DNA-Based Cryptography Approaches using Central Dogma of Molecular Biology

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Software Engineering

Submitted By
Ravi Gupta
(801231022)

Under the supervision of:
Dr. Ajay Kumar
Assistant Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

July 2014

Certificate

I hereby certify that the work which is being presented in the thesis entitled, "*DNA Based cryptography approaches using Central Dogma of Molecular Biology*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Software Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Dr. Ajay Kumar and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Signature:

(Ravi Gupta)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Ajay Kumar)
Asst. Professor
Computer Science & Engineering Department

Countersigned by

(Dr. Deepak Garg)
Head
Computer Science and Engineering Department
Thapar University
Patiala

(Dr. S. K. Mohapatra)
Dean (Academic Affairs)
Thapar University
Patiala

Acknowledgement

No volume of words is enough to express my gratitude towards my guide, **Dr. Ajay Kumar**, Assistant Professor, Computer Science and Engineering Department, Thapar University, who have been very concerned and have aided for all the material essential for the preparation of this thesis report. They have helped me to explore this vast topic in an organized manner and provided me with all the ideas on how to work towards a research oriented venture.

I am also thankful to **Dr. Deepak Garg**, Head of Department, CSED, **Dr. Maninder Singh,** Associate Professor, for the motivation and inspiration that triggered me for the thesis work.

I would also like to thank the staff members and my colleagues specifically **Rahul Kumar Singh** who were always there in the need of the hour and provided with all the help and facilities, which I required, for the completion of my thesis.

Most importantly, I would like to thank my **Parents** and the **Almighty** for showing me the right direction out of the blue, to help me stay calm in the oddest of the times and keep moving even at times when there was no hope.

Ravi Gupta

(801231022)

Abstract

Information security is one of the most significant concerned areas of communications and information transmission. The concept of making secure information is to transform a plain message that is understandable by everyone into a human unreadable format or difficult to make out by encoding message using some cryptographic algorithms. Nowadays, information security is significant and fundamental issues of information transmission, where researchers are working on the evolution of new cryptographic algorithms. Cryptography is the process of furnishing security in information while transmitting over public networks by encrypting the original information or message. Cryptography technique is used in various fields such as banking services, digital certificate, digital signature, and message & image encryption. An efficient direction of achieving information security can be termed as DNA-BASED Cryptography.

In this thesis, we propose two improved DNA-BASED symmetric key cryptographic algorithms, first extends the Triple Data Encryption Standard algorithm using the concept of DNA computing and second improved data encryption approach based on the Deoxyribonucleic acid that exploits two different techniques: Substitution technique and Central Dogma of Molecular Biology. Both algorithms convert the message into a protein (Cipher text) using the Deoxyribonucleic acid strands and Central Dogma of Molecular Biology. Deoxyribonucleic acid strands provide the robust encryption keys for first algorithm and used as information carrier for second because 163 million DNA sequences available worldwide. Central Dogma of Molecular Biology extends the complexness of the algorithms using the transcription and translation techniques. Encrypted output of the proposed algorithms is highly secure and compress because of the inclusion of the artificial DNA sequence.

Table of contents

List of figures

List of tables

Chapter 1
Introduction

Information security is the process of making information secure and transfer via networks without any attacks. Information security can be termed as cryptography. Cryptography [37] is a perspective that converts plaintext into a non-understandable format (ciphertext) by using cryptographic algorithms. Cryptography is used for the security purposes in the various areas such as banking services, digital certificate, digital signature, message and image encryption, etc. Multifarious symmetric and asymmetric key cryptography algorithms such as DES, TDES, AES, and RSA [28] exist for the metamorphosis of plaintext into the ciphertext.

Conventional cryptographic algorithms are incompetent to endow security for the huge amount of data on the growing internet and can be easily breakable by intruders for malicious purposes [2-8]. There is an exigency of new cryptographic algorithm emerged for the information security that can bestow strong ciphertext and protect unauthorized access to data. Therefore, the field of DNA computing emerges.

DNA Computing is one of the fields proposed by Adleman [1] for figuring out the directed Hamiltonian Path Problem (or Shortest Path Problem). Hamiltonian Path Problem (HPP) is to determine the shortest path by traversing each and every city (node) exactly one and return to last node in the directed graph. The solution for the directed HPP contains several steps:

i. Get random routes from the graph.
ii. Hold only those routes which start from starting city and finish with ending city.
iii. If the graph contains ‘m’ cities then hold only those routes which contain exactly ‘m’ cities.
iv. Hold only those routes which contain all the cities of the directed graph exactly one.
v. If any route persists, say, “yes," otherwise say “no” [24].

Adleman took seven cities to perform HPP using DNA computing. Adleman assigned different DNA sequence to each city in the graph. An edge between two cities was also

a DNA sequence which formed by merging of second half part DNA sequence of source city with first half part DNA sequence of destination city. Using DNA computing Adleman solved HPP problem and gave a new direction in the field of DNA computing.

Gehani *et al.* [2] worked on the DNA computing and introduced a new image encryption algorithm using one-time-pads. They proposed the substitution method where pairwise mapping executed between plaintext message and ciphertext in the algorithm. They illustrated the novel idea of DNA-chip based approach of encryption and decryption for two dimensional image using one-time-pads and also introduced an improved steganography system by decreasing the difference between plain message and disorder strands.

DNA computing is selected by various researchers as a remedy for security on a large amount of data in the recent years and has been proposed various DNA-BASED cryptographic approaches [1-3]. DNA contains some fascinating properties such as high storage capacity, vast parallelism, and high energy efficiency that are a boon for the data hiding and the solution of applications such as an optimization problem, image and signal processing, clustering and forecasting [5-6]. A gram DNA can contain 10^8 TB of data that is equal to 10^{21} DNA-bases.

1.1 DNA

DNA (Deoxyribonucleic Acid) is a source plasm in all living life, and it is a form of biological super molecule formed by nucleotides. Monomer unit of DNA is called deoxyribonucleotides. There are four types of bases or nucleotides found in DNA or DNA consists of four bases [31]. These bases or nucleotides are given below:

i. Adenine (A),
ii. Cytosine (C),
iii. Thymine (T), and
iv. Guanine (G)

Where A, C, T, and G are abbreviated form of Adenine, Cytosine, Thymine and Guanine respectively. DNA contains two nitrogen bases: Purines and Pyrimidines, where A and G bases are double-ring molecules called purines and C and T bases are single-ring molecules called pyrimidines [34].

1.2 DNA structure

DNA contains a double helical structure, which is made by two single-strands of DNA that runs in the opposite direction each other in fig. 1.1 [39]. In helical structure, A joins with T by double hydrogen bonds and C joins with G by triple hydrogen bonds, where A-T pair weaker around 30% than C-G pair. Hydrogen pairing of DNA nucleotides is known as complementary pairing discovered by Watson and Crick [9]. Pairing of bases forms base units of DNA. Structure of DNA's strand consists of a backbone that is built from sugar molecules. These sugar molecules are attached together by phosphate group. Sugar and phosphate group has ratio of 3 carbons of sugar attached with 5 carbons of the phosphate group connected with next sugar molecule. A DNA structure reads in terms of carbon ratio of sugar molecules and the phosphate group. For example, a DNA is read as 5 carbons to 3 carbons, where 5 carbons terminations represent phosphate group, and 3 carbons terminations represent sugar molecules. Sugar molecules are attached with one of the four bases i.e. (T, A, C, G) [25].

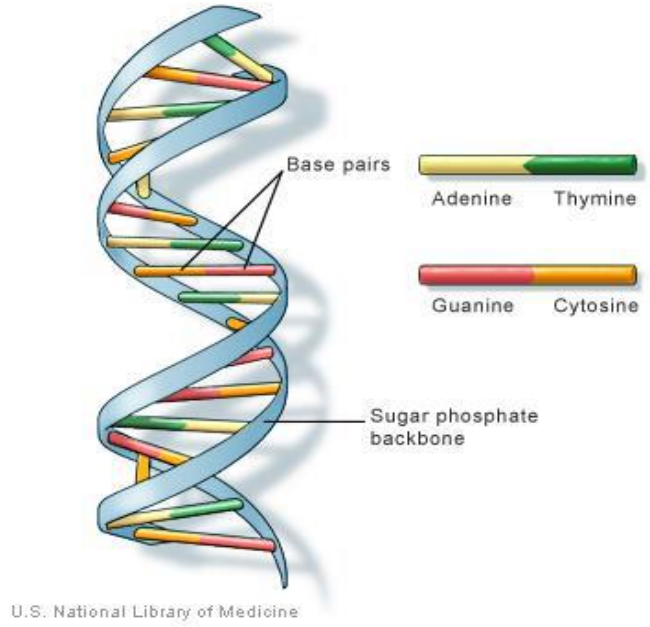

Fig. 1.1: DNA structure [14]

The backbone of nucleotide is two units negative charged for per base-pair and one unit negative charge for per phosphate group. There will be a strong repulsion occur between two strands and split up, if there is no salt present in the medium.

