Journal of Pharmaceutical, Chemical and Biological Sciences
ISSN: 2348-7658
CODEN: JPCBBG
September - November 2018; 6(3):188-198
Online available at https://www.jpcbs.info

Research Article

In Silico Analysis of β-Thalassemia Mutations in India and its Neighbouring South East Asian Countries

Dibyajyoti Rabha[1], Dipankar Baruah[2], Paresh Kumar Sarma[2], Jatin Sarmah[3]*

[1]Dibyajyoti Rabha, DBT (Govt. of India) sponsored Bioinformatics Infrastructure Facility, Bodoland University, Kokrajhar, Assam, India
[2]Dipankar Baruah, Associate Professor, Department of Pathology, Fakhruddin Ali Ahmed Medical College and Hospital, Barpeta, Assam, India
[2]Paresh Kumar Sarma, Associate Professor, Department of Medicine, Fakhruddin Ali Ahmed Medical College and Hospital, Barpeta, Assam, India
[3]Jatin Sarmah, Associate Professor, Department of Biotechnology & Co-ordinator, DBT (Govt. of India) sponsored Bioinformatics Infrastructure Facility, Bodoland University, Kokrajhar, Assam, India

***CORRESPONDING AUTHOR**
Dr.Jatin Sarmah, Associate Professor, Department of Biotechnology & Co-ordinator, DBT (Govt. of India) sponsored Bioinformatics Infrastructure Facility, Bodoland University, Kokrajhar, Assam, India
Email: jatinsarmahindia@gmail.com

ARTICLE INFORMATION
Received August 03, 2018
Revised September 12, 2018
Accepted September 14, 2018
Published October 03, 2018

ABSTRACT
Thalassemia that arises due to mutation in the ß-globin chain of Haemoglobin is called ß-thalassemia. *In silico* analysis of the 4 common mutations- CD 26(G>A), IVS1-5(G>C), IVS1-1(G>T) and CD41/42(-CTTT) occurring frequently in India as well as in its South East Asian neighbouring countries were performed. Literature search on ß-thalassemia mutation and their frequencies were performed on Pubmed database in the populations included in the study. The four common mutations found present in all the populations are analysed using bioinformatics tools for *in silico* analysis, genotype-phenotype correlation and interactome. One of the common mutations is used for predicting its secondary structure using online tool Protein Homology/analogy Recognition Engine v2.0 (PHYRE2). The pathogenicity of these mutations are assessed along with segregation and expression analysis. Structural and functional analysis of CD 26(G>A) was performed. The pathogenicity result for the mis-sense HBB variant CD 26(G>A) is predicted to have no pathogenicty expression but for the intronic mutations IVS1-5(G>C) and IVS1-1(G>T) and the frameshift mutation CD41/42(-CTTT) all are disease causing. Fatal phenotype can be identified by in silico analysis but with less accuracy in unknown or novel variants facilitating prenatal diagnosis, genetic counselling and preimplantation genetic diagnosis for the ß-thalassemia patients and their families.

KEYWORDS: ß thalassemia; *In silico*; PHYRE2; I-TASSER; South East Asia

INTRODUCTION

Thalassemia is single gene disorder that occurs due to mutation in the Haemoglobin (Hb) gene and result in microcytic anemia [1]. HBA and HBB are the two sub-units of Haemoglobin with the α and ß chain respectively. Thalassemia that arises due to mutation in the ß -globin chain of Hb is called ß-thalassemia consisting of more than 800 variants in the HBB gene described in the Beta Globin Gene Server [2-4].

Point mutations form the majority in HBB mutations. Mutations in promoter, introns, splice sites and exons affect normal functioning of the gene. Functions of the gene such as

transcription, splicing and translation can be affected.

Out of the total world thalassemics born every year, 10% of them are born in India [5]. The first case of thalassemia from India was described in a non-Mediterranean person. Subsequently in all parts of the country, cases of thalassemia were documented [6]. ß –thalassemia is the most common single gene disorder in India [7]. ß-thalassemia mutations have been found to be relatively population specific [8-10]. According to a WHO update in 2008, overall carrier frequency of ß-thalassemia in India is 3-4% [11, 12]. As India is a country with a huge population, this emphasized the need of prenatal diagnosis and carrier status detection for containing the disease. Nepal, a South Asian country shares several similarities of mutation profiles with that of India [13].

From different parts of the world over 150 different mutations causing ß thalassemia have been reported [14]. In South East Asian countries including Thailand HbE/ß-thalassemia is a common thalassemia syndrome [15].

Our study is based on the surveys performed in India, Nepal and other neighbouring South East Asian countries like Myanmar, Indonesia, Thailand and Malaysia to find out various common mutations of HBB gene. This study aimed for *in silico* analysis of common mutations to study the pathogenicity of the variants and their genotype-phenotype correlation. Such comparative analysis of mutations of HBB gene found in these countries will be useful for the ß thalassemia patients and their families for conducting genetic counselling.

