

# Bioinformatic Analysis of Insertion, Deletion, and Substitution Mutations in the Human $\beta$ -Globin Gene

Edgar Miguel Bustamante Cárdenas

Faculty of Pharmaceutical, Biochemical and Biotechnological Sciences, Catholic University of Santa María, Arequipa, Peru  
Email: edgar.bustamante@ucsm.edu.pe

**How to cite this paper:** Cárdenas, E.M.B. (2021) Bioinformatic Analysis of Insertion, Deletion, and Substitution Mutations in the Human  $\beta$ -Globin Gene. *Computational Molecular Bioscience*, 11, 1-17.  
<https://doi.org/10.4236/cmb.2021.111001>

**Received:** March 1, 2021

**Accepted:** March 28, 2021

**Published:** March 31, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Beta-thalassemia is delineated as a bunch of heritable blood disorders characterized by abnormalities within the synthesis of beta chains of hemoglobin. Most defects in hemoglobin arise directly from mutations in the structure of the  $\beta$ -globin sequence (HBB). Recent advances in computational tools have made it possible to study the relationship between genotype and phenotype in many diseases, including  $\beta$ -thalassemia. Due to the high prevalence of  $\beta$ -thalassemia, these analyses have helped to know the molecular basis in a better way. In this direction, a related database, called HbVar, was developed in 2001 by a collective academic effort to provide quality and up-to-date information on genomic variations that lead to hemoglobinopathies and thalassemia. Within present study, the aim was to investigate insertion, substitution, and deletion mutations in HBB collected from HbVar and their effects using the in-silico approach. In this comparative analysis, three sequences, one wild-type and two mutated HBB sequences were studied in the South American region. It was determined that the Brazilian nd-HPFH mutated sequence presents the formation of a beta chain in the protein structure, not being able to align or overlap in the wild-type, which results in the alteration of the function of the protein, and therefore the development of Thalassemia disease.

## Keywords

Hemoglobin, Mutation, Thalassemia

## 1. Introduction

Hemoglobinopathies are sequences disorders caused by variations of a gene within the  $\alpha$ - and  $\beta$ -type human globin gene clusters. These are the foremost

common transmissible disorders in humans, with nearly 7% of the human population acting as carriers for mutations in the globin genes. The substitutions of individual nucleotides in the coding or regulative regions of those genes will cause varied degrees of defects in their expression [1]. The HBB gene belongs to the group of  $\beta$ -globin genes that encodes the  $\beta$ -globin polypeptide. It is set on the short arm of chromosome eleven and contains two introns and three exons [2].

Molecular defects in human HBB will cause structural defects that cause abnormalities in hemoglobin, appreciate HbS, HbC and HbD, or it can cause the absence or reduced synthesis of  $\beta$ -globin chains that cause  $\beta$ -thalassemia [3]. Mutations in HBB will involve the substitution, deletion, or insertion of 1 or additional nucleotides at intervals in the gene or its flanking regions, leading to anemia and low red blood cell production [4].  $\beta$ -thalassemia is transmissible in an autosomal-recessive manner and its clinical manifestation may be divided into thalassemia major, intermedia and thalassemia minor [5].

Some mutations in HBB result in complete inactivation of the gene resulting in the absence of  $\beta$ -globin ( $\beta^0$ ) chains, that successively results in the most severe type of thalassemia. Different mutations permit the assembly of  $\beta$ -globin chains in varied proportions resulting in  $\beta$ -thalassemia. This case is most found within the Middle East, Central Asia, Mediterranean countries, India, and southern China, and in some parts of Africa and South America [6]. It is one in every of the most common genetic disorders caused by purpose mutations that cause variable phenotypical effects. These phenotypical severities will arise from defects in the transcription, ribonucleic acid processing, or translation of the HBB gene [7]. The most common mutations in most populated countries include IVSI-110 (G > A), IVSI-1 (G > A), IVSI-6 (T > C), IVSII-1 (G > A), IVSI-5 (G > C), sequence 5 (-CT) and codon 39 (C > T) [8].

Due to the high prevalence of variable thalassemia phenotypes and the remarkable heterogeneity of their molecular defects, numerous strategies were went to investigate the molecular mechanisms of this disease. Recent advances in procedure tools have created in-silico analysis one in every of the methods of option to investigate the links between phenotypical, genomic, and resultant characteristics in thalassemia. The term “in silico” is a modern word generally used to refer to computer experimentation [9]. The first published examples of the word mentioned a concise and cogent depiction of the potential of computational tools in computer chemistry, biology, and pharmacology. It is a real discovery aid when analyzing biological functions, furthermore to underline its importance as a complement to *in vivo* and *in vitro* experimentation [10] [11].

HbVar is the oldest and most appreciated information of Hb variants and thalassemia mutations established in 2001 [12]. It is a locus-specific information base, which was developed as a combined educational effort to keep up a register of hemoglobin variants, new data entries, updates, and corrections. It provides up-to-date and high-quality data on genomic variations, associated phenotypical and medical specialty effects, pathology, frequency of various mutations, ethnic

prevalence [13]. HbVar has become a primary resource for the research community acting on macromolecule proteins and for clinicians coping with patients with hemoglobinopathies, to assist them building the right diagnoses. The aim of this study is to research the results on the structure and performance of the  $\beta$ -globin protein using the in-silico method.

## 2. Material and Methods

### 2.1. Sequence Retrieval

Sequences of uncharacterized HBB gene having insertion substitution and deletion mutations are collected from HbVar (<http://globin.cse.psu.edu/globin/hbvar/>). In HbVar total 1823 variants of HBB are present. HBB variants of the South America region were retrieved from 1823 sequences. There were only 4 variants out of 1823 variants, which were from South America region, with title Venezuelan nd-HPFH, Hb Ecuador, -195 (C- > G) <sup>A</sup>gamma; the Brazilian nd-HPFH [10] and Hb Chile [14] Sequence of Venezuelan nd-HPFH, Hb Ecuador was not available while DNA sequence of -195 (C- > G) <sup>A</sup>gamma; the Brazilian nd-HPFH [11] and Hb Chile [15] was retrieved from NCBI (<https://www.ncbi.nlm.nih.gov/>). Wild type sequence of HBB gene (gene ID 3043) was also retrieved from NCBI (<https://www.ncbi.nlm.nih.gov/gene/?term=3043>).

### 2.2. Multiple Sequence Alignment (MSA)

To align the sequences, multiple sequence alignment of all 3 sequences, -195 (C- > G) <sup>A</sup>gamma; the Brazilian nd-HPFH, wild type HBB and Hb Chile was done from Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Clustal Omega is an online tool for multiple sequence alignment from EMBL-EBI web server (<https://www.ebi.ac.uk/>).