1.3 DNA operations

DNA operations contain the capability for the solution of the application in the fields as data hiding, image encryption, and steganography, etc. DNA operations are execute in the laboratory using identical conditions with care of various parameters like Pressure, Temperature, Oxygen ratio, etc. Result of operations totally dependent on the environment. DNA operations can give different outputs in different environments for the same input because DNA is a sensitive molecule. Various DNA operations can be done on DNA sequences in the test tube are as follow:

i. Synthesis

A sequence of DNA of particular length and strand can be synthesised in the laboratory. It is possible sequences up to a fixed length. Longer 'random' strands are available. A solution with four different DNA bases in it, is applied to the synthesizer. These nucleotides are formed according to a given sequence provide by the user. Instrument creates millions of copies of desired oligonucleotides and inserts them in test tube [22].

ii. Separation

Partition of DNA sequence by length.

iii. Merging

Spill out two test tubes of DNA into a single tube to perform merging.

iv. Extraction

Extract a desired sequence from a long DNA sequence.

v. Annealing/ Melting

Bond/ Break two single DNA sequence particles with complementary strands.

vi. Amplification

Use Polymerase Chain Reaction to double or replicate the DNA sequence.

vii. Cutting

Cut DNA sequence with restrict enzymes.

viii. Ligation

Join DNA sequences with complementary sticky ends using ligase.

ix. Detection

Make sure presence/ absence of DNA sequence in a given test tube [26].

x. Denature

Denature is a process used for pull two sequences of DNA apart from each other using heated up solution of DNA in dissolved water. Denature is mainly used for pull strands by heated solution of DNA [22].

xi. PCR

PCR is known as Polymerase Chain Reaction. It is used almost in every biological DNA computing. Its main works is to detect and measure vanishingly minor quantities of DNA and produce customized pieces of DNA. PCR uses polymerase enzyme as a specialized element for the process. Fig. 1.2 shows DNA amplification using PCR.

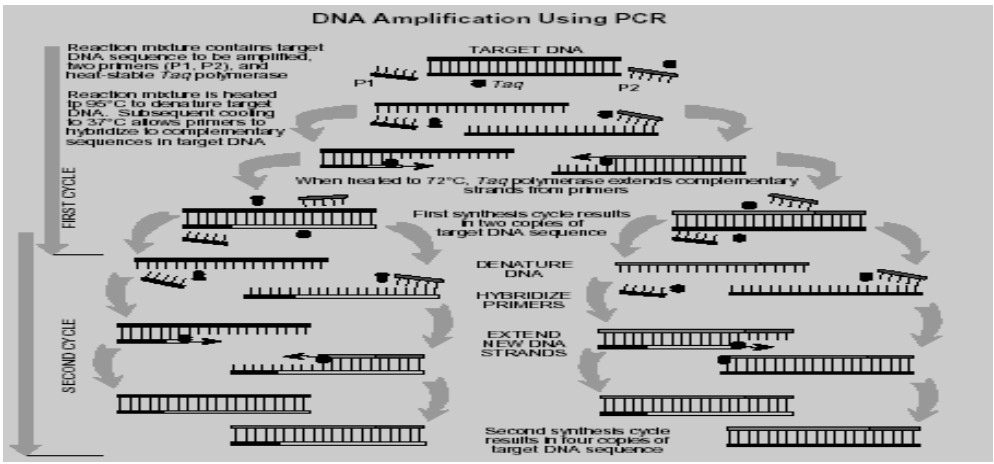

Fig. 1.2: Polymerase Chain Reaction [22]

xii. Gel-Electrophoresis

Second method for separation of specified DNA sequence by their size, not by their content is called Gel-Electrophoresis in fig. 1.3. Gel-Electrophoresis have

the capability to sort DNA sequence by size same as sorting algorithm does on the numeric or alphabetic chars. In the Gel-Electrophoresis, movement of charged molecules in the electric area is called Electrophoresis. DNA molecules hold a negative charge and when they are inserted in the electric area, they tend to move to the positive charge. Percentage of migration depends upon size and electric charge [32].

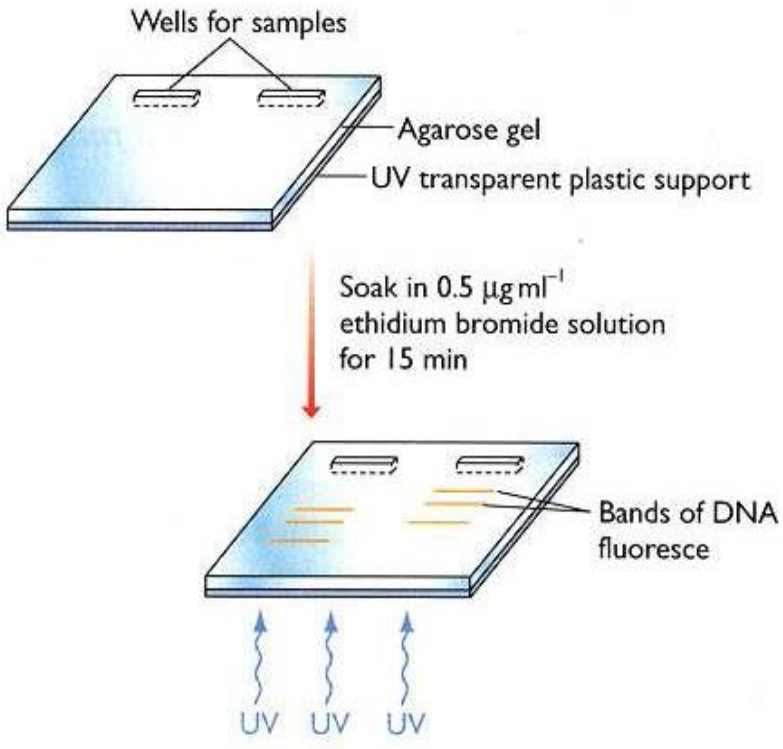

Fig. 1.3: Gel-Electrophoresis [22]

1.4 Binary coding scheme

Conversion of DNA sequence into binary form is known as binary coding scheme. In this thesis, 00, 11, 01, and 10 are used for the DNA bases A, T, C, and G respectively.

For example, 0000110110100011 is a binary sequence for "AATCGGAT" DNA sequence.

1.5 RNA

RNA (Ribonucleic Acid) consists from four bases same as DNA except that Thymine (T) replaced by Uracil (U). These bases are given below:

i. Adenine (A),

ii. Cytosine (C),

iii. Uracil (U), and

iv. Guanine (G)

Where A, C, U, and G are abbreviated form of Adenine, Cytosine, Uracil and Guanine respectively.

1.6 RNA structure

In RNA structure, each base contains a ribose sugar and carbons numbered 1′to5′. From nucleotides A, C, G, and U, one of them is tied to the 1′ position. Adenine and guanine form purines, whereas cytosine and uracil form pyrimidines. Fig. 1.4 shows RNA structure.

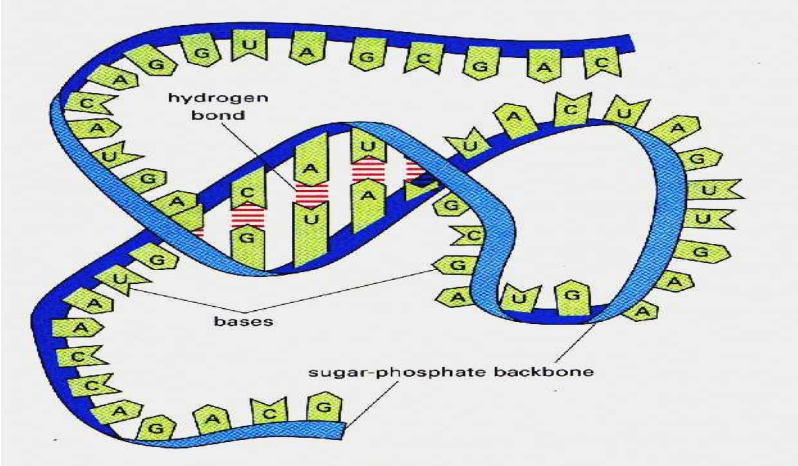

Fig. 1.4: RNA structure [20]

1.7 Protein

Proteins are large molecules or macromolecules consist of amino acids. Proteins are used to perform big functions of the array in the living organisms, it includes repeating DNA responding to stimuli, catalysing metabolic reactions, and transporting molecules from one place to another [21]. Fig. 1.5 shows protein structure in 3-D mode.

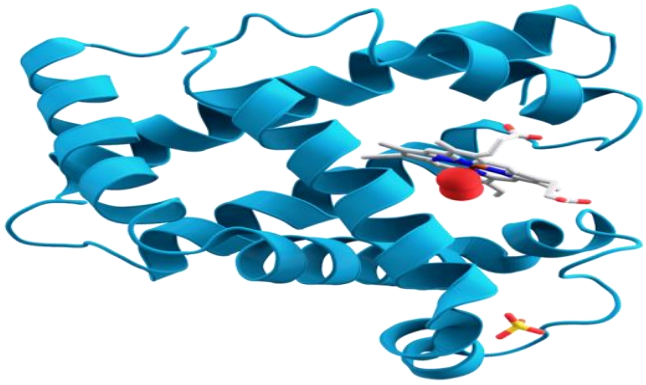

Fig. 1.5: Protein structure [21]

1.8 Central Dogma of Molecular Biology

Conversion of DNA sequences into a protein sequence is called "Central Dogma of Molecular Biology (CDMB)" in fig. 1.6. Watson and Crick [10] introduced the concept of CDMB with two different techniques known as Transcription and Translation. Transcription is a process of converting DNA to RNA, whereas RNA to Protein conversion is called as Translation [36]. Compounding of any three RNA bases out of 4 bases makes the codon table, and there is a corresponding protein for it. Using this combination $4 \times 4 \times 4=64$ codons can be generated. Each codon can be employed for one amino acid as shown in table 1.1. There is single amino acid name is referred by more than one codon, the unique code for each codon made by the suffix with integer in the amino acid first letter [10].

