

# Molecular Insights into Arylsulfatase B Mutation-Induced Instability in Mucopolysaccharidosis Type VI



Barka Basharat<sup>1</sup>, Nushrat Jahan<sup>1</sup> and Mohammad Asim Azhar<sup>2,\*</sup>

<sup>1</sup>Department of Biotechnology, School of Chemical and Life Sciences, Jamia Hamdard, Hamdard Nagar, New Delhi, India

<sup>2</sup>Organ Transplant Center of Excellence, King Faisal Specialist Hospital and Research, Riyadh, Kingdom of Saudi Arabia

\*Corresponding author: Barka Basharat, Organ Transplant Center of Excellence, King Faisal Specialist Hospital and Research, Riyadh, Kingdom of Saudi Arabia, E-mail: [azharasim@gmail.com](mailto:azharasim@gmail.com)

## Abstract

Mutations in the arylsulfatase B (ARSB) gene are directly implicated in mucopolysaccharidosis type VI (MPS VI). Several non-synonymous single-nucleotide polymorphisms (nsSNPs) in ARSB have been associated with disease pathogenesis. A comprehensive evaluation of these variants is essential to understand their structural and functional consequences. In this study, a systematic *in silico* analysis was performed to identify deleterious nsSNPs in the ARSB gene. Initially, 430 nsSNPs were evaluated using sequence-based prediction tools, including SIFT, PolyPhen-2, FATHMM, and Mutation Assessor. Subsequently, 141 nsSNPs were subjected to structure-based stability analysis using MAESTROweb, SDM2, mGSM, and DynaMut2, of which 57 variants overlapped with previous reports. High-confidence deleterious nsSNPs were further assessed for pathogenicity using PMut and MutPred2 servers. Our integrated computational approach identified 44 highly deleterious mutations. Aggregation propensity analysis revealed that 29 of these variants exhibit increased aggregation tendencies, while one variant demonstrated progressive loss of solubility. Molecular dynamics simulations further indicated that high-confidence deleterious nsSNPs significantly disrupt ARSB structural integrity, enhance molecular flexibility, reduce structural rigidity, and promote atomic-level aggregation. Overall, this study provides mechanistic insights into how pathogenic mutations destabilize the ARSB protein and contribute to MPS VI pathogenesis, highlighting potential targets for future therapeutic investigation.

**Keywords:** ARSB gene, mucopolysaccharidosis type VI, non-synonymous SNPs, pathogenic mutations, protein stability, protein aggregation, genetic variation, computational mutagenesis

Received: April 5th, 2025 Accepted: August 25th, 2025 Published: January 5th, 2026

## 1. Introduction

Lysosomes function as acidic cellular hubs for macromolecule catabolism, recycling, and signaling via hydrolytic enzymes and efflux permeases. Lysosomal storage diseases (LSD), also known as dysostosis multiplex, are due to an acquired shortage of specific lysosomal enzymes (Platt et al., 2018). These diseases are frequently linked to numerous impairments of musculoskeletal growth and formation. Deficiencies in lysosomal enzymes or associated proteins cause substrate accumulation, such as sphingolipids, glycosaminoglycans (GAGs), or oligosaccharides, leading to over 41 progressive LSDs, most of which are autosomal recessive, except for X-linked Fabry disease and MPS II (Scerra et al., 2022). These disorders exhibit heterogeneous phenotypes, from neonatal lethality (e.g., non-immune hydrops fetalis) to later-onset forms (e.g., Niemann-Pick type C), with multi-organ involvement including neurological, skeletal, ophthalmic, and visceral symptoms. Pathophysiology stems from intra-lysosomal storage disrupting signaling pathways, mitochondrial function (e.g., fragmented cristae and reduced membrane potential in MPS VI, GM1 gangliosidosis), and tissue integrity; LSDs are classified by accumulated substrates, such as sphingolipidoses,

mucopolysaccharidoses (MPS), and oligosaccharidosis (Gros & Muller, 2023).

Mucopolysaccharidosis type VI (MPS VI), also known as Maroteaux-Lamy syndrome, is an autosomal recessive lysosomal storage disorder caused by mutations in the arylsulfatase B (ARSB) gene, which encodes N-acetylgalactosamine-4-sulfatase (Tobacman & Bhattacharyya, 2022). This enzyme hydrolyzes sulfate groups from dermatan sulfate (DS) and chondroitin-4-sulfate, preventing their lysosomal accumulation (Rossi et al., 2025). GAGs are major components of the extracellular matrix and play critical roles in connective tissue structure, cell signaling, and tissue integrity. Proper ARSB activity ensures normal lysosomal function, cellular homeostasis, and continuous turnover of connective tissue components in skin, cartilage, tendons, and other organs. Deficient ARSB activity leads to progressive GAG buildup in multiple tissues, manifesting as dysostosis multiplex with musculoskeletal dysplasia, joint stiffness, corneal clouding, cardiac valve disease, hepatosplenomegaly, and respiratory complications (Leal et al., 2025).

