RESEARCH ARTICLE • OPEN ACCESS

Constructing DNA Codes with Larger Distance from Quaternary Words

Benediktus Panji Pradipta and Nikken Prima Puspita

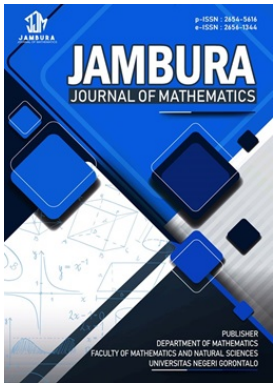
JAMBURA JOURNAL OF MATHEMATICS

Volume 7, Issue 2, Pages 151–155, August 2025

Received 1 July 2025, Revised 9 August 2025, Accepted 12 August 2025, Published 15 August 2025

To Cite this Article : B. P. Pradipta and N. P. Puspita,"Constructing DNA Codes with Larger Distance from Quaternary Words", *Jambura J. Math*, vol. 7, no. 2, pp. 151–155, 2025, *https://doi.org/10.37905/jjom.v7i2.33132*

JOURNAL INFO • JAMBURA JOURNAL OF MATHEMATICS



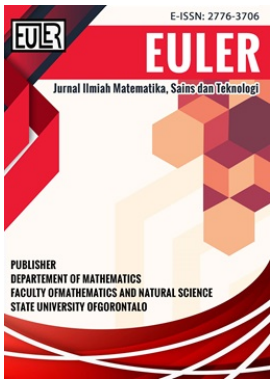
Homepage	:	http://ejurnal.ung.ac.id/index.php/jjom/index
Journal Abbreviation	:	Jambura J. Math.
Frequency	:	Biannual (February and August)
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/jjom
Online ISSN	:	2656-1344
Editor-in-Chief	:	Hasan S. Panigoro
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/jjom/oai
Google Scholar ID	:	iWLjgaUAAAAJ
Email	:	info.jjom@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

Jambura Journal of Biomathematics

Jambura Journal of Mathematics Education

Jambura Journal of Probability and Statistics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi

Jambura Journal of Mathematics, Volume 7, Issue 2, Pages 151–155, August 2025 https://doi.org/10.37905/jjom.v7i2.33132

Research Article

Constructing DNA Codes with Larger Distance from Quaternary Words

Benediktus Panji Pradipta[1,*], **Nikken Prima Puspita**[1]

[1]*Department of Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia*

ARTICLE HISTORY
Received 1 July 2025
Revised 9 August 2025
Accepted 12 August 2025
Published 15 August 2025

KEYWORDS
DNA Codes
Clique
Permutation Group
Minimum Distance

ABSTRACT. *In this work, we propose a novel method to construct DNA codes from quaternary words. The method uses permutation groups which act in the set $\{1, 2, \ldots, 4n\}$ that represents the coordinate and coordinate value of quaternary words. The DNA code is then obtained by finding a clique of the construction graph and mapping it with bijective map $0\mapsto A$, $1\mapsto C$, $2\mapsto T$, $3\mapsto G$. We then provide a novel way to construct DNA codes with a larger Hamming distance and reverse-complement distance from previously obtained DNA codes. This method uses a modified construction graph according to the desired distance parameter. As the result, we can refine a DNA code to have a better Hamming distance and reverse-complement distance while preserving the fixed GC-content constraint. This method eases the search for DNA codes with large distance parameters.*

1. Introduction

The idea to use DNA for computation was first proposed by Tom Head in 1987 [1], but it was Adleman [2] who successfully used DNA in computation, in particular to solve the Hamiltonian path problem. DNA itself is double-stranded molecule found in living organisms which contain high amount of bioinformation. DNA is formed by four basic units called nucleotides: Adenine, Cytosine, Guanine, and Thymine. Each nucleotides have its own complementary bases following the rules of Watson-Crick [3].

The core of DNA computing is DNA hybridization: the process of linking each nucleotide in the DNA with its specific complementary base. However, errors can occur in the computation process, such as mismatch pairs, formation of loops and secondary pairs, etc. Thus, minimalizing errors when designing DNA sequences plays crucial role in determining the success of DNA computing.