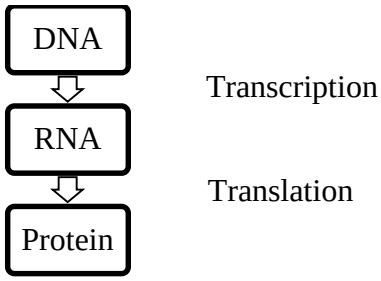

Fig. 1.6: Central Dogma of Molecular Biology

Table 1.1: Codon table of RNA for matching Amino acids [13]

Nucleotide Codon	**Amino Acid Name**	**Amino Acid Code**
UUU	**Phenylalanine**	F
UUC		F1
UUA	**Leucine**	L
UUG		L1
CUU		L2
CUC		L3
CUA		L4
CUG		L5
.	.	.
.	.	.
.	.	.
GCU	**Alanine**	A
GCC		A1
GCA		A2
GCG		A3

1.9 Research motivation

Cryptography is a perspective that converts plaintext into a non-understandable format (ciphertext) by using cryptographic algorithms. Multifarious symmetric and asymmetric key cryptography algorithms such as DES, TDES, AES, and RSA exist for

the metamorphosis of plaintext into the ciphertext. Conventional cryptographic algorithms are incompetent to endow security for the large amount of data on the growing internet and can be easily breakable by intruders for malicious purposes. There is an exigency of new cryptographic algorithm emerged for the information security that can bestow strong ciphertext and protects unauthorized access to data. Therefore, the field of DNA computing emerges. An efficient direction of achieving information security can be termed as *DNA-BASED Cryptography*.

In DNA-BASED Cryptography, DNA used in major two roles:

i. DNA as a medium for the encryption keys because DNA sequence has two fundamental properties: There is virtually no difference between real DNA sequence and fake DNA sequence, and the second is 163 million DNA sequences publically available worldwide. In last one decade, powerful DNA keys based cryptographic algorithms proposed [15-17]. Ciphertext of these algorithms is more complex than the ciphertext of normal keys.

ii. DNA operations like Hybridization, Indexing, and PCR have the capability for a solution of the application in the fields as data hiding, image encryption, and steganography [4, 6]. DNA contains some fascinating properties such as high storage capacity, vast parallelism, and high energy efficiency that are a boon for the data hiding. DNA high storage capacity provides security for fast growing internet data, whereas silicon-based computers are unable to provide security for fast growing data. Hence DNA operations based cryptographic algorithms can be the best solution in the future for growing internet data than conventional algorithms.

1.10 Thesis organization

Thesis work comprises in 5 chapters.

Chapter-1 *Introduction:* Introduction about cryptography and background details of DNA discussed in this chapter. It also contains the motivation behind DNA-BASED cryptography as an encryption algorithm.

Chapter-2 *Literature Survey:* History of cryptography and review of various DNA-BASED cryptographies, those are proposed by various researchers explained in this chapter.

Chapter-3 *Problem Formulation:* Problem statements formed in this chapter.

Chapter-4 *Proposed Approach:* DNA-BASED cryptography approaches explained in this chapter for data security.

Chapter-5 *Result and Discussion:* Result of proposed approaches and discussions between them have given in chapter 5.

Chapter-6 *Conclusion and future scope:* It contains conclusion based on this thesis work and future scope on the DNA computing.

Chapter 2

Literature Survey

2.1 Cryptography overview

Cryptography or Data hiding [12] is an approach for transforming confidential data into an unreadable format on online and offline. Data hiding protects the private or confidential data by ciphering data using various symmetric and asymmetric key cryptography methods such as DES, TDES, RSA, and AES, etc [29-30]. On the other hand, cryptanalysis [23] is reverse technique of cryptography for decipher the ciphertext by analysing ciphertext. Cryptanalysis provides strong ciphertext by applying cryptanalysis to encrypted data for strengthening ciphertext. Cryptanalysis mostly used by hackers for breaking the ciphertext for illegal use of data. Both cryptography and cryptanalysis are necessary for developing a powerful cryptographic algorithm [23]. Process of cryptography is shown in fig. 2.1 which contains both encryption and decryption processes.

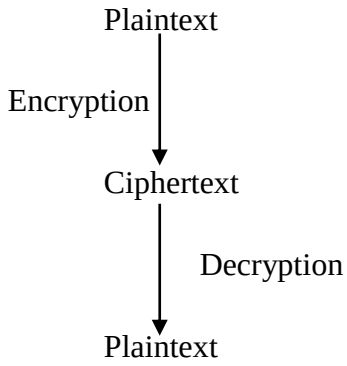

Fig. 2.1: Cryptography

Plaintext: It is source message or information which is used as an input to the encryption process.

Encryption: An approach or method which is performed by data hiding algorithms to convert plaintext into ciphertext.

Ciphertext: An encrypted message which sends to the receiver using private or public channel.

Decryption: An inverse encryption process applies on ciphertext at receiver end for obtaining original message.

Cryptanalysis breaks the ciphertext for obtaining original message whereas, Cryptography strengthen the ciphertext. In today time, many application tools are available for breaking the ciphertext like Frequency analysis, Morse code, Caesar shift, Substitution [27]. Persons involved in cryptanalysis called as cryptanalysts or hackers or intruders. Online and offline many cryptographic algorithms are available for data security that follows cryptographic procedures.

2.2 Classification of procedures

Classification of cryptographic procedures is based upon various cryptographic algorithms. Symmetric key algorithm uses identical key for encryption and decryption at both sender and receiver end whereas, asymmetric key algorithm uses distinct keys for encryption and decryption processes. On the other hand, cryptography algorithm uses hash procedure where no key required. Some cryptographic procedures are as follow:

1). Symmetric key procedure

2). Asymmetric key procedure

3). Hash procedure

2.2.1 Symmetric key algorithm

Symmetric key algorithm referred as private key algorithm. In this algorithm, plaintext gets convert into ciphertext that is non-understandable for human by using an encryption process at sender end. After that, ciphertext is shared to receiver with symmetric key using the communication channel. Receiver applies inverse encryption process to ciphertext with the same key and gets the original plaintext. Fig. 2.2 shows process of symmetric key algorithm [33].

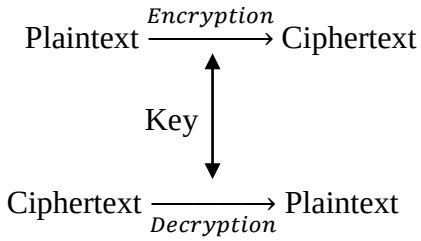

Fig. 2.2: Symmetric key cryptography

2.2.2 Asymmetric key algorithm

Asymmetric key algorithm or public key algorithm proposed in 1975, uses a combination of two keys for encryption and decryption processes at both ends. One is a private key or secret key and second one is a public key. Secret key used by receiver to decipher the encrypted data for obtaining plaintext whereas, public key used by sender to convert plaintext into ciphertext. Both ends try to make private key confidential to the outside world whereas, public key available for the outside world. Fig. 2.3 shows asymmetric key algorithm [38].

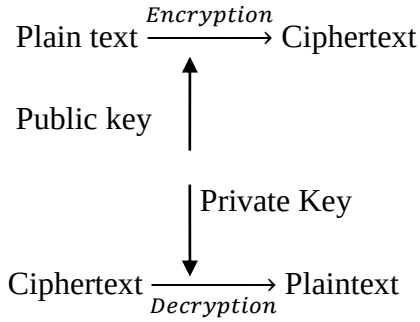

Fig. 2.3: Asymmetric key cryptography

2.2.3 Hash algorithm

Hash algorithm or one way transformations or message digests are mathematical conversion function. It takes the message of random length and transform message into binary bits and calculates the number of bits.

Following are the properties of Hash function as [23]:

i. For any random message P, finding hash function *X* (P) is an easy task because hash function takes fairly less time.
ii. It is impossible to find out message P, from any given hash function.

iii. It is more common that legion changing value of P will be converted to the identical single hash function value *X* (P), it is technically not possible to obtain two different input values that hash to the identical values.

2.3 DNA Cryptography

Researchers are working on diverse areas by DNA Computing as data and image encryption, automata theory, optimization and scheduling problem for finding a better solution. Cryptography with use of DNA as DNA operations or DNA keys or both is referred as DNA cryptography. DNA cryptography has been taken by various researchers as an emerging research in the field of security [35, 40].

2.3.1 Cryptography with DNA operations

DNA operations like Indexing, Hybridization, Annealing, and Polymerase Chain Reaction (PCR) have the ability to provide powerful data encryption technique. Details of each operation already explained in chapter 1. In the recent years, DNA operations based cryptographic algorithms have been proposed [4, 6].

2.3.2 Cryptography with DNA keys

We exploit DNA as a medium for the encryption keys because DNA sequence has two fundamental properties: There is virtually no difference between real DNA sequence and fake DNA sequence. Secondly is 163 million DNA sequences are publically available worldwide. In last one decade, powerful DNA keys based cryptographic algorithms proposed [8-9, 18]. Ciphertext of these algorithms are more complex than the ciphertext of normal keys.

2.3.3 Cryptography with DNA keys and DNA operations

Combination of both DNA keys and DNA operations make it a powerful cryptography algorithm in the laboratory [5]. These techniques are already described in section 2.2.1 and 2.2.2 respectively. Combination of techniques makes more complex algorithm than a single one.

2.4 Various algorithms used DNA Computing

Some of following algorithms are given in section 2.4.1, 2.4.2 and 2.4.3.

2.4.1 Use of DNA Operations in complex problem

In last two decades, researchers have solved complex problem like SAT, DES by use of DNA.

