• Basic Research •

Protein truncating splice site variant in *ALDH1A3* associated with bilateral anophthalmia identified in a multiplex consanguineous Pakistani family

Kashmala Samad[1], Khalid J. Alzahrani[2], Khalaf F. Alsharif[2], Fazeelat Samad[3], Farman Ullah[1], Muhammad Zeeshan Ali[1], Safeer Ahmad[1], Muhammad Muzammal[1], Sumra Wajid Abbasi[4], Jabbar Khan[5], Muzammil Ahmad Khan[1]

[1]Gomal Center of Biochemistry and Biotechnology, Gomal University, Dera Ismail Khan 29050, Pakistan
[2]Department of Clinical Laboratories Sciences, College of Applied Medical Sciences, Taif University, Taif 21944, Saudi Arabia
[3]Department of Pathology, Gomal Medical College, Dera Ismail Khan 37000, Pakistan
[4]Department of Biological Sciences, National University of Medical Sciences, Rawalpindi, Punjab 46000, Pakistan
[5]Institute of Biological Sciences, Gomal University, Dera Ismail Khan 29050, Pakistan
Correspondence to: Muzammil Ahmad Khan. Gomal Center of Biochemistry and Biotechnology, Gomal University, Dera Ismail Khan 29220, Pakistan. m.ahmad@gu.edu.pk
Received: 2025-03-11 Accepted: 2025-07-16

Abstract

● **AIM:** To identify the genetic factors underlying anophthalmia and microphthalmia (A/M), and to perform computational analysis to verify the pathophysiological mechanisms of the disease.

● **METHODS:** This study investigated a consanguineous Pakistani family with multiple affected individuals. Clinical evaluations were conducted using A-Scan and ophthalmic B-Scan ultrasonography (B-Scan). To identify the disease-causing variant, whole exome sequencing (WES) and Sanger sequencing were performed. *In silico* functional analyses were carried out using AlphaFold (for protein modeling) and ClusPro (for protein docking analysis) tools, and the hydrodynamic properties of the protein were determined *via* molecular dynamics simulations.

● **RESULTS:** Clinical analysis of the five patients revealed severe phenotypes of bilateral anophthalmia. Ocular B-Scan did not detect ocular tissue or intraocular fluid, thus confirming the diagnosis of anophthalmia in all patients. Due to these structural defects, all patients exhibited complete blindness and absence of light perception; additionally, two patients had mild to moderate intellectual disability. Genetic analysis identified a splice site variant [NM_000693.2: c.884-2_885dup; p.(Asp296SerfsTer35)] in the 9th exon of the *ALDH1A3* gene.

● **CONCLUSION:** The present study expands the genetic spectrum of *ALDH1A3* and contributes to establishing the genotype-phenotype correlation for this gene.

● **KEYWORDS:** anophthalmia; microphthalmia; whole exome sequencing; Sanger sequencing; *ALDH1A3*; Pakistani family

DOI:10.18240/ijo.2025.11.05

Citation: Samad K, Alzahrani KJ, Alsharif KF, Samad F, Ullah F, Ali MZ, Ahmad S, Muzammal M, Abbasi SW, Khan J, Khan MA. Protein truncating splice site variant in *ALDH1A3* associated with bilateral anophthalmia identified in a multiplex consanguineous Pakistani family. *Int J Ophthalmol* 2025;18(11):2048-2056

INTRODUCTION

Anophthalmia and microphthalmia are two of the most common congenital eye anomalies that can result in severe conditions and are often grouped as anophthalmia-microphthalmia (A/M). These anomalies are characterized by decreased eye volume within the orbit, which leads to vision loss[1]. Anophthalmia is defined as the complete or partial absence of some or all parts of the eye tissue, which is sometimes found to be associated with other ocular disorders as well. Microphthalmia is characterized by a reduction in the size of the eye along with other ocular anomalies, such as corneal opacity, corneal vascularization, and staphyloma. Clinically, anophthalmia and microphthalmia are two distinct genetic conditions that is separated by minor differential lines[1], where a globe with a total axial length less than two standard deviations from the age-related mean is associated with microphthalmia.

The A/M condition can be either syndromic or non-syndromic. In the syndromic form, ocular anomalies occur alongside systemic abnormalities affecting multiple organ

systems, often due to pleiotropic effect of the gene. These phenotypes may include neuropsychiatric issues, intellectual disabilities, coloboma, and optic nerve hypoplasia. Non-syndromic A/M are the sole eye abnormalities that occur independently, without associated systemic disorders. They result from disruptions in genes involved in eye development and function, without affecting other organ systems[2]. Both syndromic and non-syndromic variants of A//M are found to follow autosomal recessive, autosomal dominant, and X-linked patterns of inheritance, as well as shown substantial genetic heterogeneity[3].

A/M are relatively uncommon congenital anomalies, with combined prevalence rates ranging from 0.21 to 3.2 per 10 000 births[4] .Whereas, the projected prevalence rate of anophthalmia ranges from 0.20 to 0.40 per 10 000 births, and the prevalence rate of microphthalmia has been reported to range from 0.20 to 1.70 per 10 000 births depending on the intensity of orbital imaging. Both environmental and genetic variables may contribute to the etiology of microphthalmia[5]. Environmental variables, such as ophthalmic infection due to rubella, toxoplasmosis, cytomegalovirus (CMV), and maternal alcohol intake are thought to be the least common cause of microphthalmia[6]. Whereas aneuploidy, deletions, translocations, duplications, and trisomy 18 or trisomy 13 are among the several chromosomal disorders that can cause A/M[7].