The key idea of DNA codes is controlling the process of constructing DNA sequences by using error-correcting codes. The process led scientist to design DNA codes with regard to some certain constraints or properties. DNA code is defined as a set of DNA sequences (also known as DNA codewords). Gaborit and King [4] presented combinatorial constraints to address the issue, which then attracted classical coding theorist to develop the concept of DNA coding theory [5–9]. The development journey of DNA coding theory is summarized by Xu et al. [10]. Similar to the main problem in classical coding theory, the need to construct DNA codes with new method is highly demanded and become one of the main problems in DNA codes study.

In this paper, we propose a novel construction method of DNA code from quaternary words. We combined the concept of constructing codes from quaternary words by Laaksonen and Östergård ([11, 12]) with the formation of DNA codes in [5]. We use the aid of graph theory especially in finding clique to get the desired DNA codes. By examining the Hamming distance and reverse-complement distance of the words, we get the DNA codes.

Furthermore, with the numerous methods of DNA code construction already proposed in many studies before, which can also be seen in [10], designing new DNA codes is not necessarily by using a new method, but can also be done by refining the previously obtained DNA codes. Therefore, we propose a way to obtain DNA codes with larger Hamming distances from previously given DNA codes. We use graphs constructed by DNA codes as our main tool. This method can be used to ease the finding of DNA code especially the one with large distance parameters.

The paper is organized as follows. In Section 2, we recall some basic concepts of graph theory and DNA codes. In Section 3, we present the construction method of DNA codes from quaternary words and the method to obtain DNA codes with larger distance. We give the conclusion of our research in Section 4.

2. Preliminaries

2.1. Graph

A graph is defined as an ordered pair (V, E) of finite sets V and $E \subseteq V \times V$. The elements of V are called vertices and the elements of E are called edges. Two vertices of a graph are said to be adjacent if there is an edge between the two vertices. A graph is said to be complete if all pairs of its vertices are adjacent. A graph $H = (V_H, E_H)$ is called a subgraph of graph $G = (V_G, E_G)$ if $V_H \subseteq V_G$ and $E_H \subseteq E_G$. A clique in a graph G is a set of vertices with maximum cardinality which induces a complete subgraph of graph G.

2.2. Codes

A q-ary code C of length n is a subset of A_q^n where $A_q = \{1, 2, \ldots, q-1\}$. Elements of a code are called codewords

*Corresponding Author.

Email : *benediktuspanji@live.undip.ac.id* (B. P. Pradipta)
Homepage : http://ejurnal.ung.ac.id/index.php/jjom/index / E-ISSN : 2656-1344

and the set A_q is called the alphabet of code. Between two codewords x and y, the Hamming distance is defined as the number of coordinates where x and y differs. The minimum distance of code C is the minimum hamming distance between all pairs of codewords in C.

Two q-ary codes are said to be equivalent if one code can be obtained from the other by permuting the coordinate, the coordinate value, or both. Such permutation map between two equivalent codes is called an automorphism. Set of automorphisms together with composition forms group called the automorphism group of a code, with each element is called a group of automorphism.

2.3. DNA Codes

Deoxyribonucleic acid (DNA) stores genetic information of life in the DNA strand which consists of nucleotides. The four nucleotides are denoted by its initial: Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). Each of the nucleotides' Watson-Crick complementary base is denoted by A^C=T, T^C=A, C^C=G, and G^C=C.

A DNA code of length n is a set of sequences $x_1x_2\dots x_n$, where $x_i\in\{A,C,G,T\}$ for all i= 1, 2,…, n, with each symbol represents a corresponding nucleotide. An element of a code is called a codeword. In this paper, we use D to denote a DNA code. Given a codeword x=$x_1x_2\dots x_n$ in code D. The complement of x is the codeword $x^C = x_1^C x_2^C \dots x_n^C$, the reverse of x is the codeword $x^{RV} = x_n x_{n-1} \dots x_1$, and the reverse-complement of x is the codeword $x^{RC} = x_n^C x_{n-1}^C \dots x_1^C$. The Hamming distance between any two codewords x and y is the number of places (which also called as coordinates) where x and y differ, denoted by $d(x,y)$. After defining the reverse and the complement of codewords, we can define the reverse distance between x and y by $d(x,y^{RV})$ and the reverse-complement distance by $d(x,y^{RC})$.

