Full Length Research Article

Advancements in Life Sciences – International Quarterly Journal of Biological Sciences

Open Access

In Silico Analysis of CYP1B1 Gene Mutations in Pakistani Families Reveals Structural and Functional Characteristics

Muhammad Adnan Shan[1], Maha Munir[1], Muhammad Umer Khan[2]*, Raima Rehman[2], Tazeen Zahid[2], Muhammad Ikram Ullah[3], Samiullah Khan[4], Ayman Ali Mohammed Alameen[3], Aisha Farhana[3]

Author's Affiliation:
1. Center for Molecular Biology, University of the Punjab, Lahore – Pakistan
2. Institute of Molecular Biology and Biotechnology, University of Lahore, Lahore – Pakistan
3. Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Jouf University, Sakaka-72388, Aljouf – Saudi Arabia.
4. Faculty of Health and Life Sciences, Persiaran Perdana BBN Putra Nilai, 71800 Nilai, Negeri Sembilan – Malaysia

*Corresponding Author:
Muhammad Umer Khan
Email:
muhammad.umer4@mlt.uol.edu.pk

Abstract

Background: Mutations in the CYP1B1 gene have been associated with primary congenital glaucoma (PCG), a developmental visual disorder. CYP1B1 mutations have been linked to increased lipid peroxidation and abnormalities in trabecular meshwork. In this study, protein modeling was utilized to analyze the pathogenicity of hotspot mutations in the CYP1B1 gene in Pakistani families with PCG.

Methods: A comprehensive literature survey was conducted to identify CYP1B1 mutations in Pakistani Patients. Ensemble was used to identify missense mutations. 3-D modeling of the mutants was done using I-Tasser. Potential deleterious effects and pathogenicity were evaluated using PolyPhen-2. The effects of amino acid substitutions on protein function were assessed using SIFT analysis, scoring pathogenic mutations from 0.0 to 1.0 in decreasing order. The secondary structure of the mutated protein was predicted using PSIRED. MusiteDeep server identified potential phosphorylation sites, while Pfam database examined the potentially mutated domains. The potential effects on other genes were evaluated using Gene MANIA server and STRING databases.

Results: A total of 5 hotspot mutations—4 missense (c.716C>G/p. A115P, c.988G>T/p. A330F, c.355G>T/p. A119S, and c.1058C>T/p. E229K) and 1 frameshift mutation (c.1325delC/p. P442Efs*15)—were identified among Pakistani patients. A119S was the most common mutation. 3-D modeling of the mutant proteins using i-Tasser revealed the global topology of the variants. Moreover, the phylogenetic tree of the CYP1B1 gene revealed its closest association with Phodopus roborovskii. Secondary structure of the mutant had six alpha helices, six coils, and six extracellular structures. Potential phosphorylation was identified in two nsSNP residues.

Conclusion: This in silico analysis contributes to understanding the structural and functional implications of CYP1B1 gene mutations. These insights can potentially help in the development of new targeted and personalized treatments for PCG patients.

Introduction

Glaucoma refers to a diverse range of eye conditions marked by increased pressure on the eye, known as intraocular pressure (IOP), and damage to the optic nerve. This condition causes vision to decrease gradually and, if left untreated, may eventually lead to complete blindness [1,2]. Glaucoma is diverse class of neurodegenerative disorders which is characterized by unusual form of optic neuritis which gradually leads to deterioration and ultimate death of retinal nerve fiber layer. Such condition eventually results in substantial vision impairment [3]. A breakdown in the mechanotransduction process of eye chambers causes structural alterations, poor aqueous humors drainage, along with trabeculodysgenesis in glaucoma. Retinal irritation and activation of Muller cells, microglia, as well as astroglia are caused by such dysfunction. These variations may ultimately result in progressive eyesight loss [4]. The disease known as primary congenital glaucoma (PCG) is severe visual impairment which caused by defects in development of anterior chamber angle as well as optic nerve head. Generally, it appears in the ninth month of pregnancy. There is no known etiology for PCG, and it is not related to any additional developmental disorders. Light sensitivity, increased intraocular pressure, weakening and extension of the anterior sclera, excessive weeping, inflammation of the eyelids, optic atrophy, corneal haze, as well as globe enlargement are amongst the symptoms. Complications may include discernment membrane deterioration and conjunctival erythema. PCG can occur sporadically or run in families, and genetic counseling is recommended to identify carriers of defective genes and prevent further visual impairment [5,6].

