Chirality in a quaternionic representation of the genetic code

C. Manuel Carlevaro [a,b], Ramiro M. Irastorza [a,c], Fernando Vericat [a,d,*]

[a] Instituto de Física de Líquidos y Sistemas Biológicos, 59 Nro. 789, 1900 La Plata, Argentina
[b] Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Mozart Nro. 2300, C14071VT Buenos Aires, Argentina
[c] Instituto de Ingeniería y Agronomía, Universidad Nacional Arturo Jauretche, 1888 Florencio Varela, Buenos Aires, Argentina
[d] Grupo de Aplicaciones Matemáticas y Estadísticas de la Facultad de Ingeniería (GAMEFI), Universidad Nacional de La Plata, Calle 115 y 48, 1900 La Plata, Argentina

ABSTRACT

A quaternionic representation of the genetic code, previously reported by the authors (BioSystems 141 (10–19), 2016), is updated in order to incorporate chirality of nucleotide bases and amino acids. The original representation associates with each nucleotide base a prime integer quaternion of norm 7 and involves a function that assigns to each codon, represented by three of these quaternions, another integer quaternion (amino acid type quaternion). The assignation is such that the essentials of the standard genetic code (particularly its degeneration) are preserved. To show the advantages of such a quaternionic representation we have designed an algorithm to go from the primary to the tertiary structure of the protein. The algorithm uses, besides of the type quaternions, a second kind of quaternions with real components that we additionally associate with the amino acids according to their order along the proteins (order quaternions). In this context, we incorporate chirality in our representation by observing that the set of eight integer quaternions of norm 7 can be partitioned into a pair of subsets of cardinality four each with their elements mutually conjugate and by putting them into correspondence one to one with the two sets of enantiomers (*D* and *L*) of the four nucleotide bases adenine, cytosine, guanine and uracil, respectively. We then propose two diagrams in order to describe the hypothetical evolution of the genetic codes corresponding to both of the chiral systems of affinities: *D*-nucleotide bases/L-amino acids and *L*-nucleotide bases/D-amino acids at reading frames $5' \rightarrow 3'$ and $3' \rightarrow 5'$, respectively. Guided by these diagrams we define functions that in each case assign to the triplets of *D*- (*L*-) bases a L- (D-) amino acid type integer quaternion. Specifically, the integer quaternion associated with a given D-amino acid is the conjugate of that one corresponding to the enantiomer L. The chiral type quaternions obtained for the amino acids are used, together with a common set of order quaternions, to describe the folding of the two classes, L and D, of homochiral proteins.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Homochirality of nucleic acids and proteins is one of the attributes that characterize life on the Earth (Zucchi et al., 1999; Pályi et al., 2004). At present time all the living organisms in our planet have nucleic acids (DNA, RNA, etc.) with nucleotide bases that just take, between their two possible chiral forms, the one usually labelled as *D* (for dextro or right-handed). Also, their proteins are chains of amino acids all of which are enantiomers of (exclusively) the class L (for levo, or left-handed) except for the amino acid glycine that is not a chiral molecule.

To understand why the combination *D* for nucleotide bases and L for amino acids (and not any other) occurs in living systems is one of the greatest quests of Biology. The attempts to answer this question involves either biotic or abiotic arguments. In general the hypotheses of the first type assume that homochirality is determined by biological necessity and that it is the result of diverse selection mechanisms. Among the abiotic hypotheses are those that propose a "frozen accident" as the responsible of the homogeneous chirality and those that assume the existence of some asymmetric force that selects just one of the chiral forms (Bonner, 1997; Jorissen and Cerf, 2002; Goodman and Gershwin, 2006).

The problem with these theories is that, in general, they can not be experimentally checked. This difficulty is (at least partially) overcome by some hypotheses that claim that homochirality and

---

* Corresponding author at: Instituto de Física de Líquidos y Sistemas Biológicos, 59 Nro. 789, 1900 La Plata, Argentina.
E-mail addresses: manuel@iflysib.unlp.edu.ar (C. Manuel Carlevaro), rirastorza@iflysib.unlp.edu.ar (R.M. Irastorza), vericat@iflysib.unlp.edu.ar (F. Vericat).

the universal genetic code arose closely related. More precisely, that the genetic code, its translation direction and homochirality emerged through a common natural process of selection (Root-Bernstein, 2007). This approach has the advantage that it can incorporate the idea of an ancestral direct affinity between amino acids and nucleotide triplets. Further these triplets will acquire new functions, within a more modern translation machinery, in the form of codons and anticodons. The goal is that this early affinity between amino acids and triplets of nucleotide bases can at present be studied in standard laboratories by synthesizing small RNA-oligonucleotides (Yarus et al., 2005). Thus has been demonstrated (Yarus, 1988, 2000; Majerfeld et al., 2005; Majerfeld and Yarus, 2005; Legiewicz and Yarus, 2005; Hobish et al., 1995; Saxinger et al., 1971; Walker, 1977; Root-Bernstein, 2010) the preference of L-amino acids for D-bases triplets. Moreover, the non-biological affinity of D-amino acids for L-codons is also observed (Yarus, 1988; Root-Bernstein, 2010; Profy and Usher, 1984). Besides, the hypothesis of coevolution of homochirality with the genetic code makes a number of additional predictions such as the possibility that each codon can encode, according with the reading direction ($5' \rightarrow 3'$ or $3' \rightarrow 5'$), at least two different amino acids and thus the eventual existence of more than a single code becomes possible (Root-Bernstein, 2010).

We must mention, however, that, despite the possibility of existence of ancestral L-ribonucleic sites for D-amino acids, there is also experimental evidence of the impossibility of incorporating D-amino acids in protein structures in present biosynthetic pathways. Specifically it has been reported chiral discrimination during the aminoacylation in the active site of the aminoacyl tRNA synthetase and also during the peptide bond formation in the ribosomal peptidyl transferase center (Banik and Nandi, 2013; Englander et al., 2015).

The aim of this article is to show how a mathematical representation of the genetic code recently reported by us (Carlevaro et al., 2016) can naturally incorporate chirality in such a way that the resulting description be consistent with many of the previous observations.

Our original representation is guided by a diagram that we have proposed to sketch the evolution of the genetic code (see Fig. 2 in next section). The diagram is based on pioneering ideas by Crick (Crick et al., 1961; Crick, 1968) and include the physical concept of broken symmetry (Hornos and Hornos, 1993; Maddox, 1994; Stewart, 1994) in a very simple form that resembles the energy levels of an atom. The representation uses Hamilton quaternions (Hamilton, 1843, 1866) as main tool (see Box 1). These mathematical objects are a sort of generalization of the complex numbers and obey an algebra in many aspects similar to theirs but with the very important (for our purposes) property that the product is, in general, non commutative. In addition, the quaternions are ideal for representing rotations (see Box 2) with important advantages over the classical matrix representation (Altmann, 2005).

In our quaternionic representation we assign to each amino acid in a given protein two quaternions: (i) an integer one according to which one of the 20 standard amino acids it is (type quaternion) and (ii) a real one that determines its order inside the protein primary structure (order quaternion). The type quaternions are obtained as a quaternionic function of the codons where each of the three nucleotides bases is associated to an integer quaternion. The form of this function is inspired by the code evolution diagram. On the other hand, the order quaternions play a fundamental role in relation with the folding of the protein (Creighton, 1992; Ben-Naim, 2013).

The integer quaternions that we choose to associate with the nucleotides bases belong to a maximum cardinality subset of the set

$$\mathbf{H}_7\left(\mathbb{Z}\right) = \left\{(a_0, a_1, a_2, a_3) : a_0, a_1, a_2, a_3 \in \mathbb{Z}; \ a_0^2 + a_1^2 + a_2^2 + a_3^2 = 7, \quad a_0 > 0 \text{ and even}\right\}$$

Box 1: Hamilton quaternions

Definition

$q = (q_0, q_1, q_2, q_3)$ $q_i \in \mathbb{R} \, (i = 0, 1, 2, 3)$.

If $q_i \in \mathbb{N} \, (i = 0, 1, 2, 3)$ then q is said an *integer quaternion*.

Algebra

Let $p = (p_0, p_1, p_2, p_3)$ and $q = (q_0, q_1, q_2, q_3)$ then:

a) sum

$p + q = (p_0 + q_0, p_1 + q_1, p_2 + q_2, p_3 + q_3)$

b) scalar product

$p \cdot q = p_0 q_0 + p_1 q_1 + p_2 q_2 + p_3 q_3$

c) quaternionic product

$pq = (c_0, c_1, c_2, c_3)$ where

$c_0 = p_0 q_0 - p_1 q_1 - p_2 q_2 - p_3 q_3$

$c_1 = p_0 q_0 + p_1 q_0 + p_2 q_3 - p_3 q_2$

$c_2 = p_0 q_2 - p_1 q_3 + p_2 q_0 + p_3 q_1$

$c_3 = p_0 q_3 + p_1 q_2 - p_2 q_1 + p_3 q_0$

The quaternionic product is *not conmutative* say, in general,

$pq \neq qp$

d) conjugation

$\tilde{q} = (q_0, -q_1, -q_2, -q_3)$ ($\tilde{q}$ is said the *conjugate* of q)

e) norm

$N(q) = q \cdot \tilde{q} = q_0^2 + q_1^2 + q_2^2 + q_3^2$

If $N(q) = 1$ then we say that q is an *unitary quaternion*.

f) inverse

$q^{-1} = \tilde{q}/N(q)$ verifies $qq^{-1} = q^{-1}q = (1, 0, 0, 0)$

Box 2: Quaternions and 3D rotations

A 3D vector $\vec{v} = (v_1, v_2, v_3)$ can be thought as a quaternion $v = (0, \vec{v}) = (0, v_1, v_2, v_3)$.

