sim n$ . Coding as well as decoding are done step-by-step toward higher levels of hierarchy. On the lower level, simple symbols (alphabet) form blocks (words). On the second level, each block (word) is considered as a new symbol, and blocks group into higher-level blocks (sentences). By placement of certain points in  $n$ -dimensional space and looking at their arrangements, we determined further directions of self-assembly using purely geometrical operations.

Thus, the information theory and structure can be merged in an unified theory. All sharp configurations are  $m$ -designs, i.e. they are not simply spherical codes, but also the set of roots of some polynom (Hermit polynom) of the degree  $m$ .

The science of structure of matter is based upon self-assembly of atoms first into sharp configurations, and then,

**Table 1** Known universal optima. Energy-minimizing point configurations

$n$	$N$	$t$	Description
2	$N$	$\cos(2\pi/N)$	$N$ -gon
$n$	$N \leq n+1$	$-1/(N-1)$	Simplex
$n$	$2n$	0	Cross-polytope
3	12	$1/\sqrt{5}$	Icosahedron
4	120	$(1+\sqrt{5})/4$	Regular 600-cell
5	16	1/5	Klebsch graph
6	27	1/4	Schlaefli graph
7	56	1/3	Equilateral lines
8	240	1/2	Root lattice $E_8$
21	112	1/9	Isotropic subspaces
21	162	1/7	Sharply regular graph
22	100	1/11	Higman-Simps graph
22	275	1/6	McLaughlin graph
22	891	1/4	Isotropic subspaces
23	552	1/5	Equilateral lines
23	4600	1/3	Kissing configurations
24	196560	1/2	Leech lattice
$q \frac{q^3+1}{q+1}$	$(q+1)(q^3+1)$	$1/q^2$	Isotropic subspaces

$n$  space dimension,  $N$  number of points,  $t$  scalar product

according to the principle of a cellular automaton, into more hierarchically complex structures, into nanoparticles and then into macromolecules and macroobjects.

The structural diversity of the nanoworld is determined by the fact that finite packings may represent fragments of some “parent” structures that do not necessarily exist in 3-dimensional Euclidean space. We shall call these structures fundamental. The reciprocal transition into space with dimension  $n \leq 3$  is by the ‘cut-and-project’ procedure.

The chemical or physical nature of the subunits is of secondary importance (whether these are atoms or protein molecules). Geometry determines and forms a structural model (universal optimum) before the atoms or other subunits and their interaction potentials are chosen. Thus, in this philosophy, geometry is more than a description principle. It is a reflection of the properties of space where nanoobjects exist.

In order to understand mystical appearance of regular polytopes in the problems of applied mathematics, one has to replace direct calculations of differential geometry by more simple and general approach of symplectic and contact geometry. General principle here is as follows: one has to embed geometrical objects from ‘configurational’ space  $V$  into ‘phase’ space  $V \times F$ , where peculiarities either disappear or become simpler (in theory of differential equations and quantum theory, this approach is called ‘microlocal’). This concept transforms simple facts of differential geometry into general theorems of symplectic and contact geometry with more wide field of application. In this way, perspectives arise for very wide exploration of differential-geometric intuition (V.I. Arnold).

In the past few years, a large number of nanoparticles had been found that have structures not describable in terms of classical crystallography. To express their admiration, scientists use such terms as “unprecedented”, “unique”, “mysterious”. To explain their structure, we have used local (microlocal) approach. General principle is the same: atom configurations should be considered as geometrical objects and embedded from 3-dimensional Euclidean space into another space—of higher dimensionality and (probably) different topology. However, in local terms, this space shall still be represented as a direct product  $V \times F$ . In terms of modern geometry, this procedure corresponds to the construction of a fiber space, where  $V$  is the basis of fibering and  $F$  is the structural group of the fiber.

In general case, fibering is a manifold. Any manifold can be isometrically embedded into Euclidean space of corresponding (or even higher) dimensionality, where we have a possibility to choose a fundamental structure.

These principal results allow us, instead of search of concrete structure responsible in  $E^3$  for the minimum of some potential, to consider a whole class of derivative structures by the prior knowledge that they all have been obtained from some universally optimal non-Euclidean structures and thus

close to the local minimum for the whole class of interaction potentials.

The real type of atoms (or any other subunits) involved will influence preferential formation of derivative structures by the atoms of particular type.

