The Pathogenicity of Tafazzin Variants and their Implications in Barth Syndrome: A Bioinformatics Analysis

Markesina, Irma

UNIVERSITY OF RIJEKA

DEPARTMENT OF BIOTECHNOLOGY

Graduate program

„Biotechnology in medicine“

Irma Markesina

The Pathogenicity of Tafazzin Variants and their Implications in Barth Syndrome: A Bioinformatics Analysis

Master's thesis

Rijeka, 2023

UNIVERSITY OF RIJEKA

DEPARTMENT OF BIOTECHNOLOGY

Graduate program

„Biotechnology in medicine“

Irma Markesina

The Pathogenicity of Tafazzin Variants and their Implications in Barth Syndrome: A Bioinformatics Analysis

Master's thesis

Rijeka, 2023

Mentor: Assistant Professor Christian Reynolds, PhD

Co-mentor: Senior Research Associate Gordana Apić, PhD

Master's thesis was defended on 27th September 2023.

in front of the committee:

1. doc.dr.sc. Toni Todorovski

2. prof.dr.sc. Dean Marković

3. doc.dr.sc. Christian Reynolds

The thesis has 99 pages, 11 figures, 12 tables and 153 references.

Summary

Barth syndrome (BTHS) is an X-linked genetic disease characterized by severe cardiovascular defects, skeletal muscle weakness and neutropenia. Defects in the protein tafazzin, which is encoded by the *TAZ* gene, have been identified to give rise to BTHS through impaired capability of remodeling the mitochondrial phospholipid cardiolipin (CL). Even though the genetic defects responsible for the disease are known, the links between specific *TAZ* mutation variants and mechanisms of BTHS pathogenicity haven't been thoroughly examined. The aim of this work was to present bioinformatics methods which would facilitate this type of research and, using these methods, examine tafazzin protein features, map all known benign and pathogenic tafazzin protein changes to its corresponding domains, identify the mutations that are found in evolutionarily conserved regions, and highlight potentially important domains and mutations for future *in silico* and *in vivo* research. The variants were compiled from the Barth Syndrome Foundation's Tafazzin Human Variants Database. Databases UniProt, ClinVar, OMIM and GnomAD were used in addition. Seven pathogenicity prediction softwares were used to analyze potential pathogenicity of missense variants based on protein structure and evolutionary conservation. A species alignment analysis was done using ClustalOmega. Our results show that tafazzin is not subject to binding CL via induced-fit, but that the interaction is rather based on conformational selection, indicating a potential conformational flexibility that might enable interactions with diverse phospholipids without significant structural changes caused by ligand binding. We hypothesize that tafazzin's conformational flexibility may be influenced by the surrounding microenvironment, including membrane composition and neighboring molecules. We demonstrate pathogenic mutational enrichment in five of the protein's domains and show that 34 out of 62 pathogenic missense variants are predicted to be detrimental to the protein's function. Through further refinement, we show that two domains of

the protein, HX4D and MT1, are highlighted as significantly enriched in pathogenic variants. D74 and D75 mutations in the HX4D domain and R94 mutations in the MT1 domain, identified as pathogenic through our methods, suggest the likelihood of interactions between these domains due to the complementary charges of aspartate and arginine, hinting at a potential active site. Finally, we discuss the potential of studying these two domains and *TAZ* mutation variants in general by applying molecular dynamics simulations and using the model organism *Saccharomyces cerevisiae*.

Key words: tafazzin, *TAZ*, Barth syndrome, BTHS, bioinformatics, *in silico*, mutations, variants, enrichment, domains, conservation

Sažetak

Barthov sindrom (BTHS) je X-vezana genetska bolest karakterizirana teškim kardiovaskularnim defektima, slabošću skeletnih mišića i neutropenijom. Defekti u proteinu tafazinu kodiranom sa strane *TAZ* gena identificirani su kao uzrok BTHS-a, djelujući pogubno na sposobnost remodeliranja mitohondrijskog fosfolipida kardiolipina (CL). Iako su genetski poremećaji odgovorni za bolest poznati, poveznice između specifičnih *TAZ* varijanti i mehanizama BTHS patogenosti nisu detaljno istražene. Cilj ovog rada bio je predstaviti bioinformatičke metode koje bi omogućile takvo istraživanje, te pomoću njih proučiti proteinske značajke tafazina, mapirati sve do sada poznate benigne i patogene promjene proteina na njegove domene, identificirati mutacije koje se nalaze na evolucijski konzerviranim mjestima proteina, te istaknuti potencijalno važne domene i mutacije za buduća *in silico* i *in vivo* istraživanja. Varijante u ovom radu su primarno kompilirane iz Barth syndrome foundation baze podataka Tafazzin Human Variants, a potom su korištene i baze podataka UniProt, ClinVar, OMIM i GnomAD. Sedam programa za predviđanje patogenosti korišteno je u svrhu analize patogenosti *missense* varijanti tafazina na temelju proteinske strukture i evolucijske konzervacije. Analiza poravnanja proteinskih sekvenci između vrsti napravljena je pomoću programa ClustalOmega. Predstavljamo rezultate koji ukazuju na to da tafazin ne podliježe vezanju CL-a putem inducirane prilagodbe, već konformacijske selekcije, te da stoga postoji potencijalna konformacijska fleksibilnost proteina koja omogućuje interakcije s raznim fosfolipidima bez značajnih strukturalnih promjena uzrokovanih vezanjem liganada. Prilažemo hipotezu da konformacijska fleksibilnost tafazina podliježe utjecaju mikrookoliša, uključujući kompoziciju membrane i susjedne molekule. Pokazujemo patogeno obogaćenje u pet domena tafazina, te da su 34 od 62 patogene *missense* varijante predviđene kao pogubne za funkciju proteina. Izdvajamo dvije domene tafazina, HX4D i MT1, kao značajno obogaćene patogenim

varijantama. Identificirane kao patogene pomoću naših metoda, D74 i D75 mutacije u HX4D domeni, te R94 mutacije u MT1 domeni sugeriraju na mogućnost interakcija između ove dvije domene zbog komplementarnih naboja aspartata i arginina, što potencijalno ukazuje i na aktivno mjesto proteina. Naposlijetku, raspravljamo o potencijalu izučavanja ove dvije domene i *TAZ* mutacijskih varijanti pomoću metoda kao što su simulacije molekularne dinamike i korištenje modelnog organizma *Saccharomyces cerevisiae*.

Ključne riječi: tafazin, *TAZ*, Barthov sindrom, BTHS, bioinformatika, *in silico*, mutacije, varijante, obogaćenje, domene, konzervacija

Table of contents

1. Introduction

1.1. Tafazzin

1.1.1. Tafazzin overview

The *TAZ* gene encodes for the protein tafazzin and was first described in 1996 as the *G4.5* gene, localized on chromosome X, in distal Xq28 [1]. Its discovery was made in context of Barth syndrome (BTHS), for which mutations were identified to be responsible. The gene name *TAZ* and protein name tafazzin were introduced in 1998, after the gene was cloned and first unique mutations were found [2]. Throughout the years since its discovery, mutational and functional analyses revealed the protein's cellular functions and role in BTHS. However, the research is still in relatively early stages, with opportunity for expansion.

Tafazzin is a transacylase protein, located in the inner and outer mitochondrial membrane [1,3]. Its most studied role is the remodeling of the mitochondrial phospholipid cardiolipin (CL) [4,5]. Expression of tafazzin can be observed in most tissues, with emphasis on cardiac and skeletal muscle [6].

The *TAZ* gene consists of 11 exons [1]. However, at least 12 splice variants of tafazzin are present among different tissues in different amounts. It is suggested that either two human isoforms – the full length protein and protein lacking exon 5, or only one human isoform – the protein lacking exon 5, can functionally partake in CL remodeling [6,7]. Therefore, further consensus on the functionality of splice variants is needed.

Even though tafazzin is mostly correlated with BTHS, its aberrant expression was also revealed in different cancers [8]. It was found to play a role in apoptosis inhibition, promotion of tumorigenicity of cervical cancer, breast cancer, thyroid cancer, rectal cancer, and regulation of acute myeloid leukemia

stemness [8–12]. These findings give rise to further questions about its cellular functions.

1.1.2. Tafazzin protein domains

The protein database UniProt provides curated information on protein sequences, structures and domains. It shows that the full length human tafazzin isoform consists of 292 amino acids, while the isoform lacking exon 5 consists of 262 amino acids. The isoform lacking exon 5 is marked as the cannonical sequence and annotated with three domains – HX4D motif (aa 69-74), Mitochondrial targeting sequence (aa 82-92) and Mitochondrial targeting sequence (aa 155-190) (Figure 1).

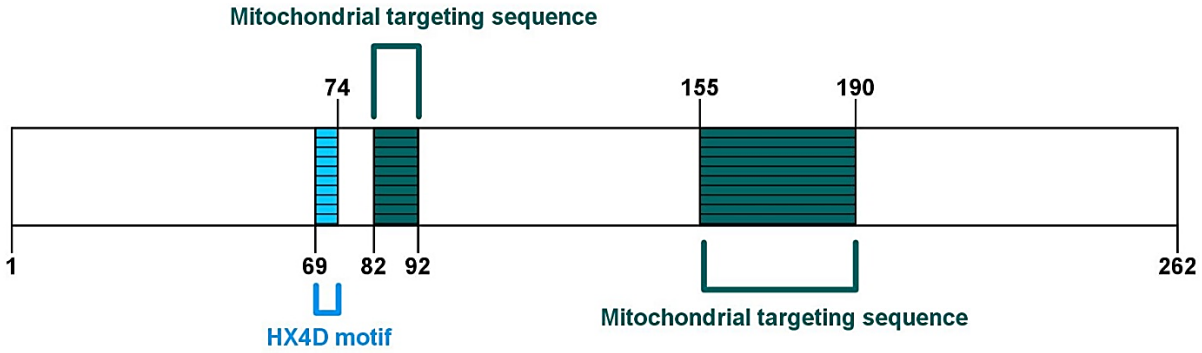

Figure 1. UniProt representation of tafazzin protein domains. UniProt suggests that tafazzin contains a HX4D motif and two Mitochondrial targeting sequence domains.

Another protein database, InterPro, classifies proteins into families and predicts their domains and important sites. InterPro suggests only one domain is present in the cannonical isoform of tafazzin – Phospholipid/glycerol acyltransferase (aa 41-217) (Figure 2).

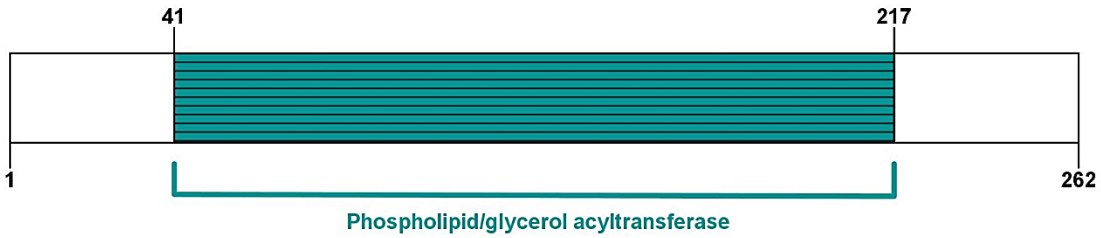

Figure 2. InterPro representation of tafazzin protein domains. InterPro predicts one Phospholipid/glycerol acyltransferase domain on tafazzin.

However, almost all relevant literature exclusively references the full length isoform, and suggests that more tafazzin domains exist. Other than the HX4D motif, mitochondrial targeting sequences and acyltransferase regions annotated by UniProt and InterPro, a transmembrane and membrane anchor domain are discussed [3,13–15]. However, the starting and ending positions of some domains slightly differ from source to source. The domains and their positions according to literature are summarized in (Table 1).

Table 1. Summary of tafazzin domains discussed in publications. TM: Transmembrane domain; AT: Acyltransferase domain; HX4D: HX4D motif; MT1: Mitochondrial targeting domain 1; MT2: Mitochondrial targeting domain 2; MA: Membrane anchor domain.

	TM	AT		HX4D	MT1	MT2	MA
Debnath et al., 2014	aa 15-35	phosphate acyltransferases	acyltransferases of phospholipid biosynthesis	aa 69-74	/	/	/
		aa 63-217	aa 41-215				
Claypool et al., 2006	/	/		/	/	/	aa 215-232 (yeast) aa 171-188 (human)
Garlid et al., 2020	aa15-aa34	aa 41-217		aa 69-74	aa 84-95	aa 185-220	aa 215-232 (human)
Dinca et al., 2018	/	/		/	aa 84-95	aa 185-220	/

The consensus domain lengths, their exact functions and share scores compared to the total protein size will be discussed in the Results section, and further addressed in the Discussion section of this study.

1.1.3. Cardiolipin remodeling

CL remodeling can occur by two distinct mechanisms [16]. In the two step deacylation-reacylation process, commonly referred to as the Lands cycle, a specific enzyme called phospholipase deacylates CL into monolyso-cardiolipin (MLCL). Following this, either an acyltransferase or a transacylase adds acyl groups from acyl-CoA or neighboring phospholipids to create remodeled CL. In an alternative mechanism, the entire process occurs in a single transacylase-catalysed step where CL and an adjacent phospholipid exchange acyl chains. As a transacylase, tafazzin has the ability to remodel CL both in the single-step transacylation process, and by reacylating MLCL to CL in the second step of the Lands cycle. CL contains acyl residues dominated by saturated chains and the process of its remodeling shifts this composition towards unsaturation, a change that is crucial for maintaining the homeostasis of the mitochondrial respiratory chain [3–5,16].

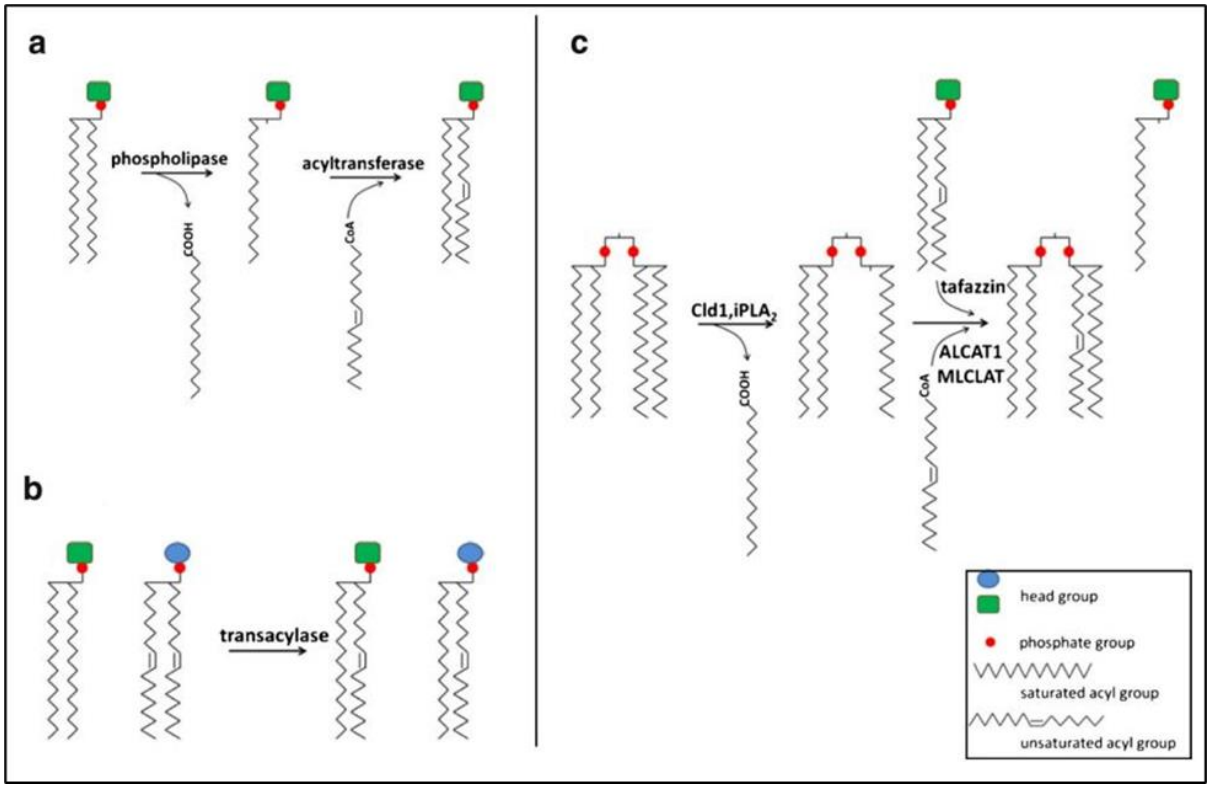

Figure 1. Phospholipid remodeling. The remodeling occurs by two mechanisms, (a) a two-step deacylation-reacylation Lands cycle, and (b) a single-step transacylation. CL remodeling (c) can occur by both mechanisms. Adopted from Ye et al., 2016 [16].

