Computers in Biology and Medicine 125 (2020) 104018

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

Computational and drug target analysis of functional single nucleotide polymorphisms associated with Haemoglobin Subunit Beta (*HBB*) gene

Opeyemi S. Soremekun [a], Chisom Ezenwa [b], Itunuoluwa Isewon [e], Mahmoud Soliman [a], Omotuyi Idowu [b,f], Oyekanmi Nashiru [b,**], Segun Fatumo [b,c,d,*]

[a] *Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban, 4001, South Africa*
[b] *Centre for Genomics Research and Innovation, National Biotechnology Agency, Nigeria*
[c] *Uganda Medical Informatics Centre and MRC/UVRI LSHTM, Uganda*
[d] *London School of Hygiene and Tropical Medicine, London, United Kingdom*
[e] *Department of Computer and Information Sciences, Covenant University, Ota, Nigeria*
[f] *Chemo-genomics Research Unit, Department of Biochemistry, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria*

ARTICLE INFO

Keywords:
SNP informatics
Single nucleotide polymorphism (SNPs)
Haemoglobin subunit beta (HBB)
Beta thalassaemia
Mutation
Alpha thalassaemia
Haemoglobinopathies

ABSTRACT

There is overwhelming evidence implicating Haemoglobin Subunit Beta (*HBB*) protein in the onset of beta thalassaemia. In this study for the first time, we used a combined SNP informatics and computer algorithms such as Neural network, Bayesian network, and Support Vector Machine to identify deleterious non-synonymous Single Nucleotide Polymorphisms (nsSNPs) present in the *HBB* gene. Our findings highlight three major mutation points (**R31G**, **W38S**, and **Q128P**) within the *HBB* gene sequence that have significant statistical and computational associations with the onset of beta thalassaemia. The dynamic simulation study revealed that **R31G**, **W38S**, and **Q128P** elicited high structural perturbation and instability, however, the wild type protein was considerably stable. Ten compounds with therapeutic potential against HBB were also predicted by structure-based virtual screening. Interestingly, the instability caused by the mutations was reversed upon binding to a ligand. This study has been able to predict potential deleterious mutants that can be further explored in the understanding of the pathological basis of beta thalassaemia and the design of tailored inhibitors.

1. Introduction

Haemoglobin (Hb) is the main carrier of oxygen in the red blood cells (RBC). Structurally, it is made up of haem groups that covalently bind to the two alpha and two beta subunits [1]. Heritable disorders of Hb synthesis have been described as the most widely spread form of human monogenic disorders, chiefly among these disorders are those altering the adult Haemoglobin Subunit Beta gene (*HBB*) [2]. Sickle cell disease and beta thalassaemia have been identified as the most clinically significant diseases that affect the *HBB* [1]. Beta thalassaemia is caused by a wide range of mutations that reduce the production of beta-globin [1]. Genetically, beta thalassaemia is caused due to mutation or deletions in the beta-globin gene, consequently resulting in reduced (beta^{+}) or absent (beta^{0}) production of beta chains of Hb. The incidence of beta thalassaemia cut across North and Sub-Saharan Africa, the Mediterranean, Southeast Asia [2]. Due to the transcontinental movement of humans, beta thalassaemia is no longer endemic to the aforementioned countries but is now a major public health issue in Europe and North America. Thalassaemia major, thalassaemia intermedia, and thalassaemia minor are the main forms of beta thalassaemia [3]. Diagnosis of beta thalassaemia is either through haematological or genetic testing [4]. The *HBB* gene is a 1.6 kb long gene, it possesses 3 exons including 5′ and 3′ untranslated regions. Regulation of the *HBB* gene is via the adjacent 5’ promoter which houses the CACCC, CAAT, and TATA boxes [5]. Several transcription factors bind and regulate the function of the *HBB* gene, the most important of which is erythroid Kruppel-like factor 1, which binds the proximal CACCC box [5]. The mutations of the *HBB* gene cause the abnormal formation of haemoglobin which leads to improper oxygen transportation and damage of red blood cells [6]. Patients with mutations in both *HBB* alleles that significantly reduce the HBB protein production suffer from severe anaemia and skeletal abnormalities [7]. Polymorphic forms of a gene having a frequency higher

* Corresponding author. London School of Hygiene and Tropical Medicine, London, United Kingdom.
** Corresponding author.
E-mail addresses: oyekan.nash@gmail.com (O. Nashiru), Segun.fatumo@lshtm.ac.uk (S. Fatumo).

https://doi.org/10.1016/j.compbiomed.2020.104018
Received 12 August 2020; Received in revised form 13 September 2020; Accepted 20 September 2020
Available online 25 September 2020

than 1% are termed single nucleotide polymorphisms (SNPs) [8]. Identification of SNPs present in a gene is important in biomedical research because they can serve as biological markers that facilitate the recognition of genes implicated in the pathogenesis of a particular disease. SPNs have also found useful application in pinpointing the position of genes between a diseased group and a controlled group in genome-wide association study (GWAS). Furthermore, they can help provide insight into the correlation between phenotypes, drug metabolism, and drug response. Most genetic variations, disorders, and abnormalities seen in humans emanate as a result of SNPs [9,10]. Taking into account the role *HBB* plays in haemoglobinopathies, it is expedient to study the implications of its polymorphic variants. Therefore, this study aims to identify deleterious and disease-causing non-synonymous Single Nucleotide Polymorphisms (SNPs) in *HBB* that could serve as molecular and genetic biomarkers for the diagnosis of beta thalassaemia. These SNPs could also be specifically targeted by inhibitors.

2. Methodology

2.1. Single nucleotide polymorphism data retrieval

The protein sequence of Human HBB with accession number P68871 was retrieved from the UniProt database [11]. HBB variants and their corresponding SNPs were retrieved from the National Centre for Biotechnology Information (NCBI) dbSNPs server [12]. The SNPs were filtered to fetch only those implicated in beta thalassaemia. Furthermore, only those reported to have clinical significance by ClinVar [13] were selected.

2.2. HBB structure elucidation, functional impact and stability analysis of predicted HBB nsSNPs

The available structures of HBB are those co-crystallized with Haemoglobin subunit alpha (HBA1 and HBA2) and protoporphyrin IX containing FE. Due to the expected structural and dynamic effect of HBA1/HBA2 and protoporphyrin IX containing FE on HBB, it was expedient for us to model the 3D structure of HBB using homology modelling technique. Homology modelling is a strategy employed in predicting the 3-dimensional structure of a protein [14]. We used Iterative Threading Assembly Refinement (I-TASSER) [15] which employs a hierarchical technique to determine protein structure. I-TASSER first pinpoints structures that could serve as a template from the Protein Data Bank (PDB) [16] by using Local Meta-Threading Server (LOMETS) [17]. Validation of the predicted HBB protein was assessed using Verify-3D [18], RAMPAGE [19], and ProSA web server [20]. Afterward, SiteMap [21] was used to identify potential binding pockets on the HBB protein. This is to facilitate the potential targeting of HBB in the drug design and development process. Point mutations were exerted on the modelled HBB 3D structure by using the “swapaa” command line in CHIMERA.

