Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

DNA Sequencing Technique for Secure Data Transfer in Cloud Computing

Dr. A. Kannaki Vasanthaazhagu [1], Devadarshini A[2], Harini C [3], Keerthana C [4], Sangavi S [5]

[1]Head of Department Computer Science Engineering, Achariya College of Engineering Technology

[2,3,4,5] Students of B.Tech , Department of Computer Science and Engineering, Achariya College of Engineering Technology

Abstract

Cloud computing is providing on-demand services over the internet, and it enables users to store and run applications on virtual servers connected through secure networks. Cryptography converts data into a format that is unreadable for an unauthorized user. This Project uses the technique for securing data using the Biological Structure of Deoxyribonucleic Acid (DNA) is called DNA Cryptography. This paper mainly focuses on securing data by using DNA Cryptography technique, that encodes information using three methods: (1) Encryption, (2) Random Key Generation and (3) Decryption. This method consists of four modules they are 1. Encryption using Hybrid Method. 2. Generation of Color Palettes by 24 Set Complementary Rule from the Encrypted DNA sequence. 3. Key Generation. 4. Decryption of Data Packets by Hybrid technique. Initially in the first stage of Encryption, when the sender sends a message, each letter of the word is converted into ASCII Code and the ASCII code is converted into Binary Code of 0's & 1's. Using Hybrid Technique, we are substituting the DNA Pairs (A = 00, C = 01, G =10, and T = 11) for the converted Binary Codes, then taking complements for the substituted values. In the second stage, the Color palettes is generated for the sender Encrypted message by 24 Set Complementary Rule. Next in the third stage, the decrypting key is generated by the Color Palettes formation for the receiver. Finally in the last stage, the Decrypting key is delivered and the receiver decrypt and reads the original message from the Sender.

Keywords: DNA Cryptography, Cloud Computing, Color palettes, 24 Set complementary Rule.

1. Introduction

Cloud computing has made data security a crucial concern, especially for organization relying on cloud storage. An emerging solution involves using DNA sequencing techniques to enhance security. By encoding digital data into DNA sequences, organizations benefit from DNA's high data density and stability—one gram can store approximately 215 petabytes. This project adopts a client-server model, where service providers (servers) and requestors (clients) interact within cloud computing. DNA-based data transfer encodes binary data into nucleotide sequences (A, T, C, G). DNA encryption is highly secure due to the complexity of decoding. This project’s encryption method involves converting data into ASCII, then binary, and mapping it to DNA bases. A hybrid method applies substitution and complementary rules, producing cipher text compressed into color palettes. A random key enhances encryption with decryption reversing the process. Challenges include the cost of the DNA synthesis, the need for standardized encoding protocols, and ethical considerations like the data privacy and compliance with regulations such as GDPR.

However, advances in synthetic biology and nanotechnology are expected to improve speed and efficiency, paving the way for DNA-based cloud storage solutions.

EXISTING SYSTEM

The existing technique of DNA cryptography uses a sequence dictionary method using a fixed spiral for securing data. The existing technique used a fix spiral transposition, so there was no flexibility. Once an attacker came to know about the technique used for encrypting the whole data, he could recover all the data. Various attacks can lead data access to unauthorized people. The limitations of the existing system are:

- Spiral transposition, trivially broken for known text attract and easily broken for cipher only attack.
- Transposition does not substitute one symbol by another, instead it changes the location of the symbol.
- The main disadvantage of this approach is that it uses a fixed spiral therefore there is no flexibility in this approach.

MOTIVATION

The motivation behind this project is to develop a highly secure cryptographic system for protecting data in cloud environments. Traditional encryption method face challenges such as vulnerability to brute force attacks and emerging quantum threats. By utilizing DNA- based cryptography, this approach leverages the complexity and randomness of nucleotide sequences, ensuring enhanced security. The integration of the hybrid encryption techniques includes methods like substitution, complementary rules, and color palette compression, strengthens data protection. This method not only ensures confidentiality and integrity but also offers scalability and adaptability for future advancements. The biological foundation of DNA cryptography makes unauthorized decryption extremely difficult, ensuring robust security.

