Chapter

DNA Computing Using Cryptographic and Steganographic Strategies

Adithya B. and Santhi G.

Abstract

Information protection and secrecy are major concerns, especially regarding the internet's rapid growth and widespread usage. Unauthorized database access is becoming more common and is being combated using a variety of encrypted communication methods, such as encryption and data hiding. DNA cryptography and steganography are used as carriers by utilizing the bio-molecular computing properties that have become more common in recent years. This study examines recently published DNA steganography algorithms, which use DNA to encrypt confidential data transmitted through an insecure communication channel. Several DNA-based steganography strategies will be addressed, with a focus on the algorithm's advantages and drawbacks. Probability cracking, blindness, double layer of security, and other considerations are used to compare steganography algorithms. This research would help and create more effective and accurate DNA steganography strategies in the future.

Keywords: DNA, Cryptography, Steganography, Bio-Molecular, DNA Computing

1. Introduction

The concept of security refers to the prevention of unauthorized access to information. In today's computer science, encryption's primary goal is to prevent confidential data from being altered, lost, hacked, or compromised by a third party [1]. Encryption and concealment of information are among the most widely used methods in networking and information security. Encryption and concealment of information (both similar concepts) are commonly used to keep communications secure [2, 3] fact that both methods have the same purpose. Still, their development and use are vastly different. Cryptography alters the sense of coded writing, while steganography is a covert way of writing that conceals the encrypted message's nature. Thus, in data transmission through an insecure public medium, the science of steganography is more reliable, necessary and often preferred over encryption [4, 5].

Various steganography systems, as well as their criteria, are discussed in this article based on the literature. Different systems use different strategies for embedding data, each with a set of benchmarks to evaluate performance and determine its advantages and disadvantages. Vulnerability to adversary attack is one of the three common criteria. To avoid arousing suspicion, the embedded data must be kept undetectable both visually and statistically. A fully reliable system with comparable carrier and stego file statistics should be considered during the message embedding process [5, 6]. The carrier's power, known as the amount of data concealed within it, is the second common prerequisite. The development of a steganography technique could allow more sensitive data to be hidden within the carrier while maintaining the properties of the stego file [1, 5]. A successful steganography strategy should keep enough information in its embedding capability [6]. Imperceptibility is the third common prerequisite, which is characterized as having a high embedding potential and the ability to resist intruders. The stego carrier should ideally be devoid of visual artifacts and the greater the stego carrier's fidelity should be better [2].

The masking theory is typically modeled by a pair of algorithms: embedding and extraction, as seen in **Figure 1**. The embedding algorithm produces a stego file containing the private data by merging two folders, secret and vector data, with an optional key. On the other hand, the extraction algorithm is used to recover the secret data from the stego file [7]. Steganography is a method of concealing data that does not require the use of a key. Its protection depends on the privacy of the algorithm. As a result, it is known as a less reliable approach [8, 9]. Another way to hide information is to hide confidential data, which uses one key for all operations (embedding and extraction). One of the most important benefits of this type is its rapid stage in all procedures [10, 11]. Unlike previous patterns, public steganography uses two keys for embedding and extraction: embedding and the other for extracting. The biggest value of this type is the durability of the system. The identification of the other key could be a concern if one of the keys is identified by a third party [10, 12]. On the other hand, this model is 100–1000 times slower than private steganography [13].

Several applications represent a container for confidential data. In steganography schemes, these programs are used as cover objects or carriers. Per carrier has its own set of characteristics that aid in the data concealment process. The carrier's field availability determines the amount of confidential information needed to hide

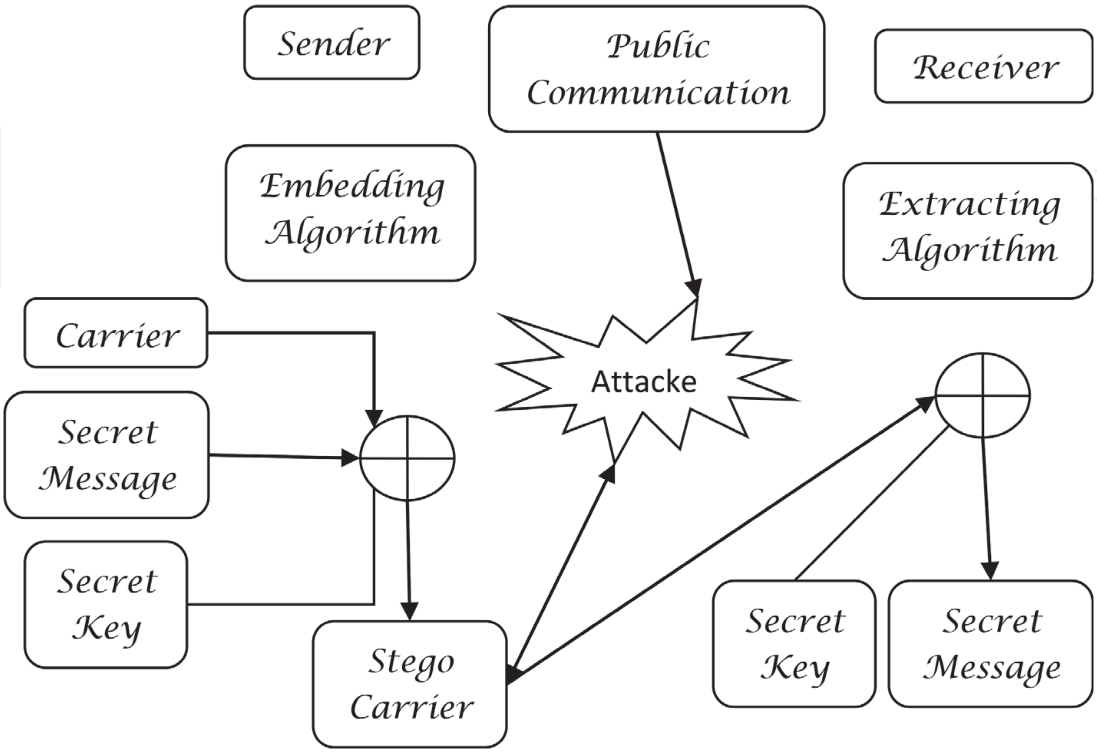


Figure 1.
Block diagram of steganography system.

data within each carrier. Text, audio, video, and photographs are examples of multimedia used to hide records. Text can be obscured by changing the text's layout, inserting an nth character from the text, or changing any of the rules, such as spacing. Text can also be hidden using a code made up of letters, lines, and page numbers. However, this process is insecure [2]. The biggest benefit of this carrier is that it does not take a lot of memory and is quick to switch.