Identification of pathological nsSNPs is essential in unravelling the potential impact of a protein in pathogenesis and the possibility of targeting such protein by leveraging on the mutated amino acids. To evaluate the disease-causing potential of HBB nsSNPs, we used Sorting Intolerant from Tolerant (SIFT) [22], Polymorphism Phenotyping v2 (Polyphen 2) [23], Predictor of human Deleterious Single Nucleotide Polymorphism (PhD-SNP) [24], and Protein Variation Effect Analyzer (PROVEAN v1.1) [25]. SIFT determines the impact of amino acid change by using a sequence-based homology algorithm. It classifies an amino acid change as either deleterious or tolerated based on the tolerance index (TI) score. SNPs having a TI score less than 0.05 are considered deleterious while those having a TI score ≥ 0.05 are considered tolerated. Just like SIFT, PolyPhen also evaluates whether an amino acid substitution is disease-causing. In addition, PolyPhen also determines if the substitution occurs in conserved regions. Based on the Position-specific counts (PSIC) score, PolyPhen classifies the mutational impact of a nsSNP as either benign or probably damaging [23]. PhD-SNP uses a supervised learning approach to classify the potential pathogenicity of a nsSNP. PhD-SNP is trained with a dataset of pathological and neutral mutation data using a gradient boosting algorithm. Three functions (mutation parameters, solvent accessibility, and residue/-sequence properties) are used to compute the pathogenicity index which ranges from 0 to 1 [24]. The structural stability of a protein is often affected by nsSNPs which consequently affect the protein structure. To access the stability of HBB nsSNPs, I-Mutant 2.0 [26], iPTREE-STAB [27], and MuPro [28] were used. I-Mutant evaluates the stability of a protein upon mutation by computing the Gibbs free energy using this simple equation. $\Delta\Delta G = \Delta G_{\text{mutant}} - \Delta G_{\text{wild protein}}$ in Kcal/mol at pH 7 and temperature 25 °C.

2.3. Dynamical and structural differences between wild and mutant HBB

To evaluate the structural and dynamic differences that may occur between the wild HBB and Mutant HBB, we used molecular dynamic simulation (MDS). The MDS protocol used has been widely discussed in our previous publications [29,30]. Briefly, the FF14SB forcefield of AMBER18 was used to parameterize the mutant and wild HBB proteins. The topology and parameter files were generated with the aid of the LEAP module. Afterward, restraint potential of 500kcal/molÅ, partial minimization of 2500 steps, full minimization of 5000 steps, gradual system thermalization from 0 to 300 k, and system equilibration of 1000 ps at 300 k without restraints while atmospheric pressure was kept constant at 1 bar using the Berendsen barostat were carried out [31]. Afterward, an MD run of 100ns was carried out [32]. At every 1ps, coordinates and trajectories generated were saved. They were further analysed using CPPTRAJ and PTRAJ [33]. Point mutations were induced at position 31, 38, and 128 of the wild HBB protein to get the mutant HBB (R31G, W38S, and Q128P). This was carried out with the aid of the MODELLER module in UCSF chimera [34].

3. Result

3.1. HBB modelling and structural characterization

The protein sequence of HBB (147 AA residues) with accession number P68871 was retrieved from UniProt [11]. This sequence was then inputted in I-TASSER and 1DXT [35] was used as the template structure to model the 3D-structure of HBB. The modelled structure has a confidence score (C-score) of 1.25, an estimated TM-Score of 0.89 ± 0.07, and an estimated Root Mean Square Deviation (RMSD) of 2.3 ± 1.8 Å. C-score is used to evaluate the standard of HBB in I-TASSER (Fig. 1). The range of the C-score is usually between −5 and 2. C-score of higher value shows that the predicted structure has high confidence. RMSD and TM-score are parameters used in evaluating the structural alikeness between a model structure and a standard structure, especially when the structure of the native protein is known. However, since the native structure is not known, the TM-score and RMSD were evaluated based on C-score. TM-score is a new parameter for structure similarity necessitated due to the sensitivity of RMSD to local error [36].

Structure validation after modelling is an important aspect of homology modelling, as the structural integrity of the crude model needs to be investigated because this would improve the reliability and usage of the modelled structure in downstream bioinformatics or dynamical analysis. Verify-3D, RAMPAGE, and ProSA (z-score) were used to validate the quality of the predicted HBB protein. Analysis using Verify-3D revealed that HBB protein had a profile score of 80%. This suggests that 80% of the amino acid residues of HBB have an average 3D-1D score of ≥ 0.2 (Fig. 2C). RAMPAGE predicted 95.2%, 2.8%, and 2.1% residues to be in favoured, allowed, and outlier regions respectively (Fig. 2B). To further validate the modelled structure, we used ProSA. ProSA uses Z-score to access the overall quality of a structure. ProSA assessment revealed that the modelled HBB structure has a Z-score of −8.53 (Fig. 2A) suggestive of the fact that the modelled structure falls within

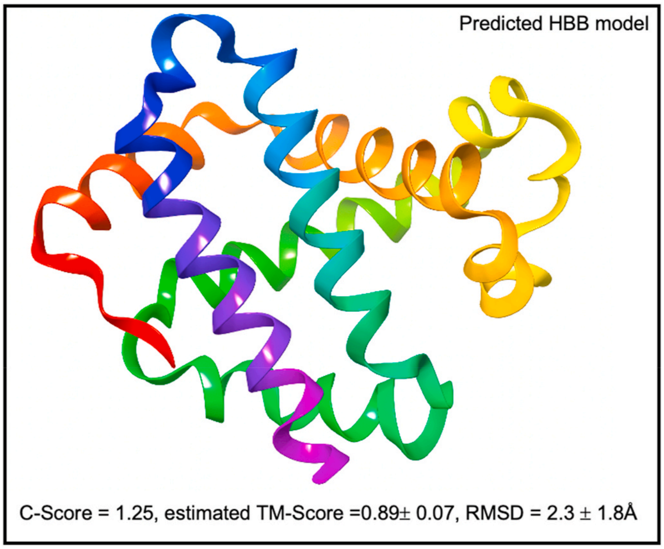

Predicted HBB model
C-Score = 1.25, estimated TM-Score =0.89± 0.07, RMSD = 2.3 ± 1.8Å

Fig. 1. *In-silico* 3-Dimensional structure of HBB modelled using *ab-initio* homology modelling.

the standard of scores attributed to proteins of similar size. Taken together, findings from RAMPAGE, ProSA, and Verify-3D validation revealed the modelled HBB structure has a very good quality and thus, can be used for further structural analysis.

3.2. Evaluation of HBB druggability via binding site analysis

SiteMap [21] predicts potential binding sites of a protein using the following parameters, site score, pocket-size, Dscore, exposure, enclosure, hydrophobicity, and hydrophilicity [21]. For a site to be considered druggable its Dscore, Sitescore, SiteSize, enclosure score, hydrophilicity score, and hydrophobicity score, are expected to be greater than 1.108, 1.091, 156, 0.807, 0.926, and 1.374 respectively. Exposure and enclosure are druggability properties used in determining how open or accessible the site is to solvent. Comparing the enclosure and exposure scores, it can be deduced that the site is not well exposed to solvent, this is evidenced by the deep cleft seen in the 3D structural depiction of the site (Fig. 3B). Furthermore, the site is highly hydrophobic, this could explain why the site had a high enclosure score, hence the burial of the site deeper in the protein structure. Dscore (1.275) also revealed that the site is highly druggable (Table 1).

The residues making up the predicted binding site are reported in Fig. 3C. Again, the high hydrophobicity score reported could also be due to the high number of hydrophobic residues such as leucine,