Almost 30 genes have been reported to be involved in non-syndromic A/M families. Variants in several genes, including *RAX*, *PAX6*, *SOX2*, *OTX2*, *VSX2*, *RARB*, *BMP7*, *BCOR*, *BMP4*, *FOXE3*, *STRA6*, *SMOC1*, *SHH*, *SNX3*, *MFRP*, *PRSS56*, *GDF3*, *GDF6*, *TENM3*, *C12orf57*, *YAP1*, *ABCB6*, *ATOH7*, *VAX1*, *NDP*, *ALDH1A3*, and *SMARCA4*, have been related to microphthalmia. While, genes that are associated with anophthalmia specifically include *RAX*, *PAX6*, *SOX2*, *OTX2*, *RARB*, *BMP7*, *BCOR*, *BMP4*, *FOXE3*, *STRA6*, *SMOC1*, *GDF6*, and *ALDH1A3*[8-10].

The genetic studies have found *SOX2* as the major causative gene, which is responsible for 10%-15% of affected individuals. While, variants in other genes are responsible for around 25% of A/M cases[11]. The underlying genetic etiology remains unknown in up to 50%-60% of the cases. Keeping in view the importance of molecular and genetic diagnosis, impacting patient care, improving the counseling of patients, and treatment of A/M, this study analyzed a consanguineous Pakistani family, suffering from bilateral anophthalmia. Exome sequence analysis in this family found a splice site variant NM_000693.2:c.884-2_885dup:p.(Asp296SerfsTer35) in the *ALDH1A3* gene. *In-silico* study has confirmed the pathogenicity of the identified variant. The study has expanded the molecular spectrum of pathogenic *ALDH1A3* variants that are associated with A/M.

PARTICIPANTS AND METHODS

Ethical Approval The study protocol was approved by the institutional ethical review board of Gomal University, D.I.Khan, Pakistan (IRB# 32/DQA/GU). A written informed consent was obtained from each family member.

In the current molecular study, a consanguineous family from a nearby village of Dera Ismail Khan in Khyber Pakhtunkhwa province of Pakistan was recruited. The blood samples were obtained from cooperative & available individuals including patients, parents and normal sibling. Later, the salting-out technique was used to extract the DNA for genetic analysis[12].

Clinical Diagnosis Utilizing A-Scan and B-Scan ultrasounds, clinically ophthalmic investigation was conducted. The axial length and intraocular power of the lens were measured using the A-Scan, and the structural integrity of the retinal cells, vitreous, and optic nerve, among other ocular tissues, was examined using the B-Scan.

Whole Exome Sequencing and Candidate Variant Analysis Whole exome sequencing (WES) was performed on two affected individuals (IV-1 and V-5), Figure 1A, belonging to different consanguineous loops. The preparation of the sequencing library was carried out with the aid of xGen Exome Research Panel v2.0 Kit (Integrated DNA Technologies, Coralville, Iowa, USA), while the sequencing was carried out on NovaSeq 6000 (Illumina, San Diego, CA, USA). To generate the sequence reads, FASTQ data files were obtained by converting the base call files using bcl2fastq v2.20.0.422 tool. The Binary Alignment Map (BAM) files was obtained after the alignment of the sequence reads to the reference genome (GRCh37/hg19) using BWA-mem 0.7.17 (arXiv:1303.3997v2 [q-bio. GN]). Variant calling was performed using Genome Analysis Toolkit (GATK; version 3.8)[13]. Variant annotation was performed independently using wANNOVAR (https://wannovar.wglab.org/)[14] and Variant Studio software (v3.0). Variants were filtered and prioritized to find rare variants predicted to cause A/M in previously reported genes. Thereafter, multiple web tools were employed to predict the biological significance of candidate pathogenic variants. These tools include Provean[15], MutationTaster[16], and varSEAK[17].

Protein Structure Modeling and Docking A structural model of ALDH3A1 was created using the AlphaFold[18] , where the model with the highest confidence score (C-score) was considered for further interaction studies. Moreover, ClusPro was employed to predict the protein-protein interactions of ALDH3A1 with its close functional interactor[19]. The prediction was performed for both the wild-type and variant. The protein interaction network was determined using the STRING database[20]. Molecular visualization was performed using LigPlot+(version 2.1)[21], and PyMOL 2.3[22].

Table 1 *ALDH1A3* gene variant's biological predictions using various online tools

Gene	rs ID	c.variant	p.variant	Software tools	Prediction
ALDH1A3	-	c.884-2_885dup	p. Asp296SerfsTer35	Mutation taster VarSeak	Disease-causing Pathogenic

Segregation and Variant Analysis To validate the segregation of the identified variant with the disease, Sanger sequencing was performed on all available family members. Primers were designed using the Primer3web tool (version 4.1.0; https://primer3.ut.ee/). Sequence alignment was performed using the BLAT tool that is accessible from the UCSC Genome Browser[23], while the sequence chromatogram was visualized using the BioEdit tool (version 7.0.5). Furthermore, we also employed NMDEscPredictor tool (https://nmdprediction.shinyapps.io/nmdescpredictor/) to predict the effect of identified splice variant on mRNA and its possible NMD pathway activation.

Molecular Dynamics Simulation The molecular dynamics (MD) simulation, which helps in predicting the movement of atoms and molecules, was accomplished through the Amber tool[24]. The stability of the variant and wild-type allele was evaluated together with docked complexes by calculating the root-mean-square deviation (RMSD), the values of the carbon atoms comprising the backbone. The RMSD value, which measures the change between the initial and final structural conformations of a protein's backbone atoms, can be used to assess the protein's stability in relation to its conformation.