Clinical severity varies widely; severe cases present symptoms by age 2–3, with loss of ambulation by age 10 and survival into

the third decades (Leal et al., 2025). Diagnosis integrates clinical evaluation, urinary GAG quantification, enzymatic assays, and ARSB genotyping. While primarily somatic, MPS VI shares skeletal and visceral features with other MPS disorders, driven by GAG-mediated connective tissue pathology, including alterations in the extracellular matrix and cardiac infiltrates (Lipiński et al., 2025). Musculoskeletal defects, joint rigidity, ocular obscuration, cardiac issues, and breathing difficulties are among the clinical symptoms of MPS VI (De Ponti et al., 2022).

Single-nucleotide polymorphisms (SNPs) represent the most common genomic variants, driving evolutionary adaptation and disease susceptibility (Sauna & Kimchi-Sarfaty, 2022). Non-synonymous single-nucleotide polymorphisms (nsSNPs) in ARSB represent a major mutational class, potentially disrupting protein stability, folding, or interactions via amino acid substitutions in conserved domains (Sinha et al., 2022). Over 200 ARSB variants have been reported, yet their structural-functional impacts remain incompletely characterized, complicating prognosis and therapy selection amid emerging options such as enzyme replacement, gene therapy, and substrate reduction (Gomez-Ospina, 2024; Tobacman & Bhattacharyya, 2022).

This study employs an integrated *in silico* pipeline comprising SIFT, PolyPhen-2, FATHMM, Mutation Assessor, MAESTROweb, SDM2, mCSM, DynaMut2, PMut, MutPred2, and molecular dynamics simulations to systematically evaluate 430 ARSB nsSNPs (Enni, 2025). Figure 1 illustrates computational methods that integrate sequence-, structure-, and dynamics-based approaches to systematically identify pathogenic variants in ARSB associated with MPS VI. We prioritize high-confidence deleterious variants, assess their aggregation propensity and changes in stability, and elucidate their mechanistic contributions to MPS VI pathogenesis, thereby informing precision diagnostics and therapeutic strategies (Sarachakov et al., 2025).

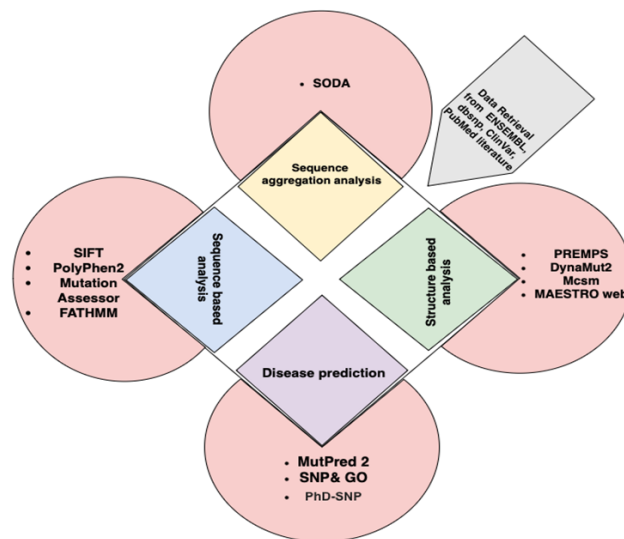
## 2. Materials and Methods

### 2.1. Data retrieval

The FASTA sequence of the human ARSB gene was obtained from the UniProt database (UniProt ID: P15848). A database of missense mutation SNPs was created using information gathered through a PubMed literature review and databases such as dbSNP, HGMD, ClinVar, and Ensemble. The list was cleared of the redundant nsSNPs. The Protein Data Bank (PDB ID: 1FSU) provided the crystal structure of human ARSB. The remaining mutations were obtained from the Ensembl databases, and all mutation types, including missense, non-synonymous, and synonymous mutations, were included (Dyer et al., 2025).

### 2.2. Sequence-Based Prediction of Deleterious Mutations

Sorting Intolerant From Tolerant (SIFT: <http://sift.jcvi.org/>) was used to predict the functional impact of amino acid substitutions based on sequence homology and evolutionary conservation (Sim et al., 2012). The underlying principle is that functionally important residues are conserved across species; therefore, substitutions at highly conserved positions are more likely to be deleterious. SIFT assigns a tolerance score ranging from 0 to 1, where scores  $\leq 0.05$  indicate damaging (intolerant) substitutions and scores  $> 0.05$  suggest tolerated variants. In addition to missense variants, SIFT can also evaluate certain 3-nucleotide indels that result in amino acid insertions or deletions. These predictions are based on conservation patterns derived from multiple sequence alignments.



**Figure 1.** A description of the computational methods used to forecast the pathogenicity of the ARSB gene mutation at the structural and functional domains.

#### 2.2.1. PolyPhen-2

PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) used to predict the potential structural and functional impact of amino acid substitutions using sequence-based and structure-based features. It incorporates evolutionary conservation, physicochemical differences between residues, and structural parameters such as solvent accessibility and proximity to functional domains (Adzhubei et al., 2013). The tool calculates the difference in Position-Specific Independent Count (PSIC) scores between wild-type and mutant residues. Variants are classified as benign, possibly damaging, or probably damaging, depending on the PSIC score difference and predictive confidence.