MATERIALS AND METHODS

Search strategies

For finding published reports of ß thalassemia mutations in India, Nepal and its neighbouring countries, a PubMed database search was conducted using the keywords: "beta globin", "gene", "mutation", and/or "beta thalassemia" and "population's name such as India, Nepal, Myanmar, Indonesia, Thailand and Malaysia". India being the largest country with a huge number of population, a thorough search for papers on ß thalassemia was performed on PubMed database. The papers were screened and included based on 3 criteria- (a) Patient's or carrier's status, (b) The frequencies of the mutations, and (c) molecular detection methods. Type of mutations, geographical location, ethnicity and the paper's year of publication were collected too.

Abstraction of mutation data

Using the keywords mentioned above, literature search was conducted in PubMed database for HBB mutations against the mentioned populations and identified the reported mutations available in the database.

Mutation selection

From the literatures collected mutation frequencies of carriers and patients were calculated found in the populations. For further studies only the most frequent mutations were collected. The frequently common mutations of Indian population were then rechecked in the population of the neighbouring countries. One such frequently observed mis-sense mutation was then selected for *in silico* and structural analysis.

In silico analysis

The selected variant was analysed using different software and servers as follows-

Structural and functional analysis

Using Phyre2 (V2.0), the structural and functional analysis of the variants in protein were determined and was compared to the HBB protein sequence (UniProtKB/Swiss-Prot P68871). For structure and function predictions I-TASSER (Iterative Threading ASSEmbly Refinement) server was used [16]. COACH server performed the functional analysis [17]. PANNZER2 (Protein ANNotation with Z-scoRE) tool was used for functional analysis.

Protein Interaction Network

The prediction for functional association of HBB in network of proteins was performed by STRING database version 10.0. STRING database consists of known and predicted protein-protein interactions. It includes direct (physical) and indirect (functional) associations. They begin from computational prediction, from the knowledge that transfers between organisms and from the interactions aggregated from other databases (primary). The five main sources from where interactions in STRING are derived are- (a) Genomic context predictions, (b) High throughput lab experiments, (c) co-expression,

(d) Automated text mining and (e) Previous knowledge in databases.

SNP annotations

The pathogenic effect of the missense mutation was assessed through application of several available bioinformatics tools like Polymorphism Phenotyping (PolyPhen-2) [18], nsSNPAnalyzer [19]. They were used for predicting functional effects of human nsSNPs. MutPred was used for predicting the pathogenicity of amino acid substitutions and their molecular mechanisms.

RESULTS

Selected studies

For analysis 59 articles were collected from India. These articles included cases of minor, major and intermedia thalassemia. These articles contained 19,318 cases of ß—thalassemia patients.

From other populations and ethnic groups 25 articles were included from the South Asia and South East Asian countries. The studies from the articles on Indian population was categorised into carriers and affected cases of individuals.

From these 84 articles we found 4 common mutations viz., IVS 1-5 (G→C), CD 41/42 (-CTTT), IVS 1-1 (G→T) and Codon 26 (G>A).

***In silico* analysis**

Among the common variants the four common mutations i.e., IVS 1-5 (G→C), CD 41/42 (-CTTT), IVS 1-1 (G→T) and Codon 26 (G>A) were focused upon. The structural and functional analysis of the missense mutation Codon 26 (G>A) or p.E27K was performed. Frameshift mutations cause truncated protein and impaired functional product. Various bioinformatics tools were then used for analyzing the other three mutations.

Protein sequence alignment

p.E27K was the subject of the protein sequence alignment among the other species of vertebrates. The degree of conservation was investigated for each amino acid (**Fig. 1**). The degree of conservation for Glutamic acid at position 27 is highly conserved. Therefore, the substitution of Glutamic acid (E) at position 27 by Lysine may cause pathogenicity.

Structure prediction

With the use of online tools, secondary structure prediction was achieved for p.E27K. The prediction of secondary structure of a protein is based on threading model [16]. The structure was predicted by I-TASSER. On I-TASSER top five final models of predicted secondary structures with 3-D predicted tertiary structures were listed (**Fig. 2A**). To simulate the final model based on pair-wise structure similarity of the decoys (using SPICKER program), decoys from a large structural conformation models were generated. C-score measures the confidence which for p.E27K is 1.24, estimated TM score = 0.88 ± 0.07 and RMSD = 2.3 ± 1.8Å. A higher value of C-score signifies a model with a high confidence and vice-versa [16, 20, 21]. Tm score > 0.5 indicates a model of correct topology while TM score < 0.17 means random similarity.