RESULTS

The current family was recruited from a remote area in Dera Ismail Khan city of Khyber Pakhtunwa. Ethnically, it belonged to a Saraiki language group. According to the documented pedigree, the family tree consisted of five generations, with three consanguineous loops, containing three affected individuals in the fifth generation, one in the fourth, and one in the third generation. The pedigree analysis revealed an autosomal recessive mode of disease inheritance (Figure 1A).

Clinical Description In the current study, five patients (III-1, IV-1, V-2, V-4, and V-5) were subjected to clinical examination. Apparently, all the patients had closed eyes due to the absence of eyeballs. Two patients (V-2 and V-5), who were siblings, had mild to moderate intellectual disability, while other patients had only anophthalmia (Figure 1B). The A-Scan and B-Scan were carried out, but neither eye tissue nor fluid was seen, hence no measurement could have been taken. These findings were suggestive of isolated bilateral anophthalmia phenotype in patients III-1, IV-1, and V-4. All patients passed the hearing exam and, were able to communicate normally, except V-2 & V-5 due to learning impairment. Apart from this, no neuromuscular phenotype was observed in any patient, as all patients had normal physique, normal walking, and no joint stiffness.

Molecular Findings A splice site variant in the 9th exon of the *ALDH1A3* gene, NM_000693.2:c.884-2_885dup: p.(Asp296SerfsTer35) was reported as a result of exome data analysis (Figure 1C). This splice site mutation suggestively led to skipping of exon, shifted reading frame, and created premature stop codon that eventually caused protein truncation. According to ACMG guideline, the identified splice site mutation lies in the category of "likely pathogenic". The variant is not reported in the population variant frequency databases and Pakistan Genetic Mutation Database (https://pakmutation.mikesoft.com.pk/). Due to protein truncation, the *ALDH1A3* suggestively losses multiple domains like catalytical domain, oligomerization domain and C-terminus domain. Sanger sequencing of available affected and normal individuals verified its perfect segregation with the disorder phenotype. Table 1 summarizes the biological predictions of the *ALDH1A3* gene variant using different software tools.

Structural and Interactional Findings of Wild-type and Variant *ALDH1A3* Protein The predicted 3D models of wild-type and variant *ALDH1A3* protein failed to superimpose with each other, which predicts the loss of its normal conformation due to splice variant (Figure 2A-2C). Moreover, significant alterations in the docking sites were also noticed during protein interaction studies involving the wild-type and variant *ALDH1A3* protein and their close interactor *CYP26B1* (Figure 2D, 2E). It was observed that wild-type protein interacted with the close interactor *CYP26B1* through 9 amino acids residues (Glu352, His128, Glu300, Tyr492, Glu491, Lys481, Lys266, Asp94, Lys446) with 10 H-bonds and 1 salt bridge, while variant *ALDH1A3* protein interacted with the close interactor *CYP26B1* through four different amino acids (Asn273, Leu74, His168, Tyr143) through 4 H-bonds. In a protein interaction network, it was determined that *CYP26B1* and *CYP26A1* were ranked as first and second interactors respectively, while CYP26C1 and RDH5 were ranked as third and fourth close interactors respectively by using the STRING database (Figure 2F). Hence, these *in silico*/computational investigations observed that wild-type and variant *ALDH1A3* proteins showed distinct interaction patterns in terms of residue number, identity, and bonding type. Due to severe structural defect of *ALDH1A3* protein, we speculated the activation of NMD pathway. Based on this notion, NMDEscPredictor tool was used, which significantly predicted the NMD pathway activation and degradation of *ALDH1A3* mRNA due to identified splice site mutation.

