Available online at www.scholarsresearchlibrary.com

Scholars Research Library

Der Pharmacia Lettre, 2015, 7 (7):58-66
(http://scholarsresearchlibrary.com/archive.html)

Scholars Research Library
ISSN 0975-5071
USA CODEN: DPLEB4

Point mutation determination using graph theory

Yamuna M. and Karthika K.

SAS, VIT University, Vellore, Tamilnadu, India

ABSTRACT

Thousands of proteins need to do their job in the right places at the right time for any cell to function properly. Sometimes, gene mutation prevent one or more of these proteins from working properly. By change in gene instruction, in protein making, a mutation can cause the protein to malfunction or to be missing completely. This may cause severe medical conditions. Any kind of treatment requires an initial identification of the mutation kind from the samples provided. Various techniques are already available and being used. In this paper we propose a method of mutation identification using graph theory properties.

Key words: Mutation, Point Mutation, Graph, Degree, Proteins.

INTRODUCTION

Graphs can be used to model many types of relations and processes in physical, biological, social and information systems. Many practical problems can be represented by graphs. Graph theory is useful in biology and conservation efforts where a vertex can represent regions where certain species exist (or habitats) and the edges represent migration paths, or movement between the regions. This information is important when looking at breeding patterns or tracking the spread of disease, parasites or how changes to the movement can affect other species [1].

In [2], Debra J Knisley et al presented a graph-theoretic model of NBD1 with the vertex-weighted hierarchical graph representation of the protein domain NBD1. They have presented a method to model the effect of a single point mutation of a protein. Using the hierarchical, vertex-weighted graph, they define novel combinatorial descriptors based on these vertex – weights. In [3], they have provided the protein structure of scTIM with a graph-theoretic model. This modeling method have provided a novel approach to the visualization of protein structures and the consequences of amino acid deletions, insertions or substitutions and also provided a new way to gain insight on the consequences of diseases caused by genetic mutations.

In [4], Gregory A Ryslik et al have given a novel methodology that increases the power to identify mutational clusters by taking into account protein tertiary structure via a graph theoretical approach.

Genetic code for amino acids using Huffman Codes and use it for encrypting any DNA sequence is described in [5]. In this paper we provide a method of identifying point mutation using graph properties.

Preliminary Note

In this section we provide the basic details required for the proposed method.

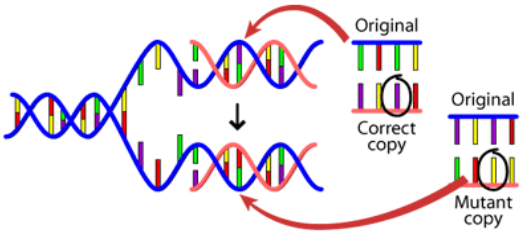
Graph theory

A linear graph (or simply a graph) G = (V, E) consists of a set of objects V = { v_1, v_2, … } called vertices, and another set E = { e_1, e_2, …}, whose elements are called edges, such that each edge e_k is identified with an unordered pair (v_i, v_j) of vertices. The vertices v_i, v_j associated with edges e_k are called the end vertices of e_k. The most

common representation of a graph is by means of a diagram, in which the vertices are represented as points and each edge as a line segment joining its end vertices. A directed graph (or a digraph for short) G consists of a set of vertices V = { v_1, v_2, …}, set of edges E = { e_1, e_2, … } and a mapping ψ that maps every edge onto some ordered pair of vertices (v_i, v_j). As in the case of undirected graphs, a vertex is represented by a point and an edge by a line segment between v_i and v_j with an arrow directed from v_i to v_j. The number of edges incident of v_i is called the out – degree of v_i. The number of edges incident into v_i is called the in – degree of v_i [6].

Mutation

When the DNA molecule in the cell replicates, several chemical changes occur during this process. These changes are sometimes able to repair but there are some changes which cannot be changed. They are called as mutations. Due to the mutations DNA molecule cannot repair itself and passes to the next generation. Mutations have the ability that they can change the entire amino acid sequences of proteins which are present inside the gene. There are some mutations which do not affect the function of the proteins but some mutations are harmful that is they do not allow the amino acids to code properly for the protein and as a result the genes do not function properly [7]. An example of point mutation is given in Snapshot – 1 [8].