Accordingly, we may consider a complex inorganic structure, e.g. paulingite,<sup>1</sup> a crystalline zeolite with several hundreds of atoms in its unit cell. Where are genes for this structure? Which information does it contain? How may we describe it? Can we describe complex inorganic structure (periodic or aperiodic) as a 1-dimensional chain with its “gene” being the basis of construction and mutation and suitable for the formation of a whole range of structures? In this case, the particular structure can be chosen by the free energy minimum but different minima should not coincide.

We shall try to use these “inorgenes” in genetic algorithm of search of structures that fulfill the requirements of the lowest energy level. This will be required to calculate the gene that corresponds to the corresponding structure. As in biology, there could be genes that may not generate a ‘living’ structure.

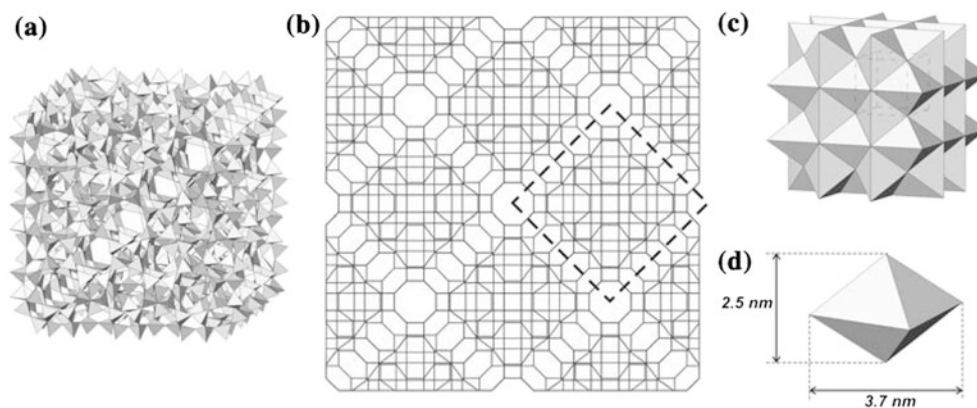
## Paulingite: Structural Architecture

Paulingite ( $K_6Ca_{16}[Al_{38}Si_{130}O_{336}](H_2O)_{113}$ ) is one of the most structurally complex minerals. Its crystal chemistry was the subject of a large number of studies in mineralogical crystallography [4, 5]. At the present time, there are several mineral species of the paulingite group that have the same root name and different suffices according to the dominating extra-framework cation: paulingite-K, paulingite-Na, and paulingite-Ca. A synthetic analogue of paulingite is known under the name ‘zeolite ECR-17’ that had been used in chemical technology [6]. However, the fundamental interest in paulingite is not in chemical diversity of extra-framework cations but in structural complexity of its tetrahedral framework. Unit cell of paulingite contains 768 tetrahedral anions  $TO_4$  ( $T=Al^{3+}$ ,  $Si^{4+}$ ) linked into complex 3-dimensional architecture (Fig. 1a).

Minerals of the paulingite group have a cubic symmetry and crystallize in the space group  $Im\bar{3}m$ , with the  $a$  parameter in the range of 35.00–35.43 Å. Topology of zeolite framework in paulingite is usually described in terms of polyhedral building units [7]. Symbolic (or nodal) description is usually used, when tetrahedra are symbolized by vertices and vertices corresponding to the linked tetrahedra are linked by an edge. Graphs obtained along this procedure are

<sup>1</sup>We have chosen paulingite because the question ‘Where are genes in paulingite?’ had been asked first in the paper of A.L. Mackay “Generalised crystallography” *Journal of Molecular Structure (Techoem)* 336 (1995) 293–303.

**Fig. 1** Tetrahedral framework in paulingite (a) and its 3-dimensional graph (b) striation denotes bipyramidal nanocomplex). Arrangement of the bipyramidal nanocomplexes (c) corresponds to the space partition into tetragonal bipyramids (= flattened octahedra) (d).



4-connected nets that are subject of a separate area of mathematical crystallography [8]. Figure 1b shows graph of the paulingite framework. In contrast to previous works [4–7], we prefer to describe this framework not in terms of polyhedral units but as a result of linkage of bipyramidal (octahedral) complexes outlined in Fig. 1b by striated line. These complexes are centered in the middle points of edges and in the centers of the cubic cell (e.g. in points (0; 0; 0.5) and (0.5; 0.5; 0)). In total, they form space partition into flattened octahedra with dimensions of  $3.7 \times 3.7 \times 2.5 \text{ nm}^3$  (Fig. 1c, d).