#### 2.2.2. Mutation Assessor

Mutation Assessor (<http://mutationassessor.org/>) evaluates the functional impact of amino acid substitutions based on evolutionary conservation patterns within protein families and subfamilies (Su et al., 2025). It distinguishes between general conservation across homologous proteins and functional specificity within subfamilies. Each variant is assigned a Functional Impact (FI) score and categorized as neutral, low, medium, or high impact. Variants with FI scores  $> 2.0$  are generally considered functionally deleterious.

#### 2.2.3. FATHMM

Functional Analysis Through Hidden Markov Models (FATHMM) predicts the pathogenicity of missense mutations using Hidden Markov Models (HMMs) and conservation-based weighting schemes (Shihab, 2021). It can incorporate disease-specific or cancer-specific models. Variants are classified as tolerated or deleterious based on a threshold score, with lower scores indicating a higher likelihood of pathogenicity.

### 2.3. Structure-Based Stability Prediction

#### 2.3.1. MAESTROweb

MAESTROweb (<https://pbwww.che.sbg.ac.at/maestro/web>) predicts mutation-induced changes in protein stability using machine-learning methods trained on experimental thermodynamic datasets. It calculates the change in Gibbs free

energy ( $\Delta\Delta G$ ) between wild-type and mutant proteins (Laimer et al., 2016). A negative  $\Delta\Delta G$  value indicates destabilization of protein structure, whereas a positive value suggests stabilization. The tool also identifies mutation “hotspots” by scanning multiple residues.

### 2.3.2. mCSM

mCSM predicts the impact of mutations on protein stability and macromolecular interactions using graph-based structural signatures derived from atomic distance patterns. It estimates  $\Delta\Delta G$  values, where negative scores indicate destabilizing mutations (Pires et al., 2014). Variants with significantly negative  $\Delta\Delta G$  values are considered likely to disrupt protein stability and function.

### 2.3.3. DynaMut2

DynaMut2 (<http://biosig.unimelb.edu.au/dynamut/>) integrates Normal Mode Analysis (NMA) with graph-based signatures to evaluate mutation-induced changes in protein stability and dynamics (Rodrigues et al., 2021). It predicts both  $\Delta\Delta G$  and changes in vibrational entropy ( $\Delta\Delta S$ ), allowing assessment of alterations in structural flexibility. This tool is particularly useful for understanding how mutations affect conformational dynamics, well to thermodynamic stability.

### 2.3.4. PremPS

PremPS (<https://lilab.jysw.suda.edu.cn/research/PremPS/>) predicts the impact of missense mutations on protein stability using a balanced dataset of stabilizing and destabilizing mutations (Chen et al., 2020). It integrates structural and evolutionary features to estimate  $\Delta\Delta G$  values and classify variants as stabilizing or destabilizing.

## 2.4. Pathogenicity Prediction Tools

### 2.4.1. MutPred2

MutPred2 (<http://mutpred.mutdb.org>) is a machine learning-based tool that predicts the pathogenicity of amino acid substitutions and provides mechanistic insights into possible molecular alterations (Pejaver et al., 2020). It assigns a pathogenicity probability score ranging from 0 to 1. Variants with scores  $> 0.5$  (commonly  $> 0.589$  for high confidence) are considered likely pathogenic. Additionally, it predicts alterations in secondary structure, post-translational modification sites, catalytic residues, and protein-protein interactions.

### 2.4.2. SNPs&GO

SNPs&GO integrates sequence features and Gene Ontology (GO) annotations using a Support Vector Machine (SVM) classifier to distinguish disease-associated variants from neutral polymorphisms. The inclusion of functional annotations improves predictive accuracy (Capriotti et al., 2013).

### 2.4.3. PhD-SNP

PhD-SNP (<https://snps.biofold.org/phd-snp/phd-snp.html>) is an SVM-based predictor that classifies missense mutations as disease-associated or neutral using sequence and evolutionary profile information. Variants with scores  $> 0.5$  are predicted to be disease-causing (Calabrese et al., 2008).

## 2.5. Evolutionary Conservation Analysis

### 2.5.1. ConSurf

ConSurf (<https://consurf.tau.ac.il/>) evaluates evolutionary conservation of amino acid residues using multiple sequence alignment and phylogenetic analysis (Ashkenazy et al., 2016). Conservation scores range from 1 (variable) to 9 (highly

conserved). Highly conserved exposed residues are often functionally important, whereas conserved buried residues are typically structural (architectural). Disease-associated mutations frequently occur at highly conserved positions.

## 2.6. Aggregation Propensity Analysis

### 2.6.1. SODA

SODA (Solubility based on Disorder and Aggregation) predicts the impact of mutations on protein solubility, secondary structure, and aggregation propensity. It helps identify variants that may promote protein misfolding or aggregation, which is relevant in disorders associated with lysosomal dysfunction (Paladin et al., 2017).

### 2.6.2. Arpeggio

Arpeggio analyzes interatomic interactions within protein structures by categorizing them into hydrogen bonds, van der Waals contacts, hydrophobic interactions, and ionic interactions. It accepts PDB structures and generates quantitative interaction profiles, enabling comparison between wild-type and mutant proteins to assess structural disruption (Jubb et al., 2017).

