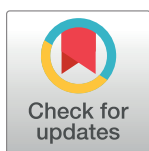


RESEARCH ARTICLE

Novel in-frame duplication variant characterization in late infantile metachromatic leukodystrophy using whole-exome sequencing and molecular dynamics simulation

Zahra Ataei^{1,2}, Zahra Nouri^{1,2}, Farial Tavakoli³, Mohammad Reza Pourreza⁴, Sina Narrei^{1,5}, Mohammad Amin Tabatabaiefar^{1,2*}



1 Department of Genetics and Molecular Biology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran, **2** Pediatric Inherited Diseases Research Center, Research Institute for Primordial Prevention of Noncommunicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran, **3** Department of Cell and Molecular Biology & Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran, **4** Core research facilities, Isfahan University of Medical Sciences, Isfahan, Iran, **5** Department of Research and Development, ERYTHROGEN Medical Genetics Lab, Isfahan, Iran

* tabatabaiefar@med.mui.ac.ir, tabatabaiefar@gmail.com

OPEN ACCESS

Citation: Ataei Z, Nouri Z, Tavakoli F, Pourreza MR, Narrei S, Tabatabaiefar MA (2023) Novel in-frame duplication variant characterization in late infantile metachromatic leukodystrophy using whole-exome sequencing and molecular dynamics simulation. PLoS ONE 18(2): e0282304. <https://doi.org/10.1371/journal.pone.0282304>

Editor: Shahid M. Baig, National Institute for Biotechnology and Genetic Engineering, PAKISTAN

Received: June 26, 2022

Accepted: February 12, 2023

Published: February 27, 2023

Copyright: © 2023 Ataei et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: This work was financially supported by Isfahan University of Medical Sciences. (grant NO.2400173). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Metachromatic leukodystrophy (MLD) is a neurodegenerative lysosomal storage disease caused by a deficiency in the arylsulfatase A (ARSA). ARSA deficiency leads to sulfatide accumulation, which involves progressive demyelination. The profound impact of early diagnosis on MLD treatment options necessitates the development of new or updated analysis tools and approaches. In this study, to identify the genetic etiology in a proband from a consanguineous family with MLD presentation and low ARSA activity, we employed Whole-Exome Sequencing (WES) followed by co-segregation analysis using Sanger sequencing. Also, MD simulation was utilized to study how the variant alters the structural behavior and function of the ARSA protein. GROMACS was applied and the data was analyzed by RMSD, RMSF, Rg, SASA, HB, atomic distance, PCA, and FEL. Variant interpretation was done based on the American College of Medical Genetics and Genomics (ACMG) guidelines. WES results showed a novel homozygous insertion mutation, c.109_126dup (p. Asp37_Gly42dup), in the ARSA gene. This variant is located in the first exon of ARSA, fulfilling the criteria of being categorized as likely pathogenic, according to the ACMG guidelines and it was also found to be co-segregating in the family. The MD simulation analysis revealed this mutation influenced the structure and the stabilization of ARSA and led to the protein function impairment. Here, we report a useful application of WES and MD to identify the causes of a neurometabolic disorder.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Leukodystrophies are a group of usually inherited disorders that affect the white matter of the central nervous system [1]. Metachromatic leukodystrophy (MLD) (OMIM 250100), as an autosomal recessive neurodegenerative disorder, is one of the most common leukodystrophies. MLD is one of the lysosomal storage diseases (LSDs) with a prevalence of 1.45 per 100,000 births worldwide [2]. MLD is a sphingolipidosis caused by lysosomal enzyme arylsulfatase A (ARSA) deficiency or its sphingolipid activator protein B (SapB) [3]. The ARSA gene is located on chromosome 22q13 with eight exons spanning a genomic region of 3kb and encoding a 509 amino acids [4, 5] ARSA is responsible for the hydrolysis of the 3-O ester bond of sphingolipid 3'-O-sulfogalactosylceramide, known as sulfatide. ARSA deficiency leads to an increase in the sulfatide within oligodendrocytes, macrophages, and some subtypes of neurons in the CNS, in Schwann cells, and in the peripheral nervous system (PNS) macrophages, which exhibit metachromatic staining characteristics [6]. Instability of the myelin sheath, change in calcium homeostasis, cell stress, apoptosis, and inflammatory response are consequences of sulfatide deposition in the cell cytoplasm [1, 7].

Depending upon the age which symptoms present, MLD is classified into three different clinical forms: late infantile, juvenile, and adult forms [8]. Late-infantile MLD, the most severe type with poor prognosis, is characterized by gait abnormalities, seizures, ataxia, hypotonia, extensor planters, optic atrophy, and regression of motor skills that lead to complete gross motor deterioration before the age of 40 months [9, 10].

There is no newborn screening test available for MLD diagnosis, and MLD is diagnosed after birth [11]. Progressive demyelination and subsequent neurological symptoms, biochemical procedures, genetic analysis, and imaging results should be applied for a specific diagnosis of MLD. Biochemical assays, including the quantification of Sulfatide accumulation in urine, and ARSA activity in peripheral blood leukocytes, are used for accurate diagnosis of MLD [12].

Next generation sequencing techniques provide an opportunity for diagnosis of hereditary genetic diseases in research and clinical settings. Whole exome sequencing (WES) has been applied in identifying genetic variants associated with a variety of diseases [13, 14]. Moreover, the biocomputational techniques have been recently applied broadly as a potential tool to investigate the variant effect on a mutant protein structure, in a rapid and cost-effective manner [15, 16]

In the current study, for accurate diagnosis of a 34-month patient, the first child of consanguineous parents with a progressive decline of motor and cognitive abilities, hypotonia, and spasticity, we used WES in the diagnosis of MLD, followed by imaging analysis and enzymatic tests aimed at confirming the molecular results, which led to a novel in-frame mutation identification. Furthermore, Molecular dynamics (MD) simulation was applied to investigate the ARSA in wild type form (WT-ARSA) and structural and conformational changes of mutant form of ARSA (mutant-ARSA) to understand the pathogenic mechanism of MLD disease at atomic level.

Material and methods

Subject and clinical evaluations and imaging results

The study was approved by the Research Ethics Committee of “Alzahra Research Center” (grant no:2400173, IR.ARI.MUI.REC.1400.011), and the patient and his parents were recruited to the study after obtaining informed consent. The proband was a 34-month boy and the first child of consanguineous Iranian parents who were first cousins without a family history of neurological disease. A detailed clinical examination and comprehensive family history were done by a clinical geneticist.

ARSA enzymatic assay

ARSA activity was estimated as $\mu\text{g}/\text{mg}$ protein in leukocytes, using 4-nitrocatechol sulfate. In brief: 0.25 mM sodium pyrophosphate was utilized to inactivate arylsulfatase B (ARSB). Then, the amount of sulfate released was measured by the absorbance of free 4-nitrocatechol at 515 nm on a spectrophotometer (Beckmann Coulter, Brea, CA, USA), which is associated with sulfatase activity [17].

Metabolic panel test

Amino Acid Profile and AcylCarnitine Profile were investigated in plasma using LC-MS/MS and MS/MS, respectively. Also, Organic Acids and Acylglycines measurements in urine were performed through GC-MS and LC-MS/MS.

