Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 401 (2008) 17–26
Theoretical Computer Science
www.elsevier.com/locate/tcs

A relation between trinucleotide comma-free codes and trinucleotide circular codes

Christian J. Michel[a,*], Giuseppe Pirillo[b,c], Mario A. Pirillo[d]

[a] *Equipe de Bioinformatique Théorique, LSIIT (UMR CNRS-ULP 7005), Université Louis Pasteur de Strasbourg, Pôle API, Boulevard Sébastien Brant, 67400 Illkirch, France*
[b] *Consiglio Nazionale delle Ricerche, Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", Unità di Firenze, Dipartimento di Matematica "U.Dini", viale Morgagni 67/A, 50134 Firenze, Italy*
[c] *Université de Marne-la-Vallée, 5 boulevard Descartes, 77454 Marne-la-Vallée Cedex 2, France*
[d] *Istituto Statale SS. Annunziata, Piazzale del Poggio Imperiale, 50134 Firenze, Italy*

Received 20 October 2007; accepted 10 February 2008

Communicated by D. Perrin

Abstract

The comma-free codes and circular codes are two important classes of codes in code theory and in genetics. Fifty years ago before the discovery of the genetic code, a trinucleotide (triletter) comma-free code was proposed for associating the codons of genes with the amino acids of proteins. More recently, in the last ten years, trinucleotide circular codes have been identified statistically in different genomes. Here, we identify a relation between these two classes of trinucleotide codes by constructing a hierarchy of comma-free and circular codes.

Keywords: Comma-free code; Circular code; Trinucleotide; Gene

1. Introduction

The genetic code associates trinucleotides (triletters) over the 4-letter alphabet $\{A, C, G, T\}$ with amino acids (letters) over a 20-letter alphabet. There are 61 trinucleotides among $4^3 = 64$ coding 20 amino acids because the three stop trinucleotides $\{TAA, TAG, TGA\}$ do not code. There are three start trinucleotides $\{ATG, GTG, TTG\}$ where ATG is the standard one that codes the methionine amino acid. These start and stop trinucleotides close a series of nucleotides (letters) in a genome which are translated from three in three nucleotides by the genetic code. This particular series of trinucleotides in a reading frame (also called codons), defines a gene which codes a series of amino acids constituting a protein.

Fifty years ago (in 1957), before the discovery of the genetic code, a class of trinucleotide codes, called comma-free codes (or codes without commas) was proposed by Crick et al. [5] for explaining how the reading of a series

* Corresponding author.
E-mail addresses: michel@dpt-info.u-strasbg.fr (C.J. Michel), pirillo@math.unifi.it (G. Pirillo), map@conmet.it (M.A. Pirillo).

0304-3975/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2008.02.049

of trinucleotides could code amino acids. The two questions of interest were: why are there more trinucleotides than amino acids and, how does one choose the reading frame?

Crick et al. [5] proposed that only 20 trinucleotides among 64 code the 20 amino acids. Such a bijective code implies that the coding trinucleotides are found only in one frame. The determination of a set of 20 trinucleotides forming a comma-free code has several constraints:

(i) A trinucleotide with identical nucleotides must be excluded from such a code. Indeed, the concatenation of *AAA* with itself (for instance) does not allow the (original) reading frame to be retrieved as there are three possible decompositions: $\ldots AAA, AAA, AAA, \ldots$, $\ldots A, AAA, AAA, AA \ldots$ and $\ldots AA, AAA, AAA, A \ldots$, the commas showing the adopted decomposition.

(ii) Two trinucleotides related to circular permutation, for example *AAC* and *ACA*, must also be excluded from such a code. Indeed, the concatenation of *AAC* with itself (for instance) does not allow the reading frame to be retrieved as there are two possible decompositions: $\ldots AAC, AAC, AAC, \ldots$ and $\ldots A, ACA, ACA, AC \ldots$.

Therefore, by excluding the four trinucleotides with identical nucleotides *AAA*, *CCC*, *GGG* and *TTT* and by gathering the 60 remaining trinucleotides in 20 classes of three trinucleotides such that, in each class, three trinucleotides are deduced from each other by circular permutations, e.g., *AAC*, *ACA* and *CAA*, we see that a comma-free code has only one trinucleotide per class and therefore contains at most 20 trinucleotides. This trinucleotide number is identical to the amino acid number, thus leading to a code assigning one trinucleotide per amino acid without ambiguity.

Some basic results on trinucleotide comma-free codes were obtained by Golomb et al. [9,10]. However, no trinucleotide comma-free codes have been identified in genes statistically. Furthermore, in the late fifties, the discovery that the trinucleotide *TTT*, an excluded trinucleotide in a comma-free code, codes phenylalanine [13], led to the abandonment of the concept of a comma-free code over the alphabet $\{A, C, G, T\}$. For several biological reasons, in particular the interaction between mRNA and tRNA, this concept was again taken up over the purine/pyrimidine alphabet $\{R, Y\}$ ($R = \{A, G\}$, $Y = \{C, T\}$) with two trinucleotide comma-free codes for primitive genes: *RRY* [4] and *RNY* ($N = \{R, Y\}$) [6].

Back in 1996, a statistical study of trinucleotide occurrences per frame has identified a set X (*EUK_PRO*) of 20 trinucleotides in the gene populations of both eukaryotes *EUK* and prokaryotes *PRO* [1]. This set is a trinucleotide circular code with several strong biomathematical properties. A circular code is a set of words over an alphabet such that any word written on a circle has at most one decomposition into words of the circular code [11]. The construction frame of a word generated by any concatenation of words of a circular code can be retrieved after the reading, anywhere in the generated word, of a certain number of nucleotides depending on the code. This series of nucleotides is called the window of the circular code. The minimal window length is the size of the longest ambiguous word that can be read in at least two frames, added with one letter. Similar to the existence of variant genetic codes (compared to the universal one), several trinucleotide circular codes have been found in genes: one code X (*MIT*) in mitochondria [2], 15 codes X ($G_{archaea}$) in archaeal genomes [7] and 72 codes X ($G_{bacteria}$) in 175 complete bacterial genomes (with several bacterial genomes having the same codes) [8].