## 3. Results and Discussion

MPS VI, a lysosomal storage disease, is caused by ARSB deficiency (Tomanin et al., 2018). The accumulation of GAGs caused by this enzyme deficiency leads to a variety of clinical symptoms, including skeletal deformities, stiff joints, and corneal clouding. MPSVI usually has serious progression, with symptoms that start in early childhood and worsen over time (Pohl et al., 2018). Determining the molecular mechanisms driving this condition requires knowledge of the variants in the ARSB gene and their effects on protein function. In this investigation, we looked at the possibility that harmful mutations in ARSB contribute to the pathophysiology of MPSVI.

We sought to identify alterations that could have a major effect on the stability and function of the ARSB protein by employing a comprehensive strategy that combines sequence- and structure-based analyses (Leal et al., 2025). Sequence-based methods, including SIFT, PolyPhen2, FATHMM, and Mutation Assessor, were used to predict the detrimental consequences of mutations. Structure-based techniques were used to evaluate the effects of changes on protein structure and aggregating propensity, including mCSM, DynaMut2, MAESTROweb, and PremPS. Our study's key finding is that these mutations may affect protein solubility, a defining characteristic of MPS VI. We thoroughly analysed 1224 ARSB gene variants obtained from the dbSNP and Ensembl databases, as well to other mutations identified. We investigated the effects of these alterations on the structure and function impacts of variations or mutations within the ARSB gene.

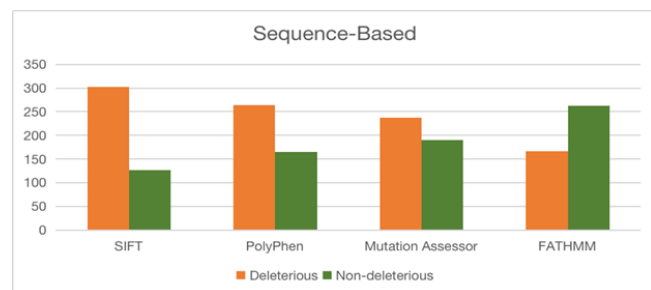
Five web-based resources were used to help the sequence-based evaluation: SIFT, PolyPhen2, FATHMM, and Mutation Assessor. Simultaneously, the structure-based method assessed single-point amino acid alterations in ARSB using mCSM, DynaMut2, MAESTROweb, and PremPS. Only mutations that the analyses showed to be high-confidence variations were sent for additional testing. We used the PhD-SNP and MutPred2 websites to explain the illness characteristics associated with these high-confidence variants, as will be covered in the sections that follow.

### 3.1. Deleterious mutations identification of nsSNPs

Sequence-based evaluation was performed on all nsSNPs utilising five online tools: SIFT, PolyPhen2, PROVEAN, Mutation Assessor, FATHMM Using the structure-based approach, 141 nsSNPs

located in the ARSB gene were analysed (Table S1) using PremPs, MAESTROweb, DynaMut2, and mCSM. As the sole region identified through experimentation and available in the PDB database, only the highly probable nsSNPs have been selected for additional investigation. Using the PMut web server and MutPred2, high-probability nsSNPs' disease phenotypes were identified. We utilised FATHMM, SIFT, PolyPhen2, and SNPs&GO. Based on protein physical characteristics, SIFT classifies variants as either intolerant or tolerated; higher scores indicate a higher likelihood of toxicity. The study of 1224 single-point amino acid changes in ARSB using the sequence-based approach produced predictions from SIFT, PolyPhen2, Mutation Assessor, and FATHMM. These tools, particularly, highlighted the dangerous substitutions 302, 264, 238, and 167 (Figure 2). By combining these sequence-based predictive techniques, we were able to examine the possible molecular effects of mutations on functioning.

We used four different structure-based prediction technologies: DynaMut2, MAESTROweb, PremPS, and mCSM (Figure 3). By computing folding free energy, these tools assess the stability of variations by analysing atomic coordinates from the PDB file of the wild-type protein. Many of these tools use a machine-learning approach, combining different biophysics-based methodologies to predict the impact of variations on protein stabilisation. Folded ( $G_f$ ) and unfolded ( $G_u$ ) forms are computed in thermodynamics using the formula  $\Delta G = G_u - G_f$ .  $\Delta\Delta G = G_m - G_w$ , wherein  $G_m$  is the energy of the protein that has been altered, and  $G_w$  is the energy of the wild-type protein, is used to assess the change in protein stability and the landscape of energy. A variation that stabilises a protein is indicated by a negative  $\Delta\Delta G$  value, whereas a mutation that destabilises the protein is suggested by a positive  $\Delta\Delta G$  score.

