Available online at www.scholarsresearchlibrary.com

Scholars Research Library

Der Pharmacia Lettre, 2014, 6 (6):175-183
(http://scholarsresearchlibrary.com/archive.html)

Scholars Research Library
ISSN 0975-5071
USA CODEN: DPLEB4

Medicine names as a DNA sequence using graph domination

Yamuna M. and Karthika K.

SAS, VIT University, Vellore, Tamilnadu, India

ABSTRACT

Had it not been for chemistry in science we may have still may deprived of many products around us used in everyday life. Research in the field of chemistry has made us aware of electrons, radio activity, atomic theory and so on. Most of the significant invention in chemistry, which are resultants of a chemical reactions are finally some kind of chemical substance. Securing these data is of high importance. Several techniques are used for this purpose DNA steganography is promising due to its high storage capacity. In this paper we propose encryption of chemical compounds using DNA steganography with graph domination as a tool for encryption.

Keywords: Drugs, DNA sequence, Domination, Domination Subdivision Stable.

INTRODUCTION

World today is fast developing in almost all domains. In this fast developing society there is an effort in all fields to improve and make human life more comfortable. In this approach new findings are common and communications regarding these findings is a must. In numerous findings like new drugs, soaps, washing powders, chemicals, etc, chemical compounds is a part. Many new finding has a compound structure, which need to be communicated. Many methods are devised for this purpose. DNA steganography is one such method which is widely due to the high storage capacity of a DNA. It can be used to carry data of any length.

A new parallel cryptography technique using DNA molecular structure, one – time – pad scheme and DNA hybridization technique which certainly minimizes the time complexity was proposed in [1]. In [2], M. Yamuna et al have provided a genetic code which was used for encrypting any DNA sequence. A method for encrypting any chemical formula was provided in [3], by using graph domination as the tool for encryption. In this paper we propose two methods of encrypting details regarding a chemical compound using graph domination as a tool.

MATERIALS AND METHODS

2.1. DNA and RNA

DNA, the major support of genetic information of any organism in the biosphere, is composed of two long strands of nucleotides, each containing one of four bases (A – adenine, C – cytosine, G – guanine, T – thymine), a deoxyribose sugar and a phosphate group. A DNA molecule has double-stranded structure obtained by two single-stranded DNA chains, bonded together by hydrogen bonds: A = T double bond and C ≡ G triple bond. The DNA strands that bond each other through A – T and C – G bonds are known as complementary strands. [4]

RNA

Ribonucleic acid (RNA) is a polymeric molecule made up of one or more nucleotides. A strand of RNA can be thought of as a chain with a nucleotide at each chain link. Each nucleotide is made up of a base (adenine, cytosine, guanine, and uracil, typically abbreviated as A, C, G and U), a ribose sugar and a phosphate [5].

The regular RNA codon table is given Table – 1, which we will use later in encryption .

Table – 1

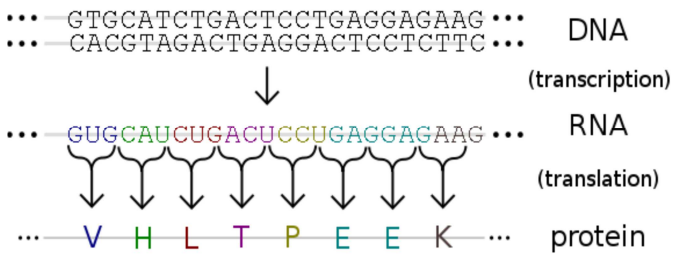
Second base

First base	U	C	A	G	Third base
U	UUU, UUC – Phenylalanine F	UCU, UCC, UCA, UCG – Serine S	UAU, UAC – Tyrosine Y	UGU, UGC – Cysteine C	U C A G
	UUA, UUC – Leucine L		UAA Stop codon, UAG Stop codon	UGA Stop codon, UGC Tryptophan W	
C	CUU, CUC, CUA, CUG – Leucine L	CCU, CCC, CCA, CCG – Proline P	CAU, CAC – Histidine H	CGU, CGC, CGA, CGG – Arginine R	U C A G
			CAA, CAG – Glutamine Q		
A	AUU, AUC, AUA – Isoleucine I	ACU, ACC, ACA, ACG – Threonine T	AAU, AAC – Asparagine N	AGU, AGC – Serine S	U C A G
	AUG Methionine start codon M		AAA, AAG – Lysine K	AGA, AGG – Arginine R	
G	GUU, GUC, GUA, GUG – Valine V	GCU, GCC, GCA, GCG – Alanine A	GAU, GAC – Aspartic acid D	GGU, GGC, GGA, GGG – Glycine G	U C A G
			GAA, GAG – Glutamic acid E		