A circular code has weaker conditions compared to a comma-free code. In particular, some trinucleotides of a circular code can be found in the non-reading frame, i.e., in the two shifted frames (the reading frame shifted by one and two nucleotides in the $5'$–$3'$ direction), while the 20 trinucleotides of a comma-free code are found only in the reading frame. On the other hand, the lengths of the minimal windows of a circular code and a comma-free code are less than or equal to 13 and four nucleotides, respectively. A comma-free code in genes is too constrained from an evolutionary point of view compared to a circular code. Perhaps, this is the reason why it is not observed in current genes statistically. From a code theory point of view, these two classes of codes are analysed separately. Here, we present several results leading to the identification of a relation between these two classes of trinucleotide codes by constructing a hierarchy of codes that are closed by the comma-free and circular codes. More precisely, all the trinucleotide codes in this hierarchy are circular, the strongest ones being comma-free.

2. Definitions

For the classical notions of *an alphabet*, *empty word*, *length*, *factor*, *proper factor*, *prefix*, *proper prefix*, *suffix*, *proper suffix*, we refer the reader to [3]. Let $\mathcal{A}$ denote a finite alphabet and let $\mathcal{A}^*$ denote the set of all words over $\mathcal{A}$. Given a subset X of $\mathcal{A}^*$, X^n is the set of the words over $\mathcal{A}$ which is the product of n words from X, i.e., $X^n = \{x_1 x_2 \cdots x_n \mid x_i \in X\}$.

There is a correspondence between the genetic and language-theoretic concepts. The *letters* (or *nucleotides* or *bases*) define the genetic alphabet $\mathcal{A}_4 = \{A, C, G, T\}$. The set of *non-empty words* (resp. *words*) over $\mathcal{A}_4$ is denoted by $\mathcal{A}_4^+$ (resp. $\mathcal{A}_4^*$). The set of the 16 words of length two (or *dinucleotides* or *diletters*) is denoted by $\mathcal{A}_4^2$. The set of the 64 words of length three (or *trinucleotides* or *triletters*) is denoted by $\mathcal{A}_4^3$. The *total order* over the alphabet $\mathcal{A}_4$ is $A < C < G < T$. Consequently, $\mathcal{A}_4^+$ is *alphabetically ordered*: given two words $u, v \in \mathcal{A}_4^+$, *u is smaller than* v in *alphabetical order*, written $u < v$, if and only if either u is a proper prefix of v or there exist $x, y \in \mathcal{A}_4$, $x < y$, and $r, s, t \in \mathcal{A}_4^*$ such that $u = rxs$ and $v = ryt$.

2.1. Two genetic maps

Definition 1. The complementarity map $\mathcal{C}$: $\mathcal{A}_4^+ \to \mathcal{A}_4^+$ is defined by $\mathcal{C}(A) = T, \mathcal{C}(T) = A, \mathcal{C}(C) = G$ and $\mathcal{C}(G) = C$ and by $\mathcal{C}(uv) = \mathcal{C}(v)\mathcal{C}(u)$ for all $u, v \in \mathcal{A}_4^+$, e.g., $\mathcal{C}(AAC) = GTT$. This map $\mathcal{C}$ is associated to the property of the complementary and antiparallel (one DNA strand chemically oriented in a 5′–3′ direction and the other DNA strand, in the opposite 3′–5′ direction) double helix. This map on words is naturally extended to word sets: a complementary trinucleotide set is obtained by applying the complementarity map $\mathcal{C}$ to all its trinucleotides.

Definition 2. The circular *permutation map* $\mathcal{P}$: $\mathcal{A}_4^3 \to \mathcal{A}_4^3$ permutes circularly each trinucleotide $l_1l_2l_3$ as follows $\mathcal{P}(l_1l_2l_3) = l_2l_3l_1$. The kth iterate of $\mathcal{P}$ is denoted $\mathcal{P}^k$. This map on words is also naturally extended to word sets: a permuted trinucleotide set is obtained by applying the circular *permutation map* $\mathcal{P}$ to all its trinucleotides.

Remark 1. Two trinucleotides u and v are *conjugate* if there exist two words s and t such that $u = st$ and $v = ts$. Therefore, if u and v satisfy $\mathcal{P}^k(u) = v$ for some k, then u and v are conjugate.

2.2. Codes, trinucleotide comma-free codes and trinucleotide circular codes

The notion of a code has very different meanings in biology and language theory. In biology, the "genetic code" associates trinucleotides with amino acids, while in language theory a "code" is a set of words with a unique decipherability condition.

Definition 3. Code: A set X of words is a code if, for each $x_1, \ldots, x_n, x'_1, \ldots, x'_m \in X$, $n, m \geq 1$, the condition $x_1 \cdots x_n = x'_1 \cdots x'_m$ implies $n = m$ and $x_i = x'_i$ for $i = 1, \ldots, n$.

The set $\mathcal{A}_4^3$ itself is a code. More precisely, it is a *uniform code* [3]. Consequently, any non-empty subset of $\mathcal{A}_4^3$ is a code called *trinucleotide codes* in this paper.

Definition 4. Trinucleotide comma-free code: A trinucleotide code X is comma-free if, for each $y \in X$ and $u, v \in \mathcal{A}_4^*$ such that $uyv = x_1 \cdots x_n$ with $x_1, \ldots, x_n \in X$, $n \geq 1$, it holds that $u, v \in X^*$.

Several varieties of trinucleotide comma-free codes were described in [12].

Definition 5. Trinucleotide circular code: A trinucleotide code X is circular if, for each $x_1, \ldots, x_n, x'_1, \ldots, x'_m \in X$, $n, m \geq 1$, $p \in \mathcal{A}_4^*$, $s \in \mathcal{A}_4^+$, the conditions $sx_2 \cdots x_n p = x'_1 \cdots x'_m$ and $x_1 = ps$ imply $n = m$, $p = \varepsilon$ (empty word) and $x_i = x'_i$ for $i = 1, \ldots, n$.

Remark 2. $\mathcal{A}_4^3$ is obviously not a circular code and even less a comma-free code (see also Propositions 1 and 2).

Definition 6. Self-complementary code: A trinucleotide code X is self-complementary if, for each $y \in X, \mathcal{C}(y) \in X$.

Definition 7. C^3 self-complementary code: A trinucleotide code X is C^3 self-complementary if X, $X_1 = \mathcal{P}(X)$ and $X_2 = \mathcal{P}^2(X)$ are codes satisfying the following properties: $X = \mathcal{C}(X)$ (self-complementary), $\mathcal{C}(X_1) = X_2$ and $\mathcal{C}(X_2) = X_1$.

Example 1. The set X (*EUK_PRO*) of 20 trinucleotides identified in the gene populations of both eukaryotes *EUK* and prokaryotes *PRO*, i.e., X (*EUK_PRO*) = {*AAC*, *AAT*, *ACC*, *ATC*, *ATT*, *CAG*, *CTC*, *CTG*, *GAA*, *GAC*, *GAG*, *GAT*, *GCC*, *GGC*, *GGT*, *GTA*, *GTC*, *GTT*, *TAC*, *TTC*} is a maximal (20 words on $\mathcal{A}_4^3$) C^3 self-complementary circular code [1].