The rotation of the vector $\vec{v}$ by an angle θ around the axis $\hat{e}$ to give a new vector $\vec{v}\prime$ can be performed according to $v\prime = qvq^{-1}$ where the unitary quaternion $q = (q_0, q_1, q_2, q_3)$ is such that $\frac{\vec{q}}{|\vec{q}|} = \hat{e}$ and $\frac{|\vec{q}|}{q_0} = \tan(\theta/2)$ with $|\vec{q}| = \sqrt{q_1^2 + q_2^2 + q_3^2}$.

---

with the property that it does not contain pairs of conjugate quaternions. The set $\mathbf{H}_7\,(\mathbb{Z})$ has $7 + 1 = 8$ elements (Davidoff et al., 2003) and so the chosen subset has 4 quaternions as it should be. This suggests to partition the set $\mathbf{H}_7\,(\mathbb{Z})$ in the form $\mathbf{H}_7\,(\mathbb{Z}) = \mathbf{H}_{7;D}\,(\mathbb{Z}) \bigcup \mathbf{H}_{7;L}\,(\mathbb{Z})$ (where the four quaternions of $\mathbf{H}_{7;D}\,(\mathbb{Z})$ and the four quaternions of $\mathbf{H}_{7;L}\,(\mathbb{Z})$ are mutually conjugate) and to associate the elements of $\mathbf{H}_{7;D}\,(\mathbb{Z})$ and $\mathbf{H}_{7;L}\,(\mathbb{Z})$ with the four D-nucleotide bases and the four L-nucleotide bases of D-RNA and L-RNA molecules, respectively. This way, chirality can be included in our formalism.

In next section two diagrams for the evolution of the genetic code are presented assuming that the two possible combinations (D, L) and (L, D) for the chirality of bases triplets and amino acid were present from beginnings. The diagrams at certain step freeze displaying two different genetic codes. The code corresponding to the combination (D, L) is the present day living systems standard code. Guided by these diagrams, in Section 3, we consider the quaternionic representation of the codes. We assign type quaternions to L- and D-amino acids in such a way that, for a given D-amino acid, the associated integer quaternion is the conjugate of the corresponding to the enantiomer L. Section 4 is devoted to show the advantages of the quaternionic representation by considering the folding of L- and D-proteins. Some remarks are finally made in Section 5.

2. Chiral diagrams for the evolution of the genetic code

Here we generalize, by including chirality, the diagram proposed in Ref. Carlevaro et al. (2016) to describe the genetic code evolution. First we take into account that Miller–Urey like experiments on the
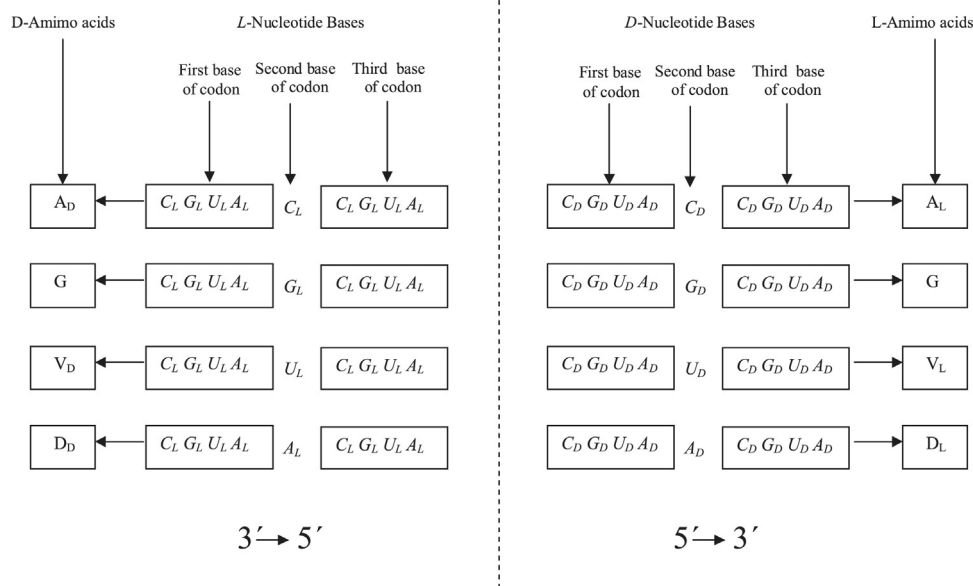
Fig. 1. Diagram for the beginnings of the genetic code evolution. The one letter convention for amino acids is used. The two chiral combinations (D-bases/L-amino acids and L-bases/D-amino acids) are identified with adequate subindices (see text). Note the achirality of the amino acid glycine (G). The translation direction considered in each case is also shown. Rectangles with four bases imply fourfold degeneration with respect to those ones, so at that moment each of the four considered amino acids was encoded by 16 triplets.

synthesis of organic molecules in a primordial environment (Miller et al., 1953; Parker et al., 2011) show the formation of amino acids in racemic mixtures. Also, as we have already mentioned, the preference of D- and L-bases triplets by L- and D-amino acids, respectively, has been observed by studying the affinity of amino acids with small RNA-oligonucleotides (Root-Bernstein, 2010). Thus we assume that from the beginnings, the two chirality systems (D-bases/L-amino acids) and (L-bases/D-amino acids) evolve without to interfere one with the other. Moreover, as we describe next, the proposed evolution diagrams for the two systems are very similar since we assume that the changes in the corresponding genetic codes are basically independent of the chirality. Chirality just manifests in settling the two affinity systems.

To construct our diagram of the code evolution in Ref. Carlevaro et al. (2016) we have followed Crick (Crick, 1968) and have considered that, in the first evolution steps, only the second base of the bases triplets was effective in codifying (binding) amino acids. Here we assume that this fact is valid for the L-bases triplets as well as for the D-ones. Thus only four L- (D-)amino acids could be codified, each one by one of the four D- (L-)bases C_I (I-cytosine), G_I (I-guanine), U_I (I-uracil) and A_I (I-adenine) ($I=D$ (L)) independently of which the first and third bases are. In Fig. 1, that sketches this first step of the code evolution, this fact is denoted with a rectangle containing the four letters. This is consistent with Crick's suggestion that only a few amino acids were coded at the beginning.

According to the diagram, C_I would codify J-alanine (A_J); G_I, glycine (G); U_I, J-valine (V_J) and A_I J-aspartic acid (D_J) ($I,J = D$,L or L,D) whatever the first and third bases are. It is worth noting here that the four amino acids that we assume were the first ones to be codified are the first four in the Trifonov (Trifonov, 2000) consensus temporal order scale for the appearance of the amino acids (column of natural numbers in Fig. 2). The four amino acids A, G, V and D were also the first four that appeared under simulation of the primitive earth conditions in Miller experiments (Miller et al., 1953; Parker et al., 2011). We must also point out the two reading frames we are considering: $5' \to 3'$ for D-triplets and $3' \to 5'$ for L-triplets.

In Figs. 2 and 3 we show how the evolution continues for the pairs (D,L) or (L,D), respectively. Since confusion is not possible, we have ignored the subindices that denote chiral class for notation simplicity.

As the left (right) part of diagram of Fig. 2 (3) shows, our version of the primitive code is highly degenerate: in principle each of the four amino acids, A_J,G,V_J and D_J, could be encoded by $4^2 = 16$ codons (see also Fig. 1). Physically the idea of degeneration is closely related with the concept of symmetry and a very illustrative form to think about these concepts is by doing an analogy with the energy levels of an atom. In our case we would have four levels indexed each one with the letter corresponding to the second codon base, say C_I,G_I,U_I and A_I (main quantum number). We thus assume that, as the code evolves, the symmetry that causes that the amino acid codification be independent of the first (third) base of the codon, disappears. Because of this symmetry breaking, a part of the degeneration also disappears. In the diagrams each of the four initial levels splits into four new levels, one for each of the possible bases (C_I,G_I,U_I and A_I) at the first (third) place of the codon (secondary quantum number). Now we have a total of 16 levels indexed each one by two letters (the first and second (second and third) bases of the codon). Each level is fourfold degenerate in the codons third (first) base. One of the new levels follows codifying the same amino acid as before that the level splits whereas the other three codify a new amino acid each. We indicate with an arrow the four groups of codons that conserve the amino acid and with a simple line those that substitute the amino acid by a new one.

As the code continues evolving it suffers new breaking of symmetry so that the third (first) base of some codons bring into use or, in the atomic analogy, some of the fourfold degenerate levels split into two levels each one twofold degenerate. Those levels pointed out with an arrow follow codifying the same amino acid whereas the other levels substitute it for a new one.