PROPOSED SYSTEM

The encryption process begins by converting the source data into ASCII format, followed by transformation into binary code. This binary sequence is then mapped onto DNA bases using a predefined system, where Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) correspond to binary pairs (A = 00, C = 01, G = 10, T = 11). DNA sequencing is the process of determining the order of Nucleotides in DNA. The principle of complementary base pairing-where Adenine pairs with Thymine, and Cytosine pairs with Guanine which ensures the accurate replication and sequence identification. Hybrid Encryption method incorporating Substitution and Complementary Rule techniques enhances complexity, generating ciphertext. The encrypted data is compressed into color palettes for added obfuscation. A randomly generated key further strengthens security. Decryption reverses these steps, restoring the original data securely. The 24 set Complementary plays a vital role in securing data by the sequence that is similarly created by using DNA Cryptography Sequence.

ENCRYPTION AND DECRYPTION MODULE

The System for DNA-based Encryption consists of three Key Methods: Encryption, Random Key Generation, and Decryption. The Encryption module, the original data is first converted to ASCII and then to Binary Code.

The raw data is converted to ASCII and then binary code in the encryption module. A mapping mechanism is used to convert the binary information into DNA bases, with A = 00, C = 01, G = 10, and T = 11.

In Hybrid Encryption function, it combines the Substitution and the Complementary Rule Techniques, it creates the Cipher text. The Encrypted text is compressed into Color Palettes for storage. A Random key is generated for an additional Encryption layer. In Decryption module, the process is reversed, restoring the original data. It contains different DNA modules under four types that is discussed below.

The DNA Cryptography has classified modules. They are:

- Encryption using Hybrid method.
- Generation of Color Palettes from the Encrypted DNA sequence.
- Decryption of Data Packets by Hybrid Technique.

MODULE 1

ENCRYPTION USING HYBRID METHOD

Encryption module converts raw sequence of text into Cipher Text. It consists of three sub divisions

BINARY CODE CONVERSION

It initially converts the raw text or the input data into its equivalent binary values. The plain text FRUIT is given as the input which is converted into their corresponding ASCII values as 70114117105116 and subsequent binary code which is equal to (01000110 01110010 01110101 01101001 011100100) is generated.

COMPLEMENTARY PAIR AND SUBSTITUTION METHOD

The Binary code generated in the above module is converted into DNA sequences using the DNA base pair coding rule i.e., (A= 00, C = 01, G = 10, T = 11). The DNA sequences are then Hybridized using Complementary Pair and the Substitution Method. Finally fake DNA sequence is induced.

COMPLEMENTARY BASE PAIRS RULE SET

DNA sequencing is defined as the process of determining the order of Nucleotides in DNA. The principle of complementary base pairing is where,

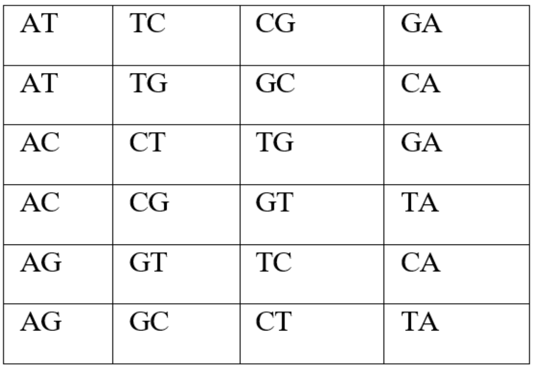
A (Adenine) pairs with **T** (Thymine)

C (Cytosine) pairs with **G** (Guanine)

Which ensures accurate Replication and Sequence Identification.

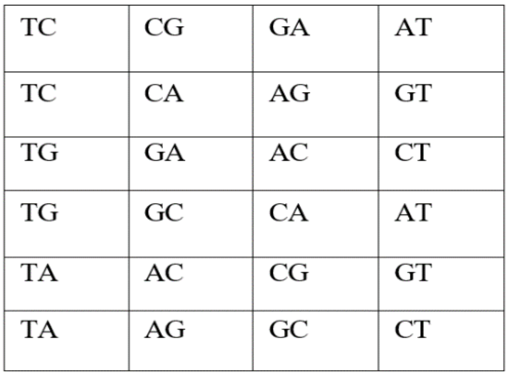
A→T→C→G

(C)



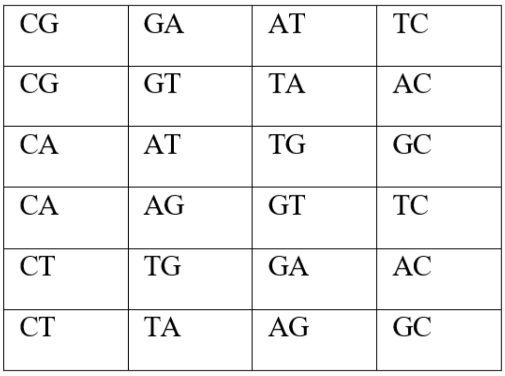
AT	TC	CG	GA
AT	TG	GC	CA
AC	CT	TG	GA
AC	CG	GT	TA
AG	GT	TC	CA
AG	GC	CT	TA

(C')