DNA codes have different errors from classical error-correcting codes. Its errors happen in the hybridization process. Hybridization is a bonding process of a DNA strand to its Watson-Crick complement strand to form a double helix [13]. A hybridization process which results in proper bonding between a DNA strand with its Watson-Crick complement strand is called specific hybridization. The result of hybridization process is said to be a false positive when a non-specific hybridization occurs, such as when a DNA nucleotide bonds a base other than its Watson-Crick complement, in which this case is called mismatches. Non-specific hybridization may also occur between a DNA strand and the reverse of a different strand. The result of hybridization process is said to be a false negative when hybridization does not happen.

The key idea to limit false positive and false negative results is by designing good DNA codes, that satisfies some combinatorial constraints on the set of DNA codewords [14]. There are four constraints to be mainly considered when designing a DNA code:

1. Hamming distance constraint (HD):
   The HD constraint of DNA codes with parameter d means that the Hamming distance between two codewords x and y is greater than or equal to d, i.e., $d(x,y) \geq d$, for every pair of distinct codewords x, y in the code.
2. Reverse-complement constraint (RC):
   The RC constraint of DNA codes with parameter d means that the reverse-complement distance between two codewords x and y is greater than or equal to d, i.e., $d^{RC}(x,y) \geq d$, for every pair of codewords x, y in the code. A code that satisfies the reverse-complement constraint is called a reverse-complement code.
3. Reverse constraint (RV):
   The RV constraint of DNA codes with parameter d means that the reverse distance between two codewords x and y is greater than or equal to d, i.e., $d^{RV}(x,y) \geq d$, for every pair of codewords x, y in the code. A code that satisfies the reverse constraint is called a reverse code.
4. Fixed GC-content constraint (GC):
   GC-weight of a codeword is the number of G and C symbols appear in the codeword. Fixed GC-content constraint is that all codewords in a code has a same GC-weight, allowing hybridization of multiple codewords simultaneously.

3. Results and Discussion

3.1. DNA Codes Construction

In this part, we present the construction method of the DNA codes. Let x=$c_1c_2\dots c_n$ be a codeword of a quaternary code C with length n. We represent the quaternary codeword $c_1c_2\dots c_n$ as the set

$$\{c_k n + k \,|\, k \in \{1,2,\dots,n\}\}, \tag{1}$$

so that each codeword is a subset with n element of $\{1,2,\dots,4n\}$.

We now recall the notion of group action. An action of G on a set S is a function $* : G \times S \to S$ such that $(g_1g_2) * s = g_1 * (g_2 * s)$ and $e * s = s$, for all $g_1, g_2 \in G$, $s \in S$, and e is the identity element of G. Since permutation is basically a function, it can be seen as an action of a set.

We then can study permutations that acts on the set $\{1,2,\dots,4n\}$. We define block l as the set $\{l, n+l, 2n+l, 3n+l\} \subset \{1,2,\dots,4n\}$, for all $l \in \{1,2,\dots,n\}$. Note that not all permutations are allowed to be used, but we only choose permutations that maps a block to some block. The orbit of word y is defined as

$$\{hy \,|\, h \in G\}.$$

A code with respect to G as its group of automorphisms consists of unions of such orbits of codewords, which acts as partition in the code. In the other words, for each orbit, it is impossible for a word in the orbit to be contained in the code while the others are not.

Example 1. Let G be a subgroup of S_{16} generated by

$$(16815)(510123)(914167)(132411)$$

that acts on the set $\{1,2,\dots,4n\}$ which represent a quaternary code. By the definition of orbits, if a codeword $c_1c_2c_3c_4$ is included in an orbit, then all other codewords obtained from the permutation are in the same orbit. For instance, take the codeword 0000, which is by eq. (1) represented as {1,2,3,4}. By the generator of our permuta-

tion group, {1,2,3,4} is mapped to the set {6,4,11,5}, which again by eq. (1), corresponds to the codeword 1120. With the same process, we have $1120 \mapsto 3221$, $3221 \mapsto 3032$, and $3032 \mapsto 0000$. Thus, the orbit of codeword 0000 is $\{0000, 1120, 3221, 3032\}$.

