Proceedings of the Fourth International Conference on Communication and Electronics Systems (ICCES 2019)
IEEE Conference Record # 45898; IEEE Xplore ISBN: 978-1-7281-1261-9

A Novel Encryption Algorithm based on DNA Cryptography

A.Vikram*, S.Kalaivani*,G.Gopinath**
*School of Computer Science and Engineering, Bharathidasan University, Tiruchirappalli.
**Professor, School of Computer Science and Engineering, Bharathidasan University, Tiruchirappalli.

Abstract: **The process of information security entails securing the information by transferring it through the networks preventing the data from attacks. This way of securing the information is known as cryptography. The perspective of converting the plain-text into non-understandable format is known as cryptography that could be possible using certain cryptography algorithms. The security could not be offered by the conventional cryptographic algorithms that lacks in their security for the huge amount of growing data, which could be easily broken by the intruders for their malicious activities. This gives rise to the new cryptographic algorithm known as DNA computing that could strengthen the information security, which does not provide any intruders to get authorized to confidential data. The proposed DNA symmetric cryptography enhances information security. The results reveal that encryption process carried out on plain-text is highly secured.**

Keywords: DNA cryptography, Information security, DNA computing.

I. INTRODUCTION

DNA computing was mainly emerged to solve the directed Hamiltonian Path Problem (HPP), which was proposed by Adleman. The concept of HPP is to obtain the shortest path that traverses along each and every node and then return to the last node in the bounded path. A DNA sequence could be formed by fixing the source node second half part to the destination node of the first half path. This DNA computing solved the problem of HPP and gave a new direction in the field of cryptography.

Figure 1: Chromosomal Sequences from a Genetic Database

The most prominent work in DNA cryptography was carried out by Gehani and they proposed an image encryption algorithm with a help of one-time-pads. Substitution method was also bought up that did the pair wise mapping among the plaintext message and ciphertext algorithm. The encryption and decryption were done by the DNA-chip based approach for the 2D images with one-time pads and observed an enhanced version of steganography by reducing the difference among the disorder strands and the plain text. This form of DNA computing was carried out by many researchers in various fields in order to provide a security for the huge amount of data as shown in figure 1. Various cryptographic approaches were also bought to idea known as DNA-BASED cryptographic approaches. Certain attractive properties were also bought in work by DNA computing such as high storage capacity, maximum efficiency in energy, high parallelism that forms high advantage for the concept of data hiding. The application could be a solution for the signal processing, optimization problems, forecasting as well as clustering. One-gram DNA could contain 108 TB of data. This is equal to 1021 DNA-bases.

The objective of this paper is to provide the following:

- To analyze various approached regarding the DNA cryptography.
- To offer a robust ciphertext with the enhanced cryptographic algorithm.
- To offer an advanced data hiding methods with the enhanced approach.

A. DNA

DNA (Deoxyribonucleic Acid) that is formed by the nucleotides is in the form of biological super molecule, which is the source plasma of all living beings. Deoxyribonucleotides is said to be known as the monomer unit of DNA. Four types of nucleotides bases are found in DNA namely,

- Adenine (A),
- Cytosine (C),
- Thymine (T), and
- Guanine (G)

There are two nitrogen bases for the DNA: 1) Purines 2) Pyrimidines. A and G are said to be the double ring molecules and are termed as purines where the single ring molecules are called pyrimidines that has the base C and T. The capability of DNA could be found in the enormous industries that could serve the purpose of image and data encryption, data hiding, steganography, etc. Several parameters such as temperature, pressure, oxygen ratio etc., should be taken care while executing the DNA operations in various laboratories (identical conditions of several parameters should be maintained). Since the DNA is the sensitive module, this could provide several outputs in various environments. Hence, it could be said that the DNA computation totally depends on the environmental conditions as shown in figure 2. Binary form of converting the DNA sequence is known as the binary coding scheme. In this paper the DNA bases such as A, T, C, and G have been used for the binary scheme of 00, 11, 01, and 10. For instance, DNA sequence “AATCGGAT” has the binary sequence of 0000110110100011. RNA (Ribonucleic Acid) also

consist of the main bases as same as the DNA but here the Thymine (T) is replaced by Uracil (U).