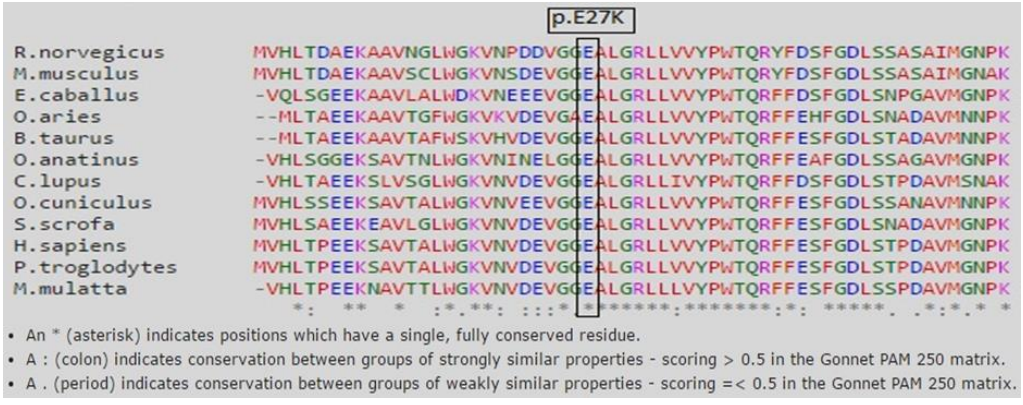

Fig. 1: Protein sequence alignment of HBB amino acid sequence among vertebrates using Clustal Omega. The position and the residue at which the missense mutation p.E27K occurred is indicated in the black box. The degree of conservation of residue is also determined. E27 is a fully conserved residue among all the other organisms

Protein sequence alignment using Clustal Omega

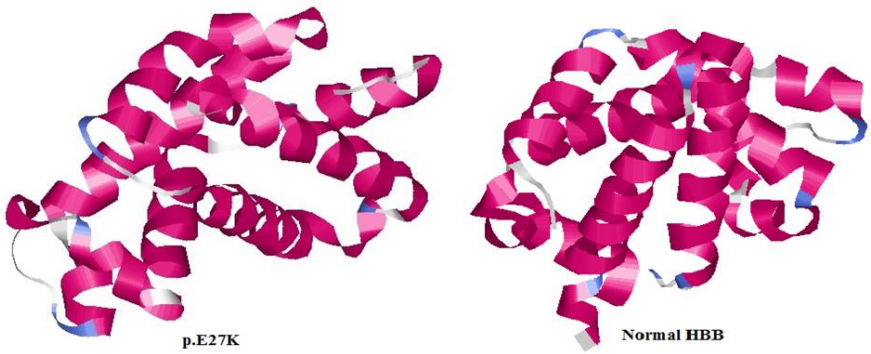

Fig. 2(A): Cartoon structural models of mutant HBB (p.E27K) and normal human HBB protein from the data provided by I-TASSER server

Structural model prediction using I-TASSER server

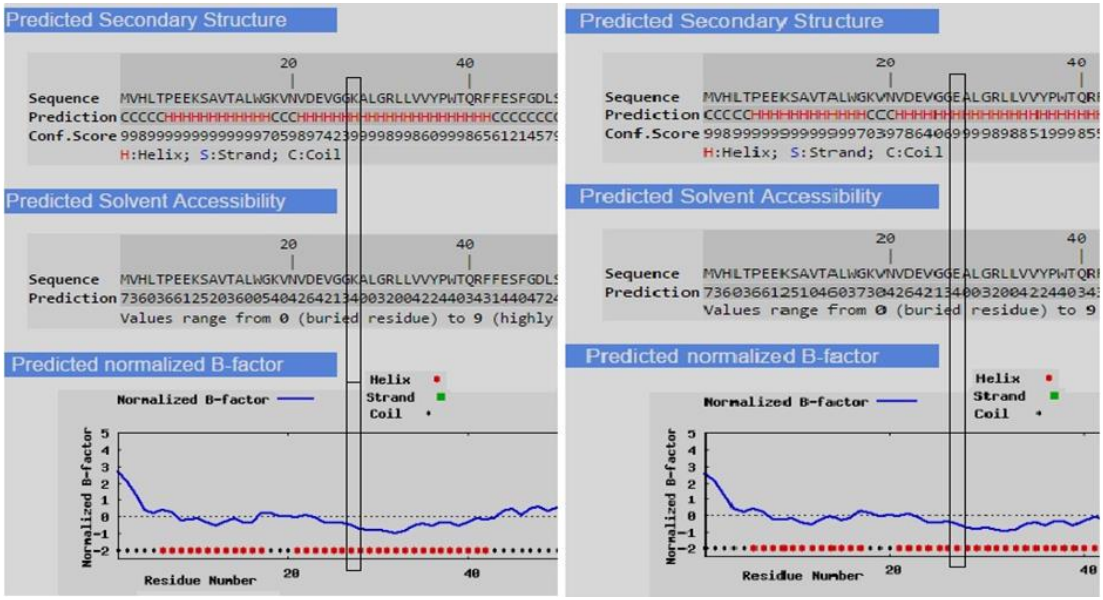

Fig. 2(B): Solvent accessibility and normalized BFP of the predicted secondary structure is shown for p.E27K (left) and normal HBB (right)

