

The Need for Early Detection of the Prototype Mutants: Sickle Cell Anemia as a Case Study

Amro Abd Al Fattah Amara*

Genetic Engineering and Biotechnology Research Institute, City for Scientific Research and Technological Applications, New Borg Al Arab, Alexandria, Egypt

Abstract

It is basic knowledge that three of the DNA nucleotides are code for one amino acid. That cause some sort for protection, while in many cases one nucleotide change did not lead to a new amino acid (silent mutant). However, we should not neglect such silent mutant particularly if it has happened in a sensitive gene such as β -globin. Such silent mutant is the base for a complete mutant, which might cause severe disease. Because protein is the macromolecules, which responsible for most of the cell functions, most studies are done so far on the protein level including, comparing protein sequences. Genetic diseases could be happened due to one nucleotide change, which could be responsible about causing dramatic illness such as the Sickle cell anemia. However, sometimes there is a need for two or even three nucleotide changes to mutate a single amino acid. Such change(s) might take longer or even generations to be happened. Alternatively, it can be simply avoided. The expected mutant can be early detected during its prototype phase. Such prototype mutation should be detected before it completely changes to full mutation. In this study, an investigation among the protein and the DNA sequences has been done aiming to prove that DNA is more suitable for detecting genetic disease and prototype mutants.

Keywords: Sickle cell anemia; prototype mutants; DNA sequences; Genetic disorder

Introduction

Human biochemical genetics and the term of the inborn errors of metabolic genetic based disease were established in 1902 [1]. Albinism is one prototypic defect studied by Garrod, where the deficiency of the enzyme tyrosinase in the hair, skin and eye prevent the synthesis of the pigment melanin. However, the most studied genetic disease is the Sickle cell anemia [2]. The World Health Organization (WHO) (1982) estimated that about five percent of the world populations are carriers of genes for clinically important disorders of hemoglobin. Third of a million severely affected homozygotes or compound heterozygotes are born each year. For more details refer to Weatherall et al. [3] and the references within. Hemoglobin is the oxygen carrier tetrameric molecule and can be found in vertebrate red blood cells, in some invertebrates and in the root nodules of legumes [1,4]. Each subunit is composed of a polypeptide chain, globin, and a prosthetic group, heme, which is an iron-containing pigment that combine with oxygen and gives the molecules its oxygen-transporting ability. Sickle cell anemia is a global disease and for the Mediterranean and the Africans communities is a local disease [5]. Livingstone, has described in detailed the roles which affect the percentage and the distribution of the Sickle cell anemia in West Africa. From the time of specifying the role of the heredity (the most critical one) till producing artificial blood and artificial oxygen carrier, the science and the scientist did not stop [6]. Additionally, scientists, particularly, those from the Sickle cell anemia endemic regions and countries, have summarized their experiences as well as the experience gained from their communities in controlling the disease side effect. There is a need for better diagnosis and simplifying the existing information for the public. Such information could be introduced as simple advice, to the communities in the west of sub-Saharan in Africa and Mediterranean region [7,8]. However, Sickle cell anemia patients will be nearly impossible to be recovered because it is a genetically based degenerative disease. That is mean each somatic cell has the responsible gene. At least its full treatment will not be in the coming years. This study tries to show that bioinformatics tools must give more concern to the DNA sequences, where DNA can show that prototype mutants are existed. The early detection of such mutants will

enable us to avoid such built in serious genetic disease.

Materials and Methods

The nucleotides sequences collection

Normal β globin DNA sequence was the start point of this study. The sequence has been obtained from the www.ncbi.nlm.nih.gov nucleotides database and has transferred to the nucleotides blast search. These nucleotide database search nucleotides using a nucleotide query (Blast.ncbi.nlm.nih.gov/Blast.cgi) [9].

The nucleotides sequences adjustment

After the search is complete twenty-six sequences have been selected based on the sickle mutant variation (Table 1). The sequences then have been saved in one file using FASTA format.

The used nucleotides sequences

Alignment has been done using ClustalX 2.1 [10]. The non-identical regions have been removed as well as the so diverse sequences. Only one diverse sequence has left for demonstration. After removing the odd sequences, the total partial sequences become as seventeen.