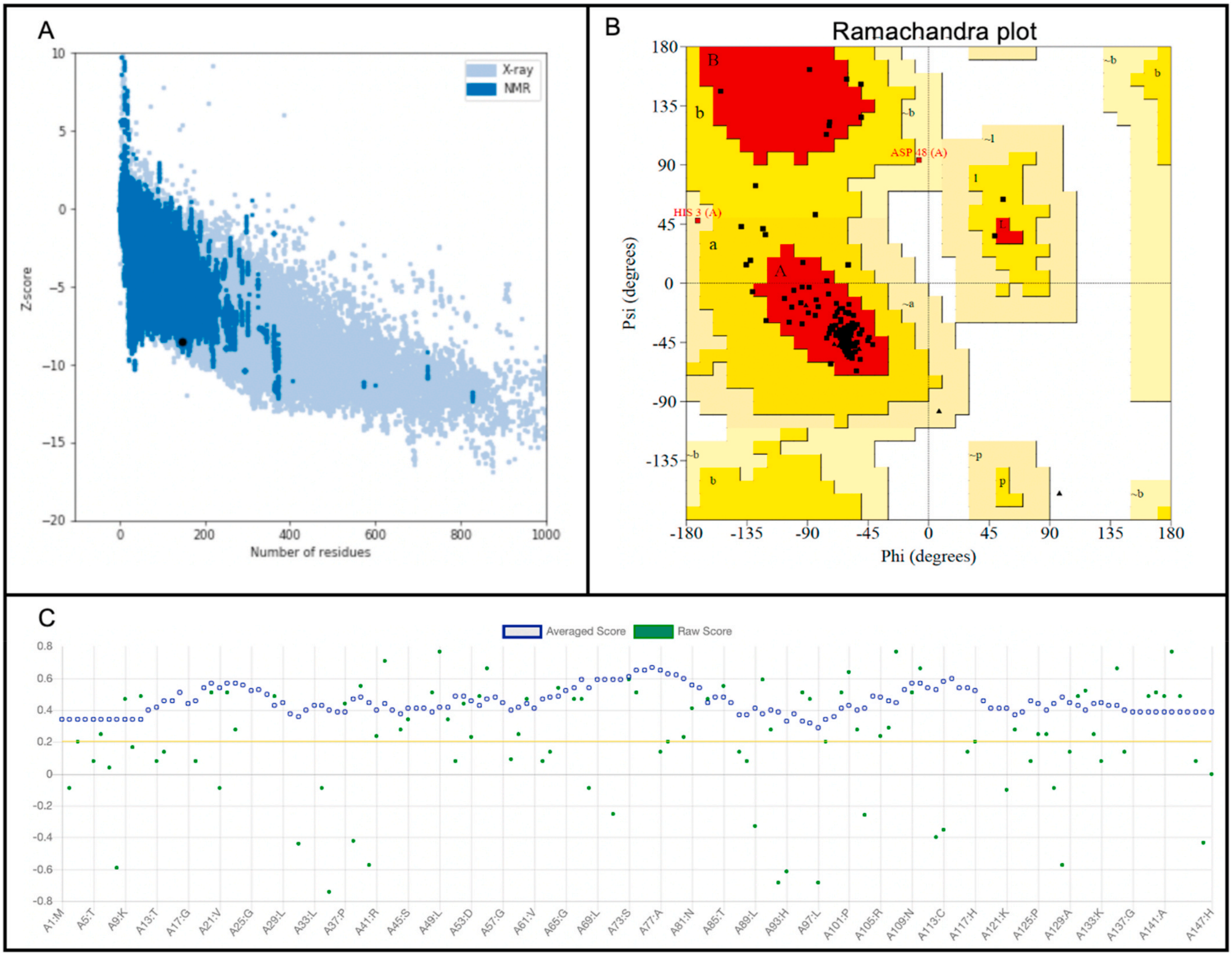


Fig. 2. Assessment and validation of HBB Protein showing ProSA plot (A), Ramachandra plot (B) and Verify3D plot (C).

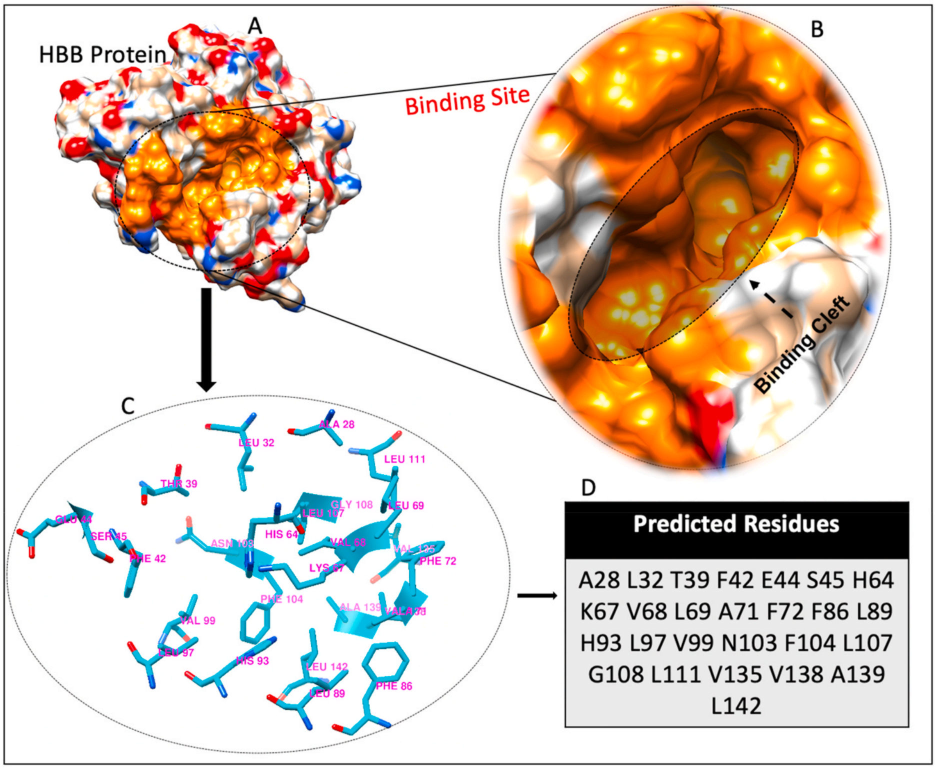


Fig. 3. Structural elucidation of the active site predicted by SiteMap highlighting the surface representation of HBB protein (**A**), binding cleft cavity (**B**), and active site residues (**C** and **D**).

Table 1
HBB druggability properties assessment and their corresponding score.

Druggability Properties	Score
Site Score	1.188
Site Size	163
DScore	1.275
Site Exposure	0.375
Enclosure	0.816
Hydrophobicity	2.970
Hydrophilicity	0.474

phenylalanine, tyrosine, alanine, and valine present in this predicted binding site.

3.3. *SNP informatics analysis*

3.3.1. *Identification of deleterious and damaging nsSNPs*

Discovery and study of protein variants are important in understanding the involvement of a protein in diseases condition. Most especially variants that are pathological [9]. A total of 78 HBB nsSNPs were fetched from the NCBI dbSNP database. The retrieval process was filtered to retrieve only nsSNPs implicated in beta thalassaemia. Furthermore, we made sure that the nsSNPs retrieved were those that have clinical significance as reported by ClinVar [13]. After rigorous screening and filtering, of the 76 SNPs retrieved from NCBI dbSNP, 3 nsSNPs with rsID rs33910569, rs33991059, and rs35684407 were jointly predicted to be disease-causing (Table 2). These three nsSNPs were further used for downstream analysis. The stability of the three predicted nsSNPs revealed that they had a decreased stability and a negative discriminatory direction of thermal stability change. By implication, it means that when compared to the wild type protein, these mutations alter the stability of HBB protein and by extension its activity. The impact of the mutations on the affinity and stability between the subunits was investigated with the aid of mCSM-PPI2 [37]. mCSM-PPI2 is a server used in evaluating the impact of mutation on protein-protein interaction and affinity [37]. Human Hemoglobin with PDB ID 1BZ0 [38] was inputted into the mCSM-PPI2 server. Upon mutation, it was observed that the predicted affinity change ($\Delta\Delta G^{Affinity}$) between the dimeric subunits of R31G, W38S, and Q128P were −2.022 kcal/mol, −1.705 kcal/mol, and −2.261 kcal/mol respectively. This is suggestive of the fact that upon mutation, the affinity between the subunits reduced, probably due to the loss of crucial bonds such as hydrogen bond that hold the subunits in their native special conformation.

This finding was corroborated by the bond loss in the mutant proteins. In the wild type protein of R31G, NH1 and NH2 atoms of HBB^{Arg31} subunit formed a strong polar bond (covalent) with the O atom of HBA^{Phe117} subunit respectively; atom CG of HBB^{Arg31} subunit elicited hydrophobic interactions with atoms CB and CG of HBA^{Pro119} subunit, while NH2 formed another polar bond with atom ND1 of HBA^{Pro119}

Table 2
Damaging and Deleterious SNPs predicted by PolyPhen, PhD-SNP, SIFT, and PROVEAN with their corresponding scores.