Eventually, in subsequent steps, a few of the twofold degenerate levels split once more given two non-degenerate levels each. This is the case of codons that codify methionine (M_J), tryptophan (W_J) and (again) the stop signal. The case of isoleucine (I_J) is a particular one since the split level coincides with the twofold one which
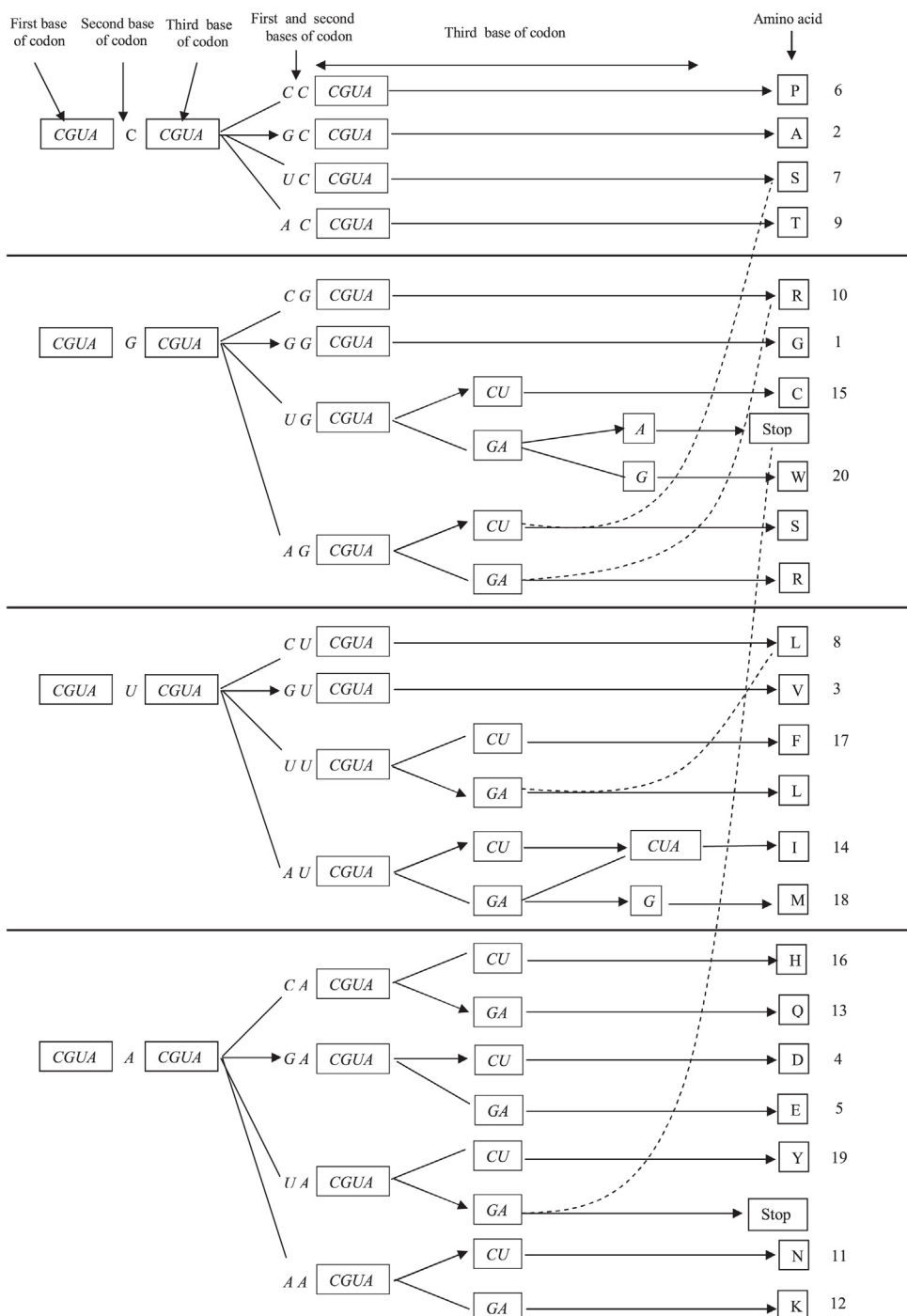
Fig. 2. Complete diagram of the genetic code evolution for the chiral combination *D*-bases/L-amino acids. The one letter convention for amino acids is used and, for simplicity, the chirality subindices in bases and amino acids have been dropped. The direction of the temporal evolution is from left to right. Rectangles with two or more bases implies degeneration with respect to those ones. The broken lines link different sets of codons that encode the same amino acid in the case of sixfold degeneration. Arrows and common lines indicate what codons follow codifying the same amino acid and what will start to codify a new one, respectively, after the symmetry is broken (see text). The natural numbers at the right of the amino acids give the temporal order of the amino acids in the Trifonov consensus scale (Trifonov, 2000).

represents the two codons that follow codifying the same amino acid. This way, isoleucine is the only amino acid which is coded by three codons. The stop signal is also threefold degenerate since it is coded by two groups of codons one twofold degenerate and the other one non-degenerate.

At this step of the evolution the code froze giving its present form. It is worth mentioning that the code evolution yields as a particular result that the amino acids serine (S_J), arginine (R_J) and leucine (L_J) are (would be) at present coded by two groups of codons

each one. In the three cases one of the groups is fourfold degenerate and the other one is twofold degenerate, so that these amino acids are the only three which are sixfold degenerate. We point out this property in the diagrams with a broken line linking the two groups of codons. The two groups of codons that codify the stop signal are also linked by a broken line. We remark again the similarity of the two evolution diagrams which means that, in our description, the symmetry breaking and the code freezing must depend of causes other than chirality.
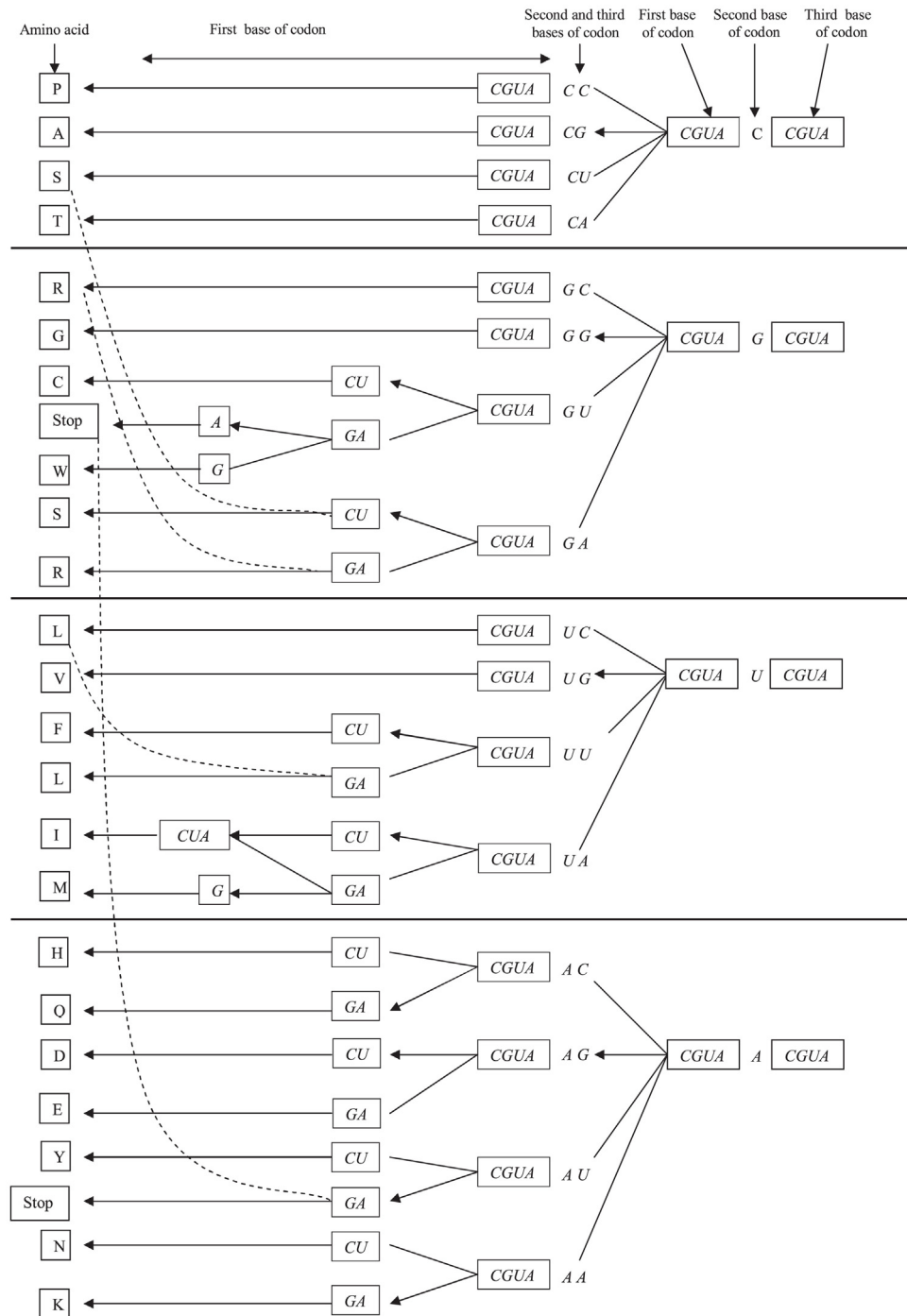
Fig. 3. Complete diagram of the genetic code evolution for the chiral combination *L*-bases/D-amino acids. The one letter convention for amino acids is used and, for simplicity, the chirality subindices in bases and amino acids have been dropped. The direction of the temporal evolution is from right to left. Rectangles with two or more bases implies degeneration with respect to those ones. The broken lines link different sets of codons that encode the same amino acid in the case of sixfold degeneration. Arrows and common lines indicate what codons follow codifying the same amino acid and what will start to codify a new one, respectively, after the symmetry is broken (see text).