I-TASSER results showing solvent accessibility and normalized BFP

I-TASSER determined the secondary structure as a helix structure with a high confidence value for p.E27K and its solvent accessibility was also predicted as 4 which is similar to the value in the normal structure (Fig. 2B). In the figure, the sequence is indicated by the first line; the second line (C: random coil; H: alpha helix; S: beta-strand) shows helix at position 27 with a confidence score of 9 (third line). A higher score indicates higher confidence prediction, the range of confidence being 0-9. The solvent accessibility of p.E27K is 4. The values range from 0 (buried residue) to 9 (highly exposed residue). The stability of the predicted secondary structure is determined by the predicted normalised B-factor. BFP values higher than 0 are less stable in experimental structures. BFP value of E27 was -0.48 and the BFP value of E27K is -0.47 (>0) which means stable in experimental structure. Using a combination of both template-based assignment and profile based prediction the normalized B-factor (BFP) is predicted. The LOMETS threading program was used for performing threading template of the query protein. With the Z-score measurement the highest significant alignment regions of the templates are chosen. The blocks and spatial positions in the assembly is the basis of the alignment. The PDB 1fhjB is the top threading template prediction for p.E27K (Aquomet haemoglobin-I of the maned wolf- Chrysocyon brachyuru – Z score > 1 showed a good alignment (E27K: identity 0.89; normalized Z-score = 2.48 and coverage = 0.99). The second template PDB 3eokA has higher Z-score (3.24) but has lower identity=0.68 for p.E27K. Model template's cartoon structure is shown here

(Fig.2A). C-score of the modelled template for p.E27K = 1.24, estimated TM score=0.88 ± 0.07, RMSD = 2.3 ± 1.8Å. C-score is typically in the range (-5, 2). A model with a high confidence is signified by a higher value of C-score and vice-versa, a TM-score > 0.5 indicates a model of correct topology and a TM-score < 0.17 means random similarity. RMSD and TM-score values indicate the similarity of the predicted structures to the native structures.

The BFP or B-factor profile value of E27 and E27K were -0.48 and -0.47 respectively. The BFP value of p.E27K is slightly higher than that of E27.

The LOMETS threading program performed the threading template of p.E27K. The top threading template prediction for it was based on PDB 1fhjB (Aquomet haemoglobin-I of the maned wolf-Chrysocyon brachyuru – identity 0.89; normalized Z-score of the threading alignments = 2.48 and coverage = 0.99 [16]. A normal sequence's predicted threading template was also based on 1fhjB – identity 0.90; normalized Z-score = 2.49 and coverage = 0.99. There is just a slight difference between the normal and p.E27K.

Protein Homology/analogy Recognition Engine V2.0 server was used for predicting the structure. The prediction was based on template-based homology modelling and fold recognition [22]. The model for p.E27K was based on the template d2dn3b1 which is a member of globin-like superfamily. The score for confidence was 100, identity = 99% and coverage = 99%. The template d2dn3b1 was a crystal structure of human haemoglobin in the carbomonoxy form. The secondary structure prediction confidence score is high (red) (Fig. 3) at the position E27K. The structure is determined to be helix at position 27. The disordered score of p.E27K is 15%.

In Fig. 3 the amino acid sequence is indicated by the first line and the secondary structure prediction in the second line is determined as alpha helix (H) extended or ß strand and coiled structure with confidence value of low to high (red, depicted in third line). In the fourth line using the disoPred program the structurally ordered sequence indicating two states of ordered and disordered was calculated. The homology modelling of p.E27K was based on the template d2dn3b1, members of globin-like superfamily model. The protein change p.E27K has 79% alpha helix structure changes and the structure disorder is about 15% with 100% confidence. The biological activity of the protein is determined by the chemical properties of the amino acids. Glutamic acid (E) is an acidic and polar (charged) amino acid. Lysine (K) is an essential amino acid and is positively charged.

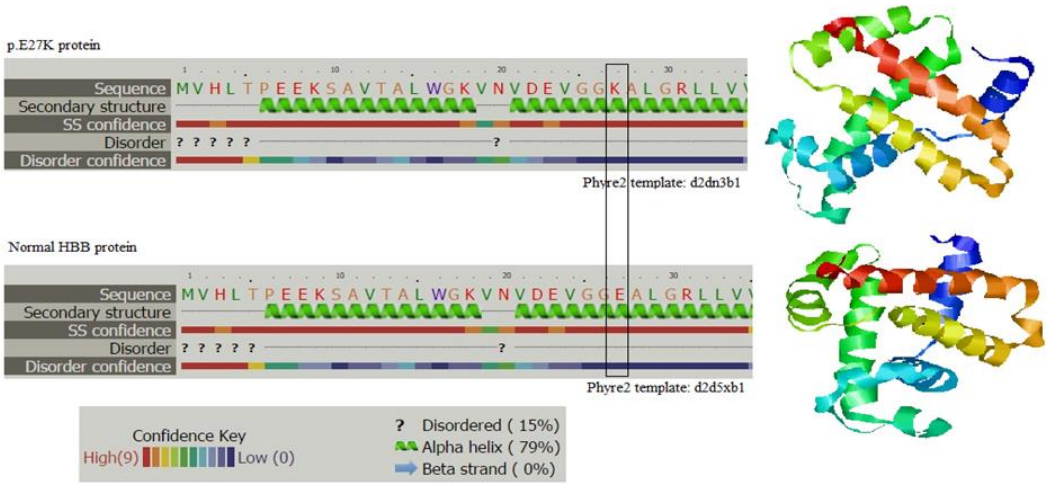

Fig. 3: Secondary structure predicted by Phyre2 server of p.E27K (above) and normal HBB (below)