S/N	rsID	SIFT	SIFT Prediction	PhD-SNP	PolyPhen	PolyPhen Prediction	PROVEAN
		SIFT Score		Prediction	PolyPhen Score		Prediction
1	rs33910569	0.002	Deleterious	Disease	1.000	Probably damaging	Deleterious
2	rs33991059	0.035	Deleterious	Disease	1.000	Probably damaging	Deleterious
3	rs35684407	0.01	Deleterious	Disease	1.000	Probably damaging	Deleterious

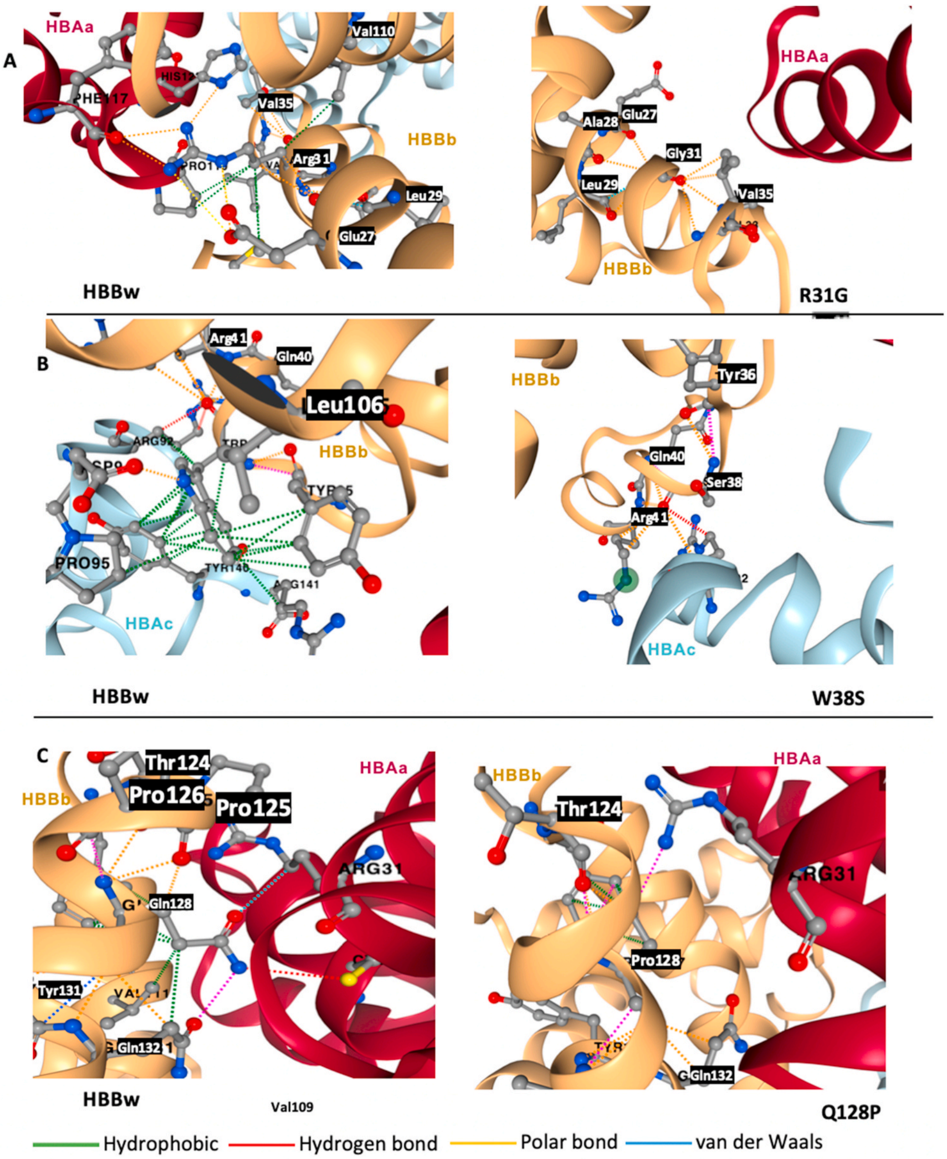


Fig. 4. Effect of mutation on the inter-subunit interactions between HBB and HBA.

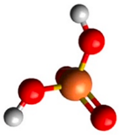
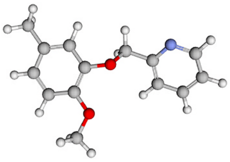
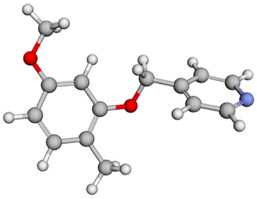
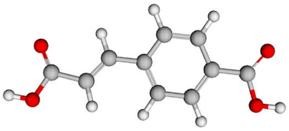
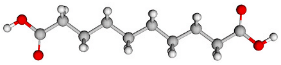
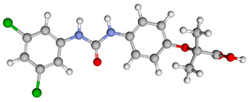
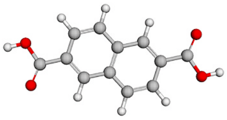
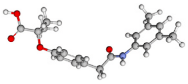
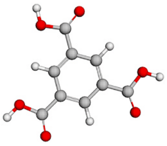
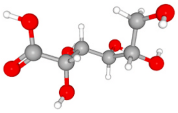
subunit. However, upon mutation, all these bonds were lost (Fig. 4A). Trp38 of HBB formed several hydrophobic interactions with Tyr40 and Arg 141 of the HBA subunit. A polar bond was also formed between atom NE1 of Trp38 and atom OD1 of HBA^{Asp94} subunit. In contrast to the bond behaviour of R31G, W38S formed a hydrogen bond between atom CD of HBA^{Arg92} and O of HBB^{Ser38}, while it still maintained a polar bond between atom CB of HBA^{Arg92} and O of HBB^{Ser38} (Fig. 4B). Gln 128 formed a strong hydrogen bond, van der Waals, and hydrophobic interactions with Cys 104, Arg31, and Val 111 of the HBA subunit, upon mutation, all these bonds were lost but for an atomic clash between Pro128 and Arg31. This clash is however not regarded as a proper bond (Fig. 4C). Aside from affecting the stability and affinity between the Hb tetramers, these mutations also altered the intra-atomic interactions within the residues.

3.4. Wild and mutant HBB as a potential drug target

Structure-based virtual screening (SBVS) is a computational technique employed in the drug design and discovery pipeline to screen small molecules or ligands libraries for a potential bioactive compound against a particular drug target. DrugBank [39] was used as our library of choice because it is an open-source and a comprehensive database. It contains a detailed pharmaceutical, pharmacological and chemical data of each drug. It is common among pharmacists, medicinal chemists, physicians, students, and users from pharmaceutical industries. 20 different molecules were reported to be related to HBB, either as an activator, inducer, oxidizer, or binder. Ten of these compounds were selected for molecular docking using AutoDockTools [40]; this is to evaluate the binding score. A grid box with coordinate (centre: x = 25.0517, y = 26.4691, z = 29.7493 and size: x = 19.5235, y = 19.8934, z = 17.8235) was set around the earlier predicted binding site. The compounds showed an appreciable amount of binding score (Table 3).

When 2, 6-dicarboxynaphthalene was docked with the mutant proteins, it was discovered that 2, 6-dicarboxynaphthalene elicited some interactions such as electrostatic, ionic, and van der Waals with some of the active site residues. The three mutant proteins elicited similar interaction with 2, 6-dicarboxynaphthalene as shown in Fig. 5D. Even though the mutated residues do not have direct interaction with the 2, 6-dicarboxynaphthalene, the stabilizing effect of 2, 6-dicarboxynaphthalene on the mutant protein could be due to the interaction between the mutated residues and active site residues. Ala 28 which is a crucial residue in the active site of HBB, formed a pi-pi stack interaction with 2,

Table 3
Binders with potential efficacy against HBB predicted by Structure-based virtual screening.