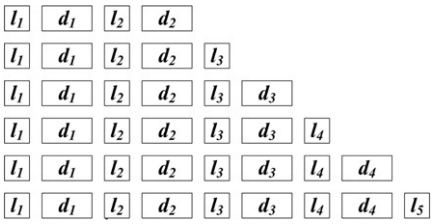

Fig. 1. A graphical representation of the regularities given in the Letter Diletter Necklaces (*LDN*) and Letter Diletter Continued Necklaces (*LDCN*) definitions.

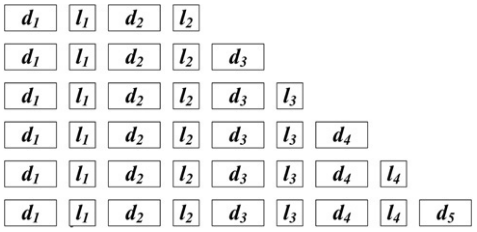

Fig. 2. A graphical representation of the regularities given in the Diletter Letter Necklaces (*DLN*) and Diletter Letter Continued Necklaces (*DLCN*) definitions.

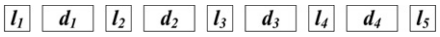

Fig. 3. A graphical representation of the 5*LDCN* definition.

2.3. Necklace concept

The concept of a necklace was introduced by Pirillo for circular codes in [14] and was used for studying the self-complementary circular codes and the trinucleotide comma-free codes in [16] and [12] respectively. Here, we extend this concept to both the trinucleotide comma-free and the circular codes with the notions of a Letter Diletter (Continued) Necklace (*LDN*, *LDCN*) and a Diletter Letter (Continued) Necklace (*DLN*, *DLCN*).

In the following definitions, $l_1, l_2, \ldots, l_{n-1}, l_n, \ldots$ are letters in $\mathcal{A}_4$, $d_1, d_2, \ldots, d_{n-1}, d_n, \ldots$ are diletters in $\mathcal{A}_4^2$ and n is an integer satisfying $n \geq 2$.

Definition 8. Letter Diletter Necklaces (*LDN*): We say that the ordered sequence $l_1, d_1, l_2, d_2, \ldots, d_{n-1}, l_n, d_n$ is an *nLDN* for a subset $X \subset \mathcal{A}_4^3$ if $l_1d_1, l_2d_2, \ldots, l_nd_n \in X$ and $d_1l_2, d_2l_3, \ldots, d_{n-1}l_n \in X$.

Definition 9. Letter Diletter Continued Necklaces (*LDCN*): We say that the ordered sequence $l_1, d_1, l_2, d_2, \ldots, d_{n-1}, l_n, d_n, l_{n+1}$ is an $(n+1)$*LDCN* for a subset $X \subset \mathcal{A}_4^3$ if $l_1d_1, l_2d_2, \ldots, l_nd_n \in X$ and $d_1l_2, d_2l_3, \ldots, d_{n-1}l_n, d_nl_{n+1} \in X$.

Definition 10. Diletter Letter Necklaces (*DLN*): We say that the ordered sequence $d_1, l_1, d_2, l_2, \ldots, l_{n-1}, d_n, l_n$ is an *nDLN* for a subset $X \subset \mathcal{A}_4^3$ if $d_1l_1, d_2l_2, \ldots, d_nl_n \in X$ and $l_1d_2, l_2d_3, \ldots, l_{n-1}d_n \in X$.

Definition 11. Diletter Letter Continued Necklaces (*DLCN*): We say that the ordered sequence $d_1, l_1, d_2, l_2, \ldots, l_{n-1}, d_n, l_n, d_{n+1}$ is an $(n+1)$*DLCN* for a subset $X \subset \mathcal{A}_4^3$ if $d_1l_1, d_2l_2, \ldots, d_nl_n \in X$ and $l_1d_2, l_2d_3, \ldots, l_{n-1}d_n, l_nd_{n+1} \in X$.

Figs. 1 and 2 give a graphical representation of the regularities in the *LDN*, *LDCN*, *DLN* and *DLCN* definitions. In particular, they show the forbidden *LD* and *DL* configurations for the codes in our hierarchy (see Definitions 8–11).

Fig. 3 gives a graphical representation of the 5*LDCN* definition. If a code X admits a 5*LDCN* then for some i, j, $1 \leq i \leq j \leq 5$, $l_i = l_j$. If $j - i = 4$ then $l_1 = l_5$ and this configuration (written on a circle in Fig. 4) is impossible for a circular code. If $j - i = 1$ (or $j - i = 2$ or $j - i = 3$ or $j - i = 4$) then there are similar configurations which are forbidden for a circular code.

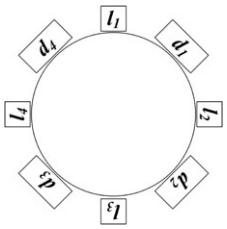

Fig. 4. The 5*LDCN* forbidden configuration for a circular code.

Proposition 1 (*[14]*). *Let X be a trinucleotide code. The following conditions are equivalent.*

(i) *X is circular code.*
(ii) *X has no 5LDCN.*

Proposition 1 of [12] gives several necessary and sufficient conditions for a trinucleotide code to be a comma-free code. In particular, the following equivalence, useful in this paper, holds.

Proposition 2 (*[12]*). *Let X be a trinucleotide code. The following conditions are equivalent.*

(i) *X is a comma-free code.*
(ii) *X has no 2LDN and no 2DLN.*

Remark 3. A circular code can have a 2*LDN*, 3*LDN*, 4*LDN*, 2*DLN*, 3*DLN* and 4*DLN* but, by Proposition 1, it has no 5*LDCN*. Given a non-circular code, by Proposition 1, for any k there exists $n \geq k$ such that X admits an *nLDN* and an *nDLN*. By Proposition 2, a comma-free code has no 2*LDN* and no 2*DLN*. A non-comma-free code must have either a 2*LDN* or a 2*DLN*.

3. A hierarchy of trinucleotide comma-free and circular codes

A hierarchy of variable length codes is presented in [15]. We propose here a hierarchy specific for trinucleotide codes. In this proposed hierarchy, all the codes are trinucleotide circular codes and the constraints ones are trinucleotide comma-free codes.