1.2. Barth syndrome

1.2.1. Barth syndrome features

BTHS was first described in 1983 as an X-linked recessive disease characterized by dilated cardiomyopathy, neutropenia and skeletal myopathy [17]. The disease occurs in approximately one in every million births, primarily affecting male population [18,19]. Cardiomyopathy, the most common clinical feature, occurs in 70% of patients within the first years of life [18,20,21]. Other clinical features include urinary excretion of 3-methylglutaconic acid, left ventricular non-compaction, endocardial fibroelastosis, ventricular arrhythmia, prolonged QTc interval, sudden cardiac death, proximal myopathy, delayed

motor milestones, compensatory monocytosis, recurrent bacterial infections, lethargy, fatigue, hypoglycaemia, growth and pubertal delay, and lactic acidosis [21].

1.2.2. Barth syndrome genetics

BTHS is caused by mutations in the *TAZ* gene, which cause defects in the protein tafazzin and alter its capability to remodel CL [1,21]. More than 120 mutations on the *TAZ* gene have been identified [21,22]. The majority of those mutations are missense mutations and small insertions or deletions. However, a small portion of patients exhibit larger deletions encompassing exons or, in one instance, the entire gene [21,23]. Additionally, frameshift mutations leading to truncation of tafazzin and mutations affecting splice donor or acceptor sites have been discovered. Mutations have been documented across all exons of *TAZ*, including an exon 5 variant of unknown significance [21,24]. Currently, no specific correlations between genotype and phenotype exist [21].

According to data from so far identified mutations, only 13% of boys carry *de novo* mutations that are not detected in their mother's somatic DNA [21,24]. Furthermore, a case of gonadal mosaicism has been documented, suggesting the remote possibility that a woman who does not possess mutations in her somatic DNA may still have more than one affected boy [21,25].

While it is theoretically possible for females to exhibit symptoms of BTHS due to skewed X-inactivation, the disease has only been reported in one female who had abnormalities in both of her X chromosomes [21,26].

1.2.3. Barth syndrome pathophysiology

CL exhibits tissue-specific variations in its fatty acyl chain configuration [21]. In tissues with high oxidative capacity, such as mammalian cardiac and skeletal muscle, the prevalent form of CL is characterized by four linoleoyl moieties. This form of CL is commonly referred to as tetralinoleoyl cardiolipin (L4-CL).

L4-CL constitutes a significant portion, accounting for up to 70-80% of the total CL content in these tissues [1,21,27].

As tafazzin is responsible for CL remodeling, mutations in the *TAZ* gene have been linked to a reduction in L4-CL production, leading to accumulation of CL molecules with different acyl compositions [21,28,29]. Consequently, an increase in intermediate species carrying three linoleoyl acyl groups, MLCL, is observed. This alteration gives rise to the significantly elevated ratio of MLCL to L4-CL, a specific diagnostic indicator for BTHS [21,30–32].

Beyond its diagnostic significance, CL has various crucial functions in mitochondrial biology. It plays an important role in maintaining mitochondrial structure and is associated with multiple mitochondrial proteins [20,21,33]. Cardiomyocyte mitochondria undergo significant morphological changes when CL is deficient, and BTHS patient biopsies exhibit mitochondrial enlargement with abnormal cristae structures [20]. Furthermore, CL has a role in stabilizing highly ordered respiratory chain supercomplexes, thus optimizing energy production within the mitochondria [21,34,35]. Changes in CL levels and species composition have been observed to affect respiratory chain structure and ROS production in patients with cardiac disease [20,36]. Finally, CL actively participates in regulating mitochondrial apoptosis [21,37]. Under pathological conditions, interaction between CL and cytochrome c results in CL peroxidation, contributing to cardiomyocyte cell death [20,38,39].

However, the correlation between pathogenicity mechanisms of BTHS and *TAZ* mutation variants remains unclear as research on the topic is still in its early stages.

1.2.4. Human Tafazzin Gene Variants Database

The Human Tafazzin Gene Variants Database by the Barth Syndrome Foundation encompasses more than 120 tafazzin mutation variants. These variants are sourced from different databases, publications, sequencing

laboratories and genetic reports shared by affected individuals and families [22]. This database also includes MLCL or CL assay results, yeast equivalent results, mRNA results in characterizing the spilicing variants, and family information regarding whether the variants are *de novo* mutations or inherited. In this study, the Human Tafazzin Gene Variants Database will serve as a comprehensive framework for all intended bioinformatics analyses, providing a foundation for data compilation and acting as a reference for assessing the coverage of other databases.

1.3. Bioinformatics analyses of proteins and genes

Standard experimental methods for protein and gene analyses involve labor-intensive and time-consuming processes. In recent years, bioinformatics has become a powerful alternative approach that is not only cost-effective and rapid but also enables targeted experimental approaches. Through the utilization of computational tools, algorithms and databases, it enables gene identification and annotation, sequence analysis, protein domain prediction, functional annotation and many other ways of helping researchers make sense of large-scale or missing data. This innovative approach is transforming the field, and the use of such tools will be the key component of this study.

1.3.1. Protein domains and 3D structure

For the analysis of protein domains, databases UniProt and InterPro offer information to a large collection of protein sequences from different organisms. While UniProt contains detailed information about the domains, including their positions within protein sequences, domain architecture and functional annotations, InterPro is based on multiple sequence alignments and predicts domain positions based on conserved sites [40].

To gain insight into protein structures, a deep learning algorithm, AlphaFold, predicts 3D structures from amino acid sequences by comparing them to a variety of known protein sequences and structures [41].

1.3.2. Intrinsically disordered proteins

Intrinsically disordered proteins are correlated to linear structures and they fold and bind mainly via induced-fit, a model of enzyme-substrate interaction which suggests that after a weak protein-ligand binding occurs, the protein undergoes conformational change in order to fit the ligand, whereas globular proteins bind to ligands mostly through conformational selection [42,43].

IUPred2A is a combined interface consisting of IUPred2 and ANCHOR2 [44]. IUPred2 is a disorder prediction method that uses an energy estimation approach to identify disordered regions in proteins. It calculates the energy contribution of intrachain interactions based on the statistical potential and amino acid composition, and the resulting scores indicate the likelihood of a residue being disordered. ANCHOR2 is another disorder prediction method that focuses on identifying disordered binding sites. It incorporates energy estimation and considers the local disordered sequence environment. The scores assigned to residues reflect their potential as disordered binding sites. This tool can help in forming hypotheses on whether the protein has a globular or linear structure, few or many interactors and whether it binds to ligands via induced-fit or through conformational selection.

1.3.3. Genetic variants

Databases UniProt, ClinVar, OMIM and gnomAD are important resources that provide an understanding of the impact of genetic variants on human health and disease.

UniProt provides a curated collection of variant annotations, including amino acid changes, positions and associated diseases, as well as providing relevant references.

ClinVar collects and curates information about genomic variants and their relationships to diseases, assigning clinical interpretations to variants, ranging from pathogenic to benign.

Online Mendelian Inheritance in Man (OMIM) catalogs information about human genes and genetic disorders. It provides detailed descriptions of the disorders, including their clinical features, inheritance patterns and associated genetic variants.

Genome Aggregation Database (gnomAD) collects and analyzes genomic data to provide information about the frequency and distribution of genetic variants in the general population [45]. It is used to determine if a variant is commonly observed in healthy individuals, to help distinguish between disease-causing and benign variants.

1.3.4. Enrichment of genetic variants

Enrichment of genetic variants is a statistical assessment on whether a specific set of genetic variants is overrepresented or underrepresented compared to what would be expected by chance. Enrichment calculations are a helpful tool to highlight crucial mutation sites and their corresponding domains.

1.3.5. Pathogenicity prediction

Pathogenicity prediction tools for genetic variants are computational programs that assess the potential impact of single amino acid substitutions in protein sequences. These tools use various features and algorithms to predict whether a variant is likely to be pathogenic or benign. In this study, seven pathogenicity prediction tools will be used:

Sorting Intolerant From Tolerant (SIFT) predicts the impact of amino acid substitutions based on sequence conservation. It compares the query sequence to a sequence database and assigns a score indicating the degree of conservation at the substituted position.

Polymorphism Phenotyping 2 (PolyPhen-2) predicts the functional impact of amino acid substitutions using a combination of sequence based and structure based features. It calculates scores based on features such as phylogenetic

conservation, protein structural properties, and sequence annotations to classify variants as either benign, possibly damaging or probably damaging.

Panther-PSEP uses evolutionary conservation information to predict the functional impact of amino acid substitutions. It assigns a score based on the observed variation at the position of interest across different species.

Meta-SNP employs a meta-analysis approach, combining the predictions of multiple pathogenicity prediction tools to provide a more accurate prediction.

Screening for Non-Acceptable Polymorphisms (SNAP) predicts the functional effects of amino acid substitutions by analyzing the protein's 3D structure and sequence conservation.

Integrated Network and Pathway Score for Multi-Dimensional Data (INPS-MD) utilizes protein interaction networks and pathway information to assess the functional impact of variants. It integrates multiple data types, including protein-protein interaction networks and functional annotations, to calculate a score for the overall impact of the variant on cellular processes.

Functional Analysis through Hidden Markov Models (FATHMM) predicts the functional consequences of variants considering features such as sequence conservation, protein domain annotations, and genomic context to estimate the likelihood of a variant being pathogenic.

1.3.6. Evolutionary conservation

Sequence alignment is a fundamental technique in bioinformatics that involves comparing and aligning protein sequences from different species to identify similarities and differences. By pinpointing evolutionarily conserved regions, variants for further investigation can be prioritized. This is crucial as the amino acid conservation indicates an important functional role of the protein, and therefore correlates with the tendency of variants at given position to be pathogenic, aiding in the understanding of disease mechanisms.

EMBL-EBI Clustal Omega is a widely used tool for sequence alignment that provides sequence comparison and conservation scores across species. This tool is important in helping researchers improve functional interpretations and pathogenicity predictions, thereby enhancing the accuracy and efficiency of their research.

1.4. Achievements and further questions in *TAZ* variant research

So far, disease models of BTHS have been made using various model organisms. The primary focus of the studies was phenotypic characterization of *TAZ* knockdowns. However, there have been investigations conducted on yeast to examine specific *TAZ* point mutations, but the precise understanding of the effects linked to these point mutations still lacks complete clarity, requiring further investigation.

TAZ knockdown in the embryonic stage of mice has resulted in an increase of MLCLs and hence the MLCL/CL ratio in heart and skeletal muscle, followed by a decreased body weight, abnormal sarcomere morphology, decreased cardiac muscle contractility, a dilated left ventricle of the heart, an abnormally high number of mitochondria and increased mitochondrial fission [46], as well as impaired skeletal muscle contractility and a disruptive lining of the inner mitochondrial membrane by mitochondrial cristae [47].

Fruit fly *TAZ* mutants exhibited compromised muscular functionality which remained unimproved with endurance exercise training [48]. Additionaly, fruit flies were employed in a study aiming to characterize splice variants of tafazzin [6].

In yeast, deletion of the *TAZ* homologue, *Taz1*, resulted in numerous defects. The null mutants were subject to chromosome instability [49], sensitive to osmotic stress [50], had abnormal mitochondria [51,52] and were impaired in respiration [53,54]. Additionally, *Taz1* point mutations have been researched in yeast and their effects are summarized in (Table 2) [55].

Table 2. *Taz1* point mutations in yeast and their effects. The effects span from lack of transacylase activity to complete destabilizatzion and inactivation of the polypeptide. Adapted from Whited et al., 2013 [55].

YEAST POINT MUTATION	EFFECT
H77Q, S79P, R102C, R102S, F112V, A118P, G124D, Y187I, V192G	Correct localization to the inner mitochondrial membrane, but absence of transacylase activity.
L90P	Significant activity with lyso-phosphatidylcholine (lyso-PC) as substrate.
N109V	No activity with lyso-PC, but measurable transacylase activity.
Q125E	Absence of mitochondrial localization and assembly, destabilization and inactivation of the mutant polypeptide.
V223D, V224R, I226P	Misincorporation into the mitochondrial matrix.
L184H	Correct localization to the inner mitochondrial membrane, but assembles into intrinsically unstable complexes.
N48D, K65L, G261R	Thermolability with significantly greater MLCL/CL ratio.

In human cells, mutations R94S, G197R, I209N, L210R, L212P and H214R did not affect mitochondrial localization of tafazzin *in vitro* [13]. However, the exact subcellular effects of human *TAZ* point mutations remain unclear. With the existing knowledge, additional questions regarding the involvement of *TAZ* mutation variants in BTHS are raised, opening opportunity for research on their relation to phenotypes and pathogenicity mechanisms.

1.5. Other methodological opportunities

On top of bioinformatics analyses that provide a basis in navigating further research, combination of methods such as *in silico* molecular dynamics and *in vivo* investigations using model organisms can form a powerful approach for next steps in research of *TAZ* variants in BTHS.

1.5.1. *In silico* molecular dynamics

In silico molecular dynamics analyses represent a computational approach that involves simulating the movements and interactions of atoms and molecules over time. This technique offers valuable insights into the behaviour and function of biological macromolecules, such as proteins and nucleic acids. To conduct *in silico* molecular dynamics simulations, a combination of a supercomputer and the CHARMM-GUI platform can be used. The CHARMM-GUI platform is particularly useful for setting up the simulation system, defining parameters and initializing the conditions, while supercomputers are crucial for their ability to distribute the computational workload across multiple processors or cores, allowing the acceleration of a complex simulation process that would otherwise be time-consuming.

This kind of analysis requires the availability of experimentally determined crystal structures of macromolecules. Recently, the crystal structure of tafazzin was obtained as part of a respiratory complex investigation in the organism *Yarrowia lipolytica* [56]. This crystal structure can provide a foundation for *in silico* exploration of alterations in tafazzin's dynamics and function across different mutation variants, thereby enhancing the understanding of its role in biological processes and disease.

1.5.2. *Saccharomyces cerevisiae* as a model organism in *TAZ* variant research

CL synthesis and remodeling occur within the inner mitochondrial membrane and are highly conserved across species, ranging from the yeast *Saccharomyces cerevisiae* to humans [57]. This conservation indicates a fundamental importance of these processes and facilitates research on various model organisms.