PHYRE2 tool predicting the secondary structure

Functional Analysis

PANNZER2 tool was use to predict Gene Ontology (GO) classes and free text descriptions about protein functionality. The molecular functions predicted were: oxygen carrier activity, oxygen binding, heme binding, metal ion binding, haptoglobin binding, haemoglobin binding and peroxidase activity. The functions were similar to that of normal haemoglobin. In comparison to normal globin the functions of

p.E27K predicted were similar. Oxygen transport and blood coagulation were the top listed biological functions predicted by PANNZER2 while the functions predicted for molecular were oxygen carrier activity, oxygen binding, heme binding, metal ion binding and to lesser extent haptoglobin binding, haemoglobin binding and peroxidase activity.

Prediction of binding sites of p.E27K was performed by the tool 3DLigandSite and no effective change in binding sites were seen (Fig. 4). Though Glutamic acid residue at position 27 was substituted by Lysine residue there was no change in the predicted binding sites.

Using COACH server I-TASSER performed the functional analysis. Determination of ligand binding sites, enzyme commission, and gene ontology were performed by functional homologous templates. On the basis of ligand binding site analysis p.E27K matched to PDB 1fhjB (Crystal structure maned wolf-Chrysocyon brachyurus-hemoglobin-I) with C-score = 1.24 as hem for its binding factor.

Enzyme commission for p.E27K was determined based on the PDB 1qvhA (C-score = 0.376 and TM-score = 0.762, identity = 0.145) with oxidoreductase activity. The PDB 1qvhA reveals an unexpected geometry of the distal heme pocket with two active sites i.e., 119 and 131. It is an X-ray structure of ferric Escherichia coli flavohemoglobin. Another comparison to PDB 1cqxA (crystal structure of the flavohemoglobin from Alcaligenes eutrophus at 1.75 A resolution) shows low enzyme commission (C-score = 0.360, TM-score 0.713 and identity = 0.152) with no active sites. From this we can say that this protein has low enzymatic activity.

The figure (Fig. 4) consists all the predicted binding sites with the number of ligands that they contact, the average distance between the residue and the residue conservation score. As illustrated we do not see a change in the residue's contact in p.E27K (left) in comparison with the normal HBB protein (right). For each residue the average distance from conservation score of each residue (range: 0-1.00) is defined. The colour shows the binding site range. Low distance indicates high accuracy and lower coverage but increase in distance indicates less accuracy and higher coverage.

Using gene ontology the functional analysis is also investigated. The PDB 1dxtB matched to p.E27K with coverage = 1.00, C-score = 0.91, TM-score = 0.9699, identity = 0.99. The second ranked template that matched p.E27K is 1g0bB with coverage = 0.99, C-score = 0.77, TM-score = 0.9817 and identity = 0.83. This shows that the protein functions were not changed due to the amino acid change. Alterations are more damaging to the quality of function of the protein and not to the ontology.

Predicted Binding Site

p.E27K protein				Normal HBB protein			
Residue	Amino acid	contact	av distance	Residue	Amino acid	contact	av distance
32	LEU	24	0.37	32	LEU	21	0.45
39	THR	24	0.35	39	THR	24	0.35
42	PHE	24	0.10	42	PHE	25	0.05
43	PHE	25	0.11	43	PHE	25	0.17
64	HIS	25	0.00	45	SER	12	0.67
67	LYS	20	0.40	46	PHE	20	0.45
68	VAL	24	0.20	64	HIS	25	0.03
71	ALA	25	0.08	67	LYS	7	0.56
72	PHE	25	0.09	68	VAL	24	0.28
86	PHE	24	0.54	71	ALA	25	0.25
89	LEU	25	0.07	72	PHE	25	0.04
92	LEU	25	0.16	86	PHE	22	0.51
93	HIS	25	0.00	89	LEU	25	0.04
97	LEU	25	0.21	92	LEU	25	0.05
99	VAL	13	0.46	93	HIS	25	0.00
103	ASN	25	0.19	97	LEU	25	0.10
104	PHE	7	0.18	99	VAL	16	0.48
107	LEU	25	0.03	103	ASN	25	0.15
138	VAL	8	0.55	104	PHE	7	0.18
142	LEU	17	0.41	107	LEU	25	0.08
				138	VAL	8	0.60
				142	LEU	25	0.36

Prediction colour legend: Other residues | Predicted Binding Site

Conservation Score Colour legend: 0-0.15 | 0.16-0.30 | 0.31-0.40 | 0.41-0.50 | 0.51-0.60 | 0.61-0.70 | 0.71-0.80 | 0.81-1.00

Fig. 4: 3DLigandSite predicts potential binding sites (cluster) in comparison to normal amino acid sequence through the data submitted by the Phyre2