Nucleotides translation

The sequences constituent of nucleotides (for each one) have been translated to amino acids using Blastx (search protein database

***Corresponding author:** Amro Abd Al Fattah Amara, Protein Research Department, Genetic Engineering and Biotechnology Research Institute, City for Scientific Research and Technological Applications, Alexandria, Universities and Research Center District, New Borg EL-Arab, Egypt, Tel: 203-4593422; Fax: 203 4593497; E-mail: amroamara@hotmail.com/amroamara@web.de

Received February 26, 2015; Accepted April 15, 2015; Published April 20, 2015

Citation: Amara AAF (2015) The Need for Early Detection of the Prototype Mutants: Sickle Cell Anemia as a Case Study. J Proteomics Bioinform S8: 006. doi:10.4172/jpb.S8-006

Copyright: © 2015 Amara AAF. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Number	DNA sequence title
1	>gi 49168543 emb CR536530.1 Homo sapiens full open reading frame cDNA clone RZPDo834D0222D for gene HBB, hemoglobin, beta; complete cds, incl. stop codon
2	>gi 164697558 dbj AK311825.1 Homo sapiens cDNA, FLJ92086, Homo sapiens hemoglobin, beta (HBB), mRNA
3	>gi 13937928 gb BC007075.1 Homo sapiens hemoglobin, beta, mRNA (cDNA clone MGC:14540
4	>gi 49456780 emb CR541913.1 Homo sapiens full open reading frame cDNA clone RZPDo834E0633D for gene HBB, hemoglobin, beta; complete cds, without stopcodon
5	>gi 28302128 ref NM_000518.4 Homo sapiens hemoglobin, beta (HBB), mRNA
6	>gi 23268448 gb AY136510.1 Homo sapiens hemoglobin beta chain variant Hb S-Wake (HBB) mRNA, complete cds
7	>gi 13549111 gb AF349114.1 AF349114 Homo sapiens beta globin chain variant (HBB) mRNA, complete cds
8	>gi 29436 emb V00497.1 Human messenger RNA for beta-globin
9	>gi 6003533 gb AF181989.1 AF181989 Homo sapiens hemoglobin beta subunit variant (HBB) mRNA, complete cds
10	>gi 40886940 gb AY509193.1 Homo sapiens hemoglobin beta mRNA, complete cds
11	>gi 187940240 gb EU694432.1 Homo sapiens hemoglobin beta chain variant (HBB) mRNA, HBB-Dothan allele, complete cds
12	>gi 29445 emb V00500.1 Human messenger RNA for beta-globin
13	>gi 183944 gb M25113.1 HUMHEMOB Human sickle beta-hemoglobin mRNA
14	>gi 6003531 gb AF181832.1 AF181832 Homo sapiens hemoglobin beta subunit variant (HBB) mRNA, partial cds
15	>gi 47124545 gb BC070282.1 Homo sapiens hemoglobin, delta, mRNA (cDNA clone MGC:88275 IMAGE:30418964), complete cds
16	>gi 46854767 gb BC069307.1 Homo sapiens hemoglobin, delta, mRNA (cDNA clone MGC:96894 IMAGE:7262103), complete cds
17	>gi 193244962 gb EU760960.1 Homo sapiens isolate TAL57 beta globin gene, partial cds
18	>gi 193244958 gb EU760958.1 Homo sapiens isolate TAL55 beta globin gene, partial cds
19	>gi 193244956 gb EU760957.1 Homo sapiens isolate TAL54 beta globin gene, partial cds
20	>gi 193244954 gb EU760956.1 Homo sapiens isolate TAL52 beta globin gene, partial cds
21	>gi 193244952 gb EU760955.1 Homo sapiens isolate TAL51 beta globin gene, partial cds
22	>gi 193244950 gb EU760954.1 Homo sapiens isolate TAL50 beta globin gene, partial cds
23	>gi 193244946 gb EU760952.1 Homo sapiens isolate TAL48 beta globin gene, partial cds
24	>gi 193244942 gb EU760950.1 Homo sapiens isolate TAL45 truncated beta globin gene, complete cds
25	>gi 193244940 gb EU760949.1 Homo sapiens isolate TAL44 beta globin gene, partial cds
26	>gi 193244938 gb EU760948.1 Homo sapiens isolate TAL42 beta globin gene, partial cds

Table 1: DNA sequences used in this study.

using a translated nucleotide query) and have been rechecked using the translation option in Bioedit software version 7.2.5 (Frame 3) [11]. The obtained translated sequences then collected and putted in one file as FASTA format and saved. The final used DNA and Protein partial sequences used in this study are represented : (1) gi-47124545, (2) gi-46854767, (3) gi-13549111, (4) gi-6003533, (5) gi-40886940, (6) gi-49456780, (7) gi-29445, (8) gi-183944, (9) gi-29436, (10) gi-23268448, (11) gi-28302128, (12) gi-164697558, (13) gi-49168543, (14) gi-187940240, (15) gi-13937928, (16) gi-6003531 and (17) gi-193244942.