In the yeast *Saccharomyces cerevisiae*, the deacylation reaction is catalyzed by a single CL-specific phospholipase known as Cld1, whereas mammalian

cells employ multiple phospholipases for this process [57–60]. The reacylation of CL is facilitated by an acyl-specific transacylase called Taz1 in yeast, the equivalent to human tafazzin [57,61,62]. The yeast Taz1 transfers acyl chains from phosphatidylethanolamine (PE) or phosphatidylcholine (PC) to MLCL. As in humans, it also plays a role in CL remodeling through direct transacylation between CL and lyso-PE or lyso-PC, independent of phospholipase activity [57,61–63].

Creation of the mutant taz1Δ, resulting from deletion of the yeast *Taz1* gene, has established the first non-patient eukaryotic model for BTHS [57]. Taz1Δ exhibits biochemical deficiencies similar to those observed in BTHS, such as diminished levels of CL, a reduction in CL species harboring unsaturated fatty acids, and an accumulation of MLCL [7,57,64,65]. Furthermore, the defect can be rescued by the expression of the human *TAZ* gene, thereby complementing the mitochondrial dysfunction associated with taz1Δ [57,64], providing a powerful tool for studying the disease.

Extensive research of *TAZ* mutations in BTHS has greatly benefited from the taz1Δ model. Notably, 21 missense mutations have been reported to occur at conserved positions in the yeast orthologue [57,66].

In a previously published study, a plasmid construct Taz1p-his was generated, encompassing the Taz1 protein sequence fused with a sequence of six repeating histidine residues (His-tag) at the C-terminus [64]. This method facilitates various experimental essays, such as detection of subcellular localization of the protein, using a specific His-tag antibody, and offers a versatile approach to protein characterization.

Another method that can be used involves examining the growth of yeast on medium supplemented with non-fermentable carbon sources instead of glucose. Since the taz1Δ mutant exhibits impaired CL composition and compromised mitochondrial energy metabolism, its ability to use these

alternative carbon sources may be affected. Growth essays from our laboratory in the Department of Biotechnology, University of Rijeka have shown a defective growth of taz1Δ on synthetic minimal medium (SMD) supplemented with pyruvate, while its growth defects on SMD with ethanol have been previously published [67]. Therefore, these types of growth essays may also provide insights into the metabolic deficiencies associated with specific *Taz1* point mutations.

With these collective findings, the yeast *Saccharomyces cerevisiae* represents a promising model organism for *TAZ* variant research, offering a platform for future investigations and potential therapeutic development.

2. Thesis goal

To date, multiple *TAZ* variants have been identified in patients with BTHS. Although several model organisms, such as mice, yeast and fruit flies, have been generated to explore the effects of BTHS, there is still limited understanding of the nature and potential effects of *TAZ* mutation variants. The lack of understanding stems from the complexity of the interactions that might arise due to the mutations. The use of bioinformatics holds promise in advancing the investigation of these mutation variants. By using computational tools and algorithms, further targeted *in silico* and *in vivo* research on their effect on phenotype and pathogenicity in BTHS can be conducted.

Therefore, we set several aims for this study, using a bioinformatics approach:

1. Investigate the 3D protein structure of tafazzin, analyze its amino acid and protein sequence, and identify its disordered regions and protein domains.
2. Use literature and databases to compile the most up-to-date list of *TAZ* variants found in individuals in BTHS and healthy subjects.
3. Examine the distribution of variants across tafazzin protein domains and assess corresponding mutational enrichment scores.
4. Evaluate the pathogenicity of missense *TAZ* variants in BTHS.
5. Analyze the conservation of tafazzin across various species, asess its evolutionary importance, and investigate the influence of conservation on the occurrence of mutations.
6. Identify and prioritize *TAZ* variants and tafazzin domains that show strong evidence of being involved in BTHS for further investigation.

With these goals, we aim to contribute in gaining insights into the molecular mechanisms underlying BTHS and forming future targeted research approaches.

3. Materials and methods

3.1. Literature curation

An expert literature curation approach was used to compile comprehensive information on the domains of tafazzin, and to confirm the data on *TAZ* variants. A thorough search was conducted using reputable scientific sources such as PubMed and Google Scholar. Specific keywords were used, including but not limited to „tafazzin domains", „*TAZ* variants", „*TAZ* variants Barth syndrome" and „tafazzin variants".

Additionally, expert literature curation was strategically used to assess the coverage of variant databases and to create a comparison between the information that is available in the literature and the information that is present in the databases.

3.2. Graphical linear representation and domain mapping of the protein

The graphical linear representation of tafazzin was generated by displaying information regarding its protein domains and variants using the Illustrator for Biological Sequences (IBS 1.0) program (Protein working mode) [68].

3.3. Generation of 3D protein structures

To obtain a 3D protein structure prediction for tafazzin, the EMBL-EBI AlphaFold Protein Structure Database was used [41]. Specifically, the precomputed structure with the UniProt accession ID Q16635 (TAZ_HUMAN), corresponding to the cannonical sequence of 262 amino acids, was used. The per-residue model confidence score (pLDDT), ranging from 0 (low confidence) to 100 (high confidence), was considered.

3.4. Analysis of disordered protein regions

To identify regions of protein disorder within two tafazzin isoform sequences, IUPred2A and AlphaFold were used [41,44], with the UniProt accession IDs Q16635 and Q16635-1 (TAZ_HUMAN). Two additional protein accession IDs,

P68871 (HBB_HUMAN) and P46527 (CDN1B_HUMAN) were used in order to perform a comparative analysis.

Disordered regions were identified by values exceeding 0.5 in IUPred2A linear plots and the findings were validated by analyzing the AlphaFold 3D structures.

3.5. Database research on *TAZ* variants

The full list of *TAZ* variants was obtained by searching the Human Tafazzin Gene Variants Database by the Barth Syndrome Foundation [22], UniProt [69], ClinVar [70], OMIM [71], and gnomAD [72].

The extraction of pathogenic and benign variants was primarily performed using the Human Tafazzin Gene Variants Database, which references the 292 amino acid sequence of tafazzin (NCBI Reference Sequence NM_000116). The obtained variants were further validated by checking the references. Additionally, cross-references to ClinVar were thoroughly examined to ensure comprehensive coverage of the variants.

TAZ variants from UniProt were obtained by searching the database with the accession ID Q16635 (TAZ_HUMAN), which represents the cannonical sequence of 262 amino acids, and extracting the queried vartiants associated with BTHS.

In ClinVar, the variants were obtained by searching the database using the term „*TAZ*“. Variants associated with BTHS were selected, considering the review status of the entries.

In OMIM, the variants were obtained by querying „300394; TAFAZZIN“. The relevant information of *TAZ* variants in BTHS was collected from the chapters „Molecular genetics“ and „Allelic variants“.

In gnomAD, the variants were obtained by searching for variants associated with the *TAZ* gene labeled as „benign“ or „likely benign“.

The refinement of the compiled variants was accomplished through a comparison of database coverage, availibility of references, and review statuses, while all variants were adapted to align with the 292 amino acid sequence. This evaluation process ensured the inclusion of reliable variants in the final dataset.

3.6. Enrichment analysis

For the analysis of pathogenic and benign variants, a mapping was performed to determine the position of the variants within the tafazzin protein sequence. The enrichment of these variants in relation to the size of each protein domain was assessed using a statistical approach.

Initially, the fraction of the protein occupied by each domain was calculated by dividing the length of the domain by the total length of the protein. This calculation represented the fraction that would be expected if the variants were randomly distributed across the protein. The following formula was used:

$$\text{protein fraction occupied by domain} = \frac{\text{length of domain}}{\text{length of protein}}$$

Subsequently, the observed fraction of variants within each domain was computed by dividing the number of variants in each domain by the total number of variants. The following formula was used:

$$\text{observed fraction of the variants} = \frac{\text{number of variants in domain}}{\text{number of total variants}}$$

The enrichment value was then determined by calculating the ratio of the observed fraction to the expected fraction. Enrichment values greater than 1 were considered to indicate a higher frequency of variants in the domain compared to what would be expected by chance, while values less than 1 were considered to indicate a lower frequency than what would be expected by chance.

3.7. Pathogenicity prediction analysis

The pathogenicity of *TAZ* missense variants in BTHS was predicted using seven different programs: SIFT [73], PolyPhen-2 [74], Panther-PSEP [75], Meta-SNP [76], SNAP [77], INPS-MD [78] and FATHMM [79]. In all programs, the 292 amino acid sequence was used as the query sequence, imported in FASTA format from the UniProt accession ID Q16635-1.

The categorization of results was achieved by the programs, with each of them using an individual scoring system. Raw results were examined, considering the specific thresholds or cutoffs of each program.

In SIFT, the scores ranged from 0 to 1, where a score less than 0.05 was considered to indicate deleteriousness, and a score greater than or equal to 0.05 was considered to indicate a tolerated variant.

PolyPhen-2 provided a score ranging from 0 to 1, where a score greater than or equal to 0.85 was considered to categorize the variant as probably damaging, a score between 0.15 and 0.85 was considered to categorize it as possibly damaging, while a score less than 0.15 was considered to classify it as benign.

Panther-PSEP scores ranged from -10 to +10, where variants with scores greater than 0 were considered to be deleterious, while variants with scores less than 0 were considered to be tolerated.

Meta-SNP scores ranged from 0 to 1, with thresholds varying for individual prediction tools used in the integrated result. Generally, scores greater than 0.5 were considered to indicate a pathogenic variant, while scores less than 0.5 were considered to indicate a benign variant.

SNAP scores ranged from -100 to 100, where scores greater than 0 were considered to predict tolerated variants, while scores less than 0 were considered indicative of pathogenic variants.

INPS-MD scores represented a stability change in kcal/mol, where no specific cutoff was universal. It rather compared the scores of the variants with scores of known pathogenic and benign variants and sorted the variants into stabilizing and unstabilizing.

FATHMM provided scores ranging from -10 to +10, where a positive score was considered to indicate a higher likelihood of being damaging, and a negative score to indicate a higher likelihood of being benign.

3.8. Evolutionary conservation analysis

The conservation of the tafazzin protein sequence across 11 species was examined. A search was conducted in the UniProt database using the keyword „*TAZ*", and only the reviewed (Swiss-Prot) sequences were manually selected for analysis, including baker's yeast (Q06510/TAZ_YEAST), fruit fly (Q9V6G5/TAZ_DROME), zebrafish (F1QCP6/TAZ_DANRE), bornean orangutan (Q6IV82/TAZ_PONPY), human (Q16635-1/TAZ_HUMAN), chimpanzee (Q6IV84/TAZ_PANTR), western lowland gorilla (Q6IV83/TAZ_GORGO), mouse (Q91WF0/TAZ_MOUSE), rhesus macaque (Q6IV77/TAZ_MACMU), red guenon (Q6IV76/TAZ_ERYPA) and common squirrel monkey (Q6IV78/TAZ_SAISC).

The FASTA format of the sequences was imported into the EMBL-EBI Clustal Omega Multiple Sequence Alignment tool [80]. Additionally, a phylogenetic tree was constructed based on the percent sequence identity among the aligned sequences, providing a comprehensive view of the evolutionary relationships between the sequences.

3.9. Impact of protein conservation on mutation occurrence analysis

Predicted-pathogenic variants were analyzed, and their localization on the aligned protein sequences was assessed. A final percentage score was calculated by dividing the number of variants observed in the conserved region by the total number of the variants, to gain information on the impact of evolutionary conservation on mutation occurrence.

4. Results

4.1. Tafazzin consists of six protein domains

Through the combinaton and refinement of data on tafazzin protein domains derived from databases and literature, we reached a following consensus – tafazzin consists of six protein domains. These include the transmembrane region (TM), acyltransferase domain (AT), HX4D motif (HX4D) [3,14], mitochondrial targeting domain 1 (MT1), mitochondrial targeting domain 2 (MT2) [3,13] and membrane anchor (MA) [3,15] (Figure 3). Approximately 32% of the protein sequence lacks domain assignment, leaving it labeled as „non-assigned". Among the identified domains, the AT domain makes up the largest portion, accounting for approximately 38% of the protein, whereas the HX4D domain represents the smallest fraction. (Table 3). As for the inconsistencies with domains TM and MA mentioned in the Introduction section of the study, we mapped the TM domain as aa 14-35 in accordance with the most recently published reference, and the MA domain as aa 171-188, as it is homologous to aa 215-232 in yeast. However, due to the overlapping regions between the AT domain and HX4D, MT1, MA, and partially with the MT2 domain, we excluded the overlapping amino acids from the AT domain in further analysis to allow a focused characterization of the remaining domains. Furthermore, we came to an observation that the MA domain and the MT2 domain overlap within the region aa 185-188 that is located in the overlap with the AT domain as well. To enhance the clarity and simplify further analysis, we designated this overlapping segment as „non-assigned" until a more precise consensus can be established. These modifications will be acknowledged during interpretation and adressed in the Discussion section of the study.

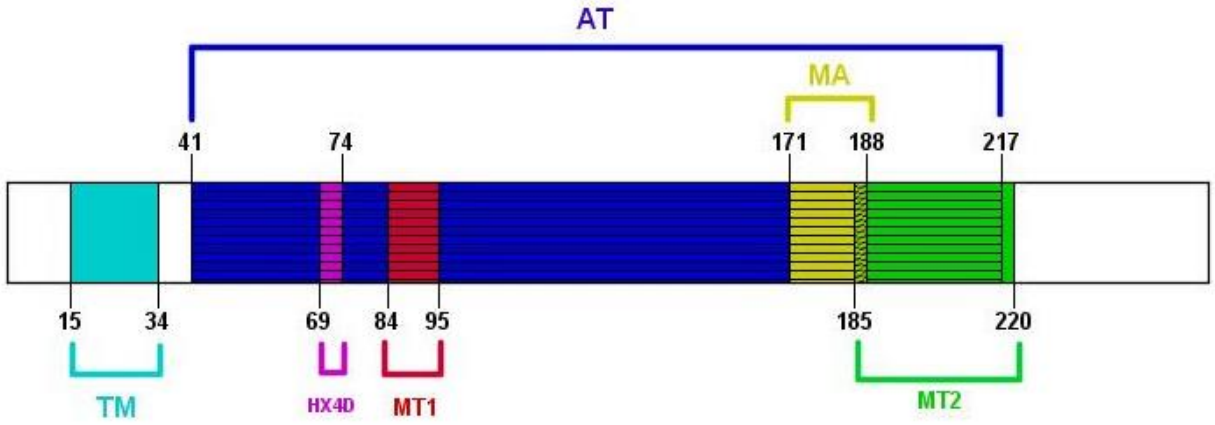

Figure 3. Tafazzin protein domains. Assigned by literature curation, tafazzin contains six domains. TM: transmembrane region; AT: acyltransferase domain; HX4D: HX4D motif; MT1: mitochondrial targeting domain 1; MA: membrane anchor; MT2: mitochondrial targeting domain 2.

Table 3. Tafazzin domain sizes compared to total protein size. The size of each protein domain was determined by calculating the total number of amino acids within each domain. To estimate the fraction occupied by a specific domain, the length of the domain was divided by the total length of the protein. n/a: non-assigned.

PROTEIN DOMAIN	DOMAIN LOCATION	AMINO ACID COUNT	% IN TOTAL PROTEIN
n/a	aa 1-14, aa 35-40, aa186-187, aa 221-292	94	32.19%
TM	aa 15-34	20	6.85%
AT	aa 41-217	112	38.36%
HX4D	aa 69-74	6	2.05%
MT1	aa 84-95	12	4.11%
MA	aa 171-185	15	5.14%
MT2	aa 188-220	33	11.30%
Σ		292	100%

The TM domain is found in the N-terminal region of the protein, implying that the N-terminal region should adopt an α-helical structure for membrane integration and anchoring [81].

