Heliyon 10 (2024) e36560

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

An extensive in silico analysis of missense mutations of the human *AIMP2* gene

Shima Farrokhi [a,b], Atieh Eslahi [a,b], Farzaneh Alizadeh [a,b], Zahra Farshchian [c], Yasamin Yousefi [d], Majid Mojarrad [a,e,f,*]

[a] Department of Medical Genetics and Molecular Medicine, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[b] Student Research Committee, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[c] Department of Medical Genetics, Faculty of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
[d] Department of Biochemistry, Mashhad University of Ferdowsi, Mashhad, Iran
[e] Genetic Center of Khorasan Razavi, Mashhad, Iran
[f] Medical Genetics Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

ARTICLE INFO

Keywords:
Hypomyelinating leukodystrophy 17
AIMP2
MSC complex
Insilico
nsSNVs

ABSTRACT

HLD17 (Hypomyelinating Leukodystrophy 17) is an inherited white matter disorder characterized by insufficient myelin production due to biallelic loss of function mutations in the aminoacyl-tRNA synthetase complex-interacting multifunctional protein 2 (AIMP2) gene. In silico analysis of SNVs (single nucleotide variants) in the AIMP2 gene is an efficient and cost-effective method for analyzing and predicting the impact of mutations on protein function and disease pathophysiology.

The study used dbSNP and Ensembl databases to obtain data on 343 nonsynonymous single nucleotide variants (nsSNVs) in the human AIMP2 gene. Six prediction algorithm tools were used to assess the effects of these nsSNVs on AIMP2's functions and structures. Results showed that 18 nsSNVs were located within functional domains, while 10 nsSNVs led to decreased protein stability. The structural and functional properties of the AIMP2 protein were investigated using databases such as Predict Protein, Mutpred2, and HOPE. ConSurf analysis provided information about conserved nsSNVs. GeneMANIA and STRING software tools were used to predict interactions between gene-gene and protein-protein, respectively. Phyre2 and I-TASSER web servers were used to predict the 3D structures of wild-type and mutant proteins.

In addition, having the challenge of probable post-translational modification sites in the AIMP2, we made predictions using various bioinformatics tools. Consecuently, three minor mutations (L138Q, V161E, and I188N) and five major mutations (C23S, D121G, I122S, P128S, and W268S) were found to affect the AIMP2 protein's structure or function. Anyway, These mutations need to be further studied and confirmed through experimental investigation and GWAS studies.

1. Introduction

HLD17 (Leukodystrophy Hypomyelinating 17) is a congenital neurodevelopmental disorder leading to numerous health conditions

* Corresponding author. Department of Medical Genetics and Molecular Medicine, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
E-mail address: majidmojarrad11@gmail.com (M. Mojarrad).

https://doi.org/10.1016/j.heliyon.2024.e36560
Received 9 April 2024; Received in revised form 8 August 2024; Accepted 19 August 2024
Available online 20 August 2024

List of abbreviations

AIMP2 aminoacyl-tRNA synthetase complex-interacting multifunctional protein 2
HLD17 Hypomyelinating Leukodystrophy 17
MSC multi-tRNA synthetase complex
SNVs single nucleotide variants
dbSNP Database for Single Nucleotide Polymorphisms
SIFT The scale-invariant feature transform
PolyPhen-2 Polymorphism Phenotyping v2
PROVEAN Protein Variation Effect Analyzer
PhD-SNP Predictor of human deleterious single nucleotide polymorphisms
SNP and GO Predicting disease associated variations using GO terms
PANTHER position-specific evolutionary preservation
Pfam protein families and domains database
the AIMP2-LysRS-bd Domain lysyl-tRNA synthetase (LysRS) binding domain of AIMP2 GST C-terminal domain: Glutathione S-transferase, C-terminal domain
DDG the value of free energy change
Mutpred Mutation predictor
ConSurf conservation surface-mapping
GeneMANIA Multiple Association Network Integration Algorithm
STRING A Search Tool for the Retrieval of Interacting Genes/Proteins
Phyre2 Protein Homology/Analogy Recognition Engine
I-TASSER Iterative Threading ASSEmbly Refinement
TM SCORE template modeling score
RMSD root mean square deviation of atomic positions
PTMs Post-translational modifications
CSS-Palm palmitoylation site prediction with a clustering and scoring strategy
iDNA-Methyl identifying DNA methylation sites
iHyd-PseAAC identify hydroxylation by pseudo amino acid composition
BDM-PUB Computational prediction of protein ubiquitination sites with a Bayesian discriminant method NetNGlyc: N-Glycosylation sites in human proteins using artificial neural networks
NetPhos 3.1 predicts phosphorylation sites in eukaryotic proteins using neural networks
GPS 3.0 Group-based Prediction System (GPS), for prediction of PK-specific phosphorylation sites integrative protein signature database

like neurodevelopmental delay, early-onset multi-focal seizures, spasticity, microcephaly (up to −10 SD), inability to walk and speak, cerebral and cerebellar atrophy, and thin corpus callosum [1].

HLD17 is caused by homozygous loss of function mutations in the AIMP2 (aminoacyl-tRNA synthetase complex-interacting multifunctional protein 2) gene located on the short arm of chromosome 7. AIMP2 encodes an auxiliary protein with 320 amino acid residues, which is necessary for the assembly and stability of the MSC (multi-tRNA synthetase complex) [1]. MSC is a house-keeping enzyme complex with nine aminoacyl tRNA synthetase and three non-enzymatic proteins (AIMP1/p43, AIMP2/p38, and AIMP3/p18) essential for the translation of mRNAs to proteins. There is some evidence that AIMP2 mutations play a crucial role in the decline of the assembly and stability of the MSC.

We can perform various cellular functions by using MSC components in response to external signals. For example, when DNA damage occurs, the AIMP2 protein is released from the MSC and, exerts pro-apoptotic activity through p53 activation [2], and also induces cell death through the ubiquitination of TRAF2 (tumor necrosis factor receptor-associated factor 2) [3].

In addition, previous studies identified some deleterious mutations in the AIMP2 gene, such as a homozygous nonsense mutation (c.105C > A; p. Tyr35Ter) (1), missense mutations such as c.A463T [4] and L350H and P390R [5], which are associated with significant neurodevelopmental disorders.

Since leukodystrophies have a low incidence and the functional impacts of their causal variants have rarely been studied, methods such as in Silico analysis could play a valuable role in the detection of pathogenic variants [6]. This study aimed to analyze the pathological variants of the AIMP2 gene.

2. Result

Based on dbSNP and ensembl results, the AIMP2 gene contains 5392 SNVs, of which 343 were missense SNVs chosen for our investigation (Fig. 1).

2.1. Identification of deleterious SNVs in the AIMP2 gene

We investigated the pathogenicity of SNVs in the AIMP2 protein using six Silico tools (SIFT, PolyPhen-2, PROVEAN, PHD-SNP, SNP&GO, and PANTHER).