Ligands	DrugBank ID	Structure	Binding Score (Kcal/mol)
Iron Dextran	DB00893		−4.4
2-[(2-methoxy-5-methylphenoxy)methyl]pyridine	DB07427		−4.7
4-[(5-methoxy-2-methylphenoxy)methyl]pyridine	DB07428		−4.6
4-Carboxycinnamic Acid	DB02126		−4.7
Sebacic acid	DB07645		−4.2
2-[4-({[(3,5-dichlorophenyl)amino]carbonyl}amino)phenoxy]-2-methylpropanoic acid	DB08077		−4.2
2,6-dicarboxynaphthalene	DB08262		−5.2
Efaproxiral	DB08486		−5.0
Trimesic acid	DB08632		−4.6
Sodium ferric gluconate complex	DB09517		−5.1

6-dicarboxynaphthalene, the residue interaction network plot revealed that Gly31 formed a hydrogen bond with Ala28 thereby providing additional stabilizing effect between Ala28 and 2, 6-dicarboxynaphthalene (Fig. 5A). A similar trend was observed in the Pro128 and Ser38 mutant proteins. Pro128 formed a van der Waals interaction with Ala139 (Fig. 5B) while Ser38 formed hydrogen bond with Glu44 (Fig. 5C). Ala139 and Glu44 formed hydrogen bond and pi-pi stack interaction with 2, 6-dicarboxynaphthalene. Just like the Gly31 Ala28 interaction, we could infer that Pro128 and Ser38 helped in providing an additional however indirect stabilizing bonds for the overall binding.

3.5. *Dynamic difference between wild HBB and mutant HBB*

We plotted the Cα backbone RMSD values during the simulation run of the mutant HBB (Mut_HBB^{R31G}, Mut_HBB^{Q128P}, and Mut_HBB^{W38S}) and ligand-bound mutants (Lig_HBB^{R31G}, Lig_HBB^{Q128P}, and Lig_HBB^{W38S}) relative to their starting coordinate (Fig. 6B). The RMSD revealed that Mut_HBB^{R31G}, Mut_HBB^{Q128P}, and Mut_HBB^{W38S} had average RMSD values of 1.69 Å, 1.94 Å, and 2.22 Å respectively, while Lig_HBB^{R31G}, Lig_HBB^{Q128P}, and Lig_HBB^{W38S} had average RMSD values of 2.00 Å, 1.89 Å, and 2.12 Å. This revealed that Mut_HBB^{R31G}, Mut_HBB^{Q128P}, and Mut_HBB^{W38S} elicited high Cα atom motional movement and instability, however, upon binding with 2,6-dicarboxynaphthalene the instability observed in the Mut_HBB^{Q128P}, and Mut_HBB^{W38S} systems were reversed. This observation is evident in the plot depiction (Fig. 6B) and the average RMSD of the systems. The Mut_HBB^{R31G} system did not exhibit the "instability-stability" behaviour of the Mut_HBB^{Q128P} and Mut_HBB^{W38S} systems. Furthermore, the four systems achieved convergence earlier in the simulation run, around 10ns. A similar trend was observed in the ROG plot (Fig. 6D). PCA is an

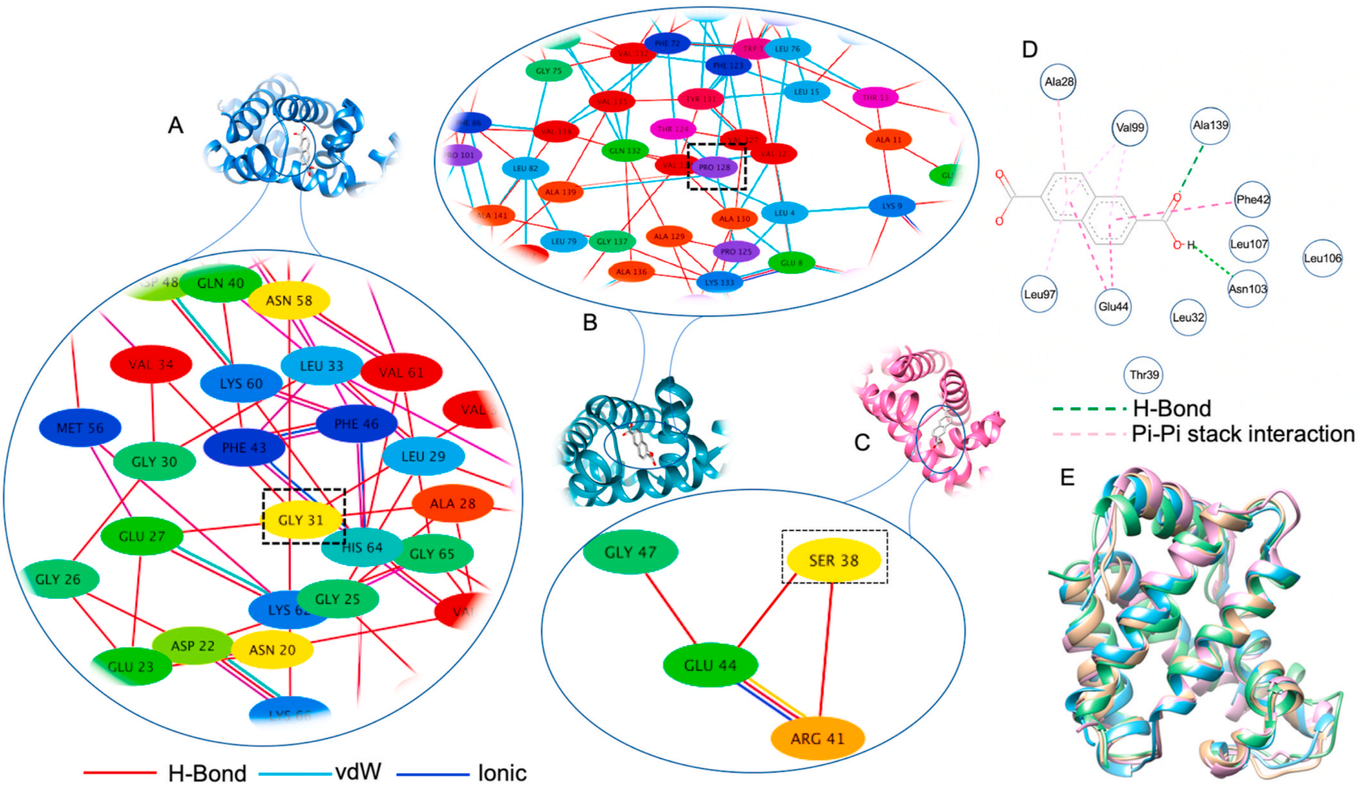


Fig. 5. A schematic describing the impact of R31G (**A**), Q128P (**B**), and W38S (**C**) on active site residues interaction with 2,6-dicarboxynaphthalene. Interaction plot of 2,6-dicarboxynaphthalene and active site residues (D). Superimposed structures of the wild and mutant proteins (Grey = wild protein, Cornflower = R31G, Purple = Q128P, and Green = W38S).

important parameter used in unravelling the conformational changes and mobility of a protein during simulation [41]. Fig. 5A shows the eigenvalues projection during the MD simulation of Mut_HBB^{R31G}, Mut_HBB^{Q128P}, Mut_HBB^{W38S}, Lig_HBB^{R31G}, Lig_HBB^{Q128P}, and Lig_HBB^{W38S} along with the principal components PC1 and PC2. The six systems exhibited unique atomic conformation and motion along the subspace ev1/PC1 vs ev2/PC2. Mut_HBB^{R31G}, Mut_HBB^{Q128P}, and Mut_HBB^{W38S} elicited the lowest dispersive motion when compared to Lig_HBB^{R31G}, Lig_HBB^{Q128P}, and Lig_HBB^{W38S}. This observation is suggestive of the fact that the mutations increased the motional movement of HBB, however, this dispersive motion was stabilized upon 2,6-dicarboxynaphthalene binding. RMSF was used to elucidate the flexibility and motion of individual residues in the proteins. Mut_HBB^{R31G}, Mut_HBB^{Q128P}, and Mut_HBB^{W38S} had overall average RMSF values and flexibility. As observed in the PCA and RMSD plots, the flexibility of the residues was altered (Fig. 6E). The hydrogen bond (Fig. S1) and SASA (Fig. S2) plots of the six systems also corroborated the results discussed above. The mutant Mut_HBB^{R31G}, Mut_HBB^{Q128P}, Mut_HBB^{W38S} proteins had an average total number of hydrogen bonds of 61, 60, and 62 respectively, however, upon binding with 2, 6-dicarboxynaphthalene, the hydrogen bonds increased to 64, 61 and 63 respectively (Fig. S1).