Multiple sequence alignment was done using default parameters.

### 2.3. Translation

For protein sequence of aligned sequences of -195 (C- > G) <sup>A</sup>gamma; the Brazilian nd-HPFH, wild type HBB and Hb Chile variants, ExPASy translate was used. ExPASy translate is an online server for translation. It inputs DNA sequences and provides protein sequence as output. Protein sequence of all 3 nucleotide sequences is generated from ExPASy translate tool by using default parameters (<https://web.expasy.org/translate/>).

### 2.4. Homology Modeling

The homology model of wild type HBB and mutant type (Hb Chile and -195 (C- > G) <sup>A</sup>gamma; the Brazilian nd-HPFH) was performed to compare 3D structure of proteins. Homology modeling was done using SWISS-MODELER [16]. SWISS-MODEL is a fully automated protein structure homology modeling server accessible via ExPASy web server, or from the program DeepView. The purpose of this server is to make protein modeling accessible to all life science

researchers worldwide (<https://swissmodel.expasy.org/>). Template was search for each sequence before building the model. Best template was chosen on the base of sequence similarity. Analysis was done using default parameters.

## 2.5. Model Evaluation

All 3 Models build from SWISS-MODELER were further evaluated from ERRAT to check quality factors of structures [17] ERRAT is an online server to analyze the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures (<https://servicesn.mbi.ucla.edu/ERRAT/>).

Furthermore, these structures were evaluated from molprobitry (<http://molprobitry.biochem.duke.edu/>) [17] [18]. Molprobitry is a general-purpose web server offering quality validation for 3D structures of proteins, nucleic acids, and complexes [19].

## 2.6. Molecular Dynamics Simulation

**GROMACS** simulation package (**GROMACS 2020.4**) was used to perform molecular dynamics simulations. Simulations of wild type and mutant proteins (HPFH and hb chile) were carried out for 30 ns in water using GROMOS96 53A6 forcefield; trajectory and energy files were written every 2 ps.

These systems were solvated in a periodically truncated octahedral box, containing 8805, 4611, 3502 SPC [20] water molecules for wild type, HPFH, and hb chile proteins, respectively. The protein was centered in the simulation box within minimum distance to the box edge of 1 nm to efficiently satisfy the minimum image convention. The protonation states of protonatable groups were set to neutral pH. Chloride ions (2 in wild type and 2 in hb chile) were added to neutralize the overall system. However, no ions were added in HPFH protein system as it was already neutralized.

Minimization of all three systems were carried out for 5000 steps using Steepest Descent Method and the convergence was achieved within the maximum force  $< 1000$  ( $\text{KJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-1}$ ), to remove any steric clashes followed by system equilibration at NPT ensemble for 1000 ps at 300K.

Production runs for all simulations were carried out at a constant temperature of 300 K and a pressure of 1 atm (NPT) using weak coupling velocity-rescaling (modified Berendsen thermostat) and Parrinello-Rahman algorithms, respectively.

Relaxation times were set to  $\tau_T = 0.1$  ps and  $\tau_P = 1.0$  ps and an estimated isothermal compressibility of water  $4.5\text{e-}5$  ( $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-3}$ )<sup>-1</sup> was used (van Gunsteren *et al.* 1996). All bond lengths involving hydrogen atom were kept rigid at ideal bond lengths using the Linear Constraint Solver (lincs) algorithm, allowing for a time step of 2 fs. Verlet scheme was used for the calculation of non-bonded interactions. Periodic Boundary Conditions (PBC) were used in all x, y, z directions. Interactions within a short-range cutoff of 1.0 nm were calcu-

lated in each time step. Particle Mesh Ewald (PME) was used to calculate the electrostatic interactions and forces to account for a homogeneous medium outside the long-range cutoff.

### 3. Results

#### 3.1. Mutational Variants

4 variants of the HBB South America region have different mutations at different locations (Table 1). Information of different HBB mutations in South America region is as following.

#### 3.2. Exploration of MSA between HB Wild and Mutated Proteins

Multiple sequence alignment of sequences was done using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Alignment showed that -195 (C- > G) <sup>A</sup>gamma; the Brazilian nd-HPFH have 587 nucleotides, wild type contains 1608 and Hb Chile contain 628 nucleotides (Figure 1). Results showed that all 3 sequences are different from each other.