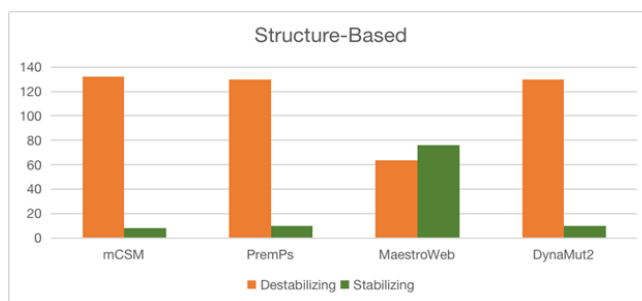


**Figure 2.** Deleterious mutations ARSB gene sequence-based approach. This graph illustrates the effects of mutation using a computational approach.

Concurrently, destabilising mutations were identified at 132, 130, 64, and 130 positions by structure-based predictions from mCSM, DynaMut2, MAESTROweb, and PremPS. To increase confidence in our results, we chose for further investigation only the modifications predicted by all sequence-based and structure-based methods to be detrimental. 44 amino acid changes that were considered harmful and destabilising were identified through this sorting process. These 44 changes were then analysed to investigate any possible connections to disease characteristics.

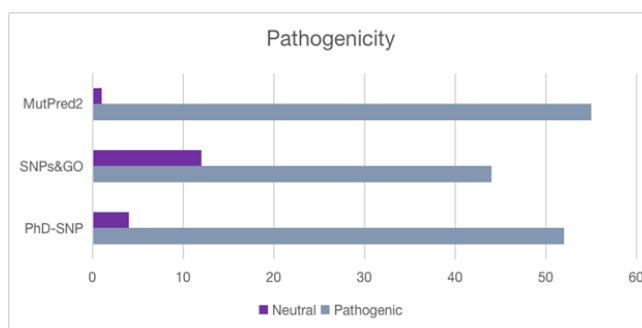
### 3.2. Identification of pathogenic mutations using a computational approach

We evaluated disease characteristics linked to mutations using PhD-SNP, SNAP & GO, and MutPred2 approaches. These tools classify variants based to potential disease associations and assign



**Figure 3.** Deleterious mutations ARSB gene structure-based approach. This graph illustrates the effects of mutation using a computational approach.

pathogenicity levels (Figure 4). Among the 57 high-confidence variations identified through a sequence- and structure-based evaluation, PMut and MutPred predicted 44 as pathogenic. Only 44 mutations, A237D, C91R, E323K, E483D, G149R, G302R, G324V, G56D, G64R, H393P, I296N, I67N, K145E, L129P, L236P, L360P, L498P, L51P, L72P, L72R, L82P, L82R, L90P, L98P, L98R, P93R, R315P, R315Q, R327G, R327Q, R388T, T92K, V277G, V80G, W146R, W146S, W353R, W438G, Y138C, Y175D, Y210C, Y266S, Y86C, Y86N) out of 57 highly probable nsSNPs were shown to be harmful using the disease phenotype assessment algorithm (Table ?? and Table ??).



**Figure 4.** Pathogenic mutations predicted in the ARSB gene were predicted using structure-based tools.

### 3.3. Analysis of conserved residue

The ConSurf tool was utilised to analyse the human ARSB genome structure and assess residue conservation (Ashkenazy et al., 2016). According to the ConSurf study, out of 44 final mutations, these 12 (A237D, G56D, G64R, L51P, I67N, L236P, W353R, V277G, T92K, R315Q, R315P, H393P) can be highly disease-causing mutations, because they belonged in high conserved areas and mutations in these regions (Figure 5).

### 3.4. Analysis of aggregation propensity

The aggregation, illness, helix, and strand propensities resulting from alterations are computed using SODA48. Out of 12 nsSNPs, 2 (A237D, W353R) were shown (Table 1) to lower the protein's solubility, and 9 of them raised it, out of the 11 mutations revealed using illness phenotype prediction. The solubility of a protein significantly impacts its function. Insoluble proteins tend to congregate, which can lead to illnesses including Parkinson's, amyloidosis, and Alzheimer's. By combining solubility data with structural and sequence-based properties, SODA provides insights into protein regions susceptible to aggregation and suggests



**Figure 5.** Conserved residue of the ARSB region.

modifications that may improve solubility (Aggidis et al., 2024; Kumar et al., 2016; Paladin et al., 2017).

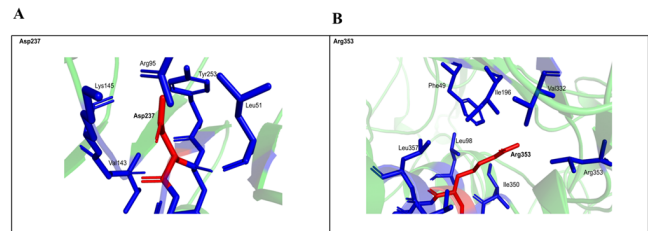
**Table 1.** Aggregation propensity prediction of mutant ARSB protein using SODA server.

Sequence	SODA	Remark
A237D	-0.471	Less soluble
G56D	0	More soluble
G64R	4.558	More soluble
H393P	4.84	More soluble
I67N	1.308	More soluble
L236P	1.72	More soluble
L51P	11.762	More soluble
R315P	4.59	More soluble
R315Q	4.03	More soluble
T92K	2.878	More soluble
V277G	3.236	More soluble
W353R	-0.676	Less soluble

Of the 44 alterations, 12 vnsSNPs decrease the protein's solubility. The fact that all three solubility-reducing alterations can contribute to the formation of amino acid clumps and the subsequent pathophysiology of illnesses makes this particularly concerning. It turned out that the dissolution of the amino acid was enhanced by the additional 29 nsSNPs.