1) Lipton [3] solved NP-complete SAT problem using the DNA computing. Lipton proposed an encoding scheme using DNA computing scheme and solved formulas of SAT problem with small numbers of variables.
2) Boneh *et al.* [4] proposed an approach using DNA technology to crack Data Encryption Standard (DES). They stated that the proposed algorithm can crack any cryptosystem that uses keys size less than 64-bits and given encryption circuit is not too large. They described that DES can be cracked in 916 steps where each step equal to 32 extractions and 10-extractions steps will take one day for processing. Hence they concluded that the proposed algorithm will take roughly 4-months to crack DES.
3) Chen *et al.* [6] proposed that hard problem can be solved by DNA computing using various DNA operations. They solved DES by using the DNA computing. Chen *et al.*'s proposed algorithm included three functions: initial function for initialization of the key space with every possibility, Encryption process, and third is to detect right corresponding key. Algorithm referred as molecular sticker algorithm which includes enzymes, short memory strands, and tubes.

2.4.2 Use of DNA sequences in cryptography

1) MingXin *et al.* [5] introduced a symmetric-key DNA cryptosystem using the inclusion of DNA biotechnology, microarray into the cryptography technologies. In MingXin *et al.*'s algorithm, encryption and decryption key formed by use of DNA probes and ciphertext embedded into microarray (DNA chip). Security of algorithm based on the advancement of DNA chip.
2) Kar *et al.* [8] proposed an algorithm for the reliable, secure, and efficient data transmission using DNA keys. They took three keys for algorithm, two keys for encryption and third session key for sharing encryption key to the receiver. In

the algorithm, initially Plaintext got convert into binary bits using string to binary conversion process and a long DNA sequence also got convert into binary bits using binary coding scheme as an encryption key. After many XOR and addition operations had applied on 64-bits plaintext and 64-bits key, they obtained ciphertext.

3) Shiu *et al.* [15] introduced three cryptographic approaches such as Insertion method, Complementary pair method, and Substitution method using the properties of DNA sequences. They showed that substitution method is better than other two approaches.
4) Liu *et al.* [16] proposed a new data hiding approach using DNA sequence and encrypted each character of ciphertext in a word file. In Liu *et al.*'s algorithm, plaintext got convert into DNA sequence using DNA coding. Chebyshev maps [16] generate two pseudorandom DNA sequence X_{XOR} and Y_{Primer}, and generate one time key using selected DNA sequence key. X_{XOR} added with message DNA sequence and then Y_{Primer} got attach with generated output. After some shift operations had applied to the previous output, they got ciphertext. This ciphertext encrypted in the word file and after conversion of word to PDF. The generated PDF sent to the receiver as final ciphertext.
5) Mandge *et al.* [9] proposed a strong ciphertext cryptographic algorithm using inclusion of matrix manipulation and key generation technology in the DNA computing. Initially, plaintext got convert into mini-cipher using various XOR and shifting operations. They stated that every time when we apply the same process to same plaintext, different mini-cipher for same plaintext because inclusion of secure key generation in this process. Secure key generation generate a different key every time. After this mini-cipher got convert into final ciphertext by biotechnologies with inclusion of DNA primers.
6) Taur *et al.* [17] improved the substitution method of Shiu *et al.* [15] using the table lookup substitution method (TLSM). Shiu *et al.* [15] method based on the 1-bit complementary rule. They proposed extended substitution method with 2-bits complementary rule. Using TLSM, 2-bits message got convert into corresponding letter through rule table that got substitute in reference DNA sequence, whereas Shiu *et al.* [15] method substitute only 1-bit in the reference DNA sequence. TLSM generates compress ciphertext than Shiu *et al.* [15] approach.

7) Huang *et al.* [19] improved Shiu *et al.* [15] approach and proposed a new reversible data hiding method using the histogram technique. They used DNA sequence for hiding plaintext or message. Initially, DNA sequence converted into binary bits, which also converted into decimal numbers. A histogram made using these decimal numbers. In this approach plaintext is encrypted into histogram. They stated that modification rate of their proposed algorithm is 69% less than Shiu *et al.* [15] approach.
8) Abbasy *et al.* [18] introduced a novel data hiding algorithm where message hidden in DNA sequence using software point of view. First of all binary message converted into DNA sequence using DNA coding then various complementary rules are applied to the DNA sequence and a fake DNA sequence is generated. Then a reference DNA sequence has been used to encrypt this fake DNA sequence into reference DNA sequence by various indexes of fake DNA sequence.

2.4.3 Use of CDMB in Cryptography

Sadeg *et al.* [7] worked on Central Dogma of Molecular Biology (CDMB) and designed an algorithm that works better than AES algorithm under various attacks. This proposed algorithm involves three phases. In the first phase, plaintext get convert into bits using the XOR operation with sub key that was generated by sub key generator. In second and third phase long sequence of bits converted into ciphertext by Transposition, CDMB, Permutation, and XOR operations.

2.5 Summary of various algorithms in DNA Computing

Summary of various cryptographic algorithms will be discussed in the table 2.1 using various factors like security, DNA keys, CDMB, DNA computing,

Table 2.1: Review of algorithms using DNA

S. No.	Algorithms	Type of operator	Advantage	Disadvantage
1.	Breaking DES Using a Molecular Computer [4]	DNA computing	It can crack any cryptosystem that uses the key size less than 64-bits.	1. It will not work for algorithm that is having key size greater than 64-bits. 2. It takes too much time for crack cryptosystem.
2.	Efficient DNA Sticker Algorithms for DES [6]	DNA computing	Ciphertext is more secure than conventional DES ciphertext.	Exists theoretically but more difficult to implement in reality.
3.	Symmetric-key cryptosystem with DNA Technology [5]	DNA sequence	Use of DNA chip ciphertext more secure than silicon computer.	Security of algorithm based on the advancement of DNA chip.
4.	An Improved Data Security using DNA Sequencing [8]	DNA sequence	Algorithm is Reliable, secure, and efficient for certain attacks.	Ciphertext contains extra bits that are not part of actual ciphertext.
5.	Data hiding methods based upon DNA sequences [15]	DNA sequence	Algorithms are simple and safe for certain attacks.	Ciphertext of Substitution algorithm is

				easily breakable.
6.	A novel data hiding method based on deoxyribonucleic acid coding [16]	DNA sequence	Use of word and PDF file ciphertext is more powerful.	Complex computation.
7.	DNA Base Data Hiding Algorithm [18]	DNA sequence	Algorithm is simple and easily implementable.	Easily breakable.
8.	A DNA-based data hiding technique with low modification rates [19]	DNA sequence	No need of compression and expansion techniques for the ciphertext. Modification rate is 17 times lower than Shiu *et al.* [15].	Need to understand the concept of histogram.
9.	DATA hiding in DNA sequences based on Table Lookup Substitution [17]	DNA sequence	Each DNA base contains information of the message in ciphertext.	Complex calculation.
10.	An encryption algorithm inspired from DNA [7]	CDMB technique	Ciphertext is robust against certain attacks.	Ciphertext is breakable.

Chapter 3
Problem Formulation

DNA computing allows researchers to find the solution of complex problems that are unable to solve on silicon-based computer in polynomial time. It is used in various fields such as optimization problem, image and signal processing, data hiding, complex problem, clustering and forecasting [8]. DNA computing provides solution in lesser time than silicon computer. This chapter provides elaborated description of problems and analysis of loopholes in existing research works.

3.1 Problem statement

DNA cryptography allows the security for a large amount on the internet where data transactions and uses of data on the internet are growing every second. Traditional cryptography algorithms are not able to provide security in the fast growing internet. This chapter provides cryptography algorithms that can be the best choice in the fastest growing internet era. Main motive of this study to gives security on data on online and offline where users should not be afraid of confidentiality of their private data.

The principal objectives of this thesis are:

i. To analyse the various DNA cryptography approaches.
ii. To provide an improve cryptographic algorithm with robust ciphertext.
iii. To provide an improve data hiding method using new approach.

In chapter 4, we have two proposed cryptographic algorithms which make data confidential from unauthorized user and provide high security than conventional algorithms. In last, a comparative study of both algorithms will be carried out.

Chapter 4
Proposed Work

In this section, two improved propose approaches are explained. Both approaches take message as a plaintext and converts into ciphertext as a protein form. Protein sent to the receiver end using public channel where receiver applies inverse encryption process to obtain message or plaintext. Each proposed cryptographic algorithm has been briefly explained with the encryption & decryption algorithms.

4.1 Approach 1-*An Improved Algorithm for Data Hiding using DNA Sequence and Central Dogma of Molecular Biology*

In this approach, sender shares binary coding scheme and codon table of RNA as a key and also use the shared symmetric key in the form of the database name (Like GENBANK) with index no. (Like 0, 1 ...) to the receiver by secure channel.

The improved propose algorithm exploits various permutation and substitution networks from DES algorithm [11]. We use encrypt and decrypt functions for the sender and receiver end.

In the encryption algorithm, initially plaintext will be converted into bits using the hexadecimal value. A long selected DNA sequence will be divided into three 32-bases sequences as S_1, S_2 and S_3 respectively. Then 32-bases DNA sequences will be converted into three 64-bits keys K_1, K_2 and K_3 respectively using binary coding scheme. Select 64-bits plaintext as block for encryption. Then encryption function will be called with the parameters K_1 (64-bits), block for encryption, and 64-bits blank cipher block. Encryption function will give an output as 64-bits cipher block. Decryption function will be called with parameters K_2, previous cipher block as input, and 64-bits blank plaintext block. After execution, call again encryption function same as previous encryption with previous decryption call output as input using key K_3. We will get 64-bits partial ciphertext block. Apply reverse binary coding scheme into partial block and gets fake DNA sequence. Next we apply CDMB technique into fake DNA sequence will produce final cipher text in the form of protein. We send this protein block to the receiver end.

