

Amino Acids in Data Encryption

Abstract

The security to a system is essential nowadays with the growth of Information technology and with the emergence of new techniques; the number of threats that a user is supposed to deal with has grown exponentially. To achieve security, it is very necessary to encode the data before sending it through the various communication channels available to make it unreadable. In this paper we proposed a method of encrypting DNA sequence using amino acids.

Keywords: Decryption; Encryption; DNA sequence; Amino acid; Deoxyribonucleic acid; Thymine; Cytosine; Guanine

Research Article

Volume 2 Issue 5 - 2016

Yamuna M* and Elakkiya A

School of Advanced Sciences, VIT University, India

*Corresponding author: Yamuna M, School of Advanced Sciences, VIT University, Vellore, India, Email: myamuna@vit.ac.in

Received: June 01, 2016 | Published: June 28, 2016

Introduction

DNA, deoxyribonucleic acid, is a molecule made out of nucleic acids that can be found in every cell in our body and forms the genetic information of each living organism. Consequently, DNA is often noted as the "blueprint of biological life", as it gives instructions for an organism's functioning and development. A single DNA molecule is double stranded and has sequences of four bases: adenine (A), thymine (T), cytosine (C), and guanine (G).

A DNA database is a collection of human DNA samples that is often derived from blood, tissue, or saliva.

DNA databases were first established in the 1980s and were initially in forensics to identify criminals and in the military to help recognize deceased military members based on their remains. Today, DNA plays an important role in military, offence and other medical research so safe transfer of DNA is important. In this paper, we have proposed a method of converting DNA sequence into binary string sequence to increase security.

Bazli et al. [1] proposed a DNA encryption scheme and the use of biological alphabets to manipulate information by employing the DNA sequence reaction, to autonomously make a copy of its threads as an extended encryption key [1]. Umalkar et al. [2] proposed a message cryptography formula supported DNA sequence using complementary rules deoxyribonucleic acid sequence [2]. Atito et al. [3] proposed a novel algorithm to communicate data securely. The proposed technique is a composition of both encryption and data hiding using some properties of DNA sequences [3]. Yamuna et al. [4] proposed a method of encrypting DNA sequence using pre - order tree traversal [4]. We have used all the papers mentioned above as base papers for our proposed method.

Discussion

DNA sequence

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. It includes any method or technology that is used to determine the order of the four bases -adenine, guanine, cytosine, and thymine - in a strand of DNA [5]. Figure 1 provides an example of DNA sequence [6].

Proposed encryption scheme

In this section we propose a method of encrypting a DNA sequence as a binary string using amino acid properties.

Construction of binary table

Amino acids play central roles both as building blocks of proteins and as intermediates in metabolism. The 20 amino acids that are found within proteins convey a vast array of chemical versatility. All amino acids found in proteins have the basic structure, differing only in the structure of the R-group or the side chain. For our conversion we are considering the distance of all the atoms such as carbon, oxygen, nitrogen, sulphur from the alpha - carbon. In this method we are converting each amino acid into binary string of length 12. The first two digits represent the chemical property with which the amino acid belongs i.e., for polar group the binary string is 00, for aromatic and non - polar 01 and 11 respectively. The next four digits represent the presence of carbon, oxygen, nitrogen, sulphur respectively. For example, if the amino acid contains carbon and nitrogen then the binary representation is 1010. The next 6 digits represent the distances of these four atoms from the alpha carbon.

For example, consider Methionine, which belongs to the polar group, so the first two digits are 00. There are also three carbons and one sulphur in the R - group of Methionine, so the next four digits are 1001. The carbon and sulphur are at the distances 1, 2, 4 and 3 respectively from the alpha - carbon. Total sum of these distances is 10; convert the number to the binary length of 6 i.e., $10 = 001010$. Therefore the binary conversion of methionine is **001001001010**.

Similarly, we are converting all the other amino acids into binary string using the above procedure. This binary string is listed in the Table 1.

Encryption algorithm

Step 1 Let Z be the sequence to be encrypted.

Let Z = ATGACGATGACTGATCGATCGATGACGTAT.

Step 2 Split the DNA sequences into codons.

For our example, Z = ATG ACG ATG ACT GAT CGA TCG ATG ACG TAT.

Step 3 Convert the codons in the DNA sequence into its corresponding amino acids.

In our example Z = M T M T D R S M T Y

Step 4 Convert the amino acids into a binary string of length 12 using Table 1.

M = 001001001010, T = 111100000101, M = 001001001010, T = 111100000101, D = 111100001001, R = 111010011011, S = 111100000011, M = 001001001010, T = 111100000101 Y = 011100011100.

Step 5 Concatenating the binary string we generate a binary sequence k.

For our example the binary string generated is k = 00100100101011110000010100100100101011

1100000101111000010011110100110111110000001100100101011110000010101110001

1100

Step 6 Send this k to the receiver.

Decryption algorithm

For decrypting the sequence, we reverse the procedure.