Thorough conformational analyses yielded a deeper understanding of the molecular effects of these pathogenic alterations on the ARSB protein. We found that the pathogenicity associated with these changes may be driven by the addition or removal of interatomic noncovalent interactions, such as

hydrophobic contacts, van der Waals forces, and hydrogen bonds (Table 2). Demonstrating structure and its interactions in Figure 6. Mutations can cause structural defects that contribute to the pathophysiology of illnesses by causing misfolding, loss of function, or enhanced agglomeration.



**Figure 6.** Mutant structure (A). Alanine mutated to Aspartic acid and (B). Tryptophan mutated to Arginine.

The significant rise in overall contacts, particularly proximal and VdW clashes, indicates increased steric hindrance and overcrowding of the altered protein. An increase in hydrogen bonds or polar interactions can counterbalance some of the destabilising effects of greater overcrowding and improve the long-term stability of the protein's structure. A denser and perhaps stressed structure is suggested by the A237D alteration, which causes a substantial rise in total acquaintances, particularly through proximal and van der Waals collisions (Table 2). This suggests that, although the A237D mutant induces structural stress, compensatory stabilising interactions may allow the protein's structure to retain its integrity (Figure 6).

Protein interactions require both hydrogen bonds and polar interactions to be specific and stable (Sheng et al., 2015). A decrease in these interactions may impair the protein's functioning

**Table 2.** Comparing changes with their interactions between wild type and mutant structure

Interactions	Wild type	Mutant (A237D)	Mutant (W353R)
Total number of contacts	[proximal+VdW clashinteraction] 71+3=74	[Vdw+VdW clash+proximal] 1+11+119=131	[Vdw+VdW clash+proximal] 4+7+119=130
Polar contacts	2	7	0
Weak polar contacts	2	4	4
Hydrogen bond	2	7	3
Ionic interaction	0	7	3
Carbonyl interaction	1	1	0
Hydrophobic contacts	3	3	0
Halogen bonds	0	0	4
Metal complex interaction	0	0	0
Hydrophobic contacts	3	3	2

and structural integrity. A decrease in hydrogen bonding can alter the shape of binding pockets or active sites, making it more difficult for a protein to interact with substrates, inhibitors, or other binding molecules (Chen et al., 2020; Khan et al., 2019; Shahid et al., 2015). This may result in decreased cellular activity overall, decreased binding affinity, and decreased enzyme activity in particular. The protein's internal integrity may be compromised by a significant decrease in these connections, exposing hydrophobic residues to the surrounding water. Because the hydrophobic residues try to minimise connection to the water-based environment, this contact could lead the protein to misfold. Proteins that are misfolded are more likely to aggregate, a process in which the exposed hydrophobic regions of many molecules of protein cling to one another to create insoluble clumps. Tryptophan, phenylalanine, and tyrosine are examples of aromatic residues that frequently take role in  $\pi$ - $\pi$  stacking connections, in which the aromatic rings correspond in a parallel or edge-to-face fashion.

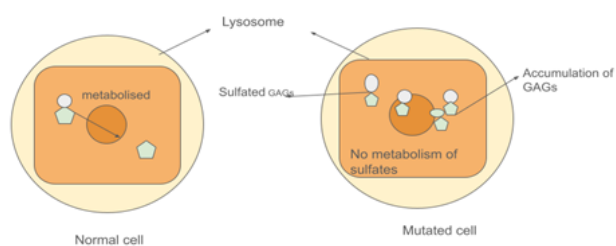
These mutations impair the stability of the gene overall, in addition to possibly causing illness by rupturing the structural integrity of highly conserved sections. After that, we evaluated the aggregation characteristics of these alterations using the SODA programme. It is quite probable that these two last mutations (A237D, W353R) will make the protein less soluble and contribute to the pathophysiology of the disease. This research provides a detailed account of the pathogenic nsSNPs in the ARSB gene and their possible effects. Determining whether particular mutations are harmful, unstable, and pathogenic provides important new insights into the molecular processes underlying the genesis of illness. The aggregation propensity study also identifies potential treatment targets by highlighting the significance of amino acid solubility in disease association.

To sum up, the A237D mutation results in a larger total number of connections, which causes structural strain, yet it also preserves overall integrity due to improved polar, hydrogen, and ionic interactions (Figure 6). On the other hand, the W353R mutation introduces notable changes that could disrupt local equilibrium by affecting aromatic interactions. Different structural results arise from the precise form and position of the alterations, even if both mutations bring compensatory interactions to preserve structural stability. While the W353R mutation has localised destabilising effects that affect the protein's functional regions, the A237D alteration better maintains its structural stability.