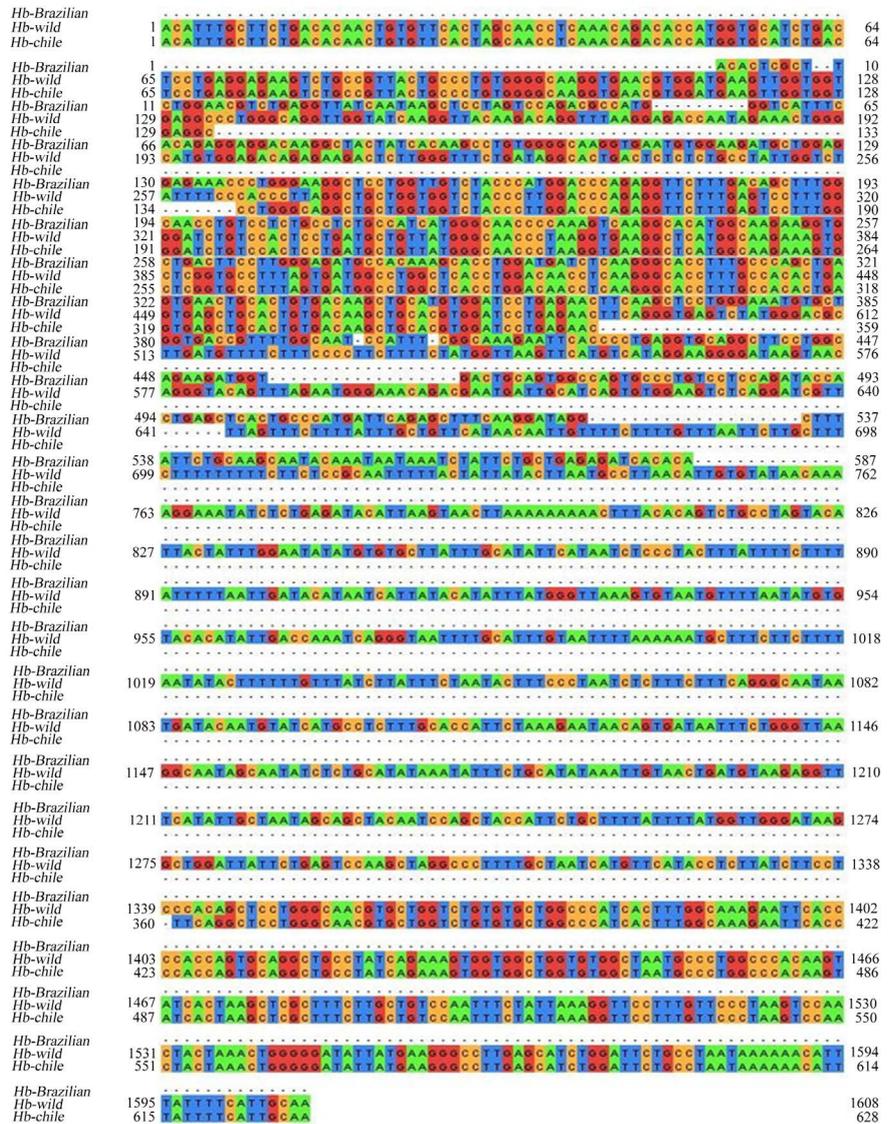
#### 3.3. Translation

Protein sequence of alignment was obtained from ExPASy Translate server (<https://web.expasy.org/translate/>). Protein sequence shows that protein sequence of -195 (C- > G) <sup>A</sup>gamma; the Brazilian nd-HPFH contain 195 amino acid residues, wild type protein contains 536 amino acid residues while Hb Chile protein contain 209 amino acid residues. Protein sequence shows that all 3 variants have different protein sequences (Figure 2).

**Table 1.** HBB Variants.

Name	HbVar ID	HGVS name	Category	Mutation location	Occurrence
Venezuelan nd-HPFH	2513	HBG1:c.-264C > T	Thalassemias	not within known transcription unit hg38: chr11 5,250,068	Ethnic background Black
Hb Ecuador	2586	HBA1:c.416C > G	Hb variants	exon hg38: chr16 177,398	found in a woman from Ecuador
-195 (C- > G) <sup>A</sup> gamma; the Brazilian nd-HPFH	1011	HBG1:c.-248C > G	Thalassemias	not within known transcription unit hg38: chr11 5,250,052	Ethnic background Brazilian
Hb Chile	287	HBB:c.85C > A	Hb variants	exon hg38: chr11 5,226,937	Ethnic background American Indian

Hb Ecuador and Hb Chile have mutations in the exon part which alter protein sequence. Hb Ecuador variant changes serine into cysteine while Hb Chile variant changes leucine into methionine.



**Figure 1.** Multiple Sequence Alignment showed Hb wild type and Hb Chile have more conserved regions, while Hb wild type and -195 (C -> G) Agamma; the Brazilian nd-HPFH have less conserved regions.

-195 (C -> G) Agamma; the Brazilian nd-HPFH

TLASGTSEVINKLLVQTPWVISQRRLTLLSQCAGAR-MWKMLEEKPEWEGSWLSTHGPRGSLTALATCPLEPLPWSWATPKSRHMARRC-LPWEVMPQSTWMSRAPLPS  
-VNQCTVSCWILRTSSWEMCW-PFWQSI SARNSPLRCLRPLGRWR-LQWVPCPPDDTELTADHSELSRIGFILQAIQINLFC-EIT

Hb Chile

TFASDITVFTSNLKQTPWCI-LLRSLPLPCGAR-TWMKLVVRPWAGCWWSLTLGPRGSLPLGICPLMLLWATLR-RLMARKCSVPLVMWMLTWTSRAPLPH-  
VSCVTVSCWILRTSSWEMCW-PFWQSI SARNSPLRCLRPLGRWR-LQWVPCPPDDTELTADHSELSRIGFILQAIQINLFC-EIT

Wild Type

TFASDITVFTSNLKQTPWCI-LLRSLPLPCGAR-TWMKLVVRPWAGWYQYKTLRRPIETHGVETKLGFLIGDLSLCLLYVFPTRLLVVVYVPTQRFESF  
GDLSTPDVAVMGNPKVKAHKKVLAGFSDGLAHLNLRGTFATLSELHCDKLVDPENFRVSLWDA-CFLSPSFLWLSCHRRKGISNRVQFRMGNRMIASVWKSOD  
RFSFFYLLFETIVFFCLLILAFFFLRNFFYYT-CLNIYVNRKYL-DLNLKKNFTQSA-YITWNICVLICIFISLLYFLFLDIT-SLYIFMG-SVMF-YV  
YTY-PNQGNAFVILKNFAFF-YTEFLFLLFLLSLSFFQGNNDI-MYHASLHHSRE-Q--FLG-GNSNISAYKYFCI-IVTDVRFILLIATLQLPFCFYFMGI  
RLDYSSEKLGPFANHVHTSYLPPTAPGQRAQLCAGPSLWQRHIFPTASGCLSESSEGGWC-CGPGQVLSLSSCCPISIKGSFVP-VQLLNWGL-RALSILWLPNK  
HLFSLQ

**Figure 2.** Protein translation showed all 3 sequences are different from each other due to mutation.

### 3.4. Homology Modeling

Homology modeling was done using SWISS-MODELER. Template search for each sequence was done before building the model. Model was built against best template which was chosen based on sequence similarity. To generate 3D structure of sequences, template having ID 1dxu.1 [21] was chosen against HBB wild type sequence having sequence identity 73.64%, template having ID 2lp6.1 [22] was chosen against -195 (C- > G)  $\gamma$ ; the Brazilian nd-HPFH having sequence identity 28.57% and template having ID 2w7k.1 [23] was chosen against Hb Chile having sequence identity 6.25%.