4. Discussion

SNPs have been described as the most common form of genetic variability in a given population. SNPs may be specific to an individual or caught across a population. Hence the study of SNPs through SNPinformatics approach can facilitate the quick identification of potential biomarkers that spread across the genome. It has been discovered that individuals respond differently to drugs, SNPinformatics approach could be used to identify the genetic variation in individuals that regulate and determine drug metabolism.

A wide spectrum of genetic mutations that affect the production of beta globin has been implicated in the onset of beta globin related pathological conditions, such as beta thalassaemia [7,42]. Genetic variants of beta thalassaemia that lead to the manifestation of anaemia and a range of clinically asymptomatic conditions attest to how a monogenic disorder can lead to different diseases [43]. Although various nsSNPs have been clinically reported to cause beta thalassaemia [5,44], we used a different SNP informatics approach to identify three nsSNPs that could be targeted in the treatment of beta thalassaemia. Several SNPs have been identified on the beta globin gene, some of them have been routinely used as genetic biomarkers such as prenatal diagnosis. Hashemi-Soteh et al., identified five SNPs (IVSII-74 (G/T), IVS11-16 (C/G), IVSII-81 (C/T), codon 2 (C/T), and IVS11-666 (T/C)), that could be used for the genetic testing of prenatal diagnosis in Iran [45]. Other variants that are currently used in genetic testing of beta thalassaemia include but are not limited to IVSI-5. Codon 41/42 (TCTT), IVS1-1 (G > T/A), 619-bp deletion, codon 8/9 (+G), codon 15 (G > A), codon 16 (-C), poly-A site (T > C), and codon 15 (-T) [46,47]. Although the computational tools used in the identification of deleterious nsSNPs employ different predictive models to evaluate pathogenicity, the predictions are sometimes not the same. Hence, it is advisable to use more than one predictive tool. The overlapping positive results or jointly predicted pathogenicity can then be used as a measure of accuracy and reliability.

There exists a co-crystallized complex of the alpha subunit, protoporphyrin IX containing FE, and a beta subunit. The existing structures have 146 amino acid length compared to the primary sequence that is translated into 147 amino acids. Valine is present at position 2 of the primary sequence while it is absent in some crystallized structures found in PDB. Due to the chemical interactions and the attendant effect of these molecules on the structural integrity of HBB, we employed *ab-initio* modelling to obtain the 3D structure of HBB independent of any chemical interaction. Determination of the potential binding pocket of a protein is essential in the drug discovery process, as this can help

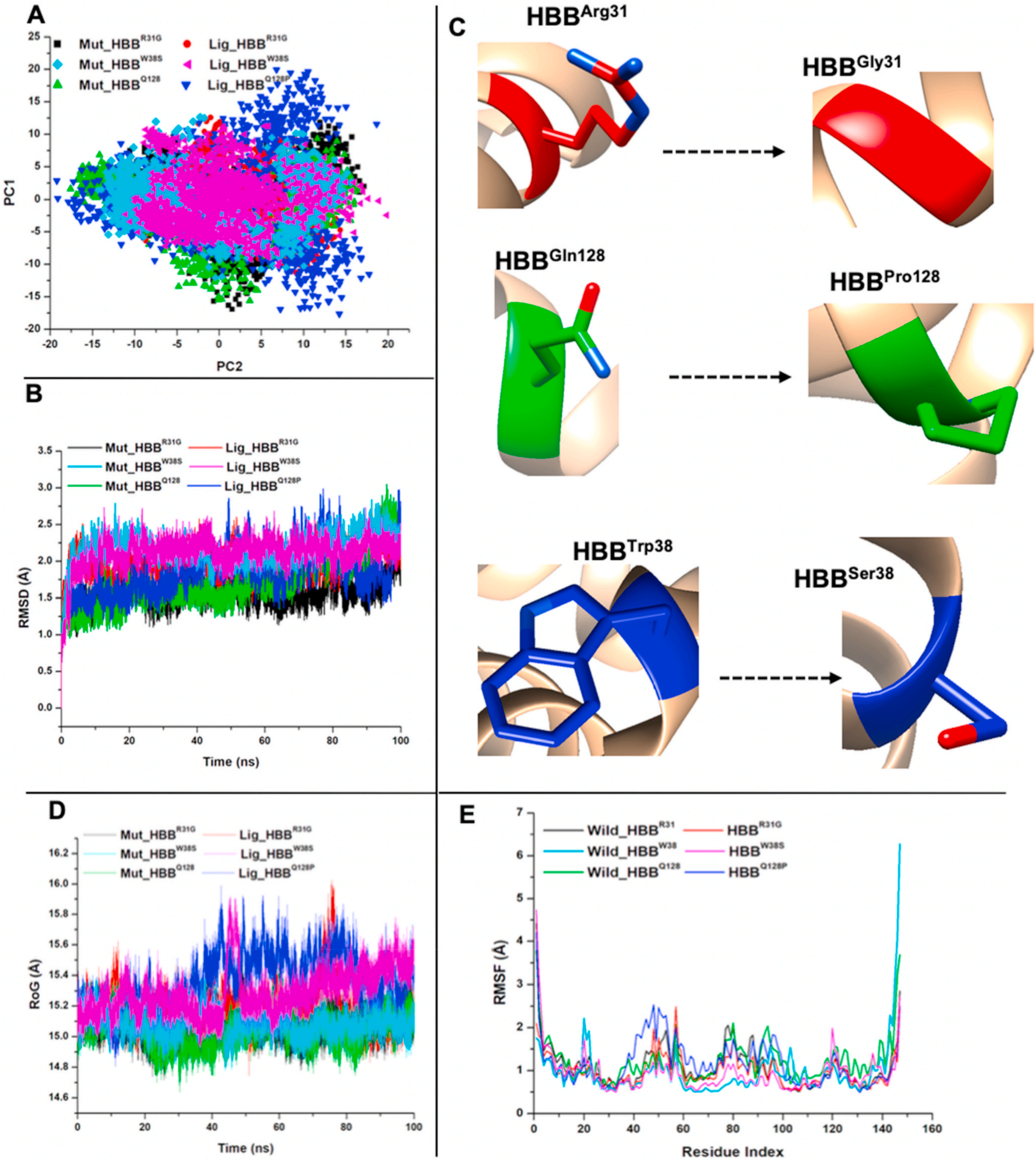


Fig. 6. PCA scatter plots depicting a distinct separation of motions between Mut_HBB^{R31G}, Mut_HBB^{Q128P}, Mut_HBB^{W38S}, Lig_HBB^{R31G}, Lig_HBB^{Q128P}, and Lig_HBB^{W38S} along the first two principal components (**A**). Backbone RMSDs are depicted as a function of time for Mut_HBB^{R31G}, Mut_HBB^{Q128P}, Mut_HBB^{W38S}, Lig_HBB^{R31G}, Lig_HBB^{Q128P}, and Lig_HBB^{W38S} (**B**) Amino acid substitution between the wild HBB and mutant HBB (**C**). Radius of gyration of C-α atoms of Mut_HBB^{R31G}, Mut_HBB^{Q128P}, Mut_HBB^{W38S}, Lig_HBB^{R31G}, Lig_HBB^{Q128P}, and Lig_HBB^{W38S} versus time at 300 k (**D**). RMSF of the C-α of Mut_HBB^{R31G}, Mut_HBB^{Q128P}, Mut_HBB^{W38S}, Lig_HBB^{R31G}, Lig_HBB^{Q128P}, and Lig_HBB^{W38S}.