PROVEAN analysis predicted that 86 and 255 nsSNVs were deleterious and neutral, respectively. Polyphen-2 analysis revealed that 106 nsSNVs were probably damaging, whereas 66 nsSNVs were predicted to be 'possibly damaging,' and the remaining 170 nsSNVs were classified as benign.

Using the SIFT Analyzer, 178 nsSNVs were predicted as damaging and 163 nsSNVs as neutral. SNP&GO analysis showed that 33 SNPs were predicted to be disease-causing, and 309 SNPs were predicted to be neutral. PhD-SNP predicted that 268 nsSNVs were neutral, while 73 nsSNVs were associated with the disease. PANTHER predicted that 80 nsSNVs were possibly damaging, 175 nsSNVs were predicted to be probably damaging, and 87 nsSNVs were categorized as probably benign (Table 1), (Supplementary Table 1).

Since each program uses different algorithms to estimate the nsSNVs, we conducted a concordance evaluation. We predicted 24 out of the 343 nsSNVs by all algorithms as deleterious (Table 1), (Supplementary Table 1).

Identification of nsSNVs distribution on the domains of AIMP2.

InterPro and Pfam predicted the domains and active sites of the AIMP2 protein. These websites indicated that 18 out of 24 deleterious nsSNVs, as identified by bioinformatics tools, were located on the three functional domains of AIMP2: the AIMP2-LysRS-bd domain (1-44), the thioredoxin 16 domain (118–208), and the GST C-terminal domain (235–309). The finding of nsSNVs with deleterious effects on protein stability.

2.2. The finding of nsSNVs with deleterious effects on protein stability

Istable software predicted the change in protein stability after a single point mutation by calculating the *change* in thermodynamic free energy (ΔΔG) and the change direction. These findings demonstrate that when $G > 0$, the *change* is stable, whereas when $G = 0$, the *change* leads to protein instability [7]. The results showed that 10 out of 18 nsSNVs decreased AIMP2 protein stability, which implies that these SNVs in the AIMP2 domain may cause maximum protein damage (Table 1), (Supplementary Table 2).

2.3. Identification of nsSNVs with deleterious effects on protein structural and functional properties

PredictProtein is a powerful tool for revealing the contributions of individual amino acids to protein structure and function. The results indicate that every 10 nsSNVs can alter the production of the AIMP2 protein, even though both normal and mutant AIMP2 proteins are expressed in the cell nucleus, with no difference in other protein properties such as disulfide bonds, transmembrane helices, or gene ontology (Table 1), (Supplementary Table 3).

Based on MutPred findings, pathogenicity is indicated by a score greater than 0.5 and a p-value <0.05 (26). Although the MutPred tool suggested that the W189L and W236R substitutions did not affect the protein, it was discovered that the other eight SNVs caused significant changes in the AIMP2 protein. The result of the MutPred server is summarized in Table 1 and, Supplementary Table 4.

Moreover, by using HOPE online software, we were able to predict how amino acid substitutions would affect protein properties, hydrophobicity, spatial structure, and function. The study revealed that all 10 pathogenic nsSNVs alter the hydrophobicity of amino

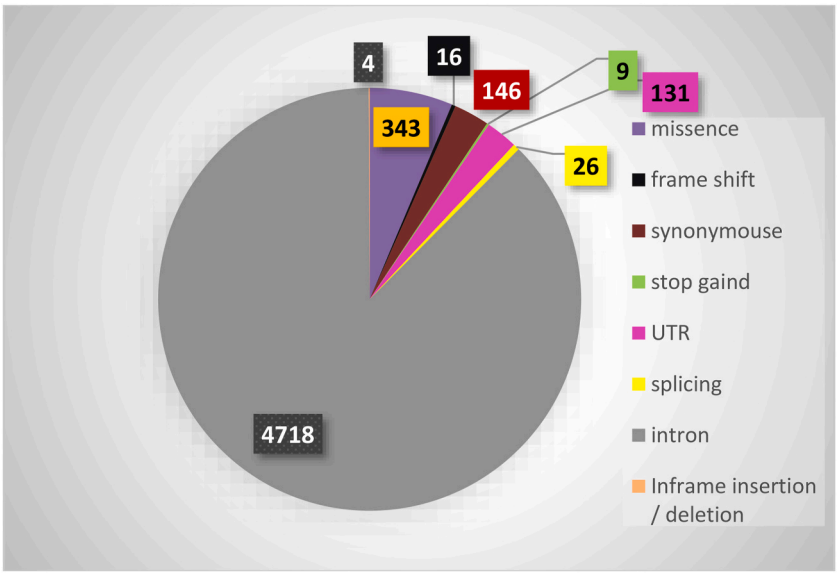

Fig. 1. Percentage of different SNP varieties in the AIMP2 gene.

Table 1
The table presents harmful SNVs identified by six software programs, findings from the Snap2 program within the PredictProtein tool, results from MutPred, and a portion of the findings from the HOPE project. The table represents disease-causing harmful variants, excluding natural variation. Additionally, it found that 8 SNVs increase protein stability, which was excluded from the study. The Snap2 program predicted the effect of each substitution on the protein, while MutPred indicated the potential impact of each mutation on protein structure and function. Furthermore, HOPE results showed the conformational change of the wild-type residue caused by the mutant residue for each mutation.

variant	mutation	SIFT	PROVEAN	PolyPhen	PhD-SNP	SNPs&GO	PANTHER	i-Mutant2.0	Mu pro	istable	Snap2 predicted	MutPred results	HOPE results
rs1371935734	C 23 S	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Decrease	Decrease	effect	Loss of Allosteric site at Y25	Reduce hydrophobicity
rs200028351	D121G	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Decrease	Decrease	effect	Loss of Phosphorylation at Y116 Altered Transmembrane protein	1. Increase hydrophobicity 2. Reduce the size of the protein 3. Alter protein charge 4. Disrupt the rigidity of the protein
rs764368950	I122S	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Decrease	Decrease	effect	Gain of Strand Gain of Ubiquitylation at K120 Altered Transmembrane protein Altered Stability	1. Reduce the size of the protein 2. Reduce hydrophobicity 3. Form a hole in the protein's center
rs778622819	P128S	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Decrease	Decrease	_	Loss of Strand Altered Transmembrane protein	1. Reduce the size of the protein 2. Reduce hydrophobicity 3. Disrupt the conformation of the protein 4. Produce a space in the core of the protein
rs747974072	P128L	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Increase	Increase	effect	_	_
rs765096358	L138Q	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Decrease	Decrease	effect	Gain of Helix	1. Alter hydrophobicity

(*continued on next page*)