Snapshot – 1

Nonsynonymous Mutations

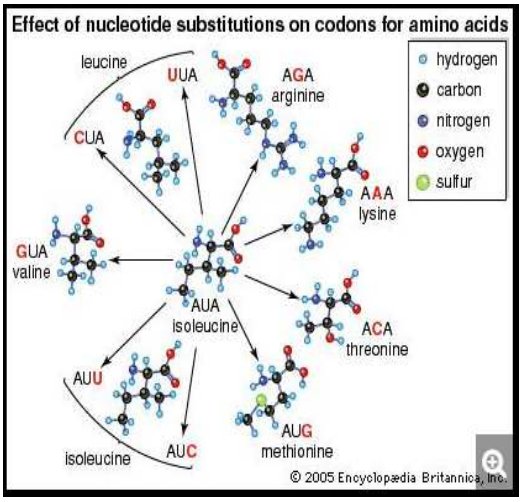
Nonsynonymous mutations have a much greater affect on an individual than a synonymous mutation. In a nonsynonymous mutation, there is usually an insertion or deletion of a single nucleotide in the sequence during transcription when the messenger RNA is copying the DNA. This single missing or added nucleotide causes a frame shift mutation which throws off the entire reading frame of the amino acid sequence and mixes up the codons. This usually does affect the amino acids that are coded for and change the resulting protein that is expressed. The severity of this kind of mutation depends on how early in the amino acid sequence it happens. If it happens near the beginning and the entire protein is changed, this could become a lethal mutation.

Another way a nonsynonymous mutation can occur is if the point mutation changes the single nucleotide into a codon that does not translate into the same amino acid. A lot of times, the single amino acid change does not affect the protein very much and is still viable. However, if it happens early in the sequence and the codon is changed to translate into a stop signal, then the protein will not be made and it could cause serious consequences.

Sometimes nonsynonymous mutations are actually positive changes. Natural selection may favor this new expression of the gene and the individual may have developed a favorable adaptation from the mutation. If that mutation occurs in the gametes, this adaptation will be passed down to the next generation of offspring. Nonsynonymous mutations increase the diversity in the gene pool for natural selection to work on and drive evolution on a micro evolutionary level [9].

Point mutation

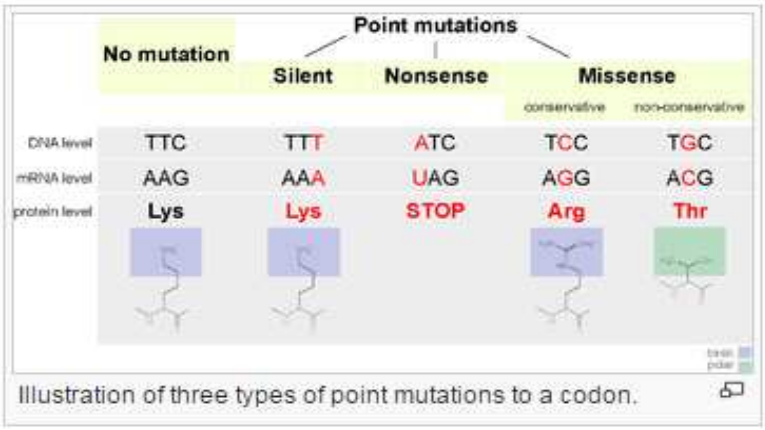
A point mutation, or single base modification, is a type of mutation that causes a single nucleotide base change, insertion, or deletion of the genetic material, DNA or RNA [10]. The effect of base substitutions, or point mutations, on the messenger – RNA codon AUS, which codes for the amino acid isoleucine is seen in the following Snapshot – 2



Snapshot – 2

Substitutions (red letters) at the first second, or third position in the codon can result in nine new codons corresponding to six different amino acids in addition to isoleucine itself [11] .

Snap shot – 3 shows the effect of different types of point mutations on the amino acids [10].



Snapshot – 3

In this paper we provide a graph theoretic approach to identify nonsense and missense mutation.

Proposed Verification Scheme
Graph Construction
Vertex set: Since codons are made of A, T, G, C only we choose four vertices and label them as A, T, G, C.

Edge set: We draw edges between vertices as follow

1.Choose any protein, say for example Alanine. The codons that contribute to Alanine are GCT, GCC, GCA, GCG. All the codons start with G. So we fix G as the start vertex. We indicate the start vertex in blue colour.