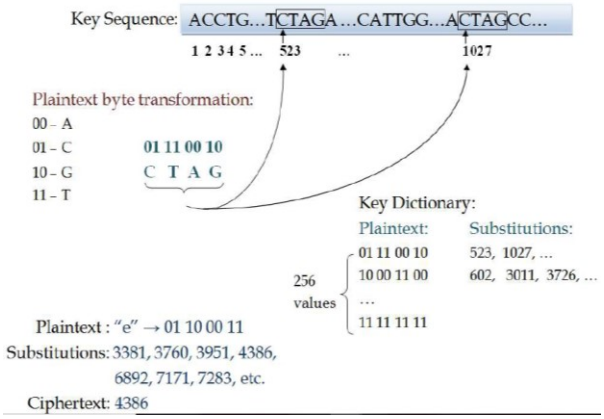

Figure 2: DNA Indexing Algorithm's encryption process

II RESEARCH MOTIVATION AND PROBLEM STATEMENT

The concept behind cryptography is to convert the plaintext into ciphertext (non-understandable format) using several algorithms that falls under cryptography concepts. Cryptography algorithms such as DES, TDES, AES, and RSA that is the multifarious symmetric and asymmetric key algorithms that helps in conversion of plaintext into ciphertext. The security could not be offered by the conventional cryptographic algorithms that lacks in their security over the growing internet usage for the huge amount of growing data, which could be easily broken by the intruders for their malicious activity. DNA could be used in two major roles under the concept of DNA cryptography. DNA acts as a medium for the encryption keys since it has two fundamental properties. The fake DNA and the real DNA sequence have virtually no major sequence. The other point is that there are 163 million DNA sequence available all over the world. Several cryptographic algorithms have been proposed over the last decade. These algorithm ciphertext is more complex than the normal key ciphertext.

III TECHNICAL BACKGROUND AND RELATED WORK

Various algorithms have been created in the concept of DNA computing. Tedious cases such as SAT, DES has been solved with the help of DNA computing in the past two decades. NP complete SAT problem has been solved by Lipton [1] using DNA computing. He also formulated an encoding scheme with the DNA scheme and the author employed a minimum number of variables to solve the SAT problem. In order to crack the Data Encryption Standard (DES), Boneh et al [2] proposed an approach with the DNA computing. Any crypto-system could be cracked with the less key size of about 64-bits with their proposed algorithm. Moreover, their encryption circuit is also small with the cracking mechanism in 916 steps. Each individual step is equal to 32 extractions and it needs at least a day processing for every 10-extraction steps. Altogether, their proposed algorithm could take at least 4 months to crack DES. With the DNA operations hard problem could be solved with the DNA computing techniques that were proposed by Chen et al. [3]. They also solved DES with the DNA technology. Three various functions have been included in the Chen et al.'s anticipated algorithm: 1) Initial function for initialization of the key space with every possibility 2) Encryption process, and 3) Detecting right corresponding key. Enzymes, tubes and short memory strands are all included in the molecular sticker algorithm.