Potential binding sites prediction using 3DLigandSite

Interactome analysis

The structure of a protein may change due to mutation and with it the function also changes. The interaction of various proteins with the mutated protein may evolve phenotypic effect. To investigate this, STRING v10.0 server was used to inspect the interaction of various proteins or genes with HBB in a network system. The proteins that were predicted to interact with HBB are as follows: NFE2 (Nuclear Factor Erythroid 2), HBG2 (Haemoglobin Gamma 2), AHSP (Alpha Haemoglobin Stabilizing Protein), HBA2 (Haemoglobin sub-unit Alpha 2), HBZ (Haemoglobin sub-unit Zeta), HBA1 (Haemoglobin sub-unit Alpha 1), HP (Haptoglobin), HPR (Haptoglobin-Related-Protein), HPX (Haemopexin) and CD163 (Cluster of Differentiation 163) (Fig. 5). The function of a protein changes with the change in its structure. Therefore, a change in B globin structure may lead to interaction with other proteins and may evolve the phenotype, hence, altering the biological functions of haemoglobin.

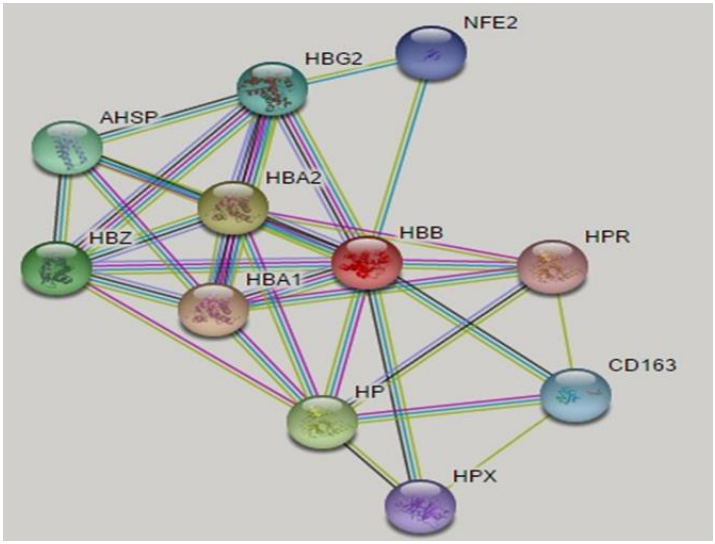

Fig.5: Protein-protein interaction network generated by STRING (v10.0)
STRING (v10.0) generated interaction network

Pathogenicity Analysis

The pathogenicity of the variants were predicted using several bioinformatics online tools depending on the type of mutations. For pathogenicity prediction of p.E27K, tools such as MutPred2, PolyPhen2 and nsSNP Analyzer were used. MutPred2 showed score = 0.428 for p.E27K. A score threshold of 0.50 would suggest pathogenicity. Polyphen2 with a score of 0.000 (sensitivity: 1.00; specificity: 0.00) predicted p.E27K to be benign. nsSNP Analyzer predicted the amino acid substitution in p.E27K to be neutral with no pathogenic phenotypic effect (Table 1).

The pathogenicity of the frameshift mutation CD 41/42(-CTTT) and two other intronic mutations viz., IVS1-1(G>T) and IVS1-5(G>C) were also predicted using Mutation Taster. For CD 41/42(-CTTT), pathogenicity was predicted as disease causing. Mutation Taster prediction of pathogenicity for the two intronic mutations: IVS1-5(G>C) and IVS1-1(G>T) were also disease causing (Table 2).

Table 1: Pathogenicity prediction of the missense mutation CD 26G > A of the HBB gene

Mutation	HGVS Nomenclature	dbSNP	Type	PolyPhen2	nsSNP Analyzer	MutPred
Codon 26G > A	c.79G > A	rs33950507	B^+	Benign (Score 0.000)	Neutral	0.428

Table 2: Bioinformatic analysis of pathogenicity of frameshift and intronic mutations of HBB gene

Mutation	HGVS Nomenclature	dbSNP	Type	Mutation Taster prediction
CD41/42(-CTTT)	c.126_129delCTTT	rs80356821	B^0	Disease causing
IVS1-5(G>C)	c.92 + 5G > C	rs33915217	B^+	Disease causing
IVS1-1(G>T)	c.92 + 1G > T	rs33971440	B^0	Disease causing

DISCUSSION

In this study; a total of 19,318 affected carriers and their families of India were investigated. From neighbouring country Nepal and countries of South East Asia like Myanmar, Indonesia, Thailand and Malaysia, total of 5,069 cases were studied. Several studies have been performed on ß thalassemia patients over the past few decades and all the reported mutations were reported in this cohort.

Due to different functional mutations on the ß globin gene, HBB gene mutations occur. Among those we have investigated 4 most frequently reported mutations.

The mutation diversity in Indian cohorts of border region is similar to neighbouring countries of Nepal and other South East Asian countries. This suggests historical immigration and emigration of these populations.