### 3.5. Molecular Dynamics Simulations of $\gamma$ the Brazilian HPFH

#### RMSD:

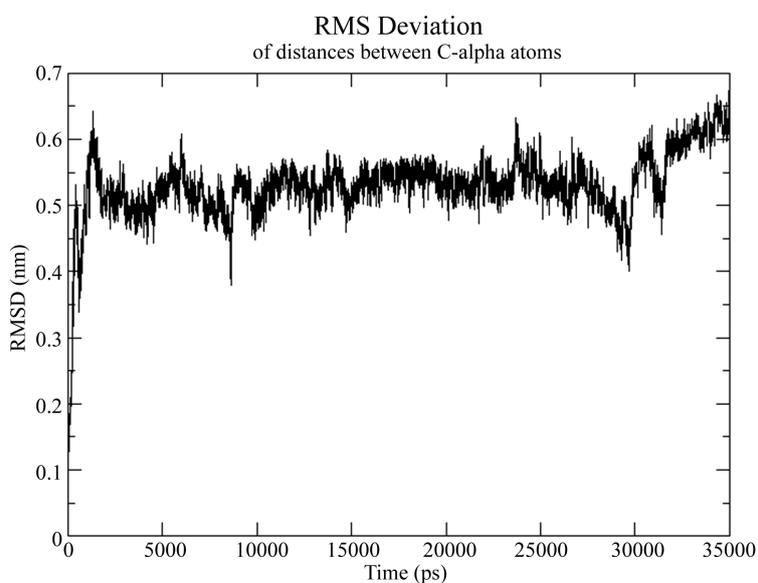
A total of 35 ns simulation was performed and for average analysis, first 5 ns (5000 ps) trajectory was discarded. The mean RMSD is 0.54 nm with standard deviation of 0.04 nm. The RMSD plot indicates that in the first 2000 ps the system shows deviation till  $\sim 0.5$  nm which remains stable until 27 ns. Then again strong fluctuation is observed from 30 to 35 ns (**Figure 3**).

#### RMSF:

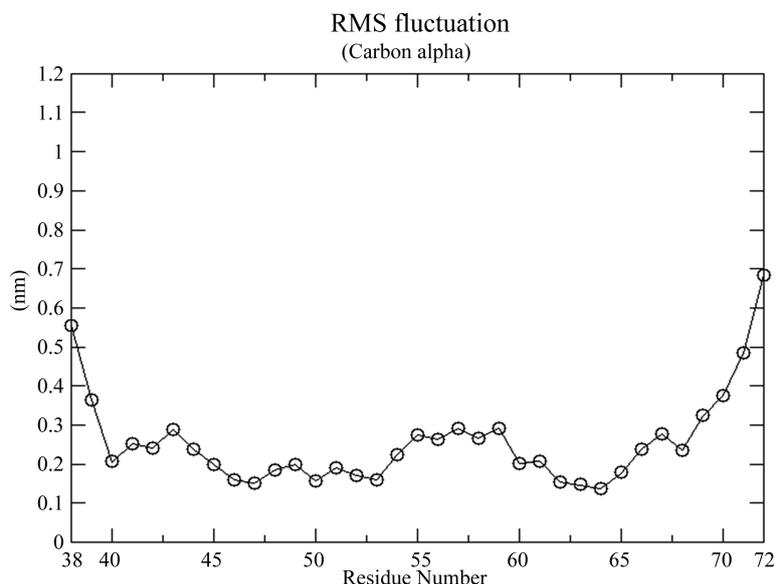
A total of 35 ns simulation was performed and for RMSF analysis, first 5 ns (5000 ps) trajectory was discarded. The RMSF plot shows that fluctuation of N and C terminals reaches up to 0.6 nm and 0.7 nm, respectively. However, the rest of residues are below 0.3 nm (**Figure 4**).

#### ROG:

A total of 35 ns simulation was performed and for average analysis, first 5ns (5000 ps) trajectory was discarded. Slight fluctuations have been observed till 30 ns. However, pronounced fluctuations were observed between 30 to 35 ns.



**Figure 3.** Root Mean Square Deviation (RMSD) plot of alpha carbon of protein.



**Figure 4.** Root Mean Square Fluctuation (RMSF) plot of alpha carbon of protein.

The mean radius of gyration is 0.98 nm with a standard deviation of 0.04 nm (Figure 5).

#### Protein Superimposition:

Protein was overlapped at different timescale *i.e.*, 0 ns, 10 ns, 20 ns, 30 ns to analyze the changes in structure and position of protein (Figure 6).

### 3.6. Molecular Dynamics Simulations of Wild Type Hb Protein

#### RMSD:

A total of 35 ns simulation was performed and for average analysis, first 5 ns (5000 ps) trajectory was discarded. RMSD of approx. 6 ns reaches 0.4 nm, and then attains a plateau with some noticeable fluctuations until 30 ns. Visible structural change was observed between 30 to 35 ns where RMSD reaches upto 0.5 nm. This strong deviation is likely to be associated with the C terminal of the protein (as indicated by its higher RMSF value). Mean RMSD is  $0.42 \pm 0.05$  representing the overall stability of the structure (Figure 7).

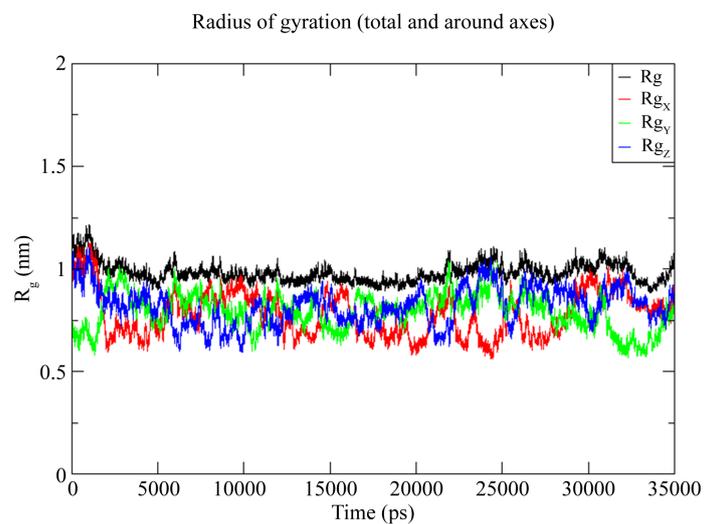
#### RMSF

A total of 35 ns simulation was performed and for RMSF analysis, first 5 ns (5000 ps) trajectory was discarded. RMSF plot shows that N terminal is quite stable. However, the residues ranging from 79 - 83, 101 - 118 and 167 - 173 are showing higher fluctuation with RMSF value greater than 0.3 nm. The C terminal is highly fluctuating and reaches upto 1.1 nm indicating some noticeable change in the structure (Figure 8).

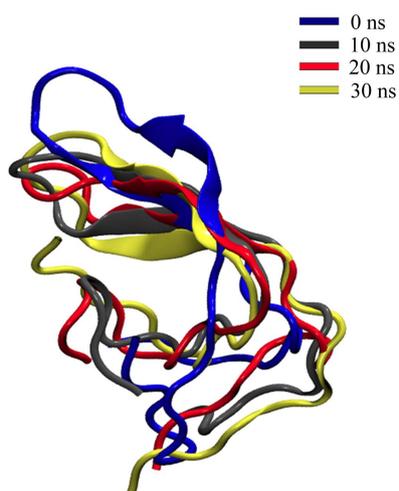
#### ROG

A total of 35 ns simulation was performed and for average analysis, first 5 ns (5000 ps) trajectory was discarded. Overall ROG is quite stable with average value of  $1.47 \pm 0.04$  nm (Figure 9).