Symmetric-key DNA cryptosystem had been introduced by MingXin et al. [4] that includes the DNA biotechnology imparted into the cryptography technologies. With the DNA probes the encryption and decryption has been formed in this algorithm and the ciphertext has been embedded into the DNA chip (microarray). The advancement of DNA chip leads to the security of the algorithm. For the efficient, reliable and secure transmission of data an algorithm has been proposed by Kar et al. [5]. This algorithm utilizes the DNA keys that utilized 3 keys: 2 keys for the encryption and the remaining one key for sharing the encryption key to the receiver. With the help of string to binary conversion process the plaintext will be converted to binary sequence in this algorithm. Binary coding scheme converts the huge DNA pattern in the form of binary data as the encryption key. The ciphertext could be obtained by applying the XOR and addition operations in the 64-bits plaintext and 64-bits key. Several cryptographic approaches had been proposed by Shiu et al. [6]. The cryptographic approaches with the utilization of DNA approach are as follows: 1) Insertion method 2) Complementary pair method, 3) Substitution method. Finally, they also proved that the substitution method is more efficient than the other two approaches. Data hiding approach has been proposed by Liu et al. [7] with the encryption of ciphertext in a word file with the DNA sequence. Their algorithm utilizes the DNA coding that converts the plaintext into DNA sequence. Chebyshev maps produce two pseudorandom DNA sequence XXOR and YPrimer, and create one time key using preferred DNA sequence key. The output generated includes the YPrimer and the DNA message sequence includes the XXOR. By applying the shift operations in the previously generated output, they could be converted to ciphertext. The word file could be converted to PDF after the encryption of ciphertext into the word file. The receiver will receive the PDF generated. A strong ciphertext algorithm had been proposed by Mandge et al. [8] that include key generation technology with the manipulation of matrix inclusion in the cryptographic algorithm. Initially the plaintext will be converted to a mini-cipher with the several shifting and XOR operations. It is said that if the process is applied to the plaintext each time, we could obtain various mini-

cipher for the similar plaintext since this process includes the secure key generation. After this mini-cipher gets converted to a final ciphertext with several biotechnologies that includes of DNA primers. The substitution method with the single bit complementary rule of Shiu et al. had been enhanced with the table lookup substitution method (TLSM) proposed by Taur et al. [9]. They extended to the two bits complementary rule for the substitution method. The two-bit message is converted to their corresponding letter with the help of TLSM that utilizes the lookup tables [10][11]. On the other hand, Shiu et al. rule could only translate one bit at a time. Moreover, compression of ciphertext is very effective in the TLSM procedure [12].

Complex problem that are difficult to solve in the silicon-based computer could be solved by the DNA computing by the researchers in a polynomial time [13]. The better results could also be generated by the DNA computing that could be used in the field of image processing, clustering and optimization problems. A brief description regarding the problem could be provided by this research work and also helps in analyzing the loopholes from the existing research works[14].DNA operations like Hybridization, Indexing, and PCR have the capacity for a solution of the application in the fields as data hiding, image encryption, and steganography. Several fascinating characteristics have also been present in the DNA that includes vast parallelism, maximum storage capacity as well as maximum efficiency in energy. This forms as an advantage for the data hiding and provides high security for the security purpose for the fast-growing internet data. Hence this could form as a greatest boon to the future for protecting the data [15]. Annealing, Hybridization, Indexing and PCR (Polymerase Chain Reaction) are said to be the DNA operations that has the ability to offer a strong data encryption technique. The powerful cryptography algorithm could be created when the DNA keys and DNA operations are combined together [16]. Cryptographic techniques thus provide greater security for the on-line as well as the off-line private security data. This offers several researchers to find an optimum solution in their research areas. The problem includes the silicon-based computer in polynomial time. Moreover, DNA computing offers greater solution in less amount of time when compared to the silicon computer [17][18].