Whole exome sequencing

Blood samples were taken from the patient and his parents. Using a Qiagen DNA extraction kit, genomic DNA was extracted from peripheral blood lymphocytes, and assessment of its purity, was done on a Nanospec Cube Biophotometer (Nanolytik®, Dusseldorf, Germany). The sample was sent to MacroGen (South Korea) (<https://www.macrogen.com/>) for WES analysis using the Novaseq 4000 platform (Illumina, San Diego, CA, USA) with the mean depth of coverage 100X. These samples were sheared into 151-bp fragments by a hydrodynamic shearing system (Covaris, Massachusetts, USA), and whole exome was captured through in-solution targeted genomic enrichment using Agilent SureSelect Human All Exon kit v6 (Agilent Technologies, CA, USA).

Bioinformatics analysis

Following sequencing, image analysis and base-calling were performed using the standard Illumina data analysis pipeline Real-Time Analysis (RTA) version (RTA) v1.12.4. CASAVA v1.8.2. The raw reads quality was assessed using the FastQC [18]. The low quality reads have been removed with TRIMMOMATIC [19] and quality control was done after trimming. Reads were mapped to the human reference genome build UCSC hg19 (<http://genome.ucsc.edu/>) using BWA (Burrows-Wheeler Aligner) (<http://bio-bwa.sourceforge.net/>). SAMtools was used to convert sequence alignment map (SAM) format to sorted, indexed binary alignment map files [20]. GATK software tools (<https://gatk.broadinstitute.org/hc/en-us>) were used to improve alignments and genotype calling. annotation was performed using ANNOVAR [21].

Missense, nonsense, start codon change, stop loss, indel variants, and splice site with Minor Allele Frequency <1% in dbSNP version 147, 1000 genomes project phase 3 database (<https://www.internationalgenome.org/>), NHLBI GO exome sequencing project (ESP) (<https://evs.gs.washington.edu/EVS/>), exome aggregation consortium (ExAC) (<https://exac.broadinstitute.org/>), Iranome database (<http://www.iranome.ir/>), and our locally developed database (GTAC) are considered for further analysis. The novelty of the variant was investigated in the Human Gene Mutation Database (HGMD) (<http://www.hgmd.cf.ac.uk/ac/index.php>) and the literature. Finally, the identified variants' pathogenicity can be interpreted according to ACMG 2015 standards and guidelines for interpreting sequence variants [22]. The MEGA6 software was utilized to investigate the conservation of the mutated region among several species [23].

Variant confirmation

The candidate variant (p.Asp37_Gly42dup) that was identified by WES was subsequently confirmed using polymerase chain reaction (PCR) and bidirectional Sanger sequencing in the proband using the SeqStudio Genetic Analyzer (Applied Biosystem Inc., Foster City, CA, USA). Then, to examine the segregation of genotype among the family members, co-segregation analysis was performed on his parents. Primers (forward primer: 5'-GTATTTGGGTCCGGGGTCTC-3' and the reverse primer: 5'-TGTGGCCTTCCCTAGAGAGA-3' (designed by Primer 3 software Input version 0.4.0 (<https://primer3.ut.ee/>) and NCBI primer BLAST software)) encompassed exon 1 of the ARSA gene. Chromatogram sequences files were compared with the reference sequence (NM_000487), via SeqMan software version 5.00© (DNASTAR, Madison, WI, USA).

Molecular modeling

Crystal structure of WT-ARSA protein in complex with NAG1 and NAG2 was obtained from Protein Data Bank (PDB ID: 1AUK with resolution 2.10 Å). Only protein molecule was retained and all excess molecules were removed from the complex using Discovery Studio 2016 Visualizer software (DS 2016) (DS 4.0, Accelrys Software Inc., San Diego, CA). Swiss Model webserver [24] was applied in order to repair the chain breaks in the WT-ARSA and construct a model for the mutant-ARSA. The modeled WT-ARSA is composed of 485 residues starting with residue Arg1 and ending with residue Pro485 while the modeled mutant-ARSA is composed of 491 residues containing the six-amino acid insertion mutation in 23–28 position (Asp23, Leu24, Gly25, Cys26, Tyr27, and Gly28) starting with residue Arg1 and ending with residue Pro491. The amino acid compositions of the active site of WT-ARSA are Ala10, Asp11, Asp12, FGly51, Arg55, Lys105, His107, His211, Asp263, Asn264, Lys284 and those of mutant-ARSA are Ala10, Asp11, Asp12, FGly57, Arg61, Lys111, His113, His217, Asp269, Asn270, and Lys290.

Molecular dynamics simulation

In order to investigate the structural and conformational changes in ARSA proteins, GROMACS 2018 package was applied to perform molecular dynamics simulation [25]. PDB2PQR web server [26] was retrieved to determine the protonation state (PH = 5) of His residues in both proteins. All the required files, topology and coordinate files, for both molecules were constructed using GROMACS [25] through CHARMM27 all-atom force field [27].

A dodecahedron TIP3P water box with a direction of 9 Å as a unit cell was constructed to solvate all models [28]. In order to calculate each system in water, the neutralization of the negative charges of the system was carried out through adding Na⁺ ions. Energy-minimization of the solvated system was performed applying 50000 steps of steepest-descent method to remove steric clashes. Equilibration of the minimized systems with position restraint on the proteins were performed by NVT and NPT ensembles for 400 ps at a temperature of 300 K and pressure of 1 bar, respectively.

V-rescale temperature [29] and Parrinello-Rahman pressure [30] coupling methods were applied to stabilize the temperature and pressure at 300 K and 1 bar for the system. 100-ns MD simulations were carried out for ARSA proteins under periodic boundary condition with the time step of 2 fs applying LINCS [31] and Partial Mesh Ewald (PME) [32] algorithms. The Coulomb and van der Waals interactions were calculated by cut-off value of 1.2 nm.

Analysis of MD simulations

Production simulations were analyzed applying trajectory analysis modules in the GROMACS simulation package and visualized using VMD [33], Root mean square deviation (RMSD),

root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), hydrogen bond (HB), secondary structure, atomic distance, principal component analysis (PCA) and gibbs free energy landscape (FEL) were obtained using GROMACS and Grace-5.1.22/QTGrace v0.2.6 program was retrieved to visualize all the graphs.

Principal component analysis. Principal component analysis (PCA) was calculated to investigate the biomolecules motion [34] applying calculating the covariance matrix C:

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (1)$$

Where x_i and x_j are the coordinate of the i th and j th atoms of the systems and $\langle x_i \rangle$ and $\langle x_j \rangle$ represent the average coordinate of the i th and j th atoms over the ensemble. Then, the principal components (PCs) are calculated by diagonalization of covariance matrix.

Free energy landscape. Free energy landscape (FEL) was calculated to investigate the conformational changes of biomolecules to identify the stable state and transient state of proteins and to assign their stability and their function [35]. The FEL can be calculated as:

$$\Delta G(X) = -K_B T \ln P(X) \quad (2)$$

Where $\Delta G(X)$ is the free energy, K_B and T , and $P(X)$ represent the Boltzmann constant, absolute temperature, and the probability distribution of the conformation ensemble along the PCs.

Results

Clinical evaluations

The proband was a 34-month boy and the first child of consanguineous Iranian parents without a family history of neurological disease. The function of kidneys, liver, thyroid, and heart was normal. He was born at full term via normal vaginal delivery after an uncomplicated pregnancy.

Prenatal screening tests did not show any abnormalities in the fetus. At the age of 15 months, he presented with spastic gait, and he could not walk independently. However, MRI results suggested a normal brain. Clonazepam was administered orally until two years, and manifestations of spasticity were controlled. The patient had no swallowing difficulties. At 32 months, the child had language disorder, weakness started in the hands, muscle spasms progressed, and muscular hypotonia was evident. Electroencephalogram (EEG) showed mildly abnormal, and MRI results show evidence of bilateral symmetrical abnormal signal intensity in deep white matter, centrum semiovale, paraventricular regions, and subcortical region as low signal intensity in T1, high signal intensity in T2W and flair images.