A genetically diverse disorder can occur sporadically or be inherited as an autosomal recessive trait within families [3]. Genetic studies on PCG began in the early 1990s, and the first locus for PCG was mapped to chromosome 2p21 in 1995. Additional loci identified include GLC3A (2p21), GLC3B, GLC3C (14q24.3), GLC3D (14q24.2-q24.3), and GLC3E (9p21.2) [7]. Whereas only three genes, namely Cytochrome P4501B1 (*CYP1B1*), latent transforming growth factor-beta binding protein 2 (*LTBP2*), and Tunica interna endothelial cell kinase (TEK), are directly implicated in the development of PCG; however, the function of the protein produced by these genes is unknown [7]. According to Abu-Amero et al. [8], PCG generated by heterozygous pathogenic mutations in TEK has an autosomal dominant inheritance pattern; however, PCG caused by *CYP1B1* or *LTBP2* has an autosomal recessive inheritance pattern.

The trabecular meshwork (TM) in the eye develops and functions appropriately only when *CYP1B1* is expressed [9]. PCG in the humans has been linked to mutations in the *CYP1B1*. In mouse knockdown experiments, decreased CYP1B1 has been confirmed to result in decreased TM collagen along with increased lesions in TM endothelial cells along with collagen. Such results point to relationship among high intraocular pressure, a crucial feature of glaucoma, as well as morphological alterations in trabecular meshwork (TM).

Furthermore, TM tissue anomalies, for instance elevated reactive oxygen species, were seen in CYP1B1 knockout mice, along with reduced periostin expression along with increased lipid peroxidation [9-11]. Capillary morphogenesis as well as angiogenesis are decreased when CYP1B1 is removed in endothelial cells because of amplified oxidative stress. CYP1B1 is vital for normal trabecular meshwork function, along with this dysregulation may contribute to the glaucoma as well as other eye diseases [10].

PCG can become evident soon after birth also varies in frequency across several demographic groups. Studies on PCG-associated mutations have been conducted in Iran, India, Saudi Arabia, South Korea, Pakistan as well as Lebanon, among other nations. The p. Arg390 in the Pakistan with 43% of PCG cases is being triggered by his mutation, it was the most predominant. Iran along with India was likewise testified to have lower frequency of such mutation. In Saudi Arabia, Pakistan, Iran, and Lebanon, as well as in Pakistan, p. Glu229Lys, another mutation has been found. The need of genetic testing for afflicted individuals along with their families is highlighted by such findings, which show regional as well as ethnic variances in PCG-related mutations.[8,12].

When consanguineal parents are present, frequency rises five to several times. About 70% of the patients faces damage to both eyes, with men frequently experiencing this [13]. PCG is declining variant of an incurable illness for which further study is compulsory to completely understand the pathophysiology. PCG as well as other forms of the glaucoma account for about 13.5% of instances of the blindness worldwide. Primary open-angle glaucoma (POAG) is one of the most common form of glaucoma, making up more than 50% of the cases worldwide [14].

It is challenging to comprehend the disease mechanism along with mode of inheritance in the PCG patients with *CYP1B1* mutations because of complexity and the variable expression of such cases. For better understanding etiology of PCG, it is critical to examine the functional significances of *CYP1B1* mutation along with regulatory variants. This information could support in identifying potential therapeutic targets and in executing proactive disease management plans. Eventually, further research on PCG along with *CYP1B1*

mutations is essential to improve our understanding of such condition [15].

The contemporary state of knowledge on genetic basis of the primary congenital glaucoma in Pakistani population is insufficient because of predominance of research conducted on cohorts from countries other than the Pakistan. To address this gap, it is critical to conduct wide-ranging genetic testing in Pakistan to recognize entire variety of the mutations. Furthermore, the development of progressive as well as tailored diagnostic methods is of chief importance for effective detection of the *CYP1B1* mutations in Pakistani patients with primary congenital glaucoma. This is crucial because current screening techniques may overlook noteworthy variations, leading to delayed interventions. The present study required to explore the functional consequences of these genetic variants on *CYP1B1* and assess their potential pathogenicity. The primary objective of this study was to employ computational tools to find hotspot as well as low-frequency *CYP1B1* gene variants in Pakistani population and evaluate their potential to cause disease.

Methods

Analysis for *CYP1B1* variants
Numerous search tools, which include Science Direct, PubMed, as well as Google Scholar, were used for literature searches using the keywords like "*CYP1B1* mutations," "primary congenital glaucoma," along with "Pakistani population." Only studies published from 2015 onwards were considered to identify mutation variants. The examined literature focused on the frequency of recognized variants and their originality within the Pakistani population.

In silico analysis:
The damaging effect of non-synonymous mutations on the structure and function of CYP1B1 was predicted and evaluated using various bioinformatics tools.

Data Collection
All *CYP1B1* data were collected from a web-based source, the National Centre for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/snp/). A comprehensive dataset of SNP mutations in *CYP1B1* was retrieved from the NCBI dbSNP database (https://www.ncbi.nlm.nih.gov/). The protein sequence (Q16678) of *CYP1B1* gene was retrieved from the UniProt KB database (https://www.uniprot.org/).