Table 1 (*continued*)

variant	mutation	SIFT	PROVEAN	PolyPhen	PhD-SNP	SNPs&GO	PANTHER	i-Mutant2.0	Mu pro	istable	Snap2 predicted	MutPred results	HOPE results
													2. Increase the size of the protein
rs765096358	L138P	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Increase	Decrease	Decrease	_	_	_
rs374990034	H155P	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Increase	Increase	Increase	_	_	_
rs746796665	S156F	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Increase	Increase	Increase	_	_	_
rs774747119	V161E	Damaging	Deleterious	Probably damaging	Disease	Disease	possibly damaging	Decrease	Decrease	Decrease	effect	gain of Intrinsic disorder Gain of SUMOylation at K159 Gain of B-factor	1. Increase the size of the protein 2. Alter protein charge 3. Reduce hydrophobicity 4. Induce protein folding problems
rs1173669660	I188N	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Decrease	Decrease	effect	Altered Stability Altered Transmembrane protein	1. Increase the size of the protein 2. Reduce hydrophobicity 3. Disrupt molecular interactions within the protein
rs377624555	W189L	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Decrease	Decrease	effect	_	1. Reduce the size of the protein 2. Produce a space in the core of the protein
rs1396282189	K190T	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Increase	Decrease	Increase	_	_	_
rs982080297	G209S	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	_	_	_	_	_	_
rs926430593	G209D	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	_	_	_	_	_	_
rs756425703	E210K	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	_	_	_	_	_	_
rs749728733	I213T	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	_	_	_	_	_	_

(*continued on next page*)

Table 1 (*continued*)

variant	mutation	SIFT	PROVEAN	PolyPhen	PhD-SNP	SNPs&GO	PANTHER	i-Mutant2.0	Mu pro	istable	Snap2 predicted	MutPred results	HOPE results
rs377482895	R215C	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	_	_	_	_	_	_
rs150887968	R215H	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	_	_	_	_	_	_
rs1041316079	W236R	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Decrease	Decrease	effect	_	1. Reduce the size of the protein 2. Alter protein charge 3. Reduce hydrophobicity 4. Induce protein folding problems
rs997976538	W268S	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Decrease	Decrease	effect	Altered Transmembrane protein Gain of Intrinsic disorder Altered Ordered interface Gain of Ubiquitylation at K265	1. Reduce the size of the protein 2. Reduce hydrophobicity 3. Loss of external interactions
rs1477363279	Q287P	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Increase	Decrease	Decrease	_	_	_
rs1291481810	C306Y	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Increase	Increase	_	_	_
rs368862792	C306W	Damaging	Deleterious	Probably damaging	Disease	Disease	Probably damaging	Decrease	Increase	Increase	_	_	_

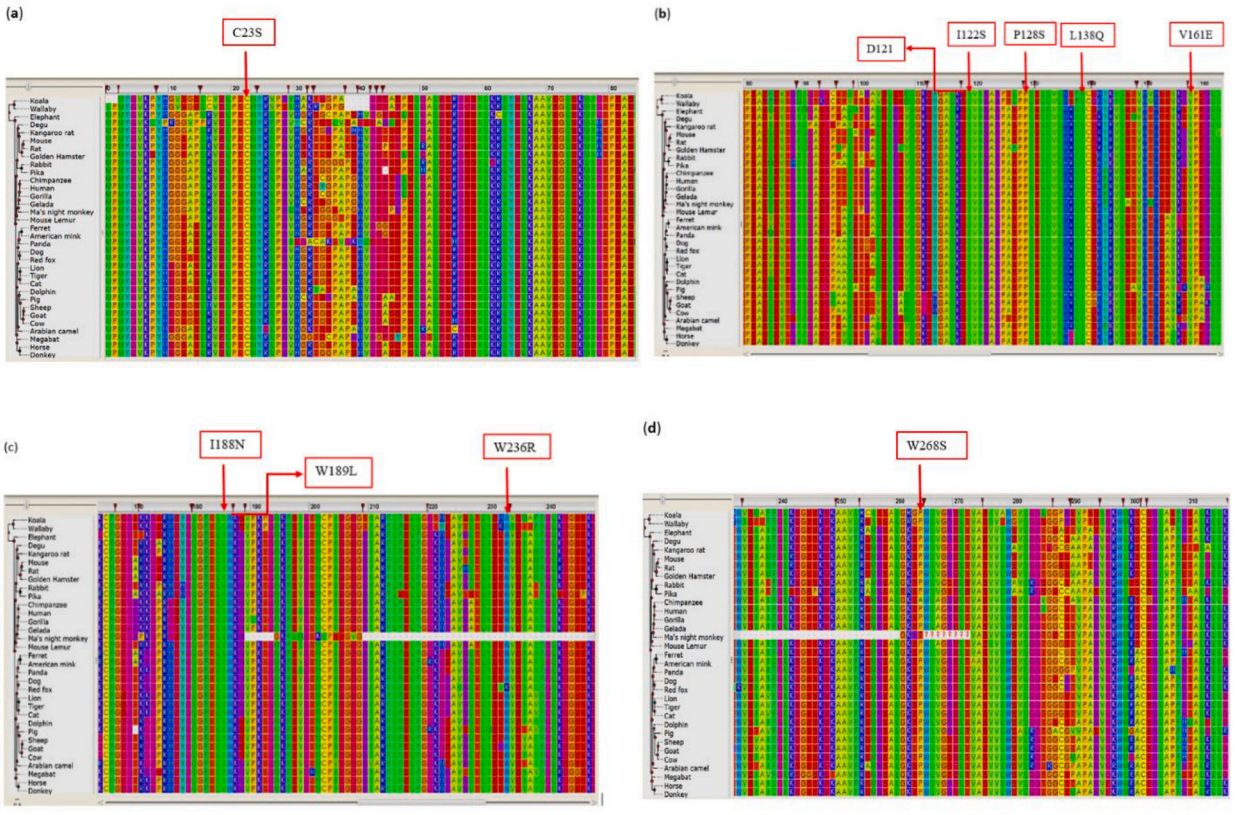

Fig. 2. Evolutionary conservation of high-risk nsSNVs in the AIMP2 protein produced by the Wasabi tool on the Consurf website. Wasabi indicates the degree of evolutionary conservation of each amino acid residue in the AIMP2 protein between 33 different species. Each amino acid is identified by a specific color. A. Evolutionary conservation of amino acid residues 1–84 B. 81–164C. 165 to 248 D. 249 to 317.

acids, thereby disrupting the H-bond interaction between adjacent molecules [7]. D121G, V161E, and W236R also alter the protein's charge. In contrast to D121G, I122S, P128S, W189L, W236R, and W268S, which change to a smaller amino acid, the mutant residues L138Q, V161E, and I188N are larger than the wild-type residues (Table 1), (Supplementary Table 5).

