

Comparative Analysis and 3D Modeling of Missense Variants in the HEXA Gene

Simran Soni

simransoni0407@gmail.com

Homestead High School

Vera Khodzhaeva

University of Cambridge

Research Article

Keywords: Tay-Sachs, HEXA, mutation

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11505011/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Comparative Analysis and 3D Modeling of Missense Variants in the HEXA Gene

¹Simran Soni, ²Vera Khodzhaeva

¹Homestead High School, Cupertino, CA

²Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom

Abstract

Tay-Sachs disease is an autosomal recessive neurodegenerative disorder caused by mutations in the HEXA gene, which lead to a deficiency of the enzyme beta-hexosaminidase A and the toxic accumulation of GM2 gangliosides in the nervous system. Currently, many HEXA missense variants fall into categories of uncertain or conflicting clinical significance. This study aimed to evaluate these diagnostic discrepancies by comparing clinical classifications from ClinVar with automated machine-learning pathogenicity predictions from AlphaMissense across 30 HEXA missense variants. A targeted subset of seven variants, 5 with conflicting classifications (Trp474Leu, Ser486Asn, Thr467Ile, Pro471Arg, and Pro475Ser) and two reference variants (Arg247Trp and Gly478Arg), were selected for three-dimensional structural modeling and rotamer strain analysis in PyMOL. Structural analysis demonstrated that buried residues within the tightly packed hydrophobic core (Ser486Asn, Thr467Ile) produced consistently higher rotamer strain values and reduced flexibility compared to surface-exposed residues (Trp474Leu, Pro475Ser), which achieved low-strain conformations without steric clashes. A boundary residue (Pro471Arg) exhibited extreme structural tension due to increases in side-chain bulk and charge. These findings suggest that HEXA protein stability is highly sensitive to core disruptions and highlight how using 3D structural modeling can help resolve diagnostic discrepancies for Variants of Uncertain Significance, supporting ongoing research into therapeutic development for Tay-Sachs disease.

Introduction

Tay-Sachs disease is a neurodegenerative disorder caused by the deficiency of the enzyme hexosaminidase-A, which helps break down fatty substances¹. The accumulation of fatty substances leads to toxic levels of buildup in the brain and spinal cord, which can affect the function of nerve cells¹. This disease primarily affects the nervous system and results in neural dysfunction, which ultimately leads to neurodegeneration¹. It is an autosomal recessive disorder that belongs to a wider group of lysosomal storage disorders called GM2 gangliosidoses². Tay-Sachs disease is rare in the general population; the carrier frequency is about 1 in 250³, compared to about 1 in 100,000 who inherit the disease⁴.

Tay-Sachs disease is caused by a mutation in the *HEXA* gene, which encodes the enzyme beta-hexosaminidase A⁵. The *HEXA* gene is located on chromosome 15q23⁶. Hexosaminidase A is a lysosomal enzyme that removes terminal N-acetylgalactosamine residues from gangliosides like GM2, which is crucial in their breakdown⁷. Mutations in the *HEXA* gene lead to Tay-Sachs disease due to GM2 accumulation, whereas proper function requires the enzyme's alpha and beta subunits

and the GM2 activator protein ⁷. Hexosaminidase A breaks down GM2 gangliosides in neurons. Without it, GM2 accumulates inside lysosomes, disrupting normal cell function. This buildup eventually causes neuron death, leading to the neurological symptoms of Tay–Sachs disease ⁷.

Tay-Sachs is most severe in the infantile stage. Most symptoms appear within 3-6 months, including motor weakness, developmental regression, seizures, vision loss (cherry-red spot in retina), and death by 4-5 years old ^{8,9}. For juveniles, it appears between 2 and 10 years old. Symptoms progress more slowly with ataxia, spasticity, and eventually severe neurological deterioration. Adult-onset is much rarer, and it appears in adolescence or adulthood. Symptoms include psychiatric manifestations, cognitive decline, and motor difficulties¹⁰. Symptoms get worse over time, and it often leads to severe disability and early death in childhood ^{9,10}.

Tay-Sachs disease occurs in all populations but is more common in certain groups, such as Ashkenazi Jewish, French-Canadian, and Cajun communities¹¹. The disorder follows an autosomal recessive mode of inheritance, meaning that a child must inherit one defective gene from each parent to develop the condition. Risk factors include having both parents as carriers of *HEXA* mutations, which increases the chance that their child will be affected. Carrier screening and genetic counseling can help identify at-risk couples before having children¹¹.

Currently, there is no cure for Tay-Sachs disease, so treatment mainly focuses on supporting patients and managing symptoms. Doctors work to control seizures, reduce muscle stiffness, and help with feeding difficulties to make patients more comfortable. Researchers are also exploring new approaches, like enzyme replacement therapy¹², gene therapy¹³, and substrate reduction therapy ^{14,15}, which aim to restore enzyme function or reduce the harmful accumulation of GM2 gangliosides in the brain. These treatments are still experimental, but they represent promising future therapeutic approaches¹⁰.

Despite advances in supportive care and the development of experimental therapies, understanding the genetic basis of Tay-Sachs disease remains essential. More than 130 pathogenic variants have been identified in the *HEXA* gene, including substitutions, insertions, deletions, and duplications ⁴. However, some variants remain classified as variants of uncertain significance or have conflicting clinical interpretations. Accurate classification of these variants is important for genetic diagnosis, carrier screening, and genetic counselling, while also contributing to future research into targeted therapies. Computational prediction tools and structural analyses can provide additional evidence to assess the potential effects of missense variants and support their interpretation. Therefore, this study aims to compare computational predictions with clinical classifications and investigate the structural consequences of selected *HEXA* missense variants.

Methods

Genetic variants associated with Tay-Sachs disease in the *HEXA* gene were obtained from the ClinVar database, a publicly available repository of clinically reported genetic variants ¹⁶. Only missense variants were included, as they result in single amino acid substitutions that can affect protein structure and function. For each variant, the DNA base change, protein change, and clinical significance were recorded (Table 1).

The same variants were analyzed using AlphaMissense, a machine learning-based tool that predicts the pathogenicity of missense mutations based on sequence and structural features ¹⁷. Predictions from AlphaMissense were compared with ClinVar classifications to identify agreement and discrepancies (Table 1).

The three-dimensional structure of *HEXA* was obtained from the Protein Data Bank (2GK1) ¹⁸ and visualized using PyMOL ¹⁹. Each mutation was mapped onto the structure to determine its spatial location. Structural effects were further analyzed using the PyMOL Mutagenesis Wizard.

To evaluate the structural changes of each mutation in PyMOL, residues were first visually classified using the protein's three-dimensional (3D) molecular surface: they were considered surface-exposed if visible on the outside of the protein, and buried if completely hidden inside the tightly packed core. When a mutation was introduced, PyMOL generated a library of possible 3D side-chain shapes, called rotamers. Every available rotamer was evaluated using a numerical strain score, which represents how poorly the new amino acid fits into the surrounding space. The rotamer with the lowest strain score was recorded as the most likely stable shape. Higher strain scores indicated worse structural fits and greater tension, which were primarily caused by steric clashes. A steric clash was defined as the physical overlap of atoms, occurring when a newly introduced amino acid is too bulky for the available space and its atoms are forced into the same space as neighboring atoms.

Results

A total of 30 missense variants in the *HEXA* gene were analysed (Table 1). ClinVar classifications were compared with AlphaMissense pathogenicity predictions, and the variants were subsequently mapped onto the *HEXA* protein structure in PyMOL to assess their predicted structural consequences.