Definition 12. Let X be a trinucleotide code. For $k \in \{2, 3, 4, 5\}$, we say that X belongs to the class C^{kLDN} if X has no *kLDN* and that X belongs to the class C^{kDLN} if X has no *kDLN*. Similarly, for $k \in \{3, 4, 5\}$, we say that X belongs to the class C^{kLDCN} if X has no *kLDCN* and that X belongs to the class C^{kDLCN} if X has no *kDLCN*.

Notation 1. $I^n = C^{nLDN} \cap C^{nDLN}$, $I^nC = C^{nLDCN} \cap C^{nDLCN}$, $U^n = C^{nLDN} \cup C^{nDLN}$, $U^nC = C^{nLDCN} \cup C^{nDLCN}$.

Proposition 3. *The following chains of inclusions hold.*

(i) $C^{2LDN} \subset C^{3LDCN} \subset C^{3LDN} \subset C^{4LDCN} \subset C^{4LDN} \subset C^{5LDCN} \subset C^{5LDN}$.
(ii) $C^{2DLN} \subset C^{3DLCN} \subset C^{3DLN} \subset C^{4DLCN} \subset C^{4DLN} \subset C^{5DLCN} \subset C^{5DLN}$.
(iii) $C^{2LDN} \subset C^{3DLCN} \subset C^{3LDN} \subset C^{4DLCN} \subset C^{4LDN} \subset C^{5DLCN} \subset C^{5LDN}$.
(iv) $C^{2DLN} \subset C^{3LDCN} \subset C^{3DLN} \subset C^{4LDCN} \subset C^{4DLN} \subset C^{5LDCN} \subset C^{5DLN}$.
(v) $I^2 \subset I^3C \subset I^3 \subset I^4C \subset I^4 \subset I^5C \subset I^5$.
(vi) $U^2 \subset U^3C \subset U^3 \subset U^4C \subset U^4 \subset U^5C \subset U^5$.

Proof. (i) We first prove that $C^{2LDN} \subset C^{3LDCN}$. By way of contradiction, suppose that $X \in C^{2LDN}$ but $X \notin C^{3LDCN}$, i.e., X has a 3*LDCN*: l_1, d_1, l_2, d_2, l_3. Immediately, l_1, d_1, l_2, d_2 is a 2*LDN* for X, i.e., $X \notin C^{2LDN}$, a contradiction. The inclusions $C^{3LDCN} \subset C^{3LDN}$, $C^{3LDN} \subset C^{4LDCN}$, etc., are proved similarly.

The proofs of (ii)–(iv) are similar to (i). Moreover, (v) and (vi) follow from (i) and (ii), respectively. □

Proposition 4. $C^{5LDN} = C^{5LDCN} = C^{5DLN}$.

Proof. We first prove that $C^{5LDN} = C^{5LDCN}$. By (i) of Proposition 3 , $C^{5LDCN} \subset C^{5LDN}$. It remains to prove that $C^{5LDN} \subset C^{5LDCN}$. By way of contradiction, suppose $X \in C^{5LDN}$ but $X \notin C^{5LDCN}$. Let $l_1, d_1, l_2, d_2, l_3, d_3, l_4, d_4, l_5$

be a $5LDCN$ for X. As $\mathcal{A}_4$ contains four letters, then for some $i, j \in \{1, 2, 3, 4, 5\}$ with $i < j$, $l_i = l_j$. So $l_i, d_i, \ldots, l_{j-1}, d_{j-1}, l_i$ is a $(j-i+1)LDCN$ for X having l_i as its first and last letter. Using this property, an $nLDCN$ with an arbitrary n can be constructed for X. A suitable "prefix" of one of them is a $5LDN$ for X, a contradiction.

The fact that $C^{5DLN} = C^{5LDCN}$ is proved similarly using (iv) of Proposition 3 and considering a "factor" of a suitable $nLDCN$ that is a $5LDN$ for X and begins with d_i. □

Remark 4. As, by Proposition 1, C^{5LDCN} is the class of circular codes, Proposition 4 shows that all the chains of inclusions of Proposition 3 end with the class of circular codes. The chain of inclusions in (v) of Proposition 3 begins with I^2 which is exactly the class of comma-free codes.

On the other hand, $C^{5DLCN} \neq C^{5LDCN}$. More precisely, there is the following inclusion.

Proposition 5. *$C^{5DLCN} \subset C^{5LDCN}$ with $C^{5DLCN} \neq C^{5LDCN}$.*

Proof. By Propositions 3 and 4, the codes in C^{5DLCN} are circular. By Proposition 1, C^{5LDCN} is the class of circular codes. So, $C^{5DLCN} \subset C^{5LDCN}$.

Now, consider the following set Y of diletters and letters $d_1 = AC, l_1 = A, d_2 = AG, l_2 = C, d_3 = AT, l_3 = G, d_4 = CG, l_4 = T, d_5 = CT$ and the following code $X = \{ACA, AAG, AGC, CAT, ATG, GCG, CGT, TCT\}$. It is circular as it is in the class C^{5LDCN} but, by construction, it is not in the class C^{5DLCN}. So the inclusion $C^{5DLCN} \subset C^{5LDCN}$ is strict. □

Remark 5. We will see in the next section that the inclusion $C^{5DLCN} \subset C^{5LDCN}$ remains strict in the class of the 528 maximal self-complementary circular codes and in the case of the 216 maximal C^3 self-complementary circular codes.

Remark 6. By Proposition 4, $C^{5LDN} = C^{5DLN}$. The strict inclusion $C^{5DLCN} \subset C^{5LDCN}$ holds too. The first level in which some codes can be in the class LD but not in the class DL and vice-versa, is $4LDN - 4DLN$. Indeed, the code $X' = \{ACA, AAG, AGC, CAT, ATG, GCG, CGT\}$ obtained from X (used in the proof of Proposition 5) by suppressing TCT, is in the class C^{4LDN} but, by construction, X' is not in the class C^{4DLN}. On the other hand, the code $X_1 = \{AAC, ACC, CAG, AGG, GAT, ATT, TCG\}$ built using the diletters and letters $l_1 = A, d_1 = AC, l_2 = C, d_2 = AG, l_3 = G, d_3 = AT, l_4 = T, d_4 = CG$ is in the class C^{4DLN} but, by construction, X_1 is not in the class C^{4LDN}.

The hierarchies in Proposition 3 concern the class of trinucleotide circular codes. The following proposition explains some symmetries of hierarchies within the 528 maximal self-complementary codes and the 216 maximal C^3 self-complementary codes.