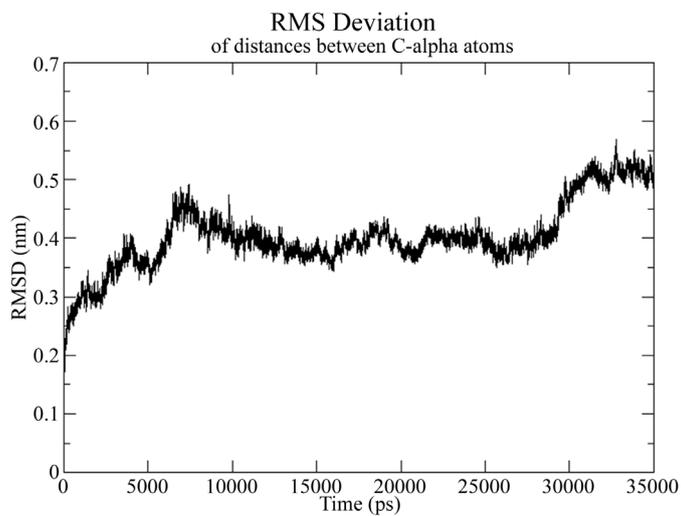
#### Protein Superimposition



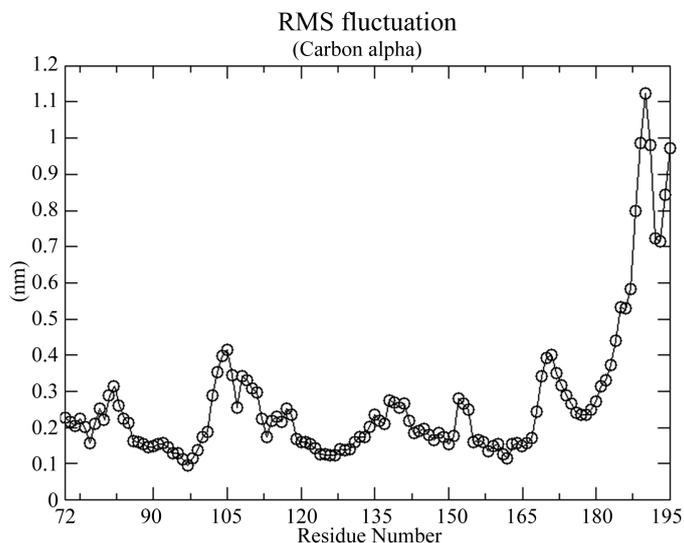
**Figure 5.** Radius of Gyration (ROG) plot of alpha carbon of protein.



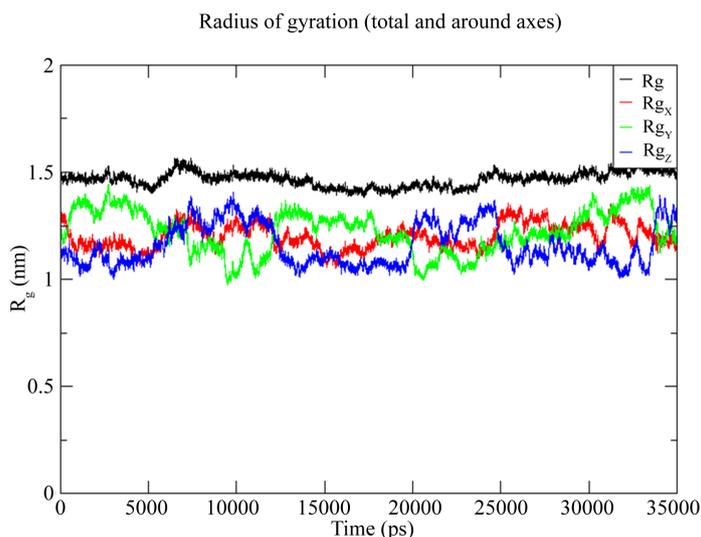
**Figure 6.** Overlapping of protein at different timescale in nanosecond.



**Figure 7.** Root Mean Square Deviation (RMSD) plot of alpha carbons of protein.



**Figure 8.** Root Mean Square Fluctuation (RMSF) plot showing alpha carbons of protein.



**Figure 9.** Radius of Gyration (ROG) plot of alpha carbons of protein.

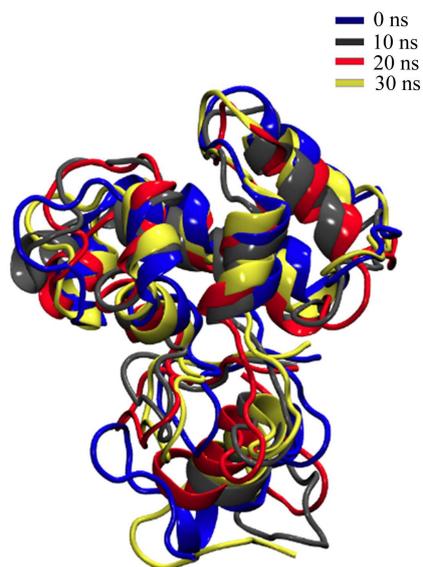
Protein was overlapped at different timescale *i.e.* 0 ns, 10 ns, 20 ns, 30 ns to analyze the changes in structure and position of protein (**Figure 10**).

