J Mol Evol (2016) 82:163–172
DOI 10.1007/s00239-016-9736-x

ORIGINAL ARTICLE

Hypothesis of Lithocoding: Origin of the Genetic Code as a "Double Jigsaw Puzzle" of Nucleobase-Containing Molecules and Amino Acids Assembled by Sequential Filling of Apatite Mineral Cellules

Nikolai E. Skoblikow[1,2] · Andrei A. Zimin[3]

Received: 25 September 2015 / Accepted: 29 March 2016 / Published online: 5 April 2016

Abstract The hypothesis of direct coding, assuming the direct contact of pairs of coding molecules with amino acid side chains in hollow unit cells (cellules) of a regular crystal-structure mineral is proposed. The coding nucleobase-containing molecules in each cellule (named "*lithocodon*") partially shield each other; the remaining free space determines the stereochemical character of the filling side chain. Apatite-group minerals are considered as the most preferable for this type of coding (named "*lithocoding*"). A scheme of the cellule with certain stereometric parameters, providing for the isomeric selection of contacting molecules is proposed. We modelled the filling of cellules with molecules involved in direct coding, with the possibility of coding by their single combination for a group of stereochemically similar amino acids. The regular ordered arrangement of cellules enables the polymerization of amino acids and nucleobase-containing molecules in the same direction (named "*lithotranslation*") preventing the shift of coding. A table of the presumed "*LithoCode*" (possible and optimal lithocodon assignments for abiogenically synthesized α-amino acids involved in lithocoding and lithotranslation) is proposed. The magmatic nature of the mineral, abiogenic synthesis of organic molecules and polymerization events are considered within the framework of the proposed "volcanic scenario".

Keywords Genetic Code · Translation · Abiogenesis · Apatite

✉ Nikolai E. Skoblikow
skoblikow@yandex.ru

[1] Laboratory of Microbiology, North-Caucasian Research Institute of Animal Husbandry, 4 Pervomayskaya Street, Znamenskiy Settlement, Krasnodar, Russia 350055

[2] Medical Laboratory "CityLab", 96 Moskovskaya Street, Krasnodar, Russia 350000

[3] Laboratory of Molecular Microbiology, Institute of Biochemistry and Physiology of Microorganisms, 5 Prosp. Nauki, Pushchino, Moscow Region, Russia 142290

Introduction (Hypothesis)

There are a number of theories explaining the origin of the genetic code (Knight et al. 2003; Yarus et al. 2009; Koonin and Novozhilov 2009; Francis 2013), most of which assume the origin of the code as a relatively late event of biogenesis. We propose a hypothesis of abiogenic origin of the code, which assumes that the sterochemical basis for coding (and translation) was established before the formation of peptides and the coding nucleic acids, and the main organizing role belongs to the regular mineral structure. The key difference of our hypothesis from all others is the idea about participation of nucleobase-containing molecules (NCMs) in coding not as part of a nucleic acid molecule (which even formed a specific secondary structure or was immobilized by some way (for example, Mellersh 1993; Mellersh and Wilkinson 2000)) but as covalently unbonded complexes, strictly located inside (not on!) fixed-size microspaces of the mineral surface.

In our opinion, such a stereochemical arrangement of amino acids and NCMs is possible under certain conditions. So, covalently unbonded NCMs contact with a stereochemically suitable amino acid side chain *in a fixed-size microspace* enabling no more than two NCMs and one amino acid side chain inside it. In the process of coding, NCMs are not covalently bound and are arranged variably with account for *partial shielding*; the first NCM is partially (or even totally) overlapped by the second one. The remaining free space of

such microspace is filled with the amino acid side chain *directly contacting* with the second (predominantly) and first NCMs. The regular ordered arrangement of a multitude of such microspaces enables the subsequent polymerization of amino acids and NCMs *in the same direction*. The very structure of such microspaces (their walls) assumes coding by certain triplets (doublets) of NCMs and *admits no shift of the coding*, i.e. no coding by combinations of NCMs from different adjacent triplets (doublets) as it could have been the case in the coding by triplets (doublets) of the already assembled RNA, even fixed on the surface of a mineral.

The described coding mechanism can be presented to be sufficiently spontaneous, if we assume the process of successive entering of NCMs and amino acids from some "primordial soup" into *superficial unit cells* (*cellules*) *of a mineral with microporous surface*. *Unit cells* (or *cellules*) are the main discrete structural elements of many minerals, including those traditionally considered in different theories of biogenesis. In our work, we will use the less widespread definition "*cellule*" (as a synonym of "unit cell") to prevent from confusing upon usage of the term "cell" relative to the living forms. We presumed that the superficial mineral cellule (probably with dislodged central ions) can serve as the abovementioned "microspace" for consequent filling of molecules, participating in the coding. According to our scheme, the first NCM will get to the lower part of the cellule (to the bottom, or will get stuck on its way); the second NCM will be piled up on it (totally or, most likely, partially shielding it), and the remaining space will be filled by the amino acid side chain according to the lock-and-key principle. We designate this coding method as *lithocoding* (ancient Greek: λίθος [lithos] for "stone, mineral"). The mineral cellule with NCMs arranged in it will be called *lithocodon*.

The mechanisms underlying the successive intrusion of NCMs and amino acids into the cellule can be different (adsorption, sedimentation, in situ synthesis or others) and will be considered below, as they are not important for the description of the coding mechanism. For the coding mechanism, it is essential that the NCMs in the cellule prove in close contact by being partially and mutually bound by functional groups, distinctly from the classical binding in the RNA molecule. Thus, the association of molecules that fill the cellule is a layered *Tetris*-like *double jigsaw puzzle* assembled under the action of physicochemical effects and structurally determined by the size of the molecules and the character of their functional groups.