Effects of variants on the relevant protein:
The effects of variants on the relevant proteins as missense mutations were evaluated using the Ensembl transcript variant database (http://asia.ensembl.org/index.html). *CYP1B1* gene variants were converted to relevant protein sequences using the Expasy Translate tool (https://web.expasy.org/translate/). 3-dimensional protein structures of the identified variants were constructed using the I-Tasser tool (https://zhanggroup.org/I-TASSER/) [16].

Predicting Functional Effects of nsSNPs:
The functional effects of non-synonymous single nucleotide polymorphisms (nsSNPs) in the *CYP1B1* gene were analyzed using PolyPhen-2, an *in silico* tool. The pathogenicity of the identified gene sequence variants was assessed using PolyPhen-2 server (http://genetics.bwh.harvard.edu/pph2/), which employs an iterative greedy algorithm to nominate sequences and structure-based analytical attributes [17]. Variants were deemed pathogenic if their score fell within the "probably damaging" range of 0.85-1.0 or the "possibly damaging" range of 0.20-0.85. The PolyPhen-2 score ranged from 1.0 to 0.0, with a score of 0-0.20 indicating a benign variant [18].

Sorting Intolerant from Tolerant:
The SIFT algorithm (http://sift.jcvi.org/) was applied to distinguish between amino acid substitutions that negatively affected protein function and those that had no effect or even enhanced protein function [19]. This was done by evaluating the physical and chemical properties of the amino acids and sequence similarity of the protein sequence. The SIFT server also helps connect genetic mutations and phenotypic changes. It can determine whether a mutation is harmful and predict the impact of amino acid substitutions on protein function using an SIFT score ranging from 0 to 1. The reference sequence identity of the non-synonymous nucleotide polymorphisms was submitted to the server to predict changes in amino acid function.

Phylogenetic Conservation Analysis:
The phylogenetic or evolutionary conservation score of an amino acid significantly influences its structure and function. This type of protein attribute is essential for interpreting four International Journal of Genomics mutations that have a deleterious effect. Hence, Multiple sequence alignment and phylogenetic trees of the sequences were created using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and Clustal W (https://www.genome.jp/tools-bin/clustalw) tools, with the default parameters for both tools [20,21].

Protein Stability Analysis:
The MUpro server (http://mupro.proteomics.ics.uci.edu/) was utilized for the stability analysis of the protein structure by

employing SVM and neural network algorithms [22]. This predicted the impact of amino acid substitutions on protein stability based on the confidence scores assigned to negative or positive numerical values. A negative score indicates decreased stability, whereas a positive score indicates increased stability [22]. The SVM tool predicted changes in protein stability for single amino acid mutations, considering both the protein sequence and structural attributes based on the ΔΔG values. To use the server, the protein sequence of CYP1B1 was obtained, and the substitution positions with wild-type and mutant residues were submitted to the appropriate boxes.

Detecting the Presence of Deleterious mutations in Secondary Structure:

The presence of harmful single nucleotide polymorphisms (nsSNPs) in secondary structures refers to the presence of deleterious nsSNPs in stabilized hydrogen-bonded protein structures, which link primary sequences to tertiary structures. A two-phase neural network was employed to analyze secondary structures using a PSI-BLAST-based position-specific scoring system. The PSIPRED web server (https://bio.tools/psipred), which consists of two feedforward neural networks, is commonly used to predict secondary structures by investigating PSI-BLAST-predicted results [23]. The FASTA sequence was entered into a server to predict the secondary structure of CYP1B1.

Presence in Functional Domain:

The functional domains of a protein were predicted using the Pfam database (https://www.ebi.ac.uk/interpro/entry/pfam/#table).
Pfam is a comprehensive database that includes protein families and domains and is typically used to analyze novel genomes and monitor research operations on specific proteins [24]. It contains multiple sequence alignments and profile hidden Markov models (HMMs) to identify these functional domains in new protein sequences. To discover the functional domains of a protein, the accession or protein ID can be entered into the server.

Prediction of Post-Translational Modification Sites

Post-translational modifications (PTMs) involve changes in amino acids through the addition of various side chains in proteins. These modifications are essential for the management of various cellular activities, and have a significant impact on the structure and function of proteins. Interference with PTM sites can disrupt important biological mechanisms, leading to various diseases. In this study, the MusiteDeep server (https://www.musite.net/) was used to predict PTM sites. MusiteDeep is an online tool

that employs a deep-learning framework for the prediction and visualization of PTM sites in proteins [25]. The FASTA sequence of CYP1B1 protein was submitted to the server to predict various PTM sites.