2.4. Evolutionary conservation analysis

A thorough understanding of evolution is essential for detecting mutations that might threaten human health. We calculated the evolutionary conservation of amino acid residues of the AIMP2 protein using the ConSurf web server to explore their potential impact. This tool provides results in several steps. In the first part, it was shown that D121G and P128S are exposed and functional, while residues C23S, I122S, L138Q, V161E, I188N, W189L, W236R, and W268S are buried and structural.

Then, using the Wasabi tool, it evaluates the evolutionary conservation of high-risk nsSNVs in 33 different species. For this purpose, each amino acid is marked with a special color. The results of this section show that the residues W236R and W268S are less

Table 2
Consurf's conservation values for each amino acid. Based on these results, 9 SNVs have high evolutionary conservation, and W236R has the lowest conservation in this list.

SNVs	substitution	Conservation score	FUNCTION	Class assignment
rs1371935734	C23S	8	Structural	Buried
rs200028351	D121G	9	Functional	Exposed
rs764368950	I122S	9	Structural	Buried
rs778622819	P128S	9	Functional	Exposed
rs765096358	L138Q	9	Structural	Buried
rs774747119	V161E	9	Structural	Buried
rs1173669660	I188N	9	Structural	Buried
rs377624555	W189L	8	Structural	Buried
rs1041316079	W236R	3	Structural	Buried
rs997976538	W268S	8	Structural	Buried

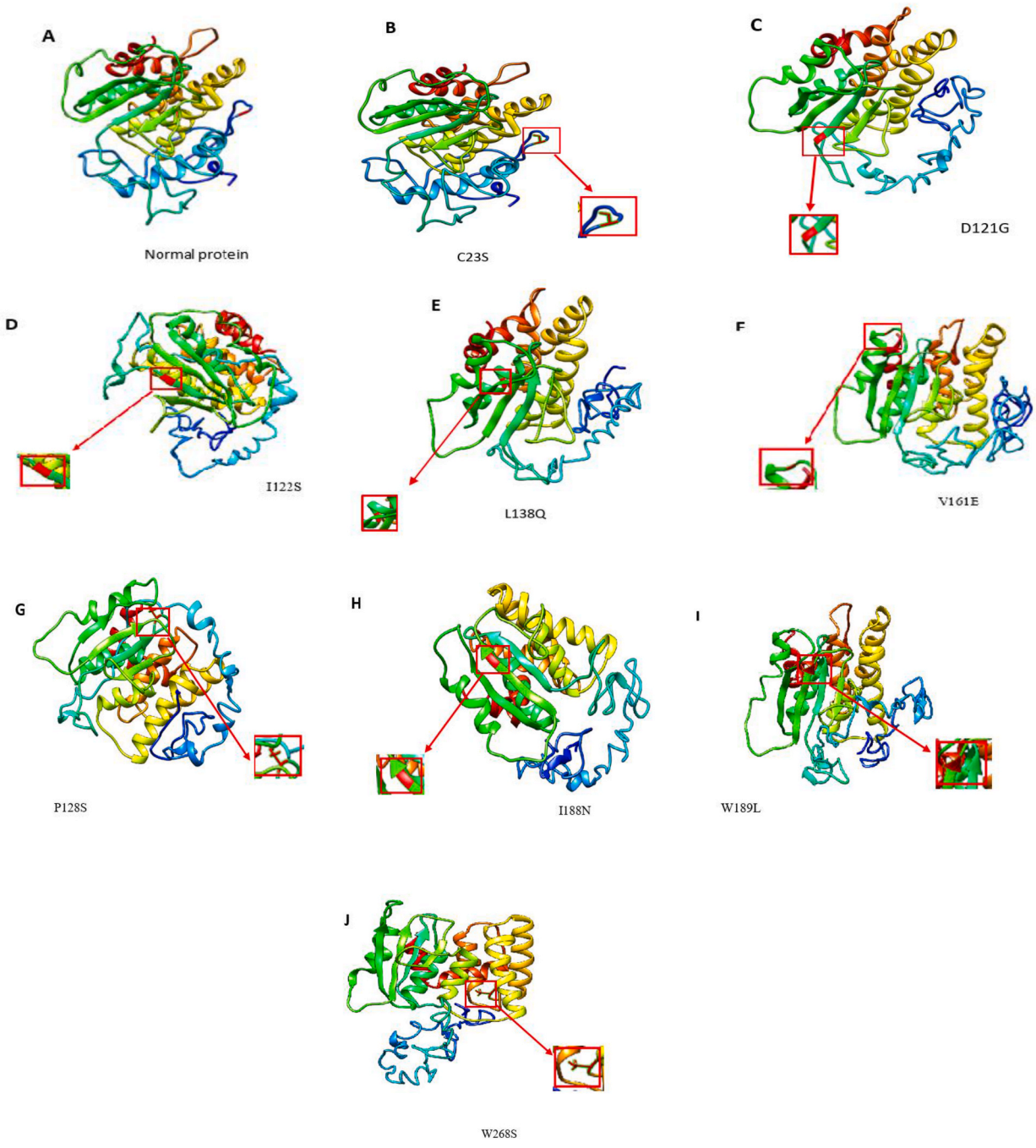

Fig. 3. A: The normal structure of the AIMP2 protein. 3B: Structure and position of the rs1371935734 variant leading to the conversion of serine to cysteine at position 23. 3C: Structure and position of the rs200028351 variant that converts aspartic acid to glycine at position 121. 3D: Structure and position of the rs764368950 variant leading to isoleucine to serine conversion at position 122. 3E: The structure and location of the rs765096358 variant leading to the conversion of leucine to glutamine at position 138. 3F: a mutant model of V161E SNV (valine converts to glutamic acid at 161). 3G: a mutant model of P128S SNV (proline converted to serine at 128). 3H: Mutant model of I188N SNV (isoleucine converts to asparagine at 188). 3I: a mutant model of W189L SNV (tryptophan converts to leucine at 189). 3J: a mutant model of W268S SNV (tryptophan converted to serine at 268). Dark blue colors indicates the AIMP2-LysRS-bd domain, light blue color indicates interdomain regions, Green colors indicate the thioredoxin 16 domain, Orange, red, and yellow colors show the GST C-terminal domain.

conservative compared to the rest of the residues. (Fig. 2a–d).

Finally, ConSurf scores conservation results ranging from 9 for most conserved amino acids to 1 for most variable amino acids, SNVs with scores between 7 and 9 should be deleterious to the genome. Consequently, W236R was excluded from further investigation

Table 3
Evaluation of RMSD, TM score, and C score for the desired mutations. C-Score shows the confidence of each model. It is usually in the range of [2,5], and a higher score indicates higher confidence. According to these results, w189l has the highest TM score and the lowest RMSD, making it more similar to the wild-type protein, and must be excluded from the study.