Principles of Modelling

For theoretical verification of the hypothesis, we designed the assumed mineral cellule conforming to the required stereochemical parameters (Fig. 1). As the particular spatial arrangement of contacting molecules in the cellule is given only approximately to demonstrate the principle of coding by partial shielding, the size of mineral cellules can as yet be indicated also only within an approximate range. Thus, the depth of the cellules can vary from 10.5 to 21.0 Å (which is comparable with the longitudinal size of one–two nucleobases of different types); the length (the distance between the centres of mineral's cellules by the axis of future polymerization) of about 5.5–7.0 Å (which is comparable with the distance between α-carbon atoms of two successive amino acids with the gap between them); the width of about 3.5–7.0 Å (the width of one–two NCMs). Minerals of a different stereometry can be considered as evolutionally dead-end sites of alternative pre-biogenesis. Thus, in a mineral with longer cellules the specific binding of α-amino acid side chains with coding molecules is possible, but, owing to a greater distance between amino acids, polymerization between them would hardly be implementable. In this case, the polymerization would be

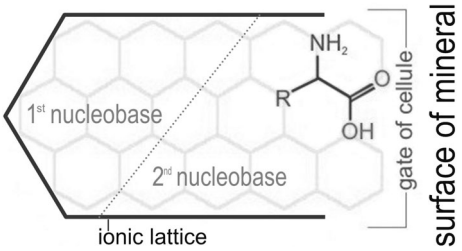

Fig. 1 Schematic structure of a mineral cellule and the order of its filling. It is necessary to be guided by the following conditions required to observe the assumed stereochemical interrelations between the contacting molecules: **a** the volume of the cellule admitting the arrangement of two nucleobase-containing molecules and one amino acid side chain; **b** the shape of the cellule (probably, slit-like) admitting its two-dimensional imaging; **c** the internal virtual structure showing the possible locations of functional groups of molecules; **d** availability of a locus for molecules to get into the cavity of the cellule ("gate"); the size of this locus should be commensurable with the size of amino acid. The fixed position of amino acid is oriented in the cellule gates in the direction corresponding to that of subsequent correct N → C polymerization, whereas the loci for nucleobases are indicated in the direction corresponding to that of subsequent 5′ → 3′ polymerization. It should be noted that the amino acid is positioned by the demonstrated way only in the "gates" of the cellule, which is already filled with both nucleobase-containing molecules—otherwise it "falls" deep into the cellule, resulting in interruption of polymerization/polycondensation of the peptide. The amino acid interacts with the coding nucleobases in the cellule only with its side chain, and not with the carboxyl or amino group. The carboxyl and amino groups avoid contacting the nucleobases, except for the cases of coding by *NA*-pairs. However, even in this case the coding involves only the amino acid side chain. The carboxyl and amino groups are involved solely in the subsequent polymerization into a peptide via the stage of interaction (includes the phosphorylation of the carboxyl group of the amino acid as an intermediate stage) with phosphate groups of the gates of the neighbouring cellules. The L-conformation of amino acid is not shown

possible for amino acids with a longer backbone, e.g. for β- or γ- (Liu and Orgel 1998). Possibly (and even for certain), such variants of coding took place at the dawn of the pre-biogenesis as the base for other versions of life.

On this scheme, we do not show carbohydrate and phosphate residues due to the following reasons. First of all, for the sake of greater visual expression, because the carbohydrate elements are, evidently, not involved in a direct stereochemical interaction with amino acids. Second, due to the indefinite arrangement of the planes of carbohydrate residues in the cellule. We do not rule out that nucleobases in cellules are arranged with respect to carbohydrate not in *trans* but in *cis* configuration as the most compact one. Third, we do not exclude (and even believe it to be highly probable) the participation of monomers lacking such large sugars as ribose in the primordial abiogenic coding. There is a number of convincing hypotheses, which speak for the possibility that simpler carbohydrates, like, for instance threose (Schöning et al. 2000) or glyceraldehyde (Zhang et al. 2005), acted as predecessors of pentoses (ribose and deoxyribose) in the prebiotic age, with the formation of Threofuranose Nucleic Acid (TNA) or Glycol-Nucleic Acid (GNA), respectively.

Also, the modelling schemes for the hypothetical arrangement of molecules in cellules have some other simplifications. The nucleobases are given only in oxo form, not in hydroxy form. Not whole amino acids but only their side chains starting from the α-carbon atom, which occupies a fixed position in the cellule gates, are indicated as molecules contacting with nucleobases. The probability of a more superficial arrangement of the second nucleobase (purine, especially adenine) without a direct contact with the first base is admitted. This is possible both in the case of the second nucleobase "stuck" for some reason in a more superficial segment of the cellule, and also if the nucleobase and amino acid simultaneously get into the cellule, thus promoting the amino acid side chain to be fixed between the first and second nucleobase.

Position		II	II	II	II
	Base	G	C	U	A
I	G	**Glycine**	**Alanine**	**Valine**	**Aspartic acid**
			α-Aminobuthyrate, Serine		*Glycine, Homoserine*
I	A	**Serine**	**α-Aminobuthyrate**	**Valine**	**Asparagine**
		2,3- Diaminopropionate	*Valine, Serine, Threonine*	*Isoleucine*	*2,4- Diaminobutyrate, Glycine*
I	C	**2,3- Diaminopropionate**	**Norvaline**	**Leucine**	**Glutamine**
		Serine	*2,4-Diaminobutyrate, Homoserine*	*Norvaline*	*Glutamate, Ornithine*
I	U	**Serine**	**NorLeucine**	**Leucine**	**Glutamic acid**
			Homoserine	*Norleucine*	*Glutamine*

