International Journal of ChemTech Research
CODEN(USA): IJCRGG ISSN : 0974-4290
Vol.5, No.6, pp 2747-2756, Oct-Dec 2013

Chemical Formula: Encryption Using Graph Domination And Molecular Biology

M. Yamuna* and K. Karthika

SAS, VIT University, Vellore, Tamilnadu, India, 632 014.

***Corres.author: myamuna@vit.ac.in and karthika.k@vit.ac.in**

Abstract: In this modern developed society new inventions have become very common. Most of the new products are based on some chemical formula. This chemical formula is the key for each product and every concern maintains this secured. But due to modernization of communication, communicating these details directly or indirectly becomes unavoidable. But the threats any user faces is growing exponentially. So secured communication of these chemical formulas is an important problem. In this article we provide a method encrypting any chemical formula using graph domination as the tool for encryption. Here every chemical formula is converted into a binary string using graph domination and later encrypted using DNA steganography.
Keywords: Domination, domination subdivision stable, DNA sequence, binary encryption.

1. Introduction

The security of a system is essential nowadays. With the growth of the information technology power, and with the emergence of new technologies, the number of threats a user is supposed to deal with grew exponentially. Cryptography is where security engineering meets mathematics. It provides us with the tools that underlie most modern security protocols. It is probably the key enabling technology for protecting distributed systems, yet it is surprisingly hard to do right. Cryptography consists in processing plain information applying a cipher and producing encoded output, meaningless to a third-party who does not know the key. In cryptography both encryption and decryption phase are determined by one or more keys.

The periodic table of the chemical elements is a table that displays all known chemical elements in a systematic way. The elements in the periodic table are ordered by their atomic number (Z) and are arranged in periods (horizontal rows) and groups (vertical columns). The layout of the periodic table is designed to illustrate periodic trends, similarities and differences in the properties of the elements. [4]

A chemical formula is a way of expressing information about the proportions of atoms that constitute a particular chemical compound, using a single line of chemical element symbols, numbers, and sometimes also other symbols, such as parentheses, dashes, brackets, and plus and minus signs. These are limited to a single typographic line of symbols, which may include subscripts and superscripts. A chemical formula is not a chemical name, and it contains no words. Although a chemical formula may imply certain simple chemical structures, it is not the same as a full chemical structural formula. Chemical formulas are more limiting than chemical names and structural formulas.[3]

The relation between cryptography, chemistry and molecular biology was originally irrelevant, but with the in-depth study of modern biotechnology and DNA computing, these two disciplines begin to work together more closely. DNA cryptography and information science was born after research in the field of DNA computing field by Adleman. In terms of hiding information, there are such results as “Hiding messages in DNA microdots,” “Cryptography with DNA binary strands” and so on. In terms of DNA algorithms, there are such results as “A DNA-based, bimolecular cryptography design,” “Public-key system using DNA as a one-way function for key distribution,” “DNASC cryptography system” and so on. However, DNA cryptography is an emerging area of cryptography and many studies are still at an early stage. [6]

In this paper we propose an encryption method of a binary string, using graph theory properties, as a false DNA sequence.

2. Research On Dna Cryptography

2.1. DNA

DNA, the major support of genetic information (genetic blueprint) 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.

There is atmost no difference between a real DNA sequence and a faked one. A large number of DNA sequence a publically available in various website. A DNA sequence is usually long. For instance, the DNA sequence of “Litmus”, its real length is with 2856 nucleotides long [7]

ATCGAATTCGCGCTGAGTCACAATTCGCGCTGAGTCACAATTCGCGCTGAGTCACAATTGTGACTC
AGCCGCGAATTCCTGCAGCCCCGAATTCCGCATTGCAGAGATAATTGTATTTAAGTGCCTAGATAC
AATAAACGCCATTTGACCATTCACCACATTGGTGTGCACCTCCAAGCTCGCGCACCGTACCGTCTC
GAGGAATTCCTGCAGGATATCTGGATCCACGAAGCTTCCCATGGTGACGTCAC.

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 is 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 — abbreviated MDS. The cardinality of any MDS 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}). C_n and P_n denotes the cycle with n vertices. For general details on domination theory we refer to [1].