Analyzing Networking of MET with Other Genes:

Gene MANIA is a web-based tool (https://genemania.org/) that analyzes and identifies associated genes with a given set of input genes by utilizing a vast array of functional association data, including protein and genetic interactions, pathways, co-expression, co-localization, and protein domain similarity [26]. This tool serves as a flexible and accessible platform for generating hypotheses on gene function, examining gene lists, and organizing genes for functional testing. The input gene name "*CYP1B1*" is used in this instance.

Predicting Interactions of MET with Other Proteins:

The Search Tool for the Retrieval of Interacting Genes (STRING) is an online database (https://string-db.org/, version 11.5) that connects gene-gene and/or protein-protein interactions into a network framework, complete with confidence scores [27]. The protein network included high-confidence interactors with scores ⩾0.700 to minimize false negatives and positives. To use the database, protein sequences were submitted, and the "minimum required interaction score" was set to "High confidence" (0.700).

Results

Literature analysis mapped different mutations in Pakistani families affected by primary congenital glaucoma to the *CYP1B1* gene. Types

Types of mutations:

Mutation analysis revealed five different hotspot mutations: four missense mutations (c.716C>G/ p.A115P, c.988G>T/ p.A330F, c.355G>T/ p.A119S, c.1058C>T/ p.E229K) and one frameshift (c.1325delC/ p. P442Efs*15). To assess the pathogenicity of each of the variations, *in silico* analyses were performed.

LOMET- a multiple threading

Also, additional structural templates from the Protein Data Bank (PDB) were identified through the use of the multiple threading approach known as LOMETS. These templates were utilized in the construction of full-length atomic models using iterative template-based fragment assembly simulations using I-Tasser.

PolyPhen-2

In addition, the analysis of all five identified mutations was conducted using Polyphen-2 in order to assess the potential detrimental impact of these changes on the protein's function and structure. PolyPhen-2 classified the A330F and E229K mutations as potentially harmful,

as they were projected to have a score closer to one (figure 1A).

SIFT algorithm- approach for mutational analysis
Simultaneously, the A119S exhibited a benign nature, as indicated by its score of 0.0006. The analysis of mutational variation was extended by employing the SIFT algorithm to ascertain the consequences of amino acid substitutions on the functional properties of the protein. A SIFT score of 1.00 was achieved through the examination of all six mutations, revealing that the replacement can be altered without impacting the functionality of the protein. It is important to note that only mutations scoring between 0.0 and 0.05 are considered to have detrimental consequences (figure 1B).

Clustal Omega and Clustal W
Clustal Omega and Clustal W were used to do additional multiple sequence alignment of CYP1B1 proteins. CYP1B1 proteins show conservation of residues at 115, 330, 229, and 442 places among seven species. Thus, three missense and one frameshift variant-affected amino acid residues were highly conserved in CYP1B1 protein orthologs. In contrast, it was 119-position in Homo sapiens, code for proline, alanine, and arginine in other species (figure 1C).
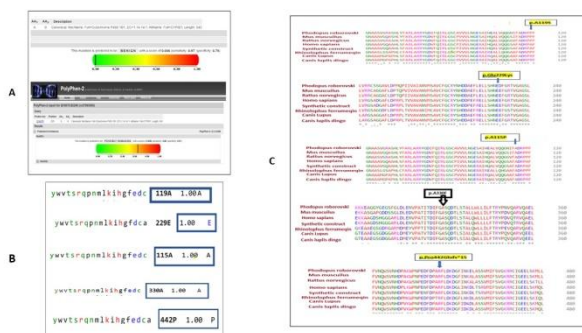

Figure 1: (A) Polyphen-2 described E229K as possibly damaging and A119S as benign. (B) SIFT described as all mutations are tolerated as score is more than 0.05. (C) Multiple sequence alignment of the CYP1B1 proteins of the affected persons with eight various species.

Clustal W showed the maximum ancestral relation to homo sapiens. *Canes Lupus Dingo*, among seven species, was the least similar to Homo *sapiens*. In contrast, *Mus Musculus* was at the end that showed ancestral relations with Homo sapiens in the phylogenetic tree (figure 2A).