Let us consider the structure of complexes in more detail. In Fig. 2, each 4-membered ring (4-MR) is replaced by a flat square. The complex can be split into five levels, each formed by a closed contour of vertically standing squares (Fig. 2b). The highest level consists of a double octagonal ring (D8-MR), the second is formed by 24 squares, the third and central level contains 40-square contour and internal D8-MR. The fourth and fifth levels repeat the first and the second, respectively. The levels are linked by inclined squares so that the whole construction reminds of hierarchical construction of Middle-Age castles that also had many-level architecture.

It is noteworthy that, in general, construction of the complexes corresponds to the principles of sharp configurations. At the first level, atoms form tetrahedra—3-dimensional simplices (see second row in Table 1), then tetrahedra are linked into 4-MRs and 8-MRs (first row in Table 1). General

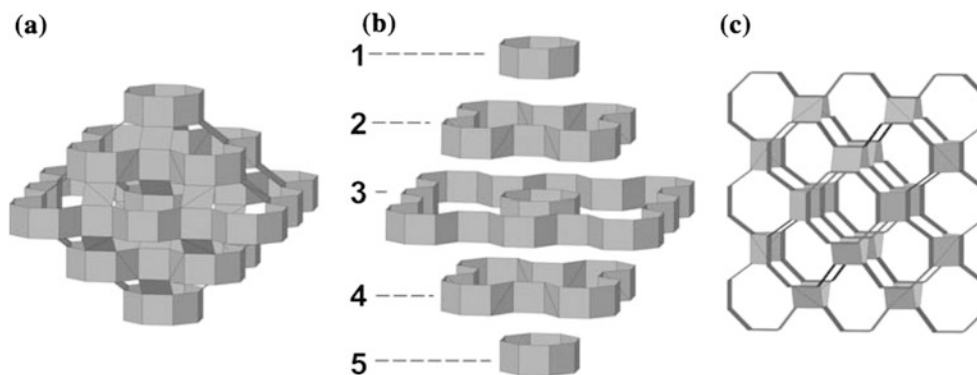
architecture corresponds to an octahedron that is a cross-polytope of 3-dimensional space (third row in Table 1).

### Paulingite: Cellular Automata Modeling

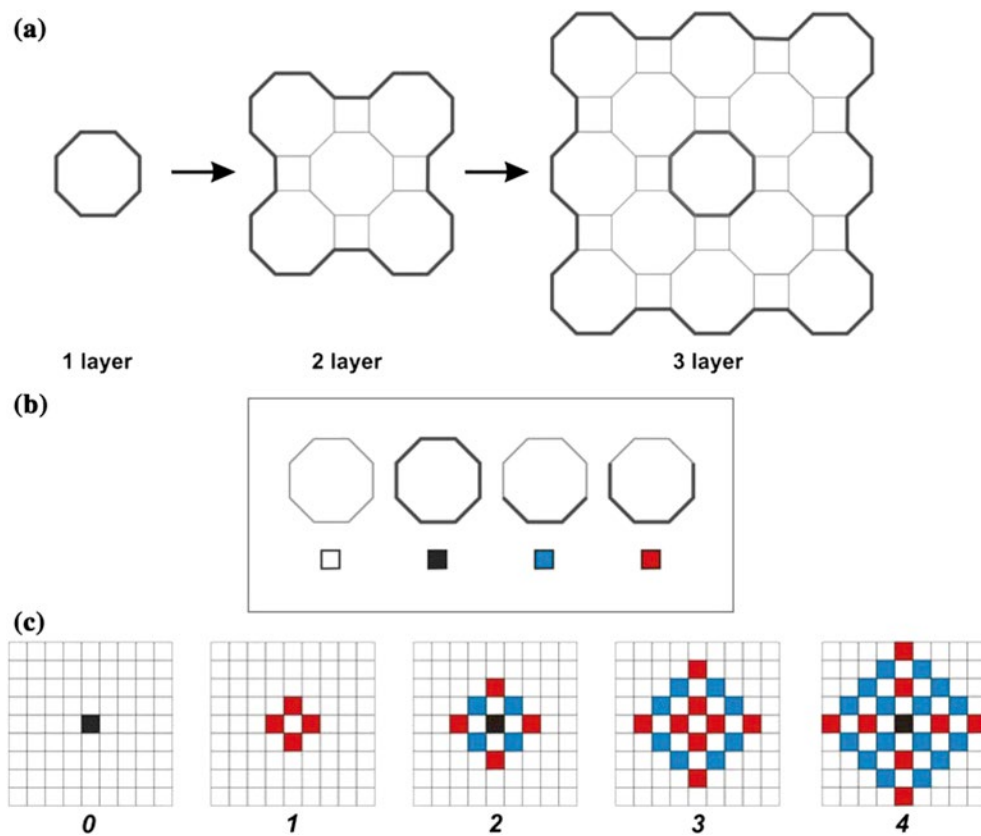
Projection of the nanocomplex along its axis is shown in Fig. 3. Obviously, the 2nd and 3rd levels are similar to the 1st level and can be obtained from the latter by some simple geometrical operations. As an instrument, we may use a cellular automaton (CA), a discrete deterministic system that develops in time and space according to simple set of transition rules [9]. The use of CAs in structural crystallography was first proposed by Mackay [10] and recently was further developed in [11].