Fig. 2 The LithoCode: optimal and possible assignments of lithocodons for abiogenically synthesized α-amino acids involved in the lithocoding. The abiogenically synthesized α-amino acids involved in the possible LithoCoding are presented. Only amino acids synthesized in discharge experiments in sufficient concentrations (including one non-coded in the modern code) are shown. Sulphur-contained amino acids are not shown. Localization of Cα-atom of amino acid is marked by *small shadowed ring*. More than one amino acid can conform to each doublet by reason of ambiguity and low specificity of the LithoCode. Most probable assignments are written in *bold* above each scheme. Other possible assignments are written in *italic* under each scheme. Assignments existing in the modern Standard Code are *underlined*. Only nucleobases of nucleobase-containing molecules are shown (without carbohydrate and phosphate residues)

Results of Modelling

Interpretation of Spatial Modelling

The considered combinations of molecules enable a conclusion with respect to not only the possibility of coding in a fixed space, but also the regularities providing for the ability of particular NCM pairs to code particular amino acids (and groups of amino acids). When modelling, we took into consideration the possibility that the spectrum of amino acids involved in primordial coding somewhat differed from the modern spectrum, as also noted by other authors (Copley et al. 2005; Hartman and Smith 2014). For this reason, we primarily considered amino acids not coded by the modern code (α-aminobutyrate, norvaline, norleucine, 2,3-diaminopropionate, 2,4-diaminobutyrate, ornithine) but formed by and detected in spark discharge experiments simulating the abiogenic synthesis of amino acids in the primitive atmosphere (Miller 1953; Miller and Urey 1959). Besides, the primary coding could be distinguished by a lower specificity, coding stereochemically homologous amino acids, i.e. one codon was used to code not only one, but a group of stereochemically similar amino acids. The ancient group coding of stereometrically similar amino acids considered within the framework of our hypothesis allows us to make assumptions about the possible assignments of some codons in the proposed table of prebiotic mineral-based code, named the *Litho-Code* (Fig. 2). Some commentaries about the results of modelling of hypothetical lithocodon–amino acid interactions are given below.

GG Two guanines being the bulkiest nucleobases fill the entire cellule, leaving no space for the amino acid side chain. A more superficial arrangement of the second guanine does not allow the further polycondensation of amino acids due to the amino group of guanosine that "sticks" out of the cellule. For this reason, the only possible amino acid capable of participating in the further synthesis of peptide is *glycine*, amino acid lacking a side chain.

GC Being a more compact base, cytosine enables arranging in the cellule the side chain that contains at least one methyl group, *alanine*. But amino acid with the side chain of two carbon atoms, α-aminobutyric acid, is also a probable candidate.

GU Uracil—an even more compact base, which, at that, partially interacts by its C_2 oxo-group with the amino group of guanine—enables arranging a branched side chain containing three carbon atoms, *valine*.

GA The positioning of adenine maximally close to guanine would have led to the blockage of the cellule gates by the hydrogen atom of the adenine amino group, making it impossible to include this combination into the coding process even for such compact amino acid as glycine. The combination of G + A NCMs could be included into the primary coding only when the amino acid side chain gets into the cellule simultaneously with (or slightly earlier than) the second adenine. In such a case, adenine should be positioned more superficially than the side chain, which could be considered as an element of the reverse coding of nucleobase by amino acid. It should be noted that no other base but adenine, which lacks the protruding functional group at the C_2 atom, "sticking" out of the cellule and preventing the further polycondensation of amino acids, could be suitable for such a configuration. Probably, this feature of adenine explains the exceptional variability of coded amino acids with particularly this nucleobase occurring in the second position of the codon. Herewith, the character of the amino acid side chain is stipulated primarily by the stereochemical properties of the first nucleobase. However, too large a size of amino acid side chain and/or its branched character would have led to the positioning of the second base in a too superficial locus of the cellule, resulting not only in the impossibility of the subsequent polycondensation of amino acids (due to the "sticking-out" state) but also in the hindrance of the following polymerization of NCMs in the nucleic acid chain. As the result, one of the most suitable side chains filling the gap between the first and the second guanines would be a short chain having a carboxyl group at the end to form a hydrogen bond with the amino group of guanine—*aspartic acid*.

AG In the case of a dense arrangement of the second guanine over the first adenine, the formed free space of the cellule allows for accommodation of a short side chain capable of interacting by the polar group either with the N_1-atom or with the oxo-group at the C_6-atom. An optimal amino acid in the most probable variant can be *serine*, especially in the case of the hydroxy (not oxo) form of guanine. Another suitable amino acid can be *2,3- diaminopropionate*.

AC It seems a possible but not an optimal combination for coding of *threonine*.

AU This combination, similar to the GU combination, admits the filling of the remaining space with the *valine* side chain. However, a more free contact of adenine with the oxygen-containing (oxo or, more likely, hydroxy) functional group at the C_4-atom, enables a more branched side chain to be located in the cellule, for instance, *isoleucine*.

AA It is an optimal combination for coding of *asparagine*. But glutamine is also possible.

CG The coding is similar to AG combination: *serine* or *2,3- diaminopropionate*.

CC Probably, the formation of cyclic *proline* (coded by this pair) is a later evolutionary acquisition. But in the proposed model it is possible that the amino acid is coded with the same three carbon atoms in the unbranched hydrophobic side chains—the norvaline.

CU The coding is similar to AU combinations to form isoleucine. However, when the functional group at C_4 atom is in the hydroxy form, uracil is in a closer contact with cytosine. This results in a changed uracil position, with *leucine* being one of the suitable amino acids. A combination including the non-branched side chains of three methyl groups—norvaline—was closer to the optimum, though. Apparently, the inclusion of atypical amino acids (for example, norvaline and norleucine) in proteins, observed in a number of cases and related to the changes in the functioning of isoleucyl- and leucyl-tRNA-synthases, is not only a reflection of a more ancient range of proteinogenic amino acids (Alvarez-Carreño et al. 2013), but also serves as an example of relict lithocoding.

CA It is an optimal combination for *glutamine*. However, ornithine is also possible.