Ensemble- approach for 3D analysis of protein
In the end, a three-dimensional structural analysis of the CYP1B1 protein was done using Ensemble to investigate the potential impact of discovered mutation variations on the three-dimensional structure of the encoded protein (Figure 2B). Ensemble was used to

analyses the effects of variations on the linked protein, and the resulting type was displayed as missense. Also, Three-dimensional protein crystal structure of CYP1B1 was visualized by Discovery Studio [28] (Figure 2C).
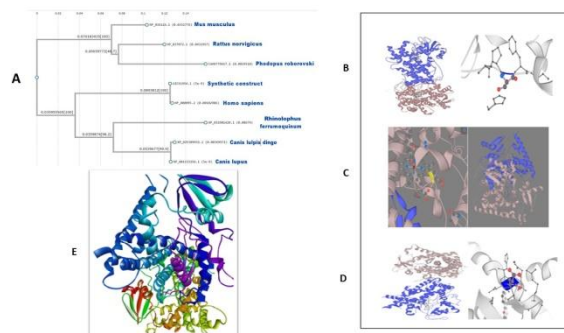
Figure 2: (A) Phylogenetic tree derived using the Clustal W algorithm illustrates the graphical depiction of the evolutionary relationships among these species. The synthetic construct and *Phodopus Roboroveskli* exhibit a significant degree of proximity to Homo sapiens, reflecting the maximum ancestral relationship with them. (B) Three-dimensional protein model by Ensemble of c.355 G > T A119S, (b) Three-dimensional protein model of c.716C>G/ p.A115P, and (c) Three-dimensional protein model of E229K. (C) Three-dimensional protein crystal structure of CYP1B1 by Discovery Studio.

Mu-Pro Analysis
Lastly, Mu-Pro is a computational tool for predicting the effects of single-site mutations on protein stability and protein-protein interactions and all mutation showed decrease stability (Figure 3A).

Predicting the Effects of nsSNPs on the Secondary Structure of MET
The PSIPRED server was used to predict the secondary structures of the CYP1B1 protein using PSI-BLAST. The predicted sequence outline indicated the organization of the alpha helix, beta sheet, extracellular, and coil structures (Figure 3B). Following the analysis of the 18 amino acid substitutions using the PSIPRED server, three distinct types of secondary structures were identified. Of the 18 residues, six (33.33%) were predicted to be in a "coil" structure, 6 (33.33%) were predicted to be in an "alpha helix" structure, and 6 (33.33%) were predicted to be in an "extracellular" structure.

Effects of Deleterious nsSNPs on Functional Domains of *CYP1B1*
The Pfam database provides information on the domains of CYP1B1 protein using its protein ID. In total, 18,869 species were included in the analysis (Figure 3C).
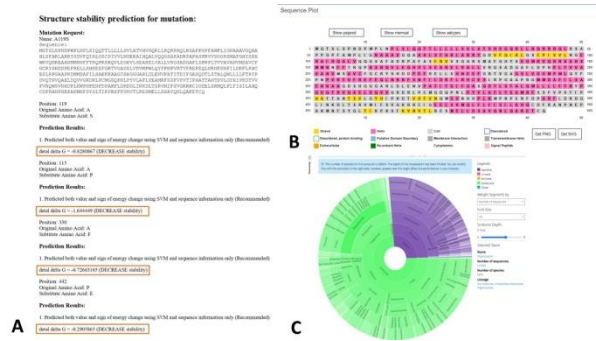
Figure 3: (A) Mu-Pro shows the decrease stability prediction for all mutations. (B) PSIPRED sequence plot of CYP1B1. The secondary sequence of CYP1B1 protein is color coded as pink, yellow and silver codes, representing helix, strand and coil structures, respectively. (C) Taxonomy of CYP1B1 protein.

Predicting the Impacts of nsSNPs on PTM Sites

The MusiteDeep server identifies potential phosphorylation sites (PTMs) on proteins by analyzing protein sequences and provides real-time estimation of multiple proteins. The output is at the level of individual amino acid residues for the selected PTM sites, with a focus on the phosphorylation of serine (S), threonine (T), and tyrosine (Y) residues.

This study analyzed the phosphorylation sites of CYP1B1 and identified amino acid residues with a score greater than 0.5. However, only two residues were identified as phosphorylated and were associated with highly deleterious non-synonymous single nucleotide polymorphisms (nsSNPs) among these residues (Figure 4A).

## Analysis of Effects of Deleterious Mutations on Other Genes

The Gene MANIA server is a comprehensive database that features interaction networks from multiple data sources, with a focus on identifying functionally connected genes (Figure 4B, C). By analyzing the gene-gene interactions of *CYP1B1*, the server was able to create a network of related genes. Gene MANIA's prediction for *CYP1B1* indicates that the gene interacts with at least 20 other genes, and mutations in these genes may impair their functions.

AA3F- an allele in maintaining core feature of *CYP1B1* gene

It is also worth noting that A330F has been detected in a single PCG patient worldwide; it is regarded as an essential allele in maintaining the core features of the *CYP1B1* gene. The frameshift mutation Pro442Glufs*15 has been found in a homozygous state in the *CYP1B1* genome.