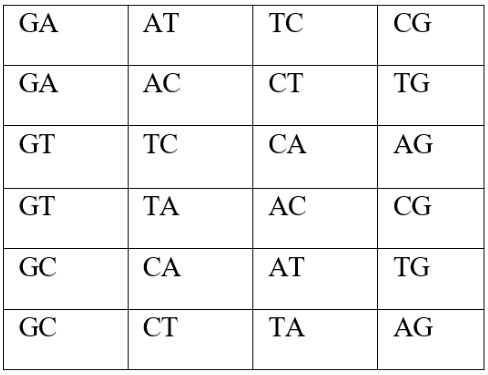
TC	CG	GA	AT
TC	CA	AG	GT
TG	GA	AC	CT
TG	GC	CA	AT
TA	AC	CG	GT
TA	AG	GC	CT

(C")



CG	GA	AT	TC
CG	GT	TA	AC
CA	AT	TG	GC
CA	AG	GT	TC
CT	TG	GA	AC
CT	TA	AG	GC

(C"')



GA	AT	TC	CG
GA	AC	CT	TG
GT	TC	CA	AG
GT	TA	AC	CG
GC	CA	AT	TG
GC	CT	TA	AG

Table 1.1 24 Set Complementary Rules

MODULE 2

GENERATION OF COLOR PALETTES

The codons (sequences of three Nucleotides in DNA) from the obtained DNA sequence. Then, it is converted to cipher text based on codebook. Cipher text will be unrecognized form since each character is represented as a group of a letter and a number. Finally, it converts these characters in Cipher Text to colors. This will improve the security because same character will be represented by different colors at different stages of encryption. The fake DNA sequence (AGCTACTTAG) generated in the initial sub module is converted into Color Palettes through ASCII values which is the required Cipher text.

MODULE 3

DECRYPTION OF DATA PACKETS BY HYBRID TECHNIQUE

The Decryption process is essentially the reverse of the Encryption process, where the encrypted ciphertext (such as color palettes or encoded data) is transformed back into its original, readable plain text format, restoring the information to its understandable state.

EXTRACTING FAKE DNA SEQUENCE

Initially the Color Palettes is converted into fake DNA sequences through their ASCII values. There are more than 10 million colors that we are able to see, 16.8 million colors can be displayed on a computer screen. These unique colors can be easily represented as corresponding ASCII values which can be utilized in our algorithm. RGB stands for the primary colors Red, Green and Blue. As explained, we can see millions of colors in the universe. The color appearance is obviously affected by many other conditions like lighting, brightness, other colors in the surroundings etc.

DECRYPTION OF ORIGINAL TEXT

Once the transformed DNA sequence is obtained, it is converted into binary code by applying a predefined DNA-to-binary rule, where each Nucleotide (A, T, C, G) is mapped to a specific binary value. The binary sequence is then translated into ASCII characters, the standard encoding for text, where each 8-bit segment of binary code corresponds to a specific character. Finally, after all the binary data is converted into ASCII characters, the original plain text is recovered. The Process relies on a Hybrid Method that integrates both the complementary base pairing rule and the substitution cipher to encrypt and decrypt the data. By combining these two methods, the encryption is strengthened, ensuring that the final output is the plain text that was originally encoded in the DNA sequence.

The Use of ASCII encoding ensures seamless integration with text data standards, making the method efficient for text-based data processing. The Complementary base pairing rule inherently supports Error Detection and Correction, enhancing the reliability of the decryption process.