### 3.7. Molecular Dynamics Simulations of Hb Chile

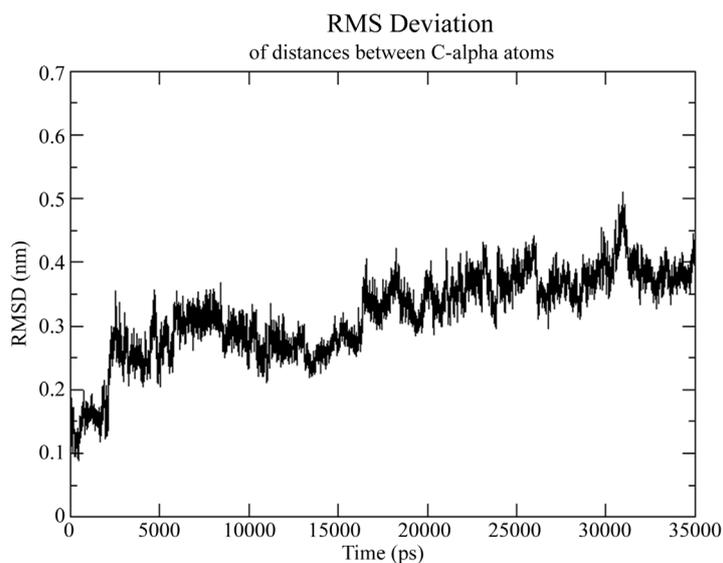
#### RMSD:

A total of 35 ns simulation was performed and for average analysis, first 5 ns (5000 ps) trajectory was discarded. Deviation reaches upto 0.3 nm in the first 2.5 ns, then after 17 ns RMSD is further escalated which remained stable throughout the rest of simulation. Overall, there are some pronounced fluctuations, which reflect structural change. The mean RMSD is 0.33 nm with a standard deviation of 0.05 nm (**Figure 11**).

#### RMSF:



**Figure 10.** Overlapping of protein at different nanosecond.



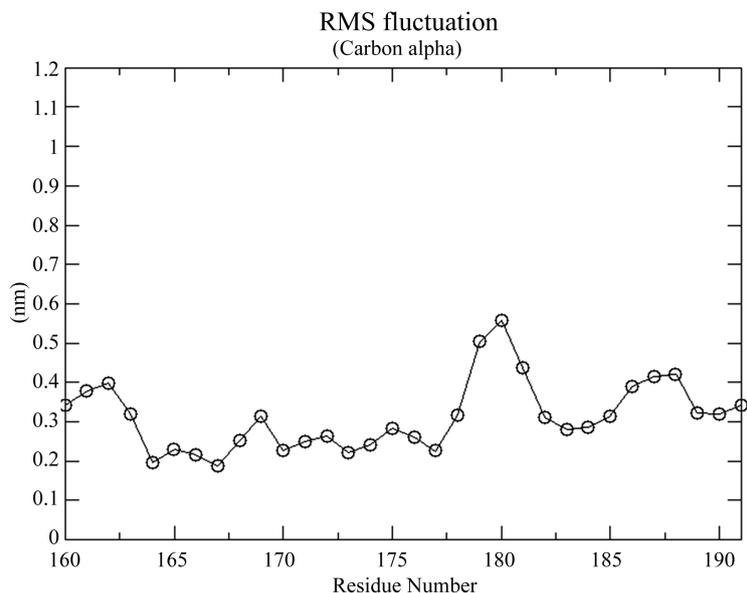
**Figure 11.** Root Mean Square Deviation (RMSD) plot.

A total of 35 ns simulation was performed and for RMSF analysis, first 5 ns (5000 ps) trajectory was discarded. Both N and C terminals are comparatively less fluctuating and remained below 0.4 nm. However, the residues 179 - 181 are highly fluctuating reaching over 0.5 nm (**Figure 12**).

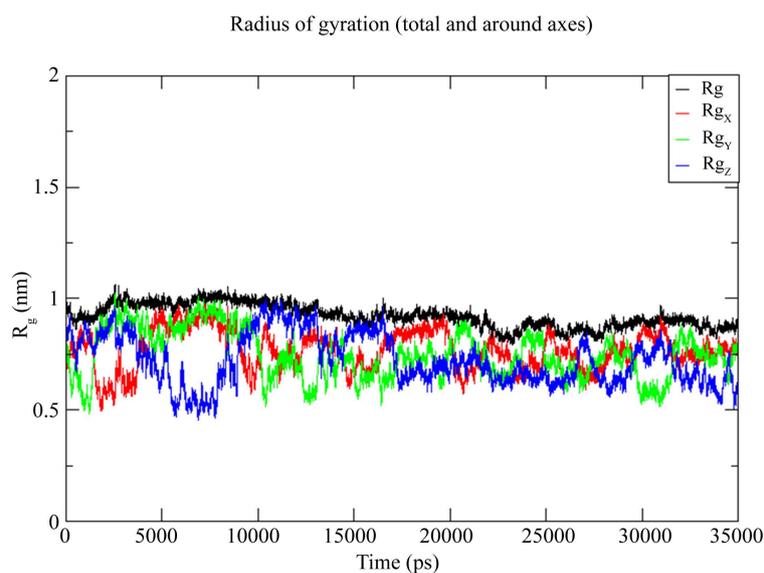
#### **ROG:**

A total of 35 ns simulation was performed and for average analysis, first 5 ns (5000 ps) trajectory was discarded. The plot shows that ROG declined in overall simulation. However, after 22 ns prominent fluctuations were seen, which indicate some structural change. The mean ROG is 0.92 nm with a standard deviation of 0.05 nm (**Figure 13**).

#### **Protein Superimposition:**



**Figure 12.** Root Mean Square Fluctuation (RMSF) plot.



**Figure 13.** Radius of Gyration (ROG) plot.

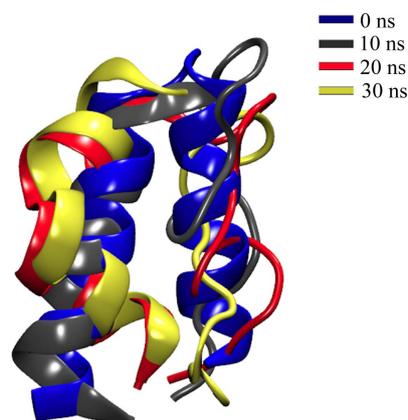
Protein was overlapped at different timescale *i.e.*, 0 ns, 10 ns, 20 ns, 30 ns to analyze the changes in structure and position of protein (Figure 14).

### 3.8. Structure Overlapping of All Proteins

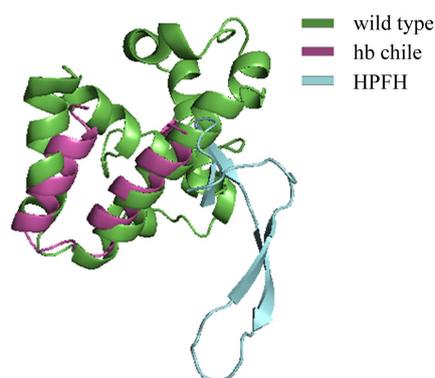
All three protein structures were overlapped using PyMol. As the HPFH has beta chain it did not overlap with other proteins (Figure 15).

