ORIGINAL PAPER

Selfcomplementary, selfreverse cyclic nucleotide sequences codonically invariant under frame shifting

Vladimir R. Rosenfeld

Abstract A cyclic nucleotide sequence is *selfcomplementary* if some circular permutation transfers it to a complementary sequence and (independently) is *selfreverse* if the cyclic order of nucleotides in both directions is the same. Moreover, it is *codonically invariant under frame shifting* if any shift of its frame does not alter the content and cyclic order of codons. We discuss sequences which may simultaneously have these three properties and exhibit color symmetry. Our considerations are also adapted for respective linear sequences.

1 Introduction

Instances of cyclic DNA and RNA are really encountered in nature. Along with these biopolymers themselves, researchers study also mathematically well-defined models of real and imagined cyclic polynucleotides. This allows to more readily determine combinatorial properties of such molecules, which may further be adapted for linear polynucleotides with consecutive repeats, as well. The present work is targeted at

V. R. Rosenfeld (✉)
Mathematical Chemistry Group, Department of Marine Sciences,
Texas A&M University at Galveston, Galveston, TX 77553–1675, USA
e-mail: rosenfev@tamug.edu; vladimir_rosenfeld@yahoo.com

V. R. Rosenfeld
Instituto de Ciencias Matematicas (ICMAT) CSIC, C/ Nicolás Cabrera,
no 13–15, Campus Cantoblanco UAM, 28049 Madrid, Spain

similar objectives and also touches in a more general context upon the color symmetry that nucleotide sequences may exhibit.

A cyclic nucleotide sequence is *selfcomplementary* if some circular permutation transfers it to a complementary sequence and (independently) is *selfreverse* if the cyclic order of nucleotides in both directions is the same. Moreover, it is *codonically invariant under frame shifting* if any shift of its frame does not alter the content and cyclic order of codons. We discuss sequences which may simultaneously have these three properties.

Let $\mathcal{A}_{n^*}$ be the set of all words w of length $|w| = n \geq 1$ over the standard genetic alphabet $\mathcal{A} = \mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}$. For an arbitrary word $u = a_1 a_2 \ldots a_s (a_j \in \mathcal{A}; j \in [1, s];$ s is not divisible by 3), one can construct a derivative word v by concatenating t copies of u, where t is necessarily divisible by 3:

$$v = \underbrace{uu \ldots u}_{t \text{ times}} = \underbrace{(a_1 a_2 \ldots a_s)(a_1 a_2 \ldots a_s) \ldots (a_1 a_2 \ldots a_s)}_{t \text{ times}}. \tag{1}$$

We denote by $\mathcal{L}_{s,t} (\mathcal{L}_{s,t} \subseteq \mathcal{A}_{s \cdot t})$ the set of all such linear words v (as in (1)) for any fixed s and t such that 3 does not divide s, but necessarily divides t, and, also, denote by $\mathcal{C}_{s,t}$ the set of respective cyclic words, where all words are defined only relative to an arbitrary circular permutation τ^p, with p considered modulo st and determining the circular shift of characters by $p \pmod{st}$ positions to the right (res. to the left). That is, for every integer p, $\tau^p v \equiv v$ in $\mathcal{C}_{s,t}$.

Earlier, we demonstrated [1] that all words of the sets $\mathcal{C}_{s,t}$ have a remarkable property that the translation of codons of any circularly permuted version $\tau^p v$ of $v \in \mathcal{C}_{s,t}$ (conserving the same circular order of codons therein) produces the same cyclic sequence of amino acids (see Proposition 3 in [1] or Proposition 1 below). Just this property of words from $\mathcal{C}_{s,t}$-sets is utilized in our paper. The following example illustrates that.

As a case in point, consider the sequence **ATCGATCGATCG**; here, the factor **ATCG** of length $s = 4$ is repeated three times. Translating this code without any shift gives *isoleucine*, *aspartic acid*, *arginine*, and *serine*, consecutively, or *IDRS* for short. The circular shift by 1 position results in *SIDR*, by 2 positions produces *RSID*, and (here) at last, the circular shift by 3 positions gives *DRSI*. Apparently, all the four translated codes of amino acids are the same relative to some circular permutation.