ARSA enzyme activity

The enzyme activity using the colorimetric method was 0.066 mu/mg protein. The reference was 0.37–1.815.

Metabolic panel

There was no abnormality in the metabolic profile of the patient in the mentioned factors.

Molecular findings

Upon exome analysis, a total number of 28135 variants was obtained. The number of variants was limited by applying the following criteria for variant filtering, and a homozygous insertion of 18 nucleotides (**c.109_126dup**) in the *ARSA* gene was found. It causes in-frame insertion in

exon 1 of this gene, resulting in a protein with 515 residues (versus 509 residues in the intact protein). The in-frame variant was absent from dbSNP version 147, 1000 genomes project phase 3, NHLBI GO ESP, ExAC, Iranome, and our locally developed (GTAC) databases. It was not found in the literature, either. The six residues are inserted into a highly conserved region of ARSA, as evident through multiple-species alignment (Fig 1). The novel in-frame variant co-segregated with the disease in the family and was heterozygous in parents but homozygous in the patient (Fig 2). Given the PM2, PM4, PP3, PP4 ACMG/AMP criteria being fulfilled and the consistent phenotype, the c.109_126dup in ARSA was considered likely pathogenic in the patient.

Molecular dynamics simulation

Molecular dynamics simulation method was employed to investigate the effect of the variant on the ARSA structure and function. To do this, 100-ns MD simulations were performed for WT-ARSA and mutant-ARSA to calculate all the required data. The obtained final structures of both ARSA proteins from 100-ns MD simulation were superimposed and are displayed in S1 Fig. No significant structural changes were found in the mutant-ARSA protein except in the loop shape region is composed of mutant residues Asp23, Leu24, Gly25, Cys26, Tyr27, Gly28.

RMSD and RMSF were calculated for each production simulation. The average all-atom RMSD values for both ARSA proteins relative to the initial structures were calculated as 0.16 and 0.17 nm, respectively (Table 1 and Fig 3A).

Residual flexibility of both proteins was obtained and is displayed in Fig 3B. The RMSF graphs reveal that this variant significantly increased the fluctuation of residues in the loop shape regions, particularly the regions are composed of mutant residues in 23–28 position and residues in 480–485 position, and the regions with residues 160–163. In addition, the RMSF graphs show that the fluctuation of active site entrance residues of mutant-ARSA including Pro77, Gly78, Val79, His139, Gly158, Asp161, Gln162, Gly163, Tyr218, Thr274, Arg276, His393 are also increased, while the fluctuation of active site residues of mutant-ARSA was small.

The radius of gyration, which is defined as R_g , shows the compression and density of protein structure and stability, and the more compact a protein is, the more stable it will be [36]. The average R_g values of WT-ARSA and mutant-ARSA were calculated as 2.18 and 2.21 nm, respectively (Table 1). The graphs of R_g data between both proteins are displayed in Fig 3C. The mutant-ARSA demonstrates a significant increment in R_g value.

The hydrophobic core region of both ARSA proteins was investigated by calculating the total and residue solvent accessible surface area (SASA). Remarkable increment in the average of total SASA and relatively small increase in the average of active site SASA were obtained in the mutant-ARSA compared to WT-ARSA (Table 1 and Fig 4A). Among the active site residues of mutant-ARSA, residues Ala10, FGly57, Arg61, Lys111 and Asp269 with SASA of 0.02, 0.38, 0.061, 0.105 and 0.072 nm² respectively indicate the most accessibility to solvent compared to those of in WT-ARSA with SASA of 0.002, 0.125, 0.011, 0.04, and 0.017 nm² respectively (Fig 4B).

Hydrogen bond (HB) analysis during MD simulation was investigated internally and the same analysis was conducted between the proteins and the solvent (protein-solvent) to understand the stability and the solubility of the ARSA proteins. The average number of intramolecular and protein-solvent HB networks and the patterns of them were obtained (Table 1 and Fig 5A and 5B).

As can be seen in Table 1 and Fig 5A, there is no substantial difference between the number of intramolecular HB of ARSA proteins, while a significant increment in the number of

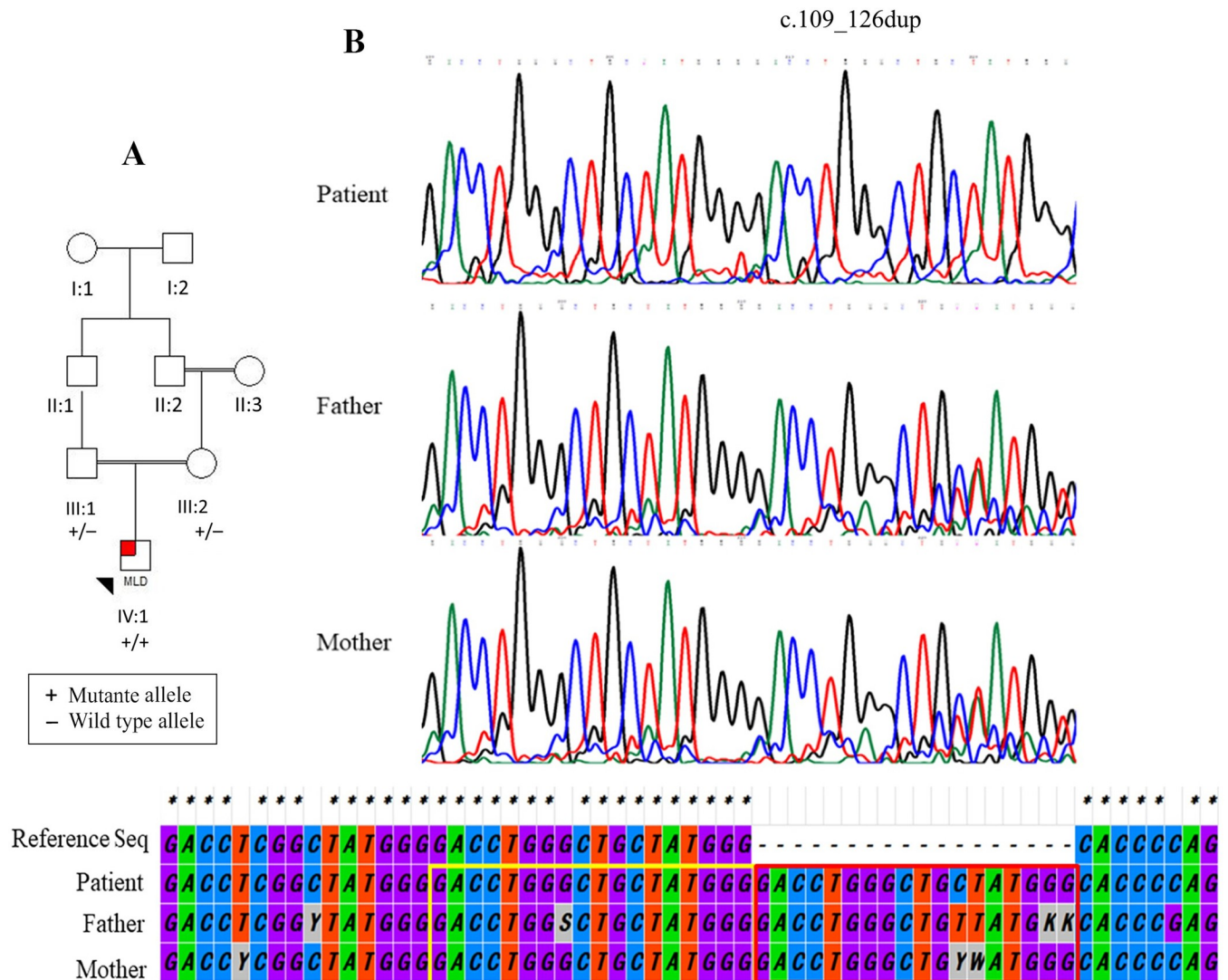


Fig 2. (A) Pedigree chart of the family. (B) Chromatogram of the homozygous sequence variant c.109_126dup detected in the ARSA gene in the proband compared with the heterozygous sequence in the parents. The comparison of three sequences with the reference sequence is given at the bottom of the electropherograms. The red and yellow boxes indicate the duplicated and the previous sequences, respectively.