Identity calculation

Antheprot v. 6.3.14 software has been used to determine the % of similarity upon alignment using the Clustal W option in the software [12].

Identifying the prototype mutants

Manual counting for the number of the different nucleotides and amino acids either in each block as in Figures 1 and 2 or for each sequence for both of the DNA and the protein alignments was performed.

Phylogenetic trees

Mega 6 software has been used to generate phylogenetic trees for each of the alignment of the DNA and the protein sequences using Maximum Parsimony method [13,14]. The Mega 6 trees for both of the DNA and protein sequences have been merged in one file using the software SplitTree 4 (version 4.13.1.) [15]. The comparison between the two trees has been done using the software Dendroscope (Version 3.2.10.) [16].

Results and Discussion

The biological system is sensitive for the chemical structure. Enzymes could be so specifics. Other protein forms could be also very sensitive. The red blood cell structure changes from ring to sickle shape without the oxygen. This cause blood to clots and deprives vital organs from their supply of blood, resulting in pain, intermittent illness, and often, a shortened life span. Single difference between normal and sickle cell hemoglobin is that in each β - chain, one glutamic acid was replaced by one valine. Valine, unlike glutamic acid, contains a nonpolar group. The result is a hydrophobic “sticky” region that can interact with hydrophobic region on neighboring molecules, producing clumping. A slight change in the β -globin 3D structure will induce a change in its configuration when it interacts with its neighboring subunits of the hemoglobin [17]. The DNA and protein alignments were represented as in Figure 1 and 2. Both of the DNA alignment and the protein alignment similarity have been calculated using Antheport 6.3.14 software. The variation in the DNA sequences show that it is more than that in the protein. However, the % of similarity between the protein sequences is higher from that within the DNA sequences. Where the similarity in the DNA sequences was 75.585% while in the protein sequences was 69.388%. That is because on calculating the % of the similarity; the number of the nucleotides is a critical factor where it exceeds the number of the amino acids and the percentage between them is 3:1. That causes the error in representing the differences between the similarities of both. For that, the number of the different nucleotides as well as the number of the different amino acids have been calculated manually. The data have been summarized in Table 2. The data represented the differences between blocks and the complete sequence(s) of the β -globin as in Table 2. For better comparison

compare both of the trees. The different trees are shown in Figures 3-5.

For the DNA sequences the differences was inferred using the Maximum Parsimony method. Tree 1 (Figure 3) out of 10 most parsimonious trees (length=72) is shown. The consistency index is 1.000000 (1.000000), the retention index is 1.000000 (1.000000), and the composite index is 1.000000 (1.000000) for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm [13] with search level 0 in which the initial trees were obtained by the random addition of sequences (10 replicates). The tree is drawn to scale, with branch lengths calculated using the average pathway method [13] and are in the units of the number of changes over the whole sequence. The analysis involved 17 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 295 positions in the final dataset. The differences analyses were conducted in MEGA 6 [14].

The differences were inferred using the Maximum Parsimony (MP) method. Tree 2 (Figure 4) out of 10 most parsimonious trees

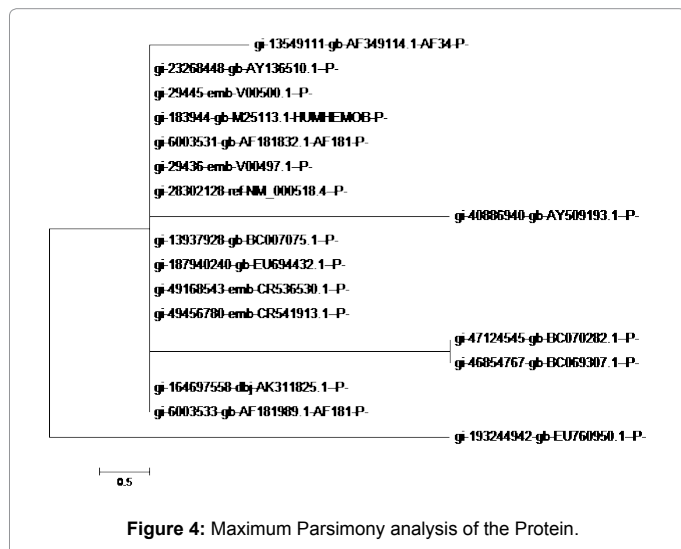


Figure 4: Maximum Parsimony analysis of the Protein.