UG In the case of a close arrangement of the second guanine over the first uracil, the formed free space of the cellule enables fitting in a short side chain capable of interacting with the N_1-atom of guanine by its polar group. An optimal variant of this side chain (especially in the case of the hydroxy (not oxo) form of guanine) can be both serine and its analogues *cysteine* and selenocysteine with the oxygen atom substituted for sulphur or selenium atom (the UGA codon at a number of conditions). Possibly, a greater compactness of uracil as compared with adenine makes it possible (in the case of adenine as the second coding nucleobase) to arrange in the cellule serine analogues containing larger atoms of sulphur and selenium instead of oxygen. But in the standard genetic code the UGA codon is terminating. The termination mechanism assumed within the framework of our hypothesis is shown and considered in more detail in Fig. 3.

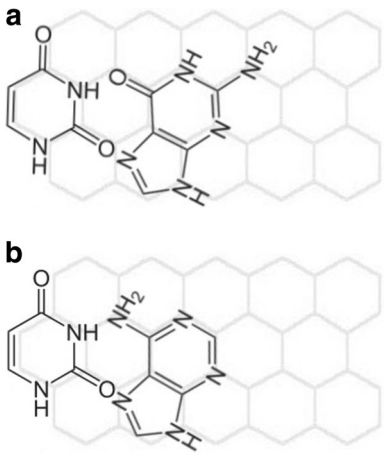

Fig. 3 Coding of termination of lithotranslation by UA and UG combinations. Formation of terminating combinations can be explained by the fact that the most compact uracil being in the first position, devoid of branched functional groups, possesses the greatest spatial mobility in the cellule, binding to nodes of the mineral lattice less rigidly. In this case, a purine nucleobase that gets into the cellule immediately afterwards has the possibility of binding to uracil (C_2-, N_3-, C_4-atoms) through the N_7-atom and the functional group at the C_6-atom. This Hoogsteen-like pairing leads to its abnormal arrangement in the cellule, not only preventing the contact of the carboxyl group of the coded amino acid with the amino group of the subsequent amino acid, but also preventing it from subsequent participation in the formation of a fully proper codon in the future nucleic acid chain. In the typical Hoogsteen pairing the group at the C_4-atom of the first uracil can form a *U–A* bond with the amino group of adenine with simultaneous bonding of the N_3-atom of uracil with the N_7-atom of adenine (Hoogsteen 1963). But in the hydroxy form of the group at the C_2-atom of the first uracil a *U–A* bond with the N_7-atom of adenine can be formed along with simultaneous bonding of the N_3-atom of uracil with the amino group of adenine (**a**). The formation of a U–G bond is possible according to a similar principle, which probably found its reflection in the stop coding by *UGA* (**b**)

UC This combination can code some amino acids including homoserine and *serine*.

UU The coding is similar to the *leucin*-coding CU combination.

UA The coding is similar to the CA combination.

Direct Stereometrical Origin of Stop Codons

Within the framework of our hypothesis, termination of the subsequent polymerization of amino acids assumes an event leading to such a translocation of the second nucleobase within the limits of the cellule, which would make impossible the coupling of the coded amino acid (its carboxyl group) with the amino group of the subsequent amino acid from the next cellule (Fig. 3).

Origin of Monomers Involved in the Coding

The synthesis of amino acids was possible outside the cellules according to the schemes and conditions

implemented in spark discharge experiments, particulary in the presence of phosphate (Saladino et al. 2006). However, the synthesis of NCMs from primary precursors (carbamide, formamide, cyanoacetylene, cyanamide, glycolaldehyde, glyceraldehyde) in situ in the presence of phosphate in the cellules of the mineral acting as a catalyst, simplifies the general scheme of events (Costanzo et al. 2012). Purine nucleotides can be synthesized abiogenically, from nucleosides; they, in turn, can be formed from nucleobases, whereas for pyrimidine nucleotides this route appeared problematic until recently. The results of experiments in this field show that in the presence of phosphate the above range of precursor molecules can form pyrimidine nucleotides avoiding the nucleoside formation stage (Powner et al. 2009). It cannot be excluded that these *precursor molecules* were the coding ones in the lithocoding.

Filling of Mineral Cellules and Types of Mineral Surface Interaction

An essential element of our hypothesis is the order and not the mechanism of filling of cellules with molecules, actors of the abiogenic coding process. Therefore our study is not aimed at establishing the particular way of filling of cellules or the type of interaction between the organic molecules and mineral surface. The issue is significantly simplified in the variant of the in situ synthesis of NCMs in mineral cellules themselves. We still think it is more probable, if not necessary, for amino acids synthesized outside the mineral to get into the cellules already containing the NCMs. In other cases, the mechanisms by means of which NCMs and amino acids successively get into the cellule can be different (adsorption, sedimentation or others).

The selective sorption and polymerization of amino acids on organophilic mineral surfaces was described earlier (Hazen and Sholl 2003; Lambert 2008). Also *encapsulation* of amino acids in silicalite was described. "The cluster of three molecules near the middle of the near-vertical channel has been optimized to interact mutually by way of hydrogen bonding and to be suspended by van der Waals bonding … ability of the zeolitic material both to accommodate large biopolymer precursors and to provide sites at which reactions may occur" (Smith et al. 1999). The removing of the central ion from the superficial mineral unit cell (cellule) enabled the intrusion of some small molecules into the remaining hollow microspaces of the superficial layer of the mineral under some destructive influences. Such mechanism was described in earlier experiments, in which the amino acid *intercalation* resulted in an *expansion of the unit cell* (Siffert and Naidja 1992; Benetoli et al. 2007). We proposed a sequential double-layer intercalation of NCMs and amino acids as the basic mechanism of our hypothesis.

Another advantage of our hypothesis is a potential simple solution of one serious problem of abiogenic protein synthesis. The separate localization of amino acid side chains within the mineral cellules relieves of need of *side chain protecting groups*, which are usually necessary for the prevention of unintended reactions in the solid-phase protein synthesis. The majority of authors speculating about the origin of translation prefer not to consider this problem.

