DNA Cryptography

Salapu Venkata Lakshmi, P.Mounika K.Divya Kalyani
Assistant Professor, Dadi Institute of Engineering and Technology
venkatalakshmisalapu@gmail.com,reddymounika010593@gmail.com
divyakalyanik@diet.edu.in

Abstract
Cryptography plays a key role in information security. Many new algorithms and techniques have been used for the same purpose. Cryptography using DNA computing is the current state of the art. DNA cryptography comes with the next level of data integrity and confidentiality to protect information from intrusions. In this project, a cipher solution is proposed with a new symmetric key generation model based upon DNA strands, nucleotides, codons base pair rules, mutation and DNA to mRNA conversion. This solution emphasizes on usage of biological processes & the random changes found in DNA and simulate those processes in the key generation model.

Introduction
Internet utilization has become a ubiquitous part of everyday life for billions of people around the world, with many relying on the internet for work, entertainment, communication, and information. The users of the internet will increases rapidly along with the increases of network technology. Securing information is an essentially difficulty in present day, and there are several instances that organizations and individuals face when it comes to protecting their data. Securing data is a complex and challenging task that requires ongoing effort and investment.

To provide secure communication and protect sensitive information we use a technique called cryptography. Cryptography plays a critical role in ensuring the security and privacy of sensitive information in many areas of real life with Encryption and Decryption techniques. As traditional cryptographic systems are now vulnerable to attacks, the concept of using DNA Cryptography has been identified as a possible technology that brings forward to provide a highly secure and efficient method of transmitting, storing, and authenticating sensitive information.

DNAcomputing:
DNA computing is a new field which is growing in the modern days. The journey of DNA cryptography started with the development of DNA

computing. DNA computing was introduced by L. Adleman [2] in the year of 1994 to solve the complex computational problem. In his study he found that DNA has high storage and computational capability. Using DNA computation, he solved a searching problem named directed Hamiltonian path problem with seven vertices where he assumed molecules as vertices and encoded them in a molecule sequence and performed computations by chemical operations in lab.

This is similar to traveling salesman path problem where large possible solutions generated to find better paths to reach from source to destination. An image encryption algorithm based on DNA sequence addition operation is presented by Wang ET. Al. A DNA sequence matrix is obtained by encoding the original image and it is divided into some equal blocks and two logistic maps, DNA complementarity and DNA sequence addition operation are utilized to add these blocks. A DNA sequence matrix is decoded to get the encrypted image DNA computing also called as Biomolecular computing.

The reason behind taking DNA in to account as a computation medium is its high storage capacity, vast parallelism and high energy efficiency. Due to these capabilities DNA can act as processor which can process large amount of data and can perform computations to get several possible solutions.

About DNA:

DNA stands for Deoxyribo Nucleic Acid. DNA represents the genetic blueprint of living creatures. DNA contains "instructions" for assembling cells. Every cell in human body has a complete set of DNA. DNA is unique for each individual. DNA) is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses.

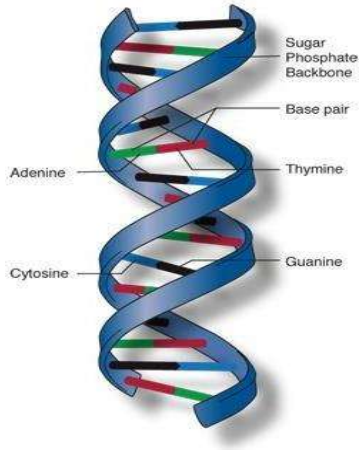


Fig.1.1. DNA Structure

DNA is a nucleic acid; alongside proteins and carbohydrates, nucleic acids compose the three major macromolecules essential for all known forms of life. [8]. in 1953, James Watson discovered the structure of DNA. Most DNA molecules consist of two biopolymer strands coiled around each other to form a double Helix.