DNA to RNA

Transcription is the first step of gene expression, in which a particular segment of DNA is copied into RNA by the enzyme RNA polymerase. Both RNA and DNA are nucleic acids, which use base pairs of nucleotides as a complementary language that can be converted back and forth from DNA to RNA by the action of the correct enzymes. During transcription, a DNA sequence is read by an RNA polymerase, which produces a complementary, anti-parallel RNA strand. As opposed to DNA replication, transcription results in an RNA complement that includes uracil (U) in all instances where thymine (T) would have occurred in a DNA complement [6].

DNA to RNA to Protein



2.2 Domination in Graph Theory

In this section the basic results of domination theory required for encryption of binary string into a DNA sequence, RNA sequence and false binary string are provided.

Dominating Set

A set of vertices D in a graph G = (V, E) is a dominating set if every vertex of V – D is adjacent to some vertex of D. If D has the smallest possible cardinality of any dominating set of G, then D is called a minimum dominating set. The cardinality of any minimum dominating set for G is called the domination number of G and it is denoted by γ (G).

A vertex in V – D is k – dominated if it is dominated by at least k – vertices in D. The private neighborhood of v ∈ D is defined by pn [v, D] = N (v) – N (D – { v }). A path is a sequence of consecutive edges in a graph and it is denoted by P_n. A path which originates and ends in the same vertex is called a cycle and it is denoted by C_n with n vertices. If there is an edge between two vertices, then we say that u is adjacent to v. For general details on domination theory we refer to [7].

Subdivision Graph

A subdivision of a graph G is a graph resulting from the subdivision of edges in G. The subdivision of some edge e with endpoints { u, v } yields a graph containing one new vertex w, and with an edge set replacing e by two new edges, { u, w } and { w, v }. We shall denote the graph obtained by subdividing any edge uv of a graph G, by G_{sd} uv. Let w be a vertex introduced by subdividing uv. We shall denote this by G_{sd} uv = w.

Domination Subdivision Stable Graphs

A graph G is said to be domination subdivision stable (DSS), if the γ - value of G does not change by subdividing any edge of G, that is γ (G) = γ (G_{sd} uv), for all u, v ∈ V (G), u adjacent to v [8].

The following result is proved in [8],

R: A graph G is DSS if and only if for every u, v ∈ V (G), either

i. there is a γ - set containing u and v or
ii. there is a γ - set D such that either
iii. pn[u, D] = { v } or
iv. v is 2 – dominated.

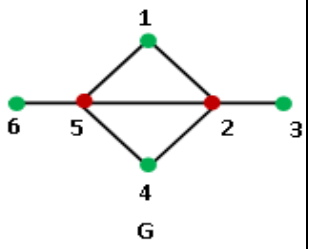
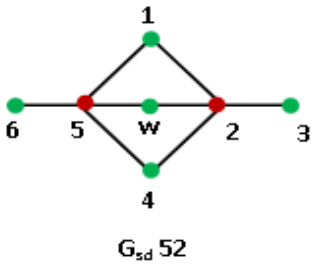
We use X_1, X_2, X_3 to represent the three properties in R.

In this paper in all the graphs

- Represent vertex that belongs to a γ – set D.
- Represent vertex that does not belong to the γ – set D.

In Table – 2 {{ 5, 2 }, { 2, 6 }, { 5, 3 }} are possible γ – sets for the graph G. We have picked the γ – set D = { 5, 2 }, for discussion. The three properties X_1, X_2, X_3 with respect to this set D is discussed, with the subdivided graph provided in the last column.

Table – 2

Graph and γ - set	X_1	X_2	X_3	Subdivided graph
 G D = { 5, 2 } is one of the γ - set.	5, 2 ∈ D	Pn[5, D] = {6}, Pn[2, D] = {3}	Vertices 1, 4 are 2 – dominated by 5 and 2	 G_{sd} 52 γ (G) = γ (G_{sd} 52) = 2. This is true for all u, v ∈ V (G), u adjacent to v.