(length = 12) is shown. The consistency index is 1.000000 (1.000000), the retention index is 1.000000 (1.000000), and the composite index is 1.000000 (1.000000) for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm [13] with search level 0 in which the initial trees were obtained by the random addition of sequences (10 replicates). The tree is drawn to scale, with branch lengths calculated using the average pathway method [13] and are in the units of the number of changes over the whole sequence. The analysis involved 17 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 80 positions in the final dataset. The differences analyses were conducted in MEGA6 [14]. Phylogenetic trees comparison for both of tree 1 and 2 has been generated as in Figure 5. The comparison show that DNA and Protein alignments some how different.

Recently, Amara (2014) has represented the hypothesis that the DNA is the subject of mutant change due to the various stresses and factors such as mutation, reproduction, adaptation, epigenetic should given more concern [18]. Yes, the protein does the function, but it might exist a prototype mutant (incomplete mutant) which is ready to be a true mutant not detected yet. And will not be detected under the protein level or even by translating the nucleotides to protein. Each nucleotide might be critical in any investigation concerning genetic disease so it is recommended to use nucleotides sequence for searching the distribution of any genetic disease in any population. This study highlight that such change is not supporting Darwin theory about the evolution, even it against it. Existence of two alleles could dilute many genetic disorders if one is correct. For, that the marriage outside the family, safe a lot and protect from the degenerative diseases as well as many other genetic related illness. In contrast, marriage from the same genetic pool increases the genetic disorder probability. Versus to Darwin hypothesis, Alleles differences enable variations (color, length, etc.) but disable species alteration. It becomes clear that Darwin has clearly succeeded to discover the role of the genetic material in the variation of the offspring phenotypes. However, he has failed in relating that to the species alteration, which might not existed at all. The hemoglobin molecule has a Quaternary structure that consists of four polypeptide chains, known as globin. While most genes are existed in pair each in one chromosome, it enables variation, as well as enable

Sequence No.	Protein Sequences				DNA sequences				Existence of prototype mutants
	Block 1	Block 2	Block 3	Total	Block 1	Block 2	Block 3	Total	
1	-	2	3	5	4 (3)	8 (?)	8 (?)	20	+
2	-	2	3	5	4 (3)	8 (?)	8 (?)	20	+
3	-	-	2	2	-	-	2 (0)	2	-
4	-	-	1	1	-	-	1 (0)	1	-
5	2	1	-	3	2 (0)	-	-	2	-
6	-	-	-	-	-	-	-	-	-
7	-	-	-	-	1 (1)	-	1 (1)	2	+
8	-	-	-	-	1 (1)	-	1 (1)	2	+
9	-	-	-	-	1 (1)	-	-	1	+
10	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	1 (1)	-	1	+
14	-	-	-	-	-	-	-	-	-
15	-	-	-	-	1 (1)	-	-	1	+
16	-	-	-	-	-	-	1 (0)	1	-
17	-	-	-	-	-	-	40 (?)	48	+

Table 2: Manual counting for nucleotides and amino acids, which show differences from the alignments.

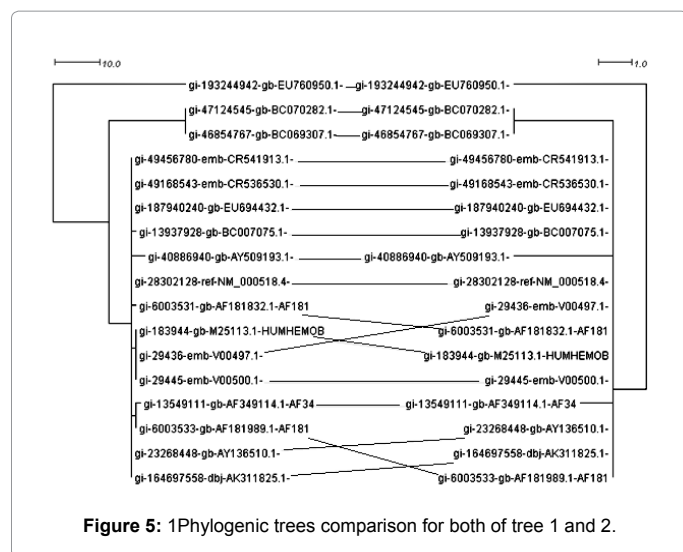


Figure 5: 1Phylogenetic trees comparison for both of tree 1 and 2.