We observe that the diagram of Fig. 2 is consistent with the present day standard genetic code (Fig. 4). The diagram of Fig. 3, on the other hand, gives the hypothetical code shown in Fig. 5. Actually, this code is not observed in Nature. We guess that, although our world (and perhaps the whole Universe) has the accurate symmetry for the existence of at least two systems of homochirality, say (*D*-nucleic acids and L-proteins) and (*L*-nucleic acids and D-proteins), the refinements in the decoding and synthesis machineries of living organisms, maybe looking for better robustness and optimization of error-correcting tools, have broken that symmetry by

discriminating between the two systems along the evolution causing the disappearance of the second one from the Earth.

3. Quaternionic representation of the genetic code and chirality

Based on the diagrams of Figs. 2 and 3, we propose, within a common formalism, quaternionic representations of the genetic

codes corresponding to both homochiral combinations (D,L) and (L,D) according to the scheme

$$
\begin{array}{ccc}
\mathcal{B}_I^3 & \longrightarrow & \mathcal{A}_J \\
\downarrow & & \downarrow \\
\mathbf{H}_{7,I}^3(\mathbb{Z}) & \longrightarrow & \mathbf{H}(\mathbb{Z})
\end{array}
\tag{1}
$$

with IJ = D,L and L,D where $\mathbf{H}(\mathbb{Z})$ denotes the set of integer quaternions (Lipschitz integers),

$$
\mathcal{B}_I = \left\{ C_I, G_I, U_I, A_I \right\} \quad (I = \text{D, L}),
\tag{2}
$$

$$
\mathcal{A}_J = \{P_J, A_J, S_J, T_J, R_J, G, C_J, W_J, L_J, V_J, F_J, I_J, M_J, H_J, Q_J, D_J, E_J,
$$
$$
Y_J, N_J, K_J, \text{Stop}_J\}
\tag{3}
$$

(J = L, D),

$$
\mathbf{H}_{7,D}(\mathbb{Z}) = \{(2, 1, 1, 1), (2, -1, 1, 1), (2, 1, -1, 1), (2, 1, 1, -1)\}
\tag{4}
$$

and

$$
\mathbf{H}_{7,L}(\mathbb{Z}) = \{(2, -1, -1, -1), (2, 1, -1, -1),
$$
$$
(2, -1, 1, -1), (2, -1, -1, 1)\}.
\tag{5}
$$

Here $\mathcal{B}_I^3$ is the set of the 64 I-codons (I = D,L) and we assume that the correspondence $\mathcal{B}_I^3 \to \mathcal{A}_J$ (IJ = D,L; L,D) are the genetic codes

as described by Figs. 4 and 5, whereas the functions $\mathcal{B}_I^3 \to \mathbf{H}_{7,I}^3(\mathbb{Z})$ assigns to each I-codon a triplet of quaternions of the sets $\mathbf{H}_{7,I.}(\mathbb{Z})$ (I = D,L).

Since the norm of all the quaternions of $\mathbf{H}_7(\mathbb{Z})$ is 7, which is a prime number, all the elements of $\mathbf{H}_7(\mathbb{Z})$ are prime quaternions (Davidoff et al., 2003). Assigning prime quaternions to the nucleotide bases gives they a certain character of *elemental* molecules in the present context.

In what follows, in order to simplify the notation, we shall assign natural numbers to identify the bases and the amino acids: $C \to 1$, $G \to 2$, $U \to 3$, $A \to 4$ and $P \to 1$, $A \to 2$, $S \to 3$, $T \to 4$, $R \to 5$, $G \to 6$, $C \to 7$, $W \to 8$, $L \to 9$, $V \to 10$, $F \to 11$, $I \to 12$, $M \to 13$, $H \to 14$, $Q \to 15$, $D \to 16$, $E \to 17$, $Y \to 18$, $N \to 19$, $K \to 20$, Stop$\to 21$.

Inspired by the diagrams of Figs. 2 and 3 we define the quaternionic functions

$$
F_{IJ} : \mathbf{H}_{7,I}^3(\mathbb{Z}) \to \mathbf{H}(\mathbb{Z})
$$
$$
\left(q_{\beta I}, q_{\gamma I}, q_{\delta I} \right) \to \alpha_{iJ} = F_{IJ}\left[\left(q_{\beta I}, q_{\gamma I}, q_{\delta I} \right) \right],
\tag{6}
$$

(I, J = D, L and L, D) by Eqs. (7) and (8):

$$
\begin{aligned}
P_L &\to \alpha_{1L} = q_{1D}q_{1D} & (\beta = 1, \gamma = 1, \delta = 1, 2, 3, 4) \\
A_L &\to \alpha_{2L} = q_{2D}q_{1D} & (\beta = 2, \gamma = 1, \delta = 1, 2, 3, 4) \\
S_L &\to \alpha_{3L} = q_{3D}q_{1D} = q_{4D}q_{2D} + \gamma_{2D;13} & (\beta = 3, \gamma = 1, \delta = 1, 2, 3, 4 \text{ or } \beta = 4, \gamma = 2, \delta = 1, 3) \\
T_L &\to \alpha_{4L} = q_{4D}q_{1D} & (\beta = 4, \gamma = 1, \delta = 1, 2, 3, 4) \\
R_L &\to \alpha_{5L} = q_{1D}q_{2D} = q_{4D}q_{2D} + \gamma_{2D;24} & (\beta = 1, \gamma = 2, \delta = 1, 2, 3, 4 \text{ or } \beta = 4, \gamma = 2, \delta = 2, 4) \\
G &\to \alpha_6 = \frac{7}{2}\left(q_{2D}q_{2D} + \tilde{q}_{2D}\tilde{q}_{2D}\right) & (\beta = 2, \gamma = 2, \delta = 1, 2, 3, 4) \\
C_L &\to \alpha_{7L} = q_{3D}q_{2D} + \gamma_{2D;13} & (\beta = 3, \gamma = 2, \delta = 1, 3) \\
W_L &\to \alpha_{8L} = q_{3D}q_{2D} + \gamma_{2D;24} + \delta_{2D;2} & (\beta = 3, \gamma = 2, \delta = 2) \\
L_L &\to \alpha_{9L} = q_{1D}q_{3D} = q_{3D}q_{3D} + \gamma_{3D;24} & (\beta = 1, \gamma = 3, \delta = 1, 2, 3, 4 \text{ or } \beta = 3, \gamma = 3, \delta = 2, 4) \\
V_L &\to \alpha_{10L} = q_{2D}q_{3D} & (\beta = 2, \gamma = 3, \delta = 1, 2, 3, 4) \\
F_L &\to \alpha_{11L} = q_{3D}q_{3D} + \gamma_{3D;13} & (\beta = 3, \gamma = 3, \delta = 1, 3) \\
I_L &\to \alpha_{12L} = q_{4D}q_{3D} + \gamma_{3D;13} = q_{4D}q_{3D} + \gamma_{3D;24} + \delta_{3D;4} & (\beta = 4, \gamma = 3, \delta = 1, 3, 4) \\
M_L &\to \alpha_{13L} = q_{4D}q_{3D} + \gamma_{3D;24} + \delta_{3D;2} & (\beta = 4, \gamma = 3, \delta = 2) \\
H_L &\to \alpha_{14L} = q_{1D}q_{4D} + \gamma_{4D;13} & (\beta = 1, \gamma = 4, \delta = 1, 3) \\
Q_L &\to \alpha_{15L} = q_{1D}q_{4D} + \gamma_{4D;24} & (\beta = 1, \gamma = 4, \delta = 2, 4) \\
D_L &\to \alpha_{16L} = q_{2D}q_{4D} + \gamma_{4D;13} & (\beta = 2, \gamma = 4, \delta = 1, 3) \\
E_L &\to \alpha_{17L} = q_{2D}q_{4D} + \gamma_{4D;24} & (\beta = 2, \gamma = 4, \delta = 2, 4) \\
Y_L &\to \alpha_{18L} = q_{3D}q_{4D} + \gamma_{4D;13} & (\beta = 3, \gamma = 4, \delta = 1, 3) \\
N_L &\to \alpha_{19L} = q_{4D}q_{4D} + \gamma_{4D;13} & (\beta = 4, \gamma = 4, \delta = 1, 3) \\
K_L &\to \alpha_{20L} = q_{4D}q_{4D} + \gamma_{4D;24} & (\beta = 4, \gamma = 4, \delta = 2, 4) \\
\text{Stop}_L &\to \alpha_{21L} = q_{3D}q_{2D} + \gamma_{2D;24} + \delta_{2D;4} = q_{3D}q_{4D} + \gamma_{4D;24} & (\beta = 3, \gamma = 2, \delta = 4 \text{ or } \gamma = 4, \delta = 2, 4)
\end{aligned}
\tag{7}
$$

$$P_D \rightarrow \alpha_{1D} = q_{1L}q_{1L} \qquad (\beta = 1, 2, 3, 4, \; \gamma = 1, \; \delta = 1)$$