The frequencies of different pathogenic HBB variations were gathered for this study. For the analysis of heterogenous and homogenous populations and subpopulations and their comparison such data is valuable. This data would also be helpful in choosing the most cost-effective strategy for screening of patients as well as in premarital and carrier counselling.

Due to incomplete analysis of some reports there maybe some bias. The reason for this may be due to methodology used in each study. ARMS-PCR and RFLPs were used in most studies for specific analysis of the patients and other mutations could have been missed though sequencing and complete analysis of the genes in some studies were needed for complete evaluation. For the phenotype genotype analysis of patients complete Sanger sequencing of the gene is helpful.

The following 4 mutations have been reported in almost all the populations taken into account in this study: IVS1-5(G>C), CD 41/42(-CTTT), IVS1-1(G>T) and CD 26(G>A).

CD 26(G>A) [HBB:c.79G>A] mutation lead to E27 amino acid substitution with Lysine (p.E27K). HbE is very common in South East Asia as well as in the Indian sub-continent. The HbE thalassemia is highly variable with some patients being asymptomatic while others being transfusion dependent. Through this study, the goal is to improve management and counselling of families.

In silico analysis

The biochemical analysis of mutations more than 800 in the HBB gene calls for an intensive effort and attention. However, all of those mutations have not been characterized and their clinical consequences remain unsolved. In silico analysis makes this straightforward [23]. The main functions of the Haemoglobin protein include oxygen transport activity; heme binding, haemoglobin binding, iron-iron binding and mutations in the ß globin subunit of the haemoglobin protein may disrupt its functions.

The main molecular activities of the HBB include oxygen carrier activity, oxygen binding, heme binding, metal ion binding and to lesser extent haptoglobin binding, haemoglobin binding and peroxidase activity. The HBB protein functions in tetramer and any variant in its gene sequence may affect the amino acid sequence, its expression as well as in the protein function. The stability may also be changed due to conformational and folding positions at mRNA and consequently at the protein level. For mis-sense mutation, I-TASSER and PHYRE2 were used for investigating its structural study at the DNA level. The structural data predicted by the servers could assign phenotypes to novel variants and as such would be helpful for practitioners and geneticists.

For previously characterized and uncharacterized mutations, structural and pathogenic aberrations were investigated computationally based on the changes brought to the protein by each variant to evaluate the confidence of the tools on novel mutations. Change in protein stability by single point

mutation was predicted and investigated by structural and functional analysis of p.E27K. Frameshift mutation lead to truncated protein. Prediction for such mutation was done by Mutation Taster. To construct model, secondary structures were determined using the native and altered amino acid sequence. For investigating protein stability, the solvent accessibility was determined. Comparison of dynamic models before and after mutation were done to evaluate the altered and native protein models and its consequences were explored in the protein's functions for biological functions (COACH), enzyme activity, ligand binding sites, gene ontology and binding sites to evaluate the changes. For additional future studies, the functional analysis in this study may be a good model.

Because of more availability of software, predictions of substitution mutations is easier using in silico analysis than the prediction in intronic and non-coding variants. Mutation Taster was used for predicting pathogenicity of the intronic variants. Pathogenicity analysis showed the variants to be disease causing i.e., of pathogenic character.

In silico evaluation of pathogenicity makes it easier and faster to predict the effect of new and novel mutations and could inform the ß-thalassemia families at risk. The data available from this study indicates that the mutations in the HBB gene's non-coding regions maybe responsible for some of the phenotypes which may be scrutinized using several bioinformatics servers (**Table 2**). Novel mutations not reported in the SNP database should be investigated if found in any thalassemic individuals. However, segregation or expression analysis should confirm this prediction. The advantage of pathogenicity prediction of a mutation allows confirmation of a defect in individuals and facilitates genetic testing and counselling of other high-risk family members.

Interactome network of proteins

From the interaction analysis, it can be seen that the interaction is with ten proteins including HBA1 and HBA2. They are mainly involved in oxygen transport [24]. The other proteins are NFE2, HBG2, AHSP, HBZ, HP, HPR, CD163 and HPX. NFE2 possess DNA binding transcription factor activity and transcription co-activator activity, HBG2 possess functions such as iron ion binding and oxygen binding, acts as a chaperone. During normal erythroid cell development it prevents the aggregation (damaging) of alpha haemoglobin. HBZ also functions in iron ion binding and oxygen binding, HP possess activities like serine-type endopeptidase activity and haemoglobin activity. HPR also possess the same functions such as the HP gene. CD163 possess the activity of scavenger receptor activity and HPX is involved in heme transporter activity. A change due to mutation in the HBB gene would have an effect on the interaction of proteins and factors, i.e., transcription, translation, development and function of the Hb gene and consequently the phenotype.