SNV	RMSD	TM score	C- score
Natural AIMP2	14.9	0.33	−3.48
C23S	15.0	0.33	−3.54
D121G	14.6	0.34	−3.40
I122S	15.3	0.32	−3.65
P128S	14.9	0.33	−3.50
L138Q	14.3	0.35	−3.26
V161E	14.3	0.35	−3.27
I188N	15.1	0.32	−3.56
W189L	13.8	0.37	−3.10
W268S	14.4	0.35	−3.30

Table 4
Analysis of post-translational modifications reveals that mutations in this protein can impact palmitoylation, hydroxylation, ubiquitination, and phosphorylation, leading to changes in protein functionality.

PTMs	Predicting tools	Positions of Aminoacides
hydroxylation	iHyd-PseAAC	21, 39, and 311
phosphorylation	NetPhos 3.1 and GPS 3.0	Predict 26 and 46 residues respectively
Ubiquitination	BDM-PUB	122 and 268
methylation	iDNA-Methyl	–
glycosylation	NetNGlyc	–
palmitoylation	CSS-palm	23

because it fell below the protection range (Table 2).

2.5. Gene-gene intractions

The study utilized the GeneMANIA program to investigate the relationship between AIMP2 genes and other genes. The results are presented in the supplementary section (Supplementary Fig. 1).

2.6. Protein–protein interactions

Protein-protein interactions have been investigated with the STRING database to understand the functional relationships between biological proteins. The findings of this section are discussed in the supplementary data (Supplementary Fig. 2).

2.7. 3D structure prediction of wild-type AIMP2 and its mutants

Two homology modeling tools, Phyre2 and I-TASSER, were used to generate the 3D models of wild-type and mutant AIMP2 proteins. As a first step, we manually introduced nsSNVs into wild-type sequences of AIMP2. Then, to model the sequences, we used I-TASSER and the Phyre2 homology modeling tools [8]. We visualized the resulting protein structure with Chimera 1.13 software (Fig. 3a-j).

In the next step, we extended our analysis to calculate TM-scores and RMSD values using I-TASSER to determine the similarity of structures. We evaluated the topological similarity between wild-type and mutant models using the TM score, which ranges from 0 to 1. A higher score indicates a closer resemblance between the two structures. Using RMSD, we can measure the average distance between the unique α-carbon backbones of wild-type and mutant models. There was a more significant deviation between mutant and normal structures when the RMSD value was high [9].

Based on the results from Phyre2 and I-TASSER, the mutant model for W189L exhibited a high TM-score and low RMSD. Therefore, we excluded them from further investigation (Table 3).

2.8. Prediction of post-translational modifications

PTMs control the structures and functions of proteins, which allows them to play a role in a variety of biological processes, such as metabolism, cell growth, and cell signaling [10, 11]. To investigate the impact of high-risk nsSNVs on PTMs, we used various in-silico tools, including CSS-Palm, iDNA-Methyl, iHyd-PseAAC, BDM-PUB, NetNGlyc, NetPhos 3.1, and GPS 3.0, to predict probable PTM sites (Table 4). As indicated by the CSS-Palm tool, cysteine at position 23 has a palmitoyl group that would be lost if cysteine changed to serine. Consequently, this mutation will increase the hydrophobicity of proteins and reduce protein-protein interactions and

trafficking. The iHyd-PseAAC server predicts a possible hydroxylation site in the AIMP2 protein. The iHyd-PseAAC results show that the wild-type protein contains hydroxyproline at three positions (21, 39, and 311). This implies that mutations on C23S remove hydroxyproline from residue 21. We used the NetPhos 3.1 and GPS 3.0 servers to predict the phosphorylation sites. GPS 3.0 predicted that 46 residues (Ser 23, Thr 16, and Tyr) would be phosphorylated, while NetPhos 3.1 indicated that 26 residues (Ser 17, Thr 7, and Tyr 2) might be phosphorylated. Based on these results, I122S, P128S, and W268S mutations result in phosphorylation, while the D121G mutation leads to the loss of phosphorylation at Y116. The BDM-PUB tool predicted that I122S and W268S lead to a gain of ubiquitination at K120 and K265, respectively. Additionally, iDNA-Methyl and NetNGlyc predicted methylation and glycosylation sites in the AIMP2 protein, and the results indicated that our high-risk nsSNVs did not affect methylation or glycosylation.

3. Discussion

Since leukodystrophies are less common than other neurodegenerative diseases, it is difficult to identify genetic variants associated with them by molecular methods. Instead, in silico methods provide an efficient, economical, fast, and reliable means of identifying pathogenic variants [6]. Thus, the purpose of the current study was to determine pathogenic mutations in the AIMP2 gene through in-silico analysis.

Since missense SNVs affect protein structure and function by causing single nucleotide substitutions [8], several studies have been performed to investigate the effects of missense SNVs on the AIMP2 protein ([4]), ([5]). In contrast to earlier studies that suggested AIMP2 SNVs are linked to different types of cancer, we investigated missense mutations leading to HLD17 in the AIMP2 gene.

> As part of this study, among 343 nsSNVs found in dbSNP and Ensembl databases, we selected 24 nsSNVs with allele frequencies <0.0001 that met the selection criteria of six in-Silico tools, including SIFT, PROVEAN, PolyPhen-2, PhD-SNP, SNP & GO, and PANTHER.

In addition, we utilized the InterPro and Pfam tools to locate these nsSNVs in various domains of AIMP2. It revealed 18 nsSNVs in three different domains of the protein, where one SNV was in the AIMP2-LysRS-bd domain, 12 SNVs were in the thioredoxin 16 domain, and five were in the GST C-terminal domain.

As a further step towards reducing the number of potentially pathogenic nsSNVs, an analysis of protein stability predictions, changes in protein structure after mutations, and evolutionary conservation of amino acids using IStable, MutPred, PredictProtein, and ConSurf tools led us to select nine deleterious nsSNVs.

The study utilized GeneMANIA and STRING software tools to identify gene-gene and protein-protein interactions. The protein-protein interaction network is crucial for understanding biological processes. Using STRING, the study analyzed functional genomics data, structural assessments, and evolutionary features of the AIMP2 protein. It was found that MARS1, LARS1, IARS1, AIMP1, EPRS1, EEFIE1, RARS1, KARS1, DARS1, and QARS1 had strong functional associations with the AIMP2 protein.

Amino acid changes can affect protein structure and function, potentially leading to phenotypic changes and altered protein expression [12]. Using GeneMANIA, a network of gene-gene interactions between the AIMP2 target gene and its 20 nearest genes was constructed. Understanding the relationships between genes can be greatly enhanced by mapping these interactions. Pairs of genes can be simultaneously perturbed to observe their interactions. AIMP2's harmful nsSNPs have the potential to disrupt the interaction between AIMP2 and its associated proteins, which could result in a variety of disease conditions [13]. This highlights the importance of these co-expressed, interconnected genes in protein synthesis.