## 4. Discussion

Building the relationship between phenotype and genotype in the clinical setting is one of the main goals of traditional research [24]. However, studies on many



**Figure 14.** Overlapping of protein at different nanosecond.



**Figure 15.** Overlapping of all three proteins using PyMol.

mutations are problematic, mainly due to experimental analyzes. In contrast, in silico analysis is faster and easier to perform, produces more results, and costs less, making it more efficient.

This type of analysis is based on alterations in the nucleotide and/or amino acid sequences. And its comparison with the native sequence to correlate the effect of these alterations on the phenotype of the individual [25]. Mutations in the HBB gene, found on chromosome 11 p15.5 [26], are responsible for several serious hemoglobinopathies, such as sickle cell anemia and  $\beta$ -thalassemia. Sickle cell anemia and  $\beta$ -thalassemia can lead to severe anemia and other life-threatening conditions [27].

Although HBB is well characterized, mutations in this gene recorded in the HbVar database are poorly understood and have not been adequately studied before. In this study, HBB variants from the South American region were collected from 1823 sequences.

The in-silico analysis showed that the variant -195 (C- > G) Agamma; the Brazilian nd-HPFH acquired a loop formation in the 3D structure of the protein (Figure 15). The Hb Chile variant, unlike the previous one, did not show the formation of a beta chain, but the molecular dynamics simulation showed evidence of effects caused by the mutation in the flexibility of the protein. The

RMSF analysis showed a high degree of fluctuations in -195 (C- > G) Agamma; the Brazilian nd-HPFH, with that of wild-type compared to Hb Chile (**Figure 4**, **Figure 8** and **Figure 12** respectively). The substitution of new amino acids can probably alter the environment of the protein, resulting in a new set of interactions between amino acids, which in turn, could have flexibility of the affected proteins.

This loss of flexibility can result in loss of protein function. Studies suggest that a change in function could be due to the folding of the protein sequence, this particularly occurs due to the change in physicochemical properties, such as hydrophobicity, charge, and geometry due to the side chain of amino acid residues. To further test the effects of these substitution mutations on protein structure and function, further analysis is required.

Studies on the radius of gyration of C atoms of wild type and mutant proteins represented in the 3D structure, the wild type looks more compact and stable compared to the two variants. These data suggest that substitution in mutations in the HBB protein may be affecting its structure and function as shown in both variants, but more intensive studies are required to fully understand the extent of these effects. We have yet to determine how these mutations affect the flexibility of the HBB protein and whether and to what extent this loss affects the function of the proteins. In vitro studies will further evaluate the functional behavior of mutant proteins.

In Peru, which represents an important ethnic complex in Latin America, knowledge about the current situation of hemoglobinopathies is limited. The high cost of carrying out screening for abnormal hemoglobins in newborns, together with the significant perinatal mortality, contributes to the scarcity of diagnoses of abnormal hemoglobin-derived syndromes that we have, influences the scarcity of diagnoses of abnormal hemoglobin-derived syndromes. The hemoglobin abnormalities described in their homo or heterozygous form are all products of immigration, it could only be demonstrated if the respective studies are carried out and the neonatal screening for abnormal hemoglobins is promoted. Due to genetic crossover, we can observe from the clinical point of view three types of thalassemias: major or homozygous, intermediate, and minor, which is the most frequent hemoglobin abnormality that exists worldwide, nevertheless under-diagnosed in our environment due to the custom that exists to give iron to all anemic patients without really investigating the cause and because this genetic alteration is confused with iron deficiency anemia. Of the larger type, we have not identified a single case, and apparently it does not yet occur in Peru. Within the intermediate and minor we have found several patients, both in their alpha and beta form.

## 5. Conclusions

The structural stability of the proteins was analyzed by DM simulations. The RMSF graph of -195 (C- > G) Agamma, the Brazilian nd-HPFH shows the fluc-

tuation of the N and C terminals reaches up to 0.6 nm and 0.7 nm, respectively. However, the rest of the residues is below 0.3 nm. Wild-type of its RMSF chart shows that the N terminal is quite stable. However, the residues ranging from 79 - 83, 101 - 118 and 167 - 173 show a greater fluctuation with a RMSF value greater than 0.3 nm. The C terminal is highly fluctuating, reaching up to 1.1 nm, indicating a marked change in structure. Hb Chile of its RMSF graph, the first trajectory of 5 ns (5000 ps) was discarded. Both N and C terminals are comparatively less fluctuating and remained below 0.4 nm. However, residues 179 - 181 are highly fluctuating, reaching greater than 0.5 nm.

Due to the study of the three sequences, one wild-type and two mutated sequences of HBB in the South American region, the -195 (C- > G) Agamma, the Brazilian nd-HPFH sequence presents the formation of a beta chain in the structure of the protein, not being able to align or overlap in the Wild-type, which indicates the alteration of the function of the protein, and consequently the development of Thalassemia disease.

### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

### References

- [1] Weatherall, D.J. (2008) Hemoglobinopathies Worldwide: Present and Future. *Current Molecular Medicine*, **8**, 592-599. <https://doi.org/10.2174/156652408786241375>
- [2] Warsy, A.S., ElHazmi, M.A., Al Momin, A.K., AlHazmi, A. and Aleem, A. (2013) Extensive Polymorphisms in Saudi Beta Thalassaemia Patients. *Biosciences Biotechnology Research Asia*, **10**, 127-132.
- [3] Weatherall, D.J. and Clegg, J.B. (2001) The Thalassaemia Syndromes. Blackwell Science, Oxford. <https://doi.org/10.1002/9780470696705>
- [4] Amjad, F., Fatima, T., Fayyaz, T., Khan, M. and Qadeer, M. (2020) Novel Genetic Therapeutic Approaches for Modulating the Severity of  $\beta$ -Thalassemia (Review). *Biomedical Reports*, **13**, 1. <https://doi.org/10.3892/br.2020.1355>
- [5] Galanello, R. and Origa, R. (2010) Beta-Thalassemia. *Orphanet Journal of Rare Diseases*, **5**, 11. <https://doi.org/10.1186/1750-1172-5-11>
- [6] Zahed, L. (2001) The Spectrum of Beta-Thalassemia Mutations in the Arab Populations. *BioMed Research Internationa*, **1**, Article ID: 918781. <https://doi.org/10.1155/S1110724301000298>
- [7] Qari, M.H., Wali, Y., Albagshi, M.H, Alshahrani, M., Alzahrani, A., Alhijji, I.A., *et al.* (2013) Regional Consensus Opinion for the Management of Beta Thalassemia Major in the Arabian Gulf Area. *Orphanet Journal of Rare Diseases*, **8**, Article No. 143. <https://doi.org/10.1186/1750-1172-8-143>
- [8] Giardine, B., van Baal, S., Kaimakis, P., Riemer, C., Miller, W., Samara, M., *et al.* (2007) HbVar Database of Human Hemoglobin Variants and Thalassemia Mutations: 2007 Update. *Human Mutation*, **28**, 206. <https://doi.org/10.1002/humu.9479>
- [9] Ekins, S., Mestres, J. and Testa, B. (2007) *In Silico* Pharmacology for Drug Discovery: Methods for Virtual Ligand Screening and Profiling. *British Journal of Pharmacology*, **152**, 9-20. <https://doi.org/10.1038/sj.bjp.0707305>