The two DNA strands are known as polynucleotides since they are composed of simpler units called nucleotides. Each nucleotide is composed of a nitrogen-containing nucleobase guanine (G), adenine (A), thymine (T), or cytosine (C) as well as a monosaccharide sugar called deoxyribose and a phosphate group. According to base pairing rules (A with T and C with G), hydrogen bonds bind the nitrogenous bases of the two separate polynucleotide strands to make double -stranded DNA. [11]. The sequence of these bases determines the information available for building or forming an organism, similar to the way in which letters of the alphabet appear in a certain order to for 'm' words and sentences [8] as shown below DNA strand made of letters (DNA bases):
ATACTTGAATATATGTCAATTAGT

Letters make words (codons):
ATA CTT GAA TAT ATG TCA ATT AGT

Words make sentences (Genes):
ATA -CTT -GAA -TAT ATG -TCA -ATT –AGT

Literature Survey
Introduction DNA Cryptography:
Cryptography came into existence when human beings started to realize the importance of information, and started get worried about its privacy Information security is based upon 3 major elements (confidentiality, integrity, authenticity). By use of a Cryptography technique one can hide some information in a way which is not publicly readable (confidentiality).

- **Biological Framework:** DNA is a molecule, inside every organism. They found out that DNA structure is double helix/ stranded like a spiral ladder. Each helix consists of other monomers (a molecule, that can be bonded to form a polymer) called nucleotides. Each nucleotide has sugar and phosphate groups and nitrogen base. Adenine always makes a bond with Thymine (T), and Guanine makes bond with Cytosine(C). These nitrogen bases are Adenine (A), Thymine (T), Guanine (G) and Cytosine (C).

- **Methodology:**DNA cryptography is a theoretical computer science field where DNA is used for information hiding. The smallest DNA is of 30 nucleotides.DNA is an information storage, few grams of DNA can consist all the data available over the internet. DNA uses nucleotides to store the information.

The methodology is based upon method discussed in biological framework section. Number of rounds and number of keys are random, depends on the user given key.

Here's one possible method for encrypting text using DNA:
Convert the text to be encrypted into binary form. For example, you could use ASCII encoding to convert each character to its corresponding binary value.

Split the binary data into small segments, such as 4 bits each.

Map each 4-bit segment to a unique DNA nucleotide sequence. For example, you could use the following mapping:
0000 = Adenine (A)
0001 = Thymine (T)
0010 = Guanine (G)
0011 = Cytosine (C)
0100 = Adenine-Thymine (AT)
0101 = Thymine-Adenine (TA)
0110 = Guanine-Cytosine (GC)
0111 = Cytosine-Guanine (CG)

To decrypt the text, the reverse process can be used. The DNA sequence is sequenced and then converted back to binary, and the binary data is then converted back to the original text.

Bits	Nucleotides
00	A
11	C
01	G
10	T

CurrentSystem:
The Block ciphers such as DES and IDEA has fixed number of rounds and fixed length of key sizes, because of that these algorithms are subjected to various attacks such as brute-forcing using some mathematical calculations performed by the cryptanalysts or attackers.

Disadvantages of Current System:-
- DES is vulnerable to a type of attack called a "meet-in-the-middle" attack, where an attacker encrypts the plaintext with all possible keys and then decrypts the resulting cipher text with all possible keys, and then looks for matching pairs of cipher text and plaintext.
- The main disadvantage of fixed-round block ciphers with fixed key sizes is their vulnerability to attacks, which can compromise the confidentiality and integrity of the encrypted data.

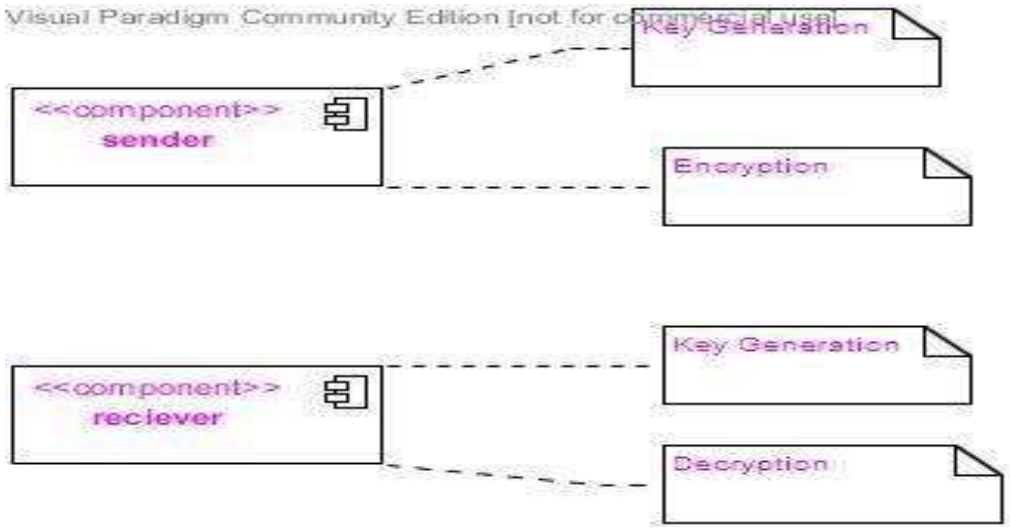
Proposed System:
The proposed system used variable size of key and variable no of rounds so security is improved. The key size and rounds depends upon user password which brings a great random behaviour in a key model, which makes cryptanalysis even harder.