In contrast to other carriers, it has a very limited number of redundant data [10, 14]. The use of inaudible frequencies and a small shift in the binary sequence of an audio file can be used to hide data in audio files [2, 15]. Data masking in video files is more efficient and effective due to the wide available space. Allowing data to be hidden within multiple video frames [16]. Uncompressed and compressed video are the two main formats of video in which data can be hidden. Digital images have been common carriers for masking confidential information due to their high redundancy, high capacity in images, low effect on exposure, and ease of manipulation [15, 17]. DNA is a relatively recent vector that has been used in the field of steganography. In this article, we look at the data hidden in DNA.

2. Deoxyribonucleic acid (DNA)

The most important molecular structure in biology is deoxyribonucleic acid (DNA), which encodes the information required to generate and direct all chemical elements in the human body. As a result, DNA has been suggested as a possible candidate for computational purposes [18].

2.1 DNA structure

DNA is described as a living creature's genetic blueprint. Each body cell has its DNA collection and a polymer made up of monomers called deoxyribose nucleotides, consisting of three components, as seen in **Figure 2** [19].

The human body is made up of trillions of cells, each with its purpose. As seen in **Figure 3**, each cell has a nucleus that comprises several chromosomes. The majority of DNA is present in a nucleus, which is known as nucleus DNA, and the remainder is found in mitochondria, which is known as mitochondria DNA (mtDNA). Each cell's activity is regulated by DNA. DNA chromosome is made up of a DNA molecule of genes. A gene is the entire genetic makeup of an organism, containing information from all chromosomes [20].

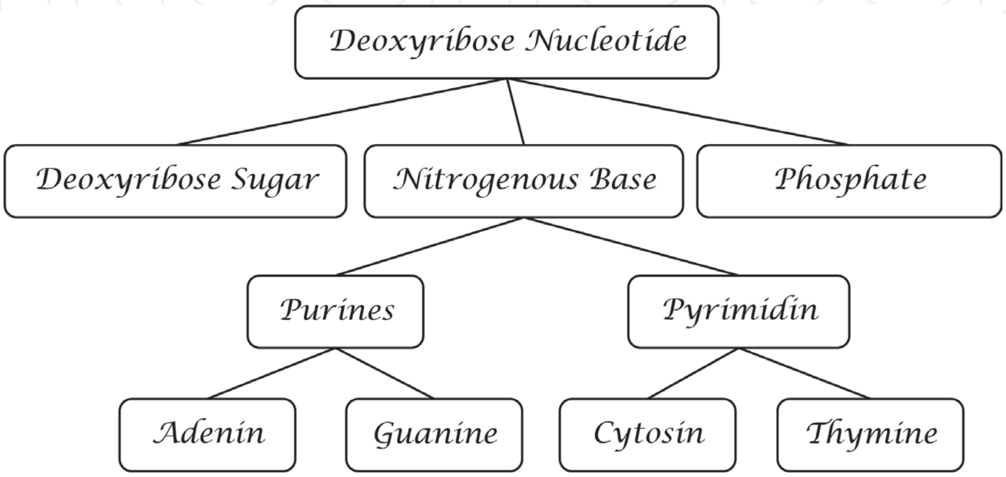


Figure 2.
Structure of deoxyribonucleic acid.

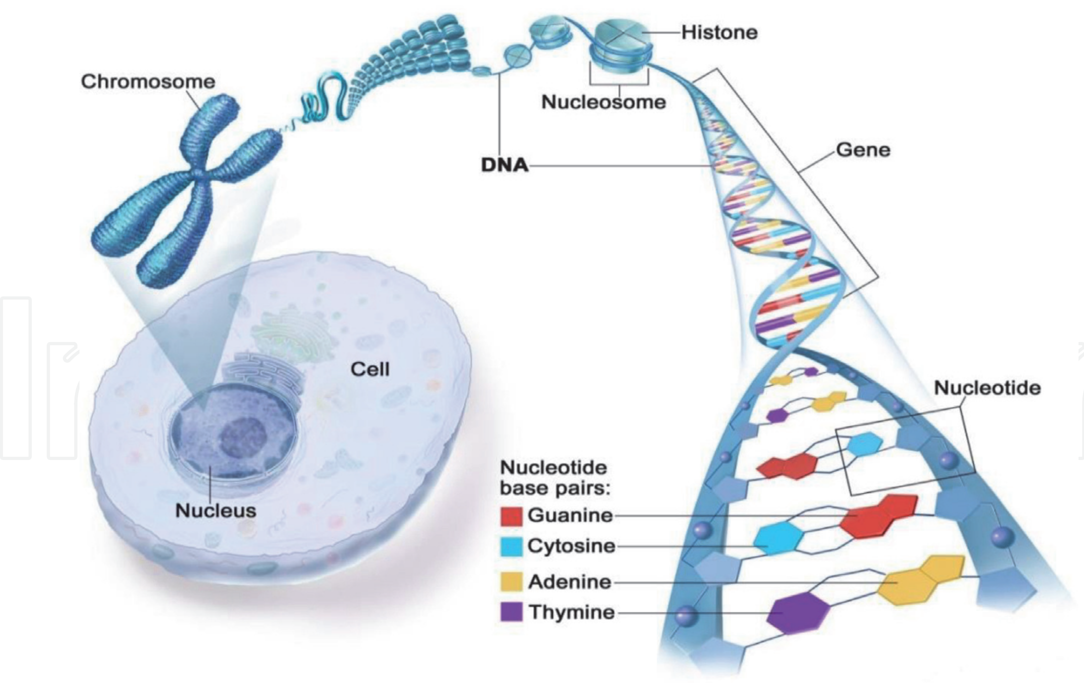


Figure 3.
Gene development cycle.

In 1953, Watson and Crick discovered DNA structure, a form of genetic material. DNA is a long molecule present in all living things' body cells. DNA is a kind of bacterial plasma that contains all lifestyles. It is made up of two simple bands that are twisted around each other in a double helix (see **Figure 4**). Each DNA chain is made up of nucleotides, which are small subunits. The four chemical bases in the chain DNA are Adenine (A), Thymine (T), Guanine (G) and Cytosine (C), which bind to sugar and phosphates in the backbone to complete the nucleotide. Purines (A and G) and Pyrimidine (T and C) are the two DNA bases in biology. Continuously (A) is bound to (T) by two hydrogen bonds, and (C) is bound to (G) by three hydrogen bonds [19, 21]. Transcription is the method for producing RNA, which is an intermediate copy of DNA instructions. Adenine (A), Cytosine (C), Uracil (U), and Guanine (G) are the four bases that makeup RNA. All 64 codons are represented in **Figure 5**. The STOP codons do not necessarily symbolize any amino acids but rather indicate the protein chain's end. The twenty amino acids are determined by the remaining 61 codons. Some amino acids are coded by several codons [11]. As a result of this codon duplication, it is possible to change the genetic sequence while keeping it functional [11, 23, 24].