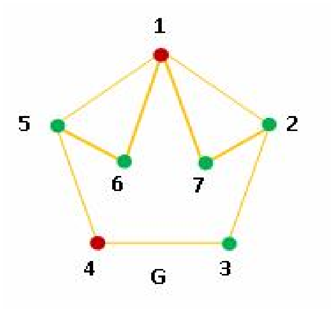


Fig. 2

In Fig. 2, G is a graph and D = { 1, 4} is a MDS for G. Here pn[1, D] = { 6, 7, 2 }, pn[4, D] = 3. Vertex 5 is 2 – dominated by 1 and 4.

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.

In [5], M. Yamuna and K. Karthika have introduced the concept of domination subdivision stable graph.

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 $\gamma (G) = \gamma (G_{sd} uv)$, for all u, v $\in$ V (G), u adjacent to v.

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.

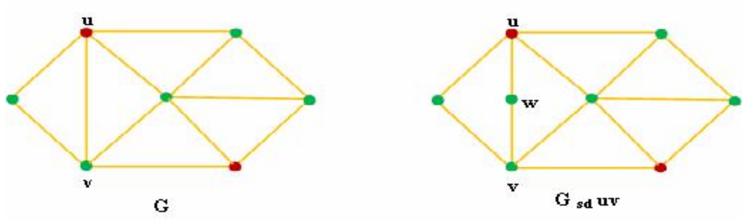


Fig. 3

In Fig. 3, $\gamma (G) = \gamma(G_{sd}uv) = 2$. This is true for all u, v $\in$ V (G), u adjacent to v. G is DSS graph.

In [5] they have proved the following result,

R_1. A graph G is DSS if and only if for every u, v $\in$V (G), either there is a γ - set containing u and v or there is a γ - set D such that either

i. pn[u, D] = { v } or

ii. v is 2 – dominated.

We use C_1, C_2, C_3 to represent these three properties. Let G be a graph and D be any γ - set for G. For any U, V $\in$ V (G) [Here after we shall use upper case letters to represent vertices], U adjacent to V let

C_1: U, V $\in$ D.

C_2: If U $\in$ D, then pn[U, D] = { v }.

C_3: If U $\in$ D, then V is 2 – dominated.

Property of DSS graph

1. If C_1 is satisfied, then D itself is a γ - set for G_{sd}UV also. Also U, V dominates W.

2. If C_2 is satisfied, then D – { U } $\cup$ { W } is a γ - set for G_{sd}UV.

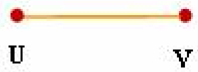
3. If C_3 is satisfied, then D itself is a γ - set for G_{sd}UV also. U dominates W here.

2.3 Edge Values For Binary Encryption

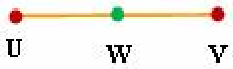
We basically use the idea that for any vertex in the MDS, we assign value 1 and for any vertex not in the MDS, we assign a value 0. Here we use the graphs P_1 to P_4 and properties of C_1, C_2 and C_3.

Case 1

C_1 is satisfied, that is if a graph G has an edge UV, where U, V belongs to a MDS.



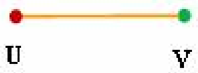
Then we know that γ (G $_{sd}$ UV) = γ (G). When we subdivide an edge UV, we get



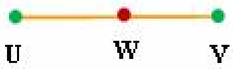
Note that U, V dominates W. Here U $\in$ D, W $\notin$ D with respect to an edge UW. Similarly W $\notin$ D, V $\in$ D with respect to an edge WV. So, we assign the edge value of UV as 1001.

Case 2

C_2 is satisfied, that is if there is a graph G has an edge UV, if U belongs to a MDS and pn[U, D] = { V }.



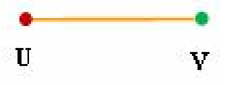
Then we know that γ (G $_{sd}$ UV) = γ (G) = D – { U } $\cup$ { W } . When we subdivide an edge UV, we get



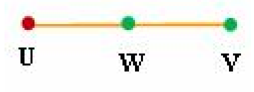
Note that now W dominates U and V. Here U $\notin$ D, W $\in$ D with respect to an edge UW. Similarly W $\in$ D, V $\notin$ D with respect to an edge WV. So, we assign the edge value of UV as 0110.