Advantage of Proposed System:-
The proposed system offers significant advantages over traditional block ciphers with fixed key sizes and rounds.

Component Diagram:
Component diagrams represent a set of parts and their relationships. These parts made up of classes, interfaces or collaborations. So Component diagrams represent the implementation view of a system. During design phase software artifacts (classes, interfaces etc) of a scheme are arranged in distinct groups depending upon their relationship. Now these groups are known as components. Finally, component diagrams are used to visualize the implementation

Algorithm

Fig 3.7 Component diagram



KEY GENERATION:

STEP 1: Convert the Text message and Password into binary format.

STEP 2: Convert the password into binary format and then convert it into the sequence of Nucleotides using the encoding table.

STEP 3: The process of anneal is started and the single helix is now bonded with its pair according to complementary rule.

T <-> A, G <-> C

STEP 4: Start the transcription process and replace T with U.

STEP 5: We get an mRNA strand and apply the point mutation on mRNA strand till a stop codons occurs.

UAG (Amber), UAA (ochre), UGA (opal)

STEP 6: After the mutation is applied, the separated proteins are called DNA keys. Number of DNA keys depends upon how many break codons are found.

STEP7: Then encode the above mRNA keys to binary it separate every block as an each 8 bits of a key.

ENCRYPTION:

STEP 1:Each DNA key creates a binary block

B= {b1, b2 …Bn}

STEP 2: In every block, each 8 bits block denoted as

Bi= {k1, K2 …KN}

STEP 3: For encryption first 8 bits of text message is picked up, and left-shifted to 1 bit and XOR function with each key, block 1, this continues for all the {k1,k2.....kn} in b1.

CM= (M << 1) b1ki

STEP 4: The second binary block shifts message to 2 bits and apply XOR

CM= (CM << 2) b2kj

That can be described as:

CM= (CM << I) biki

Decryption:

STEP 1: Each DNA key creates a binary block

B= {Bn, bn-1 ... b1}

STEP 2: In every block, each 8 bits block denoted as

Bi= {KN, kn-1 ... k1}

STEP 3: For decryption first 8 bits of encrypted message is picked - up, and XOR operation

{KN, kn-1.....k1} in b2.

CM= (M >> I) biki

STEP 4: The second binary block shifts message to 2 bits and apply XOR function.

CM= (CM >> i-1) b2ki

That can be described as:

CM= (CM >> 1) b1ki

EXAMPLE

STEP 1: Text message: AB

ASCII values: A=65, B=66
Binary format: A=01000001 B=01000010
User Password: GVp
ASCII values: G=71 V=86 p=112
Binary format: G=01000111 V=01010110
p=01100000

STEP 2**:** GAGC GGGT GTAA

Bits	Nucleotides
00	A
11	C
01	G
10	T

STEP 3: G <-> C T <-> A

CTCG CCCA CATT

=> Append Password and converted nucleotides:

GAGC GGGT GTAA CTCG CCCA CATT

STEP 4: T->U

GAGC GGGU GUAA CUCG CCCA CAUU

STEP 5: UAG (Amber), UAA (ochre), UGA (opal)

b1=GAG CGG GUG UAA CUC GCC CAC AUU

b2=CUC GCC CAC AUU

STEP 6: b1= GAG CGG GUG UAA CUC GCC CAC AUU

b2=CUC GCC CAC AUU

Drop the second variable of b1, b2:

b1=GGGG GACC CCCU

b2=CCCC CU_ _

STEP 7: Encoded the above mRNA keys to binary.

b1=01010101 01001111 11111110 => {K1, K2, K3}

b2=11111111 11100000 => {K1, K2}

ECRYPTION:

Text Message: AB

STEP 1: Binary format: A=01000001 B=01000010

Left shift to 1bit of A= 10000010 first 8bits of b1=01010101 (K1)

STEP 2: CM= (M << 1) b1ki

10000010 01010101(b1K1) = 11010111

CM= 11010111 01001111(b1K2) = 10011000

10011000 11111110 (b1K3) = 01100110

There for CM= 01100110

STEP 3: CM=(CM << 2) b2kj

CM= 01100110 shifts 2 bits left CM= 00000001 10011000

CM 1 = 00000001 11111111(b2k1) 10011000 11111111(b2k1)

= 11111110

=01100111

11111110 11100000 (b2k2) 01100111 11100000 (b2k2)

a = 00011110 b= 10000111

CM1= append a b = > 00011110 10000111

CM1 converted into decimal number = 4096+2048+1024+512+128+4+2+1

= 7815 (encrypted message)

Decryption:
STEP 1: Encrypted message=7815
Converted into binary number
00011110 10000111
STEP 2: CM= 00011110 11100000(b2k2) 10000111 11100000 (b2k2)
= 11111110 = 01100111
11111110 11111111(b2k1) 01100111 11111111(b2k1)
a1 = 00000001 b1= 10011000
STEP 3: CM= append a1b1 => 00000001 10011000
2 bits right shifted =>00000000 01100110
CM1= 01100110 11111110 (b1K3) = 10011000
= 10011000 01001111(b1K2) = 11010111
= 11010111 01010101(b1K1) = 10000010
STEP 4: CM1 = 10000010: Right shifted 1 bit
CM2= 01000001
CM = decimal number of CM2 (01000001)
=> 65
CM = A (Decrypted message)