<https://doi.org/10.1371/journal.pone.0282304.g002>

calculated by projecting the protein backbone of the MD trajectory on eigenvectors 1 and 2 (Fig 6A). The projection of the PC1 and PC2 of a trajectory can represent a cluster of stable states in protein. As can be seen in Fig 6A, the mutant-ARSA displays a transition towards a more distant region of the phase spaces compared to WT-ARSA.

Free energy landscape (FEL) was calculated for two systems applying the first two PCs as reaction coordinates to find the conformational states of ARSA proteins. The FEL maps

Table 1. The average of various parameters during 100-ns MD simulation of wild type and mutant ARSAs.

Proteins	RMSD (nm)	Rg (nm)	Total SASA (nm ²)	Active site SASA (nm ²)	No. of Intramolecular HB	No. of protein-solvent HB
WT	0.130	2.18	189.41	0.075	337	799
mutant	0.167	2.21	197.12	0.125	336	833

<https://doi.org/10.1371/journal.pone.0282304.t001>

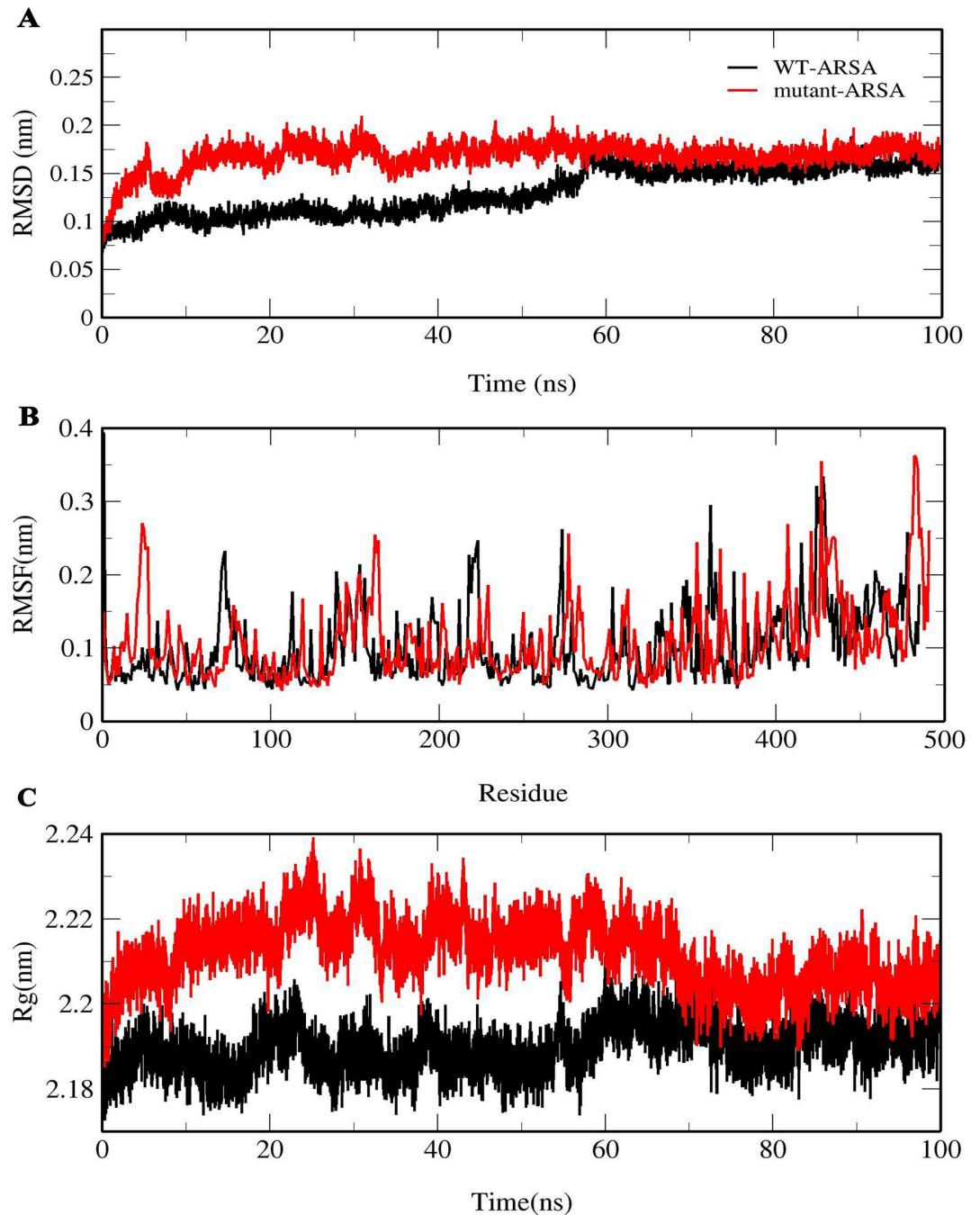


Fig 3. Conformational changes in ARSA proteins. (A) all-atom RMSD, (B) RMSF and (C) Rg of WT-ARSA and mutant-ARSA.

<https://doi.org/10.1371/journal.pone.0282304.g003>

provide noticeable data on the diverse conformational states to the ARSAs in the 100 ns-MD simulation (Fig 6B and 6C). It is observed from Fig 6 that the PC1 and PC2 motion modes of the mutant-ARSA occupied larger spaces than those of the WT-ARSA, illustrating the conformational rearrangements which lie at the root of the mutation.