2.2 DNA computing

Currently, biology methods are used in a variety of fields. DNA is a relatively new biological technology that is used in a variety of applications [25]. This is because DNA computing can solve a variety of NP-complete problems, in which the computation time increases dramatically.

There has been a considerable amount of research in this field, with significant progress made on DNA and the immune system [19]. Leonard Adelman conducted the first experiment in DNA computing (bio-molecular computing) in 1994, in which molecular biology instruments were used to solve a portion of the standard path of the Hamiltonian puzzle. Computing with molecules directly was discovered at the time, and it was regarded as a new discipline in terms of science defense [26]. The satisfaction problem (SAT), an NP-complete problem, was solved using DNA

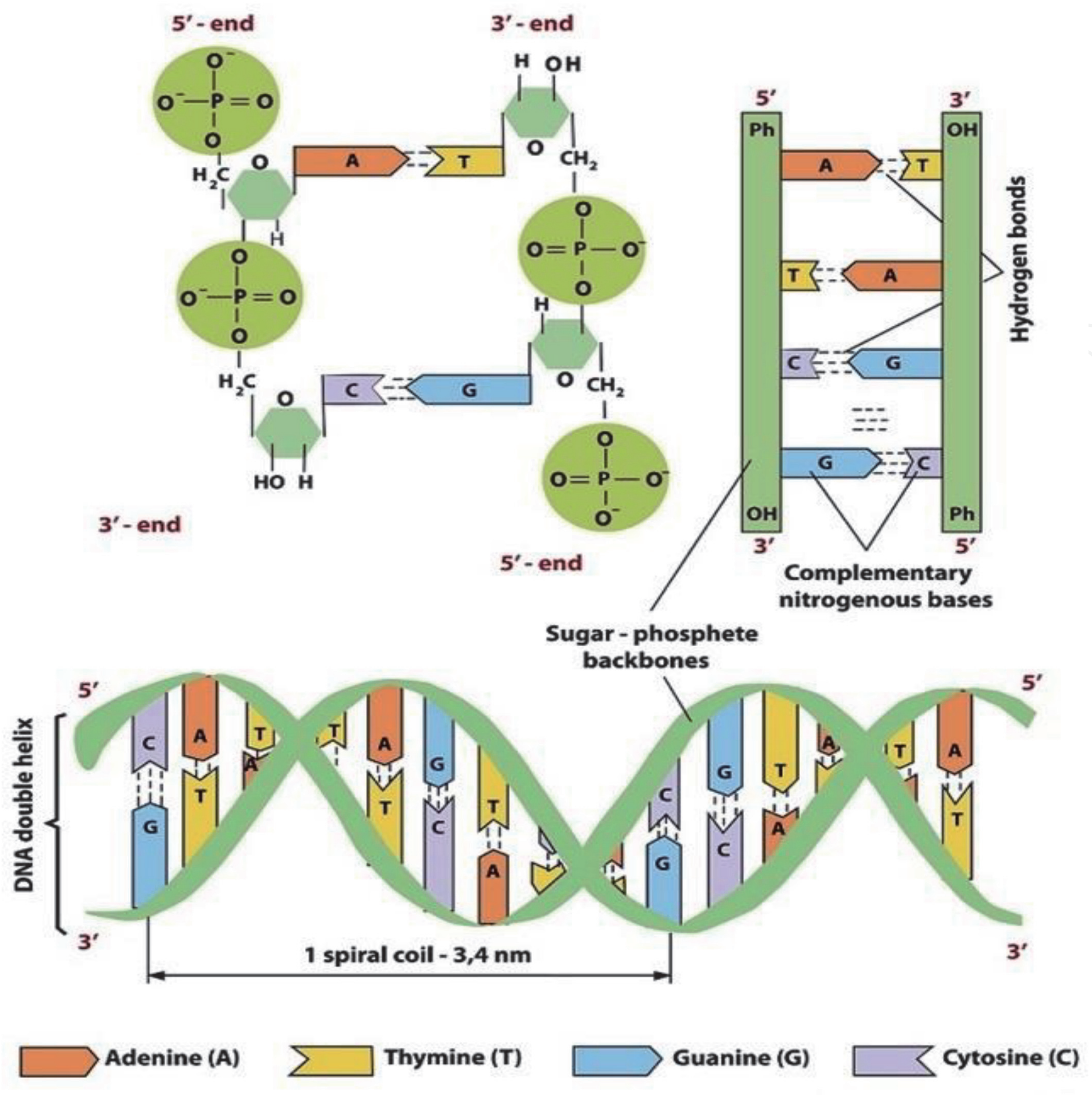
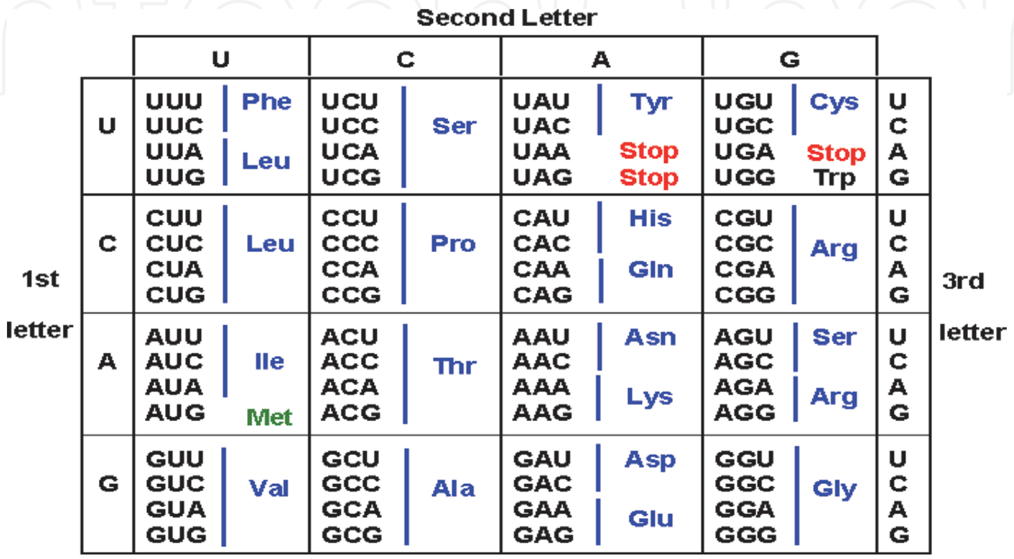


Figure 4.
Helical structure of DNA [20].