2.For each codon XYZ we draw edges directed fro X to Y and Y to Z. For example for Alanine the edge sets are { (G, C), (C, T), (G, C),(C, C), (G, C), (C, A), (G, C), (C,G) }.

The resulting graph for Alanine is as seen in Fig. 1.

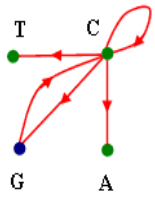


Fig. 1

Some graphs have more than one start vertex. In those cases we draw two graph one for each start vertex. For example for Arginine the codons are CGT CGC CGA CGG, AGA AGG. The two start vertices are C, A. The corresponding two graphs are as seen in Fig. 2

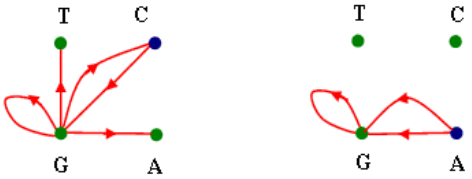


Fig. 2

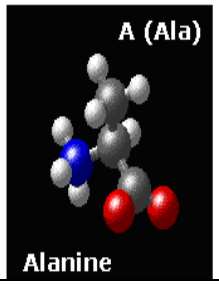
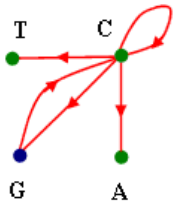
Degree Sequence Determination

For each protein we determine a degree sequence as follows.

1. For proteins with single graph we directly take the degree of each vertex as the sum of indegree + out degree.
2. The sequence is taken in the order of the vertices T, C, A, G.

For example for Alanine the degree sequence is as seen in Table – 1.

Table – 1

Protein	Graph	Vertex	In Degree	Out Degree	Sum
		T	1	0	1
		C	2	4	6
		A	1	0	1
		G	1	1	2
Resulting Degree Sequence	**1612**				

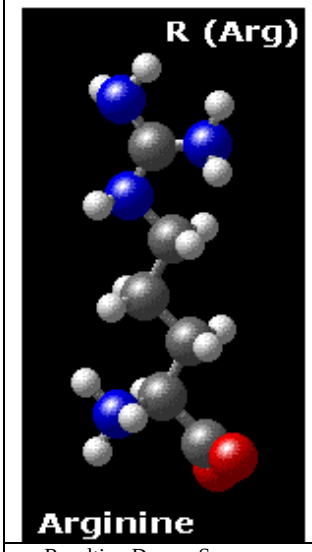
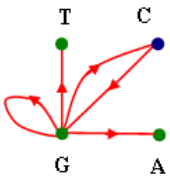
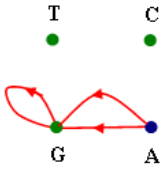
For proteins with more than one graph to have a unique degree sequence we proceed as follows. So the degree sequence of Alanine is 1612.

3. For proteins with more than one graph to have a unique degree sequence we proceed as follows

- Take the sum of the degree sequence of the two graphs.
- For the first two digits of the sum take addition modulo 9.

For example for Arginine the degree sequence is as seen in Table – 2

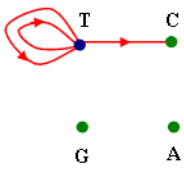
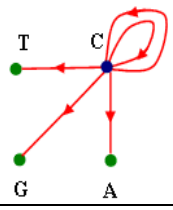
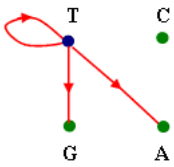
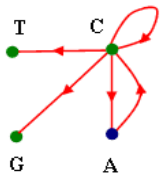
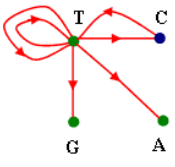
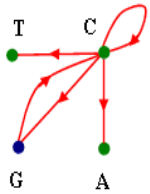
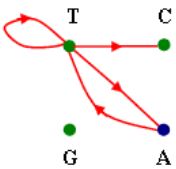
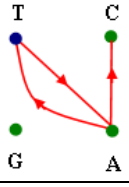
Table – 2