Of the 30 *HEXA* missense variants analyzed, 18 are classified as pathogenic or likely pathogenic in ClinVar, 9 are classified as variants of uncertain significance (VUS) or unclassified, 2 are classified as benign or likely benign, and one variant, p.Pro471Leu, is classified as likely pathogenic in ClinVar but holds an ambiguous status in AlphaMissense (Table 1). In the ClinVar

database, 'Uncertain Significance' indicates that a variant has been evaluated but the evidence remains inconclusive, whereas 'Unclassified' indicates that a variant was submitted without a formal clinical interpretation. When comparing databases, AlphaMissense predicted all 18 of the ClinVar pathogenic and likely pathogenic variants as likely pathogenic (Table 1). For the 9 uncertain or unclassified variants, AlphaMissense predicted all 9 to be likely pathogenic (Table 1). Of the two benign variants, AlphaMissense predicted one to be likely benign and one to be likely pathogenic (Table 1). Additionally, 11 of the 30 variants have no associated publications reported in the databases.

Table 1. Comparative analysis of HEXA gene missense variants and their pathogenicity using ClinVar and AlphaMissense databases

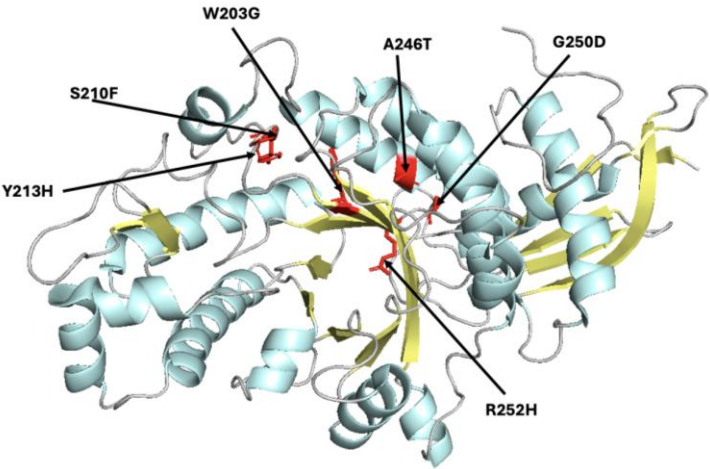
DNA Base Change	Protein Change	ClinVar	AlphaMissense	Publication
c.1511G>T	p.Arg504Leu	Pathogenic/likely pathogenic	Likely pathogenic	Gowda VK et al. (2020) ²⁰
c.607T>G	p.Trp203Gly	Pathogenic	Likely pathogenic	Montalvo AL et al. (2005) ²¹
c.629C>T	p.Ser210Phe	Pathogenic/likely pathogenic	Likely pathogenic	Akli S et al. (1991) ²²
c.637T>C	p.Tyr213His	Pathogenic	Likely pathogenic	Mistri et al. (2019) ²³
c.736G>A	p.Ala246Thr	Pathogenic/likely pathogenic	Likely pathogenic	Ou L et al. (2019) ²⁴
c.749G>A	p.Gly250Asp	Pathogenic/likely pathogenic	Likely pathogenic	Karpati M et al. (2000) ²⁵
c.755G>A	p.Arg252His	Pathogenic/likely pathogenic	Likely pathogenic	Jahnová H et al. (2019) ²⁶
c.964G>A	p.Asp322Asn	Pathogenic	Likely pathogenic	Sheth J et al. (2014) ²⁷
c.964G>T	p.Asp322Tyr	Pathogenic/likely pathogenic	Likely pathogenic	Gowda VK et al. (2020) ²⁰
c.965A>G	p.Asp322Gly	Pathogenic/likely	Likely pathogenic	Park JH et al.

		pathogenic		(2021) ²⁸
c.995A>T	p.Asn332Ile	Pathogenic	Likely pathogenic	Not reported
c.1121A>C	p.Gln374Pro	Pathogenic	Likely pathogenic	Sheth J et al. (2014) ²⁷
c.1178G>C	p.Arg393Pro	Pathogenic	Likely pathogenic	Not reported
c.1385A>T	p.Glu462Val	Pathogenic	Likely pathogenic	Mistri M et al. (2019) ²³
c.1422G>C	p.Trp474Cys	Pathogenic/likely pathogenic	Likely pathogenic	Clarke JT et al. (2011) ²⁹
c.1432G>A	p.Gly478Arg	Pathogenic/likely pathogenic	Likely pathogenic	Mistri M et al. (2019) ²³
c.1457G>A	p.Ser486Asn	Unclassified	Likely pathogenic	Not reported
c.1377G>A	p.Met459Ile	Uncertain	Likely pathogenic	Not reported
c.1395C>G	p.Asp465Glu	Likely pathogenic	Likely pathogenic	Alvarez-Rodríguez A et al. (2001) ³⁰
c.1400C>T	p.Thr467Ile	Uncertain	Likely pathogenic	Not reported
c.739C>T	p.Arg247Trp	Benign/likely benign	Likely benign	Not reported
c.1400C>A	p.Thr467Lys	Uncertain	Likely pathogenic	Not reported
c.1412C>G	p.Pro471Arg	Uncertain	Likely pathogenic	Not reported
c.1412C>T	p.Pro471Leu	Likely pathogenic	Ambiguous	Zeng BJ et al. (2008) ³¹
c.1421G>T	p.Trp474Leu	Uncertain	Likely pathogenic	Buratti E et al. (2007) ³²
c.1423C>T	p.Pro475Ser	Uncertain	Likely pathogenic	Not reported
c.1424C>G	p.Pro475Arg	Uncertain	Likely pathogenic	Mistri M et al.

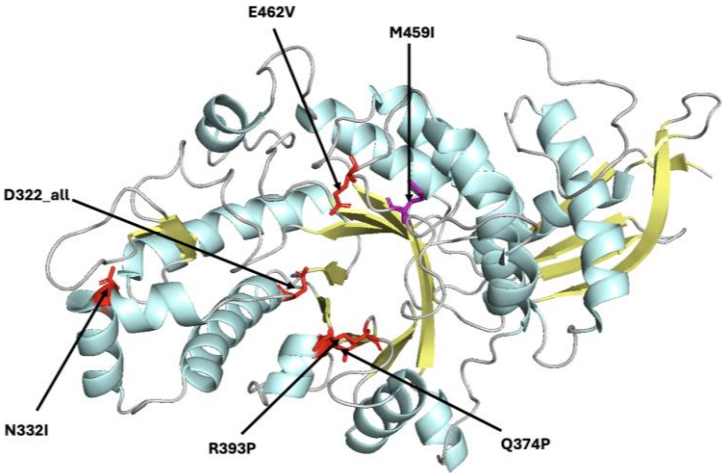
				(2019) ²³
c.1427G>A	p.Arg476Lys	Uncertain	Likely pathogenic	Not reported
c.539A>C	p.Tyr180Ser	Likely benign	Likely pathogenic	Not reported
c.616G>C	p.Val206Leu	Likely pathogenic	Likely pathogenic	Not reported

Note: variants highlighted in blue color were selected for analysis.