Firstly, I-TASSER algorithm tool was used to generate multiple structural models and assess their quality based on various scoring functions, in which

structures (a) with a C score of 0.26 and (b) with a C score of 0.01 show global topology (figure 5A).
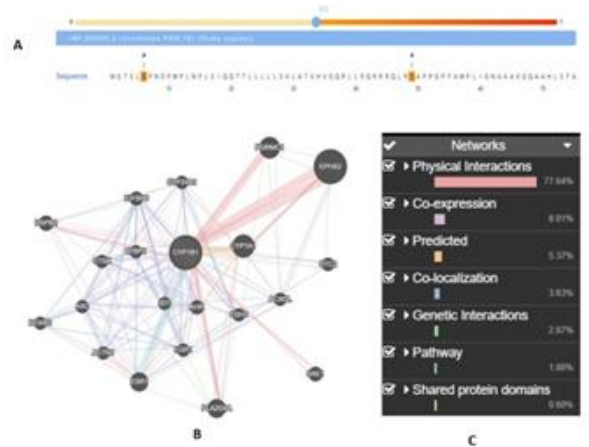
Figure 4: (A) Post-translational modification site prediction of protein CYP1B1. Highlighted serine (S) residues represent the sites of post-translational phosphorylation in CYP1B1 protein. (B) Interaction of *CYP1B1* gene with other genes.

Analysis of the Effects of Deleterious Mutations on Other Proteins

Interactions between proteins were examined to understand the functional connections among cellular proteins. STRING database prediction indicated that the CYP1B1 protein interacts with ten other proteins, whose functions may also be affected by these mutations.
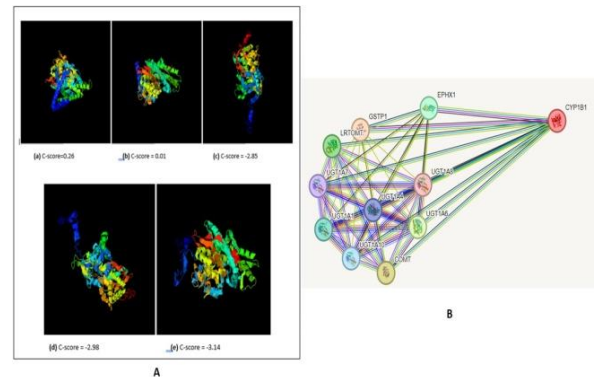
Figure 5: (A) (a) and (b) C -score showed the global topology of the *CYP1B1* gene and (B) shows Graphic representation of Protein–protein interaction network of CYP1B1 protein obtained from STRING database.

(Figure. 5B) illustrates that CYP1B1 interacts with ten different proteins, including Glutathione S-transferase P, Catechol O-methyltransferase, UDP-glucuronosyltransferase 1-6, Transmembrane O-methyltransferase, Epoxide hydrolase 1, UDP-glucuronosyltransferase 1-1, UDP-glucuronosyltransferase 1-2, UDP-glucuronosyltransferase 1-3, UDP-glucuronosyltransferase 1-4, UDP-

glucuronosyltransferase 1-7, and UDP-glucuronosyltransferase 1-8.

Discussion

In this in silico analysis, we investigated the mutational spectrum of the *CYP1B1* gene in Pakistani sufferers laid low with number one congenital glaucoma (PCG) [29]. These findings are consistent with previous research, which has found various mutations within the *CYP1B1* gene associated with PCG in different populations. The evaluation of 50 articles from 2015 to 2023 found out five specific mutations inside the *CYP1B1* gene: c.716C>G/p.A115P, c.988G>T/p.A330F, c.355G>T/p.A119S, c.1058C>T/p.E229K, and c.1325delC/p.P442Efs*15 [2,14,29-31]. These mutations have been diagnosed via complete examination and analysis of PCG sufferers from Pakistani families.

The c.1058C>T/p. E229K mutation became determined to be a singular mutation related to PCG. This mutation became also detected in Iraqi patients, suggesting its presence beyond the Pakistani population [14].

In silico evaluation using gear together with I-TASSER, Polyphen-2, SIFT, Clustal Omega, Clustal W, and Mu-Pro supplied insights into the pathogenicity, structural impact, and conservation of those mutations [2]. Polyphen-2 predicted that the A330F and E229K mutations were possibly to have deleterious effects on the protein structure, whilst A119S turned out to be expected to be benign. SIFT evaluation indicated that each of the six mutations had excessive tolerance rankings, suggesting that the substitutions may not significantly affect protein characteristics [30].