IV PROPOSED DNA CRYTOGRAPHIC ALGORITHM

The enhanced proposed approaches have been explained in this section that accepts the message in the form of a plaintext and converts it as ciphertext in the form of protein. The receiver receives the protein through the public channel and the decryption process will be applied to receive the plaintext or the message [19]. The binary coding scheme will be shared by the sender and the key will be the codon table of RNA that utilizes the database name, which will be a shared symmetric key with the index number and send through the secure channel to the receiver end in this approach. From the DES algorithm, several substitution and permutation network will be developed through the enhanced algorithm. The functions of the encryption and decryption will be used in the receiver end[20]. Generally, plaintext will be converted into binary value in the encryption algorithm with the utilization of hexadecimal values. A long pattern of DNA sequence will be divided into 32-base pattern such as S1, S2 and S3 respectively. Similarly, k1, k2, k3 will be the 64-bits keys that has the converted 32-base DNA sequence with the binary coding scheme. The encryption block is selected from the 6-bits plaintext. The function of encryption runs from the parameters K1 (64-bits), encryption block, and 64-bits blank cipher block. The output of the encryption will be in the form of cipher block of 64-bits. In the same way, a decryption process will call the parameter k2 for their function with the previous cipher block as input, and 64-bits blank plaintext block. When this process has been executed, the next encryption function will be called with the previous encryption and decryption output with the key k3 as the input. This helps us in receiving 64-bits partial ciphertext block. The fake DNA sequence could be obtained by applying the reverse binary coding scheme inside the partial block. The final ciphertext could be obtained by applying the CDMB technique into the fake DNA sequence. By doing this process, the protein could be obtained which will be send to the receiver end.

The reverse CDMB technique will be applied in the decryption technique, which will be applied to the ciphertext block and it will be first decryption level thereby creating a fake DNA sequence. In order to receive the partial cipher level in the form of bits we tend to use the binary coding technique that tends to produce three 64-bits decryption keys K1, K2 and K3 with the shared information from the sender in the receiving end. Parameter k1 could evoke the decryption function with the selection of 64-bits cipher block as well as the 64-bits plaintext block. Executing this could help in obtaining the plaintext of 64-bits. The encryption function could be evoked with the key k2 with the input of previous output block and 64-bits blank cipher block. Then again, the decryption function should be called after receiving the output of 64-bits block from the previous step that acts as an input in this function with the key k3. This helps in receiving the 64-bits plaintext block. Plaintext P could be obtained by applying the inverse hexadecimal coding.

Algorithm:

I. Procedure for Encrypt (Key K, Plaintext P, Ciphertext C):

Input: Key K, Plaintext P.

Output: Ciphertext C.

Phase 1: Receive K' as 56-bit by applying permutation on K.

Phase 2: Dividing the factor K’ as 28-bits X0 and Y0

Phase 3: Produce Xn and Yn, as 16 blocks such that 1≤n≤16. The set of block Xn and Yn will be formed with the help of the previous block left shift key.
Phase 4: Every concatenated pairs will generate 16 sub-keys, 1≤n≤16.
Phase 5: P that is of 64-bits will be applied with the initial permutation (IP).
Phase 6: IP is divided as P0 and Q0, each consisting of 32-bits.
Phase 7: The pairs of Pn and Qn are obtained for 1≤n≤16, with the help of the two equations:
Pn= Qn-1
Qn = Pn-1 ⊕X (Qn-1, Kn)
Where, X (Qn-1, Kn) = X (Qn-1) ⊕Kn; and X (Qn-1) = Expansion function.
Phase 8: The Q16 P16 blocks of 64 bits is obtained.
Phase 9: The output C is generated by applying final permutation IP-1 to Q16 P16

II Procedure for Decrypt:
Phase 1: Receive K' as 56-bit by applying permutation on K.
Phase 2: Dividing the factor K' as 28-bits X0 and Y0 each.
Phase 3: Produce Xn and Yn, as 16 blocks such that 1≤n≤16. The set of block Xn and Yn will be formed with the help of the previous block left shift key Xn-1 and Yn-1.
Phase 4: Every concatenated pairs Xn and Yn will generate 16 sub-keys, 1≤n≤16.
Phase 5: P that is of 64-bits will be applied with the initial permutation (IP).
Phase 6: IP is divided as P16 and Q16, each consisting of 32-bits.
Phase 7: The pairs of Pn and Qn are obtained for 16≥ n ≥ 1, with the help of the two equations:
Qn-1 = Pn
Pn-1 = Qn ⊕X (Pn, Kn)
Where, X (Pn, Kn) = X (Pn) ⊕ Kn; and Expansion function is X (Pn).
Phase 8: The P0 Q0 blocks of 64 bits is obtained.
Phase 9: The output P is generated by applying final permutation IP-1 to P0 Q0