The AT domain consists of an active acyltransferase segment, responsible for the transfer of acyl side chains and CL remodeling [3].

The HX4D domain comprises a histidine and aspartic acid separated by 4 amino acids and facilitates the Asp-His dyad which catalyzes chemical reactions within the protein's active site [3].

The MT1 domain provides exclusive mitochondrial localization. This region is mostly exposed and has a patch of positive residues that anchor it to the inner mitochondrial membrane [13].

The MA domain is a conserved hydrophobic sequence that plays a role in the anchorage of the membrane by providing interfacial anchoring in both the inner and outer mitochondrial membrane [3,81].

The MT2 domain is another mitochondrial targeting segment of the protein, which targets other sub-cellular compartments as well [13].

4.2. Tafazzin exhibits a globular structure

To further expand the research on tafazzin's features, we examined its 3D structure using the EMBL-EBI AlphaFold Protein Structure Database. In order to create a comparison, the 3D structure of tafazzin (Figure 4A) was generated alongside the 3D structures of haemoglobin (Figure 4B) and cyclin-dependent kinase inhibitor 1B (CDN1B) (Figure 4C). Haemoglobin, known for its globular structure, and CDN1B, known to possess a linear structure, served as references for examining tafazzin and provided insights into its structural variations. Due to the limitations of the database, we used only one isoform of tafazzin, the 262 amino acid canonical sequence.

The generated 3D structures display colored protein structures based on the per-residue confidence score (pLDDT) ranging from 1 to 100, with higher numbers indicating higher confidence. The pLDDT score predicts the Cα local-distance difference test (IDDT-Cα) score [41]. This score evaluates the differences in local distances between all atoms in a model, providing validation of stereochemical plausibility [82].

Overall, the human tafazzin model exhibits a high level of confidence. However, specific regions of the protein show low to very low confidence scores, including aa 124-153, corresponding to a segment of the AT domain, and aa 288-291, which falls into non-assigned parts of the sequence.

When compared to haemoglobin and CDN1B, tafazzin demonstrates a predominantly globular structure, with an exception in the non-assigned low confidence region, showing a potentially more open structure. This suggests that tafazzin may have fewer interacting partners, as globular proteins typically have fewer accessible interaction sites due to burial. Additionally, this should be supported by the lacking presence of disordered regions in the protein, as they are typically associated with linear proteins.

4.3. Tafazzin has little to no disordered regions

To confirm the predicted globular structure and fewer interaction sites in tafazzin, we conducted an analysis to determine the presence of disordered regions. Disordered regions in proteins lack a well-defined structure and are more often associated with linear proteins, which exhibit a higher likelihood of interaction compared to globular proteins due to their structural characteristics. We used the tool IUPred2A to assess disordered regions in the 262 and 292 amino acid isoforms of tafazzin, as well as haemoglobin and CDN1B, which again served as references for comparative analysis of intrinsically disordered regions.

Based on the threshold of 0.5, where protein regions with a score >0.5 are classified as disordered regions, both tafazzin's isoforms suggest minimal disordered regions, compared to haemoglobin with no disordered regions and CDN1B with a high number of disordered regions. This observation aligns with our previous 3D structure findings, and together, they suggest that tafazzin does not act via induced-fit, but rather through conformational selection.

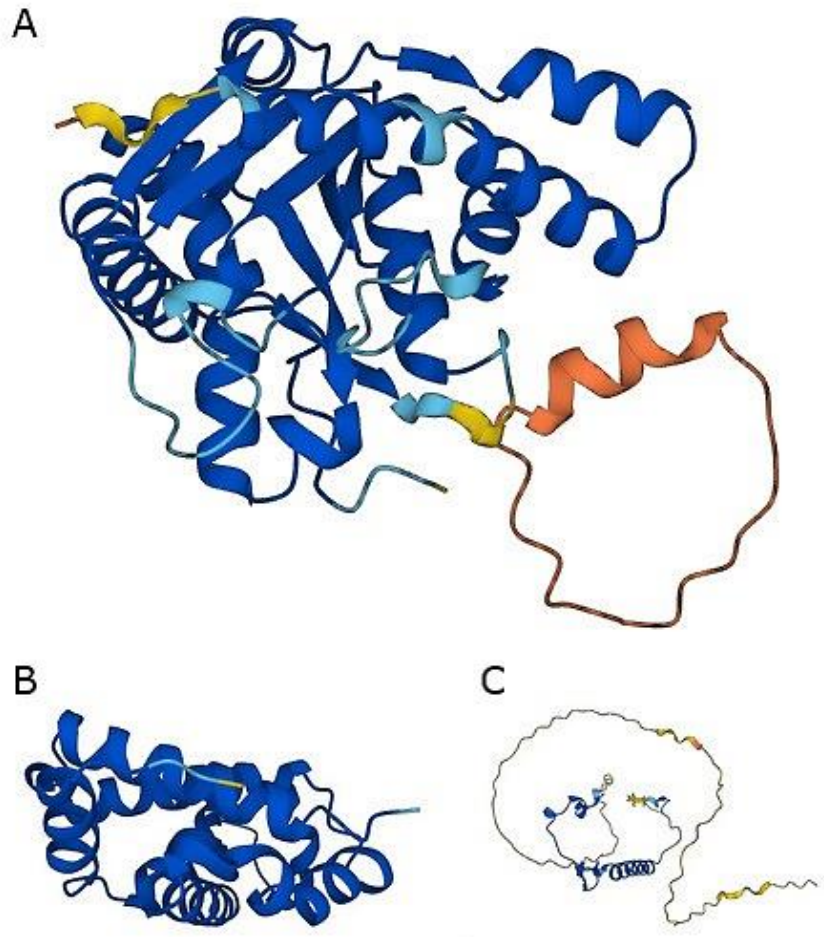

Figure 4. AlphaFold representation of protein 3D structures. Tafazzin (A) and haemoglobin (B) exhibit a globular structure. In contrast, CDN1B (C) represents a linear structure. Dark blue: very high confidence (pLDDT > 90); light blue: high confidence (90 > pLDDT > 70); yellow: low confidence (70 > pLDDT > 50); orange: very low confidence (pLDDT < 50).

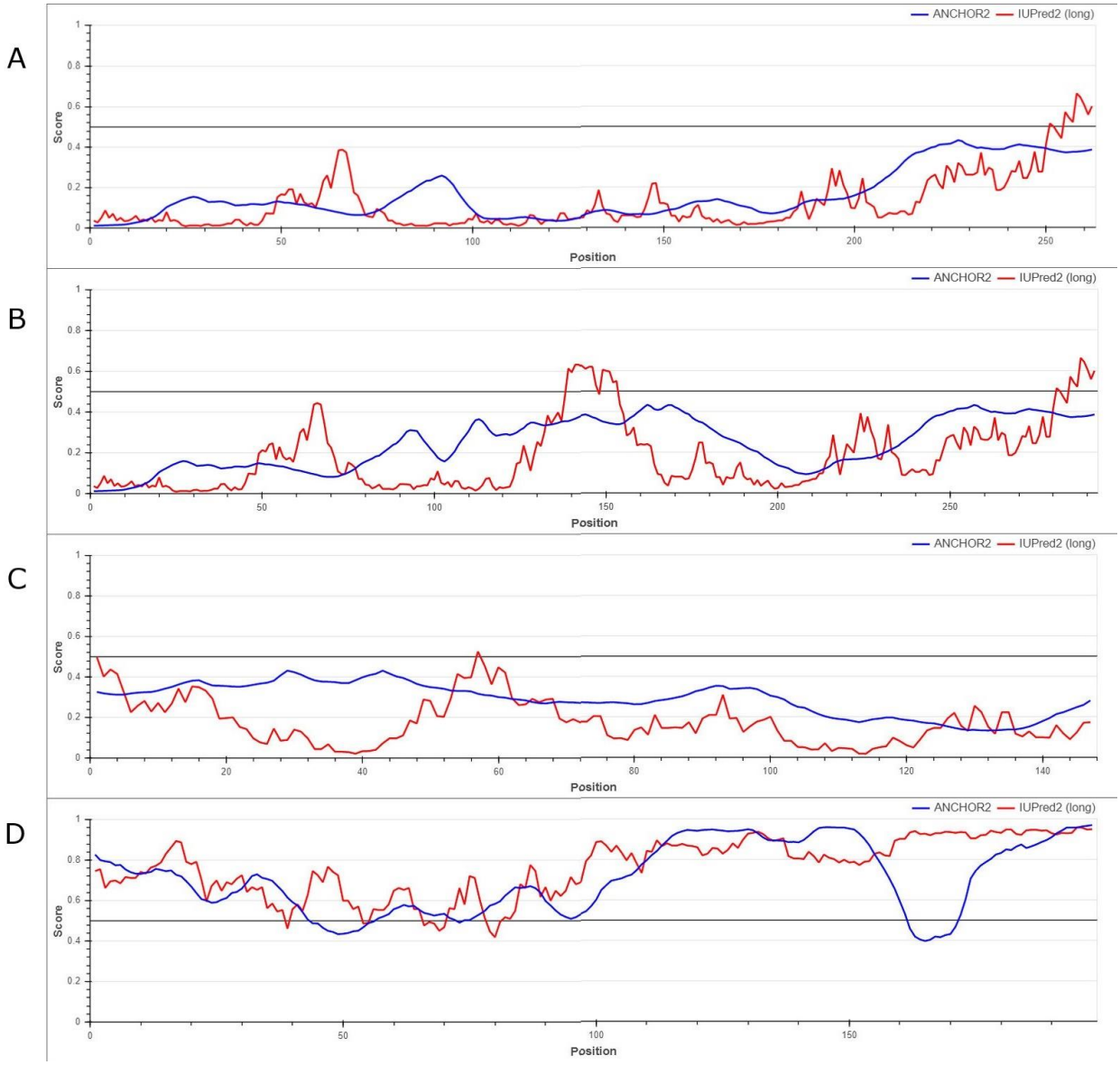

Figure 5. IUPred2A linear plots of disordered protein regions. Canonical sequence of tafazzin lacking exon 5 (A) exhibits only a small fragment of disordered regions in exon 11, while the full length isoform (B) shows an additional disordered region in exon 5. When put in comparison with haemoglobin (C) that has no disordered regions, and CDN1B (D) that is disordered throughout the whole sequence, both isoforms of tafazzin can be classified as having minimal disordered regions.

4.4. *TAZ* pathogenic variants

To establish a connection between these findings and the pathogenicity of BTHS, we conducted a comprehensive analysis of the *TAZ* gene and tafazzin protein variants in BTHS. Our research involved using the Human Tafazzin Gene Variants Database from The Barth Syndrome Foundation (BSF) as a framework, and databases UniProt, ClinVar and OMIM in addition to estimate the variant coverage. To ensure the accuracy of our findings, we also carried out an expert literature curation approach.

We documented each variant with its corresponding nucleotide change, protein change, source, MLCL/CL effect (if available) and whether it is *de novo* or inherited (if available) (Table 4).

It is important to note that the understanding of *TAZ* gene and tafazzin protein variants is constantly evolving with advancements in genomic research. We used the full length isoform of tafazzin to ensure complete coverage of the variants, however, further consensus on the functional isoform and its variants may undergo change.

The final curated list comprises a total of 139 mutation variants, including 66 missense mutations, 23 nonsense mutations, 44 frameshift mutations and 6 deletions. All of these variants are encompassed within the BSF database. Among them, 31 variants are covered by ClinVar, 7 by OMIM, and 16 by UniProt. When compared to the BSF database, the coverage of ClinVar accounts for approximately 22%, while OMIM covers around 5% and UniProt covers approximately 12% of the variants. Around 14% of the variants are categorized as strictly *de novo*, indicating they occured in the affected individual and were not inherited from their parents. The remaining mutations are classified as familial and germline, meaning they are inherited from parents and can be passed on to subsequent generations. Furthermore, an aberrant MLCL/CL profile is present in roughly 45% of the variants.

Table 4. List of *TAZ* variants in BTHS. The variant analysis was conducted through expert literature curation and database research. The variants are categorized based on their type, including missense, nonsense, frameshift and deletion variants. In addition to the variant information, MLCL/CL profile and inheritance data are provided whenever available, along with corresponding references and database coverage.

VARIANT TYPE	NUCLEOTIDE CHANGE	PROTEIN CHANGE	MLCL/CL EFFECT	INHERITANCE	DATABASE COVERAGE	REFERENCES	DIRECT BSF SUBMISSIONS
MISSENSE	c.83T>A GTG > GAG	p.Val28Glu		familial mosaicism	BSF	Zapała et al. (2015) [83]	Barbara Zapała, Teresa Staszel, Iwona Wybrańska, Beata Kieć-Wilk. Department of Clinical Biochemistry, Jagiellonian University Medical College, Krákow
	c.86G>A GGC > GAC	p.Gly29Asp		familial	BSF	Mazurova et al. (2013) [84]	
	c.118A>G AAC -> GAC	p.Asn40Asp	MLCL elevated; CL normal in fresh blood	familial	BSF	Bowron et al. (2014) [85], Hastings et al. (2009) [86]	Steward, CG. Direct submission. Bristol Royal Hospital for Sick Children, Dept. of Paediatric Oncology/BMT
	c.127A>C ACC > CCC	p.Thr43Pro		*de novo*	BSF	Bachou et al. (2009) [87]	
	c.137A>T	p.Asn46Ile	MLCL/CL = 15.52		BSF		Reported by family

	c.149T>C CTG > CCG	p.Leu50Pro	MLCL↑/CL↓	germline	BSF		Bowles, KR. John Welsh Cardiovascular Diagnostic Laboratory, Baylor College of Medicine; Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Melanie Pennock / Maggie Williams. Bristol Genetics Laboratory
	c.161T>A ATC > AAC	p.Ile54Asn	MLCL/CL = 18.37	familial	BSF	Thompson et al. (2016) [88]	Bowles, KR. John Welsh Cardiovascular Diagnostic Laboratory, Baylor College of Medicine, Houston
	c.170G>T CGA > CTA	p.Arg57Leu	MLCL elevated; CL normal in fresh blood	familial	BSF, UniProt	Kuijpers et al. (2004) [89], Bowron et al. (2014) [85], Lu et al. (2016) [90], Houtkooper et al. (2009) [91], Chatzispyrou et al. (2018) [92], Suzuki-Hatano et al. (2019) [93]	
	c.185C>T CCC > CTC	c.Pro62Leu			BSF		Medizinisch Genetisches Zentrum (MGZ)

	c.207C>G CAC > CAG	p.His69Gln	MLCL↑/CL↓	familial	BSF, ClinVar, UniProt	D'Adamo et al. (1997) [94], Steward et al. (2010) [95], Hastings et al. (2009) [86], Lu et al. (2016) [90], Abe et al. (2016) [96], Thompson et al. (2016) [88]	Reported by family
	c.211T>C TCC > CCC	p.Ser71Pro			BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Medizinisch Genetisches Zentrum (MGZ)
	c.222C>A GAC > GAA	p.Asp74Glu			BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.222C>G GAC > GAG	p.Asp74Glu		familial	BSF		Reported by family
	c.223G>C GAC > CAC	p.Asp75His	MLCL/CL = 28.04	*de novo*	BSF	Thompson et al. (2016) [88]	Reported by family
	c.223G>A GAC > AAC	p.Asp75Asn		familial	BSF		Reported by family
	c.227C>G	p.Pro76Arg		*de novo*	BSF, ClinVar		