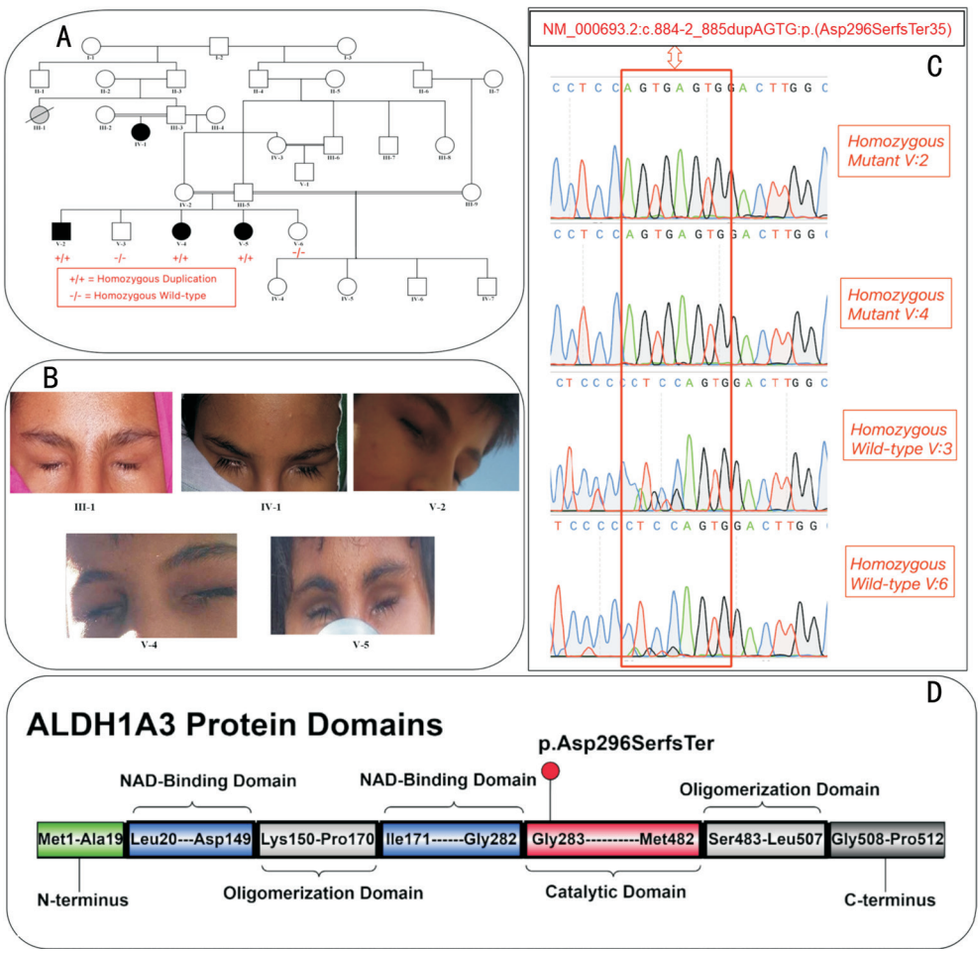


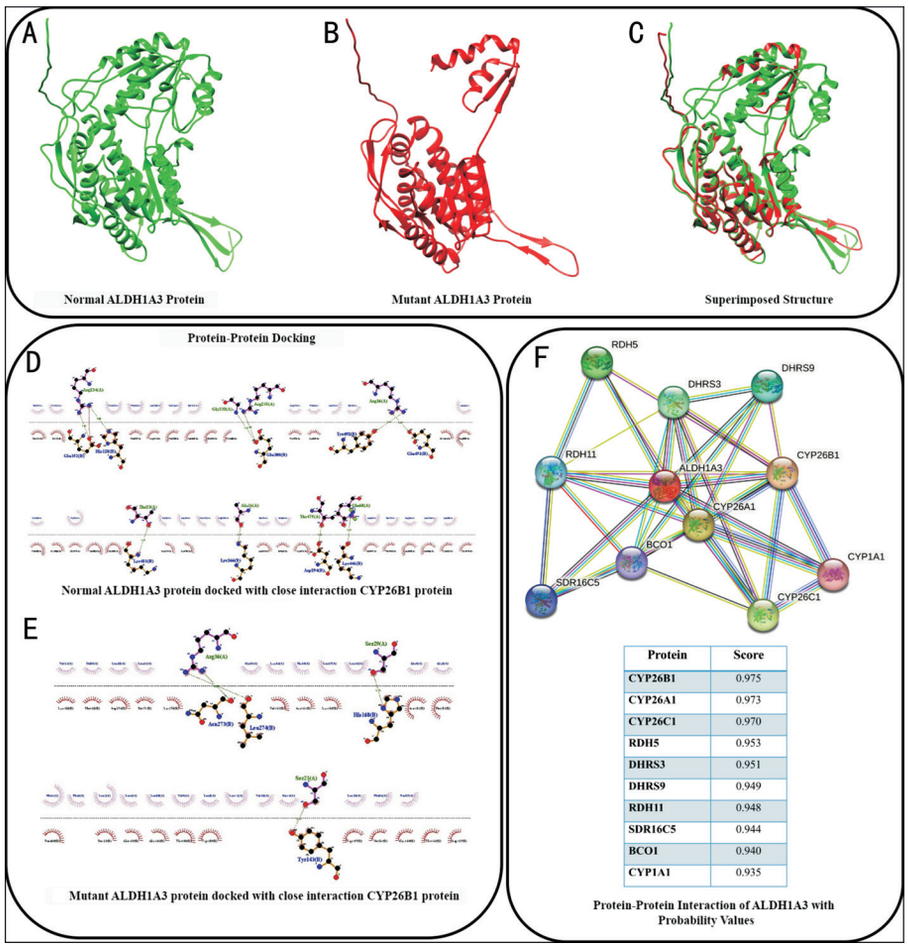
Figure 1 The image depicts family tree, patients photographs, sequence chromatogram and ALDH1A3 protein domains A: Pedigree of five generation consanguineous family along with genotype status of individuals showing homozygous variant (c.884-2_885dupAGTG) and wild-type status of normal and affected siblings; B: Patients' photographs; C: Sanger sequence chromatograms depicting homozygous duplication of "AGTG" at cDNA position 884-2_885; D: The image indicate various domains in ALDH1A3 protein with their respective positions. The pin indicates the position of current mutation.

Molecular Dynamics Simulation RMSD and RMFS were used to evaluate the MD simulation trajectories. In MD simulation, the RMSD for each protein-ligand docked complex was computed to determine the structural stability of every system. The RMSD plot, Figure 3A illustrates how the two systems under investigation fluctuated at first to different degrees before reaching equilibrium and acting consistently over the MD simulation. For the protein of the wild-type, the mean RMSD valve was measured at 10.09 Å, while for the variant, it was 3.48 Å. The highest RMSD values found were 5.19 Å for the variant protein and 13.12 Å for the wild-type protein.

The highest root mean square Fluctuation (RMSF) values, found for the variant and wild-type proteins, were 19.52 Å and 76.63 Å, respectively. However, the mean values for the variant and wild-type proteins were 3.17 Å and 44.09 Å, respectively. In the meantime, MD simulations of the wild-type protein revealed considerable fluctuation, instability, and high values of RMSD and RMSF in comparison to the variant, indicating that the variant found in *ALDH1A3* affected protein stability and function, leading to anophthalmia in the investigated patients. Thus. the outcomes of the exome investigation are further supported by the MD simulation analysis.