In the decryption algorithm, we will apply the reverse CDMB technique to the ciphertext block as the first level decryption. It will generate fake DNA sequence. After that, apply binary coding technique it will give partial cipher text in bits. Receiver will generate three 64-bits decryption keys K_1, K_2 and K_3 using sender shared information. Select 64-bits block of cipher and call the decryption function with parameters such as K_1, cipher block and blank 64-bits plaintext block. After execution, we will receive 64-bits plaintext. Call encryption function with parameters as K_2, previous output block as input and 64-bits blank cipher block and get the output as 64-bits block. At last, call again decryption function with previous encryption call output as input using key K_3. It will give final 64-bits plaintext block. Now apply inverse hexadecimal coding to get plaintext P.

Procedure. Encrypt (Key *K*, Plaintext *P*, Ciphertext *C*)

Input:	Key *K*, Plaintext *P*.
Output:	Ciphertext *C*.
Step-1:	Apply permutation on K and get K' as 56-bit.
Step-2:	Divide K' into X_0 and Y_0 as 28-bits each.
Step-3:	Create sixteen blocks X_n and Y_n, 1≤n≤16. Each pair of blocks X_n and Y_n will be formed using the left shift of the previous block X_{n-1} and Y_{n-1} respectively.
Step-4:	Create sixteen sub-keys, 1≤n≤16 from each of the concatenated pairs.
Step-5:	Apply initial permutation (*IP*) on 64-bits block of *P*.
Step-6:	Divide *IP* into P_0 and Q_0 as 32- bits each.
Step-7:	Obtain pairs P_n and Q_n, where 1≤n≤16; blocks by using two equations. $P_n = Q_{n-1}$ $Q_n = P_{n-1} \oplus X\,(Q_{n-1}, K_n)$ Where $X\,(Q_{n-1}, K_n) = X\,(Q_{n-1}) \oplus K_n$; and $X\,(Q_{n-1})$ = Expansion function.
Step-8:	Obtain $Q_{16}P_{16}$ as 64-bits block.
Step-9:	Apply final permutation IP^{-1} to $Q_{16}P_{16}$ and generate output *C*.

Procedure. Decrypt (Key *K*, Ciphertext *C*, Plaintext *P*)

Input:	Key *K*, Ciphertext *C*
Output:	Plaintext *P*
Step-1:	Apply permutation on *K* and get K' as 56-bit.
Step-2:	Divide K' into X_0 and Y_0 as 28-bits each.
Step-3:	Create sixteen blocks X_n and Y_n, $1 \le n \le 16$. Each pair of block X_n and Y_n will be formed using the left shift of the previous block X_{n-1} and Y_{n-1} respectively.
Step-4:	Create sixteen sub-keys K_n, $1 \le n \le 16$ from each of the concatenated pairs $X_n Y_n$.
Step-5:	Apply initial permutation (*IP*) on 64-bits block of *C*.
Step-6:	Divide *IP* into P_{16} and Q_{16} as 32- bits each.
Step-7:	Obtain pairs P_n and Q_n, where $16 \ge n \ge 1$; blocks by using two equations. $Q_{n-1} = P_n$ $P_{n-1} = Q_n \oplus X(P_n, K_n)$ Where $X(P_n, K_n) = X(P_n) \oplus K_n$; and $X(P_n)$ = Expansion function.
Step-8:	Obtain $P_0 Q_0$ as 64-bits block.
Step-9:	Apply final permutation IP^{-1} to $P_0 Q_0$ and get output *P*.

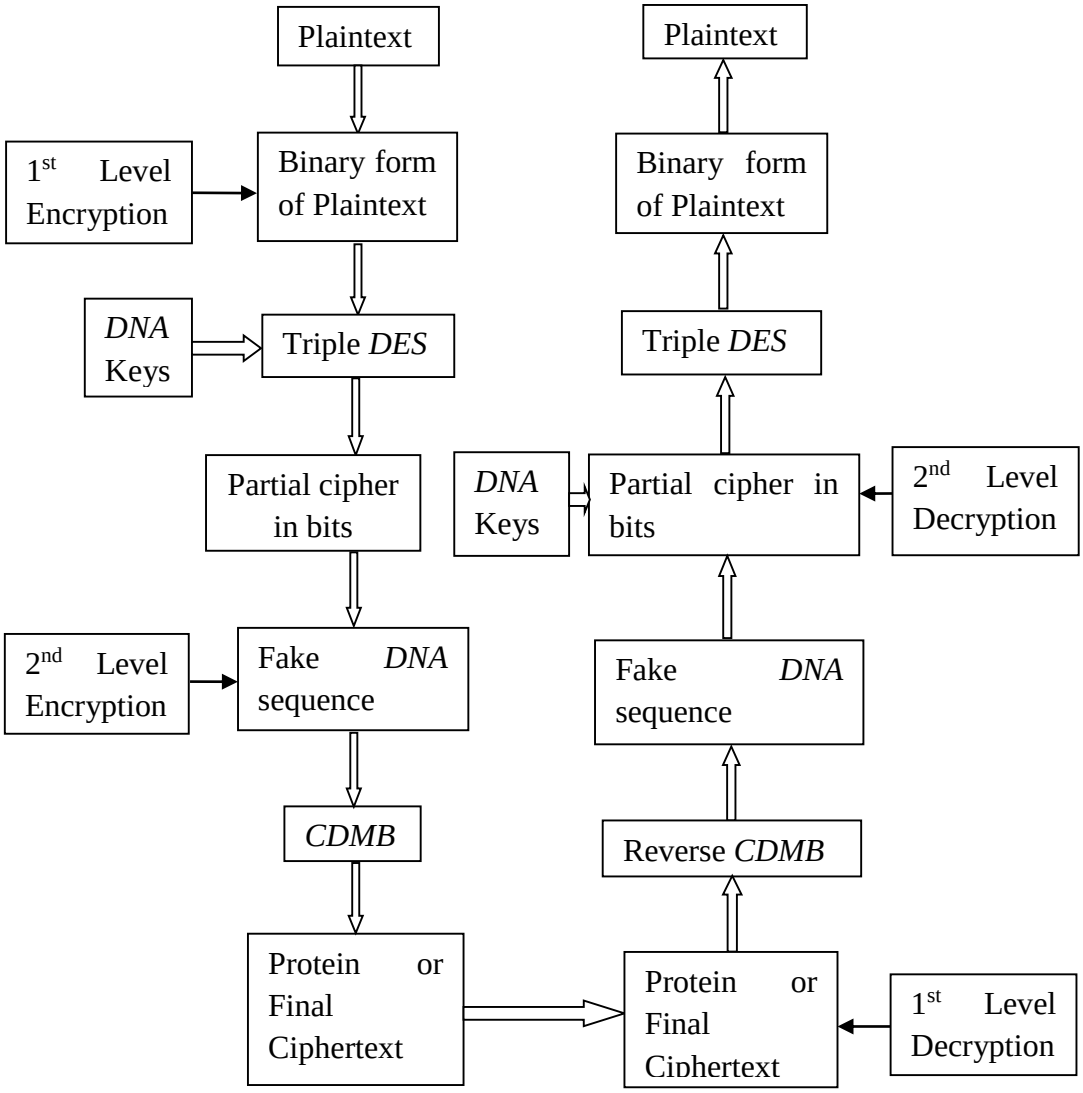

Fig. 4.1: Encryption and Decryption process

Proposed Encryption Algorithm:

Algorithm 1:	**Encryption** (Plaintext *P*, Ciphertext *C*)
Input:	Plaintext *P*, A long DNA sequence ≥ 96 DNA-bases
Output:	Ciphertext *C*
Step-1:	Convert message or plaintext (*P*) in the binary form by using Hexadecimal value.
Step-2:	Select three DNA sequences S_1, S_2 and S_3 from a given DNA sequence each of 32-bases.

Step-3:	Convert S_1, S_2 and S_3 into bits using binary coding scheme as K_1, K_2 and K_3 respectively. Select 64-bits K_1, K_3 as encryption Keys and 64-bits K_2 as the decryption key.
Step-4:	Encrypt (K_1, P, C')
Step-5:	Decrypt (K_2, C', C'')
Step-6:	Encrypt (K_3, C'', C''')
Step-7:	Apply reverses binary coding scheme into partial cipher block C''' and get fake DNA sequence.
Step-8:	Get Protein sequence by applying CDMB into fake DNA sequence.
Step-9:	Send PROTEIN sequences to the receiver as encrypted plaintext.

Proposed Decryption algorithm:

Algorithm 2: Decryption (Ciphertext *C*, Plaintext *P*, Database *D*, Index *X*)	
Input:	Ciphertext *C*, Shared key information (Database *D*, Index *X*)
Output:	Plaintext *P*
Step-1:	Apply reverse CDMB into ciphertext (*C*) and get fake DNA sequence.
Step-2:	Using binary coding scheme get partial ciphertext in bits.
Step-3:	Select three DNA sequences S_1, S_2 and S_3 from publically available DNA databases based upon the sender shared information as *D*, *X*.
Step-4:	Convert S_1, S_2 and S_3 into bits using binary coding scheme. Consider S_1, S_3 64-bits as decryption keys K_1, K_3 respectively and 64-bits S_2 work as the encryption key K_2.
Step-5:	Decrypt (K_1, C, P')
Step-6:	Encrypt (K_2, P', P'')
Step-7:	Decrypt (K_3, P'', *M*)
Step-8:	Get original plaintext *P* using reverse hexadecimal coding on *M*.