Proposition 6. *For the class of the* 528 *maximal self-complementary circular codes, the following equalities hold.*

(i) $C^{2LDN} = C^{2DLN}$.
(ii) $C^{3LDN} = C^{3DLN}$.
(iii) $C^{4LDN} = C^{4DLN}$.
(iv) $C^{5LDN} = C^{5DLN}$.

Proof. (i) Let $X \subset \mathcal{A}_4^3$ be a maximal self-complementary code. We have to prove that X is in the class C^{2LDN} if and only if X is in the class C^{2DLN}. By way of contradiction, suppose that $X \in C^{2LDN}$ and $X \notin C^{2DLN}$. Then, X has a $2DLN$ denoted by d_1, l_1, d_2, l_2. Consider the sequence $\mathcal{C}(l_2), \mathcal{C}(d_2), \mathcal{C}(l_1), \mathcal{C}(d_1)$. By the self-complementary property of X, $\mathcal{C}(l_2)\mathcal{C}(d_2) \in X$ as $d_2 l_2 \in X$, $\mathcal{C}(l_1)\mathcal{C}(d_1) \in X$ as $d_1 l_1 \in X$ and $\mathcal{C}(d_2)\mathcal{C}(l_1) \in X$ as $l_1 d_2 \in X$, i.e., X has a $2LDN$, a contradiction. So X is in the class C^{2DLN}. In a similar way, we prove that if X is in the class C^{2DLN} then X is also in the class C^{2LDN}. The proofs of (ii)–(iv) are similar to (i). □

4. Computer results

We consider the following partition of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT, ATA, TAT, CGC, GCG\}$ into 28 self-complementary pairs (Table 1). The first element of each pair is the smallest in alphabetical order and the 28 pairs are ordered according to the alphabetical order of their first components. Finally, we denote them by the following symbols $\{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, z', z''\}$ which are the letters of the English alphabet with the two additional symbols z' and z''.

Table 1
Partition of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT, ATA, TAT, CGC, GCG\}$ into 28 self-complementary pairs

$a = \{AAC, GTT\}$	$b = \{AAG, CTT\}$	$c = \{AAT, ATT\}$	$d = \{ACA, TGT\}$
$e = \{ACC, GGT\}$	$f = \{ACG, CGT\}$	$g = \{ACT, AGT\}$	$h = \{AGA, TCT\}$
$i = \{AGC, GCT\}$	$j = \{AGG, CCT\}$	$k = \{ATC, GAT\}$	$l = \{ATG, CAT\}$
$m = \{CAA, TTG\}$	$n = \{CAC, GTG\}$	$o = \{CAG, CTG\}$	$p = \{CCA, TGG\}$
$q = \{CCG, CGG\}$	$r = \{CGA, TCG\}$	$s = \{CTA, TAG\}$	$t = \{CTC, GAG\}$
$u = \{GAA, TTC\}$	$v = \{GAC, GTC\}$	$w = \{GCA, TGC\}$	$x = \{GCC, GGC\}$
$y = \{GGA, TCC\}$	$z = \{GTA, TAC\}$	$z' = \{TAA, TTA\}$	$z'' = \{TCA, TGA\}$

Table 2
Hierarchy of the 528 maximal self-complementary codes

C^{2LDN}	C^{3LDCN}	C^{3LDN}	C^{4LDCN}	C^{4LDN}	C^{5LDCN}	C^{5LDN}
0	96	96	96	$64 + 96$	$368 + 64 + 96$	$368 + 64 + 96$
C^{2DLN}	C^{3DLCN}	C^{3DLN}	C^{4DLCN}	C^{4DLN}	C^{5DLCN}	C^{5DLN}
0	0	96	$64 + 96$	$64 + 96$	$64 + 96$	$368 + 64 + 96$
I^2	I^3C	I^3	I^4C	I^4	I^5C	I^5
0	0	96	96	$64 + 96$	$64 + 96$	$368 + 64 + 96$
U^2	U^3C	U^3	U^4C	U^4	U^5C	U^5
0	96	96	$64 + 96$	$64 + 96$	$368 + 64 + 96$	$368 + 64 + 96$

The number of codes is given in each class.

Table 3
Hierarchy of the 216 maximal C^3 self-complementary codes

C^{2LDN}	C^{3LDCN}	C^{3LDN}	C^{4LDCN}	C^{4LDN}	C^{5LDCN}	C^{5LDN}
0	56	56	56	$56 + 56$	$104 + 56 + 56$	$104 + 56 + 56$
C^{2DLN}	C^{3DLCN}	C^{3DLN}	C^{4DLCN}	C^{4DLN}	C^{5DLCN}	C^{5DLN}
0	0	56	$56 + 56$	$56 + 56$	$56 + 56$	$104 + 56 + 56$
I^2	I^3C	I^3	I^4C	I^4	I^5C	I^5
0	0	56	56	$56 + 56$	$56 + 56$	$104 + 56 + 56$
U^2	U^3C	U^3	U^4C	U^4	U^5C	U^5
0	56	56	$56 + 56$	$56 + 56$	$104 + 56 + 56$	$104 + 56 + 56$

The number of codes is given in each class.

Table 4a
List of the 56 maximal C^3 self-complementary codes of the class C^{3LDCN} (class C_1)

abcegiktvx	*abcfgjlnoq*	*abcflnoqst*	*abclnoqstv*	*abciknvxz*	*abcknotvxz*
aceghijkvx	*aceghikvxy*	*acfghijlnq*	*achiknvxyz*	*acegiktuvx*	*aciknvxyz*
acknotuvxz	*anuvwxyzz′z″*	*aknvxyzz′*	*aknotuvxzz′*	*bcdefgiktx*	*bcdefgjloq*
bcdfgjlopq	*bcdflopqst*	*cdefghijkx*	*cdefghijqz″*	*cdefghijxz″*	*cdefghijlq*
cdefghikxy	*cdefghixyz″*	*cdfghijpqz″*	*cdfghijlpq*	*deghrwxyz′z″*	*dehrwxyzz′z″*
dghjpqrwz′z″	*dghpqrwyz′z″*	*dghprwxyz′z″*	*dhjpqrswz′z″*	*dhpqrswyz′z″*	*dhprwxyzz′z″*
deuvwxyzz′z″	*dektuvwxzz′*	*dpqrstuwz′z″*	*dpuvwxyzz′z″*	*bcfgjlmnoq*	*bcflmnoqst*
bclmnoqstv	*bmopqrstz′z″*	*blmnoqrstz′*	*blmnoqstvz′*	*hjmopqrsz′z″*	*hmopqrsyz′z″*
hjlmnoqrsz′	*hmnrwxyzz′z″*	*mopqrstuz′z″*	*lmnoqrstuz′*	*lmnoqstuvz′*	*mnuvwxyzz′z″*
kmntuvwxzz′	*kmnotuvxzz′*				

In this section, we present the computer results for the 528 maximal self-complementary codes and the 216 maximal C^3 self-complementary codes with a code classification according to the hierarchies of Proposition 3 (Tables 2 and 3). For both hierarchies, the classes C^{2LDN} and C^{2DLN} are empty. Moreover, the cardinalities of the classes from C^{3LDN} to C^{5LDN} and also from C^{3DLN} to C^{5DLN} are increasing. The class C^{3LDCN} is the first one that is non-empty. New codes occur in the class C^{4DLCN}, and then in the class C^{5LDCN}. According to Proposition 1, the class C^{5LDCN} contains all codes. But, for the *DL* codes, it is the class C^{5DLN} that contains all codes and not the class C^{5DLCN}.