Polymerization Events

An isolated consideration of any mechanism of amino acid coding loses sense if the further polymerization of amino acids into polypeptide is not considered (Nelsesteuen 1978). In our case, which does not assume the use of extended RNA molecule as a coding matrix, events that happen after the coding assume the combined and unidirectional polymerization of not only amino acids, but also NCMs. According to our hypothesis, polymerization of amino acids could proceed without the decomposition of the mineral cellule, owing to which after the abiogenic formation of peptide (which we propose to designate as *lithopeptide*) the cellules could be again filled with amino acids of the same specificity, which would lead to multiple synthesis of identical peptide macromolecules from the same mineral–nucleic matrix. We propose to designate this process as *lithotranslation*. It should be specifically noted that the cellules should not decompose *after single peptide synthesis*. The cellules remain intact and capable of being refilled. The binding of phosphate groups of the superficial layer with the C- and N-termini of adjacent amino acids during the peptide synthesis is temporary and reversible: the structural phosphate of the cellule acts as a catalyst and is not expended in the process of peptide chain formation. On the contrary, we postulate the potentiality of numerous filling of cellules with amino acids with side chains which conform to the NCMs sorbed in the deep of the cellule. In this case, the reproducibility of peptides with non-random sequence is ensured, which is a key and fundamental distinction of the scheme we propose from the simple mineral-dependent random polymerization of amino acids, placing it into the centre of the following biological evolution.

The unidirected lithopolymerization of amino acids and NCMs as a condition *sine qua non* for the implementation of the very principle of coding, is also well fit into the proposed "volcanic" scenario. A periodic change of temperature in a pool with the hot spring at the bottom (e.g. a fumarole) could be the main mechanism of multiple synthesis of lithopeptides (Fig. 4.). In our view, the terrestrial

anoxic geothermal fields located near volcanoes could have been the most geologically plausible environment for occurrence of the described events (Mulkidjanian et al. 2012). The general sequence of all the events described is given in the scheme which well illustrates the assumption that "*volcanic activity intruding into apatite-rich rocks might therefore have set the scene for the origin of the code*" (Cavalier-Smith 2001).

Thus, within the framework of our hypothesis the result of polymerization of NCMs is the formation of a NA molecule, not coupled with the mineral and capable to perform the ribozyme functions (if it is RNA) in the absence of a (proto)ribosomal protein-synthesizing apparatus. Herewith, immediately upon the appearance of ribosomes this molecule could play the role of an mRNA for the biosynthesis of protein products identical to products of the corresponding linear row of cellules. In case the amount of lithocodons and coded amino acids coincided, i.e. the principle "one NA molecule–one polypeptide" was implemented, it is reasonable to designate the formed polynucleotide product as *lithogen* or *lithocystrone*. Possibly, in particularly successful cases, some products of this lithotranslation (about ~25 amino acid residues long) with their respective RNA-lithocystrone could have played the main role in the organization of the functional core of the ancient ribosome (Peptidyl-Transferase Center, PTC) and the range of main proteins (Skoblikow and Zimin 2015).

Discussion

Role of Mineral's Nature

The most suitable for the role of lithocoding minerals are those of phosphate nature, particulary the minerals of the