$$A_D \rightarrow \alpha_{2D} = q_{1L}q_{2L} \qquad (\beta = 1, 2, 3, 4, \; \gamma = 1, \; \delta = 2)$$

$$S_D \rightarrow \alpha_{3D} = q_{1L}q_{3L} = q_{2L}q_{4L} + \gamma_{2L;13} \qquad (\beta = 1, 2, 3, 4, \; \gamma = 1, \; \delta = 3 \text{ or } \beta = 1, 3, \; \gamma = 2, \; \delta = 4)$$

$$T_D \rightarrow \alpha_{4D} = q_{1L}q_{4L} \qquad (\beta = 1, 2, 3, 4, \; \gamma = 1, \; \delta = 4)$$

$$R_D \rightarrow \alpha_{5D} = q_{2L}q_{1L} = q_{2L}q_{4L} + \gamma_{2L;24} \qquad (\beta = 1, 2, 3, 4, \; \gamma = 2, \; \delta = 1 \text{ or } \beta = 2, 4, \; \gamma = 2, \; \delta = 4)$$

$$G \rightarrow \alpha_6 = \frac{7}{2}(q_{2L}q_{2L} + \tilde{q}_{2L}\tilde{q}_{2L}) \qquad (\beta = 1, 2, 3, 4, \; \gamma = 2, \; \delta = 2)$$

$$C_D \rightarrow \alpha_{7D} = q_{2L}q_{3L} + \gamma_{2L;13} \qquad (\beta = 1, 3, \; \gamma = 2, \; \delta = 3)$$

$$W_D \rightarrow \alpha_{8D} = q_{2L}q_{3L} + \gamma_{2L;24} + \delta_{2L;2} \qquad (\beta = 2, \; \gamma = 2, \; \delta = 3)$$

$$L_D \rightarrow \alpha_{9D} = q_{3L}q_{1L} = q_{3L}q_{3L} + \gamma_{3L;24} \qquad (\beta = 1, 2, 3, 4, \; \gamma = 3, \; \delta = 1 \text{ or } \beta = 2, 4, \; \gamma = 3, \; \delta = 3)$$

$$V_D \rightarrow \alpha_{10D} = q_{3L}q_{2L} \qquad (\beta = 1, 2, 3, 4, \; \gamma = 3, \; \delta = 2)$$

$$F_D \rightarrow \alpha_{11D} = q_{3L}q_{3L} + \gamma_{3L;13} \qquad (\beta = 1, 3, \; \gamma = 3, \; \delta = 3) \tag{8}$$

$$I_D \rightarrow \alpha_{12D} = q_{3L}q_{4L} + \gamma_{3L;13} = q_{3L}q_{4L} + \gamma_{3L;24} + \delta_{3L;4} \qquad (\beta = 1, 3, 4, \; \gamma = 3, \; \delta = 4)$$

$$M_D \rightarrow \alpha_{13D} = q_{3L}q_{4L} + \gamma_{3L;24} + \delta_{3L;2} \qquad (\beta = 2, \; \gamma = 3, \; \delta = 4)$$

$$H_D \rightarrow \alpha_{14D} = q_{4L}q_{1L} + \gamma_{4L;13} \qquad (\beta = 1, 3, \; \gamma = 4, \; \delta = 1)$$

$$Q_D \rightarrow \alpha_{15D} = q_{4L}q_{1L} + \gamma_{4L;24} \qquad (\beta = 2, 4, \; \gamma = 4, \; \delta = 1)$$

$$D_D \rightarrow \alpha_{16D} = q_{4D}q_{2L} + \gamma_{4L;13} \qquad (\beta = 1, 3, \; \gamma = 4, \; \delta = 2)$$

$$E_D \rightarrow \alpha_{17D} = q_{4L}q_{2L} + \gamma_{4L;24} \qquad (\beta = 2, 4, \; \gamma = 4, \; \delta = 2)$$

$$Y_D \rightarrow \alpha_{18D} = q_{4L}q_{3L} + \gamma_{4L;13} \qquad (\beta = 1, 3, \; \gamma = 4, \; \delta = 3)$$

$$N_D \rightarrow \alpha_{19D} = q_{4L}q_{4L} + \gamma_{4L;13} \qquad (\beta = 1, 3, \; \gamma = 4, \; \delta = 4)$$

$$K_D \rightarrow \alpha_{20D} = q_{4L}q_{4L} + \gamma_{4L;24} \qquad (\beta = 2, 4, \; \gamma = 4, \; \delta = 4)$$

$$\text{Stop}_D \rightarrow \alpha_{21D} = q_{2L}q_{3L} + \gamma_{2L;24} + \delta_{2L;4} = q_{4L}q_{3L} + \gamma_{4L;24} \qquad (\beta = 4, \; \gamma = 2, \; \delta = 3 \text{ or } \beta = 2, 4, \; \gamma = 4, \; \delta = 3)$$

The importance of working with objects that verify a non commutative algebra is evident from these functions since otherwise amino acids A_J and R_J, and also S_J and L_J, would have associated the same quaternion. The expression for the amino acid glycine (G) takes into account that it is not chiral. The factor 7 has to do with the fact that: (a) the norm of the type quaternions can roughly be taken as a measure of the information needed to codify the corresponding amino acid in the sense that the larger the norm the larger the necessary information (see Ref. Carlevaro et al., 2016); (b) G is fourfold degenerate and the norm for all the other amino acids which are fourfold degenerate is 49 (see Eqs. (7)–(10)).

In Eqs. (7) and (8), the quaternions $\gamma_{iD;jk}$ ($\gamma_{iL;jk}$) accounts for the level splitting when the second base of codon is i and the third (first) base is jk= 13 ($C_I U_I$) or 24 ($G_I A_I$) I = D (L). Analogously, the quaternion $\delta_{iD;j}$ ($\delta_{iL;j}$) accounts for the level splitting when the second base of the codon is i and the third (first) base is j=2 (G_I) or 4 (A_I) I = D (L). Thus, in principle we have as unknown quaternions $\gamma_{2I;13}$, $\gamma_{2I;24}$, $\gamma_{3I;13}$, $\gamma_{3I;24}$, $\gamma_{4I;13}$, $\gamma_{4I;24}$ and $\delta_{2I;2}$, $\delta_{2I;4}$, $\delta_{3I;2}$ and $\delta_{3I;4}$. Of these 10 unknown quaternions we can find 5, say $\gamma_{2I;13}$, $\gamma_{2I;24}$, $\gamma_{3I;13}$, $\gamma_{3I;24}$, $\gamma_{4I;24}$, by requiring that those amino acids which are coded by two different groups of codons (case of codons sixfold degenerates or codons that codify the stop signal) have associated an unique quaternion and also that the two ways to reach isoleucine (I) give the same quaternion (see Fig. 2 (Fig. 3)). To obtain the quaternions $\delta_{2I;2}$, $\delta_{2I;4}$, $\delta_{3;2}$ and $\delta_{3I;4}$ we have assigned to those levels that can not split more (non degenerate levels) the product of the quaternions associated with each of the corresponding bases: $\alpha_{8J} = q_{3I}q_{2I}q_{2I}$; $\alpha_{13J} = q_{4I}q_{3I}q_{2J}$; $\alpha_{21J} = q_{3I}q_{2I}q_{4J}$; $\alpha_{12J} = q_{4I}q_{3I}q_{4J}$. Finally

for the remaining unknown quaternion $\gamma_{4I;13}$ we have proposed $\gamma_{4I;13} = -\gamma_{4I;24}$.

Taking: $q_{1D} = (2, 1, 1, 1)$, $q_{2D} = (2, -1, 1, 1)$, $q_{3D} = (2, 1, -1, 1)$ and $q_{4D} = (2, 1, 1, -1)$ in Eq. (7), we have explicitly obtained

$$
\begin{aligned}
&\alpha_{1L} = (1, 4, 4, 4) &\quad &\alpha_{12L} = (6, 17, 3, -3) \\
&\alpha_{2L} = (3, 0, 6, 2) &\quad &\alpha_{13L} = (18, 3, -1, 3) \\
&\alpha_{3L} = (3, 2, 0, 6) &\quad &\alpha_{14L} = (-10, 7, 5, -1) \\
&\alpha_{4L} = (3, 6, 2, 0) &\quad &\alpha_{15L} = (16, -3, 7, 1) \\
&\alpha_{5L} = (3, 0, 2, 6) &\quad &\alpha_{16L} = (-8, 3, 3, -3) \\
&\alpha_6 = (7, 0, 0, 0) &\quad &\alpha_{17L} = (18, -7, 5, -1) \\
&\alpha_{7L} = (3, -2, -6, 8) &\quad &\alpha_{18L} = (-8, 9, 1, 1) \\
&\alpha_{8L} = (6, -15, -1, 9) &\quad &\alpha_{19L} = (-12, 9, 3, -5) \\
&\alpha_{9L} = (3, 6, 0, 2) &\quad &\alpha_{20L} = (14, -1, 5, -3) \\
&\alpha_{10L} = (5, 2, 2, 4) &\quad &\alpha_{21L} = (18, -1, 3, 3). \\
&\alpha_{11L} = (2, 17, 1, 3) &\quad &
\end{aligned} \tag{9}
$$