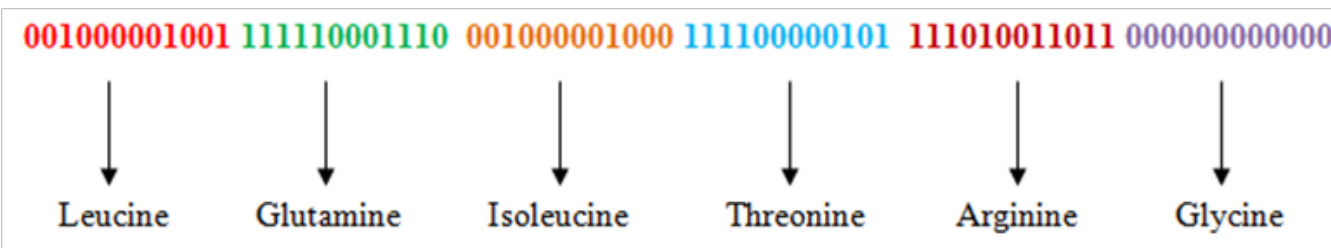
Suppose the received sequence is 00100000100111110001110001000010000100011110000010111

1010011011000000000000.

Step 1 Split this sequence into segments of length 12.

001000001001 11110001110 001000001000 111100000101 111010011011 000000000000

Step 2 Convert this binary string into amino acids using Table 1.



Step 3 The corresponding codons for the above amino acid are CTG CAG ATC ACC AGG GGG. The sequence is decrypted as **CTGCAGATCACCAGGGGG**.

Table 1: Binary Conversion of Amino Acids.

S. No	Amino Acids	Carbon	Oxygen	Nitrogen	Sulphur	Index Sum	Binary Conversion
1	Glycine	0	0	0	0	0C	000000000000
2	Alanine	1	0	0	0	1C	001000000001
3	Valine	5	0	0	0	5C	001000000101
4	Leucine	9	0	0	0	9C	001000001001
5	Methionine	7	0	0	3	7C + 3S	001001001010
6	Isoleucine	8	0	0	0	8C	001000001000
7	Phenylalanine	22	0	0	0	22C	011000010110
8	Tyrosine	22	6	0	0	22C + 6O	011100011100
9	Tryptophan	33	0	4	0	33C + 4N	011010100101
10	Serine	1	2	0	0	1C + 2O	111100000011
11	Threonine	3	2	0	0	3C + 2O	111100000101
12	Cysteine	1	0	0	2	3C	111000000011
13	Proline	5	0	0	0	5C	111000000101
14	Asparagine	3	3	3	0	3C + 3O + 3N	111110001001
15	Glutamine	6	4	4	0	6C + 4O + 4N	111110001110
16	Lysine	10	0	5	0	10C + 5N	111010001111
17	Arginine	11	0	16	0	11C + 16N	111010011011
18	Histidine	10	0	7	0	10C + 7N	111010010001
19	Aspartate	3	6	0	0	3C + 6O	111100001001
20	Glutamate	6	8	0	0	6C + 8O	111100001110

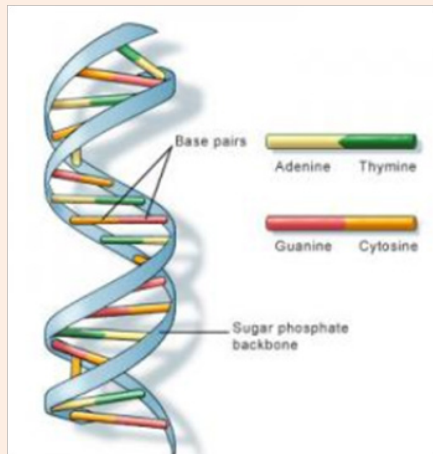


Figure 1: DNA Sequence.

Conclusion

DNA is important not only because it makes everyone biologically different from one another, but also because it is the unique identifier that humans are born with, and cannot change. Unlike other personal items which can be used to identify

individuals, DNA cannot be replaced or changed. Hospitals establish medical databases to make DNA samples available for research purposes and also private organizations establish research databases to study specific diseases and conditions. The proposed method is secure and it would be very difficult for any intruder to break the encrypted message and retrieve the actual message.

References

1. Behnam B, Mustafa AT, David LJ (2014) Data Encryption Using Bio Molecular Information. *International Journal on Cryptography and Information Security* 4(3): 21-32.
2. Amruta DU, Pritish AT (2014) Data Encryption Using DNA Sequences Based On Complementary Rules-A Review. *International Journal of Engineering Research and General Science* 2(6): 345-349.
3. Atito A, Khalifa A, Rida SZ (2012) DNA-based data encryption and hiding using play fair and insertion techniques. *Journal of Communications & Computer Engineering* 2(3): 44-49.
4. Yamuna M, Elakkiya A (2015) Amino Acid Graph Representation for Efficient Safe Transfer of Multiple DNA Sequence as Pre – Order Trees. *International Journal of Bioinformatics and Biomedical Engineering* 1(3): 292-299.
5. Anjana M (2012) DNA Sequencing - Methods and Applications. *Janeza Trdine, India*, pp. 1-184.
6. http://www.graphene-info.com/files/graphene/DNA-strand-img_assist-300x300.jpg