ENCRYPTION METHODOLOGIES

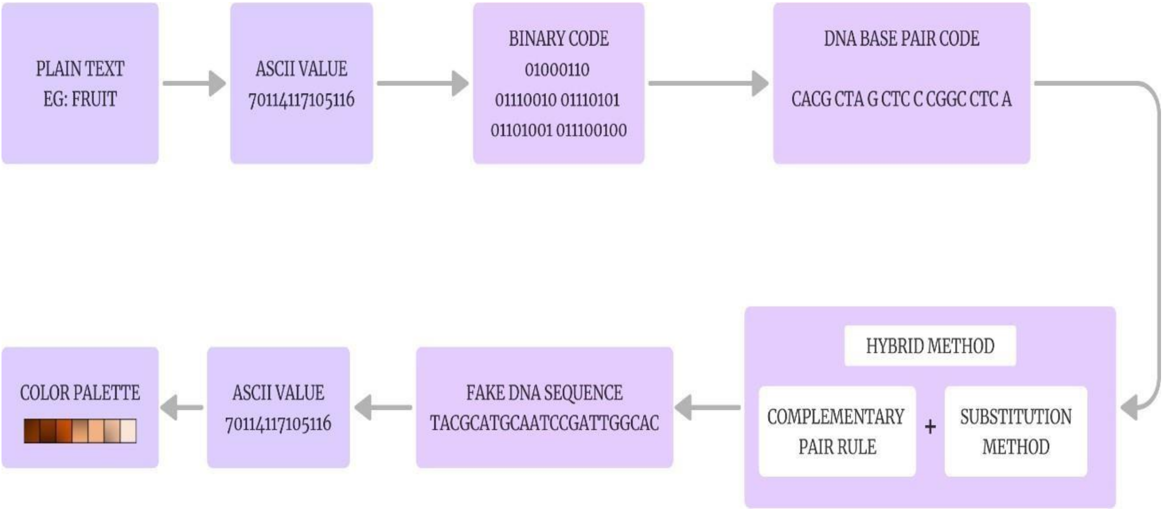


Figure 1.1 Encrypting stages of DNA Sequence

DECRYPTION METHODOLOGIES

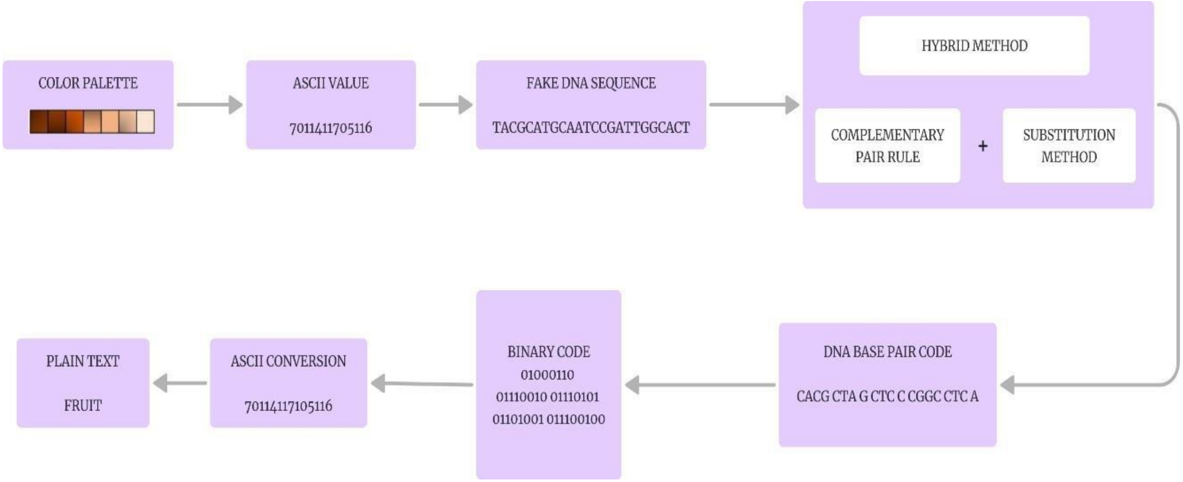


Figure 1.2 Decrypting stages of DNA Sequence

MODEL PERFORMANCE

The proposed encryption model leverages Hybrid Cryptographic Models, integrating DNA Cryptography by **substitution, complementary and 24 rule set** to enhance security.

Numerical Representation System

A DNA sequence consists of four nucleotide bases: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). These can be mapped to numerical values for computational processing:

Adenine (A) → 0

Cytosine (C) → 2

Guanine (G) → 1

Thymine (T) → 3

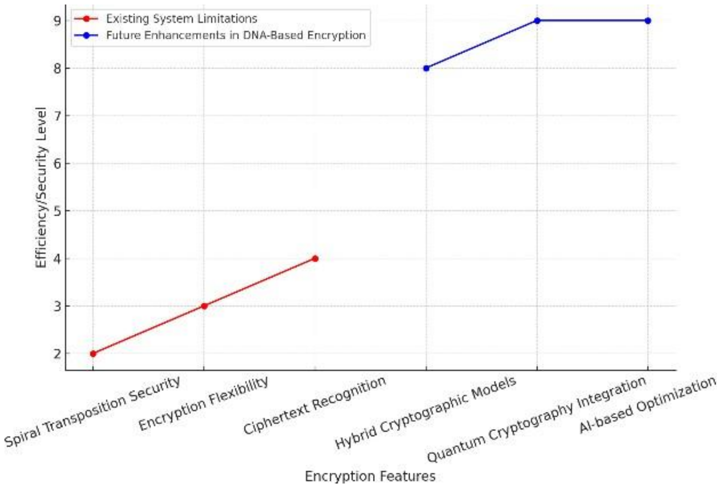


Figure 1.3 Efficiency Chart of DNA Sequence

PERFORMANCE METRICS

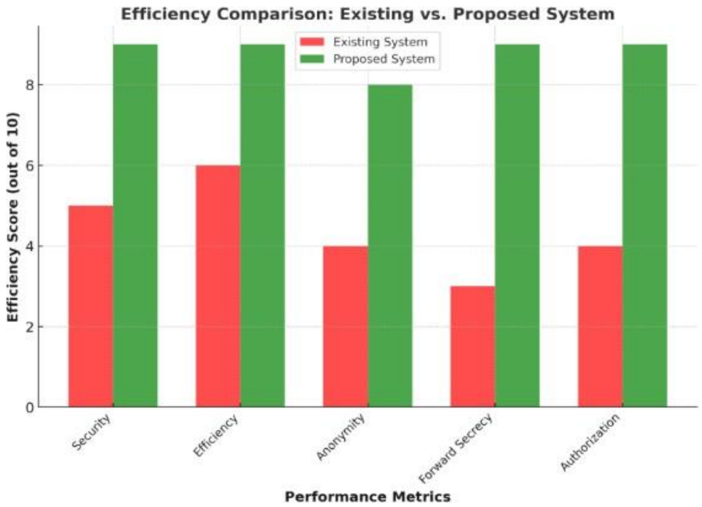


Figure 1.4 Performance Metric of DNA Sequence

FUTURE ENHANCEMENT

Future Enhancement Summary

Objective: Enhance security, scalability, and efficiency through optimized encoding and improved compression techniques.

Key Developments

HYBRID CRYPTOGRAPHIC MODEL

Combines DNA cryptography with RSA and AES for multi- layered encryption.

DNA cryptography increases complexity by leveraging biological sequence encoding.

RSA ensures secure asymmetric key distribution.

AES provides fast and strong symmetric encryption.

Ideal for cloud computing, healthcare, and secure communication.

INTEGRATION WITH QUANTUM CRYPTOGRAPHY

Combines DNA cryptography with Quantum Key Distribution (QKD) for enhanced protection.

QKD prevents eavesdropping and withstands computational attacks.

DNA-based encryption adds an additional layer of security.

Strengthens defense against both classical and quantum cyber threats.