Table 4b
List of the 40 maximal non-C^3 self-complementary codes of the class C^{3LDCN} (class C_2)

abcegikvxy	*abcfgjlopq*	*abcflopqst*	*abciknvxyz*	*abcknovxyz*	*abclopqstv*
acegijkuvx	*acfghijlpq*	*acfghijpqz″*	*achijknvxz*	*acijknuvxz*	*ahnrwxyzz′z″*
ahprwxyzz′z″	*akptuvwxzz′*	*apuvwxyzz′z″*	*bcdefgikxy*	*bcdefgixyz″*	*bcdefloqst*
bcefgjlmoq	*bceflmoqst*	*bdpqrstwz′z″*	*bdpqrswyz′z″*	*blmnoqrsyz′*	*bmopqrsyz′z″*
cdefgijkux	*cdefgiktux*	*cefghijlmq*	*cfghijlmnq*	*dgjpqruwz′z″*	*djpqrsuwz′z″*
dkptuvwxzz′	*eghmrwxyz′z″*	*ehmrwxyzz′z″*	*ekmotuvxzz′*	*ekmtuvwxzz′*	*emuvwxyzz′z″*
hlmnoqrsyz′	*jlmnoqrsuz′*	*jlmnoqsuvz′*	*jmopqrsuz′z″*		

Table 5a
List of the 56 maximal C^3 self-complementary codes of the class C^{4DLCN} that are not already in the class C^{3LDCN} (class C_3)

abcefgijkx	*abcefgijlq*	*abcefgiktx*	*abcefgjloq*	*abcefgloqt*	*abcegijkvx*
abcegkotvx	*abcfgijlnq*	*abcfglnoqt*	*abcgijknvx*	*abcgiknvtx*	*abcgjlnoqv*
abcgknotvx	*abcglnoqtv*	*aceghivxyz″*	*acehikvxyz*	*aehvwxyzz′z″*	*acegikuvxy*
aceikuvxyz	*aekuvwxyzz′*	*aeuvwxyzz′z″*	*bcdfgjopqz″*	*bcdfjlopqs*	*bdjopqrsz′z″*
cdeghikvxy	*cdeghivxyz″*	*cdfghjopqz″*	*cdfghjlopq*	*deghvwxyz′z″*	*dehvwxyzz′z″*
dghjopqrz′z″	*dhjopqrsz′z″*	*cdegikuvxy*	*deguvwxyz′z″*	*dekuvwxyzz′*	*bcfgjlmopq*
bcfjlmopqs	*bjmopqrsz′z″*	*bjlmopqrsz′*	*cfghjlmopq*	*ghjmopqrz′z″*	*hjlmopqrsz′*
mpqrsuwyz′z″	*mpqrstuwz′z″*	*mopqrsuyz′z″*	*mpruwxyzz′z″*	*mopqstuvz′z″*	*mpuvwxyzz′z″*
mptuvwxzz′z″	*mnoqrsuyz′z″*	*mnoqrstuz′z″*	*mnruwxyzz′z″*	*mnoqstuvz′z″*	*mntuvwxzz′z″*
mnouvxyzz′z″	*mnotuvxzz′z″*				

Table 5b
List of the eight maximal non-C^3 self-complementary codes of the class C^{4DLCN} that are not already in the class C^{3LDCN} (class C_4)

abcefgkotx	*abcgijlnqv*	*acehivxyzz″*	*bcdfjopqsz″*	*degkuvwxyz′*	*ghjlmopqrz′*
mnoruxyzz′z″	*mpqstuvwz′z″*				

Table 6a
List of the 104 maximal C^3 self-complementary codes of the class C^{5LDCN} that are not already in the class C^{4DLCN} (class C_5)

abcefgijqz″	*abcefgijxz″*	*abceiktvxz*	*abcekotvxz*	*abcfjlnoqs*	*abcjlnoqsv*
ablnoqrstz′	*ablnoqstvz′*	*abkntvwxzz′*	*abknotvxzz′*	*acefghijkx*	*acefghijqz″*
acefghijxz″	*acefghijlq*	*acefghikxy*	*acefghixyz″*	*aceghijvxz″*	*acfghjlnoq*
acghiknvxy	*acegkotuvx*	*aceiktuvxz*	*acekouvxyz*	*acekotuvxz*	*acgikntuvx*
acgknotuvx	*aciknuvxyz*	*acknouvxyz*	*aektuvwxzz′*	*aekouvxyzz′*	*aekotuvxzz′*
aknuvwxyzz′	*antuvwxzz′z″*	*aknouvxyzz′*	*anouvxyzz′z″*	*anotuvxzz′z″*	*bcdefgijkx*
bcdefgijqz″	*bcdefgijxz″*	*bcdefgijlq*	*bcdefgjoqz″*	*bcdegiktvx*	*bcdfgijpqz″*
bcdfgijlpq	*bcdfglopqt*	*cdefghiqyz″*	*cdefghilqy*	*cdefghjoqz″*	*cdefghjloq*
cdeghijkvx	*cdeghijvxz″*	*cdfghijkpx*	*cdfghijpxz″*	*deghqrwyz′z″*	*dehqrswyz′z″*
dghjprwxz′z″	*dghopqryz′z″*	*dhopqrsyz′z″*	*dhjprwxzz′z″*	*dghpvwxyz′z″*	*dhpvwxyzz′z″*
degruwxyz′z″	*deruwxyzz′z″*	*detuvwxzz′z″*	*dgpqruwyz′z″*	*dgpruwxyz′z″*	*dpqrsuwyz′z″*
dopqrstuz′z″	*dpruwxyzz′z″*	*dgpuvwxyz′z″*	*bcflmopqst*	*bcjlmopqsv*	*bclmopqstv*
bcfglmnoqt	*bcfjlmnoqs*	*bcgjlmnoqv*	*bcglmnoqtv*	*bcjlmnoqsv*	*blmopqrstz′*
bjlmopqsvz′	*bmopqstvz′z″*	*blmopqstvz′*	*bjlmnoqrsz′*	*bmnoqrstz′z″*	*bjlmnoqsvz′*
bmnoqstvz′z″	*ghjmpqrwz′z″*	*ghmpqrwyz′z″*	*ghmprwxyz′z″*	*ghmopqryz′z″*	*hjmpqrswz′z″*
hmpqrswyz′z″	*hmprwxyzz′z″*	*hjmnoqrsz′z″*	*hmnvwxyzz′z″*	*cflmnoqstu*	*clmnoqstuv*
cikmntuvxz	*ckmnotuvxz*	*gmpqruwyz′z″*	*gmpruwxyz′z″*	*lmopqrstuz′*	*lmopqstuvz′*
kmnuvwxyzz′	*kmnouvxyzz′*				