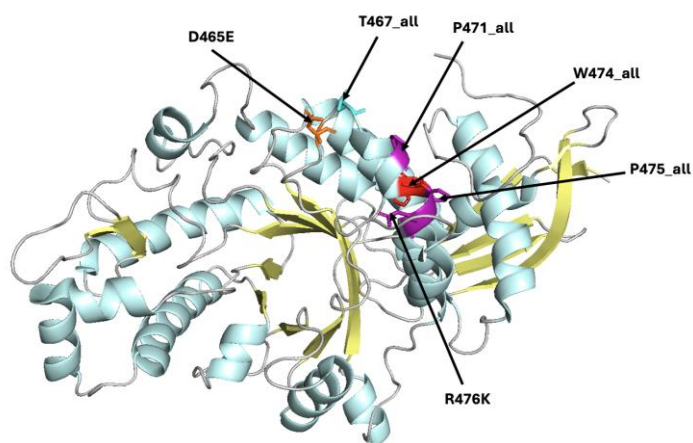
A



B



C



D

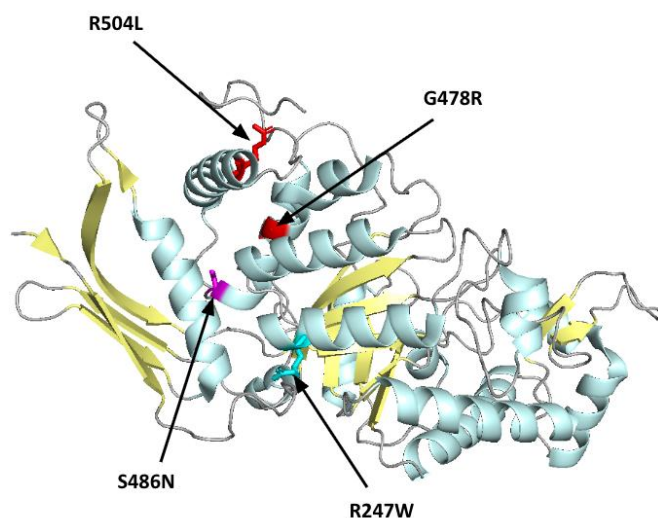


Figure 1. Structural mapping of HEXA variants.

Cartoon representation of the HEXA protein (PDB ID 2GK1), visualized in PyMOL. All missense variants introduced in Table 1 are mapped onto the protein structure to show their location. HEXA variants are color-coded by pathogenicity, integrating ClinVar, AlphaMissense, and structural analysis: red = pathogenic, orange = likely pathogenic, purple = conflicting classification, and cyan = benign. A total of 30 variants were mapped (A-D). Amino acid substitutions in the figure are labeled using standard one-letter codes (e.g., W203G), which correspond to the three-letter codes used in Table 1 (e.g., p.Trp203Gly). Labels ending in "_all" indicate positions where multiple different amino acid substitutions occur at the same site, specifically: Asp322_all, Thr467_all, Pro471_all, Trp474_all, and Pro475_all.

To establish structural baselines, two reference variants with known outcomes—Arg247Trp as a benign reference and Gly478Arg as a pathogenic reference—were first modeled in PyMOL (Figures 2–3). Next, from the 9 uncertain or unclassified HEXA missense variants in Table 1, five were selected for 3D structural modeling and rotamer strain analysis: Trp474Leu, Ser486Asn, Thr467Ile, Pro471Arg, and Pro475Ser (Figures 4–8). These five variants were chosen because their ClinVar clinical reports and AlphaMissense predictions did not match, and they are all located in different parts of the protein. This included residues buried deep inside the core (Ser486Asn, Thr467Ile), residues sitting on the outside surface (Trp474Leu, Pro475Ser), and a boundary residue sitting right in between (Pro471Arg). We also focused on variants that were clustered together in one physical area of the protein (residues 467–486) and made sure not to test the exact same location twice (for example, selecting Pro475Ser instead of testing both Pro475Ser and Pro475Arg). By choosing these specific five, we could carefully test whether 3D modeling can help clear up confusing diagnostic reports across a variety of different physical environments. All five variants are classified as likely pathogenic by AlphaMissense and uncertain or unclassified by ClinVar (Table 1).

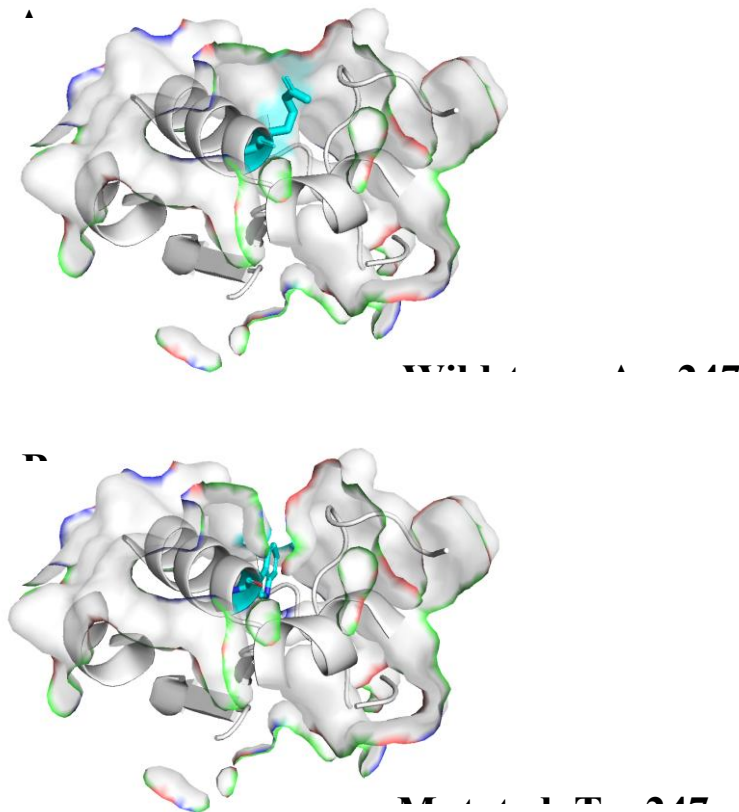


Figure 2. Arg247Trp Variant

Zoomed-in view of the **Arg247Trp mutation of HEXA visualised in PyMOL**. (A) The wild-type Arg247 residue, shown in cyan. (B) The mutated Trp247 residue, shown in cyan.

The benign reference variant Arg247Trp is classified as benign in both databases ClinVar and AlphaMissense (Table 1). Residue 247 is located on the surface of the HEXA protein (Figure 2). The mutation substitutes arginine, a positively charged amino acid, with tryptophan, a bulky nonpolar amino acid. Mutagenesis analysis in PyMOL generated a lowest-strain rotamer of 24.70 and a highest of 87.30.

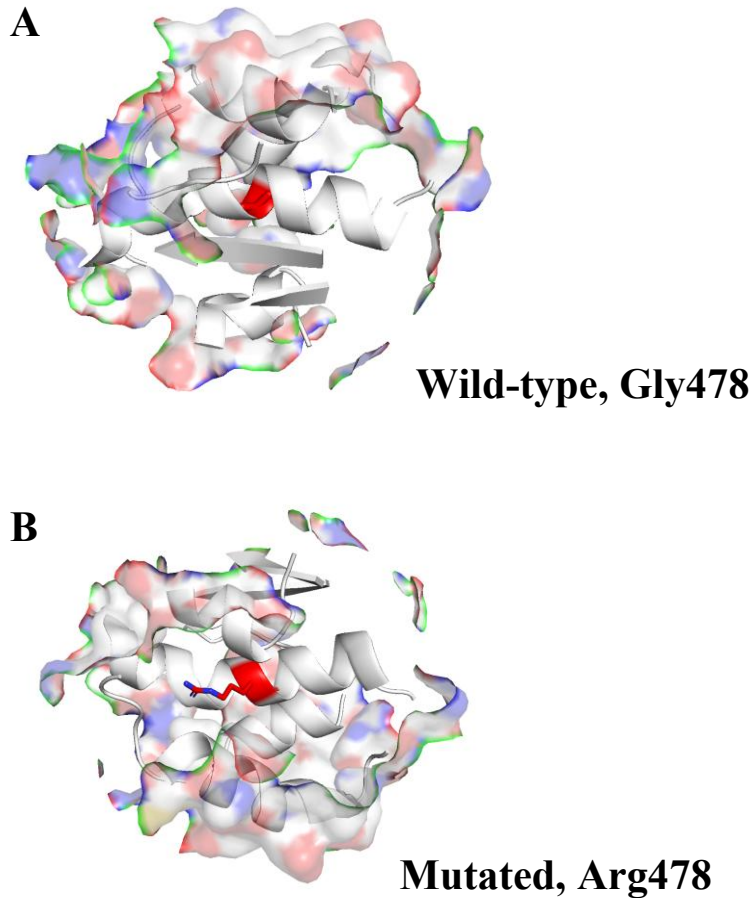


Figure 3. Gly478Arg Variant

Zoomed-in view of the **Gly478Arg mutation of HEXA visualised in PyMOL**. (A) The wild-type glycine (Gly478) residue, shown in red. (B) The mutated arginine (Arg478) residue, shown in red and blue.