1st letter		Second Letter U		Second Letter C		Second Letter A		Second Letter G		3rd letter
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Figure 5.
Codon and amino acid table [22].

DNA base	Binary code
A	00
C	01
G	10
T	11

Table 1.
Binary code of DNA.

computing in a 1995 study by Lipton. The offered approach took advantage of DNA's parallelism and its computational and storage capacities [19]. In 1997, Ogihara and Ray discovered that DNA could be used to simulate AND and OR gates [27]. Clelland [28] proposed the first successful experiment of a DNA steganography technique for concealing sensitive data using DNA microdots.

2.3 Binary code of DNA

A, C, G, and T are the four chemical bases that make up each DNA chain. A is biologically related to T, while C is related to G. T The synthesis of DNA rules can be modified in binary arithmetic by changing input judgments, such as assuming that T is related to C or T is related to G [29]. Researchers would use a binary encoding rule to translate a hidden message into DNA rules before mixing it with sequenced DNA to store data in DNA particles. For each rule (A), researchers may use the corresponding binary form: binary formulas can be "00," "01," "10," or "11." This can be expressed as in **Table 1**. The encoding of DNA and its random properties make it an ideal candidate for both coding and coding. As a result, converting DNA into the binary form will result in 4! = 24 different encoding methods [30, 31]. On DNA bases, logical operations such as addition, subtraction, XOR, AND, OR, and NOT are possible.

3. Comparative study

The aim of the comparison presented in this study is to ensure that researchers are aware of the shortcomings in current steganography systems, thus inspiring future advances in this field. **Table 2** compares the strengths and disadvantages of existing algorithms in terms of security problems such as chance of intrusion, double security layer, blindness, and more.

The derived comparison in **Table 2** aims to clarify the proposed DNA's strengths and weaknesses using data hiding algorithms. Encrypting sensitive data into encryption data before embedding, rather than including the initial data format, improves confidentiality [13, 18, 23, 34, 36, 38, 41, 44, 45, 47–49, 51, 52, 55, 57–60]. Playfair technology, adopted in [58], is the most promising encryption technology combined with DNA-based data masking technology. A thorough comparison of several encryption methods, including vigenere and Playfair, AES, and RSA ciphers, has been done in their work. Any of them was paired with a replacement tool for hiding data in DNA. The findings revealed that the Playfair cipher is not only quick and easy to use, but it also has a high level of protection and ability.

The blindness trait, which eliminates the need to give the original DNA connection to the recipient, is the primary function supported by DNA-based data masking techniques. The main goal of the blindness feature is to improve protection and avoid any intruder way of detecting it, as shown in [11, 18, 25, 35, 41, 43, 48, 49, 51, 57, 58, 62].

S.No	Reference	Strengths	Weaknesses
1	[24]	*Insertion Technique* • High embedding capacity. • Simple to bring into practice. • Modification rate is low.	• Length of Stego DNA is longer than length DNA of comparison. • The payload does not equal zero. • In the extraction process, multiple data is needed. • The amino acid function is not preserved. • The algorithm is not blind. • Increase the level of redundancy • Steganography method for purely obscuring results.
		Complementary Technique • Simple to bring into practice. • To break the hidden data, attackers must have a ton of information.	• The payload does not equal zero. • Modification rate is high. • The algorithm is not blind. • Steganography method for purely obscuring results. • After the embedding process, the length of DNA is modified.
		Substitution Technique • High embedding capacity. • Simple to bring into practice. • The payload is set to zero. • In contrast to previous approaches, this one is more efficient, dynamic, and performs better.	• The amino acid function is not preserved. • The algorithm is not blind. • Steganography method for purely obscuring results. • Modification rate is high.
2	Ref [21]	• The payload is set to zero. • High embedding capacity. • Simple to bring into practice. • Maintain the biological DNA's features. • To increase the degree of secrecy and complexity, the consequence of hiding data in the cloud is being implemented.	• The DNA reference determines the level of security. • Increase the size of the message. • The algorithm is not blind. • Steganography method for purely obscuring results.
3	Ref [25]	• Build a steganography method that is reversible. • Preserve the DNA's versatility. • The algorithm is blind. • A secret key is employed.	Does not encrypt confidential information when storing it.
4	Ref [18]	• To provide security, a map was created between DNA codons and amino acids. • Before hiding, use the playfair cipher to encrypt the hidden letter. • Improve the playfair cipher by changing it to 5*5 to prevent its pitfalls, such as the diagraphs and hidden text form remaining after encryption. • Adding a second layer of protection. • Algorithm for the blind. • Capacity and time efficiency are also improved. • Provide a high risk of cracking. • It is necessary to use a hidden key.	• Increase the length of the stego DNA. • The biological DNA's versatility is not preserved. • It must send many data to the recipient in order to retrieve the hidden message from Stego DNA. • The payload is not empty.
5	Ref [32]	• The usefulness of the initial replacement process has been improved.	• The biological DNA's versatility is not preserved.

S.No	Reference	Strengths	Weaknesses
		• The communication performance of a data hiding device on the internet can be enhanced. • In terms of power and protection, providing better results. • TLSM has been enhanced to allow secret data to be hidden in any series of letters or symbols. • The Base-t TLSM and the Extended TLSM (ETLSM) are two methods proposed to increase the efficiency of the TLSM. • Capacity has been expanded.	• It needs to submit multiple data, including DNA reference, Stego DNA, secret message site collection, table code, to extract the secret message from Stego DNA. • Modification rate is high. • The algorithm is not blind. • Steganography method for purely obscuring results.
6	Ref [13]	• Proposed a protocol for masking encrypted data to limit the use of public keys while maintaining the highest level of reliability. • The payload is set to zero. • A wide embedding capacity. • Using the cutting-edge technology of DNA data hiding, the secret key is hidden inside the DNA reference for added confidentiality.	• The biological DNA's versatility is not preserved. • The algorithm is not blind.
7	Ref [33]	• If the length of stego DNA is not extended, the payload is zero. • Algorithm is simple. • The ability to cover has been enhanced. Reduce the pace of modification. • In hiding, substitution form is used.	• The biological DNA's versatility is not preserved. • If the DNA comparison includes a number of repeated nucleotides, the modification rate would be high. • Both the sender and the receiver should be aware of the un-blind algorithm, as well as injective mapping and complementary rules. • Algorithm for simply obfuscating results.
8	Ref [34]	• Flexible algorithm that is easy to execute. • Encrypt a hidden message using a revamped Playfair algorithm that incorporates DNA and amino acids. • After the hiding process, the length of DNA does not extend. • It is necessary to use a hidden key. • In hiding, substitution form is used.	• The biological DNA's versatility is not preserved. • Algorithm that is not blind.
9	Ref [35]	• The algorithm employs three keys. • In terms of modification volume, the first and third techniques of Ref [24] have been improved. • The stego DNA is not expanded. • The algorithm is blind.	• There is no encryption method used. • Only nucleotides with marks equal to zeros after conversion to binary are used to hide hidden records.
10	Ref [36]	• It's easy to bring into effect. • Low rate of modification. • The length of stego DNA is not increased. • To encrypt hidden data before hiding it, one of the most efficient encryption techniques (RSA) is used. • A public key is employed.	• The biological DNA's versatility is not preserved. • Save the location of each DNA base that contains the hidden data and submit it to the receiver for extraction. • The hidden data's size has been expanded. • Algorithm that is not blind.