To manipulate nucleotide sequences, one may use three commuting operators: α standing for complementation of nucleotides in a string; β for inversion of the string, and the composition $\gamma = \alpha\beta = \beta\alpha$. Operators α, β, γ are also used for construction of graphs reflecting interrelations of amino acids (see Figs. 1 and 2 in [2] and Fig. 1 below). Say, in the first graph of γ-relations in [2], two amino acids Am_k and $Am_l (k, l \in [1, 21])$ are linked with an edge iff to Am_k there associates at least one codon $a_1 a_2 a_3$ such that the codon $\gamma(a_1 a_2 a_3)$ is associated with Am_l, and vice versa. In particular, an edge is reduced to a selfloop, if $Am_k = Am_l$ (see Fig. 2 in [2]).

We also use the notation $b_k := \alpha(a_k) (a_k \in \mathcal{A})$ to determine another word $u^* = b_1 b_2 \ldots b_s$, where the sequential order of complementary characters $b_k = \alpha(a_k) (k \in [1, s])$ is the same that a_k's have in u, rather than a reversed one, as in $\gamma(u) = \alpha(a_s)\alpha(a_{s-1}) \ldots \alpha(a_1) = b_s b_{s-1} \ldots b_1$ (!).

Lastly, we construct the following 'hybrid' word:

$$w = \underbrace{uu^*uu^* \ldots uu^*}_{t \text{ times}}$$
$$= \underbrace{(a_1a_2 \ldots a_sb_1b_2 \ldots b_s)(a_1a_2 \ldots a_sb_1b_2 \ldots b_s) \ldots (a_1a_2 \ldots a_sb_1b_2 \ldots b_s)}_{t \text{ times}},$$

(2)

which is an element of the set $\mathcal{T}_{s,t} \subseteq \mathcal{L}_{2s,t}$ or, in a cyclic form, is an element of $\mathcal{U}_{s,t} \subseteq \mathcal{C}_{2s,t}$, where the latter inclusion also means that all words from $\mathcal{U}_{s,t}$-sets obey Proposition 3 in [1].

Now, we turn to the main part.

2 Main part

Our first statement is due to Proposition 3 of [1]:

Proposition 1 *Let f be a cyclic sequence of nucleotides. Then, f conserves a circular order of its codons under any shift of its frame if (0) all the nucleotides are the same, (1) f has a length s not divisible by 3 and is consecutively read t times, with t a multiple of 3, or (2) f is composed of t repeated copies of a factor h of length s, where t is divisible by 3, while s is not.*

Here, we use the following technical lemma:

Lemma 2 *Let the word w^* denote the nucleotide sequence in the second strand complementary to the cyclic nucleotide sequence in the first strand represented by $w \in \mathcal{U}_{s,t} \subseteq \mathcal{C}_{2s,t}$, as in (1); that is,*

$$w^* = \underbrace{u^*uu^*u \ldots u^*u}_{t \text{ times}}$$
$$= \underbrace{(b_1b_2 \ldots b_sa_1a_2 \ldots a_s)(b_1b_2 \ldots b_sa_1a_2 \ldots a_s) \ldots (b_1b_2 \ldots b_sa_1a_2 \ldots a_s)}_{t \text{ times}}.$$

(3)

Then, $w^* = \tau^p w$, where τ^p is a circular shift of characters, of w, by $p = k|w|/2t$ positions to the right (res. left), where k is an odd integer; that is, the sequence of nucleotides in the second strand is the same relative to some circular permutation (See Remark 1).

Proof It is apparent from comparing the definitions of the words u and u^* above. □

Remark 1 Though Lemma 2 correctly describes a mutual arrangement of nucleotides in both strands, one necessary geometric clarification is required. Namely, that pyrimidine bases of two strands of DNA are projected in opposite directions, so as bases of one strand are able to meet the bases of the other and become chemically linked,

due to formation of complementary pairs of nucleotides. Such a remark is not yet so important for linear strands, since either strand simply can be rotated about its oblong axis to obey a needed orientation of nucleobases. But it is crucial for cyclic polynucleotides, because the flipping of a cyclic strand reverses the order of nucleotides, which becomes in general different. In order to conserve the order of nucleotides in a reversed cyclic strand, one should impose special restrictions on this order. See below.