Case 3

C_3 is satisfied, that is there is a graph G that has an edge UV, U belongs to a MDS D and V is 2 - dominated.



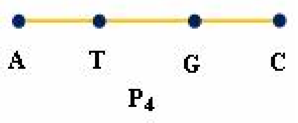
Then we know that $\gamma (G_{sd} UV) = \gamma (G)$. When we subdivide an edge UV, we get



Note that here W is dominated by U. Here U $\in$ D, W $\notin$ D with respect to an edge UW. Similarly W $\notin$ D, V $\notin$ D with respect to an edge WV. So, we assign the edge value of UV as 1000.

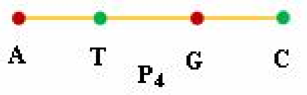
2.4 Proposed Encryption Scheme

We use result R_1 for converting a binary string as a DNA strand. Since a DNA strand has only the four combinations A, T, G, C, we choose graphs with four vertices which are DSS. The only possible DSS graphs with four vertices are P_4 and C_4. We can choose any graph for encryption. We shall use P_4. We label the vertices of P_4 as A, T, G, C.



The possible γ – sets are { A, G }, { A, C }, { T, C } and { T, G }. As we discussed earlier, we can define the edge value for γ – sets.

S_1. A, G Č D



Here pn [G, D] = { C }, that is C_2 is satisfied. Also T is 2 - dominated, that is C_3 is satisfied. So we assign the following edge values as discussed in section 2.4.

AT, GT: 1000

TA, TG: 0001

GC, CG: 0110

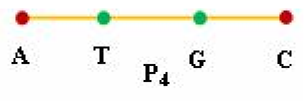
AC: 10

CA: 01

AA, GG, AG, GA: 11

TT, CC, TC, CT: 00

S_2. A, C Č D



Here pn [A, D] = { T }, pn [C, D] = { G }, that is C_2 is satisfied. So we assign the following edge values.

AT, TA, GC, CG: 0110

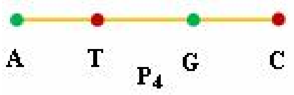
AG, CT: 10

GA, TC: 01

AA, CC, AC, CA: 11

TT, GG, TG, GT: 00

S_3. T, C Č D



Here pn [T, D] = { A }, which implies C_2 is satisfied, G is 2 – dominated, that is C_3 is satisfied. So we assign the following edge values.

AT, TA: 0110

TG, CG: 1000

GT, GC: 0001

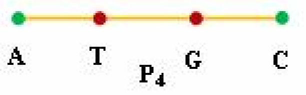
CA: 10

AC: 01

TT, CC, TC, CT: 11

AA, GG, AG, GA: 00

S_4. T, G Č D



T, G ∈ D, which implies that C_1 is satisfied, pn [T, D] = { A }, pn [G, D] = { C }, that C_2 is satisfied. So we assign the following edge values.

AT, TA, GC, CG: 0110

TG, GT: 1001

GA, TC: 10

AG, CT: 01

GG, TT: 11

AA, CC, AC, CA: 00

Method 1: Basic Method [Encryption As DNA]

Encryption Algorithm

Step 1. Consider the compound to be encoded and let S be its corresponding chemical formula.

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

Step 3. Determine the largest decimal value in S' (say k).

Step 4. Convert decimal value S' into a binary number of length m, to obtain a

binary sequence S''.

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

Step 6. Choose any one graph from S_1 to S_4.

Step 7. Divide S'' into segments, where each segment contains k – bits, k = 2 or 4.

Step 8. Replace the segments based on the choice of the graph to obtain a sequence S''' (false DNA sequence).

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 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.

By reversing the procedure we can decrypt the DNA sequence into its corresponding chemical formula.