	c.239G>A GGG > GAG	p.Gly80Glu		familial	BSF	Vanderniet et al. (2021) [97]	Steward, CG. Bristol Royal Hospital for Sick Children, Dept. of Paediatric Oncology/BMT; Melanie Pennock / Maggie Williams. Bristol Genetics Laboratory
	c.245T>C CTG > CCG	p.Leu82Pro		*de novo*	BSF	Gonzalez et al. (2005) [24]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.280C>T CGT > TGT	p.Arg94Cys	MLCL↑/CL↓	*de novo*, familial, germline	BSF, ClinVar, UniProt	Johnston et al. (1997) [98], Gonzalez et al. (2005) [24], Kirwin et al. (2013) [99], Thompson et al. (2016) [88], Houtkooper et al. (2009) [91], Sabater-Molina et al. (2013) [100], Mazurova et al. (2013) [84], Borna et al. (2017) [101], Imai-Okazaki et al. (2019) [102]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Steward, CG. Bristol Royal Hospital for Sick Children, Dept. of Paediatric Oncology/BMT; Julie Honeychurch / Maggie Williams. Bristol Genetics Laboratory
	c.280C >G CGT > GGT	p.Arg94Gly	MLCL↑/CL↓	*de novo*	BSF	Steward et al. (2010) [95]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory

	c.280C>A CGT > AGT	p.Arg94Ser	MLCL↑/CL↓		BSF, ClinVar, OMIM, UniProt	Sakamoto et al. (2002) [103], Rigaud et al. (2013) [104]	Reported by family
	c.281G>A CGT > CAT	p.Arg94His	MLCL↑/CL↓	*de novo*, familial, germline	BSF, ClinVar, UniProt	Brady et al. (2006) [105], Claypool et al. (2011) [106], Rigaud et al. (2013) [104], Thiels et al. (2016) [107], Thompson et al. (2016) [88]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; John Welsh Cardiovascular Diagnostic Laboratory, Baylor College of Medicine, Houston; Melanie Pennock / Maggie Williams. Bristol Genetics Laboratory; van Slegtenhorst M. Clinical Genetics, Erasmus MC Rotterdam; Belén Pérez. Centro de Diagnóstico de Enfermedades Moleculares, Univ. Autónoma de Madrid; Reported by family
	c.281G>CT CGT > CTT	p.Arg94Leu		germline	BSF, ClinVar, UniProt		
	c.281G>C CGT > CCT	p.Arg94Pro		germline	BSF		Reported by family
	c.302G>T GAC > TAC	p.Asp101Tyr	CL↓		BSF	Vreken et al. (2000) [108]	

	c.302A>T GAC > GTC	p.Asp101Val			BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.307T>C TGC > CGC	p.Cys103Arg		familial	BSF, ClinVar, UniProt		Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics
	c.310T>G TTC >GTC	p.Phe104Val		*de novo*	BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.310T>C TTC > CTC	p.Phe104Leu		germline	BSF, ClinVar, UniProt		Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics

	c.328T>C TCC > CCC	p.Ser110Pro		*de novo*, germline	BSF, ClinVar, UniProt	Peake et al. (2017) [109]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics
	c.346G>C GGC > CGC	p.Gly116Arg	MLCL↑/CL↓		BSF	Thompson et al. (2016) [88]	Melanie Pennock / Maggie Williams. Bristol Genetics Laboratory
	c.347G>A GGC > GAC	p.Gly116Asp		germline	BSF, ClinVar, UniProt		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics
	c.349A>G AAG > GAG	p.Lys117Glu		*de novo*	BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory

	c.352T>C TGT > CGT	p.Cys118Arg	CL↓	familial, germline	BSF, ClinVar, OMIM, UniProt	Ichida et al. (2001) [110], Schlame et al. (2003) [111]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.355G>A	p.Val119Met			BSF, ClinVar, UniProt	Al-Wakeel-Marquard et al. (2019) [112]	
	c.356T>G GTG > GGG	p.Val119Gly	MLCL↑/CL↓	familial	BSF	Rigaud et al. (2013) [104]	
	c.358C>A CCT > ACT	p.Pro120Thr	MLCL↑/CL↓		BSF		Julie Honeychurch / Maggie Williams. Bristol Genetics Laboratory
	c.370G>A GGA > AGA	p.Gly124Arg			BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.370G>C GGA > CGA	p.Gly124Arg		familial	BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.476A>C CAG > CCG	p.Gln159Pro		familial	BSF	Wang et al. (2017) [113], Hirono et al. (2018) [114]	
	c.481G>A GGG > AGG	p.Gly161Arg	MLCL/CL = 17.04	familial, germline	BSF	Thompson et al. (2016) [88], Wu et al. (2019) [115]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.505C>T	p.Leu169Phe		familial	BSF	Wang et al. (2017) [113], Hirono et al. (2018) [114]	

	c.506T>A CTC > CAC	p.Leu169His			BSF		Medizinisch Genetisches Zentrum (MGZ)
	c.526C>T CAT > TAT	p.His176Tyr	MLCL/CL = 17.26	germline	BSF, ClinVar	Wang et al. (2017) [113], Hirono et al. (2018) [114], Thompson et al. (2016) [88]	
	c.527A>G CAT > CGT	p.His176Arg		familial	BSF	Wang et al. (2017) [113]	Reported by family
	c.532T>A TTC > ATC	p.Phe178Ile		familial	BSF	D'Adamo et al. (1997) [94]	
	c.532T>C TTC > CTC	p.Phe178Leu			BSF		Ewa Pronicka & Joanna Trubicka. Dept of Medical Genetics, The Children's Memorial Health Institute
	c.548T>G GTG > GGG	p.Val183Gly		familial	BSF	Cantlay et al. (1999) [116]	
	c.578A>T AAG > ATG	p.Lys193Met		familial	BSF	Bissler et al. (2002) [117]	
	c.584G>T GGA > GTA	p.Gly195Val		*de novo*	BSF	Mazurova et al. (2013) [84]	

	c.589G>A GGG > AGG	p.Gly197Arg	MLCL↑/CL↓	familial, germline	BSF, ClinVar, OMIM	Johnston et al. (1997) [98], Christodoulou et al. (1994) [118], Cantlay et al. (1999) [116], Steward et al. (2010) [95], D'Adamo et al. (1997) [94], Bleyl et al. (1997) [119], Kirwin et al. (2013) [99], Hastings et al. (2009) [86], Donati et al. (2006) [120], Rigaud et al. (2013) [104], Wang et al. (2017) [113], Hirono et al. (2018) [114], Pronicka et al. (2016) [121]	Steward, CG. Bristol Royal Hospital for Sick Children, Dept. of Paediatric Oncology/BMT; Bowles, KR. John Welsh Cardiovascular Diagnostic Laboratory, Baylor College of Medicine, Houston; Reported by family
	c.589G>T GGG > TGG	p.Gly197Trp	MLCL↑/CL↓	germline	BSF	Rigaud et al. (2013) [104]	Medizinisch Genetisches Zentrum (MGZ)
	c.590G>T GGG > GTG	p.Gly197Val	CL↓		BSF	Rigaud et al. (2013) [104], Kuijpers et al. (2004) [89], Valianpour et al. (2002) [122], Houtkooper et al. (2009) [91]	Medizinisch Genetisches Zentrum (MGZ)

	c.590G>A GGG > GAG	p.Gly197Glu		germline	BSF, ClinVar	Johnston et al. (1997) [98], Kelley et al. (1991) [123], Gonzalvez et al. (2013) [124]	Bowles, KR. John Welsh Cardiovascular Diagnostic Laboratory, Baylor College of Medicine, Houston; Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics
	c.602C>T	p.Ala201Val		familial	BSF	Tsujii et al. (2018) [125]	
	c.622A>G ATC > GTC	p.Ile208Val		familial	BSF	Zhao et al. (2017) [126]	
	c.626T>A ATC > AAC	p.Ile209Asn	MLCL/CL = 18.85	familial	BSF	Johnston et al. (1997) [98], Schlame et al. (2003) [111], Gonzalvez et al. (2013) [124], Thompson et al. (2016) [88], Steward et al. (2010) [95]	
	c.629T>G CTG > CGG	p.Leu210Arg	CL↓	*de novo*	BSF	Cantlay et al. (1999) [116]	Melanie Pennock / Maggie Williams. Bristol Genetics Laboratory
	c.635T>C CTG > CCG	p.Leu212Pro	CL↓	familial	BSF	Johnston et al. (1997) [98], Schlame et al. (2003) [111]	
	c.641A>G CAT > CGT	p.His214Arg	MLCL/CL = 70		BSF	Ferri et al. (2013) [127]	

	c.646G>C GGA > CGA	p.Gly216Arg	MLCL/CL = 16	familial	BSF, ClinVar	D'Adamo et al. (1997) [94], Kuijpers et al. (2004) [89], Valianpour et al. (2002) [122], Takeda et al. (2011) [128], Rigaud et al. (2013) [104], Folsi et al. (2014) [129], Imai-Okazaki et al. (2017) [130], Hirono et al. (2018) [114], Baban et al. (2019) [131]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Medizinisch Genetisches Zentrum (MGZ); Bowles, KR. John Welsh Cardiovascular Diagnostic Laboratory, Baylor College of Medicine, Houston; van Slegtenhorst M. Clinical Genetics, Erasmus MC Rotterdam; Anthony Krentz. PreventionGenetics, Marshfield; Reported by family
	c.646G>A GGA > AGA	p.Gly216Arg	MLCL↑/CL↓	*de novo*, familial, germline	BSF	Baban et al. (2019) [131]	
	c.647G>T GGA > GTA	p.Gly216Val		germline	BSF, ClinVar		Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics

	c.718G>C GGG > CGG	p.Gly240Arg	CL↓	germline	BSF, ClinVar	Man et al. (2013) [132]	Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics
	c.718G>A GGG > AGG	p.Gly240Arg		familial	BSF, ClinVar, OMIM	D'Adamo et al. (1997) [94], Lindenbaum et al. (1973) [133], Bissler et al. (2002) [117], Schlame et al. (2003) [111], Vasilescu et al. (2018) [134]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Reported by family; Anthony Krentz. PreventionGenetics, Marshfield
	c.800C>A ACG > AAG	p.Thr267Lys		familial	BSF	Brión et al. (2016) [135]	
NONSENSE	c.51G>A TGG > TGA	p.Trp17*	MLCL↑/CL↓	familial	BSF	Hastings et al. (2009) [86], Sağ et al. (2019) [136]	Steward, CG. Bristol Royal Hospital for Sick Children, Dept. of Paediatric Oncology/BMT; Medizinisch Genetisches Zentrum (MGZ)
	c.124delC CTGA >TGA	p.Leu42*	MLCL/CL = 15.25	familial	BSF	Schlame et al. (2003) [111], Ichida et al. (2001) [110], Kirwin et al. (2013) [99]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory

	c.153C>G TAC > TAG	p.Tyr51*	MLCL/CL = 20	familial, germline	BSF, ClinVar, OMIM, UniProt	Bione et al. (1996) [1], D'Adamo et al. (1997) [94], Kuijpers et al. (2004) [89], Valianpour et al. (2002) [122], Yen et al. (2008) [137], Seitz et al. (2020) [138], Thompson et al. (2016) [88], Houtkooper et al. (2009) [91], Chatzispyrou et al. (2018) [92]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.154G > T	p.Glu52*		germline	BSF, ClinVar		
	c.163G>T GAG > TAG	p.Glu55*		germline	BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.208C>T CAG >TAG	p.Gln70*		germline	BSF, ClinVar, UniProt		Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics
	c.216C>A TGC >TGA	p.Cys72*	MLCL/CL = 12.98	germline	BSF		Julie Honeychurch / Maggie Williams. Bristol Genetics Laboratory
	c.236G >A TGG > TAG	p.Trp79*	MLCL/CL = 22.05	*de novo*	BSF	Gonzalez et al. (2005) [24], Schlame et al. (2003) [111], Kirwin et al. (2013) [99],	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory

						Thompson et al. (2016) [88]	
	c.264G>A TGG > TGA	p.Trp88*		familial	BSF		Anthony Krentz. PreventionGenetics, Marshfield
	c.285G>A TGG > TGA	p.Trp95*			BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.367C>T CGA > TGA	p.Arg123*	MLCL↑/CL↓	familial, germline	BSF, ClinVar, UniProt	Kirwin et al. (2013) [99], van Werkhoven et al. (2006) [32], Ferri et al. (2013) [127], Wang et al. (2017) [113], Imai-Okazaki et al. (2017) [102]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; John Welsh Cardiovascular Diagnostic Laboratory, Baylor College of Medicine, Houston
	c.370G>T GGA > TGA	p.Gly124*	CL↓	familial, germline	BSF, ClinVar, UniProt	Schlame et al. (2003) [111]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.478A>T AAG > TAG	p.Lys160*	MLCL↑/CL↓	familial	BSF	Rigaud et al. (2013) [104]	Steward, CG. Bristol Royal Hospital for Sick Children, Dept. of Paediatric Oncology/BMT

	c.497T>A TTG > TAG	p.Leu166*		familial	BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.562G>T GAA > TAA	p.Glu188*		familial	BSF	Bissler et al. (2002) [117]	
	c.581G>A	p.Trp194*	MLCL/CL = 34	germline	BSF	Thompson et al. (2016) [88]	Julie Honeychurch / Maggie Williams. Bristol Genetics Laboratory; Ewa Pronicka & Joanna Trubicka. Dept of Medical Genetics, The Children's Memorial Health Institute
	c.582G>A	p.Trp194*			BSF		Reported by family
	c.583G>T GGA > TGA	p.Gly195*		familial	BSF, ClinVar	Ronvelia et al. (2012) [139]	
	c.694G>T GGA > TGA	p.Gly232*			BSF		Bowles, KR. John Welsh Cardiovascular Diagnostic Laboratory, Baylor College of Medicine, Houston
	c.697C>T CAG > TAG	p.Gln233*	MLCL/CL = 54.06	familial, germline	BSF, ClinVar	Johnston et al. (1997) [98], Kelley et al. (1991) [123], Ances et al. (2006) [140], Thompson et al. (2016) [88]	Bowles, KR. John Welsh Cardiovascular Diagnostic Laboratory, Baylor College of Medicine, Houston

	c.748G>T GAG > TAG	p.Glu250*	MLCL/CL = 21.82	*de novo*	BSF	Thompson et al. (2016) [88], Al-Wakeel-Marquard et al. (2019) [112], Kirwin et al. (2013) [99]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.773C>A TCG > TAG	p.Ser258*	MLCL↑/CL↓		BSF, ClinVar		Melanie Pennock / Maggie Williams. Bristol Genetics Laboratory; Anthony Krentz. PreventionGenetics, Marshfield
	c.823C>T CAG > TAG	p.Gln275*		germline	BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics
FRAMESHIFT	c.9_10dupG	p.His4Alafs*130	MLCL↑/CL↓	*de novo*	BSF	Cantlay et al. (1999) [116], Hastings et al. (2009) [86], Imai-Okazaki et al. (2019) [102]	
	c.39_60del22	p.Pro14Alafs*19	MLCL↑/CL↓	familial	BSF	Borna et al. (2017) [101]	