Analogously, taking $q_{1L} = \tilde{q}_{1D} = (2, -1, -1, -1)$, $q_{2L} = \tilde{q}_{2D} = (2, 1, -1, -1)$, $q_{3L} = \tilde{q}_{3D} = (2, -1, 1, -1)$ and $q_{4L} = \tilde{q}_{4D} = (2, -1, -1, 1)$ in Eq. (8) we have

$$\alpha_{1D} = (1, -4, -4, -4) \qquad \alpha_{12D} = (6, -17, -3, 3)$$
$$\alpha_{2D} = (3, 0, -6, -2) \qquad \alpha_{13D} = (18, -3, 1, -3)$$
$$\alpha_{3D} = (3, -2, 0, -6) \qquad \alpha_{14D} = (-10, -7, -5, 1)$$
$$\alpha_{4D} = (3, -6, -2, 0) \qquad \alpha_{15D} = (16, 3, -7, -1)$$
$$\alpha_{5D} = (3, 0, -2, -6) \qquad \alpha_{16D} = (-8, -3, -3, 3)$$
$$\alpha_{6} = (7, 0, 0, 0) \qquad \alpha_{17D} = (18, 7, -5, 1) \qquad (10)$$
$$\alpha_{7D} = (3, 2, 6, -8) \qquad \alpha_{18D} = (-8, -9, -1, -1)$$
$$\alpha_{8D} = (6, 15, 1, -9) \qquad \alpha_{19D} = (-12, -9, -3, 5)$$
$$\alpha_{9D} = (3, -6, 0, -2) \qquad \alpha_{20D} = (14, 1, -5, 3)$$
$$\alpha_{10D} = (5, -2, -2, -4) \qquad \alpha_{21D} = (18, 1, -3, -3).$$
$$\alpha_{11D} = (2, -17, -1, -3)$$

We denote the set of quaternions assigned to the amino acids as given by Eqs. (9 and 10) by $\mathbf{H}_{\alpha D}(\mathbb{Z})$ and $\mathbf{H}_{\alpha L}(\mathbb{Z})$, respectively. We see that the elements of $\mathbf{H}_{\alpha D}(\mathbb{Z})$ and $\mathbf{H}_{\alpha L}(\mathbb{Z})$ verify $\alpha_{iD} = \tilde{\alpha}_{iL}$ ($i = 1, 2, \ldots, 20, 21$) say, the quaternions assigned to both enantiomers of a given amino acid are mutually conjugate.

4. Folding of L- and D-proteins in the quaternions formalism

We have presented quaternionic representations of the standard genetic code for living systems, where the chiral combination (D-bases/L-amino acids) is preferred, and also of a hypothetical genetic code for systems, in which we assume that the combination (L-bases/D-amino acids) prevails among other possible chiral combinations. These representations reproduce the structure of the corresponding codes, particularly their degeneration. However, the fact that distinguishes the quaternionic representation over most of available mathematical representations is the resulting assignation of quaternions to the amino acids. Because of the advantages of using quaternions to describe spatial rotations, the association of amino acids with quaternions opens new horizons beyond the genetic code representation (Carlevaro et al., 2016). In this context, we consider here the suitability of this association, together with our characterization of chirality, to take account of the folding of homochiral proteins formed by exclusively L or D amino acids.

The primary structure of a J-protein (J = L,D) formed by N J-amino acids is a sequence $A_{J1}, A_{J2}, \ldots, A_{JN}$ with $A_{Ji} \in \mathcal{A}_J$. Our aim is to obtain from this sequence the spatial coordinates of each one of the atoms of all the amino acids that constitute the protein in the native – or functional – state (tertiary structure). For L-proteins, we take as such the one corresponding to the protein in physiological solution whose coordinates can be obtained, after crystallization, by application of, for example, X-ray diffraction methods. That is the case of most of the proteins whose coordinates are stored at the Protein Data Bank (Bank). For D-proteins, since in the laboratories have been synthesized very few such proteins (Mandal et al., 2012; Liu et al., 2010), we take in these cases as tertiary structure the mirror-image of the corresponding experimental L-proteins structure.

In principle we restrict ourselves to determine the coordinates for just the alpha-carbon atoms of the proteins chain which is not a severe restriction since it is known that there exist very efficient algorithms for going (at least for L-proteins) from this trace representation to the full atoms one (Rotkiewicz and Skolnick, 2008). We also take into account that, in our quaternionic representation, the J-amino acids sequence is expressed as a sequence of quaternions $p_{J1}, p_{J2}, \ldots, p_{JN}$ with $p_{Ji} \in \mathbf{H}_{\alpha J}(\mathbb{Z})$. Under these conditions we proceed now to present an algorithm to determine the spatial coordinates of the alpha-carbon atoms of the protein.

Nucleotide bases: D – Amino acids: L

Second codon base

	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

First codon base (extreme 5′)

Fig. 4. Standard genetic code. The three letters convention for the amino acids is used and the third base in the codons is remarked in bold. For simplicity the subindices that identify the class of enantiomer have been ignored. The reading direction of codons is $5' \to 3'$.

Nucleotide bases: L – Amino acids: D

Second codon base

	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	CUU	Phe	CCU	Ser	CAU	Tyr	CGU	Cys
	AUU	Leu	ACU	Ser	AAU	Stop	AGU	Stop
	GUU	Leu	GCU	Ser	GAU	Stop	GGU	Trp
C	UUC	Leu	UCC	Pro	UAC	His	UGC	Arg
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	AUC	Leu	ACC	Pro	AAC	Gln	AGC	Arg
	GUC	Leu	GCC	Pro	GAC	Gln	GGC	Arg
A	UUA	Ile	UCA	Thr	UAA	Asn	UGA	Ser
	CUA	Ile	CCA	Thr	CAA	Asn	CGA	Ser
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
	GUA	Met	GCA	Thr	GAA	Lys	GGA	Arg
G	UUG	Val	UCG	Ala	UAG	Asp	UGG	Gly
	CUG	Val	CCG	Ala	CAG	Asp	CGG	Gly
	AUG	Val	ACG	Ala	AAG	Glu	AGG	Gly
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Third codon base (extreme 5′)

Fig. 5. Hypothetical genetic code for the chiral combination L-bases/D-amino acids. The three letters convention for the amino acids is used and the first base in the codons is remarked in bold. For simplicity the subindices that identify the class of enantiomer have been ignored. The reading direction of codons is $3' \to 5'$.

As it was pointed out in Carlevaro et al. (2016), although adjacent alpha-carbon atoms are not covalently bonded their distance is notably stable and take very similar values for all the pairs within a given protein and also for those belonging to different proteins. So in our calculations we assume that all these distances are equal to a unique value $d_{C_\alpha - C_\alpha} = 3.80$ Å. Thus we determine on the unit sphere with center at the origin a point for each of the amino acids (alpha-carbon atoms) in the protein sequence. To the last one we assign directly the origin, the preceding one is located at the intersection between the axis z and the sphere surface (versor $\hat{e}_z$). To each of the remaining alpha-carbon atoms we assign a point on the sphere surface that results of rotating the versor $\hat{e}_z$ (north pole) by a quaternion. For the ith alpha-carbon atom in the J-sequence, the quaternion responsible of the rotation is denoted $\hat{\beta}_{Ji}$ ($i = 3, 4, \ldots, N$).

We then expand the chain of alpha-carbon atoms from their location on the sphere into the back-bone protein three
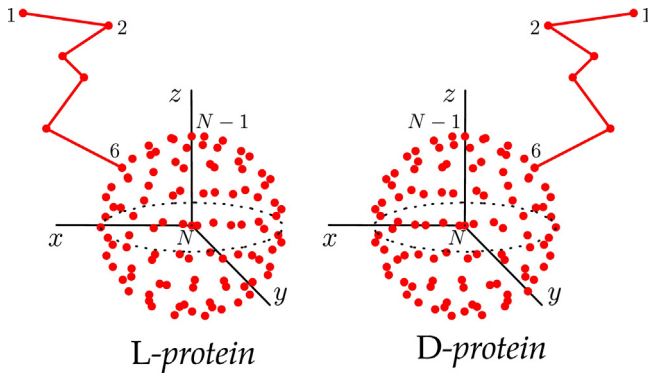
Fig. 6. Development of the alpha-carbon atoms backbones of a hypothetical L-protein of length N and the corresponding D-enantiomer from their position on the sphere surface into their spatial configuration (schematic). Both chains are the mirror-image one of the other. In each spatial chain the last two alpha-carbon atoms, as well as some of the first ones, are labelled by their order number inside the sequence.

dimensional configuration (see Fig. 6) by means of the following iterative procedure (initially the $\mathbf{r}_j$'s are on the sphere surface):

$$\text{do } i = 1, N - 2$$
$$\delta\mathbf{r} = \mathbf{r}_{i+1}$$
$$\text{do } j = 1, i$$
$$\mathbf{r}_j = \mathbf{r}_j + \delta\mathbf{r}$$
$$\text{enddo}$$
$$\text{enddo}$$

According to the algorithm, the distance between adjacent alpha-carbon atoms is the unit so, to establish the correct distance, we must multiply the final calculated coordinates by $d_{C_\alpha - C_\alpha}$.