The shortlisted 9 nsSNVs were then submitted to I-TASSER and Phyre2 to generate the 3D structure of the AIMP2 protein. Based on these results, 8 nsSNVs show a higher RMSD value and a lower TM score, indicating less deviation of the mutant protein structure from the wild-type protein.

Finally, post-translational modification analysis using different tools revealed that C23S mutations remove palmitoyl and hydroxyproline groups. The I122S, P128S, and W268S mutations all lead to a gain of phosphoryl groups; the D121G mutation results in a loss of phosphorylation; and the I122S and W268S mutations cause ubiquitination in the AIMP2 protein.

According to this data, five mutations (C23S, D121G, I122S, P128S, and W268S) in the Aimp2 gene are inferred to have caused changes in charge, size, and hydrophobicity of the protein, impacting its interaction with other molecules and disrupting the natural folding process by losing hydrogen bonds in the protein core. Furthermore, considering the role of aimp2 in the MSC complex, these mutations may have disrupted the stability and formation of the aminoacyl tRNA synthetase complex, which is essential for translating genetic information into functional proteins. As a result, the synthesis of myelin proteins and the formation of the myelin sheath is significantly reduced, leading to serious clinical effects on the patient's nervous system, especially the brain and spinal cord. On the other hand, mutations L138Q, V161E, and I188N may have had minimal influence on the protein's function.

In the end, we should take several significant limitations into consideration. Our first limitation is that we only selected SNVs reported as deleterious by all six prediction tools. Therefore, to obtain a more comprehensive picture, mutations in the "milder" spectrum should be considered. Secondly, we excluded all neutral SNVs from our analysis, but some silent or neutral mutations might alter protein sequences and influence transcript levels.

Notwithstanding these limitations, the study will introduce eight key mutations, setting a framework to evaluate leukodystrophy patients from different families and populations. This may reduce the need for exome sequencing and the cost of the study by incorporating these pathogenic variants in disease panels. In addition, the annotation of variants in-Silico prediction studies will progress in parallel with international genomic studies. Nevertheless, experimental validation needs to be conducted to understand the impact of these mutations on RNA processing, protein function, and disease pathophysiology.

4. Conclusion

The present study analyzed all non-synonymous SNVs in the AIMP2 gene to identify SNVs that negatively affect the protein. The results of our study indicate that three minor mutations (L138Q, V161E, and I188N) and five major changes (C23S, D121G, I122S, P128S, and W268S) affect the structure or functionality of the AIMP2 protein. Numerous proteomic and pharmacological studies can use these SNVs as potential targets.

5. Materials and methods

dbSNP (https://www.ncbi.nlm.nih.gov/snp/?term=aimp2) and Ensembl (https://grch37.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Table?db=core;g=ENSG00000106305;r=7:6048876-6063465;t=ENST00000223029) databases were used to obtain SNVs and related information about the AIMP2 gene. The AIMP2 gene contains 5392 SNVs, including 343 missenses, 16 frameshifts, 130 synonymous, 9 stop gained, 26 splicing, 146 synonymouse, 4718 intronic variants, and 4 in-frame insertions/deletions.

5.1. Identification of damaging nsSNVs

We retrieved the amino acid sequence of AIMP2 from UniProt (https://www.uniprot.org/uniprot/Q13155). Six bioinformatic tools were used to predict deleterious nsSNVs, including SIFT (http://sift-dna.org) [14], PROVEAN(http://provean.jcvi.org) [15], PolyPhen-2-Polymorphism Phenotyping v2 (http://genetics.bwh.harvard.edu/pph2) [16], PhD-SNP-Predictor of Human Deleterious Single Nucleotide Polymorphisms (http://snps.biofold.org/phd-snpg) [17], SNVs & GO (https://snps.biofold.org/snps-and-go/snps-and-go.html) [18], and PANTHER-PSEP (https://www.pantherdb.org) [19].

PROVEAN and SIFT are both sequence homology-based tools with own algorithms. PROVEAN can predict all types of protein sequence variations, including amino acid substitutions and in-frame insertions and deletions. If the PROVEAN score is greater than −2.5, the protein variant is predicted to be neutral [15]. SIFT determines whether an amino acid substitution in a protein will be natural or deleterious. When the score is less than or equal to 0.05, the amino acid substitution is predicted to be deleterious, but it is tolerated when the score is greater than 0.05 [20].

PolyPhen 2.0 is a popular predictor to identify the effect of substitution on proteins. It classified mutations as possibly damaging, benign, and probably damaging, where the cutoff score of 0.50 indicates a possibly damaging mutation [16].

SNP&GO predicts whether or not SNPs are disease-associated by using the protein FASTA sequence and gene ontology terms. Probability Scores above 0.5 indicate a disease-associated effect of the mutation on the function of the original protein [21]. PhD-SNP forecasts whether the provided amino acid alteration causes disease or is neutral, along with the dependability index score [20]. PANTHER predicts which nsSNVs might functionally affect a protein based on a position-specific evolution preservation score. Tracing the amino acid's reconstructed direct ancestor, it measures how long a position in a protein has been preserved (in millions of years). In general, the longer a position is maintained, the more likely it is to have a negative impact. As a result, the scores were separated into three categories: probably damaging" (time >450 my), "possibly damaging" (450 my > time >200 my), and "probably benign" (time <200 my) [7].

Furthermore, global minor allele frequency and allele frequency in the Iranian population were estimated by the gnomAD (https://gnomad.broadinstitute.org/) and Iranome (http://www.iranome.ir) databases, respectively [22], [23]. SNVs in each genotypic locus with an allele frequency of less than 0.000625 in the Iranian population and less than 0.000001 in gnomAD were predicted to be a likely pathogenic variant.

We established a scoring criterion to summarize the predictive results of the software tools. If any of the six software tools predicted that an nsSNV was "benign," they would be excluded from our analysis. However, if all software identified it as a "pathogenic" or "damaging" nsSNV, it would be examined by additional software. The nsSNV that all six software programs projected to be "pathogenic" with an allele frequency of 0.0001 would be classified as an "AIMP2 high-risk pathogenic nsSNV".

5.2. Identification of nsSNVs distribution on the domains of AIMP2

InterPro (https://www.ebi.ac.uk/interpro/) and Pfam (http://pfam.xfam.org/) predicted the AIMP2 protein's domains and active sites.

InterPro is an online tool for protein classification. Its results present similar domains, sites, families, and additional information such as a description, consistent names, and GO terms for each entry [24].

Pfam is a database of protein families that includes annotations and multiple sequence alignments generated using hidden Markov models. It was created to provide a comprehensive and accurate classification of protein families and domains and is widely used by biologists because of its extensive coverage of proteins and logical naming conventions [25].