Figure 3a shows schematical construction of paulingite nanocomplex in terms of vertical squares. The CA lattice is a dihedral partition of plane into squares and octagons. Initial condition (zero-cycle of the CA) is a regular octagon. In order to simplify the model further without the loss of generality, we may transform the initial CA lattice into a square lattice by a series of simple substitutions (Fig. 3b). The closed 8-MR will correspond to a black cell, 8-MR with one or less bold edges to a white cell, 8-MRs with three and five bold edges to blue and red cells, respectively. As a result, we shall obtain 4-colored CA shown in Fig. 3c.

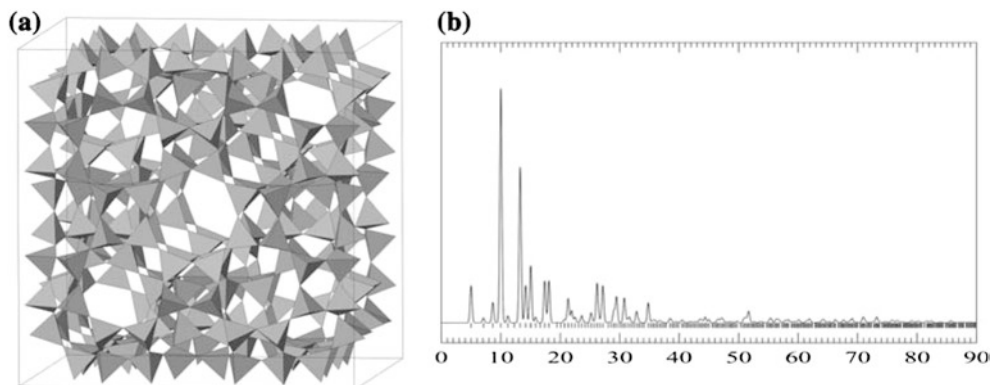
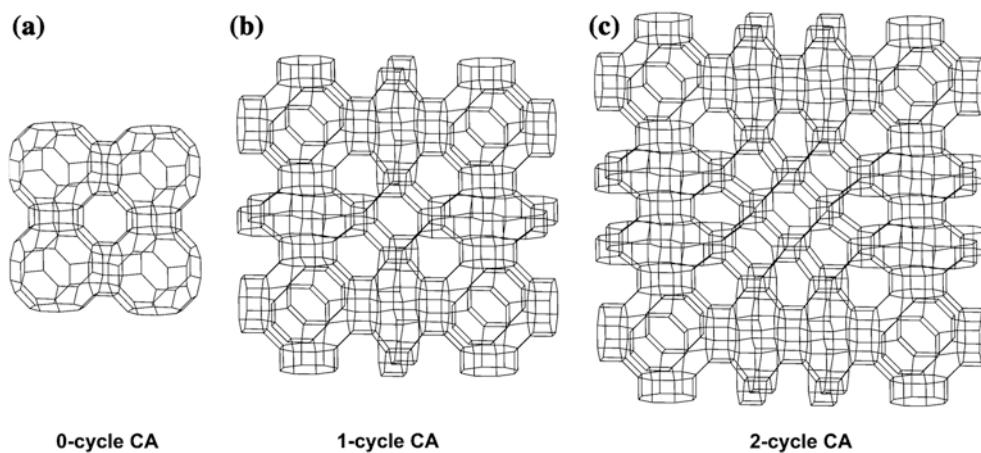
**Fig. 2** Bipyramidal nanocomplex as construction of 4-MRs (shown as squares): side view (a), stratified structure (b), top view (c).



**Fig. 3** Levels of paulingite bipyramidal nanocomplex as a result of work of the CA: schematic structure of contours in different layers (a), transition to the square lattice (b), 4 cycles of work of 4-colored CA (c).