provide more insight into the druggability and targetability of the protein. Therefore, we employed an online server, SiteMap [21] to characterize the potential binding site of HBB.

rs33910569, rs33991059, and rs35684407 were identified as disease-causing. rs33910569 has Glutamine replaced with Proline at position 128. Several studies have mapped a mutation at the position consistent with rs33910569 [48,49]. For example, Girodon et al., used a Computer-designed denaturing gradient gel electrophoresis (DGGE) to characterize and determine a missense mutation in the exon 3 of beta-globin where Arg was substituted for Gln [50]. rs33991059 has Tryptophan being substituted by Ser at position 38. Experimental studies that have confirmed the existence of this mutation include but are not limited to the researches of Fujita [51], Yamaoka [52], Kornblit et al. [53] etc. Arginine is substituted by Glycine at position 31 of the beta subunit of HBB [54]. This substitution was believed to be a β^0 -thal allele since heterozygous individuals show elevated Hb A2 β^0 -thal trait [55]. Individuals with mutations in both *HBB* alleles that significantly reduce the HBB protein production suffer from severe anaemia and skeletal abnormalities [7]. These three nsSNPs were further used for downstream analysis. The stability of the three predicted nsSNPs revealed that they had a decreased stability and a negative discriminatory direction of thermal stability change. Several studies have reported the effect of mutations in the HBB subunit of Hb on the oxygen binding-dissociation state of Hb especially in residues close to the Hb binding cleft [56–59]. 31G and 38S are close to the Hb binding site, therefore, it is expected that the change in helical distance and instability provoked by 31G and 38S mutations, will disrupt the native conformational state of the protoporphyrin IX containing FE centre. Hence, reducing the affinity of the protoporphyrin IX containing FE centre for oxygen and consequently the oxygen transport capability of Hb. This could be the reason despite the high proportion of red blood cells produced by patients with thalassaemia, they still have a high incidence of anaemia.

To elucidate the time-wise structural event of HBB protein upon mutation, we employed RMSD) and Principal Component Analysis (PCA). RMSD is a widely used quantitative variable used in the estimation of structural stability between two different structures [29]. The RMSD values can be determined for different parts of an atom, however, the Cα of the whole protein is often calculated during MD simulation. The stability plot revealed that the effect of the mutation was markedly reduced upon binding of a ligand.

5. Conclusion

One of the major aims of identifying deleterious nsSNPs is functional and structural analysis. In-depth knowledge of the conformational change of a protein can help unravel the mechanisms of disease phenotypes and in the identification of potential drugs that can therapeutically alter the function of the proteins [9]. nsSNPs have been reported to disrupt the structure-function relationship of a protein, which has led to the onset of diverse disease amongst humans and other species [60]. Due to the rapid change in the genomic landscape of humans brought about by exposure to chemical, sunlight, and radiation etc. it is becoming increasingly difficult for experimental Biologists to keep track of these nsSNPs. Computational algorithms have provided a sigh of relief in the quick identification of these deleterious nsSNPs. We used an in-silico approach to identify SNP that could serve as biomarkers for genetic testing of beta-thalassaemia. Among these SNPs, rs33910569 and rs33991059 have been experimentally validated. Furthermore, we were able to predict some drugs with potential therapeutic efficacy against HBB. Being a computational study, one of the limitations of this research is the need for experimental investigation of these genes as the result cannot be outrightly taken as a justification to be used in humans. Functional assays using cell lines could be explored as well. Another major limitation of this study is the disparity in the output of the prediction tools used in this study. This is due to the different algorithms employed by the tools. This limitation was circumvented to a point by using more than one tool for a particular analysis. The predicted compounds also need thorough therapeutic validation and toxicity evaluation. The future prospect of this research is to experimentally validate the SNPs predicted.

Funding

ON and SF are funded in part by the National Institutes of Health (NIH), United States Common Fund to the H3ABioNet Project grant number 5U24HG006941-09.

Declaration of competing interest

The authors declare none.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2020.104018.

References

[1] S. Iyer, S. Sakhare, C. Sengupta, et al., Hemoglobinopathy in India, Clin. Chim. Acta. Elsevier, 2015, pp. 229–233.

[2] S.L. Thein, The molecular basis of β-thalassemia, Cold Spring Harb. Perspect. Med. 3 (2013) 1–24.

[3] R.J.A. Trent, Diagnosis of the haemoglobinopathies, Clin. Biochem. Rev. 27 (2006) 27–38.

[4] L. Cai, H. Bai, V. Mahairaki, et al., A universal approach to correct various HBB gene mutations in human stem cells for gene therapy of beta-thalassemia and sickle cell disease, Stem Cells Transl. Med. 7 (2018) 87–97.

[5] A. Cao, R. Galanello, Beta-thalassemia, Genet. Med. Nature Publishing Group, 2010, pp. 61–76.

[6] S. Chaudhary, D. Dhawan, P.G. Bagali, et al., Compound heterozygous β+ β0 mutation of HBB gene leading to β-thalassemia major in a Gujarati family - a case study, Mol. Genet. Metab. Reports. 7 (2016) 51–53.

[7] D.J. Weatherall, Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias, Nat. Rev. Genet. Nat Rev Genet (2001) 245–255.

[8] M. Alanazi, Z. Abduljaleel, W. Khan, et al., In silico analysis of single nucleotide polymorphism (snps) in human β-globin gene, PloS One 6 (2011) 25876.

[9] O.S. Soremekun, M.E.S. Soliman, From genomic variation to protein aberration: mutational analysis of single nucleotide polymorphism present in ULBP6 gene and implication in immune response, Comput. Biol. Med. 111 (2019) 103354.

[10] D. Gurdasani, T. Carstensen, S. Fatumo, et al., Uganda genome resource enables insights into population history and genomic discovery in Africa, Cell 179 (2019) 984–1002, e36.

[11] A. Bateman, M.J. Martin, C. O'Donovan, et al., UniProt: the universal protein knowledgebase, Nucleic Acids Res. 45 (2017) D158–D169.

[12] S.T. Sherry, M. Ward, M. Kholodov, et al., dbSNP : the NCBI database of genetic variation 29 (2001) 308–311.

[13] M.J. Landrum, J.M. Lee, G.R. Riley, et al., ClinVar: public archive of relationships among sequence variation and human phenotype, Nucleic Acids Res. 42 (2014) 980–985.

[14] S.K. Panda, S. Saxena, L. Guruprasad, Homology modeling, docking and structure-based virtual screening for new inhibitor identification of Klebsiella pneumoniae heptosyltransferase-III, J. Biomol. Struct. Dyn. 38 (2020) 1887–1902.

[15] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinf. 8 (2008) 1–8.

[16] S. Parasuraman, Protein data bank, J. Pharmacol. Pharmacother. 3 (2012) 351–352.

[17] S. Wu, Y. Zhang, LOMETS: a local meta-threading-server for protein structure prediction, Nucleic Acids Res. 35 (2007) 3375–3382.

[18] C.T. Kresge, M.E. Leonowicz, W.J. Roth, J.C. Vartuli, J.S. Beck, ^{2}»ÉøÍÐÔµÄ © 19 9 2 nature publishing group, Nature 359 (1992) 710–713.

[19] S.C. Lovell, I.W. Davis, W.B. Adrendall, , et al.S. Altschul, W. Gish, W. Miller, E. W. Myers, D.J. Lipman, Basic local alignment search tool. Journal of molecular Biology.etry: phi,psi and C beta deviation, Structure validation by C alpha geomF, Proteins-Structure Funct. Genet. 50 (1990) 437–450, 2003.