The pathogenic variant Gly478Arg, a buried residue, was chosen as a reference for pathogenic classification (Figure 3). It is described as pathogenic or likely pathogenic in both ClinVar and AlphaMissense databases (Table 1). The mutation results in a substitution of glycine, the smallest amino acid with no side chain, with arginine, a much larger, positively charged amino acid. Furthermore, residue 478 is located within an alpha helix; substituting a flexible

glycine with a highly bulky arginine structurally disrupts this helix²³. This substitution significantly increases the steric bulk and introduces a new positive charge into the local environment. Mutagenesis analysis in PyMOL generated consistently high strain values across 29 rotamers, demonstrating severe structural tension. The lowest-strain rotamer had a strain value of 23.89, while the highest-strain rotamer had a value of 50.64. These elevated strain values suggest the structural disruption caused by introducing a bulky residue into this position.

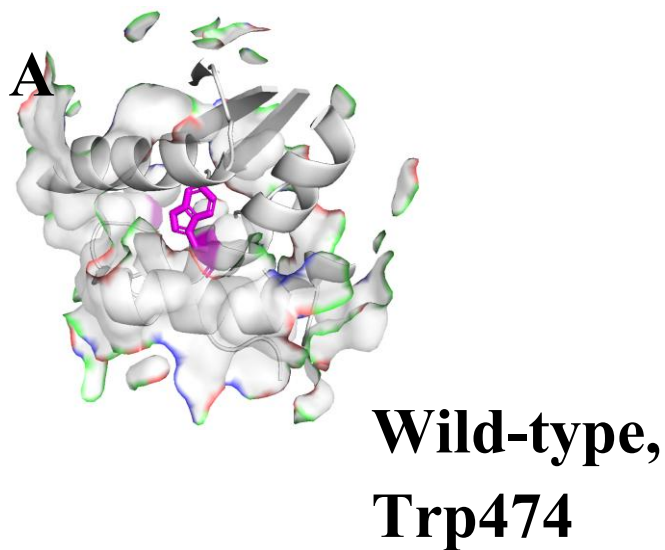
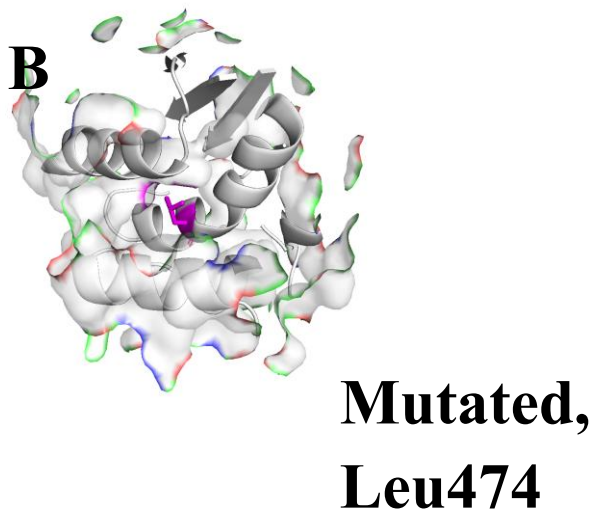


Figure 4. Trp474Leu Variant

Zoomed-in view of the Trp474Leu **variant of HEXA visualised in PyMOL, shown in purple.** (A) The wild-type tryptophan (Trp474) residue, shown in purple. (B) The mutated leucine (Leu474) residue, shown in purple.

Next, we analysed our selected five variants. Trp474Leu is classified as uncertain in ClinVar and likely pathogenic in AlphaMissense (Table 1). Located on the surface of the HEXA protein, residue 474 was fully visible when the molecular surface was displayed, indicating that it is not buried within the protein core (Figure 4). The Trp474Leu variant causes the substitution of tryptophan, a large aromatic hydrophobic amino acid, with leucine, a smaller hydrophobic amino acid. This substitution decreases the size of the side chain and removes the aromatic ring that was present in tryptophan, which then reduces steric bulk. Mutagenesis analysis in PyMOL revealed the lowest-strain rotamer had a strain value of 5.28, while the highest had a strain value of 35.23. Specifically, three rotamers showed low strain values of 5.28, 7.36, and 5.67, while one rotamer exhibited a higher strain value of 35.23. Because residue 474 sits on the flexible protein surface, no steric clashes were observed.

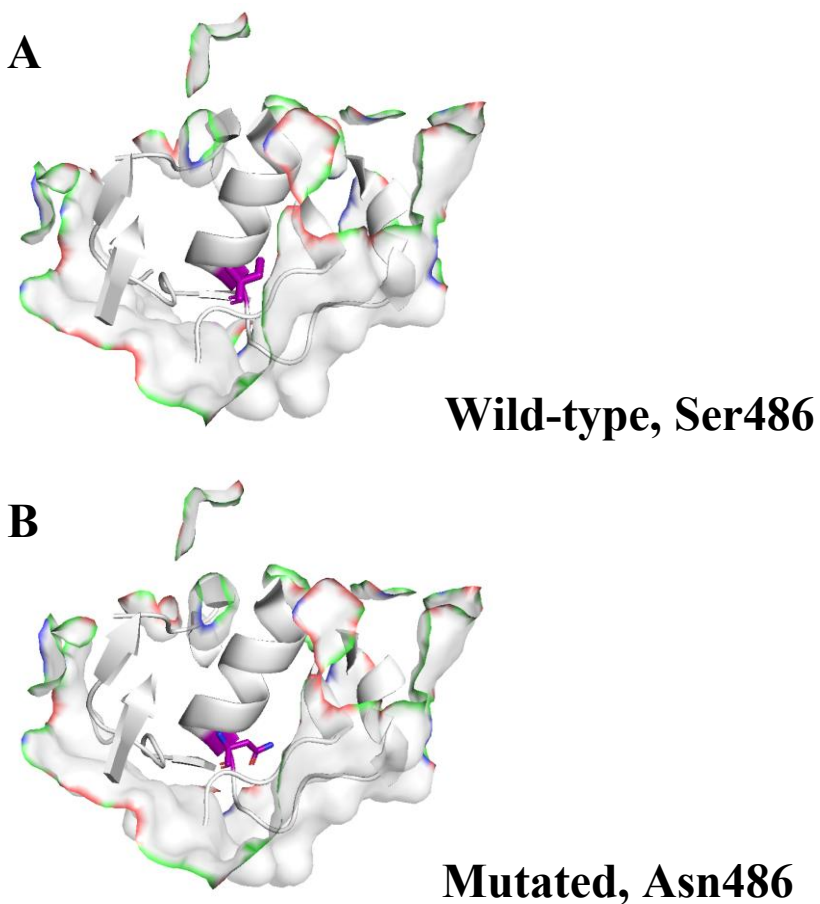


Figure 5. Ser486Asn Variant

Zoomed-in view of the **Ser486Asn mutation of HEXA visualised in PyMOL, shown in purple.** (A) The wild-type serine (Ser486) residue, shown in purple. (B) The mutated asparagine (Asn486) residue, shown in purple and blue.