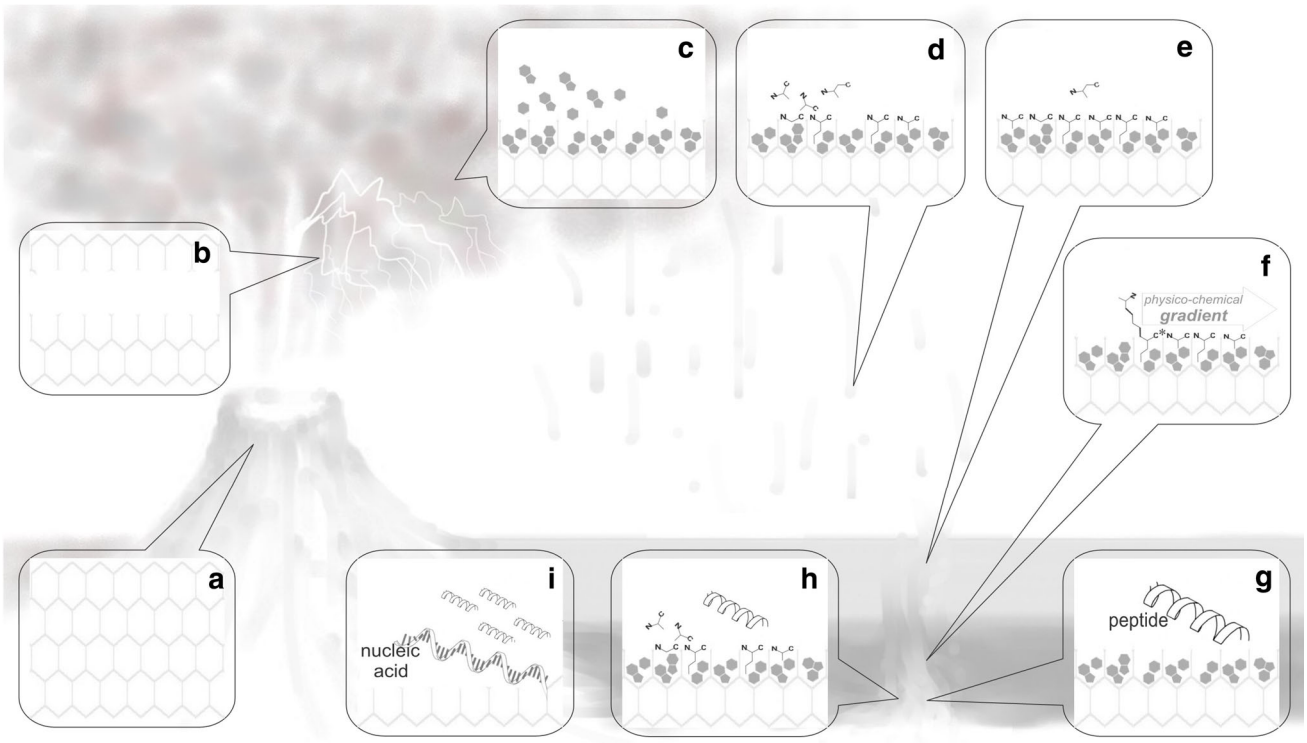

Fig. 4 The "volcanic scenario" of abiogenic mineral-dependent coding (lithocoding) and translation (lithotranslation). The overall process of lithocoding and lithotranslation should include the following stages. Magmatic *origin of mineral* with certain stereochemical parameters of cellules (**a**). Volcanic eruption. Mechanic (and/or thermoclastic) mechanism of fragmentation of mineral and deformation of fragments with *formation of a cellular surface* (**b**). Motion of fragments with high kinetic energies through more light organic clowd. Multiplicity of polar particles formed in eruptions as a factor stimulating electric discharges (lightnings are typical of eruptions) for *synthesis of amino acids* (**c**). Rapid sorption of primary reagents on the surface of particles and *primary filling of cellules*. High temperature of the surface of mineral fragments decreases the threshold energy of the reactions of *abiogenic synthesis of coding* (*nucleobase-containing*) *molecules* from simpler precursors on a superficial layer of the mineral (**d**). Falling of particles through a heavier (gas–liquid) phase containing amino acids and the *secondary filling* of the remaining cellular space according to the lock-and-key principle with *triply selected* amino acids (**e**). Final falling of particles with completely filled cellules into the liquid-phase containing both amino acids and other organic and inorganic compounds. The placing of particles in vertical or inclined position (**f**). The pH gradient or gradual cooling down of the falling particles as a factor creating a temperature gradient for the successive *polymerization of amino acids* on the mineral surface (**g**). The *release of polypeptides* (**h**). Potential repeating of the cycle (**e**–**g**) and *reproduction of the same polypeptides*. Slow dissolution of phosphate walls of the mineral superficial cellules (probably by inorganic acids H_3PO_4, H_2S, HCl), simultaneous *polymerization of nucleobase-containing molecules* and the subsequent *releasing of NA molecules* (**i**)

apatite group. There are dozens of kinds of different apatites and hundreds of other less suitable minerals (Pasero et al. 2010). Among apatite-group minerals, whitlockite ($Ca_9(MgFe)(PO_4)_6PO_3OH$) seems to be especially attractive, containing, apart from calcium, the ions of magnesium and iron, which play an important role in maintaining the basic biological processes. The key role of phosphorus compounds in the functioning of virtually all biological processes has long ago led the investigators to the idea of the importance of such minerals in the emergence of life. Herewith, attention was mainly paid to the metabolic and energetic functions of phosphates (Kulaev and Vagabov 1983; Kornberg et al. 1999). A number of investigators have established an important role of phosphates as catalytic sites for polymerization of amino acids (Ferris et al. 1996) and nucleotides (Acevedo and Orgel 1987). Such facts allowed for an assumption that "replication began on the surfaces of high-energy phosphate minerals" (Cavalier-Smith 2001). Phosphate minerals were considered not only as a catalytic site but also as a source of energy substrate ("inorganic phosphate mineral fuels") for polymerization of biomolecules. This metabolic process designated in this work as *lithophosphorylation* was considered to be "by far the simplest bioenergetic system using known enzymes". Finally, analysis and revelation of the regularities of the structural (the ability to form helical shapes) and stereometric characteristics (the correspondence of some sizes of mineral unit cellules and the size of nucleobases) of apatite and nucleic acids became the basis for assuming the emergence of the primordial coding system and the coding mechanism at the entry of organic molecules into the internal space of the mineral (Kosteckiy 2008). However, this hypothesis explains the irregularity of nucleic acids and peptides involved in the process by additional inclusions in the mineral, without disclosing the very mechanism of coding. But none of the mentioned works (including the most recent and very detailed, e.g. Norris et al. 2014) considered mineral phosphate as a structural, stereometrically determined basis of the coding, based on the principle of direct contact in the cellule with partial shielding of covalently unbonded NCMs.

The suggested nature of prebiotic hereditary material gives an additional benefit in the plausible way of prevention of the ancient hereditary material from damage by such unfavourable environmental factors as UV, X-rays and cosmic radiation. The main problem of DNA (and RNA) preservation in extreme conditions is unrepaired disruptions of NA molecules (Ronto et al. 2002). But in the case of *absence of covalent bonds between nucleotides* the damaging agent simply has no subject for disruption. The proposed way of organization of coding molecules can provide stabilization, protection, relatively long existence and distant transfer of this unusual and "raw" hereditary material in non-terrestrial conditions, which may be interesting for astrobiology. Moreover, the supposed litho-translational system can provide not only a possibility for storage and transfer of hereditary material, but also primitive peptide synthesis in very harsh abiogenic conditions typical for the earliest stage of planetogenesis.

Due to the key role in the correct positioning of molecules participating in coding, the apatite cellules can be considered as prebiotic aminoacyl-tRNA-synthases. Such point of view was already suggested earlier by other authors: "in prebiotic world the aminoacyl-tRNA synthetase was a copper-montmorillonite" (Hartman 1995). This led to the suggestion that a "PolyP-peptide synergy linking phosphate metabolism with peptide synthesis was one of the key steps in the emergence of life" (Milner-White and Russell 2010). One of the results of this linkage could have been the "PolyP-template-directed synthesis of the peptide precursors of the modern translational machinery and the various kinases, phosphatases and ion channels that are incorporated into acidocalcisomes" (Norris et al. 2014). Probably, the intracellular granular inorganic phosphates (acidocalcisomes) are the relict apatite mineral fragments, which provided the archaic form of translation, not simple polymerization. In this case, the acidocalcisome is not just the main partner of the ribosome in living cells (Norris et al. 2014; Root-Bernstein and Root-Bernstein 2015), but is the *direct evolutionary and functional ribosomal ancestor*. Moreover, possibly, the granular inorganic polyphosphate (acidocalcisomes) can still play the relict (and spare) genetic role in modern cells, but this function is yet to be discovered and confirmed experimentally. Perhaps, this way of primitive translation still can be used for synthesis of some short peptides necessary for DNA protection in extreme conditions and for existence in viable but non-culturable bacterial forms.