saving the host if one gene is defect. Even so if the two essential genes are defecting, that cause fatal disease such as the degenerative disease and put a big question mark about our understanding to the evolution. Genes constituents have narrow range of change if compared by their mobility within the same species, and partially with the sexually reproduced (have x and y chromosome) organisms including the human. Narrow range of change enables variation but major change is restricted, that safe the whole species feature. It has been expected that the regions where Sickle cell anemia become as an endemic disease that the healthy individuals must be subjected to DNA analysis before marriage. For better investigation of any pro-mutant, individuals which expected to be in risk, such those in endemic area must subjected to examination using complete sequencing for the globin genes. While the scientific research is aiming to solve problems, here this study introduces some steps should be followed by the government where Sickle disease is existed:

1. Marriage from relatives should be avoided.
2. Marriage from Foreigner should be encouraging.
3. Balanced foods contain antioxidants must be used.
4. Natural Edible plants, which proved traditionally to be effective with the Sickle cell anemia patient should be investigated and should be available in the market.
5. Control Sickle cell disease selective agent, particularly the malaria, to increase the Genetic pool where the correct genes should be higher than the defect ones. In such case the defect ones will be disappeared by correct marriage as a result of the time factor (more than one successful generations).
6. More sophisticated determination roles are in need to determine both of the Sickle anemia patients or those are more able to acquire the genetic illness or in better word to transfer it to their generations. Complete sequence for $\beta 1$, $\beta 2$, $\alpha 1$ and $\alpha 2$ globins are in need; particularly in the endemic regions.

Conclusion

This study concerning with Sickle cell anemia disease. Different bioinformatics software have been used to investigate the various types of β globin under both of the DNA and the protein level. The

sequences, which have been used in this study have been reduced aiming to investigate the differences in sequences have nearly no gaps. The DNA and the protein have been compared using phylogenetic trees. DNA sequence is more sensitive to detect the disease and complete globins genes sequencing strategy must be followed.

References

1. Thompson MW, McInnes RR, Willard HF (1991) Chapter 11 "The hemoglobinopathies models of molecular disease" In "Genetics in medicine". USA: W. B. Saunders Company. Harcourt Brace Jovanovich, Inc.
2. McKusick VA (1990) Mendelian inheritance in man (9th edn) Johns Hopkins University Press, Baltimore.
3. Weatherall DJ, Clegg JB, Higgs DR, Wood WG (1989) The hemoglobinopathies. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic basis of inherited disease, (6th edn), McGraw-Hill, New York.
4. Aguilera G, Bielawski JP, Yang Z (2006) Evolutionary rate variation among vertebrate beta globin genes: implications for dating gene family duplication events. *Gene* 380: 21-29.
5. Kumar R, Sagar C, Sharma D, Kishor P (2015) β^2 -globin genes: mutation hot-spots in the global thalassemia belt. *Hemoglobin* 39: 1-8.
6. Livingstone FB (1989) Simulation of the diffusion of the beta-globin variants in the Old World. *Hum Biol* 61: 297-309.
7. Ciminelli BM, Pompei F, Relucanti M, Lum JK, Simporé J, et al. (2002) Confirmation of the potential usefulness of two human beta globin pseudogene markers to estimate gene flows to and from sub-Saharan Africans. *Hum Biol* 74: 243-252.
8. Shohat M, Bu X, Shohat T, Fischel-Ghodsian N, Magal N, Nakamura Y, et al. (1992) The gene for familial Mediterranean fever in both Armenians and non-Ashkenazi Jews is linked to the alpha-globin complex on 16p: evidence for locus homogeneity. *Am J Hum Genet* 51: 1349-1354.
9. Madden T (2002) The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 16.
10. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882.
11. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95-98.
12. Deléage G, Clerc FF, Roux B, Gautheron DC (1988) ANTHEPROT: a package for protein sequence analysis using a microcomputer. *Comput Appl Biosci* 4: 351-356.
13. Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
14. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725-2729.
15. Huson DH, Scornavacca C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* 61: 1061-1067.
16. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23: 254-267.
17. Efimov AV (1979) Packing of alpha-helices in globular proteins. Layer-structure of globin hydrophobic cores. *J Mol Biol* 134: 23-40.
18. Amara AA (2014) Proteases and their adaptation. *IIOABJ* 5 (4):12-29.