Multiple sequence alignment the use of Clustal Omega and Clustal W revealed that the affected amino acid residues had been conserved among unique species, highlighting their purposeful importance. This conservation provides similar proof of the function of these residues within the structure and characteristics of the CYP1B1 protein. Three-dimensional structural evaluation using Ensemble showed missense modifications within the protein structure because of the identified mutations. In addition to this, Mu-Pro evaluation anticipated a decline in balance for all the mutations, signifying capability disturbances in the protein balance as well as protein-protein interactions [32].Such discoveries enhance our knowledge of genetic terrain of the PCG, along with this it shed light on conceivable pathways which may underlie the illness. The assorted array of the mutations present within the *CYP1B1* underscores its fundamental role in pathogenesis of the PCG, while at the same time revealing the disparities which exist between numerous populations with regard to prevalence of such mutations [33]. Identifying the imperative nature of in silico evaluation, it is important to acknowledge the trials which are associated with this particular approach. Despite these challenges, insights derived from the analyses of in silico can notify future research also contribute to development of the PCG genetic testing protocols, that necessitates the need for the experimental authentication to confirm its utility [31].

The discovery of the *CYP1B1* gene variations contributes to our understanding of genetic basis of the PCG as well as provides insights into its pathogenesis. Development of the genetic diagnostic methods for the PCG also the potential implications for personalized treatment strategies are supported by such findings. Though, it is important to identify that in silico analysis has limitations as well as experimental validation is important to authorize the functional significance of such changes. On comparing our results with those of prior studies, we detected both likenesses and variances. The A119S mutation, that was previously thought to be benign, was also perceived in unrelated PCG patients from the Brazil. The E229K mutation was formerly reported in Pakistani families also it was related with a severe phenotype [1]. These findings highlight the diversity of *CYP1B1* mutations in particular populations and their erratic effects on disease presentation. This study carried out an widespread analysis of five *CYP1B1* gene mutations recognized in Pakistani patients, out of which four were missense (c.716C>G/p. A115P, c.988G>T/p.A330F, c.355G>T/p.A119S, c.1058C>T/p.E229K), and the one is frameshift mutation (c.1325delC/p.P442Efs*15) with A119S being the most dominant variant. The mutation with the maximum frequency was A119S. This study used i-Tasser for structural modeling, PolyPhen-2 for functional forecasts, and MUpro for stability assessments. These analyses show that these mutations have noteworthy unfavorable effects and decline protein stability. Though the SIFT score of 1.0 suggests that A119S is the silent mutation with no change in sequence of amino acid, the structural changes specify interruption of protein function. The multiple sequence alignments highlighted p.A119 as the mutation hotspot, whereas phylogenetic investigation showed a close connotation with *Phodopus Roborovskii*, underscoring evolutionary conservation. Further analysis of the mutated protein revealed three distinct secondary structures, alpha helix, coils, and extracellular structures, in the protein. Moreover, MusiteDeep server identifies two new highly deleterious nsSNP as potential phosphorylation sites (PTMs). These mutations were identified to impact the interactions of the protein with atleast ten other genes and their function. Such findings enhance our understanding of genetic landscape of *CYP1B1*, and highlight the need for the further *in vitro* and *in vivo* validation, offering potential pathways for the

diagnostic as well as therapeutic advancements in the *CYP1B1*-related diseases.

Author Contributions

Muhammad Adnan Shan and Maha Munir wrote the initial manuscript. Raima Rehman and Tazeen Zahid performed the *Insilco* analysis. Samiullah Khan and Muhammad Ikram Ullah revised and finalized the manuscript. Ayman Ali Mohammed Alameen and Aisha Farhana constructed the figures and critically reviewed the manuscript. Muhammad Umer Khan conceptualized and supervised the research project. All the authors approved the final version of the manuscript

Conflict of Interest

All the authors declare no competing interests, either financial or personal.