4.2 Approach 2-*An Improved Substitution method for data encryption using DNA Sequence and CDMB*

Proposed scheme uses a DNA encoding table 4.1 for plaintext to DNA sequence conversion. In the table, each character is formed by a combination of three DNA bases. Proposed algorithm also uses various complementary rules for boosting the ciphertext because a right complementary rule follows property like $V(W) \neq V(V(W)) \neq V(V(V(W))) \neq V(V(V(V(W))))$, where 'W' is any string or alphabet. Following complementary rules for the proposed approach are used: (AT) (CA) (GC) (TG), Means V (A) = T

Table 4.1: DNA Encoding [16]

A=CGA	H=CGC	O=GGC	V=CCT	2=TAG	...	9=GCG
B=CCA	I=ATG	P=GGA	W=CCG	3=GCA	…	=ATA
C=GTT	J=AGT	Q=AAC	X=CTA	4=GAG	…	,=TCG
D=TTG	K=AAG	R=TCA	Y=AAA	5=AGA	…	.=GAT
E=GGT	L=TGC	S=ACG	Z=AAT	6=GGG	…	:GCT
F=ACT	M=TCC	T=TTC	0=TTA	7=ACA	…	;=ATT
G=TTT	N=TCT	U=CTG	1=ACC	8=AGG	…	_=ATC

Receiver knows DNA encoding table, complementary rules, binary coding scheme and RNA codon table using a secure channel made by the sender. Proposed algorithm exploits substitution method from Shiu *et al.* [15].

In the encryption process, Get a plaintext and convert it into the artificial DNA sequence using DNA encoding. Binary coding scheme will transform artificial DNA sequence into bits. Select a long DNA sequence (from GENBANK), where length should be greater or equal to the number of bits. After that, a set of random integers is generated, where number of elements in the set equal to binary bits and range of each element should be from 1 to length of DNA sequence. Now substitution method is applied using

the complementary rules to the DNA sequence, and it will provide fake DNA sequence. CDMB will convert fake DNA sequence to protein sequence. Send protein sequence to the receiver end.

Encryption process:

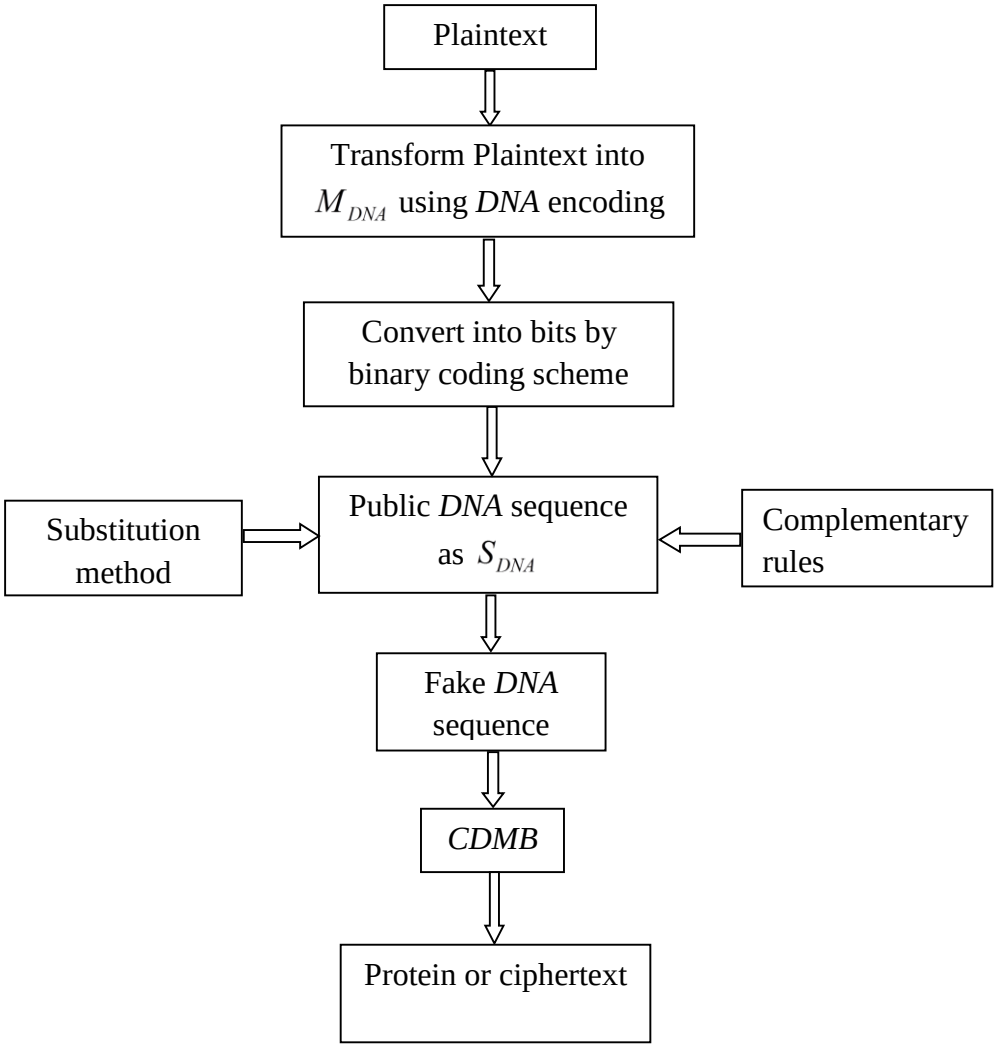

Fig. 4.2: Encryption flow chart

In the decryption process, received protein sequence will be converted into the fake DNA sequence using inverse CDMB technique. Now apply the substitution method to the fake DNA sequence using the sender's shared DNA sequence and complementary rules, it will generate a binary sequence. It uses reverse binary coding into the binary sequence; it will produce another fake DNA sequence. At last, inverse DNA encoding will generate final plaintext to the fake DNA sequence.

Decryption Process:

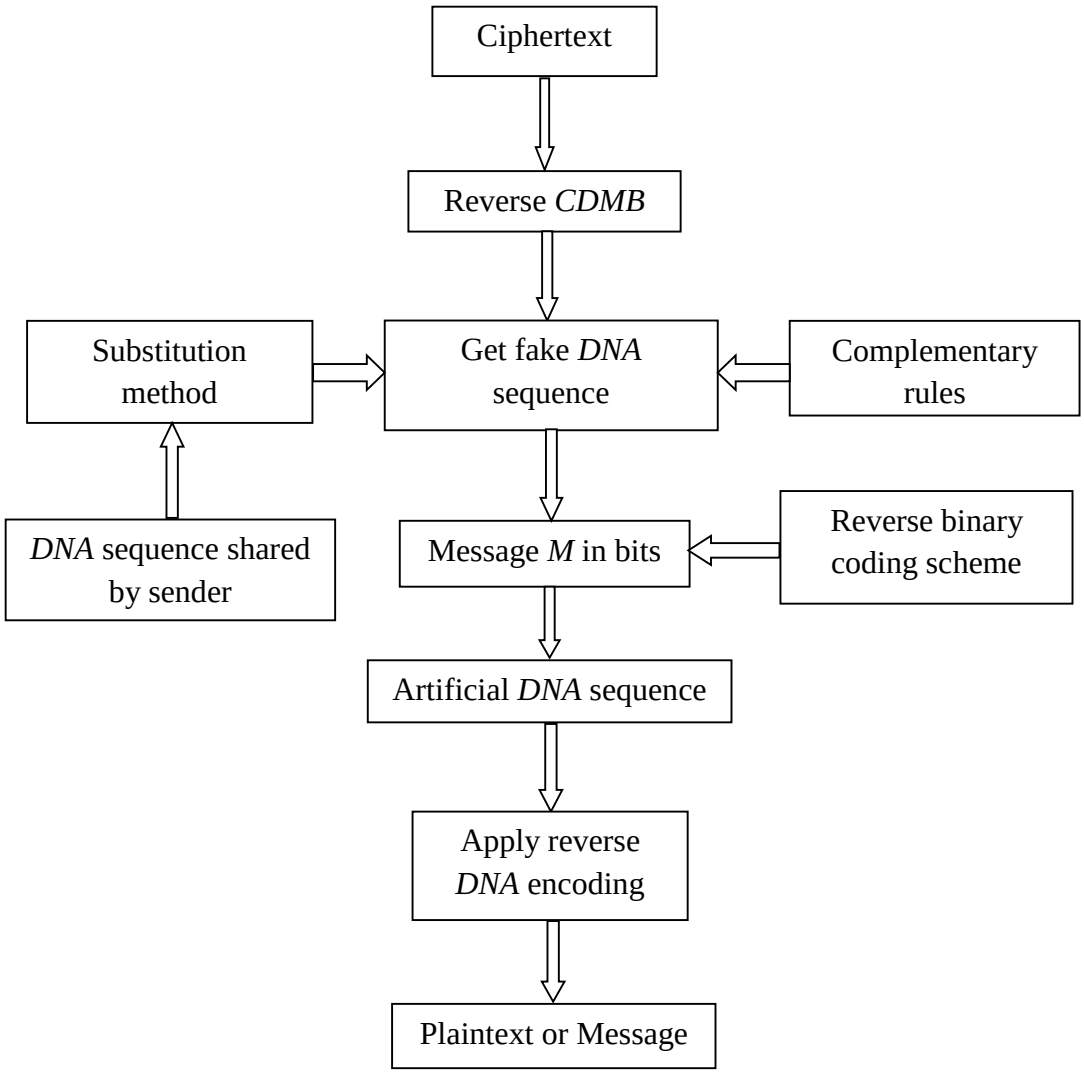

Fig. 4.3: Decryption flow chart

Proposed Encryption Algorithm:

Algorithm 3: Encryption (Message *M*, Ciphertext *C*)	
Input:	Message *M*
Output:	Ciphertext *C*
Step-1:	Get message *M* and convert into DNA sequence as M_{DNA} using DNA encoding.
Step-2:	Convert M_{DNA} into bits using binary coding scheme as M' and calculate the length of M' as *X*.
Step-3:	Select a DNA sequence as S_{DNA} from publically available DNA database (Like GENBANK) where $S_{DNA} \geq X$.