Following Remark 1, we consider special conditions for our cyclic polynucleotide sequences to be invariant under reversion of their order. Namely, using the notation of (2) and (3), we additionally assume $u_\pi = a_1 a_2 \ldots a_s$ and $u_\pi^* = a_1^* a_2^* \ldots a_s^*$ to be *palindromes* ($\pi\alpha\lambda\iota\nu\delta\rho\acute{\omega}\mu\sigma\varsigma$), where $a_j = a_{s+1-j}$ and $a_j^* = a_{s+1-j}^*$ ($j \in [1, s]$), respectively. Also, let w_π and w_π^* be words produced by substitution of u_π for u and u_π^* for u^* on the second side of (2) and (3), correspondingly. We denote by $\mathcal{U}_{s,t}^\pi$ the subset of all such cyclic words: $w_\pi \in \mathcal{U}_{s,t}^\pi \subseteq \mathcal{U}_{s,t} \subseteq \mathcal{C}_{2s,t}$, where s is not divisible by 3, while t is divisible by 3, and both equalities hold just under $s = 1$.

The introduced palindromic restrictions allow us to state the following common corollary of Proposition 1 and Lemma 2:

Proposition 3 *Let $w_\pi \in \mathcal{U}_{s,t}^\pi$ be a cyclic nucleotide sequence as above. Then,*

- (i) w_π *is codonically invariant under frame shifting;*
- (ii) w_π *is selfcomplementary;*
- (iii) w_π *is selfreverse;*
- (iv) w_π *contains for each sort of codons $a_1 a_2 a_3$ the same number of codons $b_3 b_2 b_1$ ($a_k \in \mathcal{A}$; $b_k = \alpha(a_k)$; $k = 1, 2, 3$) (as is also in an ideal Watson and Crick DNA).*

Proof It follows from the construction, where (i) is based on Proposition 1, (ii) on Lemma 2, (iii) is easily deduced from the palindromicity of alternated factors u_π and u_π^*, in w_π, while (iv) is due to a simultaneous application of (ii) and (iii) to $w_\pi \in \mathcal{U}_{s,t}^\pi$, which completes the proof.   □

In an ideal DNA, positions symmetric about the center of a strand are occupied by complementary nucleotides a_k and b_k. In our 2-strand cyclic polynucleotide, each copy of a_j ($j \in [1, s]$) in either strand is accompanied with b_{s+1-j}, which is not in general a complementary nucleotide for a_j. Thus, such a model, based on the set $\mathcal{U}_{s,t}$ of words, does not necessarily describe an intact DNA really existing in nature. But it describes linear factors of DNA having the form

$$w_+ = b_{s-1} b_s w a_1 a_2$$
$$= b_{s-1} b_s \underbrace{(a_1 a_2 \ldots a_s \ldots b_1 b_2 \ldots b_s) \ldots (a_1 a_2 \ldots a_s \ldots b_1 b_2 \ldots b_s)}_{t \text{ times}} a_1 a_2.$$

$$(4)$$

Evidently, a frame shift along the entire strand by 1 or 2 positions to the right (res. left) does not alter the circular order of a codon sequence within $2st$ consecutive positions of the factor w in w_+ (while the latter has the total length $|w_+| = 2st + 4$).

Hence, also the translation of the w_+-factor will produce the same circular sequence of amino acids associated with its factor w.

Due to (iv) in Proposition 3 above, Corollary 3.1 and Proposition 4 of [2] may be adapted for amino-acid sequences translated from the cyclic nucleotide sequences $w_\pi \in \mathcal{U}_{s,t}^\pi$. Two respective adaptations are:

Proposition 4 *In amino-acid sequences translated from cyclic nucleotide sequences $w_\pi \in \mathcal{U}_{s,t}^\pi$, numbers of amino-acid residues are related*

$$\#_{Asp} + \#_{Asn} + \#_{Tyr} + \#_{His} = \#_{Ile} + \#_{Met} + \#_{Val}.$$

Proposition 5 *In amino-acid sequences translated from cyclic nucleotide sequences $w_\pi \in \mathcal{U}_{s,t}^\pi$, there are inequalities on the numbers of different amino acids:*