Tables 4a, 4b, 5a, 5b, 6a and 6b list the 528 maximal self-complementary codes organized according to their occurrences in the hierarchy. Precisely, we define a partition of these 528 codes into six classes C_1, C_2 , C_3, C_4, C_5 and C_6.

In a compact way, by using the above partition of the 528 maximal self-complementary codes, the following hierarchies are observed (Table 7).

Table 6b
List of the 264 maximal non-C^3 self-complementary codes of the class C^{5LDCN} that are not already in the class C^{4DLCN} (class C_6)

abcefgikxy	*abcefgixyz″*	*abcefgkoxy*	*abcefloqst*	*abcegivxyz″*	*abcegkovxy*
abceikvxyz	*abceivxyzz″*	*abcekovxyz*	*abcfgijlpq*	*abcfgijpqz″*	*abcfgjopqz″*
abcfglnoqy	*abcfglopqt*	*abcfjlopqs*	*abcfjopqsz″*	*abcflnoqsy*	*abcfopqstz″*
abcgijlpqv	*abcgiknvxy*	*abcgikptvx*	*abcgjlopqv*	*abcgknovxy*	*abcglopqtv*
abcijknvxz	*abcikptvxz*	*abcinvxyzz″*	*abcjlopqsv*	*abkptvwxzz′*	*ablnoqrsyz′*
abnoqrstz′z″	*abnoqrsyz′z″*	*abnoqstvz′z″*	*abnorxyzz′z″*	*abnotvxzz′z″*	*abnovxyzz′z″*
abnrwxyzz′z″	*abntvwxzz′z″*	*abnvwxyzz′z″*	*abopqrstz′z″*	*abopqrsyz′z″*	*abopqstvz′z″*
abpqrstwz′z″	*abpqrswyz′z″*	*abpqstvwz′z″*	*abprwxyzz′z″*	*abptvwxzz′z″*	*abpvwxyzz′z″*
acefgijkux	*acefgiktux*	*acefgkotux*	*acehijkvxz*	*acehijvxzz″*	*aceijkuvxz*
acfghilnqy	*acfghjlopq*	*acfghjopqz″*	*acfghlnoqy*	*acghijknvx*	*acghijkpvx*
acghijlnqv	*acghijlpqv*	*acghijpvxz″*	*acgijknuvx*	*acgijkpuvx*	*acgikptuvx*
achijkpvxz	*achijnvxzz″*	*achijpvxzz″*	*achinvxyzz″*	*acijkpuvxz*	*acikptuvxz*
aehrwxyzz′z″	*ahjlnoqrsz′*	*ahjnoqrsz′z″*	*ahjnrwxzz′z″*	*ahjnvwxzz′z″*	*ahjopqrsz′z″*
ahjpqrswz′z″	*ahjprwxzz′z″*	*ahjpvwxzz′z″*	*ahlnoqrsyz′*	*ahnoqrsyz′z″*	*ahnorxyzz′z″*
ahnvwxyzz′z″	*ahopqrsyz′z″*	*ahpqrswyz′z″*	*ahpvwxyzz′z″*	*ajknuvwxzz′*	*ajkpuvwxzz′*
ajnuvwxzz′z″	*ajpuvwxzz′z″*	*akpuvwxyzz′*	*anoruxyzz′z″*	*anruwxyzz′z″*	*apruwxyzz′z″*
aptuvwxzz′z″	*bcdefgkotx*	*bcdefgkoxy*	*bcdefgloqt*	*bcdefgloqy*	*bcdefgoqyz″*
bcdefjloqs	*bcdefjoqsz″*	*bcdefloqsy*	*bcdefoqstz″*	*bcdefoqsyz″*	*bcdegikvxy*
bcdegivxyz″	*bcdfgikptx*	*bcdfopqstz″*	*bcdgikptvx*	*bcefgijlmq*	*bcefglmoqt*
bcefglmoqy	*bcefjlmoqs*	*bceflmoqsy*	*bcfgijlmnq*	*bcfglmnoqy*	*bcflmnoqsy*
bcgijlmnqv	*bdektvwxzz′*	*bdeoqrstz′z″*	*bdeoqrsyz′z″*	*bdeqrstwz′z″*	*bdeqrswyz′z″*
bderwxyzz′z″	*bdetvwxzz′z″*	*bdevwxyzz′z″*	*bdjpqrswz′z″*	*bdkptvwxzz′*	*bdopqrstz′z″*
bdopqrsyz′z″	*bdpqstvwz′z″*	*bdprwxyzz′z″*	*bdptvwxzz′z″*	*bdpvwxyzz′z″*	*belmoqrstz′*
belmoqrsyz′	*bemoqrstz′z″*	*bemoqrsyz′z″*	*blmopqrsyz′*	*bmnoqrsyz′z″*	*bmpqrstwz′z″*
bmpqrswyz′z″	*bmpqstvwz′z″*	*cdefghloqy*	*cdefghoqyz″*	*cdefgijlqu*	*cdefgikuxy*
cdefgjloqu	*cdefgkotux*	*cdefgloqtu*	*cdefloqstu*	*cdegijkuvx*	*cdegiktuvx*
cdfgijkpux	*cdfgijlpqu*	*cdfgikptux*	*cdfgjlopqu*	*cdfglopqtu*	*cdflopqstu*
cdghijkpvx	*cdghijpvxz″*	*cdgijkpuvx*	*cdgikptuvx*	*cefghijkmx*	*cefghikmxy*
cefghilmqy	*cefghjlmoq*	*cefghlmoqy*	*cefgijkmux*	*cefgijlmqu*	*cefgikmtux*
cefgjlmoqu	*cefgkmotux*	*cefglmoqtu*	*ceflmoqstu*	*ceghijkmvx*	*ceghikmvxy*
cegijkmuvx	*cegikmtuvx*	*cegkmotuvx*	*cfghijlmpq*	*cfghilmnqy*	*cfghjlmnoq*
cfghlmnoqy	*cfgijlmnqu*	*cfgjlmnoqu*	*cfglmnoqtu*	*cghijkmnvx*	*cghijlmnqv*
cghikmnvxy	*cgijkmnuvx*	*cgijlmnquv*	*cgikmntuvx*	*cgjlmnoquv*	*cgkmnotuvx*
cglmnoqtuv	*chijkmnvxz*	*chikmnvxyz*	*cijkmnuvxz*	*deghoqryz′z″*	*degktuvwxz′*
dehoqrsyz′z″	*deoqrstuz′z″*	*deqrstuwz′z″*	*dghjpvwxz′z″*	*dgjkpuvwxz′*	*dgjopqruz′z″*
dgjpuvwxz′z″	*dgkptuvwxz′*	*dgkpuvwxyz′*	*dhjpvwxzz′z″*	*djkpuvwxzz′*	*djopqrsuz′z″*
djpqsuvwz′z″	*djpuvwxzz′z″*	*dkpuvwxyzz′*	*dpqstuvwz′z″*	*dptuvwxzz′z″*	*eghlmoqryz′*
eghmoqryz′z″	*eghmvwxyz′z″*	*egkmtuvwxz′*	*egkmuvwxyz′*	*egmruwxyz′z″*	*egmuvwxyz′z″*
ehlmoqrsyz′	*ehmoqrsyz′z″*	*ehmorxyzz′z″*	*ehmvwxyzz′z″*	*ekmouvxyzz′*	*ekmuvwxyzz′*
elmoqrstuz′	*emoqrstuz′z″*	*emoruxyzz′z″*	*emotuvxzz′z″*	*emouvxyzz′z″*	*emruwxyzz′z″*
emtuvwxzz′z″	*ghjlmnoqrz′*	*ghlmnoqryz′*	*ghlmopqryz′*	*gjlmnoqruz′*	*gjlmopqruz′*
gjmopqruz′z″	*gjmpqruwz′z″*	*hjmnrwxzz′z″*	*hjmnvwxzz′z″*	*hlmopqrsyz′*	*hmnoqrsyz′z″*
hmnorxyzz′z″	*jkmnuvwxzz′*	*jlmopqrsuz′*	*jlmopqsuvz′*	*jmnoqrsuz′z″*	*jmnoqsuvz′z″*
jmnuvwxzz′z″	*jmopqsuvz′z″*	*jmpqrsuwz′z″*	*jmpqsuvwz′z″*	*kmptuvwxzz′*	*lmnoqrsuyz′*