DISCUSSION

ALDH1A3 gene is a protein-coding gene that is located on chromosome 15q26.3. This gene encodes 512 amino acids long NAD-dependent aldehyde dehydrogenase protein. The catalytic activity of the aldehyde dehydrogenase protein is involved in the oxidation of aldehydic substrates to their corresponding carboxylic acids like all-trans-retinal and all-trans-13,14-dihydroretinal to all-trans-retinoate and all-trans-13,14-dihydroretinoate respectively. The ALDH1A3 protein is involved in the synthesis of retinoic acid through the oxidation of retinaldehyde to retinoic acid in the ocular and nasal regions at the embryonic level[25]. In the vertebrate's retinoic-acid signalling pathway (RASP), retinoic acid functions as a ligand for retinoid binding receptors, which in turn regulates the transcription of target genes[26]. The nuclear hormone receptors



Protein	Score
CYP26B1	0.975
CYP26A1	0.973
CYP26C1	0.970
RDH5	0.953
DHRS3	0.951
DHRS9	0.949
RDH11	0.948
SDR16C5	0.944
BCO1	0.940
CYP1A1	0.935

Figure 2 The image represents protein structure modelling and docking A: Wild-type protein model; B: Variant protein model; C: Superimposed models of wild-type and variant; D: Normal protein docked with close interactor protein; E: Variant protein docked with close interactor protein; F: STRING protein interaction depicting close interactor of ALDH1A3 protein with probability values.

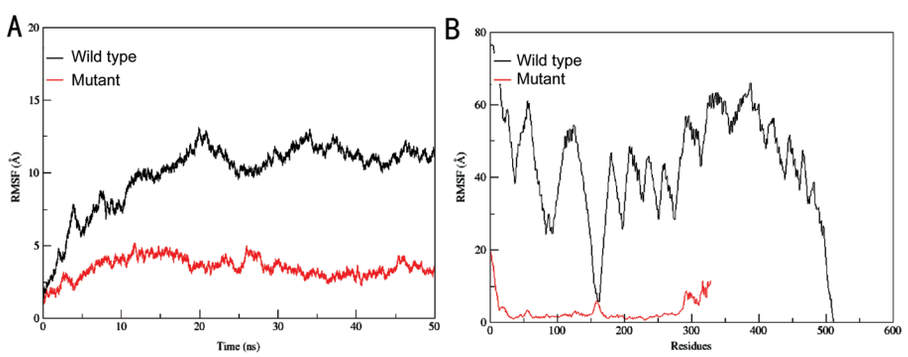


Figure 3 The graphs show outcome of molecular dynamic simulation A: The graph shows root mean square deviation; B: Root mean square fluctuation to indicate the altered simulation status of wild-type and variant ALDH1A3 protein.

family of retinoid receptors, which includes retinoic acid receptors (RARs) and retinoid X receptors (RXRs), are used in the retinoic-acid signaling pathway in vertebrates. These are transcription factor proteins that bind with cis or trans-retinoic acid by using a ligand-binding domain and also use their DNA-binding domain to regulate the transcription of targeted genes[27]. Retinoic acid is one of the potent derivatives of vitamin A (retinal), which affects the development of the retina and spinal cord during embryogenesis. In adult tissues, it also retains epithelial cells and helps in the development of neural cells. Two steps are involved in the production of retinoic acid. In the first step, alcohol dehydrogenase oxidizes retinal to produce relative aldehyde. While, in the next phase, retinal dehydrogenases are used to irreversibly convert aldehyde to its corresponding acid[28].

The mouse retinaldehyde dehydrogenase gene (*raldh2)* is expressed in the periocular mesenchyme cells, a subset of neural cells necessary for the development of anterior structures, including the anterior part of the eye. This gene is similar to the *ALDH1A2* gene in humans[29]. The optic

cup invagination is inhibited by a functional loss of variant in *Raldh2*. Hence, it is thought that *ALDH1A2* controls the synthesis of retinoic acid in humans by controlling the signal that periocular mesenchyme sends to invade the optic vesicle[30]. A mouse variant with *raldh3* (*ALDH1A3* of humans counterpart) initiates the formation of the optic cup but the absence of retinoic acid in the ventral optic cup resulted in the shortening of the ventral retina. Consequently, it appears that ALDH1A3 plays a key role in the formation of the optic cup[29].

The isozymes ALDH1A3, 1A2, and 1A1 exhibit sequence similarity of about 70%. Though, ALDH1A1, 1A2, and 1A3 identify the same substrate, but their expression patterns are slightly different which indicates their substrate preference. Certainly, ALDH1A1 specifically processes aldehydes that are produced through lipid peroxidation, whereas ALDH1A2 and ALDH1A31 use retinal as a primary substrate[31]. The structural study by Wong *et al*[32] revealed a strong degree of structural homology between ALDH1A3 and other aldehyde dehydrogenases including ALDH1A2 and ALDH1A1. These enzymes showed a remarkable degree of conservation in the overall fold and active site configuration of ALDH1A3. The structural alignment highlighted the functional associations between these dehydrogenases by revealing a high degree of conservation in the catalytic domain and the cofactor-binding site[32].

Although, ALDH1A3 exhibits a tetramer structure, its monomeric unit can independently use NAD as a cofactor to oxidize retinaldehyde into retinoic acid. The functional domains of ALDH1A3 are the NAD-binding domains, these include the catalytic domain and the C-terminal oligomerization domains. These domains are formed by folding monomeric units into 13 α-helices, 19 β-sheets, as well as other connecting loops[27]. Variants in the retinaldehyde dehydrogenase gene are thought to be responsible for 10% of cases of A/M[33]. Biallelic variants may occur alone or in complex forms and are primarily associated with bilateral A/M[10]. Neurocognitive signs and symptoms are also caused by a biallelic pathogenic *ALDH1A3* variant, while other extraocular abnormalities are rare[25].