The Ser486Asn variant is unclassified by ClinVar, while AlphaMissense predicts it as likely pathogenic. This residue is located within the interior and is completely buried within the protein core (Figure 5). The mutation results in a substitution of serine, a small polar amino acid containing a hydroxyl group, with asparagine, a larger polar amino acid containing an amide group. This substitution increases the size of the side chain and replaces the hydroxyl group with an amide functional group. Mutagenesis analysis showed the lowest-strain rotamer had a value of 26.58, while the highest-strain rotamer had a value of 39.63. Although no steric clashes were observed for the lowest-strain rotamer, the elevated strain reflects the tension introduced into the tightly packed core environment.

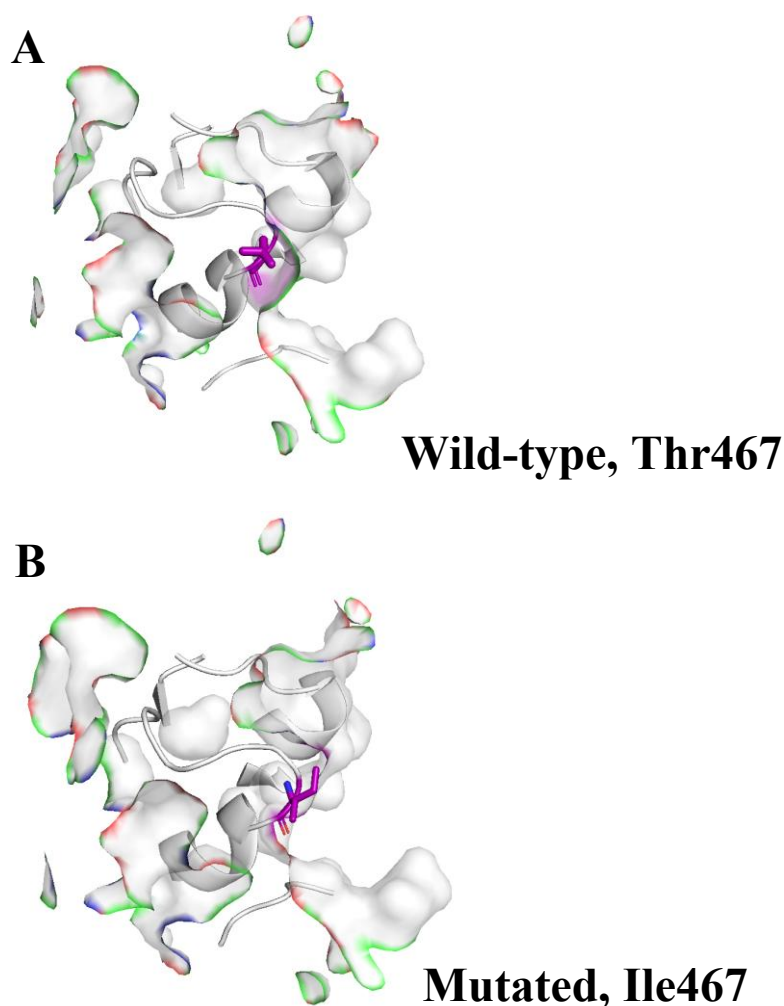
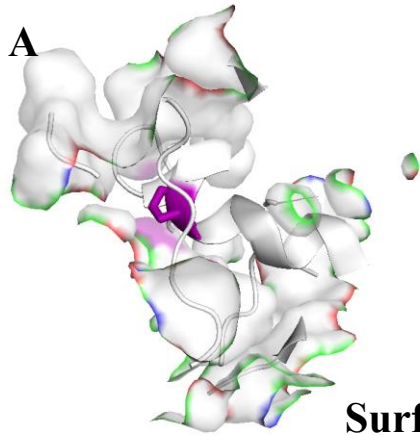


Figure 6. Thr467Ile Variant

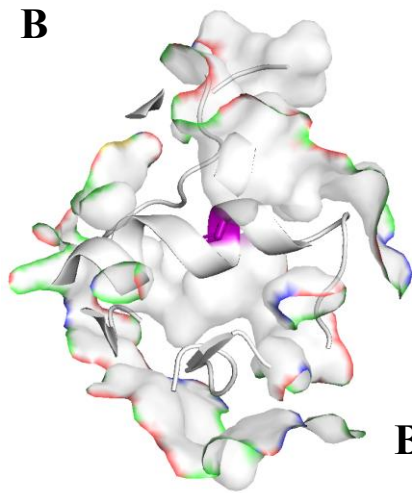
Zoomed-in view of the **Thr467Ile mutation of HEXA visualised in PyMOL**. (A) The wild-type threonine (Thr467) residue, shown in purple. (B) The mutated isoleucine (Ile467) residue, shown in purple and blue.

Mutagenesis analysis of the Thr467Ile mutation in PyMOL generated strain values of 15.95, 20.73, and 25.33. All rotamers showed moderate strain values, which likely reflects the tightly packed

environment of the protein core. Residue 467 is buried (Figure 6). The mutation results in a substitution of threonine, a small polar amino acid containing a hydroxyl group, with isoleucine, a larger hydrophobic amino acid. This substitution increases the size of the side chain and removes the hydroxyl group. While no major steric clashes were present, the increased strain scores directly increased structural tension.



**Surface exposed wild-type,
Pro471,**



Buried wild-type, Pro471

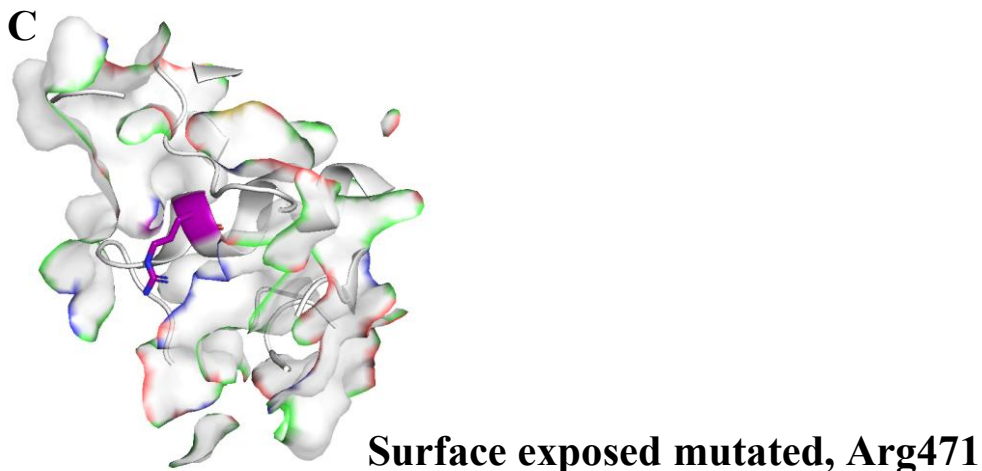
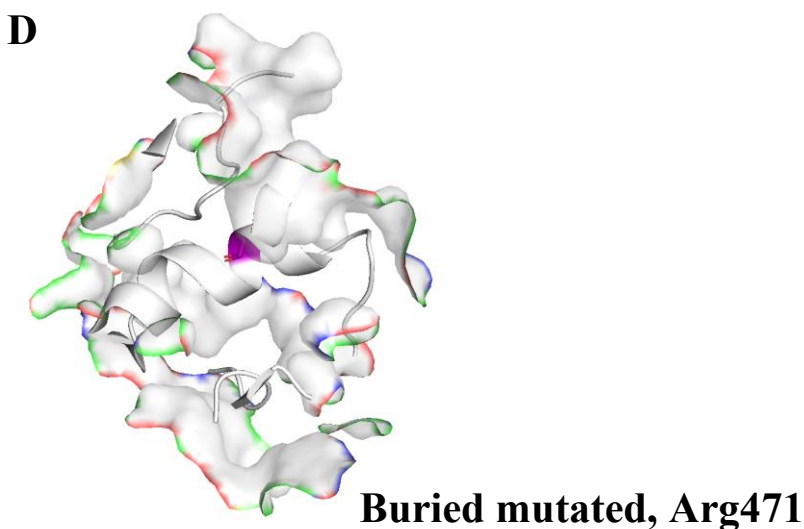


Figure 7. Pro471Arg Variant

Zoomed-in view of the **Pro471Arg mutation of HEXA visualised in PyMOL**. (A) The wild-type residue in a surface-exposed microenvironment, shown in purple. (B) The wild-type residue in a buried microenvironment, shown in purple. (C) The mutated residue in the surface-exposed conformation, shown in purple and blue. (D) The mutated residue in the buried conformation, shown in purple and red.