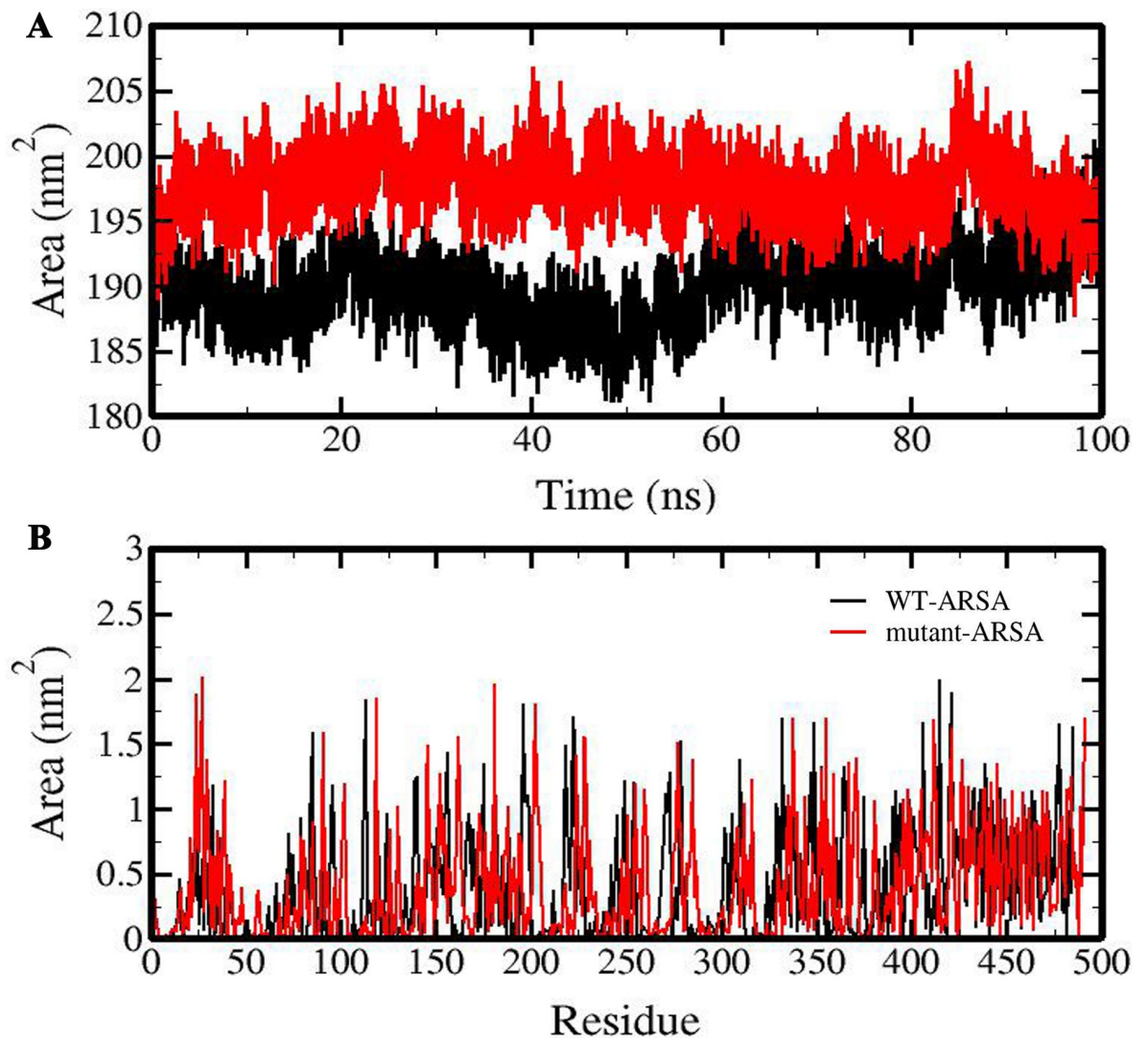


Fig 4. Graphical representation of solvent accessible surface area (SASA) of ARSA proteins. (A) Total SASA and (B) active site SASA of WT-ARSA and mutant-ARSA.

<https://doi.org/10.1371/journal.pone.0282304.g004>

Discussion

Owing to the rarity of MLD and the heterogeneity in its presentation, early diagnosis of MLD is challenging. To treat these individuals at an early stage of disease, accurate methods will be required to obtain an early diagnosis. Patients with MLD who get Haematopoietic Stem Cell Transplantation (H SCT) or gene therapy earlier in the disease have a better prognosis [11, 37, 38].

As a common molecular diagnostic tool, WES had a significant impact on identifying causative variations and determining the appropriate illness management strategy [39]. In the case of MLD, NGS can assist in illness identification, differentiating Pseudodeficiency diagnosis to avoid erroneous diagnoses based on ARSA activity, and identifying juvenile and adult variants of MLD to consider early treatment.

According to the Human Gene Mutation Database (HGMD) (<http://www.hgmd.cf.ac.uk/ac/index.php>), 303 ARSA mutations have been reported. Amongst Iranian MLD patients, c.931G>A(p.Gly311Ser) and c.465+1G>A variants are the most frequent alleles [40].

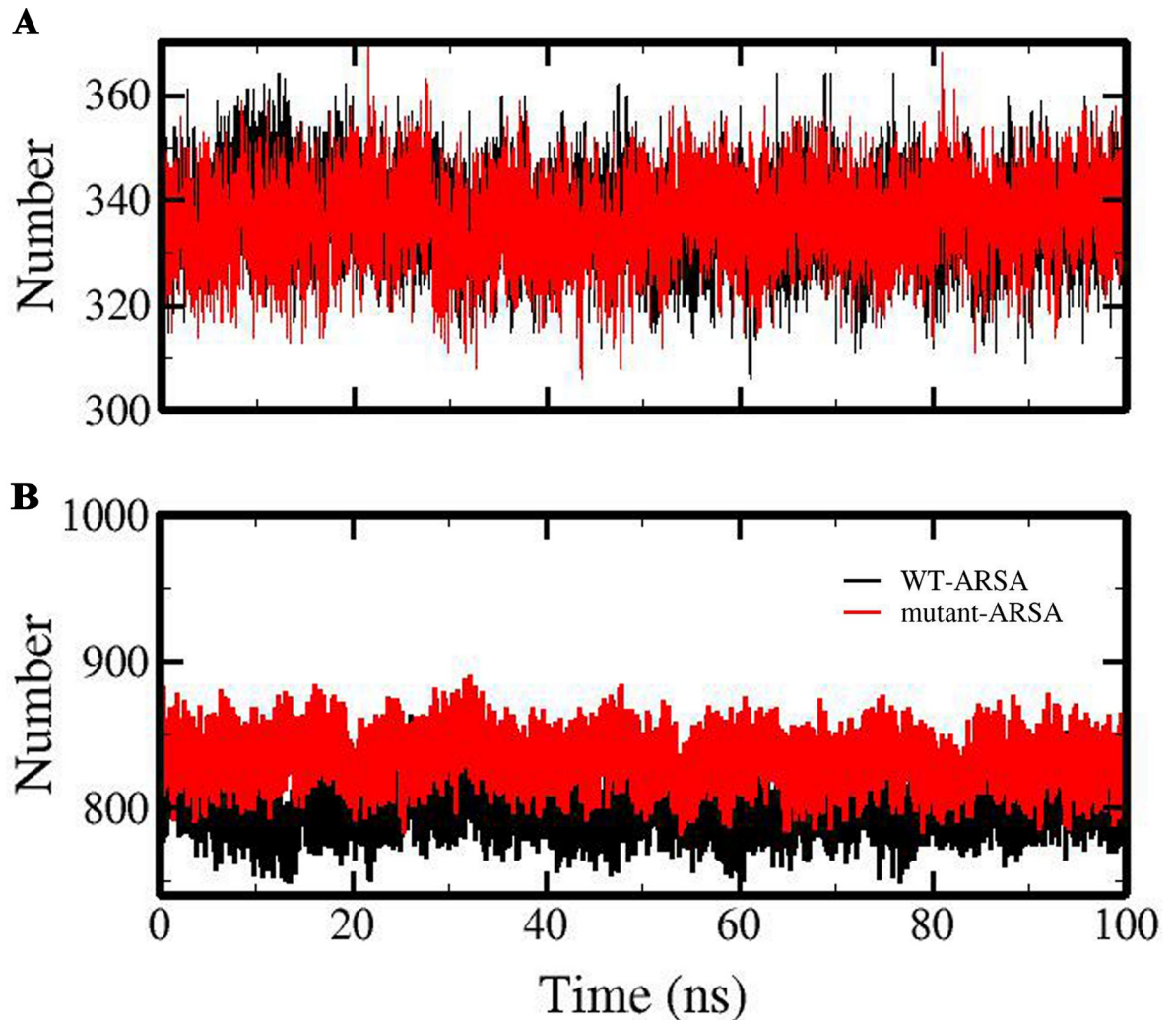


Fig 5. The pattern of hydrogen bonds during 100 ns MD simulation for ARSA proteins. (A) Intramolecular and (B) protein-solvent hydrogen bonds of WT-ARSA and mutant-ARSA.