Thirty-three distinct variants have been identified so far, including five splice sites, one small deletion, one tiny insertion, twenty-one missense/nonsense, and five other variants. The reported variants occasionally result in unilateral or bilateral A/M. ALDH1A3 has also been associated with neurocognitive defects and intellectual disability (Human Genome Mutation Database: HGMD.cf.ac.uk/ac/all.php). The currently identified splice site mutation [p.(Asp296SerfsTer35)] is present in the catalytic domain, which spans from amino acids 283 to 482, and is thought to be involved in the alteration of tertiary structure and affect the folding of the protein. The catalytic domain contains the majority of variants, followed by the NAD binding domains and the oligomerization domains. Notably, there was no apparent association between the variants' positions and certain phenotypic traits or differences in the severity of the disease[34].

Although, the variants in *ALDH1A3* gene, which have previously been reported in many consanguineous families, are globally responsible for the development of anophthalmia and microphthalmia. However, there is some partial similarity between the clinical appearance of the current family and other families that were previously associated with the same *ALDH1A3* gene variant. In addition to bilateral A/M, individuals with biallelic *ALDH1A3* variants have also been documented to have other varied systemic characteristics, such as neurodevelopmental delay and autism. Kesim *et al*[34] have studied seven families with biallelic pathogenic *ALDH1A3* variants. These patients, apart from A/M, has revealed facial dysmorphic, autism, intellectual disability and developmental delays. This study exhibited notable intra- and interfamilial clinical heterogeneity among the patients. Whereas, in the current study, we found that all patients had anophthalmia (total absence of eyeballs). Furthermore, two siblings (V-2 and V-5), from different parental loop, had both anophthalmia and mild to moderate intellectual disability, while, the other patients had only isolated anophthalmia with no developmental delays or intellectual impairment.

Previously, a study was conducted by Lin *et al*[3], on two consanguineous Pakistani families, in which seven patients had bilateral non-syndromic anophthalmia and identified c.1240G >C, p.(Gly414Arg; family 1) and a frameshift variants c.172dup p.(Glu58Glyfs*5; family 2) in *ALDH1A3* gene. Semerci and his co-workers in 2014, identified homozygous missense variants c.666G>A and c.1398C>A in two Turkish families with A/M. The c.666G>A variant skips exon 6 of the *ALDH1A3* gene. While c.1398C>A variant was substituting the highly conserved amino acid Asparagine at position 466 with Lysine[35]. Mory *et al*[36] reported a homozygous missense variant c.211G>A; p.(Val71Met) in the *ALDH1A3* gene in nine affected individuals of a family from Israel with severe A/M, while no other clinical features were found in all the patients. Recently in 2023, Piryaei *et al*[37] reported a novel frameshift variant c.1441delA; p.(Met482Cysfs*8) in *ALDH1A3* gene in Iranian proband with clinical features of severe microphthalmia, cyst in the eye, and blindness. Additionally, even for identical variants, there is significant diversity in traits both within and between families. Roos *et al*[38] reported neurodevelopmental intra-familial heterogeneity within a large consanguineous family with microphthalmia/coloboma and reported a homozygous *ALDH1A3* variant *i.e.* p.(Cys174Tyr). Similarly, Kesim *et al*[34] documented seven

families with biallelic pathogenic *ALDH1A3* variants in 2023. The affected brothers in family 1, who had identical compound heterozygous variants, showed variable phenotypes. The younger sibling had a severe neurodevelopmental delay, while the older sibling had isolated bilateral anophthalmia and a normal intellect. In all such cases, *ALD1H3* variants are consistently responsible for showing syndromic condition of neurodevelopmental impairments with significant intra-familial variability.

The MD simulation is an effective method to observe the physical movement of atoms and molecules with time[39]. The difference in the RMSD values of the variant and wild-type protein represents the structural stabilization of the protein. In this study, it was observed that the mean RMSD value for the wild-type protein was 10.09 Å, whereas the variant's was 3.48 Å. While the wild-type protein had the highest RMSD values at 13.12 Å, and the variant protein had the highest at 5.19 Å. The lowered RMSD values of the variant indicate that the protein's flexibility has been reduced, which could have an impact on the protein's dynamic characteristics and interactions with other molecules as high rigidity in the variant may result in reduced activity or changed functioning[40]. Whereas the rapid increase in the RMSD values of the wild-type in the first 10 ns, and then fluctuating values after 20 ns highlights the higher flexibility and instability in the wild-type protein structure, which is important to the wild-type biological role.

The flexibility and fluctuation of individual residues within the proteins were shown by the RMSF results[41]. In the current study, the variant and wild-type proteins had the highest RMSF values, which were 19.52 Å and 76.63 Å, respectively. Whereas 3.17 Å and 44.09 Å were the mean RMSF values for the variant and wild-type proteins, respectively. Low values of RMSF are related to less structural movement and a more rigid structure[42], which were observed for variant protein as compared to wild-type, hence being more flexible. The dynamic characteristics of the variant may be affected by its protein's reduced flexibility, altering its capacity to go through conformational changes which are essential for its function and hence impairing its biological role.

In contrast to the variant, MD simulations of the wild-type protein showed significant deviation, instability, and high values of RMSD and RMSF. These findings suggest that the variant identified in this study may impact protein stability and function, potentially resulting in anophthalmia in the patients under study.