Sitting uniquely at the boundary between the protein surface and the interior of the HEXA protein, residue 471 is structurally constrained. When the molecular surface was displayed in PyMOL, part of the side chain of residue 471 was visible while another portion was buried within the protein structure (Figure 7). The Pro471Arg mutation results in a substitution of proline, a nonpolar amino acid with a cyclic side chain, with arginine, a larger positively charged amino acid with a longer side chain. This substitution drastically increases the size of the side chain and creates a new positive charge. Consequently, mutagenesis analysis generated strain values ranging from 35.13 to 94.70. No major steric clashes were observed for the lowest-strain rotamer.

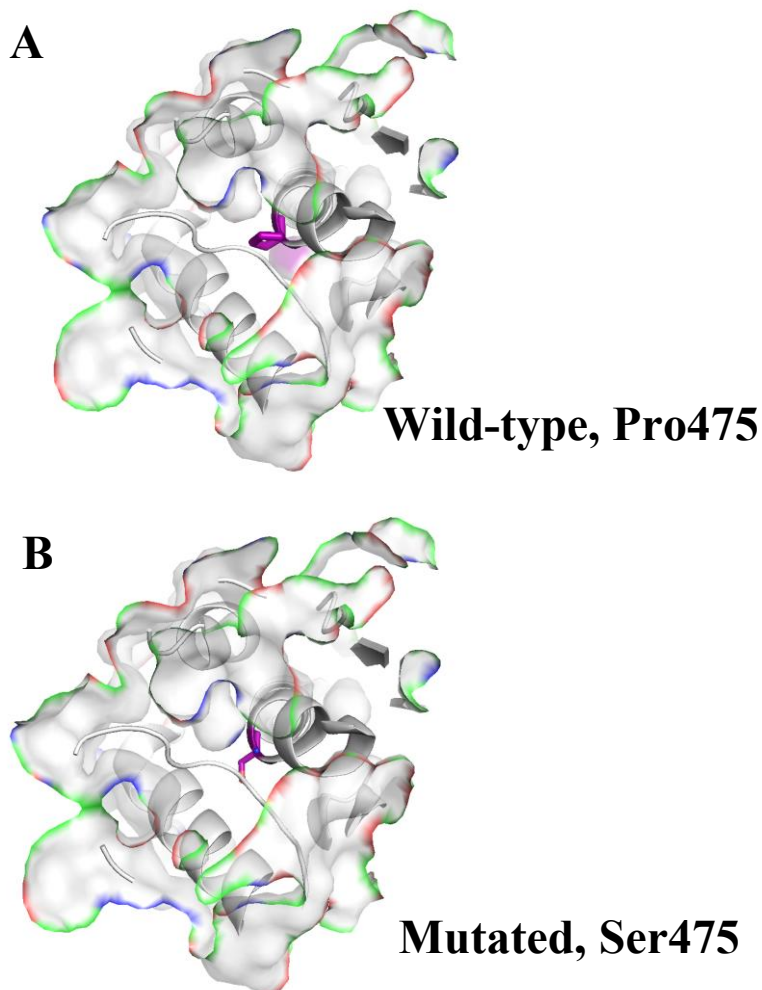


Figure 8. Pro475Ser Variant

Zoomed-in view of the **Pro475Ser mutation of HEXA visualised in PyMOL**. A) The wild-type threonine (Thr467) residue, shown in purple. (B) The mutated isoleucine (Ile467) residue, shown in purple and blue.

Finally, the Pro475Ser variant causes a substitution of proline, a rigid cyclic nonpolar amino acid, with serine, a smaller polar amino acid containing a hydroxyl group. This substitution reduces steric bulk and shifts the residue's chemistry to polar. Residue 475 is located on the surface of the HEXA protein (Figure 8). Mutagenesis analysis in PyMOL generated clustered strain values of 40.65, 40.96, and 41.16. No major steric clashes were observed for this lowest-strain rotamer. All data obtained after analysis were summarised in Table 2.

Table 2. Rotamer strain values for modeled HEXA variants.

Variant	Residue Location	Amino Acid Change	Lowest Strain	Highest Strain	Clashes

Arg247Trp	Surface	Charged → Nonpolar	24.70	87.30	None
Gly478Arg	Buried	Nonpolar → Charged	23.89	50.64	Yes
Trp474Leu	Surface	Nonpolar → Nonpolar	5.28	35.23	None
Ser486Asn	Buried	Polar → Polar	26.58	39.63	None
Thr467Ile	Buried	Polar → Nonpolar	15.95	25.33	None
Pro471Arg	Boundary	Nonpolar → Charged	35.13	94.70	None
Pro475Ser	Surface	Nonpolar → Polar	40.65	41.16	None

Note: Variant names are color-coded based on pathogenicity, integrating ClinVar and AlphaMissense data: Red = Pathogenic; Light yellow = Conflicting classification; Cyan = Benign.

3D modeling of the five selected variants showed that buried core mutations and boundary mutations generally produced higher structural strain than surface-exposed mutations. While strain values varied based on residue location, no major steric clashes were observed in the lowest-strain conformations for any of the variants. Overall, the analyzed HEXA missense variants showed variation in both predicted pathogenicity and structural effects depending on residue location and substitution type.

Discussion

Out of the 30 HEXA missense variants evaluated in this study, ClinVar and AlphaMissense demonstrated general agreement on 63.3% (19 out of 30) of the variants, particularly aligning on established pathogenic and benign classifications. However, notable disagreements arose on the remaining 36.7% (11 out of 30) of variants, primarily driven by the subset of variants classified by ClinVar as uncertain of significance or unclassified, where AlphaMissense consistently predicted them as likely pathogenic. This proportion of disagreement highlights a key limitation in clinical databases: many rare or newly discovered variants lack sufficient clinical or functional evidence, leaving them as variants of uncertain significance (VUS). While computational predictors like AlphaMissense can flag these variants as potentially damaging based on structural and evolutionary patterns, the discrepancy highlights the limitations of relying on computational predictions alone, particularly in the absence of experimental validation.

This study compared the classifications of HEXA variants in ClinVar and AlphaMissense and used structural modeling to investigate the potential effects of selected missense variants. The comparison between ClinVar classifications and AlphaMissense predictions identified several discordant HEXA variants. Of the 9 variants classified by ClinVar as uncertain or unclassified, AlphaMissense predicted all 9 to be likely pathogenic. This suggests that computational

predictions may help prioritise variants for further investigation, particularly when clinical or functional evidence is limited. However, computational predictions alone cannot establish pathogenicity. Three-dimensional structural modeling can provide additional mechanistic insight into the potential effects of variants and may therefore help inform the selection of candidates for further experimental investigation.