References

1. Curry SM, Daou AG, Hermanns P, Molinari A, Lewis RA, et al. Cytochrome P4501B1 mutations cause only part of primary congenital glaucoma in Ecuador. Ophthalmic Genetics, (2004); 25(1): 3-9.
2. Zahid T, Khan MU, Zulfiqar A, Jawad F, Saleem A, et al. Investigation of mutational spectrum in cytochrome P4501B1 (CYP1B1) as the principal cause of primary congenital glaucoma. Pakistan Journal of Medical Sciences, (2023); 39(2): 409.
3. Rauf B, Irum B, Kabir F, Firasat S, Naeem MA, et al. A spectrum of CYP1B1 mutations associated with primary congenital glaucoma in families of Pakistani descent. Human genome variation, (2016); 3(1): 1-4.
4. Križaj DJWTOotR. What is glaucoma? Webvision: The Organization of the Retina Visual System, (1995); ID: NBK543075
5. Jin SW, Ryu WY. Clinical manifestations of strabismus in patients with primary congenital glaucoma; 2019. Taylor & Francis. pp. 451-457.
6. Pedersen KB, Kappelgaard P, Kessel L, Sandfeld L, Zibrandtsen N, et al. Primary congenital glaucoma in Denmark, 1977–2016. Acta Ophthalmologica, (2020); 98(2): 182-189.
7. Imran A, Khan MU, Nasir U, Qayyum Q, Hector R, et al. The genetics associated with Primary Congenital Glaucoma. Advancements in Life Sciences (2020); 7(2): 106-112.
8. Abu-Amero KK, Kondkar AA, Khan AO. Molecular karyotyping of a dysmorphic girl from Saudi Arabia with CYP1B1-negative primary congenital glaucoma. Ophthalmic Genetics, (2016); 37(1): 98-101.
9. Zhao Y, Sorenson CM, Sheibani N. Cytochrome P450 1B1 and primary congenital glaucoma. Journal of Ophthalmic Vision Research, (2015); 10(1): 60.
10. Shah BR, Xu W, Mraz J. Cytochrome P450 1B1: Role in health and disease and effect of nutrition on its expression. RSC Advances, (2019); 9(36): 21050-21062.
11. Tehreem R, Arooj A, Siddiqui SN, Naz S, Afshan K, et al. Mutation screening of the CYP1B1 gene reveals thirteen novel disease-causing variants in consanguineous Pakistani families causing primary congenital glaucoma. PloS one, (2022); 17(9): e0274335.
12. Haddad A, Ait Boujmia OK, El Maaloum L, Dehbi H. Analysis of CYP1B1 gene mutations in primary congenital glaucoma patients. European Journal of Ophthalmology, (2021); 31(6): 2796-2807.
13. Hitchings R, Migdal C, Bechetoille A. Terminology and guidelines for glaucoma. European Glaucoma Society (4th edition) Savona: PubliComm, (2014); 79-90.
14. Khan MU, Rehman R, Kaul H, Mahmood S, Ammar A. Mutational analysis of CYP1B1 gene in Pakistani pediatric patients affected with Primary Congenital Glaucoma. Advancements in Life Sciences, (2019); 7(1): 32-37.
15. Faiq MA, Dada R, Qadri R, Dada T. CYP1B1-mediated pathobiology of primary congenital glaucoma. Journal of Current Glaucoma Practice, (2015); 9(3): 77.
16. Yang J, Zhang Y. Protein Structure and Function Prediction Using I-TASSER. Current Protocols in Bioinformatics, (2015); 52:5.8.1–5.815
17. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics, (2013); Chapter 7Unit7.20.
18. Wiggs JL, Langgurth AM, Allen KF. Carrier frequency of CYP1B1 mutations in the United States (an American Ophthalmological Society thesis). Transactions of the American Ophthalmological Society, (2014); 11294-102.
19. Zhu D. SIFT algorithm analysis and optimization; 2010 9-11 April 2010. pp. 415-419.
20. 20. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular System Biology, (2011); 7539.
21. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, (1994); 22(22): 4673-4680.
22. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins: Structure, Function, and Bioinformatics, (2006); 62(4): 1125-1132.
23. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics, (2000); 16(4): 404-405.
24. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, et al. Pfam: The protein families database in 2021. Nucleic Acids Research, (2021); 49(D1): D412-D419.
25. Wang D, Liu D, Yuchi J, He F, Jiang Y, et al. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Research, (2020); 48(W1): W140-w146.
26. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Research, (2010); 38: W214-W220.
27. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Research, (2023); 51(D1): D638-d646.
28. Sakhawat A, Khan MU, Rehman R, Khan S, Shan MA, et al. Natural compound targeting BDNF V66M variant: insights from in silico docking and molecular analysis. Amb Express, (2023); 13(1): 134.
29. Rashid M, Yousaf S, Sheikh SA, Sajid Z, Shabbir AS, et al. Identities and frequencies of variants in CYP1B1 causing primary congenital glaucoma in Pakistan. Molecular vision, (2019); 25144.
30. Ou Z, Liu G, Liu W, Deng Y, Zheng L, et al. Bioinformatics analysis of CYP1B1 mutation hotspots in Chinese primary congenital glaucoma patients. Bioscience reports, (2018); 38(4): BSR20180056.

31. Sheikh SA, Waryah AM, Narsani AK, Shaikh H, Gilal IA, et al. Mutational spectrum of the CYP1B1 gene in Pakistani patients with primary congenital glaucoma: novel variants and genotype-phenotype correlations. Molecular vision, (2014); 20991.

32. Mashima Y, Suzuki Y, Sergeev Y, Ohtake Y, Tanino T, et al. Novel cytochrome P4501B1 (CYP1B1) gene mutations in Japanese patients with primary congenital glaucoma. Investigative ophthalmology visual science, (2001); 42(10): 2211-2216.

33. Khan MU, Tabassum W. An insight into primary congenital glaucoma. Critical Reviews™ in Eukaryotic Gene Expression, (2020); 30(1): 39-43.