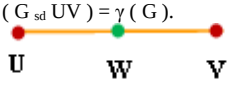
From the table it can be noticed that, with respect to D, for every edge in the graph one of the properties X_1, X_2, X_3 is satisfied. We also notice that D is a γ – set for G_{sd} 52 also. This is true for all u, v ∈ V (G), u adjacent to v. So, the graph G is DSS.

RESULTS AND DISCUSSION

3.1. Edge Values for Binary Encryption

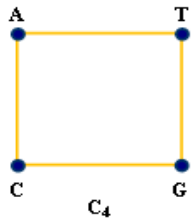
Let G be a graph and D be any γ - set for G. We basically use the idea that for any vertex in the γ – set we assign value 1, and for any vertex not in the γ – set we assign a value 0. The properties of DSS graphs and edge values assigned are discussed in the Table – 3. For discussion purpose we have taken an edge e = (u v) in a graph G. Here after we shall use upper case letters to represent vertices.

Table – 3

S. No.	γ – set	γ – set after subdivision of e = (U V)	Edge values after subdivision	Final edge value assigned to e = (U V)
1.	**X_1:** U, V ∈ D. 	γ (G_{sd} UV) = γ (G). 	U, V ∈ D, W ∉ D. Edge value of UW = 10 Edge value of WV = 01	1001
2.	**X_2:** U ∈ D, pn [U, D] = { V }. 	γ (G_{sd} UV) = D – { U } ∪ { W }. 	U, V ∉ D, W ∈ D. Edge value of UW = 01 Edge value of WV = 10	0110
3.	**X_3:** U ∈ D, V is 2 – dominated. 	γ (G_{sd} UV) = γ (G). 	U ∈ D, W, V ∉ D. Edge value of UW = 10 Edge value of WV = 00	1000

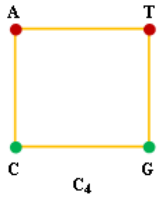
3.2. Proposed Encryption Scheme

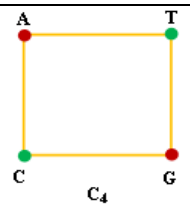
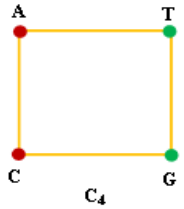
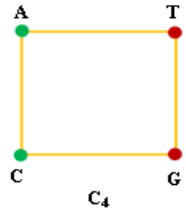
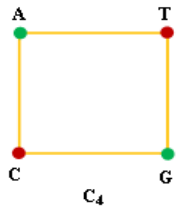
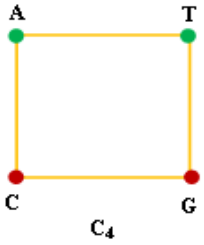
A DNA sequence has only the four combinations A, T, G, C. So, we choose graphs with four vertices which are DSS. The only possible DSS graphs with four vertices are P_4 and C_4 [8]. We can choose any graph for encryption. We shall use C_4. We label the vertices of C_4 as A, T, G, C.



The possible γ – sets are { A, T }, { A, G }, { A, C }, { T, G }, { T, C } and { G, C }. The edge values assigned to these γ – sets is given in Table – 4 using Table – 3.

Table – 4

γ - sets	DSS properties	Edge values
S_1: A, T ∈ D 	i. pn [A, D] = { C }, pn [T, D] = { G }, implies X_2 is satisfied. ii. A, T ∈ D, implies X_1 is satisfied.	AT, TA: 1001 AC, CA, TG, GT: 0110 AG, TC: 10 GA, CT: 01 AA, TT: 11 CC, CG , GC, GG: 00
S_2: A, G ∈ D	T and C are 2 – dominated, implies X_3 is satisfied	AT, GT, GC: 1000 TA, TG, CG: 0001