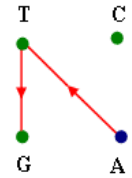
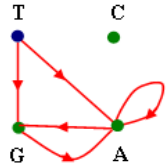
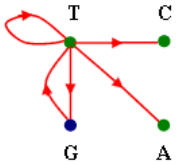
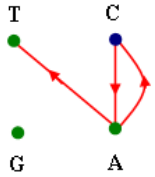
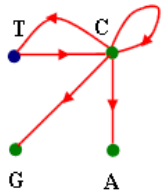
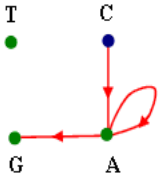
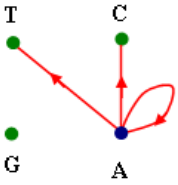
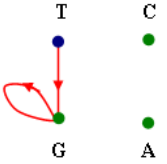
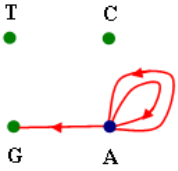
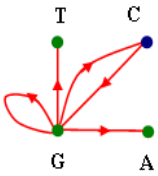
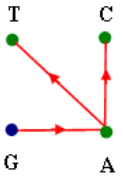
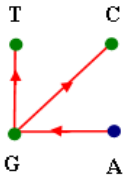
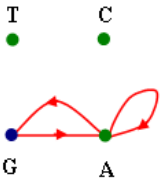
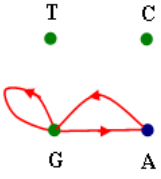
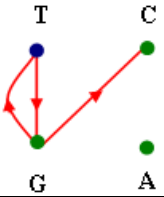
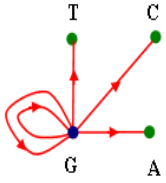
Protein	Graph	Vertex	In Degree	Out Degree	Sum
		T	1	0	1
		C	1	1	2
		A	1	0	1
		G	2	4	6
		T	0	0	0
		C	0	0	0
		A	1	1	2
		G	3	1	4
Resulting Degree Sequence	**1216 + 0024 = 1240**				

So the degree sequence of Arginine is 1240.

Graphs for all the twenty amino acids and the corresponding degree sequences is as seen in Table – 3.

Table – 3

S. No.	Protein	Codon	Graph Representation	Degree Sequence	S. No.	Protein	Codon	Graph Representation	Degree Sequence
1	F	TTT TTC		5100	**8**	P	CCT CCC CCA CCG		1711
2	L	TTA TTG		2222	**9**	T	ACT ACC ACA ACG		1621
3	L	CTT CTC CTA CTG		2222	**10**	A	GCT GCC GCA GCG		1612
4	I	ATT ATC ATA		5120	**11**	Y	TAT TAC		2130

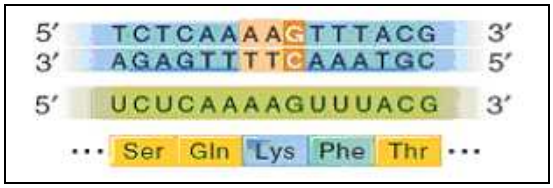
5	M	ATG		2011	**12**	STOP	TAA TAG TGA		2053
6	V	GTT GTC GTA GTG		6112	**13**	H	CAT CAC		1230
7	S	TCT TCC TCA TCG		3724	**14**	Q	CAA CAG		0141
15	N	AAT AAC		1140	**20**	W	TGG		1003
16	K	AAA AAG		0051	**21**	R	CGT CGC CGA CGG		1240
17	D	GAT GAC		1131	**22**	S	AGT AGC		3724
18	E	GAA GAG		0042	**23**	R	AGA AGG		1240
19	C	TGT TGC		2103	**24**	G	GGT GGC GGA GGG		1117

Point Mutation Verification

Silent Mutation

Silent mutations are DNA mutations that do not significantly alter the phenotype of the organism in which they occur. Silent mutations can occur in non-coding regions, or they may occur within exons. When they occur within exons they either do not result in a change to the amino acid sequence of a protein, or result in the insertion of an alternative amino acid with similar properties to that of the original amino acid, and in either case there is no significant change in phenotype [12]. An example of silent mutation is given Snapshot – 4 and 5 [13].

Snapshot – 4



Snapshot – 5

One of the nucleotide (AAA) in an amino acid sequence in Snapshot – 4 is replaced by another nucleotide (AAG) in an amino acid sequence in Snapshot – 5 of the DNA sequence. But both amino acid sequences do not affect the protein Lysine.