CONCLUSION

Mutations of high frequencies reported in this region of the world was analyzed. Such study is valuable for regions with no access to advanced genetic technologies for analysis of panel based studies. For performing prenatal diagnosis of beta thalassemia this article will be of interest. In silico analysis can be used to identify the fatal phenotype but not with great accuracy for unknown variants. For assessing pathogenicity in novel mutations, computational analysis can be used alongwith segregation and expression analysis. In case of prenatal diagnosis of at-risk patients, computational analysis could be used to predict phenotype but careful analysis should also be done. Through such studies functional consequences of various mutations are known which would give a high confidence for at-risk families to undergo genetic counselling to know the probability of the affected child. It gives the families various options such as termination, treatment of patients or PGD (preimplantation genetic diagnosis).

ACKNOWLEDGEMENT

We gratefully acknowledge the Department of Biotechnology (DBT), Government of India sponsored Bioinformatics Infrastructure Facility for providing the infrastructure to pursue the work and Mr. Dibyojyoti Rabha was offered 6-months studentship by the DBT.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest in this research article.

REFERENCES

1. Sarmah J, Baruah D, Sarma PK, Das D. Dietary Habit, Anthropometric measurements and haematological parameters in correlation with prevalence of iron-deficiency anaemia among never married tribal female postgraduates of Assam, India. Res Rev J Health Prof 2018; 8(2): 19-26.
2. Giardine B, Borg J, Viennas E, Pavlidis C, Moradkhani K, Joly P, et al. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. Nucleic Acids Res 2014; 42: D1063-D1069.
3. Huisman THJ, Carver MFH, Baysal E, Efremov GD. Available from http://globin.bx.psu.edu/hbvar/menu. html. [(A database of human hemoglobin variants and thalassemias)]; 2014.
4. Patrinos GP, Giardine B, Riemer C, MillerW, Chui DH, Anagnou NP, et al. Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. Nucleic Acids Res 2004; 32: D537- D541.
5. Bashyam MD, Bashyam L, Savithri GR, Gopikrishna M, Sangal V, Devi ARR. Molecular genetic analyses of beta thalassemia in South India reveal rare mutations in the beta globin gene. J Hum Genet 2004; 49: 408-413.
6. Verma IC, Saxena R, Kohli S. Past, present & future scenario of thalassaemic care & control in India. Indian J Med Res 2011; 134: 507-521.
7. Verma IC. The challenge of genetic disorders in India. Molecular genetics and gene therapy- the new frontier. Amsterdam: Scientific Communications; 1994; p11.
8. Setianingsih I, Williamson R, Daud D, Harahap A, Marzuki S, Forrest S. Phenotypic variability of Filipino beta (0)-thalassemia/HbE patients in Indonesia. Am J Hematol 1999; 62:7-12.
9. Colah R, Gorakshakar A, Nadkarni A, Phanasgaonkar S, Surve R, Sawant P, Mohanty D, Ghosh K. Regional heterogeneity of beta-thalassemia mutations in the multi ethnic Indian population. Blood Cells Mol Dis 2009; 42: 241-246.
10. Colah R et al. Epidemiology of beta thalassemia in western India: mapping the frequencies & mutations in subversion of Maharashtra and Gujrat. Br J Haematol 2010; 149: 739-747.
11. Population Reference Bureau (PRB). World population data sheet. Washington DC: Population Reference Bureau; 2010.
12. World Health Organization (WHO). Joint WHO-TIF meeting on management of hemoglobin disorders (2nd: 2008: Nicosia, Cyprus). Geneva: World Health Organization; 2008.
13. Mishra A et al. Distribution and ethnic variation of α-thalassemia mutations in Nepal. Nepal Med Coll J 2012; 14(1): 49-52.
14. Baysal E, Carver MFH. The beta and delta thalassemia repository (eight edition). Haemoglobin 1995; 19(3-4):213-236.
15. Wasi P, Pootrakul S, Pootrakul P. Thalassemia in Thailand. Ann N Y Acad Sci 1980; 344: 352-363.
16. Roy A, Kucukural A, Zhang Y. I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc 2010; 5: 725-738.
17. Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 2013; 29: 2588-2595.
18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P. A method and server for predicting damaging missense mutations. Nat Methods 2010; 7: 248-249.
19. Saunders CT, Baker D. Evaluation of structural and evolutionary contributions to deleterious mutation prediction. J Mol Bio 2012; 322: 891-901.
20. Roy A, Yang J, Zhang. Cofactor: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res 2012; 40: W471–W477.
21. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformat 2008; 9: 40.
22. Kelley LA, Sternberg MJ. Protein structure prediction on the web: a case study using the PHYRE server. Nature Protoc 2009; 4: 363-371.

23. Alanazi M, Abduljaleel Z, Khan W, Warsy AS, Elrobh M, Khan Z, Amri AA, Bazzi, MD. In silico analysis of single nucleotide polymorphism (SNPs) in human beta-globin gene. PLoS One 2011; 6(10), e25876.

24. Franceschini A et al. STRING v9.1: protein–protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013; 41: D808-D815.

Cite this article as:
Dibyajyoti Rabha, Dipankar Baruah, Paresh Kumar Sarma, Jatin Sarmah. In Silico Analysis of ß-Thalassemia Mutations in India and its Neighbouring South East Asian Countries. J Pharm Chem Biol Sci 2018; 6(3): 188-198