		AC: 10 CA:01 AA, GG, AG, GA: 11 TT, CC, TC, CT: 00
S_3: A, C ∈ D 	i. pn [A, D] = { T }, pn [C, D] = { G }, implies X_2 is satisfied. ii. A, C ∈ D, implies X_1 is satisfied	AC, CA: 1001 AT, CG, TA, GC: 0110 AG, CT: 10 GC, TA: 01 AA, CC: 11 TT, TG, GT, GG: 00
S_4 : T, G ∈ D 	i. pn [T, D] = { A }, pn [G, D] = { C }, implies X_2 is satisfied. ii. T, G ∈ D, implies X_1 is satisfied.	TG, GT: 1001 TA, GC, AT, CG: 0110 TC, GA: 10 CT, AG: 01 TT, GG: 11 AA, AC, CA, CC: 00
S_5 : T, C ∈ D 	A and G are 2 – dominated, implies X_3 is satisfied.	AC, GT, GC: 0001 CA, TG, CG: 1000 TA: 10 AT: 01 TT, CC, TC, CT: 11 AA, GG, AG, GA: 00
S_6 : G, C ∈ D 	i. pn [C, D] = { A }, pn [G, D] = { T }, implies X_2 is satisfied. ii. C, G ∈ D, implies X_1 is satisfied.	CG, GC: 1001 CA, GT, AC, TG: 0110 GA, CT: 10 TC, AG: 01 CC, GG: 11 AA, AT, TA, TT: 00

3.3. Method 1 [Encryption as RNA]

In this method the chemical formula of the medicine to be encrypted is converted into a RNA sequence using the results in Section – 3.1 and 3.2. The encryption algorithm for the same is provided here.

Encryption Algorithm

Step 1. Consider the medicine to be encoded. Let S be its chemical formula for this medicine.

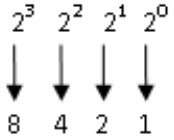
For illustrate let **ASACOL** be the medicine to be encrypted. Its chemical formula is

S: $C_7H_7NO_3$

Step 2. Using periodic table replace the chemical elements by its corresponding atomic number to obtain S'.

S': 6 7 1 7 7 1 8 3

Step 3. Determine the largest decimal value in S' (say p). Replace each decimal value into binary representation.



The largest decimal value in S' is 8, so the length of binary string is 4.

Step 4. Convert decimal value of S' into a binary number of length p, to obtain a binary sequence S".
S'': 0110 0111 0001 0111 0111 0001 1000 0011

Step 5. If S" is of odd length suffix 0 to S".

Step 6. Divide S" into segments, where each segment contains k – bits, k = 2 or 4.

Splitting this into two bit strings we get

S'': 01 10 01 11 00 01 01 11 01 11 00 01 10 00 00 11 (or)

Suppose we prefer to use strings of length four also (if length four is not represent in S_3, then consider length two.

So, the string will be the combination of length 4 and 2), then the given strings can also be represented as

S'': 0110 01 11 00 01 01 11 01 11 00 01 10 00 00 11.

Step 7. Choose any graph from S_1 to S_6 from Table – 4. (let us choose S_1)

Step 8. Generate a DNA sequence S''' using Table – 4.

S"can be replaced by **S''': GA AG GA TT CC CT CT AA GA AA GG GA AG GG GG TT**, if we decide to use encoding as strings of size two only. (or)

S"can be replaced by **S''': AC GA TT CC CT CT AA GA AA GG GA AG GG GG TT** (the string will be the combination of length 4 and 2).

Step 9. Send the sequence S^{iv} = < X > < Y > < S''' > to the receiver, where X represents the γ - set and hence one of the graphs S_1, S_2, S_3, S_4 , S_5, S_6 used for encryption

$$Y=\begin{cases}AT, \text{ if length of } S' \text{ is odd}\\ GC, \text{ if length of } S' \text{ is even}\end{cases}$$

S''' is the false DNA sequence obtained.

S^{iv}: ATGCGAAGGATTCCCTCTAAGAAAGGGAAGGGGGTT or as

S^{iv}: ATGCACGATTCCCTCTAAGAAAGGGAAGGGGGTT

where, the first two entry AT represents S_1 and third fourth entry GC represents that string is of even length.

These are two possible sequence among numerous different combinations available for encrypting the message.

Step 10. Convert each bit in S^{iv} into the corresponding RNA codon to generate a fake RNA sequence S^v.