S.No	Reference	Strengths	Weaknesses
			• Cracking with a low probability
11	Ref [36]	• High embedding capacity. • Simple to bring into practice.	• The biological DNA's versatility is not preserved. • Algorithm that is not blind. • Algorithm for simply obfuscating results. • Cracking with a low probability
12	Ref [11]	• It's easy to bring into effect. • Ensure the biological DNA's functionality is maintained. • Low rate of modification. • The algorithm is blind. • The secret key is hidden in the DNA guide, which adds to the protection. • After hiding sensitive details, the DNA reference is not extended.	• Due to the use of LSB in the hiding operation, the potential is low. • There was no encryption on the confidential data until it was hidden. • Cracking with a low probability.
13	Ref [23]	• Exhibit DNA amino acids to encrypt hidden records. • Before hiding the secret key inside the DNA reference, encrypt it using the RSA algorithm. • The public key is used, and the capability is high. • Cracking with a high probability.	• Algorithm that is not blind. • A high degree of modification. • The payload is not empty. • The versatility of amino acids is not maintained.
14	Ref [37]	• Preservation of protein translation in the protein coding DNA (PcDNA). • Data encoding is consistent and near optimal. • Keep track of the codon statistics. • Embedding data came close to being perfect. • Embedding data in DNA in a reliable and effective manner. • A secret key is employed.	• Estimation that is difficult. • Unconstrained ncDNA hiding can be estimated by intruders.
15	Ref [38]	• Before using the Playfair algorithm to hide hidden data, encrypt it. • High-level surveillance. • Since hiding in an audio at the last stage would not draw attackers. • Hide the secret data and translate it into an audio file so that it is impossible to show that all data is inside the audio. • Provide two layers of concealment. • A secret key is employed.	• The hidden data must be extracted using several data sources. • The algorithm is not blind.
16	Ref [38]	• Key area is wide enough to resist negative intruders using brute force. • Before hiding secret data in host text, encrypt it. • The algorithm is blind. • The embedding power ratio is 100 percent. • Provide two layers of concealment. • Chebyshev maps are used to establish DNA references. • In hiding, the substitution method is used.	• Calculation is difficult.

S.No	Reference	Strengths	Weaknesses
17	Ref [39]	• Ref [40] algorithm's hidden key was modified to use the secret key. As well as keeping all of Ref [40] high points.	• Pure steganography algorithm. • Complex calculation.
18	Ref [41]	• The initial replacement technique's capability and protection have been increased. • The algorithm is blind. • Method of replacement has been improved.	• Pure steganography algorithm. • The biological DNA's versatility is not preserved. • If multiplied by 6, if the result is not equal to zero, additional zeros are added. • The length of Stego DNA is extended.
19	Ref [42]	• High embedding capacity. • Simple to bring into practice. • Secret data is sent in the (ABCD) format.	• Pure steganography algorithm. • Cracking with a low probability • Algorithm is not blind. • The receiver should obtain a random DNA sequence and a complementary pair law. • There is no encryption on the data until it is embedded. • Cracking with a low probability • Steganography method for purely obscuring results.
20	Ref [43]	• Only the correct value of Stego DNA is sent to the recipient. • High level protection. • Hackers have a tough time spotting the seeds of the random numbers generated. • Hackers have a hard time deciding how many packets to split, in addition to the number of DNA message bits and binary in each packet. • The secret message bits and DNA comparison bits are randomly combined. • The algorithm is blind. • A secret key is employed. • Cracking with a high probability	• Redundancy has been increased. • The message size has been increased. • The DNA functionality is not preserved. • Increase the size of stego DNA.
21	Ref [44]	• A secret key is employed. • Until hiding a secret document, encrypt it with RC4. • Exceptional ability. • Providing a safe environment. • Provide two layers of concealment. • Build DNA from a picture.	• During the extraction process, the algorithm needs several keys.
22	Ref [45]	• A secret key is employed. • Classified data protection has increased dramatically. • Extra grids of different sizes may be used to store additional data. • BASE64 encoding is used to encrypt confidential info. • Provide two layers of concealment. • Secret text is used to build DNA.	• Complex calculation.
23	Ref [46]	• A secret key is employed. • High levels of protection. • High capacity. • Since the key of prime duration is between 20 and 40, the possible prime range is 420–440.	• The extraction header and data extractions are two aspects of the algorithm.

S.No	Reference	Strengths	Weaknesses
		• Increased payload capability thus reducing image distortion. • Until being hidden, sensitive data is encrypted using RC4 encryption. • Provide two layers of concealment. • Develop DNA from the cover image.	
24	Ref [47]	• A secret key is employed. • Ensure that there are two levels of protection. • AES-128 is used to encrypt secret files. • AES has provided a strong degree of protection. • Before and after encryption, separate operations such as XOR and HASH-512 were performed on sensitive data. • Microdot has DNA embedded it to improve security.	• Several types of data are needed during the extraction process. • The DNA functionality is not maintained.
25	Ref [48]	• Modification rate is low. • After embedding confidential details, the DNA reference does not extend. • It makes use of two DNA references. • The initial DNA reference's usefulness was preserved. • Algorithm for blind people. • The non-labeled nucleotides do not shift. • High ability. • Until embedding plain text, encrypt everything. • Cracking with a high probability.	• Steganography method for purely obscuring results. • The receiver should be sent substitution rules. • Only uppercase letters, lowercase letters, 0, …., 9, period, and dots) are allowed in plain text. • It cannot have any other punctuation marks in it.
26	Ref [49]	• In the suggested algorithm, three DNA references are used. • Before hiding the plain text, encrypt it. • A secret key is employed. • Cracking with a high probability. • The algorithm is blind.	• Modification rate is high. • The biological DNA's versatility is not preserved.
27	Ref [50]	• Any programming language can be used to execute it. • To translate a hidden message to DNA format, build a random codon table. • Because of the insertion technique, there is a lot of duplication.	• There is no encryption. • May not keep records of an organism's life knowledge. • After embedding, lengthen the DNA reference. • The algorithm is not blind. • Algorithm for purely hiding records.
28	Ref [51]	• The algorithm is blind. • A secret key is employed. • Encrypt the hidden message using Playfair's algorithm. • After hiding the hidden data, there was no extension to the DNA reference. • In concealment, the replacement form is used. • Modification rate is poor. • The initial DNA reference's usefulness was preserved.	• Cracking with a low probability • The alteration rate would be high if the DNA comparison has several repetitive bases.
29	Ref [52]	• A secret key is employed. • High embedding capacity.	• The biological DNA's versatility is not preserved.