Step-4: Calculate the length S_{DNA} as *Y*. Generate a set *P* with *X* distinct random numbers range from 1 to *Y* and sort *P* in increasing order.

Step-5: Convert S_{DNA} into S'_{DNA} using following substitution code:

Begin

For i←1 to *Y*

If i == P_j and M'_j==1 and 1≤ j ≤ *X*, then S^i_{DNA} ←V (S^i_{DNA})

Else

If i == P_j and M'_j==0 then $S^i_{DNA} \leftarrow S^i_{DNA}$

Else

If i≠ P_j then S^i_{DNA} ← V (V (S^i_{DNA}))

End If

End If

End If

End For

End

Step-6: Convert fake DNA sequence S'_{DNA} into RNA sequence by replacing *T* with *U*.

Step-7: By the amino acid code convert RNA sequence into PROTEIN sequence.

Step-8: Send protein sequence to the receiver end.

Proposed Decryption Algorithm:

Algorithm 4: Decryption (Ciphertext *C*, DNA-sequence *S*, Plaintext *P*)

Input: Ciphertext *C* as Protein, DNA sequence as *S*

Output: Plaintext *P*

Step-1: Convert ciphertext *C* (Protein) into RNA sequence using RNA codon table.

Step-2: Get fake DNA sequence as S' by replacing U with T in RNA sequence.

Step-3: Calculate the length of S' and S as X and Y respectively.

Step-4: Get message M from S'using following substitution code:

Begin

d ←1

For c ← 1 to X

If S'_c == S_c then M_d ← 0 and d ←d + 1

Else

If S'_c == V (S_c) then M_d ← 1 and d ←d + 1

End If

End If

End For

End

Step-5: Convert M into DNA sequence using reverse binary coding.

Step-6: Get plaintext by convert DNA sequence into P using reverse DNA encoding.

Limitation of proposed algorithms: In the proposed approach, left one and two RNA nucleotides will be same in the final ciphertext (Protein) for RNA to Protein conversion, because there is no amino acid code for left one and two RNA nucleotides.

4.3 Comparison between the proposed approaches

Similarities and differences of both proposed approaches are given in table 4.2 based on the parameters like Ciphertext, CDMB technique, DNA sequence, etc.

Table 4.2: Comparison between proposed approaches

Parameters	Improved TDES	Improved Substitution Method
CDMB technique	Yes	Yes
Ciphertext	In protein	In protein
DNA sequence	As key	As information carrier
Calculation	Very complex	Simple
Binary Coding Scheme	Yes	Yes

DNA Encoding	No	Yes
Complementary rules	No	Yes

Chapter 5

Result and Discussion

Results and discussions of both proposed approaches are given below with the help example of algorithms on the input data.

Approach 1-*An Improved Algorithm for Data Hiding using DNA Sequence and Central Dogma of Molecular Biology:*

The proposed algorithm is implemented in C Language. Fig. 5.1 shows the result of encryption using proposed algorithm at sender end.

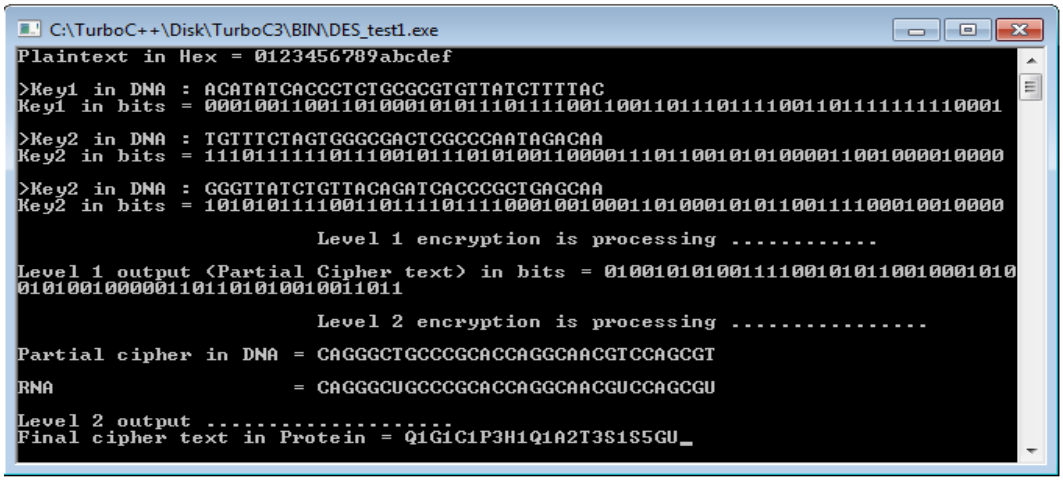

Fig. 5.1: Encryption output

Fig. 5.2 shows the result of decryption using the inverse of the encryption process at the receiver end.

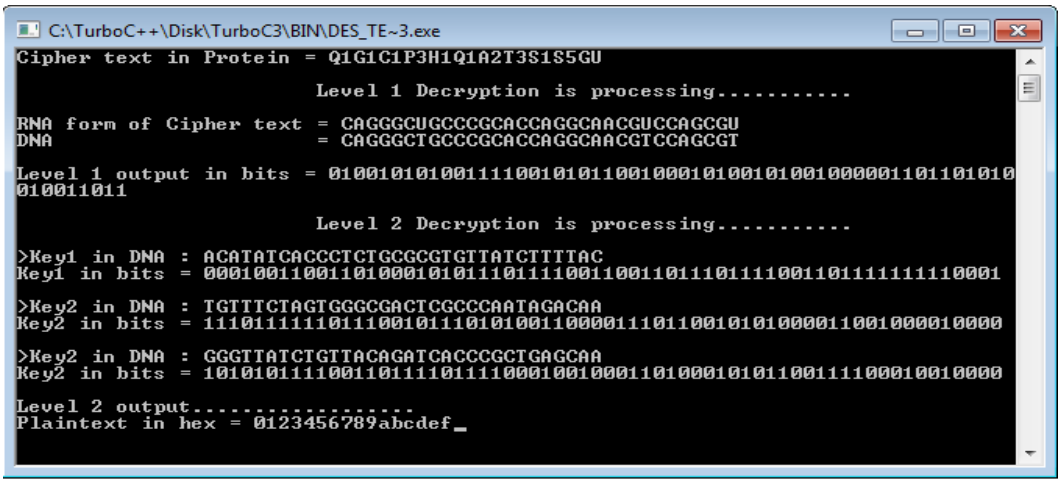

Fig. 5.2: Decryption output

Table 5.1 shows a comparison between the TDES and Improved TDES (proposed algorithm).

Table 5.1: Comparison between TDES and Improved TDES

Factor	TDES	Improved TDES
Security	Lesser than improved TDES	More secure because use of powerful DNA keys and CDMB technique.
Cryptographic strength	Lesser than improved TDES.	More because use of 2 levels encryption.
Encryption and Decryption running time	Less than improved TDES.	Slow, because the use of biotechnology.
Power	Power is required.	No power need, because DNA is power efficient.[9]
Availability of keys	Shortage of keys	163 million DNA keys are available worldwide.[15]

Vulnerability of key	Easily deductible and breakable.	No weakness
Strength of key	Simple than DNA keys	No one can virtually guess the encryption keys. [15]
Complexness of algorithm	Lesser than improved TDES.	More complex using inclusion of biotechnology.

Approach 2-*An Improved Substitution method for data encryption using DNA Sequence and CDMB:*

Encryption process of proposed approach has explained using plaintext "AB" and public *DNA* sequence "AAGCTTACAGACTCCAGGTATGGACTTCAAGT". Result of the encryption process comes in the form of protein that will be sent to the receiver end.

ORIGINAL MESSAGE *M*: **AB**

M_{DNA}: CGACCA

M': 100100101000

S_{DNA}: AAGCTTACAGACTCCAGGTATGGACTTCAAGT

P: {2,3,5,7,8,11,13,15,17,19,20,23}

S'_{DNA}: GTGTTCTCGAATGTCGCATACAG

RNA sequence: GUGUUCUCGAAUGUCGCAUACAG

PROTEIN: V3F1S3N1V1A2Y1AG

Possibility of attack in the proposed approach: For a successful attack in the proposed approach, intruder must attempt successful guess for the following terms:

For (1): $\frac{1}{1.63 \times 10^8}$= right *DNA* sequence from 163 million *DNA* sequence

For (2): $\frac{1}{6}$ = right complementary rules in the 6 possibilities given below:

(AT) (TC) (CG) (GA),

(AT) (TG) (GC) (CA),

(AC) (CT) (TG) (GA),

(AC) (CG) (GT) (TA),

(AG) (GT) (TC) (CA), and (AG) (GC) (CT) (TA)

For (3): $\frac{1}{64}$ = correct guess of *DNA* encoding table possibilities.

For (4): $\frac{1}{24}$ = correct guess of binary coding scheme possibilities.

For (5): $\frac{1}{X}$ = right match of *RNA* codon code.

For example, GCU, GCC, GCA, and GCG codons come in Alanine group, where A, A1, A2, and A3 are Amino acid codes respectively for codons. Therefore, we have 4! Possibilities for Amino acid code for Alanine codons. Hence, value of *X* depends on the number of codons in the specified Amino acid, here *X* =4!

Table 5.2: Showed the comparison between the Shiu *et al.* [15] and proposed approach.

Factor	**Shiu *et al.* Approach [15]**	**Improved Approach**
Security	Lesser than improved approach.	More secure, because the use of DNA encoding and CDMB technique.
Cryptographic strength	Lesser than improved approach.	More because CDMB strengthens the encrypted output.
Complexness of algorithm	Simple Calculation.	More complex by inclusion of CDMB and DNA encoding.