5.3. Analysis of protein stability

Istable (http://predictor.nchu.edu.tw/iStable) [26] is an integrated predictor that reports results based on mupro and i-Mutant2 algorithms. It uses a support vector machine to analyze the effect of mutations on thermodynamic free energy (ΔΔG). We submitted the reference protein sequence and the desired substituted amino acid to the Istable website at 37 °C and a pH of 7.0.

6. Identification of structural and functional properties

PredictProtein (http://ppopen.rostlab.org) [27], Mutpred2 (http://mutpred.mutdb.org/#dload) [28], and HOPE (https://www3.cmbi.umcn.nl/hope/) [29], are web-based tools that predict whether a missense mutation affects the structure and function of a protein.

PredictProtein is an online tool that predicts various protein properties, including secondary structure analysis, solvent accessibility, transmembrane helices, disulfide bonding, the impact of point mutations, gene ontology, and subcellular localization.

This software generates two types of results: it not only predicts the effects of point mutations on the AIMP2 gene using the Snap2 software but also compares the amino acid sequences of wild-type and mutant AIMP2 proteins.

Mutpred2 is another functional predictor that consists of two parallel subroutines: one predicts any change in the function and structure of a protein by aligning the protein with a series of homologous proteins, while the other predicts the molecular cause of diseases by comparison to the human proteome [8].

We submitted the AIMP2 protein sequence and specific mutations to Mutpred2, and we evaluated the results based on the p-value and the score [$p = 0.50$ indicates confidence, $p = 0.01$ indicates strong hypotheses, and a score >0.5 indicates pathogenicity].

Furthermore, a protein sequence or an accession code of a specific protein along with the wild-type and mutant residues, was imported into the HOPE (Have Our Protein Explained) pipeline. Results are based on structural changes between mutant and wild-type residues.

6.1. Analyzing protein evolutionary conservation

We assessed the protein conservation using the ConSurf tool (http://consurf.tau.ac.il), which builds a phylogenetic tree for each protein residue and calculates a conservation score for it. The input data was submitted as an amino acid or nucleic acid sequence along with a MODELER license key and specific parameters to generate the Multiple Sequence Alignment (MSA). ConSurf employs the MSA algorithm to create a phylogenetic tree and CSI-BLAST, PSI-BLAST, or BLAST to find relative homologous sequences. It also uses Bayesian or ML algorithms to compute positions' conservation scores [30], [31]. Furthermore, in ConSurf highly conserved residues are classified as functional or structural depending on where they are located on the surface or inside the protein. Some amino acids are more preserved than others because they have vital biological functions. As a result, SNVs found in these areas will be extremely harmful [32].

6.2. Prediction of gene-gene interaction

The Gene-MANIA (Gene Multiple Association Network Integration Algorithm) is an online tool that can be accessed at (https://genemania.org/). This tool uses an extensive dataset of functional relationships, including pathways, co-expression, co-localization, protein domain similarity, and interactions between genes and proteins, to identify additional genes associated with a given input set of genes. It is an accessible platform for making assumptions about gene function, analyzing gene lists, and preparing genes for functional tests [33]. In this specific case, the gene AIMP2 was inputted into the tool.

6.3. Protein–protein interaction

The STRING software program (https://string-db.org/) was used to determine protein-protein interactions. The STRING database intends to collect and consolidate this knowledge by bringing together data from several organisms that participate in known or potential protein-protein interactions [34]. Confidence scores were assigned to each, with an emphasis on high-confidence interactors (scores ≥ 0.700). When submitting protein sequences, the "minimum required interaction score" at "high confidence" (0.700) has to be specified [33].

6.4. 3D protein modeling

The 3D models of wild-type and mutant AIMP2 proteins of missense SNVs were predicted by two homology model generators: Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) [35], I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER) [9], [36].

The hidden Markov models were aligned using Phyre2 through HHsearch to improve the accuracy of arrangement and detection rate. Additionally, based on alignment to well-known protein structures, it produces several potential 3D models of the desired mutated protein. This algorithm generates findings in five categories, including secondary structure and disorder prediction, domain analysis, specific model and template information, alignment, and the prediction of the transmembrane helix [35].

Itasser is another modeling tool that searches a PDB library for template proteins with similar folds using a meta-pathway method (LOMETS). Using Monte Carlo simulation, we reassembled the replica fragments into complete models. The Low-temperature replicas generated during the simulation were clustered by SPICKER. The hydrogen bridge network was then optimized using REMO to produce the final model. The output of the Itasser was based on the RMSD, C-score, and TM-Score [9], [36].

Finally, Chimera 1.13 [37] was used to open the resultant structures. This software is a potent tool for molecular structure visualization and study of several protein features like an electrical charge, hydrogen bonding, angle and spacing between atoms and bands, exhibiting multiple protein structures, etc..

6.5. Predicting post-translational modification (PTM) sites

In post-translational modification, chemical groups are added or removed from protein structure in ways that affect cell signaling, cell organization, cell localization, and protein function [8]. Various databases are available for the analysis of specific post-translational modifications. We used iDNA-Methyl (http://www.jci-bioinfo.cn/iDNA-Methyl) [38] CSS-Palm-Sites (http://csspalm.biocuckoo.org/online.php) [39] iHyd-PseAAC (http://app.aporc.org/iHyd-PseAAC/) [40] BDM-PUB(http://bdmpub.biocuckoo.org/prediction.php) [41] NetNGlyc (https://services.healthtech.dtu.dk/services/NetNGlyc1.0/) [42], NetPhos3.1(https://services.healthtech.dtu.dk/services/NetPhos-3.1/) [43], and GPS (http://973-proteinweb.ustc.edu.cn/gps/gps_web/) [44] To analyze palmitoylation, methylation, hydroxylation, ubiquitination, glycosylation, and phosphorylation, respectively.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the dbSNP, Ensembl genome browse,r, and UniProt repository,

(https://www.ncbi.nlm.nih.gov/snp/?term=aimp2),(https://grch37.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Table?db=core;g=ENSG00000106305;r=7:6048876-6063465;t=ENST00000223029) , (https://www.uniprot.org/uniprot/Q13155).

Also, Data included in the article/supp. material/referenced in article.

Funding

Not applicable.

CRediT authorship contribution statement

Shima Farrokhi: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Atieh Eslahi:** Writing – review & editing, Supervision. **Farzaneh Alizadeh:** Data curation. **Zahra Farshchian:** Resources. **Yasamin Yousefi:** Data curation. **Majid Mojarrad:** Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e36560.