S.No	Reference	Strengths	Weaknesses
		• Using a modified Playfair algorithm, encrypt a secret letter. • After the hiding process, the length of DNA does not extend. • Easy, fast to implement, and performs better than Ref [32]. • Ref [32] hiding mechanism has been improved. • In hiding, the substitution form is used.	• The algorithm is not blind. • Cracking with a low probability
30	Ref [53]	• Technique that is almost imperceptible. • Before hiding a hidden message, encrypt it. • Provide two layers of concealment.	• The algorithm is not blind. • Algorithm for purely hiding records. • Only one part of the cover image is used to hide the DNA message.
31	Ref [54]	• A secret key is employed. • Without distorting the picture, two secret images may be hidden within it. • Provide two layers of concealment.	• The algorithm is not blind. • On secret records, no encryption technique was used.
32	Ref [55]	• Protection has been improved. • By reducing picture noise bits, the double carrier has been improved. • Enable for a fair amount of space. • Using a two-dimensional 2D logistic map with many parameters. • RC4 is a cryptographic algorithm that is used to encrypt sensitive information. • Provide two layers of concealment. • Image is used to create DNA. • A secret key is employed. • In hiding, the substitution form is used.	• Multiple data are required in the embedding and extraction processes.
33	Ref [56]	• Technique that is almost imperceptible. • This is an effective method. • By hiding in a random video frame, you can have protection. • Provide two layers of concealment.	• The algorithm is not blind. • Algorithm for purely hiding records. • The extraction method necessitates the use of numerous data sources.
34	Ref [57]	• The algorithm is blind. • Method that is both safe and efficient. • Until embedding, encrypt hidden data using the RSA algorithm. • Provide two layers of concealment. • A public key is employed.	• The biological DNA's versatility is not preserved.
35	Ref [58]	• Keeping track of an organism's life records. • The length of stego DNA is not increased. • The hidden data is encrypted using XOR and PRBG. • Reed-Solomon (RS) programming is used to measure and correct errors.	• It's not easy to put into practice. • Modification rate is high.
36	Ref [59]	• The hidden data and the key may be of any form and dimension.	• Algorithm is not blind.

S.No	Reference	Strengths	Weaknesses
		• Until hiding, using various encryption methods and analyzing them to choose the best one. • The normal key is used to select English characters to create more stable playfair cipher network. • There is no redundancy in the operation. • Strong results in a limited period of time. • In hiding, the substitution form is used.	• The amino acid functionality is not maintained. • A high degree of modification. • Cracking with a low probability
37	Ref [60]	• Using the vigenere or playfair cipher, encrypt hidden info. • The sum of data that is hidden is doubled. • High levels of security. • Until submitting to the recipient, the DNA connection will be hidden in a microdot on a piece of paper. • If the paper is unsafe, recreate a new key and sequence DNA, and the hiding process will start again. • Maintain the DNA sequence's functionality while avoiding mutations.	• Different data sets are sent to the receiver for retrieval. • Non-coding area has a high degree of alteration.
38	Ref [61]	• High-level security. • Random key generator for two levels of randomness. • It is necessary to use a hidden key. • The risk of cracking is incredibly high.	• Algorithm is not blind. • The functionality of DNA is not maintained. • The payload is not empty.

Table 2.
A comparison of the strengths and weaknesses of DNA steganography techniques.

This is accomplished by minimizing the requisite data that is transmitted to the recipient as much as possible. One of the strengths is to biologically preserve the DNA relationship's original features during the inclusion step while maintaining a fair data load. The reference DNA is used to mask hidden data while preserving protein processing functions. As shown in [11, 21, 25, 37, 48, 51, 52, 58, 60], some DNA characteristics such as silent mutation and codon repetition can mask details and alter the genetic sequence without changing the protein chain.

After most data-masking algorithms, the carrier can experience some distortion. Data masking techniques take care of embedding and embedded data; that is why it is communicated invisibly. As a result, it is important to minimize conveyor distortion. When data is entered into a string of stego DNA, the sequenced DNA's length and the degree of change are used to determine stego DNA precision. The low rate of change and lack of expansion rate results in high-quality DNA, which attracts less interest from potential attackers. [11, 33, 35, 36, 48, 51] reaches a low modulation frequency. Moreover, the expansion rate characteristic of DNA stego is not achieved at [11, 13, 21, 33–36, 48, 51, 58], which means that the payload is equal to zero.

It is recommended to use a two-stage steganography technique to hide sensitive data with more detail than previous data masking methods. Using two separate vectors in the same manner, increases confidentiality and makes it difficult for criminals to ingest or recover hidden data. Several methods [38, 44, 46, 54–57, 62] used the ref. DNA with another multimedia player to cover the hidden data. Some built DNA from cover images or confidential information, as shown in [44–46, 55, 62], while others used a random sample or selected from an online database, as shown in [38, 54–57].

The main factor is one of the most important aspects of data masking strategies. Data masking schemas are centered on the key used and can be classified into three categories. As shown in [21, 24, 32, 33, 40–42, 48, 50, 53, 56], pure data masking is less reliable because it does not use any key. As a result, using a key increases the device's usability by complicating the data-masking mechanism attack. Even if the perpetrators figure out what data-masking scheme is being used, they are unable to retrieve it. The carrier's sensitive information is not protected by the key. The secret is only in the hands of the sender and receiver. As a result, it is advisable to use a strong key when encrypting files, which ensures a more stable method. The second form is the hidden key [11, 13, 18, 25, 34, 35, 37–39, 43–47, 49, 51, 52, 54, 55, 58–61], which was accomplished in [11, 13, 18, 34, 35, 37–39, 43–47, 49, 51, 52, 54, 55, 58, 59]. The third form is classified as a public key, as shown by [23, 36, 57]. The public key is more secure than the private key in general, but it is still slower.