<https://doi.org/10.1371/journal.pone.0282304.g005>

The structure of the DNA sequence around the 18-base-pair tandem duplication allele was investigated. In the normal sequence, we found a region with two short repeated sequences (Fig 7), which is indicative of a replication slippage. Polymerase slippage, that is expected to be as the possible cause of up to 75% of all indels, might explain the mechanism through which the mutation was generated [41]. This repeat, in our case is imperfect (GACCTCGGC and GACCTGGGC) (Fig 7), and the duplicated region is (GACCTGGGCTGCTATGGG). As a hypothesis, it appears polymerase slippage can cause this duplication at the genomic level, and subsequently leading to protein instability, and the resultant clinical manifestation in patients. We suggest further functional studies for interpreting the pathogenic mechanism of MLD.

In computational part of our study, the protein structure of the mutant-ARSA was constructed using Swiss Model server, then structural and functional effects of the six-amino acid insertion mutation in the ARSA protein were studied using 100 ns MD simulation to investigate the stability of mutant protein and the molecular mechanism of disease.

The RMSD plots of both proteins showed that they are constant during the simulation, with RMSD value of approximately 0.15 nm. The RMSF plots displayed that the most mutation

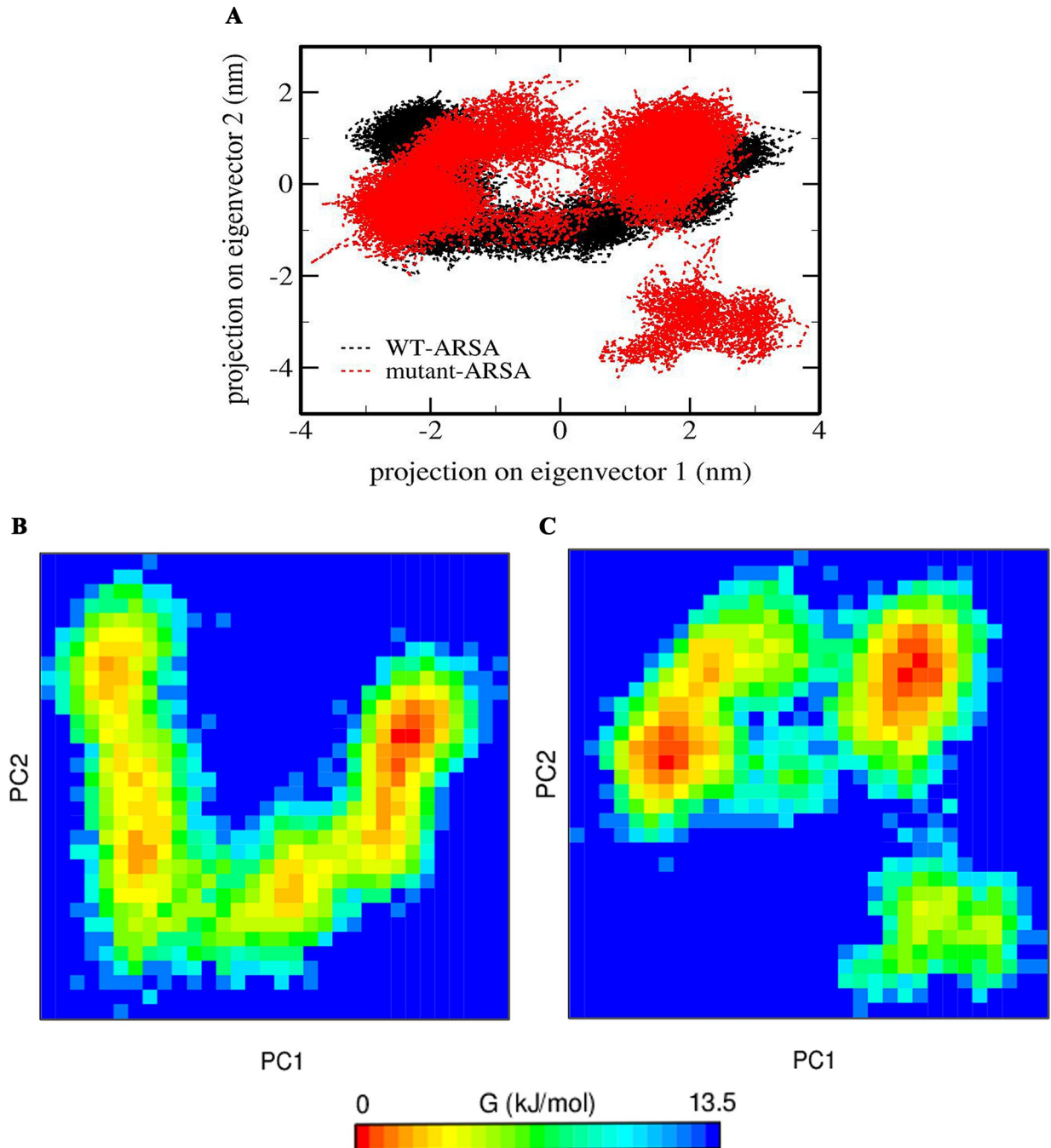


Fig 6. Conformational sampling and Free energy landscape (FEL) analysis of ARSA proteins. (A) Conformational sampling of WT-ARSA and mutant-ARSA proteins by 2D projection of the MD trajectory on PC1 and PC2. (B) FEL of WT-ARSA and (C) FEL of mutant-ARSA.

<https://doi.org/10.1371/journal.pone.0282304.g006>

effect on the fluctuation of residues was related to the loop shape regions and the active site entrance residues. It revealed that the regions with high fluctuation have a low contribution to creating contact with other amino acids [42], particularly the loop regions in positions 160–163, 480–485 and 23–28 with mutant residues. A comparison of Rg data of both proteins revealed that the Rg value of mutant-ARSA increased significantly, indicating a loss in compactness of the protein.

Wilde Type c.94 **GACCTCGGC**TATGGGG**GACCTGGGCT**GTCTATGGG

Mutated allele c.94 **GACCTCGGC**TATGGGG**GACCTGGGCT**GTCTATGGGG**GACCTGGGCT**GTCTATGGG

Fig 7. DNA sequences of Wilde Type ARSA and the mutated allele. The 5' one is shown in red (GACCTCGGC) and the 3' one is shown in blue (GACCTGGGC). The underlined sequence is duplicated in the mutated allele.

<https://doi.org/10.1371/journal.pone.0282304.g007>

The investigation of the hydrophobic core region, a key parameter to assess protein folding and stability [43] in both proteins, displayed a noticeable increment in average total SASA and a small increase in average active site SASA in mutant-ARSA, demonstrating a large surface exposed to the solvent and this could be due to the solvent exposure of hydrophobic amino acids and subsequently influencing the protein folding [35, 43]. Furthermore, the HB analysis revealed that although there was no significant change in intramolecular HB of both proteins, the average number of protein-solvent HB in mutant-ARSA increased considerably, which indicates higher solubility than that of the WT-ARSA [44]. The insignificant reduction in the intramolecular HB formation in mutant- ARSA might be due to the fact that the variant occurs in a highly flexible loop at the protein surface and, subsequently it has minimal effect on the existing interactions [42, 45, 46], as clarified from the RMSF data. As elucidated from our Rg data as a measure of the protein stability, change in compression and size of mutant-ARSA could have been a plausible reason behind the changing in protein folding, increasing its solubility and following instability of mutant-ARSA [15]. Above all, the examination of the atomic distance between amino acid residues pairs of the active site in both proteins as a function of time demonstrated clearly that the mutation causes instability in mutant-ARSA active site, whereas the active site structure of WT-ARSA was stable in the course of simulation.