	c.53_54delCC	p.Leu19Glyfs*114	MLCL↑/CL↓	germline	BSF	Johnston et al. (1997) [98], Gonzalez et al. (2005) [24], Schlame et al. (2003) [111], Kirwin et al. (2013) [99]	
	c.134_136delinsCC	p.His45Profs*38		familial	BSF	Wang et al. (2017) [113]	
	c.140_152del13 [GGGAGGTGCTGTA]del	p.Arg47Thrfs*32	CL↓	*de novo*	BSF	D'Adamo et al. (1997) [94], Gonzalez et al. (2005) [24], Schlame et al. (2003) [111], Vreken et al. (2000) [108], Kirwin et al. (2013) [99], Suzuki-Hatano et al. (2019) [93]	
	c.143 delGG	p.Glu48Aspfs85*	MLCL↑/CL↓	germline	BSF	Rigaud et al. (2013) [104]	
	c.157dupC CTC > CCTC	p.Leu53Profs*81		*de novo*	BSF	Hirono et al. (2018) [114]	
	c.158_162del TCATC	p.Leu53Argfs*79		familial	BSF	Hastings et al. (2009) [86]	
	c.171delA	p.Gly58Alafs*25	MLCL/CL = 27.41	*de novo*	BSF	Johnston et al. (1997) [98], Gonzalez et al. (2005) [24], Schlame et al. (2003) [111], Kelley et al. (1991) [123], Kirwin et al. (2013) [99], Thompson et al. (2016) [88]	chlame
	c.227delC	p.Pro76Leufs*7		germline	BSF	Kim et al. (2013) [141]	
	c.239_240insC	p.Ile81Aspfs*53	CL↓		BSF	Valianpour et al. (2002) [122], Vreken et al. (2000) [108]	

	c.253insC CGC > CCG	p.Arg85Profs*54	MLCL↑/CL↓	familial	BSF	Avdjieva-Tzavella et al. (2016) [142]	
	c.284dupG TGgtg > TGGgtg	p.Thr96Aspfs*37	MLCL↑/CL↓	germline	BSF	Ferri et al. (2013) [127]	
	c.348delC	p.Lys117Serfs*22		germline	BSF	Nie et al. (2019) [143]	
	c.478del A AAG > AGG	p.Lys160Argfs*24			BSF		Medizinisch Genetisches Zentrum (MGZ)
	c.481_482ins20	p.Gly161Glufs*30		familial	BSF	Woiewodski et al. (2017) [144], Finsterer et al. (2017) [145]	Anthony Krentz. PreventionGenetics, Marshfield
	c.492insC	p.Ile165Hisfs*39			BSF		Reported by family
	c.517delG GAC > ACT	Asp173Thrfs*11	MLCL↑/CL↓	*de novo*, germline	BSF	Gonzalez et al. (2005) [24], Schlame et al. (2003) [111], Kirwin et al. (2013) [99]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory
	c.536delC (CCA G> CAG)	p.Pro179Glnfs*5		familial	BSF	Vesel et al. (2003) [146]	
	c.553A>G	p.Lys182Glnfs*4	MLCL↑/CL↓	familial	BSF	Fan et al. (2013) [147], Bowron et al. (2014) [85]	Julie Honeychurch / Maggie Williams. Bristol Genetics Laboratory
	c.580_581dupT	p.Trp194Leufs*9	MLCL↑/CL↓	familial	BSF	Bione et al. (1996) [1], D'Adamo et al. (1997) [94], van Werkhoven et al. (2006) [32]	
	c.605_608delAGTG	p.Glu202Valfs*15		familial	BSF, ClinVar, OMIM	Marziliano et al. (2007) [148]	
	c.634delC CCCC > CCC	p.Leu212Cysfs*6	MLCL↑/CL↓	familial	BSF, ClinVar, OMIM	D'Adamo et al. (1997) [94], van Werkhoven et al. (2006) [32]	

	c.654_655dupGAAT	p.Asp219Glufs*1	MLCL↑/CL↓		BSF	Thiels et al. (2016) [107]	Medizinisch Genetisches Zentrum (MGZ)
	c.658_661dup (dupGTCC)	p.Leu221Argfs*4	MLCL↑/CL↓	*de novo*	BSF	Rigaud et al. (2013) [104]	
	c.678_691del14	p.Tyr227Trpfs*79		germline	BSF	Ferri et al. (2013) [127]	
	c.686_690delCCCGC	p.Pro229Leufs*79		familial	BSF		Surita Meldau. IMD Laboratory, Chemical Pathology, National Health Service
	c.688delC	p.Arg230Alafs*9		germline	BSF	Johnston et al. (1997) [98], Christodoulou et al. (1994) [118]	
	c.703dupA	p.Ile235Asnfs*76	MLCL/CL = 32.23; 23.58	familial	BSF	Schlame et al. (2003) [111], Kirwin et al. (2013) [99], Thompson et al. (2016) [88]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Medizinisch Genetisches Zentrum (MGZ)
	c.710_711delTG	p.Val237Alafs*73		familial, germline	BSF, ClinVar	Wang et al. (2017) [113]	John Christodoulou. Western Sydney Genetics Program; Laboratory for Molecular Medicine (LMM), Harvard Partners Center for Genetics and Genomics

	c.735_736insCCCT	p.Leu246Profs*66		familial	BSF		Medizinisch Genetisches Zentrum (MGZ)
	c.738_739insG	p.Pro247fs*64			BSF		Reported by family
	c.747_747delCGinsA	p.Glu250Serfs*17		*de novo*	BSF		Reported by family
	c.747_748ins19CGGGCGCCGGCGGCTCTCC	p.Glu250Argfs*67		familial	BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Medizinisch Genetisches Zentrum (MGZ)
	c.748delGinsAAA	p.Glu250fs*18			BSF		Reported by family
	c.754_763del	p.Leu252Argfs*12		germline	BSF	Kirwin et al. (2013) [99], Thompson et al. (2016) [88], Al-Wakeel-Marquard et al. (2019) [112]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Medizinisch Genetisches Zentrum (MGZ)
	c.774_775insGCTCC	p.Val260Profs*9		familial	BSF		Reported by family

	c.800_818delinsGGG	p.Thr267Argfs*6		*de novo*	BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Medizinisch Genetisches Zentrum (MGZ)
	c.798-801delGACG	p.Asp268Serfs*9			BSF	Tikhomirov et al. (2008) [149]	
	c.821_824del,	p.Ile274Lysfs*7			BSF		Reported by family
	c.836delC	p.Thr279Ilefs*60		germline	BSF, ClinVar		
	c.837_838delTC	p.Gln280Glyfs*30	MLCL↑/CL↓	familial	BSF	Steward et al. (2010) [95]	Julie Honeychurch / Maggie Williams. Bristol Genetics Laboratory
	c.873_874insCCTGG	p.Arg292Leufs*49	MLCL/CL = 2.67; 4.75	*de novo*	BSF	Kirwin et al. (2007) [150], Kirwin et al. (2013) [99], Thompson et al. (2016) [88]	Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Medizinisch Genetisches Zentrum (MGZ)
	c.873delG	p.Arg292Aspfs*47			BSF	Sweeney et al. (2008) [151]	
DELETION	c.82_84del GTG	p.Val28del	MLCL↑/CL↓	*de novo*	BSF	Kirwin et al. (2013) [99]	Steward, CG. Bristol Royal Hospital for Sick Children, Dept. of Paediatric Oncology/BMT

	c.164_178del15	p.Glu55Ala (delKRGP)	MLCL/CL = 8.21	familial	BSF		Reported by family
	c.327_332delCTCCCA	p.Ser110_His111del		familial	BSF		Kirwin, S, Gonzalez, IL. A I duPont Hospital for Children, Molecular Diagnostics Laboratory; Medizinisch Genetisches Zentrum (MGZ)
	c.337_339delTTC	p.Phe113del	MLCL↑/CL↓		BSF		Medizinisch Genetisches Zentrum (MGZ); Julie Honeychurch / Maggie Williams. Bristol Genetics Laboratory
	c.348C>T	p.Lys117_Gly124del	MLCL/CL = 3.9	familial	BSF	Baban et al. (2019) [131], Ferri et al. (2016) [127]	
	c.603_605delTGA	p.Glu202del		germline	BSF	Aljishi et al. (2010) [152]	Julie Honeychurch / Maggie Williams. Bristol Genetics Laboratory

4.5. Tafazzin pathogenic protein changes

In order to understand the impact of these changes on tafazzin protein domains, we conducted a mapping of all pathogenic tafazzin protein changes to the established tafazzin domains (Figure 7). It is worth mentioning that certain protein changes were associated with multiple gene variants, resulting in a smaller number of unique protein changes compared to the total number od gene variants. The distribution of these protein changes across individual domains is visualized in a pie chart (Figure 6).

The TM domain exhibits 5 protein change variants, while the AT domain contains 52 variants. In the HX4D domain, 5 variants are present, followed by 10 variants in the MT1 domain, 8 variants in the MA domain, and 23 variants in the MT2 domain. 31 variants are located in non-assigned parts of the sequence. It is noteworthy that the number of variants in the HX4D and MT1 domains is relatively significant, considering the smaller size of these domains, which indicates mutational enrichment.

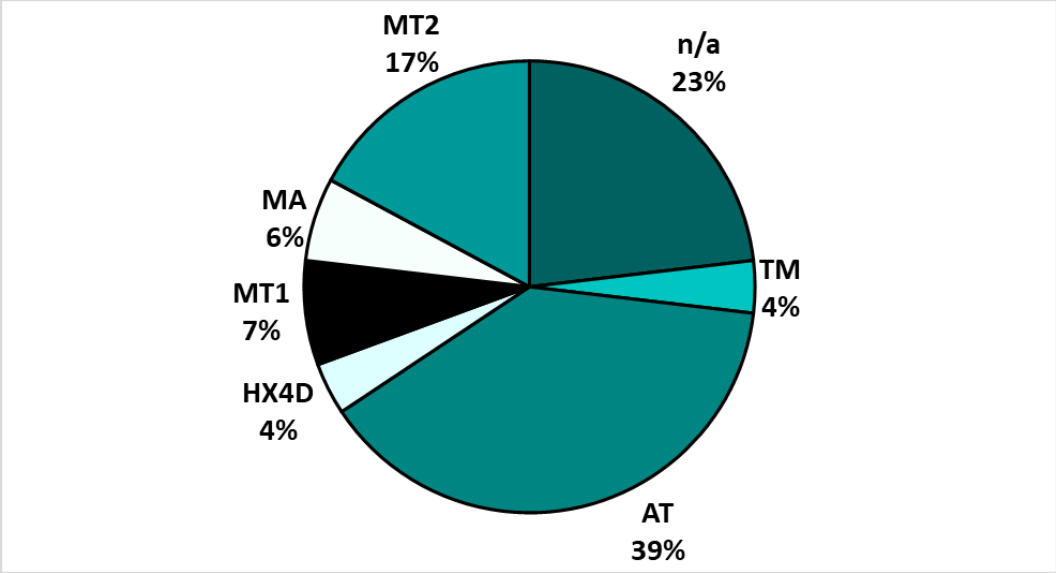

Figure 6. Distribution of tafazzin variants in Barth snydrome across protein domains. The distribution is represented as percentage in the pie chart, reflecting the exact share score of variants in each domain.

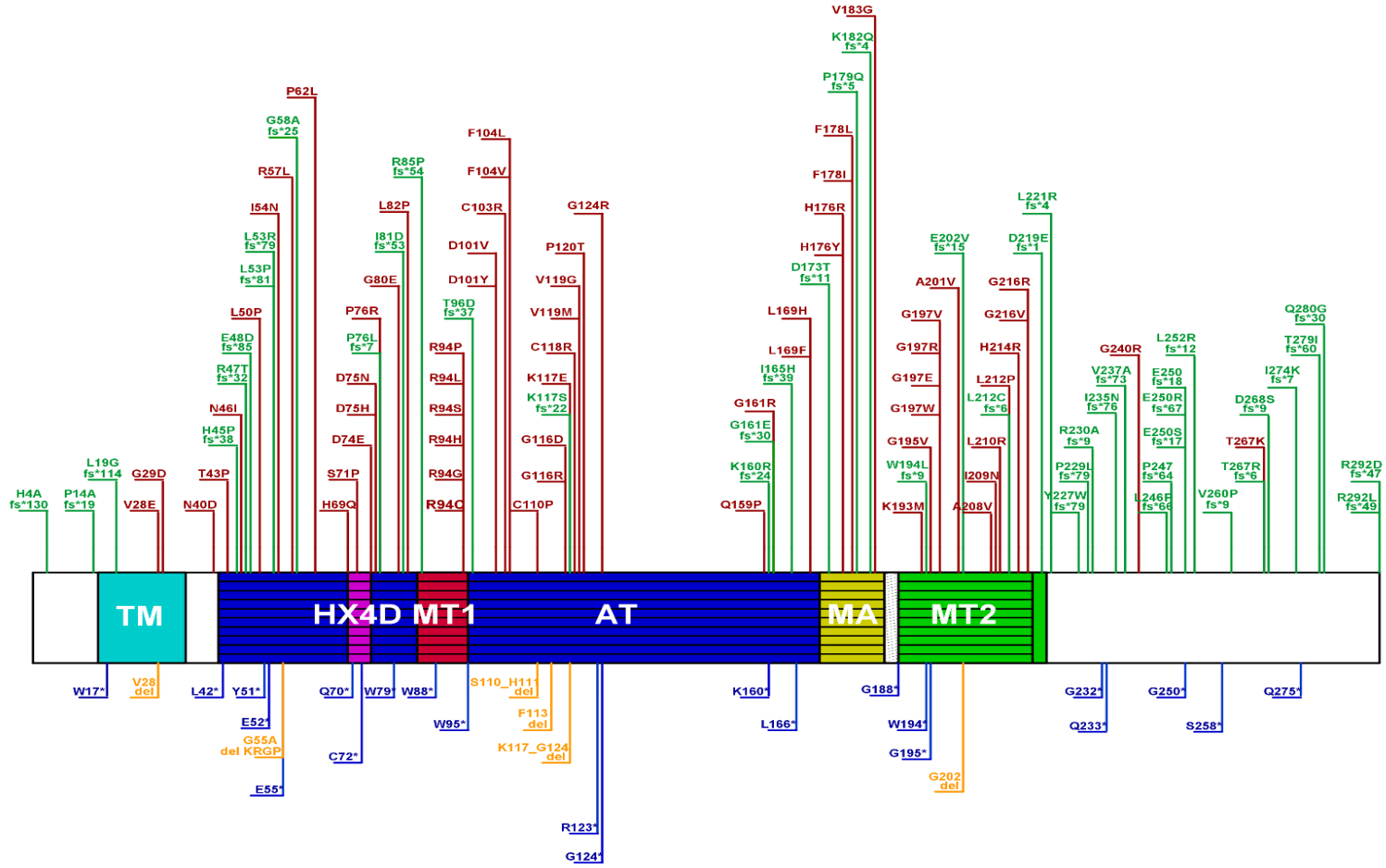

Figure 7. Tafazzin variants in BTHS mapped on protein domains. The protein changes were mapped on appropriate tafazzin protein domains. Red: missense; green: frameshift; blue: nonsense; orange: deletions.

4.6. Enrichment of tafazzin pathogenic protein changes

To analyze significant variants based on the domain they occupy, we further analyzed their distribution relative to the size of each protein domain, following the protocol described in section 3.6 of Materials and Methods.

The HX4D and MT1 domains exhibit the highest enrichment, both with a score of 1.82, confirming our previous hypothesis. Another domain showing enrichment is the MT2 domain with the enrichment score of 1.52. The AT and MA domains show a slightly less significant enrichment with scores 1.01 and 1.16, respectively. The TM domain and non-assigned regions are not enriched in pathogenic protein change variants. (Table 5).

Table 5. Enrichment of tafazzin variants in BTHS. Enrichment calculations were performed for each domain individually, with enrichment values below 1 indicating lower than expected occurrence, while values above 1 indicate enrichment.