The probability of splitting the code and accessing confidential, sensitive data is known as the algorithm-cracking potential. Studying the probability of a striatum fracture aims to identify the variables that ensure that the risk of rupture is reduced. The likelihood of a leak is determined by the inclusion of certain unknown variables in the algorithm used to mask sensitive data, not by the amount of attempts made before the attacker gained access to the secret data. High probability penetration leads to high protection of the data-masking strategy described in [18, 23, 43, 48, 49, 58, 61]. The replacement strategy is believed to be a more powerful means of concealing data in DNA. The DNA sequence length can be preserved using this process as long as the payload is kept at zero. It also has more power as seen in [32–34, 41, 51, 52, 55, 59, 62], because it substitutes certain DNA nucleotides with cached data blocks or other nucleotides based on confidential data.

Capacity is a vital aspect of any data masking strategy, and it is one of the main criteria for data masking techniques. A steganography strategy must have broad data anonymization potential. This capacity can be measured in absolute terms, such as the hidden message's volume (for example, the data embedding rate, the bit per pixel, the bit per non-zero discrete cosine, the conversion factor, or the ratio of the secret message to a medium). The strength of DNA is calculated in bits per nucleotide (bpn). Thus, one of the main concerns for researchers in this area is improving the potential of secret results, which has previously been accomplished in [13, 18, 21, 23, 32, 33, 40–42, 44, 46, 48, 52, 55, 58–60].

As a result, it can be inferred that the primary goal of DNA-based double-layer masking algorithms is to encode sensitive data before hiding it in a high-power, blind, bio-stored, low moderation rate, load-free algorithm, not a pure method, with a high probability crack. In [48, 51, 58] suggested a low moderation rate, preservation of stretch length DNA for contrast, blindness, preservation of DNA versatility, double layer of security, high strength, and not a pure algorithm.

4. Conclusions

An increase in storage demand has generated a massive demand for creating new and evolving strategies for storing large amounts of data. DNA has recently been recognized as an efficient data carrier with the additional benefit of dependable data storage. DNA's bio-molecular computing capabilities are being used in cryptography and steganography. This research compares some recent DNA-based steganography algorithms and points out their security flaws. Each algorithm's advantages and disadvantages are also listed. Some crucial issues are discussed in terms of chance breaking, double layer security, single and double hiding layers, blindness, biologically retained DNA, alteration rate, an extension of DNA comparison, not a pure algorithm, substituting operation, and capacity. This study's comparison aims to provide researchers with the information they need to perform future tasks on more effective and accurate stable DNA steganography techniques.

Conflict of interest

"The authors declare no conflict of interest."

Author details

Adithya B.[1]* and Santhi G.[2]

1 Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India

2 Department of Information Technology, Pondicherry Engineering College, Puducherry, India

*Address all correspondence to: adithya27.07@pec.edu

References

[1] Singh G. A study of encryption algorithms (RSA, DES, 3DES and AES) for information security. International Journal of Computer Applications, 2013; 67(19).

[2] Subhedar MS, Mankar VH. Current status and key issues in image steganography: A survey. Computer science review, 2014; 13:95–113.

[3] Hamed G, et al. Comparative study for various DNA based steganography techniques with the essential conclusions about the future research. 11th International Conference on Computer Engineering & Systems (ICCES); IEEE; 2016.

[4] Amin MM, et al. Information hiding using steganography. 4th National Conference on Telecommunication Technology; IEEE; 2003.

[5] Al-Mohammad A. Steganography-based secret and reliable communications: Improving steganographic capacity and imperceptibility [thesis]. Brunel University, School of Information Systems, Computing and Mathematics; 2010.

[6] Santoso KN, et al. Information Hiding in Noncoding DNA for DNA Steganography. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences; 2015; 98(7):1529–1536.

[7] Kumari P, Kapoor R. Image Steganography for Data Embedding & Extraction using LSB Technique. International Journal of Computer Applications & Information Technology; 2016; 9(2):192.

[8] Ashok J, et al. Steganography: an overview. International Journal of Engineering Science and Technology; 2010; 2(10):5985–5992.

[9] Nickfarjam AM, Azimifar Z. Image steganography based on pixel ranking and Particle Swarm Optimization. 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP); IEEE; 2012.

[10] Sheelu AB. An Overview of Steganography. IOSR Journal of Computer Engineering (IOSR-JCE); 2013; 11(1):15–19.

[11] Khalifa A. LSBase: A key encapsulation scheme to improve hybrid crypto-systems using DNA steganography. 8th International Conference on Computer Engineering & Systems (ICCES); IEEE; 2013.

[12] Jain S, Bhatnagar V. Analogy of various DNA based security algorithms using cryptography and steganography. International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT); IEEE; 2014.

[13] Torkaman MRN, Kazazi NS, Rouddini A. Innovative approach to improve hybrid cryptography by using DNA steganography. International Journal of New Computer Architectures and their Applications (IJNCAA); 2012; 2(1):224–235.

[14] Bansod S, Bhure G. Data encryption by image steganography. Int. J. Inform. Comput. Technol; Int. Res. Publ. House; 2014; 4:453–458.

[15] Singh KU. Video steganography: text hiding in video by LSB substitution. International Journal of Engineering Research and Applications; 2014;4(5): 105–108.

[16] Chandel B, Jain S. Video Steganography: A Survey. IOSR Journal of Computer Engineering (IOSR-JCE); 2016; 18(1):11–17.

[17] Yang Y. Information analysis for steganography and steganalysis in 3D polygonal meshes [thesis]. Durham University; 2013.

[18] Atito A, Khalifa A, Rida S. DNA-based data encryption and hiding using playfair and insertion techniques. Journal of Communications and Computer Engineering; 2012; 2(3):44.

[19] Al-Wattar AHS, Mahmod R, Zukarnain ZA, Udzir N. Review Of Dna And Pseudo Dna Cryptography. International Journal of Computer Science and Engineering (IJCSE); 2015; 4(4):65–76.

[20] Tornea O. Contributions to DNA cryptography: applications to text and image secure transmission [Thesis]. Université Nice Sophia Antipolis; 2013.

[21] Abbasy MR, et al. DNA base data hiding algorithm. International Journal of New Computer Architectures and their Applications (IJNCAA); 2012; 2(1): 183–192.