The *in silico* studies on protein-protein docking (ALDH1A3-CYP26B1) have shown disruptive interaction. CYP21B1 is a member of cytochrome P450 proteins that catalyzes several biochemical reactions involving retinoic acid metabolism. Intracellularly, CYP26B1 is localized in the endoplasmic reticulum, where it acts as a critical regulator of all-trans retinoic acid levels by the specific inactivation of all-trans retinoic acid to hydroxylated forms. Genetic mutation in the *CYP26B1* gene leads to a spectrum of phenotypes affecting the bone, teeth, nervous and eye[43]. Being the member of retinoic acid metabolism, ALDH1A3 & CYP26B1 interaction is important for regulating the level of retinoic acid. Hence, their interactional loss may alter the retinoic acid level and thus may affect eye development.

In the current study, we report a splice site variant NM_000693.2: c.884-2_885dup: p.(Asp296SerfsTer35) in the 9th exon of *ALDH1A3* that results in bilateral anophthalmia. Genotype-phenotype association has observed that the missense variant generally causes microphthalmia, whereas protein-truncating variants (*i.e.*, nonsense and frameshift variants) are significantly associated with anophthalmia. The findings of the current study may be useful in developing a genetic diagnostic test for Pakistani families suffering from the A/M phenotype.

ACKNOWLEDGEMENTS

We appreciate the voluntary cooperation of all patients and family members in the current study. The authors also extend their appreciation to Taif University, Saudi Arabia, for supporting this work through project number (TU-DSPP-2024-05).

Author's Contributions: Samad K, Muzammal M, Ahmad S, Ullah F, Samad F, and Ali MZ were involved in patient recruitment and their clinical analysis. Samad K, Alzahrani KJ, and Alsharif KF, performed experiments and exome data analysis. Abbasi SW and Muzammal M performed protein modeling, docking, and simulation analysis. Khan MA, Alzahrani KJ, and Alsharif KF conceptualized, supervised the study, and also drafted the manuscript as well.

Data Availability: Reference sequences were obtained from freely available genome databases. While sequence, photographs, and pedigrees will be provided on request.

Foundation: Supported by Taif University, Taif, Saudi Arabia (TU-DSPP-2024-05).

Conflicts of Interest: Samad K, None; **Alzahrani KJ,** None; **Alsharif KF,** None; **Samad F,** None; **Ullah F,** None; **Ali MZ,** None; **Ahmad S,** None; **Muzammal M,** None; **Abbasi SW,** None; **Khan J,** None; **Khan MA,** None.

REFERENCES

1 Li JJ, Yang W, Wang YJ, *et al*. Exome sequencing identifies genetic variants in anophthalmia and microphthalmia. *Am J Med Genet A* 2022;188(8):2376-2388.

2 Reis LM, Costakos D, Wheeler PG, *et al*. Dominant variants in PRR12 result in unilateral or bilateral complex microphthalmia. *Clin Genet* 2021;99(3):437-442.

3 Lin SY, Harlalka GV, Hameed A, *et al*. Novel mutations in ALDH1A3 associated with autosomal recessive anophthalmia/microphthalmia, and review of the literature. *BMC Med Genet* 2018;19(1):160.

4 Stoll C, Dott B, Alembik Y, *et al*. Associated anomalies in anophthalmia and microphthalmia. *Eur J Med Genet* 2024;67:104892.

5 Harding P, Gore S, Malka S, *et al*. Real-world clinical and molecular management of 50 prospective patients with microphthalmia, anophthalmia and/or ocular coloboma. *Br J Ophthalmol* 2023;107(12):1925-1935.

6 Jackson D, Malka S, Harding P, *et al*. Molecular diagnostic challenges for non-retinal developmental eye disorders in the United Kingdom. *Am J Med Genet C Semin Med Genet* 2020;184(3):578-589.

7 Fahnehjelm C, Dafgård Kopp E, Wincent J, *et al*. Anophthalmia and microphthalmia in children: associated ocular, somatic and genetic morbidities and quality of life. *Ophthalmic Genet* 2022;43(2):172-183.

8 Ullah E, Nadeem Saqib MA, Sajid S, *et al*. Genetic analysis of consanguineous families presenting with congenital ocular defects. *Exp Eye Res* 2016;146:163-171.

9 Yahyavi M, Abouzeid H, Gawdat G, *et al*. ALDH1A3 loss of function causes bilateral anophthalmia/microphthalmia and hypoplasia of the optic nerve and optic chiasm. *Hum Mol Genet* 2013;22(16):3250-3258.

10 Williamson KA, FitzPatrick DR. The genetic architecture of microphthalmia, anophthalmia and coloboma. *Eur J Med Genet* 2014;57(8):369-380.

11 Li L, Cui YJ, Zou YC, *et al*. Genetic association study of SOX2 gene polymorphisms with high myopia in a Chinese population. *Eur J Ophthalmol* 2021;31(2):734-739.

12 Ahmed I, Muzammal M, Khan MA, *et al*. Identification of four novel candidate genes for non-syndromic intellectual disability in Pakistani families. *Biochem Genet* 2024;62(4):2571-2586.

13 Zhou Y, Kathiresan N, Yu ZC, *et al*. A high-performance computational workflow to accelerate GATK SNP detection across a 25-genome dataset. *BMC Biol* 2024;22(1):13.

14 Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc* 2015;10(10):1556-1566.

15 Siddalingappa R, Kanagaraj S. A novel ML approach for computing missing sift, provean, and mutassessor scores in Tp53 mutation pathogenicity prediction. *Int J Adv Comput Sci Appl* 2023;14(6):1038-1047.