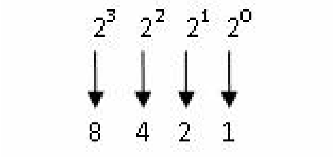
Example

If we want to encode the chemical formula of Caffeine, then consider a chemical formula of **CAFFEINE** and label it as S, that is **S: $C_8H_{10}N_4O_2$**.

By using periodic table [by table - 1], replace each chemical element into atomic number to generate

S': 6 8 1 10 7 4 8 2

Replace each decimal value into binary representation.



The largest decimal value in S' is 10, so the length of binary string for each decimal is 4. Hence from S' we get the binary string 0110 1000 0001 1010 0111 0100 1000 0010

Concatenating the above string we get S'': 011010000001101001110100100000010 to be encoded. The length of the string is 36. Splitting this into two bit strings we get

01 10 10 00 00 01 10 10 01 11 01 00 10 00 00 10

Suppose we prefer to choose S_3, then the given string can be replaced by

AC CA CA AG AA AC CA CA AC CT AC GG CA AG GA CA

if we decide to use encoding as strings of size two only.

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

AT TG GT CA CA AC TT AC AG CA GA AG CA

The given string will be encrypted as

TCGCACCACAAGAAACCACAACCTACGGCAAGGACA or as

TCGCATTGGTCACAACTTACAGCAGAAGCA.

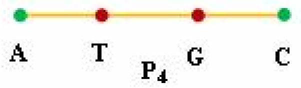
where, the first two entry TC represents S_3 and third fourth entry GC represents that string is of even length. These are two possible sequence among numerous different

combinations available for encrypting the message.

Suppose the message received is

TGGCATAAGGCCAGCTTTCTGGACCTGACCACGA

The first two entries TG indicates that the graph under consideration for encryption is



The third fourth entry GC indicates that the string is of even size. Splitting the message for decoding we get

AT AA GG CC AG CT TT CT GG AC CT GA CC AC GA

Using the string values from S_4, we get

0110 00 11 00 01 01 11 01 11 00 01 10 00 00 10.

Splitting the binary string of length four.

0110 0011 0001 0111 0111 0001 1000 0010

Replace the binary string into decimal values

6 3 1 7 7 1 8 2.

Replace the alternate decimals by their corresponding chemical element from the periodic table we get the formula $\mathbf{C_3H_7NO_2}$, which is the chemical formula for the amino acid **ALANINE**.

Method 2 [Encryption As DNA Using Insertion Method]

Chose a random DNA sequence. The DNA sequence obtained by method – 1 can be inserted into this sequence using the insertion method [2].

Encryption Algorithm

Step 1. Consider a DNA sequence M.

Step 2. Divide M into segments, whereby each segment contains r – bits.

Step 3. Insert bits from S^{iv} one at a time, into the beginning of segment of M, to generate a fake DNA sequence S^{v}.

Step 4. Send S^{v} to the receiver.

Example

Consider a random DNA sequence

M : ATGATAGATCGGTAGCGTAGCGTAGGTACAGTGTACGCGTAGCGTAGGTACAGTGTAGCG.

Splitting this M into two bit strings we get,

AT GA TA GA TC GG TA GC GT AG CG TA GG TA CA GT GT AC GC GT AG CG TA GG TA CA GT GT AG CG.

Consider a chemical formula of caffeine S: $C_8H_{10}N_4O_2$ to be encoded.

By using method - 1we get

S^{iv} : TCGCATTGGTCACAACTTACAGCAGAAGCA.

Insert bits from S^{iv} one at a time, into the beginning of segment of M, to generate a fake DNA sequence AT**T** GA**C** TA**G** GA**C** TC**A** GG**T** TA**T** GC**G** GT**G** AG**T** CG**C** TA**A** GG**C** TA**A** CA**A** GT**C** GT**T** AC**T** GC**A** GT**C** AG**A** CG**G** TA**C** GG**A** TA**G** CA**A** GT**A** GT**G** AG**C** CG**A**.

We can send S^{v}: ATTGACTAGGACTCAGGTTATGCGGTGAGTCGCTAAGGCT

AACAAGTCGTTACTGCAGTCAGACGGTACGGATAGCAAGTAGTGAGCCGA

Suppose the message received is

S^{v}: GCTCTGCTGATCTTAAATTATAGGGTATGAACAATTTACCCCAGTGCC

CTATGACGATGGTGCCGA.