References

[1] A. Shukla, et al., Homozygosity for a nonsense variant in AIMP2 is associated with a progressive neurodevelopmental disorder with microcephaly, seizures, and spastic quadriparesis, J. Hum. Genet. 63 (1) (2018) 19–25.
[2] S.G. Park, E.C. Choi, S. Kim, Aminoacyl-tRNA synthetase–interacting multifunctional proteins (AIMPs): a triad for cellular homeostasis, IUBMB Life 62 (4) (2010) 296–302.
[3] J.W. Choi, et al., Multidirectional tumor-suppressive activity of AIMP2/p38 and the enhanced susceptibility of AIMP2 heterozygous mice to carcinogenesis, Carcinogenesis 30 (9) (2009) 1638–1644.
[4] M. Mazaheri, et al., Case report: mutation in AIMP2/P38, the scaffold for the multi-trna synthetase complex, and association with progressive neurodevelopmental disorders, Front. Genet. 13 (2022) 24.

[5] Z. Hei, et al., Retractile lysyl-tRNA synthetase-AIMP2 assembly in the human multi-aminoacyl-tRNA synthetase complex, J. Biol. Chem. 294 (13) (2019) 4775–4783.
[6] G.C. Megalencephalic, Leukodystrophy using in silico prediction methods, Genomics 2 (3) (2017) 1023.
[7] Z. Wang, et al., In silico analysis and high-risk pathogenic phenotype predictions of non-synonymous single nucleotide polymorphisms in human Crystallin beta A4 gene associated with congenital cataract, PLoS One 15 (1) (2020) e0227859.
[8] M. Arshad, A. Bhatti, P. John, Identification and in silico analysis of functional SNPs of human TAGAP protein: a comprehensive study, PLoS One 13 (1) (2018) e0188143.
[9] J. Yang, et al., The I-TASSER Suite: protein structure and function prediction, Nat. Methods 12 (1) (2015) 7–8.
[10] Y. Shiloh, Y. Ziv, The ATM protein kinase: regulating the cellular response to genotoxic stress, and more, Nat. Rev. Mol. Cell Biol. 14 (4) (2013) 197–210.
[11] C. Dai, W. Gu, p53 post-translational modification: deregulated in tumorigenesis, Trends Mol. Med. 16 (11) (2010) 528–536.
[12] A. Ajith, U. Subbiah, In silico prediction of deleterious non-synonymous SNPs in, Asian Biomed. 17 (4) (2023) 185–199.
[13] M.M. Hassan, et al., In silico analysis: HLA-DRB1 gene's variants and their clinical impact, Cell Transplant. 32 (2023) 09636897231184473.
[14] P. Kumar, S. Henikoff, P.C. Ng, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm, Nat. Protoc. 4 (7) (2009) 1073–1081.
[15] Y. Choi, et al., Predicting the Functional Effect of Amino Acid Substitutions and Indels, 2012.
[16] I.A. Adzhubei, et al., A method and server for predicting damaging missense mutations, Nat. Methods 7 (4) (2010) 248–249.
[17] E. Capriotti, R. Calabrese, R. Casadio, Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information, Bioinformatics 22 (22) (2006) 2729–2734.
[18] E. Capriotti, et al., WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation, BMC Genom. 14 (2013) 1–7.
[19] H. Tang, P.D. Thomas, PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation, Bioinformatics 32 (14) (2016) 2230–2232.
[20] F. Akhoundi, N. Parvaneh, E.-B. Modjtaba, In silico analysis of deleterious single nucleotide polymorphisms in human BUB1 mitotic checkpoint serine/threonine kinase B gene, Meta gene 9 (2016) 142–150.
[21] R. Calabrese, et al., Functional annotations improve the predictive score of human disease-related mutations in proteins, Hum. Mutat. 30 (8) (2009) 1237–1244.
[22] K.J. Karczewski, et al., The mutational constraint spectrum quantified from variation in 141,456 humans, Nature 581 (7809) (2020) 434–443.
[23] K. Banihashemi, Iranian human genome project: overview of a research process among Iranian ethnicities, Indian J. Hum. Genet. 15 (3) (2009) 88.
[24] M. Blum, et al., The InterPro protein families and domains database: 20 years on, Nucleic acids research 49 (D1) (2021) D344–D354.
[25] S. El-Gebali, et al., The Pfam protein families database in 2019, Nucleic acids research 47 (D1) (2019) D427–D432.
[26] C.-W. Chen, J. Lin, Y.-W. Chu, iStable: off-the-shelf predictor integration for predicting protein stability changes, in: BMC Bioinformatics, BioMed Central, 2013.
[27] B. Rost, G. Yachdav, J. Liu, The predictprotein server, Nucleic acids research 32 (suppl_2) (2004) W321–W326.
[28] V. Pejaver, et al., Inferring the molecular and phenotypic impact of amino acid variants with MutPred2, Nat. Commun. 11 (1) (2020) 5918.
[29] H. Venselaar, et al., Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, BMC Bioinf. 11 (1) (2010) 1–10.
[30] A. Biegert, J. Söding, Sequence context-specific profiles for homology searching, Proc. Natl. Acad. Sci. USA 106 (10) (2009) 3770–3775.
[31] T. Pupko, et al., Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues, Bioinformatics 18 (suppl_1) (2002) S71–S77.
[32] C. Berezin, et al., ConSeq: the identification of functionally and structurally important residues in protein sequences, Bioinformatics 20 (8) (2004) 1322–1324.
[33] F.S. Laskar, et al., An in silico approach towards finding the cancer-causing mutations in human MET gene, International Journal of Genomics 2023 (2023).
[34] D. Szklarczyk, et al., The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible, Nucleic acids research (2016) gkw937.
[35] L.A. Kelley, et al., The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (6) (2015) 845–858.
[36] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinf. 9 (2008) 1–8.
[37] E.F. Pettersen, et al., UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (13) (2004) 1605–1612.
[38] Z. Liu, et al., iDNA-Methyl: identifying DNA methylation sites via pseudo trinucleotide composition, Anal. Biochem. 474 (2015) 69–77.
[39] J. Ren, et al., CSS-Palm 2.0: an updated software for palmitoylation sites prediction, Protein Eng. Des. Sel. 21 (11) (2008) 639–644.
[40] Y. Xu, et al., iHyd-PseAAC: predicting hydroxyproline and hydroxylysine in proteins by incorporating dipeptide position-specific propensity into pseudo amino acid composition, Int. J. Mol. Sci. 15 (5) (2014) 7594–7610.
[41] A. Li, et al., BDM-PUB: computational prediction of protein ubiquitination sites with a Bayesian discriminant method, in: BDM-PUB: Computational Prediction of Protein Ubiquitination Sites with a Bayesian Discriminant Method, 2009.
[42] R. Gupta, E. Jung, S. Brunak, Prediction of N-glycosylation sites in human proteins, NetNGlyc 1.0 (2004). NetNGlyc website. Retrieved from, http://www.cbs.dtu.dk/services/NetNGlyc.
[43] N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, Journal of molecular biology 294 (5) (1999) 1351–1362.
[44] Y. Xue, et al., GPS: a comprehensive www server for phosphorylation sites prediction, Nucleic acids research 33 (suppl_2) (2005) W184–W187.