III Proposed Encryption Algorithm:
Algorithm: Encryption (Plaintext P, Ciphertext C)
Input: Plaintext P, A long DNA sequence > 96 DNA-bases
Output: Ciphertext C
Phase 1: The plaintext or the message is converted in binary form with the help of Hexadecimal value.
Phase 2: S1, S2, and S3 are the 3 DNA sequences chosen from a given 32 bases DNA sequences.
Phase 3: S1, S2, and S3 are converted as K1, K2, and K3 in the form of bits with the binary coding scheme. K1 and K3 are chosen as the encryption key. On the other hand, K2 is made as the decryption. Each key are 64 bits.
Phase 4: Encrypt (K1, P, C')
Phase 5: Decrypt (K2, C', C'')
Phase 6: Encrypt (K3, C'', C''')
Phase 7: C''' is said to be the partial cipher block that is applied with the reverse binary coding scheme and the fake DNA sequence is obtained.
Phase 8: Apply CDMB to obtain the protein sequence from the fake DNA sequence.
Phase 9: PROTEIN sequence is sent to the receiver in the form of encrypted message.

III Proposed Decryption algorithm:
Algorithm: Decryption (Plaintext P, Ciphertext C, Index X and Database D)
Input: Ciphertext C, Key information shared (Index X and Database D)
Output: Plaintext P
Phase 1: Implement reverse CDMB to obtain the fake DNA sequence from the Ciphertext C.
Phase 2: Get the partial ciphertext with the binary coding scheme in the form of bits.
Phase 3: S1, S2, and S3 are the 3 DNA sequences chosen from the shared information by the sender from the DNA databases.
Phase 4: S1, S2, and S3 are converted as K1, K2, and K3 in the form of bits with the binary coding scheme. K1, and K3 are chosen as the decryption key. On the other hand, K2 is made as the encryption key. Each key are 64 bits.
Phase 5: Decrypt (K1, C, P')
Phase 6: Encrypt (K2, P', P'')
Phase 7: Decrypt (K3, C'', M')
Phase 8: Apply the reverse hexadecimal coding to receive the original plaintext P from M.

Conversion of DNA sequence is possible with the proposed scheme that utilizes the DNA encoding table that is shown below. Here the plaintext is been used and each character indicates a DNA pattern with the combination of 3 bases. Several complementary rules has been followed by the proposed algorithm that boost a ciphertext due to perfect complementary rule that follows property like V(W) V(V(W)) V(V(V(W))) V(V(V(V(W)))), where 'W' is any string or alphabet. The complementary rules that have been used in this approach are as follows: (AT) (CA) (GC) (TG), Means V (A) = T.

Table 1: DNA Encoding [6]

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. Exploitation of substitution method for the algorithm is obtained from Shiu et al.

V EVALUATION

Complexity analysis for the proposed algorithm was carried out since it exposes the security level of the algorithm for real-time applications.In this work the computational time of the algorithm was analyzed employing complexity theory methods. The input consists of alphabets; special character and alphanumeric.The obtained conclusions were tested to be true through the implementation results.

Table 2: Measurement of Key-Table Computation Run-Time

Nucleotides (key length)	Runtime (ms)
3000	66
4000	300
6000	436
8000	566
9000	636
10000	866

Table 3: Measurement of Encryption Runtime

Size of Plain-Text	Run-time (ns)
86 KB	6
100 KB	13
300KB	36
1000KB	66
13000KB	76

Table 4: Measurement of Decryption Runtime

Cipher-text Size (KB)	Run-time (ns)
86 KB	16
100 KB	33
300KB	46
1000KB	53
13000KB	66