### 3.5. Aggregation Propensity

Treatment plans to mitigate the effects of such harmful mutations can be developed using the knowledge gained from this study (Israil et al., 2025; Nag and Tripathi, 2022; Paladin et al., 2017). Creating medications or tiny compounds that stabilise the peptide, improve its ability to dissolve, or stop it from aggregating are some possible strategies. Additionally, by understanding the structural alterations caused by these mutations, tailored therapies that restore lost or normal protein function may be developed.

In the final analysis, our thorough examination of nsSNPs in the ARSB gene not only advances our understanding of the genetic underpinnings of associated illnesses but also lays the groundwork for developing effective treatments. The integration of structural, solubility, and sequencing research offers a comprehensive knowledge of the effects of pathogenic mutations. Two of the discovered mutations are located in highly conserved regions that may affect the gene's function, according to our ConSurf analysis. Using the Arpeggio web server, we further evaluated the aggregation properties and conducted additional structural analyses (Figure 7). We discovered two changes that reduce protein solubility, suggesting a possible role in ARSB aggregation and the ensuing illness. The identification of mutations with the greatest potential to contribute to disease pathophysiology was made possible by sequence-based analysis. Simultaneously, the predictions based on structure reinforced our belief that mutations could be crucial to the onset and progression of the disease.

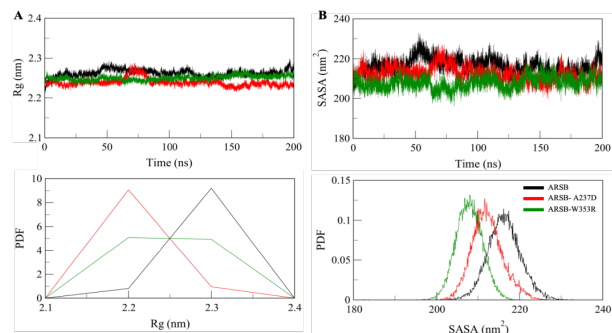


**Figure 7.** ARSB mutation can cause accumulation of sulfated GAG in the lysosome, which may result in a lysosomal disorder.

### 3.6. Molecular Dynamics Simulation Analysis

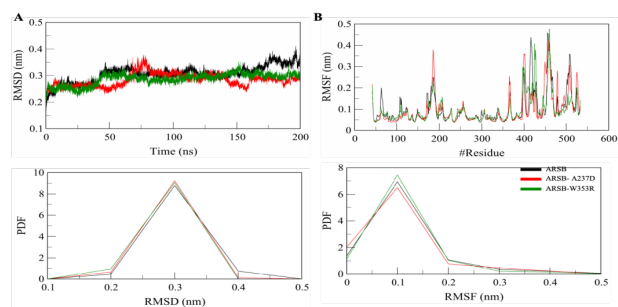
The wild-type ARSB protein, along with a few selected mutant variants, was subjected to molecular dynamics (MD) simulations

to confirm the structural consequences of high-confidence deleterious substitutions. MD simulations provide dynamic insights into the stability, flexibility, and conformational behavior of proteins under physiologically relevant conditions, in contrast to static protein structure investigations.



**Figure 8.** Radius of Gyration (Rg) analysis indicating changes in compactness of wild-type and mutant ARSB proteins over time.

RMSD, RMSF, Radius of Gyration (Rg), SASA, and hydrogen-bond evaluations were included in the trajectory evaluations (Naqvi et al., 2018). In contrast to the natural type, mutant proteins displayed delayed equilibration and higher RMSD values, suggesting decreased structural reliability. RMSF analysis revealed higher adaptability in conserved and functionally significant regions, suggesting a disturbance in catalytic activity (Figure 8). Many mutants had elevated Rg and SASA values, consistent with unfolding and increased aggregation (Figure 9). These values also indicated decreased compactness and higher solvent exposure.



**Figure 9.** RMSD plot showing overall structural stability and conformational deviations of wild-type and mutant ARSB proteins during the simulation.

#### 4. Conclusions

SNPs, or single-nucleotide polymorphisms, are thought to be among the most common genetic variations linked to several human disorders. 139 of the 429 mutations found are harmful and destabilising, according to sequence- and structure-based studies. A research investigation on pathogenicity found that 44 of the total number of variants are harmful. After consurf analysis of aggregation propensity, we found that 2 final mutations may be factors to causing disease. A thorough study of SNPs can provide insights into the mechanisms underlying disease development and inform the development of efficient therapeutic approaches. Our results indicate the importance of computational mutational analysis for understanding the genetic underpinnings

of complex diseases such as MPS VI (Karageorgos et al., 2007; Prokop et al., 2022). This study lays the foundation for future research to identify targeted therapy strategies and contributes to the expanding body of knowledge on the pathophysiology of MPS VI. In the end, the study highlights the importance to using cutting-edge computational methods to better understand the molecular pathways underlying disease pathology (Dissanayake et al., 2025).

#### Conflicts of Interest

The authors declare no conflict of interest.

#### Funding

This work received no funding.

#### Data availability statement

All data generated or analyzed during this study are included in this manuscript.

#### Declaration on the Use of AI Tools

The authors declare that ChatGPT (OpenAI) was used solely to refine the language, improve grammar, and enhance the clarity of the manuscript.