Consider **S^{iv}: ATGCACGATTCCCTCTAAGAAAGGGAAGGGGGTT.** By using Table – 1, we can convert each bit into the corresponding RNA segment. The given string is encrypted as

S^{v}: GCU ACU GGU UGC GCC UGC GGU GCA ACU ACU UGC UGC UGU ACU UGC ACU GCU GCU GGU GCG GCG GCG GGU GGU GGU GCC GCC GGA GGA GGA GGA GGA ACU ACU.

Step 11. Send S^{v} to the receiver.

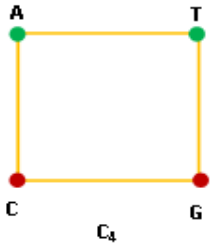
The procedure is reversed for decryption

Suppose the message received is,

UGCGGAGGCUGUGCUGCCUGUUGCACUACAUGUGGUGCGGCAACAACGGGA
UGCUGUUGCGCUGCGGGGGGUUGUGGCUGUGCCACAACGGGUGCUACUACUGGUGCU.

Split this RNA codon into 3 – bits for decoding, UGC GGA GGC UGU GCU GCC UGU UGC ACU ACA UGU GGU GCG GCA ACA ACG GGA UGC UGU UGC GCU GCG GGG GGU UGU GGC UGU GCC ACA ACG GGU GCU ACU ACU GGU GCU.

By using Table – 1, we can encode the RNA segments into DNA segments. So, S": CGGCAACCTTCGAATTGCCCAAGGCGCATTGATTGA. The first two entries CG indicates that the graph under consideration for encryption is



The third fourth entry GC indicates that the string is of even size. Split the message for decoding we get AA CC TT CG AA TT GC CC AA GG CG CA TT GA TT GA.

Using the string values from S_6, we get 00110010010000100111001110010110001000100

Splitting the binary string into segments of length five, 00110 01001 00001 00111 00111 00101 10001 00010

Replace the binary string into decimal values S': 6 9 1 7 7 5 17 2

By using periodic table replace the decimal values into its corresponding chemical element. Apply this procedure for alternate numbers, we get the chemical formula **S: $C_9H_7N_5Cl_2$**

S is the chemical formula for the medicine **LAMOTRIGINE**.

Method 2 [Encryption as a Binary String]
We transform the alphabets A, T, G and C into binary code as in [9].

Table – 5

Alphabet	Binary Representation
A	00
Tor (U)	11
G	10
C	01

Using Table – 1 and 5, we generate the following binary conversion table, by replacing A, U, G, C by their respective binary values.

From the Table – 6, we observe that each codon has a unique binary string conversion of size 6. Each codon obtained by method – 1 can be replaced by the respective binary string from the table. So the given binary sequence will again be encrypted as a new binary sequence.

Encryption Algorithm

Step 1. Obtain S^v from Method – 1.

S^v: GCUACUGGUUGCGCCUGCGGUGCAACUACUUGCUGCUGUACUUGCACUGC
UGCUGGUGCGGCGGCGGGUGGUGGUGCCGCCGGAGGAGGAGGAGGAACUACU.

Step 2. Split S^v into segments of length 3 and convert each codon in S^v into the corresponding binary string from the RNA codon binary string Table – 6, to obtain a sequence S^{vi}.

Table – 6

Binary Conversion Table

	U	C	A	G	
U	UUU – 111111 UUC – 111101 UUA – 111100 UUG – 111110	UCU – 110111 UCC – 110101 UCA – 110100 UCG – 110110	UAU – 110011 UAC – 110001 UAA – 110000 UAG – 110010	UGU – 111011 UGC – 111001 UGA – 111000 UGG – 111010	**U** **C** **A** **G**
C	CUU – 011111 CUC – 011101 CUA – 011100 CUG – 011110	CCU – 010111 CCC – 010101 CCA – 010100 CCG – 010110	CAU – 010011 CAC – 010001 CAA – 010000 CAG – 010010	CGU – 011011 CGC – 011001 CGA – 011000 CGG – 011010	**U** **C** **A** **G**
A	AUU – 001111 AUC – 001101 AUA – 001100 AUG – 001110	ACU – 000111 ACC – 000101 ACA – 000100 ACG – 000110	AAU – 000011 AAC – 000001 AAA – 000000 AAG – 000010	AGU – 001011 AGC – 001001 AGA – 001000 AGG – 001010	**U** **C** **A** **G**
G	GUU – 101111 GUC – 101101 GUA – 101100 GUG – 101110	GCU – 100111 GCC – 100101 GCA – 100100 GCG – 100110	GAU – 100011 GAC – 100001 GAA - 100000 GAG – 100010	GGU – 101011 GGC – 101001 GGA – 101000 GGG – 101010	**U** **C** **A** **G**

Step 2. Split S^v into segments of length 3 and convert each codon in S^v into the corresponding binary string from the RNA codon binary string Table – 6, to obtain a sequence S^{vi}.