The structural analysis of the selected HEXA gene variants suggests that both the location of a mutation within a protein and the chemical properties of substituted amino acids may influence their stability and their potential disease-causing effects in Tay-Sachs disease. Rather than showing a simple pattern based on strain values alone, the data show different disruptions based on the protein microenvironment. Because the five structurally modeled variants were deliberately selected to represent surface, boundary, and buried environments, the structural findings should be considered hypothesis-generating rather than representative of the broader HEXA missense variant landscape.

The benign reference variant Arg247Trp showed a relatively high minimum strain of 24.70. While this indicates local structural tension, it likely reflects the rigid-body limitations of the PyMOL mutagenesis tool, which cannot model how surrounding surface residues naturally flex to accommodate a bulky amino acid. Furthermore, Arg247Trp is clinically known as a pseudodeficiency allele³³. Pseudodeficiency is a phenomenon where specific HEXA variants cannot efficiently process the synthetic substrates used in laboratory tests, leading to false-positive deficiency results³³. However, these variants do not reduce the enzyme's ability to break down natural GM2 gangliosides in the human body. Because the enzyme still functions normally *in vivo*, individuals with this mutation remain healthy and do not develop Tay-Sachs disease³⁴. Thus, Arg247Trp provides an important example in which relatively high predicted structural strain does not correspond to loss of physiological enzyme function or clinical disease.

Across the analyzed variants, Trp474Leu, Ser486Asn, Thr467Ile, Pro471Arg, and Pro475Ser, differences in residue exposure and the physicochemical properties of the substitutions were associated with varying levels of predicted structural strain in PyMOL. For example, substituting a small polar serine with a bulkier asparagine within the tightly packed protein core (Ser486Asn) resulted in relatively high-strain values (Table 2), whereas substituting a surface-exposed tryptophan with a smaller leucine (Trp474Leu) resulted in low-strain conformations (Table 2). This suggests that both the local environment and the properties of the amino-acid substitution play important roles in determining structural compatibility.

Among the variants analysed, surface-exposed residues appeared to influence the predicted structural compatibility of some substitutions. While both the Trp474Leu and Pro475Ser mutations fit without major clashes, they were not exactly the same. Trp474Leu was highly flexible and had a very low strain score (5.28). Pro475Ser was a bit more restricted, with a moderate strain score

(40.65) (Table 2). This happens because proline has a rigid, ring-like structure that locks the protein backbone in place. Even though both residues are on the surface, position 475 has tighter local backbone constraints than position 474. Therefore, surface exposure alone doesn't guarantee low strain, as local backbone geometry and rigidity still create structural tension. Despite this difference, both show that mutations on the surface are generally well tolerated. However, even when steric clashes are minimal, substitutions at surface residues may still alter protein behavior by affecting interactions with other proteins, membranes, or ligands. This is particularly relevant for HEXA, which functions in the lysosome, where protein interactions and enzyme accessibility are critical for GM2 ganglioside breakdown⁷.

In contrast to the surface variants, the buried variants (Ser486Asn and Thr467Ile) occupy a much more constrained environment. Although direct comparisons of raw strain values across different positions should be interpreted with caution, both variants exhibit elevated strain levels comparable to the reference pathogenic variant, Gly478Arg. Just as introducing a bulky arginine at residue 478 generated severe structural tension, substituting a smaller residue with bulky asparagine inside the core (Ser486Asn) results in similarly high strain. The tension observed in the Thr467Ile variant reflects the increased size of isoleucine, which leaves less available space within the tightly packed core. These structural observations provide additional support for the AlphaMissense predictions that Ser486Asn and Thr467Ile may be likely pathogenic variants. However, the modeling does not directly demonstrate protein misfolding, reduced stability, or loss of enzyme activity. Experimental studies would be needed to determine whether the predicted structural constraints result in functional changes.

The Pro471Arg variant at the boundary region showed a wide range of rotamer strain values, with scores ranging from 35.13 to 94.70. Such a wide range reflects the structural constraints of partially exposed residues positioned at the interface between the core and the surface environment, where intermediate microenvironments often experience high local sensitivity to side-chain alterations¹⁸. For example, studies on proteins like FUS show that boundary residues sitting between structured and unstructured regions act as critical anchors that help hold RNA in place³⁵. The substitution of proline with the larger, positively charged arginine side chain introduces both increased steric bulk and a change in charge, which may make some rotamer orientations less favourable within the surrounding structure. The results suggest that mutations at boundary regions can also experience substantial structural constraints, even when part of the residue is exposed to the protein surface.

In the selected variants, several buried and boundary substitutions produced rotamers with relatively high local strain scores, whereas Trp474Leu which is located on the protein surface included a particularly low-strain rotamer. This pattern is consistent with the idea that substitutions within the tightly packed protein core may have greater potential to disrupt local structural interactions, as there is less space to accommodate changes in amino acid size or chemical properties. Therefore, even small disruptions can affect proper folding and function³⁶. These

observations support the hypothesis that mutations affecting buried residues may be more likely to disrupt the local protein structure and potentially contribute to loss of function. However, these scores represent local steric compatibility and should not be interpreted as direct measures of protein stability or pathogenicity. Moreover, the structural modeling does not directly measure protein stability, folding, or enzyme activity, so further experimental studies would be needed to determine whether these predicted structural effects have a functional impact on HEXA.

These findings highlight the potential value of integrating structural modeling with genetic variant classification. Computational analysis can provide additional information about how specific amino-acid substitutions may affect the protein structure. Together, these approaches can give useful supporting evidence for interpreting variants when clinical or functional information is limited.

While 3D modeling provides great insights, it is important not to overstate what these static models can prove. This study has several limitations that should be considered when interpreting the results. First, the structural analysis was performed using a single protein structure from the Protein Data Bank, which does not capture the flexibility of the HEXA protein in physiological conditions. Proteins naturally undergo conformational changes, especially in lysosomal environments, which may influence the actual impact of the mutations. Second, PyMOL rotamer and strain calculations provide estimates of steric compatibility but do not directly measure functional enzyme activity or protein stability *in vivo*. Third, pathogenicity predictions from AlphaMissense and ClinVar classifications may not always agree, and both are based on different methodologies. Only missense mutations were analyzed in this study, meaning that other relevant mutation types, such as insertions or deletions, were not included. Finally, structural analysis was performed on only five variants and therefore may not represent all HEXA missense mutations.

Future studies could expand this analysis by incorporating molecular dynamics simulations to better understand how HEXA mutations affect protein flexibility over time in more realistic environments. Additionally, experimental validation using *in vitro* expression systems could be used to measure enzyme activity levels and confirm whether predicted structural disruptions correspond to functional loss^{18,36}. Expanding the dataset to include a bigger number of variants and adding different mutation types would also provide a better understanding of HEXA function.

Understanding the structural effects of HEXA mutations is directly relevant to the development of emerging therapeutic approaches for Tay-Sachs disease. Gene therapy strategies aim to restore enzyme activity and prevent GM2 ganglioside accumulation¹³. However, identifying which mutations lead to severe structural instability can help prioritize which variants warrant further experimental validation. Additionally, structural modeling may contribute to future studies that investigate variant effects and therapeutic development.

Overall, combining computational pathogenicity prediction with structural modeling provides complementary information that may improve the interpretation of HEXA missense variants and help prioritize variants for future experimental investigation.

Conclusion

This study indicates that combining ClinVar classifications, AlphaMissense predictions, and 3D structural modeling can provide additional insight into *HEXA* missense variants. In the variants analyzed, buried and partially buried residues generally showed greater predicted rotamer strain than some surface-exposed residues. However, structural modeling alone cannot confirm pathogenicity or directly measure protein stability or enzyme activity. Overall, these findings suggest that 3D modeling can provide useful supporting evidence for interpreting uncertain *HEXA* variants and identifying variants that may benefit from further experimental investigation.