#### References

- De Ponti, G., Donsante, S., Frigeni, M., et al. 2022, *International Journal of Molecular Sciences*, 23, 11168
- Enni, M. A. 2025, *International Journal of Scientific Interdisciplinary Research*, 6, 88
- Gomez-Ospina, N. 2024, *Arylsulfatase A deficiency*
- Gros, F., & Muller, S. 2023, *Nature Reviews Nephrology*, 19, 366
- Leal, A. F., Prieto, L. E., Pachajoa, H., & Tomatsu, S. 2025, *Molecular Genetics and Metabolism*, 109255
- Lipiński, P., Rózdżyńska-Świątkowska, A., Wiśniewska, K., et al. 2025, *Biomolecules*, 15, 1448
- Platt, F. M., d'Azzo, A., Davidson, B. L., Neufeld, E. F., & Tiffit, C. J. 2018, *Nature Reviews Disease Primers*, 4, 27
- Rossi, A., Romano, R., Fecarotta, S., et al. 2025, *Med*, 6
- Sarachakov, A., Yudina, A., Svelkolkin, V., et al. 2025, *Human Genetics*, 144, 1245
- Sauna, Z. E., & Kimchi-Sarfaty, C. 2022, *Single nucleotide polymorphisms: human variation and a coming revolution in biology and medicine* (Springer)
- Scerra, G., De Pasquale, V., Scarcella, M., et al. 2022, *Open Biology*, 12
- Sinha, A., Dinakarkumar, Y., Al-Qahtani, W. H., et al. 2022, *Human Gene*, 34, 201079
- Tobacman, J. K., & Bhattacharyya, S. 2022, *International Journal of Molecular Sciences*, 23, 13146

**Table 3.** Disease phenotype evaluation of higher certainty nsSNPs in the ARSB gene applying PMut, MutPred, and SNAP & GO estimation methods.

Mutations	PhD-SNP	SNP & GO	MutPred2	Remark
A237D	Disease	Disease	0.9	Pathogenic
C91R	Disease	Disease	0.923	Pathogenic
E323K	Disease	Disease	0.943	Pathogenic
E483D	Disease	Disease	0.871	Pathogenic
G149R	Disease	Disease	0.947	Pathogenic
G302R	Disease	Disease	0.967	Pathogenic
G324V	Disease	Disease	0.946	Pathogenic
G527R	Disease	Neutral	0.931	Benign
G56D	Disease	Disease	0.964	Pathogenic
G64R	Disease	Disease	0.715	Pathogenic
H393P	Disease	Disease	0.889	Pathogenic
I296N	Disease	Disease	0.944	Pathogenic
I67N	Disease	Disease	0.818	Pathogenic
K145E	Disease	Disease	0.921	Pathogenic
L129P	Disease	Disease	0.937	Pathogenic
L132P	Disease	Neutral	0.735	Benign
L236P	Disease	Disease	0.925	Pathogenic
L360P	Disease	Disease	0.952	Pathogenic
L498P	Disease	Disease	0.909	Pathogenic
L51P	Disease	Disease	0.957	Pathogenic
L72P	Disease	Disease	0.939	Pathogenic
L72R	Disease	Disease	0.935	Pathogenic
L82P	Disease	Disease	0.938	Pathogenic
L82R	Disease	Disease	0.94	Pathogenic
L90P	Disease	Disease	0.836	Pathogenic
L98P	Disease	Disease	0.91	Pathogenic
L98R	Disease	Disease	0.901	Pathogenic
N84K	Neutral	Neutral	0.64	Benign
P248A	Neutral	Neutral	0.515	Benign
P531R	Neutral	Neutral	0.947	Benign
P93R	Disease	Disease	0.713	Pathogenic
R102H	Disease	Neutral	0.614	Benign
R315P	Disease	Disease	0.967	Pathogenic
R315Q	Disease	Disease	0.89	Pathogenic
R327G	Disease	Disease	0.96	Pathogenic
R327Q	Disease	Disease	0.902	Pathogenic
R388T	Disease	Disease	0.775	Pathogenic
R484G	Neutral	Neutral	0.611	Benign
R66C	Disease	Neutral	0.301	Benign
T92K	Disease	Disease	0.639	Pathogenic
V277G	Disease	Disease	0.882	Pathogenic
V332G	Disease	Neutral	0.934	Benign
V48A	Disease	Neutral	0.765	Benign
V80G	Disease	Disease	0.746	Pathogenic
W146R	Disease	Disease	0.967	Pathogenic
W146S	Disease	Disease	0.97	Pathogenic
W353R	Disease	Disease	0.914	Pathogenic
W438G	Disease	Disease	0.908	Pathogenic
W450C	Disease	Neutral	0.773	Benign
Y138C	Disease	Disease	0.946	Pathogenic
Y175D	Disease	Disease	0.957	Pathogenic
Y210C	Disease	Disease	0.837	Pathogenic
Y266S	Disease	Disease	0.861	Pathogenic
Y86C	Disease	Disease	0.922	Pathogenic
Y86N	Disease	Disease	0.93	Pathogenic