[22] Adithya B, Santhi G. Bio-inspired Deoxyribonucleic Acid based data obnubilating using Enhanced Computational Algorithms. In: Proceedings of the International Conference on Computer Networks, Big Data and IoT; Springer; 2020. p. 597–609

[23] Skariya M, Varghese M. Enhanced double layer security using RSA over DNA based data encryption system. International Journal of Computer Science& Engineering Technology (IJCSET); 2013; 4(6):746–750.

[24] Shiu H, et al. Data hiding methods based upon DNA sequences. Information Sciences; 2010; 180(11): 2196–2208.

[25] Mousa H, et al. Data hiding based on contrast mapping using DNA medium. Int. Arab J. Inf. Technol.; 2011; 8(2): 147–154.

[26] Adleman LM. Molecular computation of solutions to combinatorial problems. Nature; 1994; 369:40.

[27] Ogiwara M. Simulating Boolean Circuits on DNA Computers. In Proceedings of the 1st International Conference on Computational Molecular Biology; ACM Press; 1997.

[28] Clelland CT, Risca V, Bancroft C. Hiding messages in DNA microdots. Nature; 1999; 399(6736):533–534.

[29] Sureshraj D, Bhaskaran VM. Automatic DNA sequence generation for secured cost-effective multi-cloud storage. 2012.

[30] Singh A, Singh R. Information hiding techniques based on DNA inconsistency: An overview. 2nd International Conference on Computing for Sustainable Global Development (INDIACom); IEEE; 2015.

[31] Bhateja A, Mittal K. DNA Steganography: Literature Survey on its Viability as a Novel Cryptosystem. Journal of Computer Science and Engineering; 2015; 2(1):8–14.

[32] Taur J.-S, et al. Data hiding in DNA sequences based on table lookup substitution. International Journal of Innovative Computing, Information and Control; 2012; 8(10):6585–6598.

[33] Guo C, Chang C-C, Wang Z-H. A new data hiding scheme based on DNA sequence. Int. J. Innov. Comput. Inf. Control; 2012; 8(1):139–149.

[34] Khalifa A, Atito A. High-capacity DNA-based steganography. 8th International Conference on Informatics and Systems (INFOS); IEEE; 2012.

[35] Huang YH, Chang CC, Wu CY. A DNA-based data hiding technique with

low modification rates. Multimedia tools and applications; 2014; 70(3):1439–1451.

[36] Mitras BA, Abo A. Proposed steganography approach using DNA properties. International Journal of Information Technology and Business Management; 2013; 14(1):96–102.

[37] Haughton D, Balado F. Biocode: Two biologically compatible algorithms for embedding data in non-coding and coding regions of dna. BMC bioinformatics; 2013; 14(1):121.

[38] Shyamasree C, Anees S. Highly secure DNA-based audio steganography. International Conference on Recent Trends in Information Technology (ICRTIT); IEEE; 2013.

[39] Haughton D, Balado F. Security study of keyed DNA data embedding. Global Conference on Signal and Information Processing (GlobalSIP); IEEE; 2013.

[40] Bhattacharyya D, Bandyopadhyay SK. Hiding secret data in dna sequence. International Journal of Scientific &Engineering Research; 2013; 4(2).

[41] Agrawal R, Srivastava M, Sharma A. Data hiding using dictionary based substitution method in DNA sequences. 9th International Conference on Industrial and Information Systems (ICIIS); IEEE; 2014.

[42] Menaka K. Message encryption using DNA sequences. World Congress on Computing and Communication Technologies (WCCCT); IEEE; 2014.

[43] Manna S, et al. Modified technique of insertion methods for data hiding using DNA sequences. International Conference on Automation, Control, Energy and Systems (ACES); IEEE; 2014.

[44] Das P, Kar N. A DNA based image steganography using 2d chaotic map. International Conference on Electronics and Communication Systems (ICECS); IEEE; 2014.

[45] Majumdar A, Sharma M, Kar N. An Improved Approach to Steganography using DNA Characteristics. IEEE; 2014. p. 138–145

[46] Das P, Kar N. A highly secure DNA based image steganography. International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE); IEEE; 2014.

[47] Chaudhary H, Bhatnagar V. Hybrid approach for secure communication of data using chemical DNA. 5th International Conference on Confluence the Next Generation Information Technology Summit (Confluence); IEEE; 2014.

[48] Ibrahim FE, Abdalkader H, Moussa M. Enhancing the Security of Data Hiding Using Double DNA Sequences. Industry Academia Collaboration Conference (IAC).

[49] El-Latif EIA, Moussa MI. Chaotic Information-hiding Algorithm based on DNA. International Journal of Computer Applications (0975–8887); 2015; 122 (10).

[50] Yamuna M, Elakkiya A. Codons in Data Safe Transfer. International Journal of Engineering Issues; 2015; (2): 85–90.

[51] Hamed G, et al. Hybrid technique for steganography-based on DNA with n-bits binary coding rule. 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR); IEEE; 2015.

[52] Marwan S, Shawish A, Nagaty K. An Enhanced DNA-based Steganography

Technique with a Higher Hiding Capacity. Bioinformatics; 2015.

[53] Manisha B, Mohit P. Double Layered Dna Based Cryptography. IJRET: International Journal of Research in Engineering and Technology; 2015; 4 (4):2321–7308.

[54] Chakraborty S, Bandyopadhyay KS. Data Hiding by Image Steganography Appling DNA Sequence Arithmetic. International Journal of Advanced Information Science and Technology (IJAIST); 2015; 44(44).

[55] Das P, et al. An Improved DNA based dual cover steganography. Procedia Computer Science; 2015; 46: p. 604–611

[56] Indora S. Cascaded DNA cryptography and steganography. International Conference on Green Computing and Internet of Things (ICGCIoT); IEEE; 2015.

[57] Tank RM, Vasava HD, Agrawal V. DNA-Based Audio Steganography. Oriental journal of Computer Science and Technology; 2015; 8:43–48.

[58] Santoso K, et al. Sector-based DNA information hiding method. Security and Communication Networks; 2016; 9 (17):4210–4226.

[59] Marwan S, Shawish A, Nagaty K. DNA-based cryptographic methods for data hiding in DNA media. Biosystems; 2016; 150:110–118.

[60] Marwan S, Shawish A, Nagaty K. Utilizing DNA Strands for Secured Data-Hiding with High Capacity. International Journal of Interactive Mobile Technologies; 2017; 11(2).

[61] Malathi P, et al. Highly Improved DNA Based Steganography. Procedia Computer Science; 2017; 115: p. 651–659

[62] Liu H, Lin D, Kadir A. A novel data hiding method based on deoxyribonucleic acid coding. Computers & Electrical Engineering; 2013; 39(4):1164–1173.