FLOW CHART

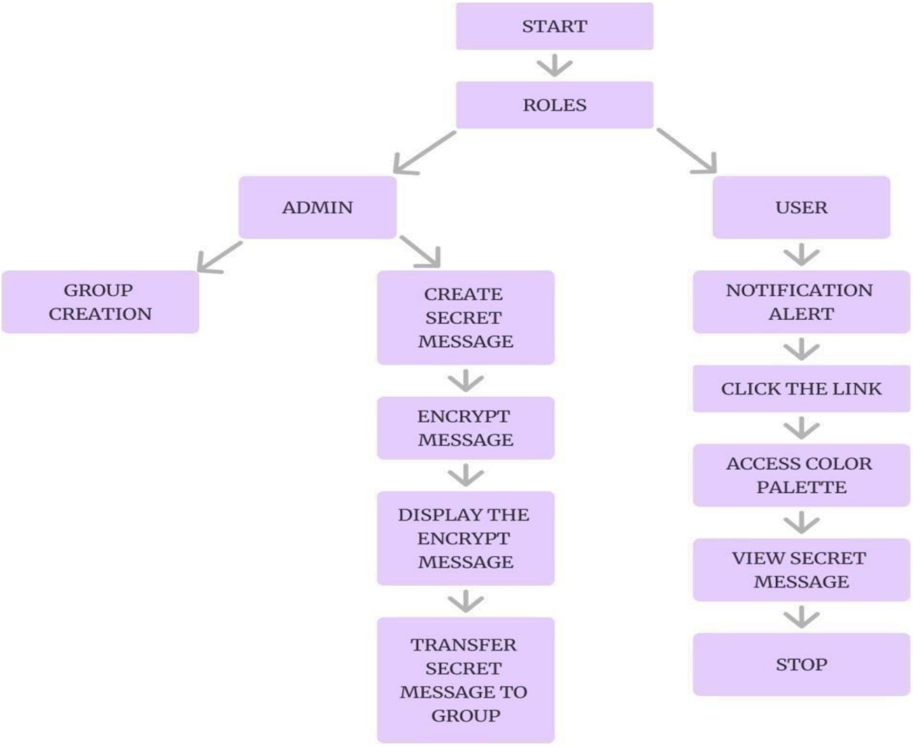


Figure 1.5 Flow Chart of DNA Sequence

AI-BASED OPTIMIZATION

AI improves DNA cryptographic efficiency for real-time cloud applications.

Enhances DNA sequence encoding by reducing redundancy and maximizing efficiency.

Automates binary-to-DNA mapping for faster encryption and decryption.

OUTCOME

This enhancement fortifies secure cloud data transfer, ensuring resilience against evolving cyber threats through advanced cryptographic techniques.

CONCLUSION

The future of DNA-based encryption in cloud computing lies in the integration of hybrid cryptographic models, quantum cryptography, and artificial intelligence- driven optimization. These advancements will significantly enhance security, scalability, and efficiency, making DNA cryptography a viable solution for secure data transfer in modern cloud environments. These techniques will continue to be refined through ongoing research and development, ensuring their practical application across various domains requiring robust data protection.

ACKNOWLEDGEMENT

We express our sincere gratitude to Teacher, Trainer, Transformer, Trends Setter, Total Reformer. J. Arawindhan, Chief Mentor of Achariya Educational Trust, for providing an environment conducive to completing this project. We are thankful to our Principal, Dr. S. Gurulingam, for his encouragement, and to Dr. A. Kannaki @ VasanthaAzhagu, Head of the Department, for permitting us to undertake this work. Special thanks to our guide, Dr. A. Kannaki @ VasanthaAzhagu, for her invaluable guidance and support. We also appreciate the assistance of Dr. S. S. Boomiga, our project coordinator, and the teaching and non-teaching staff. Finally, we thank our parents and friends for their unwavering support.

References

1. Kyelim Lee et al., 15 May 2024, Digital Object Identifier Biological Moleculessss 10.1109/ACCESS.2024.3401464
2. Needleman-Wunsch Attention: A Framework for Enhancing DNA Sequence Embedding.
3. Maria Fernandes et al., VOL. 24, NO. 3, MARCH 2020, IEEE, DNA- SeAl:
4. Sensitivity Levels to Optimize the Performance of Privacy-Preserving DNA Alignment.
5. Ilan Shomorony et al., IEEE Transactions on Information Theory, Vol. 67, No. 6, June 2021 DNA-Based Storage: Models and Fundamental Limits.
6. Xingyuan Wang et al., date of publication April 9, 2020, Digital Object Identifier10.1109/ACCESS.2020.2986831,
7. IEEE, Image Encryption Based on Hash Table Scrambling and DNA Substitution.
8. Milena M. Arruda et al., date of publication May 6, 2021, Digital Object Identifier10.1109/ACCESS.2021.3078138,
9. IEEE, Is BCH Code Useful to DNA Classifications an Alignment- Free Method.

10. Jiyun Zhou et al., VOL. 17, NO. 1, January/February 2020, IEEE, Prediction of DNA-Binding Residue from Protein Sequence by Combining Long Short- Term Memory and Ensemble Learning.
11. NA Deng et al., August 5, 2019, Digital Object Identifier is the FP- Algorithm 10.1109/ACCESS.2019.2933044 Frequent Patterns Mining in DNA Sequence.
12. Inbal Preuss et al., VOL. 10, NO. 2, JUNE 2024, IEEE, Sequencing Coverage Analysis for commercial DNA- Based Storage Systems.
13. Zeina Salman et al., was published in the Open Journal of Nanotechnology, Digital Object Identifier 10.1109/ OJNANO.2024.3451954
14. Simulation of the Interaction between DNA Nucleotides and One- Dimensional Carbon Chain.
15. YUNFEI GAO et al., date of publication 12 January 2024, Digital Object Identifier 10.1109/ACCESS.2024.3353305
16. IEEE, Adaptable DNA Storage Coding: An Efficient Framework for Homopolymer Constraint Transitions.