**Fig. 4** Graphs of zeolite frameworks consisting of nanocomplexes formed as a result of work of paulingite CA: zeolite Rho RHO (a), hypothetical zeolite ISC-1 (b) and paulingite (c).



**Fig. 5** Tetrahedral framework of the hypothetical material ISC-1 (a) and its calculated X-ray powder diffraction pattern (for  $\lambda = 1.54178 \text{ \AA}$ ) (b).

**Table 2** Atom coordinates and geometrical parameters of tetrahedral framework in hypothetical zeolite ISC-1 ( $a=25 \text{ \AA}$ , sp. gr.  $Im\bar{3}m$ )

Atom coordinates			
Atom	x	y	z
Si1	0.2400	0.0632	0.1520
Si2	0.3640	0.0632	0.1520
Si3	0.4368	0.0632	0.25
O1	0.2291	0	0.1317
O2	0.2256	0.1000	0.1000
O3	0.3014	0.0698	0.1710
O4	0.2015	0.0765	0.2015
O5	0.3756	0	0.1370
O6	0.3755	0.0980	0.0980
O7	0.4260	0	0.2661
O8	0.4020	0.0820	0.1999
Geometry of tetrahedra			
Bond	Bond length (Å)	Angle	Angle value (°)
Si1–O1	1.682	O1–Si1–O2	104.64
Si1–O2	1.633	O1–Si1–O3	109.81
Si1–O3	1.615	O1–Si1–O4	109.29
Si1–O4	1.603	O2–Si1–O3	112.74
		O2–Si1–O4	111.42
		O3–Si1–O4	108.81
Si2–O3	1.644	O3–Si2–O5	109.25
Si2–O5	1.650	O3–Si2–O6	110.67
Si2–O6	1.631	O3–Si2–O8	108.66
Si2–O8	1.599	O5–Si2–O6	106.95
		O5–Si2–O8	110.26
		O6–Si2–O8	111.01
Si3–O7	1.652 2x	O7–Si3–O7	111.74
Si3–O8	1.596 2x	O7–Si3–O8	112.53 2x
		O7–Si3–O8	106.40 2x
		O8–Si3–O8	107.21
Intertetrahedral angles			
Angle	Angle value (°)	Angle	Angle value (°)
Si1–O1–Si1	139.89	Si2–O5–Si2	146.55
Si1–O2–Si1	148.00	Si2–O6–Si2	148.38
Si1–O3–Si2	144.04	Si3–O7–Si3	145.95
Si1–O4–Si1	152.21	Si2–O8–Si3	145.65

It is of interest that infinite work of this CA will result in formation of a periodic 3-dimensional framework with an axial structure. Along its axis, with the period of 1 nm, will appear 8-MRs. We do not see any principal reasons why this structure cannot be formed under natural or experimental conditions. However, if they really form, their identifications would be a complex and separate task.

It is quite probable that paulingite is the by-product of the work of the CA described above that has been stopped on the 2nd cycle. In order to check this hypothesis, it is of interest to consider related structures

The zero-cycle of the CA shown in Fig. 3 corresponds to the structure of zeolite Rho [12–14] that is depicted in Fig. 4a. As paulingite, zeolite Rho crystallizes in cubic symmetry,

**Table 3** Theoretical X-ray diffraction pattern of hypothetical zeolite ISC-1 calculated from the data given in Table 2 ( $\lambda=1.54178 \text{ \AA}$ ,  $2\theta_{\max}=30^\circ$ ; 0-intensity lines are listed as well)