The most important parameters, principal component, and free energy landscape analysis were retrieved to investigate the global motion, folding, function, and, eventually stability of the protein [34, 47]. PCA demonstrated that mutant-ARSA covered a wide range of phase spaces compared to WT-ARSA. Given that the proteins perform their function via collective atomic motions and the stability of a protein is related to its collective atomic motion, PCA results indicate that the underlying reason for the impairment of the protein function might be given rise to an increase in the overall motion in mutant-ARSA. FEL analysis revealed that, compared to the WT-ARSA which displayed a single global energy minimum basin, mutant-ARSA illustrated wide global energy minima via a transition state which demonstrates the mutation caused the conformational rearrangement in protein. Indeed, the existence of several energy minima in the conformational space of mutant-ARSA proves the significant destabilization of the protein [47].

According to the *in silico* study, although no remarkable changes were observed in the overall structure, the secondary structure content and internal HB of both proteins, due to the position of the mutation in the flexible loop region far from the active site [45, 46, 48], changes in compactness, solvent accessibility, protein-solvent hydrogen bond, atomic distance measurement, protein motion, and FEL analysis obviously demonstrate the instability and dysfunction of the mutant-ARSA, which is in agreement with a remarkable reduction in ARSA Activity (0.066 mu/mg pro).

Conclusion

In this paper, we report the successful application of WES for diagnosis and proper genetic counseling of MLD. The *in silico* data analysis of the ARSA protein provided an in-depth comprehension of destabilization and loss of conformational dynamics of the mutant-ARSA and

elucidated clearly that the structure and function of mutant-ARSA are negatively influenced due to the mutation. The approach would be highly useful in pre-clinical studies for the discovery of novel therapeutic and management strategies for MLD.

Supporting information

S1 Fig. Schematic representation of obtained final structures of ARSA proteins after 100 ns MD simulation. (A) Superimposition of final structure WT-ARSA over mutant-ARSA. (B) Ribbon view showing the position of mutant amino acid residues. WT-ARSA and mutant-ARSA are shown in golden and green color respectively.
(JPG)

S2 Fig. Secondary structure changes of ARSA proteins during 100 ns MD simulation. (A) WT-ARSA and (B) mutant-ARSA.
(JPG)

S3 Fig. Distance calculation between C α atoms of active site residues of WT-ARSA and mutant-ARSA as function of time (100 ns). Amino acid residues pairs of WT-ARSA active site are Ala10—Asp11, Asp11—Asp12, Asp12—FGL51, FGL51—Arg55, Arg55—Lys105, Lys105—His107, His107—His211, His211—Asp263, Asp263—Asn264, Asn264—Lys284, Lys284—Ala10 and those of mutant-ARSA are Ala10—Asp11, Asp11—Asp12, Asp12—FGL57, FGL57—Arg61, Arg61—Lys111, Lys111—His113, His113—His217, His217—Asp269, Asp269—Asn270, Asn270—Lys290, Lys290—Ala10.
(ZIP)

Acknowledgments

We would like to thank all individuals who cooperated in this study.

Author Contributions

Conceptualization: Mohammad Amin Tabatabaiefar.

Data curation: Zahra Ataei, Zahra Nouri, Mohammad Amin Tabatabaiefar.

Formal analysis: Fariyal Tavakoli, Mohammad Reza Pourreza, Sina Narrei.

Investigation: Zahra Ataei, Zahra Nouri, Fariyal Tavakoli.

Methodology: Zahra Ataei, Mohammad Amin Tabatabaiefar.

Project administration: Zahra Ataei, Mohammad Amin Tabatabaiefar.

Software: Zahra Nouri, Fariyal Tavakoli, Mohammad Reza Pourreza, Sina Narrei.

Supervision: Mohammad Amin Tabatabaiefar.

Validation: Zahra Ataei, Zahra Nouri, Mohammad Amin Tabatabaiefar.

Visualization: Zahra Nouri, Fariyal Tavakoli.

Writing – original draft: Zahra Ataei, Fariyal Tavakoli.

Writing – review & editing: Zahra Nouri, Mohammad Amin Tabatabaiefar.

References

1. van der Knaap MS, Bugiani M. Leukodystrophies: a proposed classification system based on pathological changes and pathogenetic mechanisms. *Acta neuropathologica*. 2017; 134(3):351–82. <https://doi.org/10.1007/s00401-017-1739-1> PMID: 28638987
2. Giugliani R. Newborn screening for lysosomal diseases: current status and potential interface with population medical genetics in Latin America. *Journal of inherited metabolic disease*. 2012; 35(5):871–7. <https://doi.org/10.1007/s10545-011-9436-z> PMID: 22231381
3. Barboura I, Maire I, Ferchichi S, Miled A. Metachromatic Leukodystrophy Clinical, Biological and Therapeutic Aspects. Latest Findings in Intellectual and Developmental Disabilities Research: IntechOpen; 2012.
4. Cesani M, Lorioli L, Grossi S, Amico G, Fumagalli F, Spiga I, et al. Mutation update of ARSA and PSAP genes causing metachromatic leukodystrophy. *Human mutation*. 2016; 37(1):16–27. <https://doi.org/10.1002/humu.22919> PMID: 26462614
5. Stein C, Gieselmann V, Kreysing J, Schmidt B, Pohlmann R, Waheed A, et al. Cloning and expression of human arylsulfatase A. *Journal of Biological Chemistry*. 1989; 264(2):1252–9. PMID: 2562955
6. Biffi A, Lucchini G, Rovelli A, Sessa M. Metachromatic leukodystrophy: an overview of current and prospective treatments. *Bone Marrow Transplantation*. 2008; 42(2):S2–S6. <https://doi.org/10.1038/bmt.2008.275> PMID: 18978739
7. Barkovich AJ. Concepts of myelin and myelination in neuroradiology. *American Journal of Neuroradiology*. 2000; 21(6):1099–109. PMID: 10871022
8. Van Rappard DF, Boelens JJ, Wolf NI. Metachromatic leukodystrophy: disease spectrum and approaches for treatment. *Best Practice & Research Clinical Endocrinology & Metabolism*. 2015; 29(2):261–73. <https://doi.org/10.1016/j.beem.2014.10.001> PMID: 25987178
9. Lynch DS, Wade C, de Paiva ARB, John N, Kinsella JA, Merwick Á, et al. Practical approach to the diagnosis of adult-onset leukodystrophies: an updated guide in the genomic era. *Journal of Neurology, Neurosurgery & Psychiatry*. 2019; 90(5):543–54.
10. Kehr C, Blumenstock G, Gieselmann V, KRÄGELOH-MANN I, LEUKONET G. The natural course of gross motor deterioration in metachromatic leukodystrophy. *Developmental Medicine & Child Neurology*. 2011; 53(9):850–5. <https://doi.org/10.1111/j.1469-8749.2011.04028.x> PMID: 21707604
11. Sessa M, Lorioli L, Fumagalli F, Acquati S, Redaelli D, Baldoli C, et al. Lentiviral haemopoietic stem-cell gene therapy in early-onset metachromatic leukodystrophy: an ad-hoc analysis of a non-randomised, open-label, phase 1/2 trial. *The Lancet*. 2016; 388(10043):476–87. [https://doi.org/10.1016/S0140-6736\(16\)30374-9](https://doi.org/10.1016/S0140-6736(16)30374-9) PMID: 27289174
12. Shaimardanova AA, Chulpanova DS, Solovyeva VV, Mullagulova AI, Kitaeva KV, Allegrucci C, et al. Metachromatic leukodystrophy: diagnosis, modeling, and treatment approaches. *Frontiers in medicine*. 2020; 7:576221. <https://doi.org/10.3389/fmed.2020.576221> PMID: 33195324
13. Wang Z, Lin Y, Zheng D, Yan A, Tu X, Lin J, et al. Whole-exome sequencing identifies compound heterozygous mutations in ARSA of two siblings presented with atypical onset of metachromatic leukodystrophy from a Chinese pedigree. *Clinica Chimica Acta*. 2016; 460:135–7. <https://doi.org/10.1016/j.cca.2016.06.038> PMID: 27374302
14. Sarmadi A, Mohammadi A, Tabatabaei F, Nouri Z, Chaleshtori MH, Tabatabaiefar MA. Molecular genetic study in a cohort of Iranian families suspected to maturity-onset diabetes of the young, reveals a recurrent mutation and a high-risk variant in the CEL gene. *Advanced biomedical research*. 2020;9.
15. Sadr AS, Eslahchi C, Ghassempour A, Kiaei M. In silico studies reveal structural deviations of mutant profilin-1 and interaction with riluzole and edaravone in amyotrophic lateral sclerosis. *Scientific reports*. 2021; 11(1):1–14.
16. Tavakoli F, Ganjalikhany MR. Structure-based inhibitory peptide design targeting peptide-substrate binding site in EGFR tyrosine kinase. *PloS one*. 2019; 14(5):e0217031. <https://doi.org/10.1371/journal.pone.0217031> PMID: 31116768
17. Wenger D, Williams C. Screening for lysosomal disorders. *Techniques on Diagnostic Human Biochemical Genetics A Laboratory Manual* New York, NY. 1991; 10003:587–617.
18. Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.
19. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943

21. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*. 2010; 38(16):e164–e. <https://doi.org/10.1093/nar/gkq603> PMID: 20601685
22. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine*. 2015; 17(5):405–23. <https://doi.org/10.1038/gim.2015.30> PMID: 25741868
23. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*. 2013; 30(12):2725–9. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122
24. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*. 2014; 42(W1):W252–W8. <https://doi.org/10.1093/nar/gku340> PMID: 24782522
25. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1:19–25.
26. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic acids research*. 2004; 32(suppl_2):W665–W7. <https://doi.org/10.1093/nar/gkh381> PMID: 15215472
27. Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl E. Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *Journal of chemical theory and computation*. 2010; 6(2):459–66. <https://doi.org/10.1021/ct900549r> PMID: 26617301
28. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A*. 2001; 105(43):9954–60.
29. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *The Journal of chemical physics*. 2007; 126(1):014101. <https://doi.org/10.1063/1.2408420> PMID: 17212484
30. Nosé S, Klein M. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*. 1983; 50(5):1055–76.
31. Hess B. P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation*. 2008; 4(1):116–22.
32. Sagui C, Darden TA. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure*. 1999; 28(1):155–79. <https://doi.org/10.1146/annurev.biophys.28.1.155> PMID: 10410799
33. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics*. 1996; 14(1):33–8. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570
34. David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. *Protein dynamics*: Springer; 2014. p. 193–226.
35. Tanwar H, George Priya Doss C. An Integrated Computational Framework to Assess the Mutational Landscape of α -L-Iduronidase IDUA Gene. *Journal of Cellular Biochemistry*. 2018; 119(1):555–65.
36. Mlu L, Bogatyreva N, Galzitskaia O. Radius of gyration is indicator of compactness of protein structure. *Molekuliarnaia biologii*. 2008; 42(4):701–6. PMID: 18856071
37. van Rappard DF, Boelens JJ, van Egmond ME, Kuball J, van Hasselt PM, Oostrom KJ, et al. Efficacy of hematopoietic cell transplantation in metachromatic leukodystrophy: the Dutch experience. *Blood, The Journal of the American Society of Hematology*. 2016; 127(24):3098–101. <https://doi.org/10.1182/blood-2016-03-708479> PMID: 27118454
38. van der Knaap MS, Schiffmann R, Mochel F, Wolf NI. Diagnosis, prognosis, and treatment of leukodystrophies. *The Lancet Neurology*. 2019; 18(10):962–72. [https://doi.org/10.1016/S1474-4422\(19\)30143-7](https://doi.org/10.1016/S1474-4422(19)30143-7) PMID: 31307818
39. Di Resta C, Galbiati S, Carrera P, Ferrari M. Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities. *Ejifcc*. 2018; 29(1):4. PMID: 29765282
40. Mahdih N, Sharifi A, Rabbani A, Ashrafi M, Tavasoli AR, Badv RS, et al. Novel disease-causing variants in a cohort of Iranian patients with metachromatic leukodystrophy and in silico analysis of their pathogenicity. *Clinical Neurology and Neurosurgery*. 2021; 201:106448. <https://doi.org/10.1016/j.clineuro.2020.106448> PMID: 33385934
41. Montgomery SB, Goode DL, Kvikstad E, Albers CA, Zhang ZD, Mu XJ, et al. The origin, evolution, and functional impact of short insertion–deletion variants identified in 179 human genomes. *Genome research*. 2013; 23(5):749–61. <https://doi.org/10.1101/gr.148718.112> PMID: 23478400

42. Yu H, Yan Y, Zhang C, Dalby PA. Two strategies to engineer flexible loops for improved enzyme thermostability. *Scientific reports*. 2017; 7(1):1–15.
43. Savojardo C, Manfredi M, Martelli PL, Casadio R. Solvent accessibility of residues undergoing pathogenic variations in humans: from protein structures to protein sequences. *Frontiers in molecular biosciences*. 2021; 7:460. <https://doi.org/10.3389/fmolb.2020.626363> PMID: 33490109
44. Mohammadian H, Mahnam K, Sadeghi HM, Ganjalikhany MR, Akbari V. Rational design of a new mutant of tobacco etch virus protease in order to increase the in vitro solubility. *Research in Pharmaceutical Sciences*. 2020; 15(2):164. <https://doi.org/10.4103/1735-5362.283816> PMID: 32582356
45. Batori V, Koide A, Koide S. Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. *Protein engineering*. 2002; 15(12):1015–20. <https://doi.org/10.1093/protein/15.12.1015> PMID: 12601141
46. Khan MT, Ali S, Zeb MT, Kaushik AC, Malik SI, Wei D-Q. Gibbs free energy calculation of mutation in pncA and rpsA associated with pyrazinamide resistance. *Frontiers in molecular biosciences*. 2020; 7:52. <https://doi.org/10.3389/fmolb.2020.00052> PMID: 32328498
47. Amir M, Mohammad T, Kumar V, Alajmi MF, Rehman M, Hussain A, et al. Structural analysis and conformational dynamics of STN1 gene mutations involved in coat plus syndrome. *Frontiers in molecular biosciences*. 2019; 6:41. <https://doi.org/10.3389/fmolb.2019.00041> PMID: 31245382
48. Studer RA, Dessailly BH, Orengo CA. Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *Biochemical journal*. 2013; 449(3):581–94. <https://doi.org/10.1042/BJ20121221> PMID: 23301657