Compactness of ciphertext	Less Compact.	More Compress using CDMB technique.
Possibility of attack	$\frac{1}{1.63 \times 10^8} \times \frac{1}{6}$ or $\frac{1}{3^n}$ [15]	$\frac{1}{1.63 \times 10^8} \times \frac{1}{6} \times \frac{1}{64} \times \frac{1}{24} \times \frac{1}{X}$

Chapter 6

Conclusion and Future Scope

DNA computing is a computing technique based on the DNA, DNA operations and biotechnologies instead of the traditional computing by silicon based computer. DNA computing or bio-molecular computing contains the capability for the solution of the application in the fields as data hiding, image encryption, and steganography, etc. DNA operations are execute in the laboratory using identical conditions with care of various parameters like Pressure, Temperature, Oxygen ratio, etc. Result of operations totally dependent on the environment. In this thesis, we proposed two improved DNA based symmetric key cryptographic algorithms, first extends the Triple Data Encryption Standard algorithm using the concept of DNA computing and second improved data encryption approach based on the Deoxyribonucleic acid that exploits two different techniques: Substitution technique and Central Dogma of Molecular Biology. Both algorithms convert the message into a protein (Ciphertext) using the Deoxyribonucleic acid strands and Central Dogma of Molecular Biology. Vast DNA key space and CDMB technique provide robust ciphertext to resist all brute force attacks.

In the research work, only two DNA-BASED cryptographic algorithms – improved TDES with DNA and improved substitution with DNA using complementary rules have proposed. In the future,

1. We can propose a novel algorithm and improve conventional cryptographic approaches using the DNA operations and DNA sequences. Both DNA operations and DNA sequences are part of DNA computing for problem formulation.
2. Various complex problems such as Graph colouring problem, SAT problem, and TSP, etc. These are unsolvable by silicon computers, but can be solvable by DNA computer in polynomial time.

References

[1]. L.M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science, JSTOR,* 266(5187), 1021-1024, 1994.

[2]. A. Gehani, T. LaBean, and J. Reif, "DNA-based cryptography," *Dismacs Series in Discrete Mathematics and Theoretical Computer Science,* 54, 233-349, 2000.

[3]. R.J. Lipton, "DNA solution of hard computational problems," *Science, New Series,* 268(5210), 542-545, 1995.

[4]. D. Boneh, C. Dunworth, and R.J. Lipton, "Breaking DES Using a Molecular Computer," *DIMACS workshop on DNA computing*, 27, 37-51, 1995.

[5]. L. MingXin, L. XueJia, X. GuoZhen, and Q. Lei, "Symmetric-key cryptosystem with DNA technology," *Science in China Series F: Information Sciences*, 50(3), 324-333, 2007.

[6]. Z. Chen, X. Geng, and J. Xu, "Efficient DNA Sticker Algorithms for DES," *IEEE 3rd International Conference on Bio-Inspired Computing (BICTA),* 15-22, 2008.

[7]. S. Sadeg, M. Gougache, N. Mansouri, and H. Drias, "An encryption algorithm inspired from DNA," *IEEE International Conference on Machine and Web Intelligence (ICMWI),* 344-349, 2010.

[8]. N. kar, A. Majumder, A. Saha, A. Jamatia, K. Chakma, and M. C. Pal, "An Improved Data Security using DNA Sequencing," *Proceedings of the 3rd ACM MobiHoc workshop on Pervasive wireless healthcare,* 13-18, 2013.

[9]. T. Mandge and V. Choudhary, "A DNA Encryption Technique Based on Matrix Manipulation and Secure key Generation Scheme," *IEEE International conference on Information Communication and Embedded Systems (ICICES),* 47-52, 2013.

[10]. U.N. Hussain, T. Chithralekha, A.N. Raj, G. Sathish, and A. Dharani, "A Hybrid DNA Algorithm for DES using Central Dogma of Molecular Biology (CDMB)," *International Journal of Computer Applications*, 42(20), 1-4, 2012.

[11]. Data Encryption Standard (DES), Available at: http://csrc.nist.gov/publications/fips.

[12]. M. Rusia and R.H. Makwana, "Review on DNA Based Encryption Algorithm for Text and Image Data," *International Journal of Engineering Research & Technology (IJERT)*, 3(1), 3182-3186, 2014.

[13]. RNA codon table, Available at: http://en.wikibooks.org/wiki/Proteomics/Protein_Primary_Structure/Genetic_Code.

[14]. DNA structure, Available at: http://ghr.nlm.nih.gov/handbook/illustrations/DNAstructure.

[15]. H.J. Shiu, K.L. Ng, J.F. Fang, R.C.T. Lee, and C.H. Huang, "Data hiding methods based upon DNA sequences," *Information Sciences*, 180, 2196-2208, 2010.

[16]. H. Liu, D. Lin, A. Kadir, "A novel data hiding method based on deoxyribonucleic acid coding," *Computers and Electrical Engineering*, 39, 1164-1173, 2013.

[17]. J.S. Taur, H.Y. Lin, H.L. Lee and C.W. Tao, "Data hiding in DNA sequences based on Table Lookup Substitution," *International Journal of Innovative Computing, Information and Control*, 8(10), 6585-6598, 2012.

[18]. M.R. Abbasy, P. Nikfard, A. Ordi, M.R.N. Torkaman, "DNA Base Data Hiding Algorithm," *International Journal on New Computer Architectures and Their Applications (IJNCAA)*, 2(1), 183-192, 2012.

[19]. Y.H. Huang, C.C. Chang, C.Y. Wu, "A DNA-based data hiding technique with low modification rates," *Multimedia Tools Applications*, 1-13, 2012.

[20]. RNA Structure, Available at: http://www.uic.edu/classes/phys/phys461/phys450/ANJUM04/.

[21]. Protein Structure, Available at: http://en.wikipedia.org/wiki/File:Myoglobin.png.

[22]. M. Sharma, "DNA Computing for Signal Matching," *M.E. Thesis, Thapar University*, 2010.

[23]. A.P. Thiruthuvadoss, "Comparison and Performance Evaluation of Modern Cryptography and DNA Cryptography," *M.S. Thesis, Royal Institute of Technology*, 2012.

[24]. G. Cui, L. Qin, Y. Wang, X. Zhang, "Information Security Technology Based on DNA Computing," *International Workshop on Anti-counterfeiting, Security, Identification*, 288-291, 2007.

[25]. DNA Structure, Available at: http://tigger.uic.edu/classes/phys/phys461/phys450/ANJUM04/.

[26]. DNA Operations, Available at: http://www.thestudymaterial.com/presentation-seminar/computer-presentation/212-DNA-computer-presentation.html?start=2.

[27]. Ciphertext breaking tools, Available at: http://www.richkni.co.uk/php/crypta/.

[28]. M. Zareen, "Enhancement on Implementation of Multi-Prime and Multi-Power RSA Algorithm," *M. E. Thesis, Thapar University*, 2011.

[29]. M. Zareen, A. Kumar, "Comparisons among the Different Algorithms of RSA and its Implementation Issues," *International Conference on Communication and Computing Technologies (ICCCT)*, 2011.

[30]. M. Zareen, A. Kumar, "Implementing Efficient RSA with BigInteger on 2048-bit," *National Conference on the Recent Advances in Electronics and Communication Technologies (RAECT)*, 2011.

[31]. M. Babaei, "A novel text and image encryption method based on chaos theory and DNA computing," *Natural computing*, 12(1), 101-107, 2013.

[32]. T. Mandge and V. Choudhary, "A Review on Emerging Cryptography Technique: DNA Cryptography," *In IJCA Proceedings on International Conference Technology and Computer Science*, 9-13, 2012.

[33]. S. Jain and V. Bhatnagar, "Analogy of Various DNA Based Security Algorithms Using Cryptography and Steganography," *International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 285-291, 2014.

[34]. L. MingXin, L. XueJia, X. GuoZhen, and Q. Lei, "Asymmetric encryption and signature method with DNA technology," *Science China Information Sciences*, 53(3), 506-514, 2010.

[35]. A. Leier, C. Richter, W. Banzhaf, and H. Rauhe, "Cryptography with DNA binary strands," *BioSystems*, 57(1), 13-22, 2000.

[36]. W.L. Tai, C.C.N. Wang, P.C.Y. Sheu, and J.J.P. Tsai, "Data Hiding in DNA for Authentication of Plant Variety Rights," *Journal of Electronic Science and Technology*, 1, 010, 2013.

[37]. T. Hodorogea and I.S. Otto, "Deriving DNA Cryptographic Keys Based on Evolutionary Models," *12th International on Carpathian Control Conference (ICCC)*, 144-147, 2011.

[38]. A. Adhikari, "DNA Secret Sharing," *IEEE Congress on Evolutionary Computation Sheraton Vancouver Wall Centre Hotel*, 1407-1411, 2006.

[39]. M. Borda and O. Tornea, "DNA Secret Writing Techniques," *IEEE conferences on DNA secret writing Techniques*, 452-456, 2010.

[40]. B. Roy, G. Rakshit, R. Chakraborty, "Enhanced key Generation Scheme based Cryptography with DNA Logic," *International Journal of Information and Communication Technology Research*, 1(8), 370-374, 2011.

List of Publications

ACCEPTED

1. R. Gupta and A. Kumar, "An Improved Algorithm for Data Hiding using DNA Sequence and Central Dogma of Molecular Biology," Paper ID 967, International Conference on Emerging Research in Computing, Information, Communication and Applications (ERCICA-14).

2. R. Gupta and A. Kumar, "An Improved Substitution method for data encryption using DNA Sequence and CDMB," Paper ID 1569973413, Second International Symposium on Security in Computing and Communications (SSCC-2014).