S^v: GCU ACU GGU UGC GCC UGC GGU GCA ACU ACU UGC UGC UGU ACU UGC ACU GCU GCU GGU GCG GCG GCG GGU GGU GGU GCC GCC GGA GGA GGA GGA GGA ACU ACU.

S^{vi}: 100111000111101011111001100101111001101011100100000111000111111100111100111101100011111110010001111001111001111010111001101001101001101010111010111010111001011001101000101000101000101000101000000111.

Step 3. Send S^{vi} to the receiver.

The procedure is reversed for decryption

Suppose the message received is,

S^{vi}: 111001101000101001111011100111100101111011111001000111000100111011101110101111

00110100100000100000110101000111001111011111001100111100110101010101011111011101001111011100101000100000011010101110011100011100011110101110011.

Split this S^{vi} into 6 – bits for decoding, we get
S^{vi}: 111001 101000 101001 111011 100111 100101 111011 111001 000111 000100 111011 101011 100110 100100 000100 000110 101000 111001 111011 111001 100111 100110 101010 101011 111011 101001 111011 100101 000100 000110 101011 100111 000111 000111 101011 100111.

By using Table – 6, we can convert each bit into the corresponding RNA codon. The given binary string decrypted as

S^{v}: UGC GGA GGC UGU GCU GCC UGU UGC ACU ACA UGU GGU GCG GCA ACA ACG GGA UGC UGU UGC GCU GCG GGG GGU UGU GGC UGU GCC ACA ACG GGU GCU ACU ACU GGU GCU.

Again by applying method – 1, we can encode the above DNA Segments. From this we can obtain the original chemical formula for the medicine **LAMOTRIGINE**.

CONCLUSION

Chemical formula of drugs can be of any length. A DNA sequence and binary string have an advantage of being of any length. A formula when converted into these sequences can be made available in public domain. Numerous strings are available in public domain, and it is not possible to find the fake strings. Graph properties are used as key for encryption. Numerous graphs are available. So a correct guess of the graph used is not practical. Even a correct guess of graph will not lead to breaking the code for unless the γ – set and the property is known breaking the code becomes impossible. This encryption has different stages of encryption before being coded into a string. So the proposed method is safe for carrying details of the chemical formula of any drug.

REFERENCES

[1] Sabari Pramanik, Sanjit Kumar Setua, DNA Cryptography, 7th International Conference on Electrical and Computer Engineering, 20 – 22 December, **2012**, Dhaka, Bangladesh.
[2] M. Yamuna, B. Joseph Sasikanth Reddy, Nithin Kumar Reddy, Paladugula Raghuram, *International Journal of ChemTech Research*, Jan – March **2014**, Vol.6, No.1, pp 53 – 63.
[3] M. Yamuna, K. Karthika, *International Journal of ChemTech Research*, Oct – Dec **2013**, Vol.5, No. 6, pp. 2747 – 2756.
[4] Yunpeng Zhang and Liu He Bochen Fu, Research on DNA Cryptography, Applied Cryptography and Network Security, Dr. Jaydip Sen (Ed.), **2012**.
[5] http://exploringorigins.org/rna.html
[6] http://en.wikipedia.org/wiki/Transcription_%28genetics%29
[7] T. W. Haynes, S. T. Hedetniemi, P. J. Slater, Fundamental of Domination in Graphs, Marcel Dekker, New York;**1998**.
[8] M. Yamuna, K. Karthika, K, *International Journal of Mathematical* Archive, **2012**, Vol. 3, No. 4, pp. 1467 – 1471.
[9] M. I. Youssef, A. Emam and M. Abd Elghany, *International Journal of Journal of Computer Applications*, **2012**, Vol. 45, No. 10, pp: 19 – 24.