- [10] Sieburg, H.B. (1990) Physiological Studies *in Silico*. *Studies in the Sciences of Complexity*, **12**, 321-342.
- [11] Danchin, A., Medigue, C., Gascuel, O., Soldano, H. and Henaut, A. (1991) From Data Banks to Data Bases. *Research in Microbiology*, **142**, 913-916. [https://doi.org/10.1016/0923-2508\(91\)90073-J](https://doi.org/10.1016/0923-2508(91)90073-J)
- [12] Desmet, F.O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M. and Bérout, C. (2009) Human Splicing Finder: An Online Bioinformatics Tool to Predict Splicing Signals. *Nucleic Acids Research*, **37**, Article No. e67. <https://doi.org/10.1093/nar/gkp215>
- [13] Giardine, B., Borg, J., Viennas, E., Pavlidis, C., Moradkhani, K., Joly, P., Bartsakoulia, M., Riemer, C., Miller, W., Tzimas, G., Wajcman, H., Hardison, R.C. and Patrinos, G.P. (2013) Updates of the HbVar Database of Human Hemoglobin Variants and Thalassemia Mutations. *Nucleic Acids Research*, **42**, D1063-D1069. <https://doi.org/10.1093/nar/gkt911>
- [14] Costa, F.F., Zago, M.A., Cheng, G., Nechtman, J.F., Stoming, T.A. and Huisman, T.H. (1990) The Brazilian Type of Nondeletional A Gamma-Fetal Hemoglobin Has a C—G Substitution at Nucleotide-195 of the A Gamma-Globin Gene. *Blood*, **76**, 1896-1897.
- [15] Hojas-Bernal, R., McNab-Martin, P., Fairbanks, V.F., Holmes, M.W., Hoyer, J.D., McCormick, D.J. and Kubik, K.S. (1999) Hb Chile (b28 (B10) Leu → Met): An Unstable Hemoglobin Associated with Chronic Methemoglobinemia and Sulfonamide or Methylene Blue-Induced Hemolytic Anemia. *Hemoglobin*, **23**, 125-134. <https://doi.org/10.3109/03630269908996157>
- [16] Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. (2003) SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucleic Acids Research*, **31**, 3381-3385. <https://doi.org/10.1093/nar/gkg520>
- [17] Colovos, C. and Yeates, T.O. (1993) Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Science*, **2**, 1511-1519. <https://doi.org/10.1002/pro.5560020916>
- [18] Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., *et al.* (2018) MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Science*, **27**, 293-315. <https://doi.org/10.1002/pro.3330>
- [19] Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., *et al.* (2007) MolProbity: All-Atom Contacts and Structure Validation for Proteins and Nucleic Acids. *Nucleic Acids Research*, **35**, W375-W383. <https://doi.org/10.1093/nar/gkm216>
- [20] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermans, J. (1981) Interaction Models for Water in Relation to Protein Hydration. In: Pullman, B., Ed., *Intermolecular Forces*, Springer, Dordrecht, 331-342. [https://doi.org/10.1007/978-94-015-7658-1\\_21](https://doi.org/10.1007/978-94-015-7658-1_21)
- [21] Kavanaugh, J.S., Rogers, P.H. and Arnone, A. (1992) High-Resolution X-Ray Study of Deoxy Recombinant Human Hemoglobins Synthesized from. Beta.-Globins Having Mutated Amino Termini. *Biochemistry*, **31**, 8640-8647. <https://doi.org/10.1021/bi00151a034>
- [22] Snyder, D.A., Aramini, J.M., Yu, B., Huang, Y.J., Xiao, R., Cort, J.R., Shastry, R., Ma, L.-C., Liu, J., Rost, B., *et al.* (2012) Solution NMR Structure of the Ribosomal Protein RP-L35Ae from *Pyrococcus furiosus*. *Proteins: Structure, Function, and Bioinformatics*, **80**, 1901-1906. <https://doi.org/10.1002/prot.24071>

- [23] Bhavani, B.S., Rajaram, V., Bisht, S., Kaul, P., Prakash, V., Murthy, M.R.N., Rao, N.A. and Savithri, H.S. (2008) Importance of Tyrosine Residues of *Bacillus stearothermophilus* Serine Hydroxymethyltransferase in Cofactor Binding and L-*allo*-Thr cleavage. *The FEBS Journal*, **275**, 4606-4619. <https://doi.org/10.1111/j.1742-4658.2008.06603.x>
- [24] Kadian Singh, P. and Mistry, K.N. (2016) A Computational Approach to Determine Susceptibility to Cancer by Evaluating the Deleterious Effect of nsSNP in XRCC1 Gene on Binding Interaction of XRCC1 Protein with Ligase III. *Gene*, **576**, 141-149. <https://doi.org/10.1016/j.gene.2015.09.084>
- [25] Lettre, G. (2012) The Search for Genetic Modifiers of Disease Severity in the  $\beta$ -Hemoglobinopathies. *Cold Spring Harbor Perspectives in Medicine*, **2**, a015032. <https://doi.org/10.1101/cshperspect.a015032>
- [26] Onda, M., Akaishi, J., Asaka, S., Okamoto, J., Miyamoto, S., Mizutani, K., *et al.* (2005) Decreased Expression of Haemoglobin Beta (HBB) Gene in Anaplastic Thyroid Cancer and Recovery of Its Expression Inhibits Cell Growth. *British Journal of Cancer*, **92**, 2216-2224. <https://doi.org/10.1038/sj.bjc.6602634>
- [27] Carlice-dos-Reis, T., Viana, J., Moreira, F.C., Cardoso, G.L., Guerreiro, J., Santos, S. and Ribeiro-dos-Santos, Â. (2017) Investigation of Mutations in the HBB Gene Using the 1,000 Genomes Database. *PLoS ONE*, **12**, e0174637. <https://doi.org/10.1371/journal.pone.0174637>