Splitting the message for decoding we get

GCT CTG CTG ATC TTA AAT TAT AGG GTA TGA ACA ATT TAC CCC AGT GCC CTA TGA CGA TGG TGC CGA

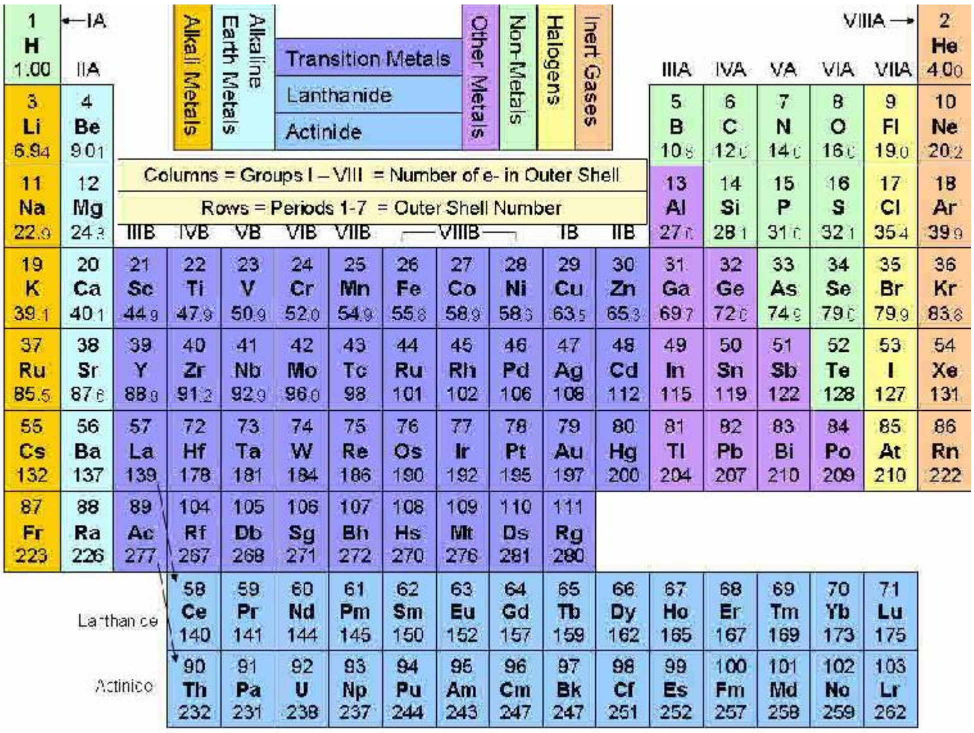
From the above bits, we can get S^{iv}: TGGCATTGAAATCCTCAAAGCA

Again by applying method – 1, we can encode the above DNA segment. From this we can obtain the original chemical formula for the amino acid **ALANINE**.

Remark

P_4 and C_4 are the only graphs that are domination subdivision stable with four vertices. C_4 can also be used for encryption by assigning edge values as discussed in the case of P_4.

Table: 1



Alkali Metals | Alkaline Earth Metals | Transition Metals | Lanthanide | Actinide | Other Metals | Non-Metals | Halogens | Inert Gases