Ways for Experimental Confirmation

Both any element of the hypothesis and the complete cycle of the described events can (and should) be experimentally verified. The key issue, which requires experimental verification, is the possibility of abiogenic formation of superficial mineral cellules with the required stereochemistry and capability of being filled with organic (primarily nucleobase-containing) molecules. But, at least, the phenomenon of the affinity of nucleic acids to crystalline apatite particles is an established fact (Okazaki et al. 2001) and is successfully used for the transfer of genes and plasmids into cells for therapeutic aims (Chowdhury et al. 2004, Batard et al. 2001). Herewith, it is interesting that acidic pH facilitates the release of DNA from dissolving particles (Chowdhury 2013). Probably, some elements of the experiments targeting the confirmation of our

hypothesis would require the use of methodic approaches applied in the cited works. As for the assumed mechanism of the entry of organic acids into mineral cellules, when modelling the mechanism, one should probably pay attention to CVD (chemical vapour deposition).

The further experimental confirmation of our hypothesis represents a successive series of step-by-step experiments aiming to simulate its various stages. Ideally, each successive stage assumes testing after successful implementation of the preceding stage. Herewith, the general sequence of experiments will presumably be as follows. First of all, an experiment for complementation in cellules with four types of nucleobase-containing molecules and abiogenically synthesized proteinogenic and non-proteinogenic amino acids (10–11 amino acids)—testing of the coding principle. Secondly, an experiment for polymerization of amino acids after the coding took place—testing of the abiogenic formation of a lithopeptide. Thirdly, an experiment for polymerization of NCMs accompanied by the decomposition of mineral cellules—testing of the abiogenic formation of a lithocystrone. Experimental verification of even some of the aspects of our theory will be a significant step forward in understanding the primordial mechanisms of the origin of the genetic code. Satisfactory implementation of the complete experiment will show the possibility of initial abiogenic replication of major biological polymers. Subsequently, the binding of these biopolymers by structures forming the primary cell envelope could lead to the construction of early cell formations.

Conclusion

All known stereochemical hypotheses of the origin of the genetic code are operating with some nucleic acid molecules of various natures and structures. Instead of this, we are operating with *covalently unbonded NCM complexes* strictly fixed in the mineral structure. In contrast with the other hypotheses, our concept explains and details the following aspects of the origin of the code: (1) the role of the primary prebiotic conditions not only in formation of organic monomers but also in implementation of the coding mechanism (the *sequence of filling* the cellules); (2) the mechanism of multiple isomeric selection including not only chiral selection of L-amino acids *but also selection of α-amino acids* with the formation of a genuine range of amino acids involved in the "*LithoCode*"; (3) the mechanism of *direct interaction* of coding molecules with amino acid side chains *in limited microspaces* of the mineral cellule; (4) the mechanism *preventing the interaction between adjacent amino acids* during abiogenic peptide synthesis; (5) the prebiotic mechanisms *preventing the* "*coding shift*" (which is quite probable in consideration of a fixed preformed NA as the coding matrix); (6) the stereochemically stipulated variability of *U*NN and N*A*N codon assignments; (7) the prebiotic *origin of stop codons*; (8) the origin, physicochemical nature and approximate stereometry of the *mineral*, *which acts as the primorial* (*de*)*coder—an apatite*; (9) the mechanism of the *successive unidirectional polymerization* of amino acids (named "*Lithotranslation*") and NCMs as the fundamental basis of code fixation.

All aspects of the hypothesis we considered fit well into the framework of a unified "volcanic scenario" bringing together the prebiotic mechanisms of the synthesis of amino acids and nucleobase-containing molecules, the primary coding mechanism itself and the abiogenic synthesis of polypeptides and early coding nucleic acids.

Acknowledgments We sincerely thank Vladimir Selivanov and Konstantin Shavkunov for providing substantial assistance in preparing the English version of this text.

References

Acevedo OL, Orgel EL (1987) Template-directed ligation of oligonucleotides on hydroxy-apatite: a model for complexation in a primitive ocean. Orig Life 18:441

Alvarez-Carreño C, Becerra A, Lazcano A (2013) Norvaline and norleucine may have been more abundant protein components during early stages of cell evolution. Orig Life Evol Biosph 43:363–375. doi:10.1007/s11084-013-9344-3

Batard P, Jordan M, Wurm F (2001) Transfer of high copy number plasmid into mammalian cells by calcium phosphate transfection. Gene 270:61–68

Benetoli LOB, de Souza CMD, da Silva KL, de Souza IG Jr, de Santana H, Paesano A Jr, da Costa ACS, Zaia CTBV, Zaia DAM (2007) amino acid interaction with and adsorption on clays: FT-IR and Mössbauer spectroscopy and X-ray diffractometry investigations. Orig Life Evol Biosph 37(6):479–493

Cavalier-Smith T (2001) Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. J Mol Evol 53(4–5):555–595

Chowdhury EH (2013) pH-responsive magnesium- and carbonate-substituted apatite nano-crystals for efficient and cell-targeted delivery of transgenes. Open J Gen 3:38–44

Chowdhury EH, Kunou M, Nagaoka M, Kundu AK, Hoshiba T, Akaike T (2004) High-efficiency gene delivery for expression in mammalian cells by nanoprecipitates of Ca–Mg phosphate. Gene 341:77–82