Verification Using Degree Sequence

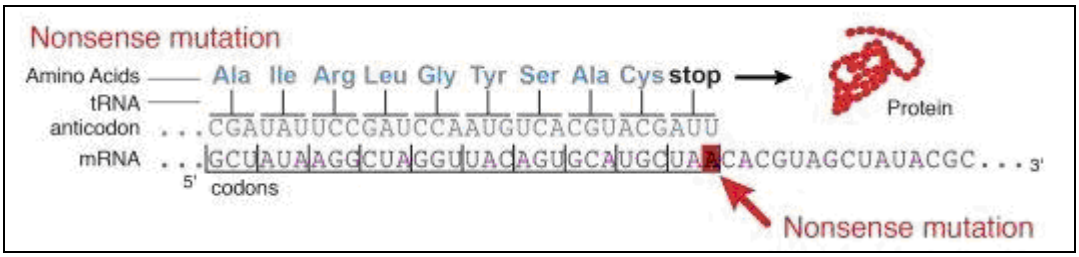
For the sequences provided in Snapshot – 4 and 5, sequence verification is done as seen in Table – 4

Table – 4

Sequence 1	TCT CAA AAA TTT ACG
Sequence 2	TCT CAA AAG TTT ACG
Degree Sequence 1	3724 0141 0051 5100 1621
Degree Sequence 2	3724 0141 0051 5100 1621
Conclusion	Since degree sequences are equal they represent silent mutation.

Nonsense Mutation

In genetics, a nonsense mutation is a point mutation in a sequence of DNA that results in a premature stop codon, or a nonsense codon in the transcribed mRNA, and in a truncated, incomplete, and usually nonfunctional protein product [14]. An example of nonsense mutation is given in Snapshot – 6 [15].



Snapshot – 6

In Snapshot – 6, the stop codon is introduced at the 10 th place in amino acid sequences.

Verification Using Degree Sequence

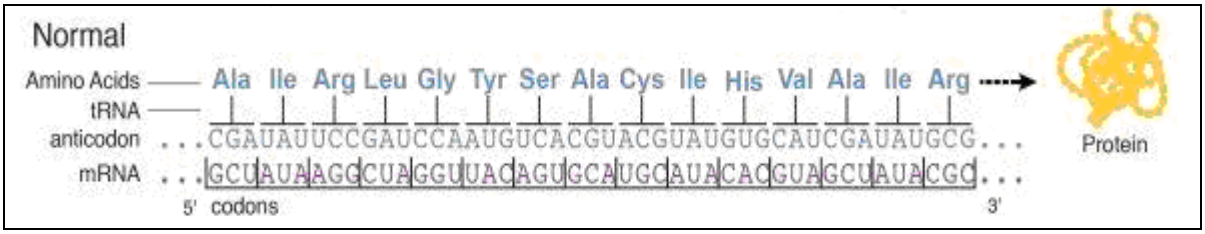
For the sequences provided in Snapshot – 6, sequence verification is done as seen in Table – 5

Table – 5

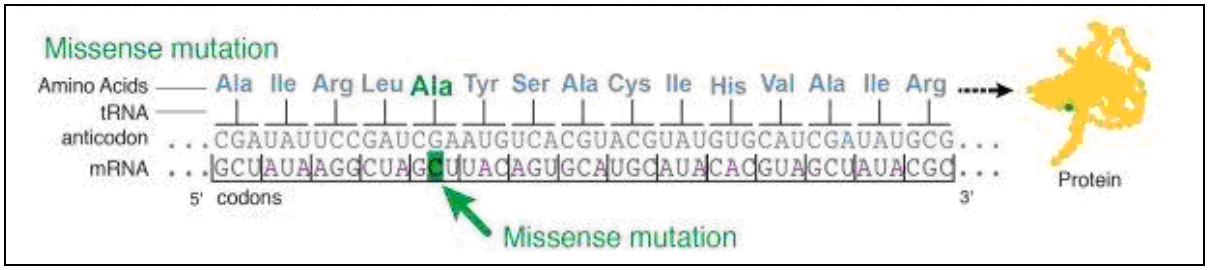
Sequence 1	GCU AUA AGG CUA GGU UAC AGU GCA UGC UAA CAC
Degree Sequence 1	1612 5120 1240 2222 1117 2130 3724 1612 2103 2053 1230
Conclusion	Since the stop codon appears in the degree sequence it represents nonsense mutation.