We then define a bijective mapping from the set $\{0, 1, 2, 3\}$ to the set of symbols of nucleotides $\{A, C, T, G\}$ by

$$0 \mapsto A,\ 1 \mapsto C,\ 2 \mapsto T,\ 3 \mapsto G.$$

The Watson-Crick complement of DNA can be applied to elements 0,1,2,3 by defining the complement as the corresponding nucleotides of DNA, that is,

$$0^C = 2,\ 2^C = 0,\ 1^C = 3,\ 3^C = 1.$$

The reverse-complement distance of codewords in a quaternary code C over the set $\{0, 1, 2, 3\}$ then can be defined by this complement.

Now we are ready to present our construction of DNA codes from quaternary words. We choose a permutation $\pi \in S_{4n}$ that acts in the set $\{1,\ 2,\ \ldots,\ 4n\}$ of the coordinate and coordinate value of quaternary words with length n. From this action group, we obtain orbits of words. Two criteria for constructing the codes are as follows:

1. for any orbit W_i in the codes, the Hamming distance and the reverse-complement distance of any two distinct words in the orbit have to be greater than or equal to d.
2. for any two orbits W_i and W_j in the code, the Hamming distance and the reverse-complement distance between words in W_i and W_j have to be greater than or equal to d.

We call orbits that satisfy 1 as admissible orbits. From the admissible orbits, we construct the graph as follows:

1. The admissible orbits become the vertices of the graph.
2. Weight of the vertices is the length of the orbit corresponding to the vertices.
3. Two vertices are adjacent if and only if the two corresponding orbits satisfy criterion 2.
4. Clique on graph G becomes the code and the weight of the cliques is the size of the code.

Let C be the quaternary code obtained from the clique of the graph. We recall back the mapping

$$0 \mapsto A,\ 1 \mapsto C,\ 2 \mapsto T,\ 3 \mapsto G$$

to get the corresponding DNA Codes D. Then we have the following property.

Proposition 1. *The obtained DNA code D has minimum Hamming distance and reverse complement distance greater than or equal to d.*

The proof of Proposition 1 follows directly from the two criteria of constructing the codes and the adjacency rule of the graph. Example of the construction is as follows.

Example 2. Suppose we want to construct a DNA code of length 6 with minimum Hamming distance and reverse-complement distance equal to 4. We choose the permutation $\pi = (1)(7)(13)(19)(2\,11\,15\,4)(8\,17\,21\,10)(14\,23\,3\,16)(20\,5\,9\,22)(6)(12)(18)(24)$. By acting π on the set of all quaternary words of length 6, i.e., on $\{0, 1, 2, 3\}^6$, we obtain 1120 orbits, of which 232 are admissible. From these admissible orbits we build a graph with 232 vertices and 10,248 edges.

We use the open-source program *Cliquer* [15] to find cliques of this graph. Cliquer returns a clique of 8 vertices with total length 16, consisting of the following orbits:

$$\{001220, 012330, 023000, 030110\}, \{003133, 021313\},$$
$$\{010203, 032023\}, \{301003, 312113, 323223, 330333\},$$
$$\{303310\}, \{310020\}, \{321130\}, \{332200\}.$$

Let the mapping $\varphi : \{0, 1, 2, 3\} \to \{A, C, T, G\}$ be given by $0 \mapsto A$, $1 \mapsto C$, $2 \mapsto T$, and $3 \mapsto G$. Under φ, the clique above corresponds to a DNA code $\mathcal{C}$ of 16 codewords, each of length 6, having minimum Hamming distance and reverse-complement distance 4:

$$\{\text{AACTTA, ACTGGA, ATGAAA, AGACCA}\},$$
$$\{\text{AAGCGG, ATCGCG, ACATAG, AGTATG}\},$$
$$\{\text{GACAAG, GCTCCG, GTGTTG, GGAGGG}\},$$
$$\{\text{GAGGCA, GCAATA, GTCCGA, GGTTAA}\}.$$