Results

Encryption and Key Generation

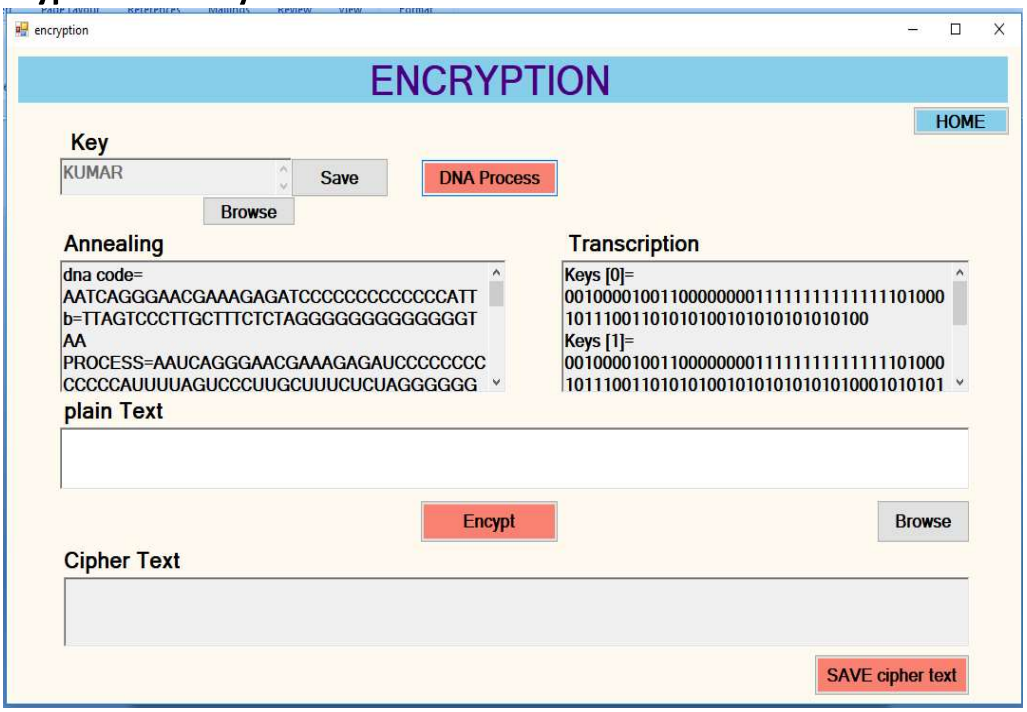


Fig 7.4 Encryption

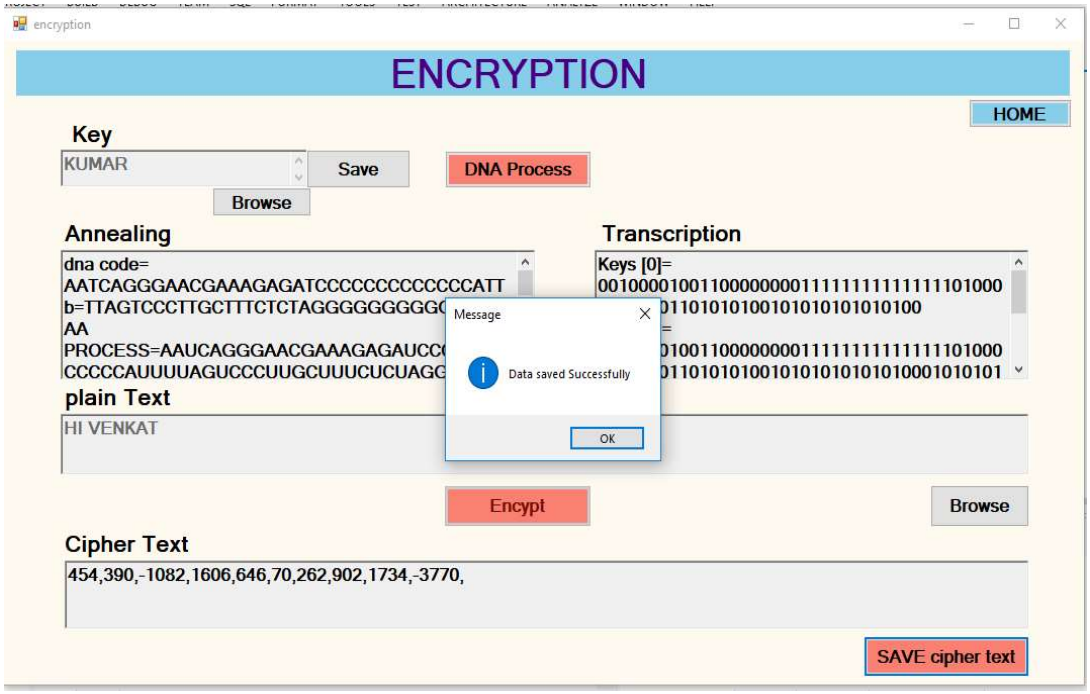


Fig 7.5 Save Cipher text

Decryption and Key Generation:

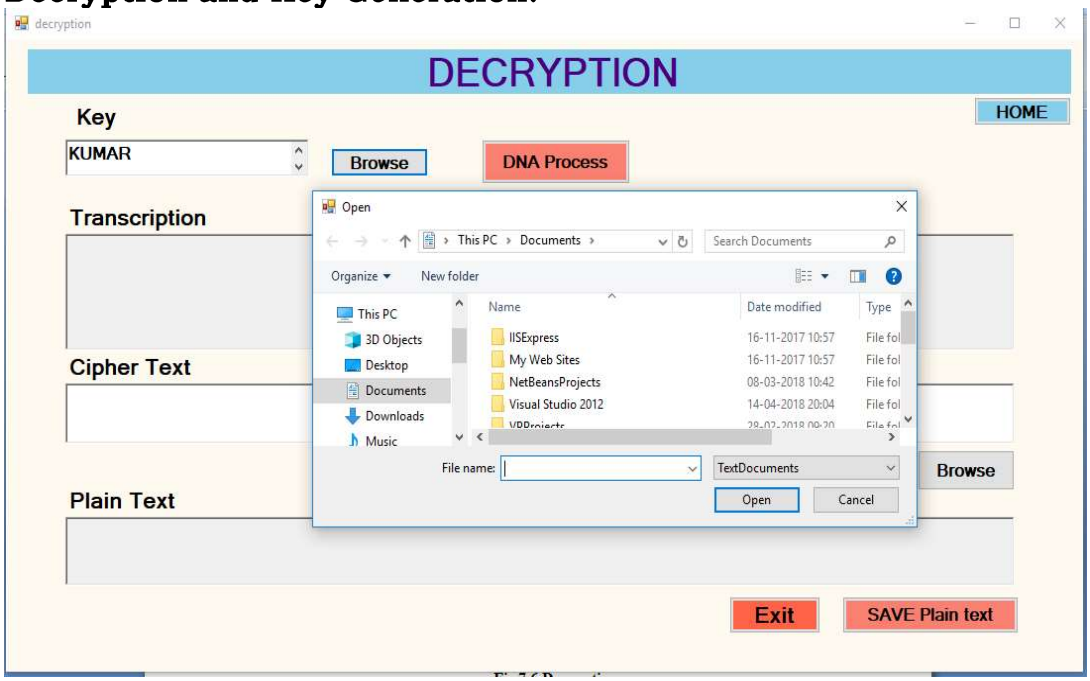


Fig 7.6 Browse key

Fig 7.7 Browse Cipher text

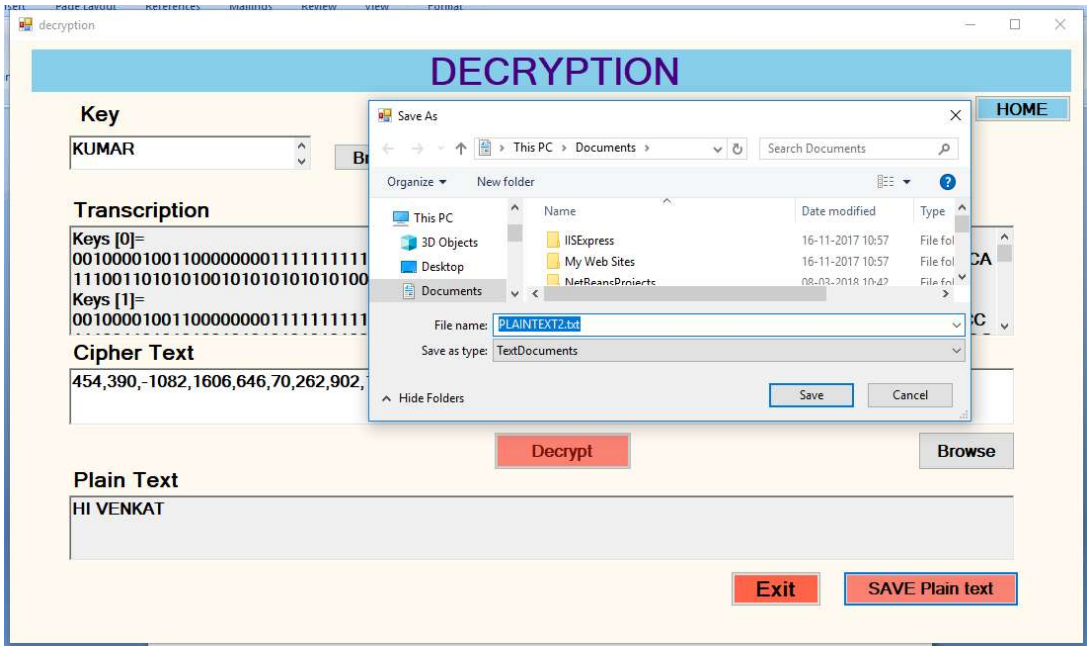


Fig 7.8 save Plaintext

Conclusion

In this approach the key size is not fixed unlike other block ciphers where key size is fixed like AES. The focus of the methodology was that the number of rounds should be random. The key size and rounds depends upon user password, which brings a great random behavior in key model, which makes cryptanalysis even harder. This

can be used for those networks where processing power is not a question because user password may produce a great variety of keys.

References

[1] Adleman, L. M. 1994. Molecular computation of solutions to combinatorial problems.

[2] Chen, J.2003. A DNA-based, bio molecular cryptography design. Circuits and Systems.

[3] Church house, R. F. 2002. Codes and ciphers: Julius Caesar, the Enigma, and the Internet

[4] Husain, E. M. 2016. A DNA cryptographic technique based on dynamic DNA sequence table.

[5] Leuenberger, M. N. 2001. Quantum computing in molecular magnets.

[6] Li, X.-s. L.-p. 2008. A novel generation key scheme based on DNA.

[7] Needleman, S. B. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins.

[8] Ning, K. A. A pseudo DNA cryptography method.

[9] Ochani, A. D. 2016. DNA image encryption using Modified Symmetric Key (MSK).

[10] Varma, P. S. 2014. Cryptography based on DNA using random key generation scheme.

[11] Watson, J. D. 1953. The structure of DNA.