Missense Mutation

In genetics, a missense mutation (a type of nonsynonymous substitution) is a point mutation in which a single nucleotide change results in a codon that codes for a different amino acid [16]. An example of missense mutation is given in Snapshot – 7 and 8 [15].



Snapshot – 7



Snapshot – 8

Compare Snapshot – 7 and 8. One of the nucleotide (Glycine) in Snapshot – 7 is replaced by another nucleotide (Cytosine) in Snapshot – 8. This results in an incorrect amino acid (Alanine) being incorporated into the protein sequence of Snapshot – 8.

Verification Using Degree Sequence

For the sequences provided in Snapshot – 7 and 8, sequence verification is done as seen in Table – 6.

Table – 6

Sequence 1	GCU AUA AGG CUA GGU UAC AGU GCA UGC AUA CAC GUA GCU AUA CGC
Sequence 2	GCU AUA AGG CUA GCU UAC AGU GCA UGC AUA CAC GUA GCU AUA CGC
Degree Sequence 1	1612 5120 1240 2222 1117 2130 3724 1612 2103 5120 1230 6112 1612 5120 1240
Degree Sequence 2	1612 5120 1240 2222 1612 2130 3724 1612 2103 5120 1230 6112 1612 5120 1240
Conclusion	Since degree sequences are not equal they represent missense mutation.

CONCLUSION

Graph theory is growing as a promising field in various fields of biotechnology, medicine due its compatibility in representing any structure as a graph. It is compatible for coding also as various algorithms are developed using graph theory. Protein point mutations are an essential component of the evolutionary and experimental analysis of protein structure and function. While many manual databases attempt to index point mutations, our approach is to identify the different types of point mutation in an easy way through a number sequences.

In the proposed method each amino acid is represented as a graph. Moreover degree sequence of the graphs are used for mutation verification. Since umbers can be programmed easily using MATLAB, the proposed method can be used as a means of initial verification for the possible occurrence of the mutations in the sequences provided.

REFERENCES

[1] https://en.wikipedia.org/wiki/Graph_theory.
[2] Debra J. Knisley; Jeff R. Knisley; Andrew Cade Herron. *Computational Biology Journal*, **2013**, vol. 2013, Article ID 938169, 9 pages. doi:10.1155/2013/938169.

[3] Debra J Knisley; Jeff R Knisley. Seeing the Results of a Mutation With a Vertex Weighted Hierarchical Graph, From 3rd IEEE Symposium on Biological Data Visualization, Atlanta, GA, USA. 13 -14 October **2013**, BMC Proceedings **2014**, 8(Suppl 2):S7. http://www.biomedcentral.com/1753-6561/8/S2/S7.
[4] Gregory A Ryslik; Yuwei Cheng; Kei-Hoi Cheung; Yorgo Modis; Hongyu Zhao. *MC Bioinformatics* **2014**, 15:86. http://www.biomedcentral.com/1471-2105/15/86.
[5] M. Yamuna; B. Joseph Sasikanth Reddy; Nithin Kumar Reddy; Paladugula Raghuram. *International Journal of ChemTech Research*, **2014**, Vol.6, No.1, pp 53 – 63.
[6] Narsing Deo. Graph theory with Applications to Engineering and Computer Science, Prentice Hall India, **2010**.
[7] http://www.biotecharticles.com/Genetics-Article/Types-of-Mutations-Frameshift Chromosomal-and-Point-Mutation-120.html.
[8] http://evolution.berkeley.edu/evolibrary/article/0_0_0/mutations_04.
[9] http://evolution.about.com/od/Overview/a/Synonymous-Vs-Nonsynonymous-Mutations.htm.
[10] https://en.wikipedia.org/wiki/Point_mutation.
[11] http://www.britannica.com/science/point-mutation.
[12] https://en.wikipedia.org/wiki/Silent_mutation.
[13] https://www.studyblue.com/notes/note/n/chapter-7/deck/2223355.
[14] https://en.wikipedia.org/wiki/Nonsense_mutation.
[15] https://www.genome.gov/dmd/img.cfm?node=Photos/Graphics&id=85225.
[16] https://en.wikipedia.org/wiki/Missense_mutation.