$$\#_{Met} \leq \#_{His};$$
$$\#_{His} \leq \#_{Met} + \#_{Val};$$
$$\#_{Ile} \leq \#_{Tyr} + \#_{Asn} + \#_{Asp};$$
$$\#_{Tyr} + \#_{Asn} + \#_{Asp} \leq \#_{Ile} + \#_{Val};$$
$$\#_{Gln} \leq \#_{Leu};$$
$$\#_{Trp} \leq \#_{Pro};$$
$$\#_{Ter} \leq \#_{Leu} + \#_{Ser};$$
$$\#_{Pro} \leq \#_{Trp} + \#_{Arg} + \#_{Gly};$$
$$\#_{Ser} \leq \#_{Ter} + \#_{Arg} + \#_{Gly} + \#_{Ala} + \#_{Thr};$$
$$\#_{Cys} \leq \#_{Ala} + \#_{Thr};$$
$$\#_{Ala} + \#_{Thr} \leq \#_{Ser} + \#_{Arg} + \#_{Gly} + \#_{Cys};$$
$$\#_{Leu} + \#_{Phe} \leq \#_{Gln} + \#_{Lys} + \#_{Glu} + \#_{Ter};$$
$$\#_{Gln} + \#_{Lys} + \#_{Glu} \leq \#_{Phe} + \#_{Leu};$$
$$\#_{Gln} + \#_{Lys} + \#_{Ter} + \#_{Glu} \leq \#_{Leu} + \#_{Phe} + \#_{Ser};$$
$$\#_{Trp} + \#_{Arg} + \#_{Gly} + \#_{Cys} \leq \#_{Pro} + \#_{Ser} + \#_{Ala} + \#_{Thr};$$
$$\#_{Pro} + \#_{Ser} + \#_{Cys} \leq \#_{Trp} + \#_{Ter} + \#_{Arg} + \#_{Gln} + \#_{Ala} + \#_{Thr};$$

where the number $\#_{Ter}$ of "stops" is conveniently identified to the number of different proteins.

Remark 2 We make one remark about the analogy of cyclic nucleotide sequences $w_\pi \in \mathcal{U}_{s,t}^\pi$ and strands of an ideal DNA, in the sense of common relations for the numbers of amino-acid residues. Namely, such an analogy is due to the fact that two antiparallel strands of an ideal DNA are identical and have the same content of codons that has the entire DNA. However, in natural DNAs, there may be deviations from the complementarity of codons. This may result in disagreement with the relations for the numbers of amino-acid residues translated from the ideal DNA and, more generally, from *codonic palindromic conglomerates*—see Corollary 3.1 and Proposition 4 in [2]. Whereas our combinatorial considerations for ideal objects—as are the ideal Watson
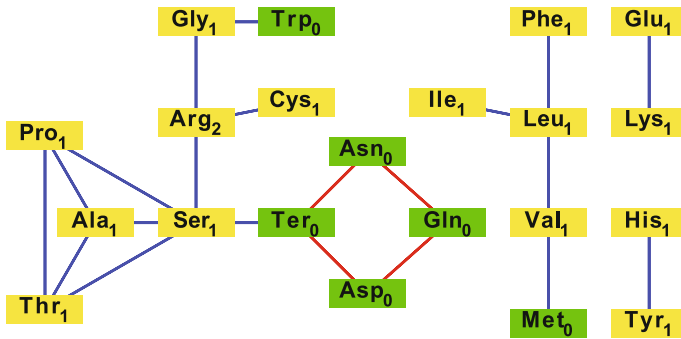
Fig. 1 The graph of β-relations of amino acids, where indices count the number of symmetric codons

& Crick DNA and cyclic nucleotides $w_\pi \in \mathcal{U}_{s,t}^\pi$—are rigorous and do not depend on any deviations which occur due to nonideal complementarity of strands in natural DNAs.

Note that both Corollary 3.1 and Proposition 4 in [2] are due to the graph of γ-relations in Fig. 1 (see p. 171 in [2]), while the graph of α-relations in Fig. 2 (see p. 174 in [2]) remains not used. Without delving into details already described in [2], we utilize here that second graph and arrive at:

Proposition 6 *In amino-acid sequences translated from cyclic nucleotide sequences* $w_\pi \in \mathcal{U}_{s,t}^\pi$, *numbers of amino-acid residues are (additionally) related*

$$\#_{Phe} = \#_{Lys};$$
$$\#_{Gly} = \#_{Pro};$$
$$\#_{Val} = \#_{Gln} + \#_{His};$$
$$\#_{Leu} = \#_{Asp} + \#_{Asn} + \#_{Glu};$$
$$\#_{Met} + \#_{Ile} + \#_{Thr} = \#_{Tyr} + \#_{Trp} + \#_{Cys} + \#_{Ter};$$
$$\#_{Arg} \leq \#_{Ala} + \#_{Ser}.$$