References

1. Ferreira, C. R. & Gahl, W. A. Lysosomal storage diseases. *Transl. Sci. Rare Dis.* **2**, 1–71 (2017).
2. Gangliosidoses. in *Handbook of Clinical Neurology* vol. 113 1707–1708 (Elsevier, 2013).
3. Lew, R., Burnett, L., Delatycki, M. & Proos, A. Tay-Sachs disease: current perspectives from Australia. *Appl. Clin. Genet.* **19** (2015) doi:10.2147/TACG.S49628.
4. Myerowitz, R. Tay-Sachs disease-causing mutations and neutral polymorphisms in the Hex A gene. *Hum. Mutat.* **9**, 195–208 (1997).
5. Dersh, D., Iwamoto, Y. & Argon, Y. Tay–Sachs disease mutations in HEXA target the α chain of hexosaminidase A to endoplasmic reticulum–associated degradation. *Mol. Biol. Cell* **27**, 3813–3827 (2016).
6. Mistri, M. *et al.* Identification of Novel Mutations in HEXA Gene in Children Affected with Tay Sachs Disease from India. *PLoS ONE* **7**, e39122 (2012).
7. Solovyeva, V. V. *et al.* New Approaches to Tay-Sachs Disease Therapy. *Front. Physiol.* **9**, 1663 (2018).
8. Bley, A. E. *et al.* Natural History of Infantile GM2 Gangliosidosis. *Pediatrics* **128**, e1233–e1241 (2011).

9. Jarnes Utz, J. R. *et al.* Infantile gangliosidoses: Mapping a timeline of clinical changes. *Mol. Genet. Metab.* **121**, 170–179 (2017).
10. Maegawa, G. H. B. *et al.* The Natural History of Juvenile or Subacute GM2 Gangliosidosis: 21 New Cases and Literature Review of 134 Previously Reported. *Pediatrics* **118**, e1550–e1562 (2006).
11. Leal, A. F. *et al.* GM2 Gangliosidoses: Clinical Features, Pathophysiological Aspects, and Current Therapies. *Int. J. Mol. Sci.* **21**, 6213 (2020).
12. Tropak, M. B. *et al.* Construction of a hybrid β -hexosaminidase subunit capable of forming stable homodimers that hydrolyze GM2 ganglioside in vivo. *Mol. Ther. - Methods Clin. Dev.* **3**, 15057 (2016).
13. Cachón-González, M. B. *et al.* Effective gene therapy in an authentic model of Tay-Sachs-related diseases. *Proc. Natl. Acad. Sci.* **103**, 10373–10378 (2006).
14. Bembi, B. *et al.* Substrate reduction therapy in the infantile form of Tay-Sachs disease. *Neurology* **66**, 278–280 (2006).
15. Platt, F. M. *et al.* Prevention of Lysosomal Storage in Tay-Sachs Mice Treated with *N*-Butyldeoxynojirimycin. *Science* **276**, 428–431 (1997).
16. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
17. Cheng, J. *et al.* Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).
18. Lemieux, M. J. *et al.* Crystallographic Structure of Human β -Hexosaminidase A: Interpretation of Tay-Sachs Mutations and Loss of GM2 Ganglioside Hydrolysis. *J. Mol. Biol.* **359**, 913–929 (2006).

19. Hale, R. C., Seeley, M. E., La Guardia, M. J., Mai, L. & Zeng, E. Y. A Global Perspective on Microplastics. *J. Geophys. Res. Oceans* **125**, e2018JC014719 (2020).
20. Gowda, V. K. *et al.* Clinical and Laboratory Profile of Gangliosidosis from Southern Part of India. *J. Pediatr. Genet.* **11**, 034–041 (2022).
21. Montalvo, A. L. E. *et al.* Molecular analysis of the *HEXA* gene in Italian patients with infantile and late Onset Tay-Sachs disease: detection of fourteen novel alleles. *Hum. Mutat.* **26**, 282–282 (2005).
22. Akli, S., Chelly, J., Lacorte, J.-M., Poenaru, L. & Kahn, A. Seven novel Tay-Sachs mutations detected by chemical mismatch cleavage of PCR-amplified cDNA fragments. *Genomics* **11**, 124–134 (1991).
23. Mistri, M. *et al.* Identification of novel variants in a large cohort of children with Tay–Sachs disease: An initiative of a multicentric task force on lysosomal storage disorders by Government of India. *J. Hum. Genet.* **64**, 985–994 (2019).
24. Ou, L., Kim, S., Whitley, C. B. & Jarnes-Utz, J. R. Genotype-phenotype correlation of gangliosidosis mutations using in silico tools and homology modeling. *Mol. Genet. Metab. Rep.* **20**, 100495 (2019).
25. Karpati, M., Peleg, L., Gazit, E., Akstein, E. & Goldman, B. A novel mutation in the *HEXA* gene specific to Tay–Sachs disease carriers of Jewish Iraqi origin. *Clin. Genet.* **57**, 398–400 (2000).
26. Jahnová, H. *et al.* Amyotrophy, cerebellar impairment and psychiatric disease are the main symptoms in a cohort of 14 Czech patients with the late-onset form of Tay–Sachs disease. *J. Neurol.* **266**, 1953–1959 (2019).

27. Sheth, J. *et al.* Burden of Lysosomal Storage Disorders in India: Experience of 387 Affected Children from a Single Diagnostic Facility. in *JIMD Reports - Volume 12* (eds Zschocke, J., Gibson, K. M., Brown, G., Morava, E. & Peters, V.) vol. 12 51–63 (Springer International Publishing, Cham, 2013).
28. Park, J. H. *et al.* Novel *HEXA* variants in Korean children with Tay–Sachs disease with regression of neurodevelopment from infancy. *Mol. Genet. Genomic Med.* **9**, e1677 (2021).
29. Clarke, J. T. R. *et al.* An open-label Phase I/II clinical trial of pyrimethamine for the treatment of patients affected with chronic GM2 gangliosidosis (Tay–Sachs or Sandhoff variants). *Mol. Genet. Metab.* **102**, 6–12 (2011).
30. Álvarez-Rodríguez, A., Triggs-Raine, B., Barros-Núñez, P. & Lozano, C. M. A novel *HEXA* mutation [1393G>A (D465N)] in a Mexican Tay-Sachs disease patient: Mutation and Polymorphism Reports. *Hum. Mutat.* **17**, 437–437 (2001).
31. Zeng, B. J. *et al.* Spontaneous appearance of Tay–Sachs disease in an animal model. *Mol. Genet. Metab.* **95**, 59–65 (2008).
32. Buratti, E. *et al.* Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res.* **35**, 4250–4263 (2007).
33. Triggs-Raine, B. L. *et al.* A pseudodeficiency allele common in non-Jewish Tay-Sachs carriers: implications for carrier screening. *Am. J. Hum. Genet.* **51**, 793–801 (1992).
34. Cecchi, A. C. *et al.* Screening for Tay-Sachs disease carriers by full-exon sequencing with novel variant interpretation outperforms enzyme testing in a pan-ethnic cohort. *Mol. Genet. Genomic Med.* **7**, e836 (2019).

35. Balasubramanian, S., Maharana, S. & Srivastava, A. “Boundary residues” between the folded RNA recognition motif and disordered RGG domains are critical for FUS–RNA binding. *J. Biol. Chem.* **299**, 105392 (2023).
36. Dersh, D., Iwamoto, Y. & Argon, Y. Tay–Sachs disease mutations in HEXA target the α chain of hexosaminidase A to endoplasmic reticulum–associated degradation. *Mol. Biol. Cell* **27**, 3813–3827 (2016).