[20] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007) 407–410.

[21] T.A. Halgren, Identifying and characterizing binding sites and assessing druggability, J. Chem. Inf. Model. 49 (2009) 377–389.

[22] R. Vaser, S. Adusumalli, S.N. Leng, et al., Protocol UPDATE SIFT missense predictions for genomes, Nat. Protoc. 11 (2015) 1–9.

[23] I.A. Adzhubei, S. Schmidt, L. Peshkin, et al., HHS Public Access 7 (2010) 248–249.

[24] E. Capriotti, P. Fariselli, PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants, Nucleic Acids Res. 45 (2017) W247–W252.

[25] Y. Choi, A.P. Chan, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels, Bioinformatics 31 (2015) 2745–2747.

[26] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (2005) 306–310.

[27] L.T. Huang, M.M. Gromiha, S.Y. Ho, iPTREE-STAB: interpretable decision tree based method for predicting protein stability changes upon mutations, Bioinformatics 23 (2007) 1292–1293.

[28] J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, Proteins Struct. Funct. Genet. 62 (2006) 1125–1132.

[29] O.S. Soremekun, F.A. Olotu, C. Agoni, et al., Drug promiscuity: exploring the polypharmacology potential of 1, 3, 6-trisubstituted 1, 4-diazepane-7-ones as an inhibitor of the 'god father' of immune checkpoint, Comput. Biol. Chem. 80 (2019) 433–440.

[30] R.O. Kumi, O.S. Soremekun, A.R. Issahaku, et al., Exploring the ring potential of 2,4-diaminopyrimidine derivatives towards the identification of novel caspase-1 inhibitors in Alzheimer's disease therapy, J. Mol. Model. 26 (2020) 1–17.

[31] H.J.C. Berendsen, J.P.M. Postma, WF Van Gunsteren, et al., Molecular dynamics with coupling to an external bath Molecular dynamics with coupling to an external bath, J. Chem. Phys. 3684 (2012) 926–935.

[32] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341.

[33] D.R. Roe, T.E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular synamics trajectory data, J. Chem. Theor. Comput. 9 (2013) 3084–3095.

[34] Z. Yang, K. Lasker, D. Schneidman-Duhovny, et al., UCSF Chimera, MODELLER, and IMP: an integrated modeling system, J. Struct. Biol. 179 (2012) 269–278.

[35] J.S. Kavanaugh, P.H. Rogers, A. Arnone, High-resolution X-ray study of deoxy recombinant human hemoglobins synthesized from β-globins having mutated amino termini, Biochemistry 31 (1992) 8640–8647.

[36] Y. Zhang, J. Skolnick, Scoring function for automated assessment of protein structure template quality 710 (2004) 702–710.

[37] C.H.M. Rodrigues, Y. Myung, D.E.V. Pires, et al., MCSM-PPI2: predicting the effects of mutations on protein-protein interactions, Nucleic Acids Res. 47 (2019) W338–W344.

[38] J.S. Kavanaugh, W.F. Moo-Penn, A. Arnone, Accommodation of insertions in helices: the mutation in hemoglobin catonsville (Pro 37α-glu-thr 38α) generates a 310→ α bulge, Biochemistry 32 (1993) 2509–2513.

[39] D.S. Wishart, C. Knox, A.C. Guo, et al., DrugBank: a knowledgebase for drugs, drug actions and drug targets, Nucleic Acids Res. 36 (2008) 901–906.

[40] O. Trott, A. Olson, AutoDock Vina:improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461.

[41] F. Sittel, A. Jain, G. Stock, Principal component analysis of molecular dynamics: on the use of Cartesian vs. internal coordinates, J. Chem. Phys. 141 (2014).

[42] R.C. Tine, M. Ndiaye, H.H. Hansson, et al., The association between malaria parasitaemia, erythrocyte polymorphisms, malnutrition and anaemia in children less than 10 years in Senegal: a case control study, BMC Res. Notes 5 (2012) 1–10.

[43] A.W. Nienhuis, D.G. Nathan, Pathophysiology and clinical manifestations of the β-thalassemias, Cold Spring Harb. Perspect. Med. 2 (2012) 1–13.

[44] C. Badens, P. Joly, I. Agouti, et al., Variants in genetic modifiers of β-Thalassemia can help to predict the major or intermedia type of the disease, Haematologica 96 (2011) 1712–1714.

[45] M.B. Hashemi-soteh, S.S. Mousavi, A. Tafazoli, Haplotypes inside the Beta-Globin Gene : Use as New Biomarkers for Beta-Thalassemia Prenatal Diagnosis in North of Iran, vols. 4–8, 2017.

[46] T. Owen, M.D. Chan, K. Westover, L. Dietz, J.L. Zehnder, Schrijver, Comprehensive and efficient HBB mutation analysis for detection of beta-hemoglobinopathies in a pan-ethnic population, Am. J. Clin. Pathol. 133 (2010) 700–707.

[47] X. Shang, Z. Peng, Y. Ye, et al., Rapid targeted next-generation sequencing platform for molecular screening and clinical genotyping in subjects with hemoglobinopathies, EBioMedicine 23 (2017) 150–159.

[48] H.H. Kazazian, S.H. Orkin, C.D. Boehm, et al., Characterization of a spontaneous mutation to a β-thalassemia allele, Am. J. Hum. Genet. 38 (1986) 860–867.

[49] H.H. Kazazian, S.H. Orkin, C.D. Boehm, et al., β-Thalassemia due to a deletion of the nucleotide which is substituted in the β(s)-globin gene, Am. J. Hum. Genet. 35 (1983) 1028–1033.

[50] E. Girodon, N. Ghanem, M. Vidaud, et al., Rapid molecular characterization of mutations leading to unstable hemoglobin β-chain variants, Ann. Hematol. 65 (1992) 188–192.

[51] S. Fujita, Oxygen equilibrum characteristics of abnormal hemoglobins:hirose (alpha2beta237ser), L ferrara (alpha247Glybeta2), broussais (alpha290Asnbeta2), and dhofar (alpha2beta258Arg), J. Clin. Invest. 51 (10) (1972) 2520–2529.

[52] K. Yamaoka, Hemoglobin Hirose: 2 237(C3) tryptophan yielding serine, Blood 38 (6) (1971) 730–738.

[53] B. Kornblit, P. Taaning, H. Birgens, β-thalassemia due to a novel nonsense mutation at codon 37 (TGG→TAG) found in an Afghanistani family, Hemoglobin 29 (2005) 209–213.

[54] Waye JS, Eng B, Patterson M, Chui DHK and Fernandes BJ. Novel β^0 -thalassemia mutation in a Canadian woman. Hemoglobin. 21:4, 385-387..

[55] M. Schmugge, J.S. Waye, R.K. Basran, et al., THE Hb S/β+-thalassemia phenotype demonstrates that the IVS-I (-2) (A>C) mutation is a mild β-thalassemia allele, Hemoglobin 32 (2008) 303–307.

[56] G.D. Efremov, Dominantly inherited β-thalassemia, Hemoglobin 31 (2007) 193–207.

[57] R.S. Human, A.J. Irlathews, R.J. Rohlfss, et al., The Effects of E7 and E 11 Mutations on the Kinetics of Ligand, 1989, pp. 16573–16583.

[58] A. Allen, C. Fisher, A. Premawardhena, et al., Adaptation to anemia in hemoglobin E-β thalassemia, Blood 116 (2010) 5368–5370.

[59] L. Mouawad, D. Perahia, C.H. Robert, et al., New insights into the allosteric mechanism of human hemoglobin from molecular dynamics simulations [Internet], Biophys. J. 82 (2002) 3224–3245, https://doi.org/10.1016/S0006-3495(02)75665-8. Available from.

[60] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2, Curr. Protoc. Hum. Genet. 7 (20) (2013).