Proof It follows from the graph of α-relations in Fig. 2 of [2] and (ii) of Proposition 3. □

Now, we use the graph of β-relations in Fig. 1 to state:

Proposition 7 *In amino-acid sequences translated from cyclic nucleotide sequences* $w_\pi \in \mathcal{U}_{s,t}^\pi$, *numbers of amino-acid residues are (moreover) related*

$$\#_{Met} \leq \#_{Val};$$
$$\#_{Trp} \leq \#_{Gly};$$
$$\#_{Gln} \leq \#_{Asp} + \#_{Asn};$$
$$\#_{Ter} \leq \#_{Asp} + \#_{Asn} + \#_{Ser};$$
$$\#_{Asp} + \#_{Asn} \leq \#_{Gln} + \#_{Ter};$$

Proof It follows from the graph of β-relations in Fig. 1 above and (iii) of Proposition 3.

□

By virtue of Proposition 3, one may consider a simultaneous merger of all relations stated in Propositions 4–7, which are all true for cyclic nucleotide sequences $w_\pi \in \mathcal{U}_{s,t}^\pi$; moreover, recall that all numbers in Propositions 4–7 are nonnegative integers whose total is equal to $2st/3$.

As an illustration of that Propositions 4–7 are practically obeyed, construct a sequence w_π. First, prepare palindromic words $u_\pi = v_1 v_2$ and $u_\pi^* = v_1^* v_2^*$, where v_1 (v_1^*) and v_2 (v_2^*) are mirror-symmetric halves of u_π (u_π^*).

$v_1 = $ **gcatgcgacgaattcggacacataaaattaatgaacccacaaagaagcacagtatggta**

$v_2 = $ **atggtatgacacgaagaaacacccaagtaattaaaatacacaggcttaagcagcgtacg**

$v_1^* = $ **cgtacgctgcttaagcctgtgtattttaattacttgggtgtttcttcgtgtcataccat**

$v_2^* = $ **taccatactgtgcttctttgtgggttcattaattttatgtgtcctaattcgtcgcatgc**

where ordinary characters **a, c, g, t** denote four nucleotides (instead of capitals **A, C, G, T**).

The minimum word $w_\pi \in \mathcal{U}_{s,t}^\pi$ which can be constructing using (5) is

$$w_\pi = u_\pi u_\pi^* u_\pi u_\pi^* u_\pi u_\pi^* = v_1 v_2 v_1^* v_2^* v_1 v_2 v_1^* v_2^* v_1 v_2 v_1^* v_2^*. \tag{5}$$

Thus, $s = 118, t = 3$, and the length of w_π is $|w_\pi| = 2st = 708$. This sequence contains 144 cytosine and 144 guanine residues intermixed with 210 adenine and 210 thymine residues, which complies with Chargaff's second parity rule, experimentally corroborated [3–6]. Accordingly, a single strand of an ideal Watson and Crick DNA (or codonic palindromic conglomerate [2], in general) should contain equal numbers of cytosine and guanine as well as adenine and thymine residues. Since all codonic palindromic conglomerates obey (iv) of Proposition 3, Chargaff's second parity rule is a direct consequence of (iv) in Proposition 3. This rule holds practically true for single strands of most natural DNAs [3–6], which means that natural DNA's strands may closely resemble codonic palindromic conglomerates – though not necessarily obey (5).

Translation of w_π produces 236 amino-acid residues. With the aid of Maple, we perform an imagined translation of w_π and obtain the following numbers of amino-acid residues: $\#_{Ala} = 13, \#_{Cys} = 12, \#_{Asp} = 3, \#_{Glu} = 6, \#_{Phe} = 12, \#_{Gly} = 8, \#_{His} = 13, \#_{Ile} = 13, \#_{Lys} = 12, \#_{Leu} = 20, \#_{Met} = 7, \#_{Asn} = 11, \#_{Pro} = 8, \#_{Gln} = 5, \#_{Arg} = 16, \#_{Ser} = 15, \#_{Thr} = 18, \#_{Val} = 18, \#_{Trp} = 4, \#_{Tyr} = 11, \#_{Ter} = 11$. It is easy to establish that this set of numbers obeys all equalities and inequalities in Propositions 4–7. Note that codons **GAT** (*Asp*), **GAG** (*Glu*), **ATC** (*Ile*), **CTA** (*Leu*), **CTC** (*Leu*), **TAG** (*Ter*) are missing in our cyclic sequence but respective five amino acids are represented by other codons.