Copley SD, Smith E, Morowitz HJ (2005) A mechanism for the association of amino acids with their codons and the origin of the genetic code. Proc Natl Acad Sci USA 102(12):4442–4447

Costanzo G, Saladino R, Botta G, Giorgi A, Scipioni A, Pino S, Di Mauro E (2012) Generation of RNA molecules by a base-catalysed click-like reaction. ChemBioChem 13(7):999–1008. doi:10.1002/cbic.201200068

Ferris JP, Hill AR Jr, Liu R, Orgel LE (1996) Synthesis of long prebiotic oligomers on mineral surfaces. Nature 381(6577):59–61

Francis BR (2013) Evolution of the genetic code by incorporation of amino acids that improved or changed protein function. J Mol Evol 77(4):134–158. doi:10.1007/s00239-013-9567-y

Hartman H (1995) Speculations on the origin of the genetic code. J Mol Evol 40(5):541–544
Hartman H, Smith TF (2014) The evolution of the ribosome and the genetic code. Life 4:227–249. doi:10.3390/life4020227
Hazen MR, Sholl DS (2003) Surface and thin films Chiral selection on inorganic crystalline surfaces. Nat Mater 2:367–374
Hoogsteen K (1963) The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. Acta Crystallogr A 16:907–916
Knight R, Landweber L, Yarus M (2003) Tests of a stereochemical genetic code. In: Lapointe J, Brakier-Gingras L (eds) Translation mechanisms. Landes Bioscience, Georgetown, pp 115–128
Koonin EV, Novozhilov AS (2009) Origin and evolution of the genetic code: the universal enigma. IUBMB Life 61:99–111
Kornberg A, Rao NN, Ault-Riché D (1999) Inorganic polyphosphate: a molecule of many functions. Annu Rev Biochem 68:89–125
Kosteckiy EYa (2008) How life had arisen? Bull Pacif State Econ Univ 2:86–103 **(in Russian)**
Kulaev IS, Vagabov VM (1983) Polyphosphate metabolism in microorganisms. Adv Microbiol Physiol 24:83–171
Lambert J-F (2008) Adsorption and polymerization of amino acids on mineral surfaces: a review. Orig Life Evol Biosph. doi:10.1007/s11084-008-9128-3
Liu R, Orgel LE (1998) Polymerization on the rocks: beta-amino acids and arginine. Orig Life Evol Biosph 28(3):245–257
Mellersh AR (1993) A model for the prebiotic synthesis of peptides which throws light on the origin of the genetic code and the observed chirality of life. Orig Life Evol Biosph 23(4):261–274
Mellersh AR, Wilkinson AS (2000) RNA bound to a solid phase can select an amino acid and facilitate subsequent amide bond formation. Orig Life Evol Biosph 30(1):3–7
Miller S (1953) A production of amino acids under possible primitive earth conditions. Science 117:528–529
Miller SL, Urey HC (1959) Organic compound synthesis on the primitive earth. Science 130(3370):245–251
Milner-White EJ, Russell MJ (2010) Polyphosphate-peptide synergy and the organic takeover at the emergence of life. J Cosmol 10:3217–3229
Mulkidjanian AY, Bychkov AY, Dibrova DV, Galperin MY, Koonin EV (2012) Origin of first cells at terrestrial, anoxic geothermal fields. Proc Natl Acad Sci USA 109:E821–E830
Nelsesteuen GL (1978) Amino acid-directed nucleic acid synthesis. J Mol Evol 11:109–120
Norris V, Reusch RN, Igarashi K, Root-Bernstein R (2014) Molecular complementarity between simple, universal molecules and ions limited phenotype space in the precursors of cells. Biol Direct 9:28
Okazaki M, Yoshida Y, Yamaguchi S, Kaneno M, Elliott JC (2001) Affinity binding phenomena of DNA onto apatite crystals. Biomaterials 22:2459–2464
Pasero M, Kampf AR, Ferraris C, Pekov IV, Rakovan J, White TJ (2010) Nomenclature of the apatite supergroup minerals. Eur J Miner 22:163–179
Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459(7244):239–242
Ronto G, Gaspar S, Fekete A, Kerekgyarto T, Berces A, Grof P (2002) Stability of nucleic acid under the effect of UV radiation. Adv Space Res 30(6):1533–1538
Root-Bernstein M, Root-Bernstein R (2015) The ribosome as a missing link in the evolution of life. J Theor Biol 367(21):130–158. doi:10.1016/j.jtbi.2014.11.025
Saladino R, Crestini C, Ciciriello F, Costanzo G, Di Mauro E (2006) About a formamide-based origin of informational polymers: syntheses of nucleobases and favourable thermodynamic niches for early polymers. Orig Life Evol Biosph 5–6:523–531
Schöning K, Scholz P, Guntha S, Wu X, Krishnamurthy R, Eschenmoser A (2000) Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3′ → 2′) oligonucleotide system. Science 290(5495):1347–1351
Siffert B, Naidja A (1992) Stereoselectivity of montmorillonite in the adsorption and deamination of some amino acids. Clay Min 29:109–118
Skoblikow NE, Zimin AA (2015) A search for relict ribonucleotide and amino acid sequences that played a key role in the development of the ribosome and modern protein diversity. Math Biol Bioinform 10(1):116–130. doi:10.17537/2015.10.116
Smith JV, Arnold FP Jr, Parsons I, Lee MR (1999) Biochemical evolution III: polymerization on organophilic silica-rich surfaces, crystal–chemical modeling, formation of first cells, and geological clues Proc. Natl Acad Sci USA 96:3479–3485
Yarus M, Widmann JJ, Knight R (2009) RNA-amino acid binding: a stereochemical era for the genetic code. J Mol Evol 69(5):406–429
Zhang L, Peritz A, Meggers E (2005) A simple glycol nucleic acid. J Am Chem Soc 127(12):4174–4175. doi:10.1021/ja042564z