h	k	l	d (Å)	$2\theta$ (°)	I <sub>0</sub>
0	1	1	17.6777	5.00	16
0	0	2	12.5000	7.07	2
1	1	2	10.2062	8.66	9
0	2	2	8.8388	10.01	100
0	1	3	7.9057	11.19	3
2	2	2	7.2169	12.26	0
1	2	3	6.6815	13.25	66
0	0	4	6.2500	14.17	16
1	1	4	5.8926	15.03	0
0	3	3	5.8926	15.03	24
0	2	4	5.5902	15.85	2
2	3	3	5.3300	16.63	0
2	2	4	5.1031	17.38	18
0	1	5	4.9029	18.09	13
1	3	4	4.9029	18.09	5
1	2	5	4.5644	19.45	0
0	4	4	4.4194	20.09	0
0	3	5	4.2875	20.72	1
3	3	4	4.2875	20.72	1
0	0	6	4.1667	21.32	10
2	4	4	4.1667	21.32	0
1	1	6	4.0555	21.92	0
2	3	5	4.0555	21.92	5
0	2	6	3.9528	22.49	2
1	4	5	3.8576	23.05	0
2	2	6	3.7689	23.61	3
1	3	6	3.6860	24.14	0
4	4	4	3.6084	24.67	0
0	1	7	3.5355	25.19	1
0	5	5	3.5355	25.19	3
3	4	5	3.5355	25.19	0
0	4	6	3.4669	25.70	1
1	2	7	3.4021	26.19	12
2	5	5	3.4021	26.19	0
3	3	6	3.4021	26.19	5
2	4	6	3.3408	26.68	1
0	3	7	3.2827	27.16	16
1	5	6	3.1750	28.10	0
2	3	7	3.1750	28.10	0
0	0	8	3.1250	28.56	0
1	1	8	3.0773	29.02	2
1	4	7	3.0773	29.02	3
4	5	5	3.0773	29.02	0
0	2	8	3.0317	29.46	11
4	4	6	3.0317	29.46	0
3	5	6	2.9881	29.90	0

space group  $Im\bar{3}m$  with  $a \sim 15 \text{ \AA}$ . In this zeolite, D8-MRs are linked directly by 4-MRs. Thus, whereas paulingite corresponds to the 2nd cycle of the CA (Fig. 4c), zeolite Rho corresponds to the zero cycle of the CA.

## ISC-1: A New Theoretical Zeolite Framework

We could not find in literature any data concerning zeolite intermediate between paulingite and zeolite Rho. However, the model developed above allows to predict its hypothetical structure (Fig. 4b). It has a cubic unit cell,  $Im\bar{3}m$  with  $a \sim 25 \text{ \AA}$ . We were able to construct model of its tetrahedral framework (Fig. 5a). Atom coordinates and basic geometrical parameters are given in Table 2. We denote this zeolite as ISC-1 (Institute of Silicate Chemistry-1). Theoretical X-ray diffraction pattern of ISC-1 is shown in Fig. 5b (diffraction lines till  $2\theta_{\max} = 30^\circ$  are listed in Table 3).

Thus, zeolite Rho, hypothetical material ISC-1, and paulingite form a family of zeolites that result from the CA shown in Fig. 3. It can be proposed that more complex zeolites of the family can be synthesized that correspond to higher cycles of the CA work. Ideal symmetry of these zeolites will be cubic, space group  $Im\bar{3}m$ ,  $a = 15 + 10n \text{ [\AA]}$ , where  $n$  is a number of the CA cycle.

## Conclusions

1. By means of the concept of universal optimum and general principles of inorganic gene, structures of paulingite-related zeolites and minerals have been constructed.
2. A new hypothetical zeolite of the paulingite family has been predicted and named ISC-1 (Institute of Silicate Chemistry-1).

3. It appears to be very probable that universal optimum contains all necessary information for its material realization (sharp configurations) and construction of materials with certain type of interaction potential (even with restrictions existing in theory for this potential).

## References

1. Mackay AL (1967) Nature 216:159
2. Cohn H, Kumar AJ (2007) Am Math Soc 20:99
3. Ballinger B, Bleknerman G, Cohn H, Giansiracusa N, Kelly E, Schurmann A. Experimental study of energy-minimizing point configurations on spheres. arxiv:math.MG/0611451
4. Gordon EK, Samson S, Kamb WB (1966) Science 154:1004
5. Passaglia E, Gualtieri AF, Marchi E (2001) Eur J Miner 13:113
6. Vaughan DEW, Strohmaier KG (1999) Microporous Mesoporous Mater 28:233
7. Baerlocher Ch, Meier WM, Olson DH (2001) Atlas of zeolite framework types. Amsterdam, Elsevier
8. Friedrichs OD, Dress AWM, Huson D, Klinowski J, Mackay AL (1999) Nature 400:644
9. Toffoli T, Margolus N (1987) Cellular automata machines: a new environment for modeling. Boston, MIT Press
10. Mackay AL (1976) Phys Bull 495
11. Krivovichev SV (2004) Acta Crystallogr A60:257
12. Robson HE, Shoemaker DP, Ogievie RA, Manor PC (1973) Adv Chem Ser 121:106
13. McCusker LB (1984) Zeolites 4:51
14. Parise JB, Gier TE, Corbin DR, Cox DE (1984) J Phys Chem 88:1635