Remark 3 In general, there exist nucleotide sequences having any composition of codon types, where a combinatorial number of codon subsets is equal to 2^{64}–10^{20}, which excludes a full analysis of all possible instances. However, if missing codons

are known, there may be made some simplifications in the graph representing relationships of amino acids (for operators α, β, γ) and extensions of systems of numeric relationships given by Propositions 4–7. Here, we use "may be", because not always some missing codons imply such changes. In particular, the absence of codons **GAT** and **TAG** implies that the graph in Fig. 1 loses the link connecting aspartic acid (Asp) and "stop-codon acid" (Ter). Then, in addition to nonstrict inequalities of Proposition 7, there occurs $\#_{Asp} \leq \#_{Gln}$. Such considerations may be continued and involve also two graphs of [2]. Recall that stop codons may sometimes be translated into amino acids; otherwise, their presence should imply termination of a growing protein chain. However, we consider purely combinatorial, imagined sequences of nucleotides which admit any number of stop codons at any place or, alternatively, allow not to have stop codons at all. The general combinatorial treatment herein encompasses both natural and imagined sequences.

We want also to refer to certain considerations on p. 174–175 in [2], in which the results of this paper might find some retrospective application. Besides, a more complex case may be considered, in which two circularly equivalent words w_π and w_π^* are additionally (or alternatively) equivalent in some other sense, as is the rigorous matter of two independent texts [7,8].

3 Discussion

In crystallography, they use the notion of *color symmetry* introduced by Shubnikov in 1950 (see [9] and Internet), which means, in particular, that a certain passage within the constructional motif of a crystal can be made by either a shift in space or by permuting some sorts of atoms. For instance, this can be done in a crystal of NaCl, whose Na^+ and Cl^- sublattices are identical, with accuracy to sorts of atoms therein. Or even a simpler, noncrystallographic, example is a chessboard, where a shift by one line is equivalent to interchanging black and white colors of cells. We find much the same feature with the word w (w^*) in our text, where the circular permutation by s positions is equivalent to replacement of all nucleotides by complementary ones (and is tantamount to consideration of the second strand).

Accordingly, the *automorphism (color symmetry) group* $G = \mathrm{Aut}(w_\pi) = \langle\rho, \sigma, \tau^s\rangle$ of a cyclic word $w_\pi \in \mathcal{U}_{s,t}^\pi$ is the group generated by three its elements ρ, σ, τ^s standing for the color-symmetry operation of complementation of nucleotides, reversion of the cyclic order of w_π, and circular permutation thereof clockwise (counterclockwise) by s positions, consecutively; $\rho^2 = \sigma^2 = (\tau^s)^{2t} = e$, where $2t$ is the order of τ^s, and e is an identity. Obviously,

$$G = \langle\rho\rangle \times \langle\sigma, \tau^s\rangle = \langle\sigma, \tau^s\rangle \times \langle\rho\rangle = ((\langle\tau^s\rangle \rtimes \langle\sigma\rangle) \times \langle\rho\rangle, \tag{6}$$

where $\langle\tau^s\rangle \rtimes \langle\sigma\rangle$ is the semidirect product of a cyclic group $\langle\tau^s\rangle = \{e, \tau^s, \tau^{2s}, \ldots, \tau^{2s(t-1)}\}$ and group $\langle\sigma\rangle = \{e, \sigma\}$, with the former as a proper normal subgroup, $\langle\tau^s\rangle \lhd \langle\sigma, \tau^s\rangle$; and $\langle\rho\rangle = \{e, \rho\}$; apparently, $\langle\sigma, \tau^s\rangle \cong D_{2t}$, and G has $|G| = 8t$ elements. Since we define cyclic words w_π relative to an arbitrary circular permutation τ^p, $G \subset T = \langle\rho, \sigma, \tau\rangle$ ($|T| = 8st$), where T is isomorphic to the automorphism

group $\mathrm{Aut}\overset{\circlearrowleft}{\mathcal{U}}^{\pi}_{s,t}$ of the set $\overset{\circlearrowleft}{\mathcal{U}}^{\pi}_{s,t}$ of all circularly permuted versions $\overset{\circlearrowleft}{w}_{\pi}$ of words $w_{\pi} \in \mathcal{U}^{\pi}_{s,t}$.