	TOTAL NUMBER OF VARIANTS IN DOMAIN (134)	% (OBSERVED FRACTION OF THE VARIANTS)	% (PROTEIN FRACTION OCCUPIED BY DOMAIN – EXPECTED FRACTION)	DIFFERENCE	ENRICHMENT VALUE (OBSERVED/EXPECTED)
n/a	31	23.13%	32.19%	**9.06% (-)**	**0.72**
TM	5	3.73%	6.85%	**3.12% (-)**	**0.54**
AT	52	38.81%	38.36%	**0.45% (+)**	**1.01**
HX4D	5	3.73%	2.05%	**1.68% (+)**	**1.82**
MT1	10	7.46%	4.11%	**3.35% (+)**	**1.82**
MA	8	5.97%	5.14%	**0.83% (+)**	**1.16**
MT2	23	17.16%	11.30%	**5.86% (+)**	**1.52**
	Σ	100.00%	100.00%		

4.7. *TAZ* naturally occurring (benign/likely benign) variants

Different *TAZ* variants have also been observed in healthy individuals, indicating the presence of naturally occurring or bening/likely benign variants. To analyze these variants and compare their occurrence with the occurrence of pathogenic variants, we used the BSF database and the gnomAD database to compile a comprehensive list of the variants and mapped the corresponding protein changes to tafazzin protein domains.

We documented each variant in the curated list with its corresponding nucleotide change, protein change, and database source. For variants covered by gnomAD, we included a corresponding categorization.

The final list consists of a total of 92 mutation variants, with all of them encompassed within the BSF database. Among them, 27 variants are covered by gnomAD, representing a coverage of approximately 29%. Roughly 61% of the variants represent synonymous mutations, suggesting that they are indeed unlikely to have an impact on the phenotype. The remaining variants are missense mutations that result in an amino acid substitution. (Table 6).

Table 6. List of naturally occurring (benign/likely benign) *TAZ* variants. The variant analysis was conducted through database research. The variants are listed in order of amino acid sequence, along with the corresponding nucleotide change, protein change, database coverage and category.

NUCLEOTIDE CHANGE	PROTEIN CHANGE	DATABASE COVERAGE	CATEGORY
c.1A>T	p.Met1Leu	BSF	
c.2T>G	p.Met1Arg	BSF	
c.12C>A	p.His4Gln	BSF	
c.18G>T	p.Lys6Asn	BSF	
c.17_18insA	p.Trp7Val	BSF	
c.20G>C	p.Trp7Ser	BSF	
c.21G>T	p.Trp7Cys	BSF	
c.22C>T	p.Pro8Ser	BSF	
c.28C>T	p.Pro10Ser	BSF	

c.29C>G	p.Pro10Arg	BSF	
c.32C>T	p.Ala11Val	BSF	
c.33G>C	p.Ala11=	BSF	
c.45C>T	p.Leu15=	BSF	
c.48C>G	p.Thr16=	BSF, gnomAD	Likely benign
c.54C>T	p.Thr18=	BSF	
c.62G>T	p.Ser21Ile	BSF	
c.69C>T	p.Val23=	BSF, gnomAD	Likely benign
c.79T>C	p.Leu27=	BSF, gnomAD	Likely benign
c.90C>G	p.Thr30=	BSF	
c.102C>G	p.Phe34Leu	BSF	
c.108C>T	p.Thr36=	BSF	
c.123C>T	p.His41=	BSF, gnomAD	Likely benign
c.147G>A	p.Val49=	BSF	
C>G	p.Pro62=	BSF, gnomAD	Likely benign
c.198G>A	p.Val66=	BSF, gnomAD	Likely benign
c.216C>T	p.Cys72=	BSF	
c.231T>C	p.His77=	BSF	
c.240G>T	p.Gly80=	BSF	
c.240G>A	p.Gly80=	BSF	
c.250C>T	p.Leu84Phe	BSF	
c.255C>T	p.Arg85=	BSF	
c.267C>T	p.Asn89=	BSF, gnomAD	Likely benign
c.270G>A	p.Leu90=	BSF	
c.321G>A	p.Glu107=	BSF	
c.331C>T	p.His111Tyr	BSF	
c.351G>A	p.Lys117=	BSF	
c.371G>A	p.Gly124Glu	BSF	
c.383T>C	p.Phe128Ser	BSF, gnomAD	Benign/Likely benign
c.403A>G	p.Lys135Glu	BSF	
c.405A>G	p.Lys135=	BSF	
c.407G>T	p.Gly136Val	BSF	
c.419C>T	p.Thr140Ile	BSF	
c.436G>C	p.Gly146Arg	BSF	
c.436G>A	p.Gly146Arg	BSF	
c.437G>T	p.Gly146Val	BSF	
c.442G>A	p.Gly148Arg	BSF	
c.456G>C	p.Glu152Asp	BSF	
c.468C>T	p.Gly156=	BSF, gnomAD	Benign/Likely benign
c.469G>A	p.Val157Ile	BSF	
c.480G>A	p.Lys160=	BSF	

c.497T>C	p.Leu166=	BSF	
c.504G>A	p.Lys168=	BSF	
c.510C>T	p.Asn170=	BSF	
c.534C>T	p.Phe178=	BSF, gnomAD	Likely benign
c.543G>A	p.Gly181=	BSF, gnomAD	Likely benign
c.557G>A	p.Ser186Asn	BSF	
c.561 C>T	p.Ser187=	BSF	
c.567C>T	p.Phe189=	BSF, gnomAD	Likely benign
c.588C>T	p.Ile196=	BSF	
c.591G>T	p.Gly197=	BSF, gnomAD	Likely benign
c.591G>A	p.Gly197=	BSF, gnomAD	Likely benign
c.594C>A	p.Arg198=	BSF	
c.494C>T	p.Arg198=	BSF	
c.634C>T	p.Leu212=	BSF, gnomAD	Benign/Likely benign
c.645C>T	p.Val215=	BSF	
c.657C>T	p.Asp.219=	BSF, gnomAD	Likely benign
c.658G>C	p.Val220Leu	BSF	
c.658G>A	p.Val220Ile	BSF	
c.660C>T	p.Val220=	BSF, gnomAD	Likely benign
c.670C>T	p.Asn223=	BSF	
c.675G>A	p.Pro225=	BSF, gnomAD	Benign/Likely benign
c.682T>C	p.Phe228Leu	BSF	
c.690C>A	p.Arg230=	BSF, gnomAD	Likely benign
c.696A>G	p.Gly232=	BSF	
c.715A>G	p.Ile239Val	BSF	
c.717C>T	p.Ile239=	BSF	
c.738G>A	p.Leu246=	BSF	
c.747C>G	p.Leu249=	BSF, gnomAD	Benign/Likely benign
c.747C>T	p.Leu249=	BSF, gnomAD	Benign/Likely benign
c.748G>A	p.Glu250Lys	BSF	
c.762G>A	p.Ala254=	BSF, gnomAD	Likely benign
c.771G>A	p.Lys257=	BSF, gnomAD	Likely benign
c.774G>A	p.Ser258=	BSF, gnomAD	Likely benign
c.791A>G	p.Lys264Arg	BSF	
c.801G>A	p.Thr267=	BSF, gnomAD	Likely benign
c.807C>A	p.Phe269Leu	BSF	
c.816G>A	p.Glu272=	BSF	
c.846G>A	p.Glu282=	BSF, gnomAD	Likely benign
c.855C>T	p.His285=	BSF	

c.861C>A	p.His287Gln	BSF	
c.873G>A	p.Gly291=	BSF, gnomAD	Benign/Likely benign
c.874A>G	p.Arg292Gly	BSF	

4.8. Tafazzin benign/likely benign protein changes

To assess the occurrence of benign/likely benign protein changes in context of protein domains, we performed a mapping of all benign/likely benign tafazzin protein changes to the established domains (Figure 9). As previously, some protein changes were linked to multiple gene variants, resulting in a smaller number of unique protein changes compared to gene variants. The distribution of these protein changes across domains is depicted in a pie chart (Figure 8).

The non-assigned parts of the sequence indicate enrichment with 41 protein change variants. The TM domain shows 7, whereas the AT domain has 24 variants. The HX4D domain has 1 variant, followed by 4 variants in the MT1 domain, 2 variants in the MA domain, and 7 variants in the MT2 domain.

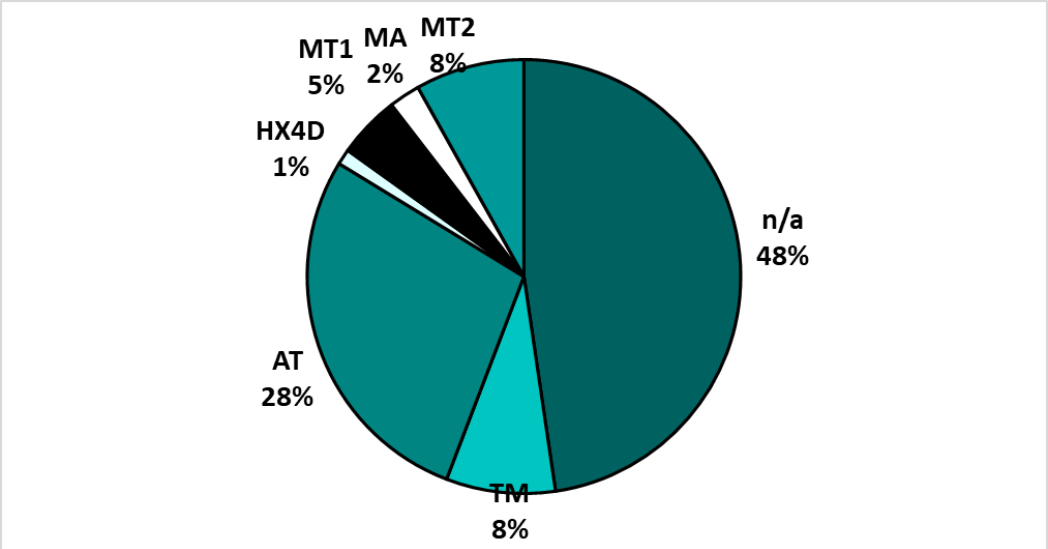

Figure 8. Distribution of tafazzin variants in Barth snydrome across protein domains. The distribution is represented as percentage in the pie chart, reflecting the exact share score of variants in each domain.

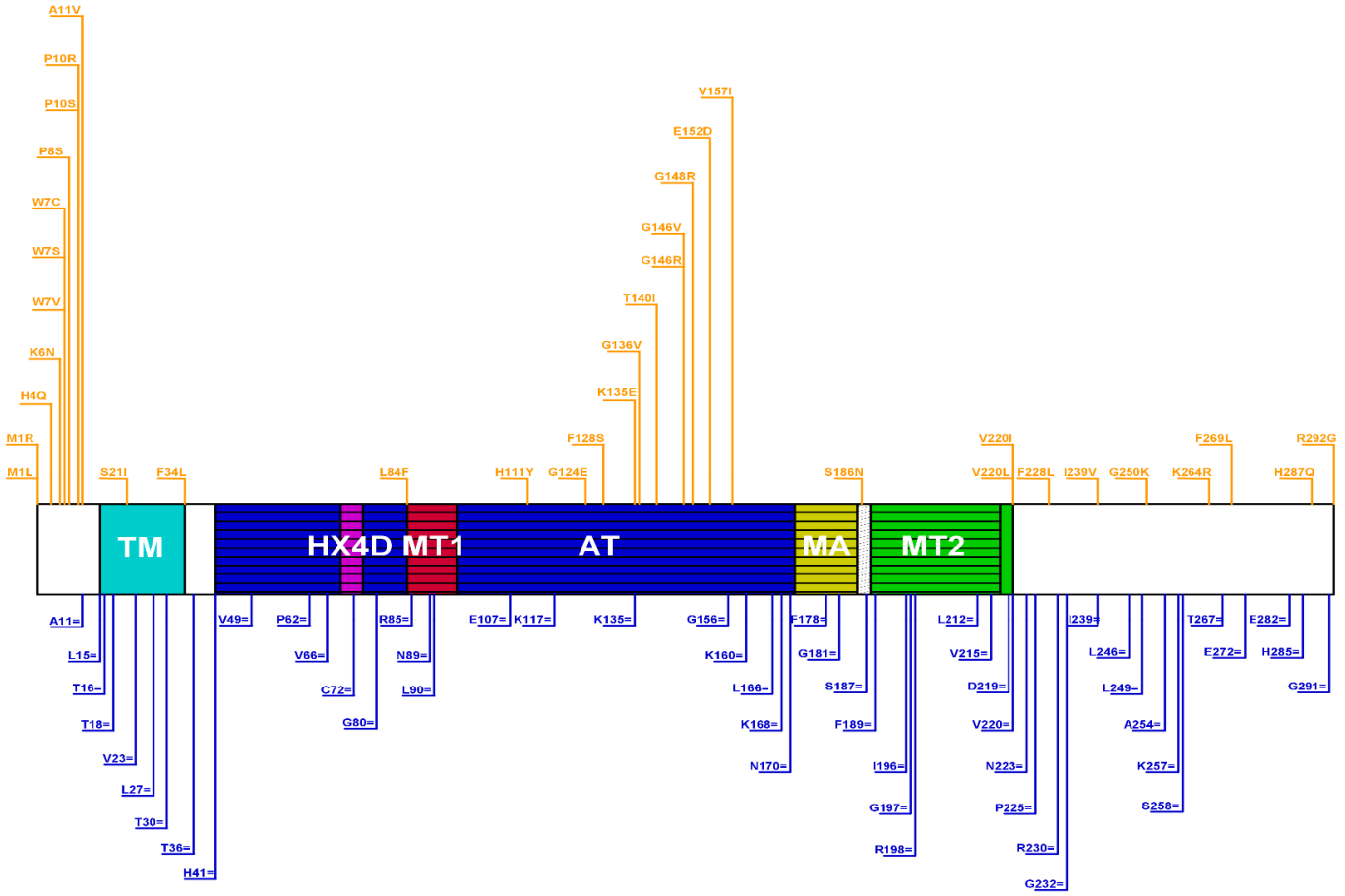

Figure 9. Tafazzin benign/likely benign variants mapped on protein domains. The protein changes were mapped on appropriate tafazzin protein domains. Orange: missense; blue: synonymous.

4.9. Enrichment of tafazzin benign/likely benign protein changes

For domain-specific analysis of benign/likely benign tafazzin variants, we used the statistical method for enrichment calculation from section 3.6 in Materials and Methods. The TM domain and non-assigned regions of the protein show the highest enrichment with scores of 1.89 and 1.48, respectively. The MT1 domain also exhibits enrichment, albeit less significant, with a score of 1.13. Domains AT, HX4D, MA and MT2 do not show enrichment in protein change variants. (Table 7).

The observed pattern shows that domains enriched in pathogenic variants do not display enrichment in benign variants and vice versa, with an exception in the MT1 domain, which shows enrichment in both pathogenic and benign variants. Nevertheless, it is important to note that three out of the four benign protein changes in this domain are synonymous, similar to over half of the benign protein changes found in other domains. Therefore, further investigation on the level of RNA is needed to understand their significance.

Table 7. Enrichment of benign/likely benign tafazzin variants. Enrichment values below 1 indicate lower than expected occurrence, while values above 1 indicate enrichment.