3.2. DNA Codes with Larger Distances

We have arrived at the construction of DNA codes with larger distances by using graphs. We developed the construction by Laaksonen and Östergård [11, 12] to make it applicable in the sense of DNA codes. Suppose we have a DNA code D with a minimum Hamming distance and reverse-complement distance d. We want to construct a DNA code with minimum Hamming distance and reverse-complement distance of d', with $d' > d$. For every codeword $x \in D$, we compute the reverse-complement distance x^{RC} of codeword x. We then construct the following graph Γ:

1. The vertex set of Γ is the family of all set $\{x, x^{RC}\}$ for all $x \in D$, with every vertex of Γ corresponds with a set $\{x,\ x^{RC}\}$.
2. Two vertices in Γ are adjacent if and only if all pairwise distances between the codewords in the two different vertices are greater than or equal to d'. For instance, two vertices correspond with set $\{x,\ x^{RC}\}$ and $\{y,\ y^{RC}\}$ are adjacent if $d(x, y)$, $d(x, y^{RC})$, $d(x^{RC}, y)$, $d(x^{RC}, y^{RC}) > d'$.

Then we have the following property.

Proposition 2. *The clique of graph Γ mentioned above guarantees all pairwise codewords in the vertices in the clique has distance at least d'. By selecting only codeword x from every vertex $\{x, x^{RC}\}$ in the clique and removing the codeword x^{RC}, the obtained DNA code has minimum Hamming distance and minimum reverse-complement distance d', with $d' > d$. Furthermore, the*

fixed-GC content constraint is preserved.

Proof. By the adjacency rule of graph Γ, we make sure that all pairs of codewords in the two vertices have not only Hamming distance but also reverse-complement distance greater than or equal to d'. Therefore, all pairs of codewords from different vertices within the clique of the graph will also have Hamming distances and reverse-complement distances greater than or equal to d', thus satisfying the Hamming distance and reverse-complement constraints, and implying the satisfaction of reverse constraint. By only choosing one codeword from each vertex in the clique, we obtain a DNA code with minimum Hamming distance and minimum reverse-complement distance greater than or equal to d'. Since all codewords in this new DNA code with minimum Hamming distance d' are obtained from the previous DNA code, the fixed GC-content constraint is preserved. □

Examples of the construction are given below.

Example 3. Let C be a reversible self-dual code over GF(4) with generator matrix

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & \omega & \overline{\omega} \\ 0 & 1 & 0 & 0 & \overline{\omega} & \omega \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right]$$

Using the algorithm presented by Kim *et al.* [5], we obtain a DNA code with 14 codewords satisfying the RC constraint with $d = 2$ and fixed GC content $k = 4$ as follows:

{CCAACC, CCTTCC, ATCCGC, ATGGGC, CGCCTA, CGGGTA, ACCCTG, ACGGTG, GTCCCA, GTGGCA, AGCCCT, TCCCGA, CTCCAG, GACCTC}.

These codewords are then numbered 1 to 14 respectively. By applying our construction, we obtain a construction graph Γ with 14 vertices and 24 edges as illustrated in Figure 1.

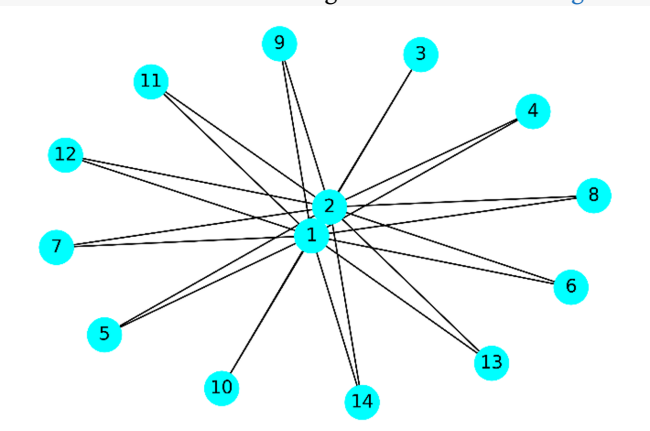


Figure 1. Graph Γ with 14 vertices and 24 edges

From Figure 1, we can see that the clique of graph Γ is a set of two vertices, where one vertex is either vertex 1 or 2, and the other vertex is chosen from the remaining vertices. We choose the vertex set $\{2, 14\}$ as our clique. This clique corresponds to the DNA code

{CCTTCC, GACCTC}.