It remains to determine how to calculate the quaternions $\hat{\beta}_{Ji}(i = 3, 4, \ldots, N)$. In Ref. Carlevaro et al. (2016) we do this in a somewhat heuristic way. We take into account that the ith amino acid interacts in some way with the $i - 1$ previous amino acids in the sequence and also with the $N - i$ subsequent ones. In these interactions the effect of the medium should be incorporated in some form, for example in the form of effective interactions between amino acids. Actually, we are trying for a sort of decodification and so we are not directly interested into the detailed form of the interactions, but we recognize that in any codification of information that involves those interactions, some trace of their general form should be. In general it is reasonable to think that the global interaction includes just two body (effective) interactions so by analogy we choose with generality for $\hat{\beta}_{Ji}$ the normalized version of the quaternion

$$\beta_{Ji} = \sum_{\substack{r = 1, 2, \ldots, N \\ r \neq i}} c_{Jr} \left(p_{Jr} \cdot p_{Ji} \right) \quad (J = L, D; \, i = 1, 2, \ldots, N) \quad (11)$$

where $\cdot$ denotes the quaternionic dot product:

$$p_{Jr} \cdot p_{Ji} = \left(p_{Jr} \right)_0 \left(p_{Ji} \right)_0 + \left(p_{Jr} \right)_1 \left(p_{Ji} \right)_1 + \left(p_{Jr} \right)_2 \left(p_{Ji} \right)_2 + \left(p_{Jr} \right)_3 \left(p_{Ji} \right)_3$$

and $c_{Jr} \in \mathbf{H}(\mathbb{R}) \, (r = 1, 2, \ldots, N)$ are unknown real quaternions (order quaternions) which are determined by means of an optimization technique. As such we use the particle swarm optimization (PSO) procedure of Kennedy and Eberhart (1995) taking as function of fitness the difference between the coordinates of the alpha-carbon atoms calculated following the previous procedure and the corresponding experimental ones as read from the PDB. We take the rmsd (root-mean-square deviation) as a measure of this
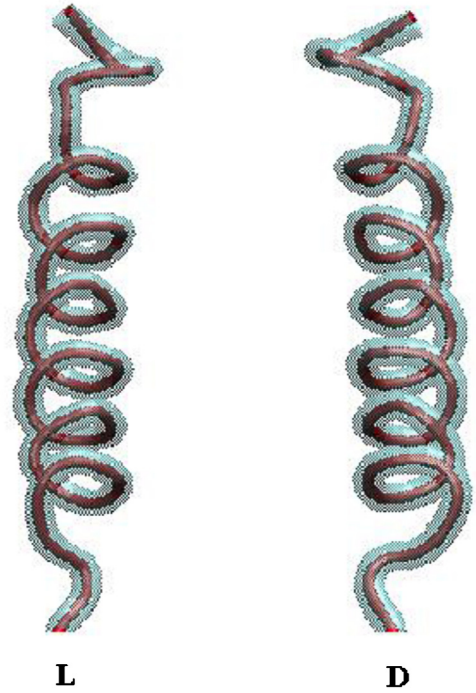

Fig. 7. Trace representation of the alpha-carbon atoms backbone for L- and D-1GCN. Red (dark grey) inner tube: from the coordinates obtained using our procedure. Cyan (light grey) external transparent tube: from the coordinates stored at PDB for L and its mirror-image for D. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

difference, using to that effect Bosco K. Ho's implementation of Kabsch algorithm (Kabsch, 1976).

Actually, it is enough to consider J = L since $p_{Dr} = \tilde{p}_{Lr}$ so, if we take

$$c_{Dr} = \mathbf{i} c_{Lr} \tilde{\mathbf{i}} = \tilde{\mathbf{i}} c_{Lr} \mathbf{i}, \quad (12)$$

with $\mathbf{i} = (0, 1, 0, 0)$, then a similar relationship is verified by β_{Di}:

$$\beta_{Di} = \mathbf{i} \beta_{Li} \tilde{\mathbf{i}} = \tilde{\mathbf{i}} \beta_{Li} \mathbf{i} = \left(\left(\beta_{Li} \right)_0, \left(\beta_{Li} \right)_1, -\left(\beta_{Li} \right)_2, -\left(\beta_{Li} \right)_3 \right). \quad (13)$$

We observe that if we denote with $\mathbf{x}_{Li} = ((\mathbf{x}_{Li})_1, (\mathbf{x}_{Li})_2, (\mathbf{x}_{Li})_3)$ the point on the sphere that results of rotating the versor $\hat{e}_z$ by the quaternion $\hat{\beta}_{Li}$: $(0, \mathbf{x}_{Li}) = \hat{\beta}_{Li} \left(0, \hat{e}_z \right) \tilde{\hat{\beta}}_{Li}$, then the sphere point that results of rotating the north pole by $\hat{\beta}_{Di}$ is given by

$$(0, \mathbf{x}_{Di}) = \hat{\beta}_{Di} \left(0, \hat{e}_z \right) \tilde{\hat{\beta}}_{Di} = \mathbf{i}(0, \mathbf{x}_{Li}) \mathbf{i} = (0, -(\mathbf{x}_{Li})_1, (\mathbf{x}_{Li})_2, (\mathbf{x}_{Li})_3),$$

$$(14)$$

say $\mathbf{x}_{Di}$ is the mirror-image of $\mathbf{x}_{Li}$ with respect to the plane $x = 0$ in a cartesian axis (x, y, z) (see Fig. 6). When we expand the two chains (L and D) of alpha-carbon atoms from their location on the sphere by using the previous iterative procedure they result to be the mirror-image one of the other.

In Figs. 7–9 we show the L and D enantiomers of three small proteins as obtained by using our procedure: in Fig. 7 of the hormone glucagon (PDB ID:1GCN – length: 29 amino acids); in Fig. 8 of the ion channel inhibitor osk1 toxin (PDB ID: 2CK5 – length: 31 amino acids) and, in Fig. 9, of a type III antifreeze protein (PDB ID:1HG7 – length: 66 amino acids). In all the cases the theoretic curves are compared with the corresponding experimental ones as obtained directly from the PDB for L-proteins and from the mirror-image of these for D-proteins.
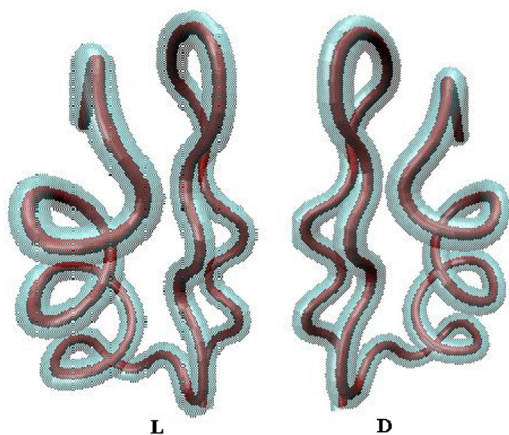
Fig. 8. Trace representation of the alpha-carbon atoms backbone for L- and D-2CK5. Red (dark grey) inner tube: from the coordinates obtained using our procedure. Cyan (light grey) external transparent tube: from the coordinates stored at PDB for L and its mirror-image for D. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)
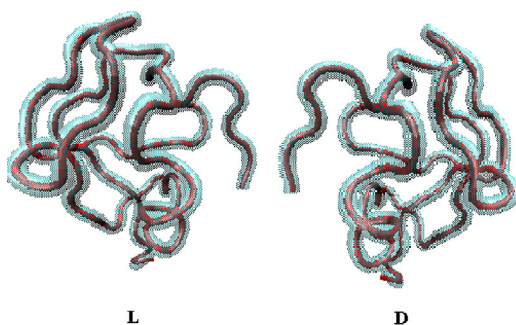

Fig. 9. Trace representation of the alpha-carbon atoms backbone for L- and D-1HG7. Red (dark grey) inner tube: from the coordinates obtained using our procedure. Cyan (light grey) external transparent tube: from the coordinates stored at PDB for L and its mirror-image for D. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

5. Remarks

Assuming that at the beginnings amino acids and nucleotide bases were synthesized from primordial elements in racemic mixtures and that *D*-bases were always more affine for *L*-amino acids whereas *L*-bases have preferred D-amino acids, we have proposed diagrams for the evolution of genetic codes which at present would settle the correspondence between codons and amino acids looking for those affinities. Actually, of both codes only that corresponding to the affinity system (*D*-bases/*L*-amino acids) is nowadays observed on the Earth. Although the existence of the chiral combination (*L*-bases/*D*-amino acids) is in principle possible, none of the organisms that live in our planet shows *L*-nucleic acids or D-proteins. However, the existence of life in other planets, including the possibility that it be governed by such a genetic code (Fig. 5), is an open issue.

Our evolution diagrams (Figs. 2 and 3) are based on pioneering ideas by Crick and introduce in a very simple way the concept of broken symmetry. The fact that the diagrams for both affinities systems, (*D*-bases/*L*-amino acids) and (*L*-bases/*D*-amino acids), follow a common pattern for the degeneration breaking and for the code freezing, means that we are considering that these aspects basically do not depend of the molecules chirality but of other physical, chemical and/or biological causes.