By cutting a cyclic word $w_{\pi} \in \mathcal{U}^{\pi}_{s,t}$ at one position, one can produce a respective linear word $\overset{\times}{w}_{\pi}$, whose (color) symmetry depends on the choice of such a position. There are three mutually exclusive instances:

(1) the cut is inside a factor u (u^*) of w_{π} (see (2) or (3)) and divides it into two unequal parts, which produces asymmetric word $\overset{\times}{w}_{\pi}$ with a trivial symmetry group $H_1 = \{e\}$;
(2) the cut divides a factor u (u^*) of w into equal parts, which corresponds to the mirror-symmetry group $H_2 = \{e, \sigma\}$ of $\overset{\times}{w}_{\pi}$ (without the color symmetry);
(3) the cut is made just between two adjacent factors u and u^*, which produces a word $\overset{\times}{w}_{\pi}$ with the group $H_3 = \{e, \rho\}$ of color symmetry (but without the mirror symmetry).

Here, we leave the symmetry analysis of two-strand (cyclic and linear) polynucleotides formed from single strands (2) and (3) as an exercise to the reader. But note that, in this case, the number of symmetry elements is doubled, due to the 'antiparallelism' of two equal strands in such constructions. In general, one may consider all possible symmetry groups (and semigroups) of two-strand (spiral) cyclic polynucleotides and respective nucleotide sequences.

For imagined cyclic nucleotide sequences considered above, it is also of interest to specially discuss a theoretical (or even practical) possibility of translation of such a sequence or its linear analog into an amino-acid sequence using both directions of reading the former. If this may be of a biological relevance, an organism having such DNA's strands would be a champion among all organisms, in the sense of its conceptual survival in nature.

The numbers of amino-acid residues in our illustrative sequence w_{π} are tantamount to the following content of amino acids: *alanine* 5.508 %, *cystein* 5.085 %, *aspartic acid* 1.271 %, *glutamine acid* 2.542 %, *phenylalanine* 5.085 %, *glycine* 3.390 %, *histidine* 5.508 %, *isoleucine* 5.508 %, *lysine* 5.085 %, *leucine* 8.475 %, *methionine* 2.966 %, *asparagine* 4.661 %, *proline* 3.390 %, *glutamine* 2.119 %, *arginine* 6.780 %, *serine* 6.356 %, *threonine* 7.627 %, *valine* 7.627 %, *tryptophan* 1.695 %, *tyrosine* 4.661 %, *stop* 4.661 %. It is interesting to check if such a content may correspond to any natural case or be suitable for any (special) applications.

Of practical importance may be nanotechnological applications of cyclic polynucleotides and proteins. Say, it is worth trying to perform construction (selfassembly) of nanotubules and nanotori from identical cyclic polynucleotides. In this case, regular polymers (say, such as considered above) normally exhibit the lower minimum of energy than irregular ones. So, even if our objects may deviate from natural standards for living organisms, there still remains a certain reason to study these for perspective nanotechnological uses.

References

1. V.R. Rosenfeld, D.J. Klein, Cyclic nucleotide sequences codonically invariant under frame shifting. Studia Univ. Babes-Bolyai Chemia **55**(4), 177–182 (2010). (Texts are available from the authors!)
2. V.R. Rosenfeld, D.J. Klein, Implications of sense/antisense nucleic-acid codons on amino-acid counts. Studia Univ. Babes-Bolyai Chemia **55**(4), 167–176 (2010)
3. N. Sueoka, Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J. Mol. Evol. **40**, 318–325 (1995)
4. J.R. Lobry, Properties of a general model of DNA evolution under no-strand-bias conditions. J. Mol. Evol. **40**, 326–330 (1995)
5. S.J. Bell, D.R. Forsdyke, Accounting units in DNA. J. Theor. Biol. **197**, 51–61 (1999)
6. K. Sorimachi, A proposed solution to the historic puzzle of Chargaff's second parity rule. Open Genomics J. **2**, 12–14 (2009)
7. V.R. Rosenfeld, Using semigroups in modeling of genomic sequences. MATCH Commun. Math. Comput. Chem. **56**(2), 281–290 (2006)
8. V. R. Rosenfeld, Equivalent genomic (proteomic) sequences and semigroups. J. Theor. Chem. (submitted)
9. A.V. Shubnikov, N.V. Belov, W.T.E. Holser, *Color Symmetry* (Pergamon Press, Oxford, 1964)