This code satisfies the Hamming distance and RC constraints with $d = 5$ and the GC content constraint with $k = 4$.

Example 4. Let C be a reversible self-dual code over GF(4) with generator matrix

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & \omega & \omega \\ 0 & 1 & 0 & 0 & 1 & \overline{\omega} & 1 & \omega \\ 0 & 0 & 1 & 0 & 1 & \omega & \omega & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{array}\right]$$

Using the algorithm presented by Kim et al. [5], we obtain a DNA code with 70 codewords satisfying the RC constraint with $d = 4$ and fixed GC content $k = 4$ as follows:

{ATGCCGTA, ACTGGTCA, AGCTTCGA, CAGTTGAC, CTAGGATC, CGTAATGC, AAACACCC, ATGACTGC, ACTTGGAC, AGCGTATC, CAGGTTCA, CTATGCGA, CCCACAAA, CGTCAGTA, AAAGAGGG, ATGTCACG, ACTAGCTG, AGCCTTAG, GATTCCGA, GTCGATCA, GCACTGTA, GGGAGAAA, ATCCGACA, ACAGCCTA, CACTAAGC, CGAATCAC, ATCGGTGT, TGTGGCTA, CACAATCG, GCTAACAC, ATCAGCAC, ACATCAGC, CACGACTA, CGACTACA, ATCTGGTG, ACAACTCG, GCGGTCTA, GCTCAACA, ACGGTAGA, AGATGGCA, CATTGCTC, CTCGTTAC, ACGCTTCT, TCTTCGCA, CATAGGAG, GAGGATAC, ACGTTCTC, AGAGGTAC, CATGGAGA, CTCTTGCA, ACGATGAG, AGACGATG, GAGTAGCA, GTAGCAGA, AAGAGCCA, ACCGAGAA, CAATCTCC, CCTCTAAC, AAGTGGGT, TGGGTGAA, CAAACAGG, GGACAAAC, AAGCGAAC, ACCTATCC, CAAGCGAA, CCTATCCA, AAGGGTTG, ACCAAAGG, GTTGGGAA, GGAAACCA}

These codewords are then numbered 1 to 70 respectively, and by applying our construction, we obtain a construction graph Γ with 70 vertices and 1063 edges. In search of the clique from this graph, we use the `Cliquer` program developed by Niskanen and Östergård [15]. From `Cliquer`, we obtain a clique of 8 vertices corresponding to the DNA code:

{ACTGGTCA, CAGTTGAC, AAACACCC, ATGACTGC, AGCGTATC, CTATGCGA, CCCACAAA, CGTCAGTA}.

This code satisfies the Hamming distance and RC constraints with $d = 6$ and the GC content constraint with $k = 4$.

4. Conclusion

From this research, we have presented a novel construction method for DNA codes from quaternary words by benefiting from group automorphism and graph theory. We can choose permutations that act on the set $\{1, 2, \ldots, 4n\}$. This set represents the coordinate and coordinate value of quaternary words. Then, we construct a graph satisfying two criteria. By finding the clique of the graph and mapping the words contained in the clique with the bijective maps $0 \mapsto$ A, $1 \mapsto$ C, $2 \mapsto$ T, $3 \mapsto$ G, we obtain our DNA code.

Moreover, by examining the Hamming distance and reverse-complement distance to construct the graph Γ, we can design DNA codes with higher distance parameters while preserving the fixed GC content constraint, thus obtaining a DNA code with better parameters. This method can ease the process of finding DNA codes with large distances. Future research includes applying from this construction method to obtain new bounds of DNA codes by examining DNA codes with lower minimum Hamming distance and reverse-complement constraints to design ones with larger distances.