Columns = Groups I – VIII = Number of e- in Outer Shell

Rows = Periods 1-7 = Outer Shell Number

IA	IIA	IIIB	IVB	VB	VIB	VIIB	VIIIB	VIIIB	VIIIB	IB	IIB	IIIA	IVA	VA	VIA	VIIA	VIIIA
1 **H** 1.00																	2 **He** 4.00
3 **Li** 6.94	4 **Be** 9.01											5 **B** 10.8	6 **C** 12.0	7 **N** 14.0	8 **O** 16.0	9 **Fl** 19.0	10 **Ne** 20.2
11 **Na** 22.9	12 **Mg** 24.3											13 **Al** 27.0	14 **Si** 28.1	15 **P** 31.0	16 **S** 32.1	17 **Cl** 35.4	18 **Ar** 39.9
19 **K** 39.1	20 **Ca** 40.1	21 **Sc** 44.9	22 **Ti** 47.9	23 **V** 50.9	24 **Cr** 52.0	25 **Mn** 54.9	26 **Fe** 55.8	27 **Co** 58.9	28 **Ni** 58.3	29 **Cu** 63.5	30 **Zn** 65.3	31 **Ga** 69.7	32 **Ge** 72.0	33 **As** 74.9	34 **Se** 79.0	35 **Br** 79.9	36 **Kr** 83.8
37 **Ru** 85.5	38 **Sr** 87.6	39 **Y** 88.9	40 **Zr** 91.2	41 **Nb** 92.9	42 **Mo** 96.0	43 **Tc** 98	44 **Ru** 101	45 **Rh** 102	46 **Pd** 106	47 **Ag** 108	48 **Cd** 112	49 **In** 115	50 **Sn** 119	51 **Sb** 122	52 **Te** 128	53 **I** 127	54 **Xe** 131
55 **Cs** 132	56 **Ba** 137	57 **La** 139	72 **Hf** 178	73 **Ta** 181	74 **W** 184	75 **Re** 186	76 **Os** 190	77 **Ir** 192	78 **Pt** 195	79 **Au** 197	80 **Hg** 200	81 **Tl** 204	82 **Pb** 207	83 **Bi** 210	84 **Po** 209	85 **At** 210	86 **Rn** 222
87 **Fr** 223	88 **Ra** 226	89 **Ac** 277	104 **Rf** 267	105 **Db** 268	106 **Sg** 271	107 **Bh** 272	108 **Hs** 270	109 **Mt** 276	110 **Ds** 281	111 **Rg** 280							
		Lanthance	58 **Ce** 140	59 **Pr** 141	60 **Nd** 144	61 **Pm** 145	62 **Sm** 150	63 **Eu** 152	64 **Gd** 157	65 **Tb** 159	66 **Dy** 162	67 **Ho** 165	68 **Er** 167	69 **Tm** 169	70 **Yb** 173	71 **Lu** 175	
		Actinico	90 **Th** 232	91 **Pa** 231	92 **U** 238	93 **Np** 237	94 **Pu** 244	95 **Am** 243	96 **Cm** 247	97 **Bk** 247	98 **Cf** 251	99 **Es** 252	100 **Fm** 257	101 **Md** 258	102 **No** 259	103 **Lr** 262	

3. Conclusion

Any chemical formula can be encrypted as a DNA sequence. This when available in public domain looks like one among the hundreds of fake DNA strands and it is difficult to identify that a chemical formula is been encrypted. A binary string of any length can be encoded using this technique. So any formula of any length can be encoded. Even combined chemical formulas can be encrypted using this method. Moreover graph properties are used for encryption and hence even if one knows that a DNA strand has been used for encrypting breaking the code is not possible unless the property is known. Four different γ - sets provides us more secured encryption as decoding would be possible only if the graph and its γ - set is known. Also together with the six cycles totally ten graphs can be used for encryption. So the proposed method is safe for encryption of any formula.

4. References

1. Haynes T. W, Hedetniemi S. T and Slater P. J., Fundamentals of Domination in Graphs, Marcel Dekker, New York , 1998.
2. Shiu H.J. et al., Data hiding methods based upon DNA sequences, Information Sciences, 2010, Vol. 180, No. 11, pp: 2196 – 2208.
3. http://en.wikipedia.org/wiki/Chemical_formula.
4. http://www.webqc.org/periodictable.php
5. Yamuna M. et al., Domination Subdivision Stable Graphs, International Journal of Mathematical Archive, 2012, Vol. 3, No. 4, pp: 1467 – 1471.
6. Youssef M. I. et al., Multi-Layer Data Encryption Using Residue Number System in DNA Sequence, International Journal of Security and Its Applications, 2012, Vol. 6, No. 4, pp: 1 – 12.
7. Yunpeng Zhang and Liu He Bochen Fu, Research on DNA Cryptography, Applied Cryptography and Network Security.

★★★★★