Table 7
Hierarchy of the 528 maximal self-complementary codes according to the classes C_1, C_2, C_3, C_4, C_5 and C_6

C^{3LDCN}	C^{4LDN}	C^{5LDCN}
$C_1 \cup C_2$	$C_1 \cup C_2 \cup C_3 \cup C_4$	$C_1 \cup C_2 \cup C_3 \cup C_4 \cup C_5 \cup C_6$
C^{3DLN}	C^{4DLCN}	C^{5DLN}
$C_1 \cup C_2$	$C_1 \cup C_2 \cup C_3 \cup C_4$	$C_1 \cup C_2 \cup C_3 \cup C_4 \cup C_5 \cup C_6$

Remark 7. The maximal C^3 self-complementary circular code X (*EUK_PRO*) belongs to the class C^{5LDCN} and is the code *acekotuvxz* in 6a.

Acknowledgment

We thank Amy Glen and a referee for their advices.

References

[1] D.G. Arquès, C.J. Michel, A complementary circular code in the protein coding genes, J. Theoret. Biol. 182 (1996) 45–58.
[2] D.G. Arquès, C.J. Michel, A circular code in the protein coding genes of mitochondria, J. Theoret. Biol. 189 (1997) 273–290.
[3] J. Berstel, D. Perrin, Theory of Codes, Academic Press, London, 1985.
[4] F.H.C. Crick, S. Brenner, A. Klug, G. Pieczenik, A speculation on the origin of protein synthesis, Origins Life 7 (1976) 389–397.
[5] F.H.C. Crick, J.S. Griffith, L.E. Orgel, Codes without commas, Proc. Natl. Acad. Sci. 43 (1957) 416–421.
[6] M. Eigen, P. Schuster, The hypercycle. A principle of natural self-organization. Part C: The realistic hypercycle, Naturwissenschaften 65 (1978) 341–369.
[7] G. Frey, C.J. Michel, Circular codes in archaeal genomes, J. Theoret. Biol. 223 (2003) 413–431.
[8] G. Frey, C.J. Michel, Identification of circular codes in bacterial genomes and their use in a factorization method for retrieving the reading frames of genes, J. Comput. Biol. Chem. 30 (2006) 87–101.
[9] S.W. Golomb, B. Gordon, L.R. Welch, Comma-free codes, Canad. J. Math. 10 (1958) 202–209.
[10] S.W. Golomb, L.R. Welch, M. Delbrück, Construction and properties of comma-free codes, Biol. Medd. Dan. Vid. Selsk. 23 (1958).
[11] J.-L. Lassez, Circular codes and synchronization, Int. J. Comput. Syst. Sci. 5 (1976) 201–208.
[12] C.J. Michel, G. Pirillo, M.A. Pirillo, Varieties of comma-free codes, Comput. Math. Appl. 55 (2008) 989–996.
[13] M.W. Nirenberg, J.H. Matthaei, The dependance of cell-free protein synthesis in E. Coli upon naturally occurring or synthetic polyribonucleotides, Proc. Natl. Acad. Sci. 47 (1961) 1588–1602.
[14] G. Pirillo, A characterization for a set of trinucleotides to be a circular code, in: C. Pellegrini, P. Cerrai, P. Freguglia, V. Benci, G. Israel (Eds.), Determinism, Holism, and Complexity, Kluwer, 2003.
[15] G. Pirillo, A hierarchy for circular codes, RAIRO-Theor. Inf. Appl. (2008) (in press).
[16] G. Pirillo, M.A. Pirillo, Growth function of self-complementary circular codes, Biology Forum 98 (2005) 97–110.