The sum of all the operations represents the execution time of the proposed algorithm. The number of operations contains variable as well as constant based on input parameters. Based on the approximations from complexity theory, the smallest possible class of functions is employed to represent the growing rate of the algorithm's runtime, for instance, the number of operations for 1 + 2n would be O(n) in-terms of complexity; if the number of operations is 4 + n + n3, then the complexity would be O(n3). The experimental results have proved the correctness of the estimated complexity. In order to see the progression of the runtime, the program was executed at different, progressively increasing values of n and m. Figure 3 depicts graphics of the runtime growing rate for the key table computation, encryption and decryption processes. The execution time measurements are presented in Tables 2 - 4.

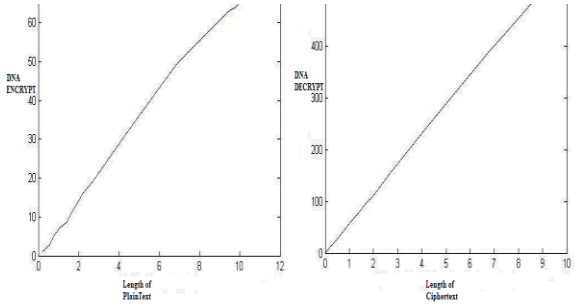

Figure 3: Results of DNA Encryption and DNA Decryption

VI CONCLUSION AND FUTURE WORK

Cryptography based security services could be seen in many sectors that includes several areas that includes digital certificate, banking sectors, image and message encryption etc. Certain algorithms such as DES, TDES, AES, and RSA are multifarious symmetric and asymmetric key cryptography algorithms that exists metamorphosis of plaintext into ciphertext. As the next level of evaluating the efficiency of the proposed algorithms, the following methods will be conferred as future work:

- Statistical measurements
- Key space analysis

REFERENCES

[1] R.J. Lipton, "DNA solution of hard computational problems," Science, New Series, 268(5210), 542-545, 1995.
[2] D. Boneh, C. Dunworth, and R.J. Lipton, "Breaking DES Using a Molecular Computer," DIMACS workshop on DNA computing, 27, 37-51, 1995.
[3] 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.
[4] 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.
[5] 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.
[6] 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.
[7] H. Liu, D. Lin, A. Kadir, "A novel data hiding method based on deoxyribonucleic acid coding," Computers and Electrical Engineering, 39, 1164-1173, 2013.
[8] 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.
[9] 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.
[10] 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.
[11] 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.
[12] O. Tornea, M. Antonini, M. Borda, "Multimedia Data Compression and Encryption using DNA Cipher", Communications Department, Technical University of ClujNapoca, winner of 2nd Prize, album SSET 2013.
[13]. W.M. Shih, J.D. Quispe, G.F. Joyce, "1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron", Nature, Vol. 427, pp. 618-621, 2004.
[14] . H. Shiu, K. Ng, J.F. Fang, et al., "Data hiding methods based upon DNA sequences", Elsevier Inc., pp. 2196–2208,2010.
[15] M. Reza Najaf Torkaman et al, "Innovative Approach to Improve Hybrid Cryptography by using DNA Steganography", IJNCAA 2012
[16] Zhihua Chen et el. "Efficient DNA Sticker Algorithms for DES", IEEE 3rd international conference on Bio-inspired computing, 2012
[17] A.P. Thiruthuvadoss, "Comparison and Performance Evaluation of Modern Cryptography and DNA Cryptography," M.S. Thesis, Royal Institute of Technology, 2012.
[18] 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
[19] M. Babaei, "A novel text and image encryption method based on chaos theory and DNA computing," Natural computing, 12(1), 101-107, 2013
[20] M.F Vaida, R. Terec, O. Tornea, L. Chiorean, A. Vanea, "DNA Alternative Security", Advances in Intelligent Systems and Technologies Proceedings ECIT2010 – 6th European Conference on Intelligent Systems and Technologies, pp. 1-4, October 2010