16 Schwarz JM, Cooper DN, Schuelke M, *et al*. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014;11(4):361-362.

17 Yari A, Ali-Nejad RM, Saleh-Gohari N. A novel homozygous splice-site mutation in SCARB2 is associated with progressive myoclonic epilepsy with renal failure. *Neurol Sci* 2021;42(12):5077-5085.

18 Peng ZL, Wang WK, Wei H, *et al*. Improved protein structure prediction with trRosettaX2, AlphaFold2, and optimized MSAs in CASP15. *Proteins* 2023;91(12):1704-1711.

19 Ghani U, Desta I, Jindal A, *et al*. Improved Docking of Protein Models by a Combination of Alphafold2 and ClusPro. 2021. https://www.biorxiv.org/content/10.1101/2021.09.07.459290v2.article-info. Posted July 27, 2022.

20 Szklarczyk D, Kirsch R, Koutrouli M, *et al*. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 2023;51(D1).

21 Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* 2011;51(10):2778-2786.

22 Schiffrin B, Radford SE, Brockwell DJ, *et al*. PyXlinkViewer: a flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system. *Protein Sci* 2020;29(8):1851-1857.

23 Raney BJ, Barber GP, Benet-Pagès A, *et al*. The UCSC Genome Browser database: 2024 update. *Nucleic Acids Res* 2024;52(D1).

24 Case DA, Aktulga HM, Belfon K, *et al*. AmberTools. *J Chem Inf Model* 2023;63(20):6183-6191.

25 Slavotinek A. Genetics of anophthalmia and microphthalmia. Part 2: Syndromes associated with anophthalmia-microphthalmia. *Hum Genet* 2019;138(8-9):831-846.

26 Mori Y, Masuda M, Yoshida-Shimizu R, *et al*. All-trans retinoic acid induces lipophagy through the activation of the AMPK-Beclin1 signaling pathway and reduces Rubicon expression in adipocytes. *J Nutr Biochem* 2024;126:109589.

27 Moretti A, Li JF, Donini S, *et al*. Crystal structure of human aldehyde dehydrogenase 1A3 complexed with NAD^+ and retinoic acid. *Sci Rep* 2016;6:35710.

28 Duester G. Genetic dissection of retinoid dehydrogenases. *Chem Biol Interact* 2001;130-132(1-3):469-480.

29 Duester G. Keeping an eye on retinoic acid signaling during eye development. *Chem Biol Interact* 2009;178(1-3):178-181.

30 Dai J, Jiang HT, Yang ZQ, *et al*. A functional variant of ALDH1A2 is associated with hand osteoarthritis in the Chinese population. *Int J Immunogenet* 2024;51(3):157-163.

31 Maly IP, Crotet V, Toranelli M. The so-called “testis-specific aldehyde dehydrogenase” corresponds to type2 retinaldehyde dehydrogenase in the mouse. *Histochem Cell Biol* 2003;119(2):169-174.

32 Wong WR, Maher S, Oh JY, *et al*. Conserved missense variant in *ALDH1A3* ortholog impairs fecundity in *C. elegans*. *MicroPubl Biol* 2021;2021.

33 Slavotinek AM, Garcia ST, Chandratillake G, *et al*. Exome sequencing in 32 patients with anophthalmia/microphthalmia and developmental eye defects. *Clin Genet* 2015;88(5):468-473.

34 Kesim Y, Ceroni F, Damián A, *et al*. Clinical and genetic analysis further delineates the phenotypic spectrum of ALDH1A3-related anophthalmia and microphthalmia. *Eur J Hum Genet* 2023;31(10):1175-1180.

35 Semerci CN, Kalay E, Yıldırım C, *et al*. Novel splice-site and missense mutations in the ALDH1A3 gene underlying autosomal

recessive anophthalmia/microphthalmia. *Br J Ophthalmol* 2014;98(6):832-840.

36 Mory A, Ruiz FX, Dagan E, *et al*. A missense mutation in ALDH1A3 causes isolated microphthalmia/anophthalmia in nine individuals from an inbred Muslim kindred. *Eur J Hum Genet* 2014;22(3):419-422.

37 Piryaei F, Pakmanesh R, Salehirad M, *et al*. ALDH1A3-related congenital microphthalmia-8 due to a novel frameshift variant. *Eur J Med Genet* 2023;66(8):104801.

38 Roos L, Fang M, Dali C, *et al*. A homozygous mutation in a consanguineous family consolidates the role of ALDH1A3 in autosomal recessive microphthalmia. *Clin Genet* 2014;86(3):276-281.

39 Bai QF, Liu S, Tian YN, *et al*. Application advances of deep learning methods for *de novo* drug design and molecular dynamics simulation. *Wires Comput Mol Sci* 2022;12(3):e1581.

40 Ullah A, Shah AA, Syed F, *et al*. Molecular dynamic simulation analysis of a novel missense variant in *CYB5R3* gene in patients with methemoglobinemia. *Medicina* (*Kaunas*) 2023;59(2):379.

41 Mohamad Yunus AA, Abdul Wahab R, Bahaman AH, *et al*. In-silico alanine scanning analysis on the catalytic residues of a novel β-glucosidase from trichoderma asperellum uc1. *J Teknologi* 2021;83(3):61-73.

42 Oyewusi HA, Wahab RA, Akinyede KA, *et al*. Bioinformatics analysis and molecular dynamics simulations of azoreductases (AzrBmH2) from Bacillus megaterium H_2 for the decolorization of commercial dyes. *Environ Sci Eur* 2024;36(1):31.

43 Silveira KC, Fonseca IC, Oborn C, *et al*. CYP26B1-related disorder: expanding the ends of the spectrum through clinical and molecular evidence. *Hum Genet* 2023;142(11):1571-1586.