Inspired by the evolution diagrams we propose a quaternions based mathematical representation of the corresponding genetic codes. We assign to each nucleotide base an integer quaternion, so the codons are triplets of such quaternions. The representation assigns to each triplet another integer quaternion that is associated with one of the 20 amino acids (type quaternions). The bases quaternions belong to the set of eighth prime integer quaternions of norm 7 and the nucleotide bases chirality is introduced by partitioning this set into two subsets (Eqs. (4) and (5)) of cardinality 4 each with their elements mutually conjugates and associating their elements with the *D*- and *L*-bases, respectively. The correspondences between triplets of quaternions and type quaternions for both chiral combinations, (*D*/L) and (*L*/D), are given by functions (Eqs. (7) and (8)) that use the sum and ordinary product of quaternions and is such that the type quaternions assigned to both enantiomers of a given amino acid are mutually conjugate.

Apart of preserving the degeneration of the genetic codes as it should be, our representations distinguish among other mathematical representations of the genetic code because they assign quaternions to the amino acids as a final result so that, in view of the close relationship between quaternions and spatial rotations, a door towards the study of the proteins folding opens. In this context we propose an algorithm to go from the primary to the tertiary structure of L- as well as D-proteins. The algorithm uses, besides the integer type quaternions, a set of real quaternions associated with the order of the amino acids in the protein sequence. These order quaternions are basically the same ones for L- and D-proteins so the algorithm is such that for a given primary sequence the 3D structure of the L- and D-proteins are the mirror-image one of the other.

Finally, another observation about this algorithm whose critical step is the building of the quaternions β_{Ji} (Eq. (11)). In Ref. Carlevaro et al. (2016) we use for them an expression that involves the ordinary product between the type quaternion corresponding to the position i and all the others. However such expression is not adequate for describing with a common set of order quaternions the folding of L- and D-proteins. In this article to overcome this problem we change the ordinary quaternionic product by a dot product between type quaternions. This way the number of order quaternions sets adjusting a given protein diminishes. Thus the search for the unique set that describes the folding of all the proteins (if it exists!) would be facilitated. We are currently working on this issue.

References

Altmann, S.L., 2005]. Rotations, Quaternions, and Double Groups. Dover Books on Mathematics. Dover Publications, ISBN:9780486317731.

Banik, S.D., Nandi, N., 2013]. Chirality and Protein Biosynthesis. Springer, Berlin Heidelberg, pp. 255–305, http://dx.doi.org/10.1007/128_2012_369, ISBN:978-3-642-37626-9.

Protein Data Bank. http://www.rcsb.org/pdb/.

Ben-Naim, A., 2013]. The Protein Folding Problem and its Solutions. World Scientific, ISBN:9789814436359.

Bonner, W.A., 1997]. Homochirality and life. EXS 85, 159–188.

Carlevaro, C.M., Irastorza, R.M., Vericat, F., 2016]. Quaternionic representation of the genetic code. Biosystems 141, 10–19.

Creighton, T.E., 1992]. Protein Folding. WH Freeman and Company.

Crick, F.H.C., 1968]. The origin of the genetic code. J. Mol. Biol. 38 (3), 367–379.

Crick, F.H.C., Barnett, L., Brenner, S., Watts-Tobin, R.J., 1961]. General nature of the genetic code for proteins. Nature 192 (December (4809)), 1227–1232, http://dx.doi.org/10.1038/1921227a0.

Davidoff, G., Sarnak, P., Valette, A., 2003]. Elementary Number Theory, Group Theory and Ramanujan Graphs. Cambridge University Press, ISBN:9780521531436.

Englander, M.T., Avins, J.L., Fleisher, R.C., Liu, B., Effraim, P.R., Wang, J., Schulten, K., Leyh, T.S., Gonzalez, R.L., Cornish, V.W., 2015]. The ribosome can discriminate the chirality of amino acids within its peptidyl-transferase center. Proc. Natl. Acad. Sci. U. S. A. 112 (19), 6038–6043, http://dx.doi.org/10.1073/pnas.1424712112 http://www.pnas.org/content/112/19/6038.abstract.

Goodman, G., Gershwin, M.E., 2006]. The origin of life and the left-handed amino-acid excess: the furthest heavens and the deepest seas? Exp. Biol. Med. 231 (10), 1587–1592.

Hamilton, W.R., 1843]. On quaternions; or on a new system of imaginaries in algebra (letter to John T. Graves, dated October 17, 1843). Philos. Mag. 25, 489–495.

Hamilton, W.R., 1866]. Elements of Quaternions. Longmans, Green, & Company.

Hobish, M.K., Wickramasinghe, N.S.M.D., Ponnamperuma, C., 1995]. Direct interaction between amino acids and nucleotides as a possible physicochemical basis for the origin of the genetic code. Adv. Space Res. 15 (3), 365–382, http://dx.doi.org/10.1016/S0273-1177(99)80108-2, 0273–1177. http://www.sciencedirect.com/science/article/pii/S0273117799801082.

Hornos, J.E.M., Hornos, Y.M.M., 1993]. Algebraic model for the evolution of the genetic code. Phys. Rev. Lett. 71 (26), 4401.

Jorissen, A., Cerf, C., 2002]. Asymmetric photoreactions as the origin of biomolecular homochirality: a critical review. Orig. Life Evol. Biosph. 32 (2), 129–142.

Kabsch, W., 1976]. A solution for the best rotation to relate two sets of vectors. Acta Crystallogr. Sect. A: Cryst. Phys. Diffr. Theor. Gener. Crystallogr. 32 (5), 922–923.

Kennedy, J., Eberhart, R., 1995, November. Particle swarm optimization. In: IEEE International Conference on Neural Networks, 1995. Proceedings, vol. 4, pp. 1942–1948, http://dx.doi.org/10.1109/ICNN.1995.488968.

Legiewicz, M., Yarus, M., 2005]. A more complex isoleucine aptamer with a cognate triplet. J. Biol. Chem. 280 (20), 19815–19822.

Liu, M., Pazgier, M., Li, C., Yuan, W., Li, C., Lu, W., 2010]. A left-handed solution to peptide inhibition of the p53-MDM2 interaction. Angew. Chem. Int. Ed. 49 (21), 3649–3652.

Maddox, J., 1994]. The genetic code by numbers. Nature 367 (6459), 111.

Majerfeld, I., Yarus, M., 2005]. A diminutive and specific rna binding site for L-tryptophan. Nucl. Acids Res. 33 (17), 5482–5493.

Majerfeld, I., Puthenvedu, D., Yarus, M., 2005]. RNA affinity for molecular L-histidine; genetic code origins. J. Mol. Evol. 61 (2), 226–235.

Mandal, K., Uppalapati, M., Ault-Riché, D., Kenney, J., Lowitz, J., Sidhu, S.S., Kent, S.B.H., 2012]. Chemical synthesis and X-ray structure of a heterochiral {D-protein antagonist plus vascular endothelial growth factor} protein complex by racemic crystallography. Proc. Natl. Acad. Sci. U. S. A. 109 (37), 14779–14784.

Miller, S.L., et al., 1953]. A production of amino acids under possible primitive earth conditions. Science 117 (3046), 528–529.

Pályi, G., Zucchi, C., Caglioti, L., 2004]. Progress in Biological Chirality. Elsevier Science, Oxford, ISBN:9780080474038.

Parker, E.T., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Callahan, M., Aubrey, A., Lazcano, A., Bada, J.L., 2011]. Primordial synthesis of amines and amino acids in a 1958 miller H2S-rich spark discharge experiment. Proc. Natl. Acad. Sci. U. S. A. 108 (14), 5526–5531.

Profy, A.T., Usher, D.A., 1984]. Stereoselective aminoacylation of a dinucleoside monophosphate by the imidazolides of DL–alanine and n-(tert-butoxycarbonyl)-DL-alanine. J. Mol. Evol. 20 (2), 147–156.

Root-Bernstein, R., 2007]. Simultaneous origin of homochirality, the genetic code and its directionality. BioEssays 29 (7), 689–698.

Root-Bernstein, R., 2010]. Experimental test of L-and D-amino acid binding to L-and D-codons suggests that homochirality and codon directionality emerged with the genetic code. Symmetry 2 (2), 1180–1200.

Rotkiewicz, P., Skolnick, J., 2008]. Fast procedure for reconstruction of full-atom protein models from reduced representations. J. Comput. Chem. 29 (9), 1460–1465.

Saxinger, C., Ponnamperuma, C., Woese, C., 1971]. Evidence for the interaction of nucleotides with immobilized amino-acids and its significance for the origin of the genetic code. Nature 234 (49), 172–174.

Stewart, I., 1994]. Broken Symmetry in the Genetic Code?

Trifonov, E.N., 2000]. Consensus temporal order of amino acids and evolution of the triplet code. Gene 261 (1), 139–151.

Walker, G.W.R., 1977]. Nucleotide-binding site data and the origin of the genetic code. Biosystems 9 (2), 139–150.

Yarus, M., 1988]. A specific amino acid binding site composed of RNA. Science 240 (4860), 1751–1758, http://dx.doi.org/10.1126/science.3381099 http://science.sciencemag.org/content/240/4860/1751.

Yarus, M., 2000]. RNA-ligand chemistry: a testable source for the genetic code. RNA 6 (04), 475–484.

Yarus, M., Caporaso, J.G., Knight, R., 2005]. Origins of the genetic code: the escaped triplet theory. Annu. Rev. Biochem. 74, 179–198.

Zucchi, C., Caglioti, L., Pályi, G., 1999]. Advances in Biochirality. Elsevier Science, Oxford, ISBN:9780080526621.