Author Contributions. **B. P. Pradipta**: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing–original draft preparation, writing–review and editing, visualization. **N. P. Puspita**: Conceptualization, methodology, validation, writing–review and editing, supervision, project administration, funding acquisition. All authors participated in the discussion of the results and collaborated on the final version of the manuscript.

Acknowledgement. The authors extend their sincere thanks to the editor and anonymous reviewers for their meticulous evaluation, valuable comments, and constructive suggestions, which have substantially contributed to the enhancement of this work.

Funding. This research was supported by the 2025 DIPA PNBP UNDIP funding.

Conflict of interest. The authors declare that there is no conflict of interest concerning this article.

Data availability. Not applicable.

References

[1] T. Head, "Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors," *Bull. Math. Biol.*, vol. 49, no. 6, pp. 737–759, 1987, doi: 10.1016/S0092-8240(87)90018-8.

[2] L. M. Adleman, "Molecular Computation of Solutions to Combinatorial Problems," *Science*, vol. 266, no. 5187, pp. 1021–1024, Nov. 1994, doi: 10.1126/science.7973651.

[3] D. Limbachiya and M. K. Gupta, "On Designing DNA Codes and their Applications," Dhirubhai Ambani Institute of Information and Communication Technology, 2019.

[4] P. Gaborit and O. D. King, "Linear constructions for DNA codes," *Theor. Comput. Sci.*, vol. 334, no. 1–3, pp. 99–113, Apr. 2005, doi: 10.1016/j.tcs.2004.11.004.

[5] H. J. Kim, W. H. Choi, and Y. Lee, "Designing DNA codes from reversible self-dual codes over GF(4)," *Discrete Math.*, vol. 344, no. 1, Jan. 2021, doi: 10.1016/j.disc.2020.112159.

[6] S. Houghten and S. Banik, "Effective decoders for DNA codes," *BioSystems*, vol. 211, Jan. 2022, doi: 10.1016/j.biosystems.2021.104583.

[7] A. Nicolaidis and F. Psomopoulos, "DNA coding and Gödel numbering," *Physica A: Stat. Mech. Appl.*, vol. 594, May 2022, doi: 10.1016/j.physa.2022.127053.

[8] S. V. Petoukhov, "Genetic code, the problem of coding biological cycles, and cyclic Gray codes," *BioSystems*, vol. 246, Dec. 2024, doi: 10.1016/j.biosystems.2024.105349.

[9] B. Feng, S. S. Bai, B. Y. Chen, and X. N. Zhou, "The constructions of DNA codes from linear self-dual codes over $\mathbb{Z}_4$," 2015, doi: 10.2991/cisia-15.2015.135.

[10] J. Xu, W. Liu, K. Zhang, and E. Zhu, "DNA coding theory and algorithms," *Artif. Intell. Rev.*, vol. 58, no. 6, Jun. 2025, doi: 10.1007/s10462-025-11132-x.

[11] A. Laaksonen and P. R. J. Östergård, "New lower bounds on q-ary error-correcting codes," *Cryptogr. Commun.*, vol. 11, no. 5, pp. 881–889, Sep. 2019, doi: 10.1007/s12095-018-0302-9.

[12] A. Laaksonen and P. R. J. Östergård, "Constructing error-correcting binary codes using transitive permutation groups," *Discrete Appl. Math.*, vol. 233, pp. 65–70, Dec. 2017, doi: 10.1016/j.dam.2017.08.022.

[13] A. Marathe, A. E. Condon, and R. M. Corn, "On combinatorial DNA word design," *J. Comput. Biol.*, vol. 8, no. 3, pp. 201–219, Jun. 2001, doi: 10.1089/10665270152530818.

[14] D. Limbachiya, B. Rao, and M. K. Gupta, "The art of DNA strings: Sixteen years of DNA coding theory," Jul. 2016, [Online]. Available: http://arxiv.org/abs/1607.00266.

[15] S. Niskanen and P. R. J. Östergård, *Cliquer User's Guide, Version 10*, Espoo, Finland: Communications Laboratory, Tech. Rep. T48, Helsinki Univ. of Technology, 2003.