	TOTAL NUMBER OF VARIANTS IN DOMAIN (86)	% (OBSERVED FRACTION OF THE VARIANTS)	% (PROTEIN FRACTION OCCUPIED BY DOMAIN – EXPECTED FRACTION)	DIFFERENCE	ENRICHMENT VALUE (OBSERVED/EXPECTED)
n/a	41	47.67%	32.19%	**15.48% (+)**	**1.48**
TM	7	8.14%	6.85%	**1.29% (+)**	**1.89**
AT	24	27.91%	38.36%	**10.45% (-)**	**0.73**
HX4D	1	1.16%	2.05%	**0.89% (-)**	**0.57**
MT1	4	4.65%	4.11%	**0.54% (+)**	**1.13**
MA	2	2.33%	5.14%	**2.81% (-)**	**0.45**
MT2	7	8.14%	11.30%	**3.16% (-)**	**0.72**
	Σ	100.00%	100.00%		

4.10. Tafazzin has 34 predicted-pathogenic missense variants

In order to further validate and gain a more comprehensive understanding of tafazzin protein changes related to BTHS, we examined their level of pathogenicity through seven different *in silico* pathogenicity prediction programs. Given that the programs allow analysis exclusively for missense variants, our focus was limited to this variant type. In total, we analyzed 62 tafazzin protein changes caused by missense variants.

The classification of pathogenicity levels among the seven programs varies, with some representing three, and some two categories. Their respective categorizations are as follows: SIFT – „affects protein function"/„tolerated"; PolyPhen2 – „probably damaging"/„possibly damaging"/„benign"; Panther-PSEP – „probably damaging"/„possibly damaging"/„probably benign"; Meta-SNP – „disease"/„neutral"; SNAP – „disease"/„neutral"; INPS-MD – „destabilizing"/„stabilizing"; FATHMM –„damaging"/ „benign". The results of each program are shown in (Table 8).

To identify missense variants that are strongly considered pathogenic, we focused on those predicted as pathogenic by 100% of the programs. For this assessment, we considered the cutoff values and corresponding categorizations of each program (SIFT – <0.05/„affects protein function"; PolyPhen-2 – ≥0.85/„probably damaging"; Panther-PSEP – >10/„probably damaging"; Meta-SNP – >0.5/„disease"; SNAP – <0/„disease"; INPS-MD – „destabilizing"; FATHMM – >0/„damaging").

In total, 34 of the missense variants show a high likelihood of pathogenicity in 100% of the pathogenicity prediction programs, exhibiting 55% of consistency.

Table 8. Results of *in silico* pathogenicity prediction of tafazzin protein changes caused by missense variants. The pathogenicity levels are categorized according to cutoff values of each program, with an exception of INPS-MD which utilizes comparison methods for automatic sorting.

PROTEIN CHANGE	SIFT	PolyPhen-2	Panther-PSEP	Meta-SNP	SNAP	INPS-MD	FATHMM
p.Val28Glu	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly29Asp	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.Asn40Asp	AFFECTS PROTEIN FUNCTION	POSSIBLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Thr43Pro	TOLERATED	PROBABLY DAMAGING	PROBABLY BENIGN	DISEASE	NEUTRAL	DESTABILIZING	DAMAGING
p.Asn46Ile	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	NEUTRAL	DESTABILIZING	DAMAGING
p.Leu50Pro	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Ile54Asn	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	POSSIBLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.Arg57Leu	TOLERATED	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
c.Pro62Leu	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.His69Gln	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Ser71Pro	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Asp74Glu	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Asp75His	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Asp75Asn	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Pro76Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly80Glu	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Leu82Pro	AFFECTS PROTEIN FUNCTION	POSSIBLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING

p.Arg94Cys	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Arg94Gly	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Arg94Ser	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Arg94His	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Arg94Leu	AFFECTS PROTEIN FUNCTION	POSSIBLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.Arg94Pro	AFFECTS PROTEIN FUNCTION	BENIGN	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Asp101Tyr	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Asp101Val	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Cys103Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Phe104Val	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Phe104Leu	AFFECTS PROTEIN FUNCTION	POSSIBLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Ser110Pro	AFFECTS PROTEIN FUNCTION	BENIGN	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly116Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly116Asp	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Lys117Glu	TOLERATED	PROBABLY DAMAGING	POSSIBLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Cys118Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Val119Met	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	NEUTRAL	DISEASE	DESTABILIZING	DAMAGING
p.Val119Gly	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Pro120Thr	TOLERATED	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly124Arg	AFFECTS PROTEIN FUNCTION	POSSIBLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.Gln159Pro	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly161Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.Leu169Phe	TOLERATED	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Leu169His	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING

p.His176Tyr	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.His176Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.Phe178Ile	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Phe178Leu	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Val183Gly	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Lys193Met	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly195Val	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly197Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly197Trp	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.Gly197Val	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	STABILIZING	DAMAGING
p.Gly197Glu	TOLERATED	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Ala201Val	TOLERATED	PROBABLY DAMAGING	PROBABLY DAMAGING	NEUTRAL	NEUTRAL	STABILIZING	DAMAGING
p.Ile208Val	TOLERATED	BENIGN	PROBABLY BENIGN	NEUTRAL	NEUTRAL	DESTABILIZING	DAMAGING
p.Ile209Asn	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Leu210Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	POSSIBLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Leu212Pro	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.His214Arg	TOLERATED	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly216Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly216Val	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Gly240Arg	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	DISEASE	DISEASE	DESTABILIZING	DAMAGING
p.Thr267Lys	AFFECTS PROTEIN FUNCTION	PROBABLY DAMAGING	PROBABLY DAMAGING	NEUTRAL	NEUTRAL	DESTABILIZING	DAMAGING

As the alteration in MLCL/CL ratio is a significant indicator of pathogenicity in BTHS, we analyzed the presence of documented MLCL/CL changes. Out of 28 missense variants that were reported to have defective MLCL/CL profiles, 16 variants can be found in the list of variants that are predicted pathogenic in all programs, indicating a consistency of approximately 57%. (Table 9). For the remaining 18 pathogenic predicted variants, no MLCL/CL changes have been reported. However, the absence of documented changes could be attributed to the fact that MLCL and CL levels might not have been measured in patients with these specific variants.

Table 9. Pathogenic predicted protein changes with corresponding MLCL/CL effect. Only protein changes that were predicted pathogenic by all 7 pathogenicity programs are included. The consistency between prediction programs and so far reported variants with an aberrant MLCL/CL ratio is roughly 57%.

PROTEIN CHANGE	MLCL/CL EFFECT
V28E	not reported
L50P	MLCL↑/CL↓
H69Q	MLCL↑/CL↓
S71P	not reported
D74E	not reported
D75H	MLCL/CL = 28.04
D75N	not reported
P76R	not reported
G80E	not reported
R94C	MLCL↑/CL↓
R94G	MLCL↑/CL↓
R94S	MLCL↑/CL↓
R94H	MLCL↑/CL↓
D101Y	CL↓
D101V	not reported
C103R	not reported
F104V	not reported
G116R	MLCL↑/CL↓
G116D	not reported

C118R	CL↓
V119G	MLCL↑/CL↓
Q159P	not reported
L169H	not reported
F178I	not reported
F178L	not reported
V183G	not reported
K193M	not reported
G195V	not reported
G197R	MLCL↑/CL↓
I209N	MLCL/CL = 18.85
L212P	CL↓
G216R	MLCL/CL = 16
G216V	not reported
G240R	CL↓

4.11. Enrichment of tafazzin predicted-pathogenic protein changes

To further refine our enrichment analysis, we performed the calculations described in section 3.6 of Materials and methods, exclusively with the subset of 34 missense variants that were predicted as pathogenic by all prediction programs.

The results align with the enrichment analysis of pathogenic variants conducted earlier. The TM domain and non-assigned regions still exhibit no enrichment, while the AT, HX4D, MT1, MA, and MT2 domains display enrichment in variants. The process of refining variants using pathogenicity prediction programs before conducting this enrichment analysis allowed us to pinpoint two specific domains with the highest enrichment: the HX4D domain with a score of 4.30, and the MT1 domain with a score of 2.86.

Table 10. Enrichment of tafazzin variants in BTHS after refinement through pathogenicity prediction. Enrichment values below 1 indicate lower than expected occurrence, while values above 1 indicate enrichment.

	TOTAL NUMBER OF VARIANTS IN DOMAIN (34)	% (OBSERVED FRACTION OF THE VARIANTS)	% (PROTEIN FRACTION OCCUPIED BY DOMAIN – EXPECTED FRACTION)	DIFFERENCE	ENRICHMENT VALUE (OBSERVED/EXPECTED)
n/a	1	2.94%	32.19%	**29.25% (-)**	**0.10**
TM	1	2.94%	6.85%	**3.91% (-)**	**0.43**
AT	15	44.12%	38.36%	**5.76% (+)**	**1.15**
HX4D	3	8.82%	2.05%	**6.77% (+)**	**4.30**
MT1	4	11.76%	4.11%	**7.65% (+)**	**2.86**
MA	3	8.82%	5.14%	**3.68% (+)**	**1.72**
MT2	7	20.59%	11.30%	**9.29% (+)**	**1.82**
	Σ	100.00%	100.00%		

4.12. Tafazzin differs across species

To highlight functionally crucial tafazzin variants, our next objective was to investigate the similarity and conservation of the tafazzin protein sequence across species at various evolutionary levels. For enhanced precision, we exclusively used Swiss-Prot protein sequences from UniProt. In total, we imported 11 sequences into the Clustal Omega program, belonging to baker's yeast (TAZ_YEAST), fruit fly (TAZ_DROME), zebrafish (TAZ_DANRE), bornean orangutan (TAZ_PONPY), human (TAZ_HUMAN), chimpanzee (TAZ_PANTR), western lowland gorilla (TAZ_GORGO), mouse (TAZ_MOUSE), rhesus macaque (TAZ_MACMU), red guenon (TAZ_ERYPA), and common squirrel monkey (TAZ_SAISC). (Figure 10).

Tafazzin displays notable variations across these species. Human tafazzin exhibits the highest degree of similarity with the western lowland gorilla, and the lowest degree of similarity with yeast. (Table 11).

Figure 10. Sequence alignment of tafazzin from different species. The Swiss-Prot sequences were taken from UniProt and imported to Clustal Omega. Asterisks (*) – positions with a fully reserved residue; colons (:) – conservation between groups of strongly similar properties; periods (.) – conservation between groups of weakly similar properties. Supplementary information regarding arrows can be found in section 4.13. of Results.

Table 11. Pairwise identity scores matrix. Human tafazzin shares the most similarity with tafazzin from the western lowland gorilla. All species are represented with a % similarity score between each other.

SPECIES	Baker's yeast	Fruit fly	Zebrafish	Bornean orangutan	Human	Chimpanzee	Western lowland gorilla	Mouse	Rhesus macaque	Red guenon	Common squirrel monkey
Baker's yeast	100.00	29.73	26.69	28.97	28.97	28.97	28.97	28.69	28.69	28.69	28.69
Fruit fly	29.73	100.00	44.40	45.56	46.33	45.56	45.56	45.35	45.35	45.35	45.35
Zebrafish	26.69	44.40	100.00	67.18	67.18	67.18	67.18	67.05	67.05	67.05	67.05
Bornean orangutan	28.97	45.56	67.18	100.00	98.29	98.97	99.32	98.09	99.24	99.62	99.62
Human	28.97	46.33	67.18	98.29	100.00	99.32	98.97	97.33	98.47	98.85	98.85
Chimpanzee	28.97	45.56	67.18	98.97	99.32	100.00	99.66	98.09	99.24	99.62	99.62
Western lowland gorilla	28.97	45.56	67.18	99.32	98.97	99.66	100.00	98.47	99.62	100.00	100.00
Mouse	28.69	45.35	67.05	98.09	97.33	98.09	98.47	100.00	98.09	98.47	98.47
Rhesus macaque	28.69	45.35	67.05	99.24	98.47	99.24	99.62	98.09	100.00	99.62	99.62
Red guenon	28.69	45.35	67.05	99.62	98.85	99.62	100.00	98.47	99.62	100.00	100.00
Common squirrel monkey	28.69	45.35	67.05	99.62	98.85	99.62	100.00	98.47	99.62	100.00	100.00

To illustrate evolutionary relationships among these species, we also generated a phylogenetic tree of similarity between their tafazzin sequences (Figure 11).

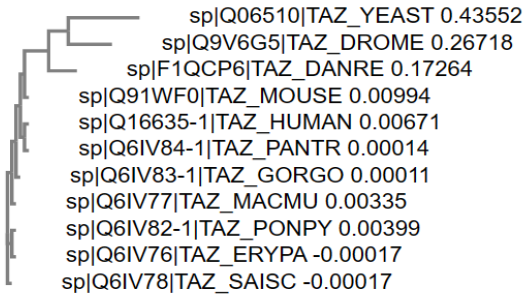

Figure 11. Phylogenetic tree of similarity between species. Species that are closer to each other and depicted by narrower lines indicate a higher degree of evolutionary similarity. Most of the species are closely related, with humans having the strongest similarity with the chimpanzee and western lowland gorilla.

4.13. Impact of tafazzin conservation on mutation occurrence

In light of our goal to highlight crucial tafazzin variants based on the significance of evolutionary similarity in conserved protein regions, we used the results from the analysis of tafazzin missense variants in BTHS, focusing on the variants predicted as pathogenic by all programs. Specifically, we examined which of these damaging variants fell within the conserved regions of the human tafazzin protein, using the aligned sequence presented in (Figure 10).

The analysis reveals that approximately 65% of the variants are mapped to fully conserved regions of the protein. Furthermore, the majority of the remaining variants show conservation across all species except for yeast. With removal of yeast from the equation, a striking 91% of the variants are found to be fully conserved in the remaining examined species.

The pathogenic variants within the fully conserved regions, including yeast, are: L50P, H69Q, S71P, D74E, D75H, D75N, P76R, R94C, R94G, R94S, R94H, C103R, F104V, G116R, G116D, Q159P, V183G, K193M, G195V, G216R, G216V, and G240R.

4.14. Enrichment of tafazzin predicted-pathogenic protein changes on evolutionarily conserved regions

Lastly, we again used the statistical method for enrichment calculation, as described in section 3.6 of Materials and Methods, to examine the enrichment patterns across domains following the additional refinement through the selection of variants located in evolutionarily conserved regions of tafazzin.

As observed previously, the TM domain and non-assigned regions still do not exhibit any enrichment. The AT and MT2 domains demonstrate a relatively modest enrichment, with scores of 1.07 and 1.60, respectively. Notably, the MA domain, which previously showed enrichment, now presents no enrichment. On the other hand, domains HX4D and MT1 are further

emphasized with significantly higher enrichment scores of 6.65 and 4.42, respectively, indicating their critical functional importance of harboring pathogenic variants. (Table 12).

Table 12. Enrichment of tafazzin variants in BTHS after refinement through analysis of evolutionarily conserved regions. Enrichment values below 1 indicate lower than expected occurrence, while values above 1 indicate enrichment.

	TOTAL NUMBER OF VARIANTS IN DOMAIN (22)	% (OBSERVED FRACTION OF THE VARIANTS)	% (PROTEIN FRACTION OCCUPIED BY DOMAIN – EXPECTED FRACTION)	DIFFERENCE	ENRICHMENT VALUE (OBSERVED/EXPECTED)
n/a	1	4.55%	32.19%	**27.64% (-)**	**0.14**
TM	0	0.00%	6.85%	**6.85% (-)**	n/a
AT	9	40.91%	38.36%	**2.55% (+)**	**1.07**
HX4D	3	13.64%	2.05%	**11.59% (+)**	**6.65**
MT1	4	18.18%	4.11%	**14.07% (+)**	**4.42**
MA	1	4.55%	5.14%	**0.59% (-)**	**0.89**
MT2	4	18.18%	11.30%	**6.88% (+)**	**1.60**
	Σ	100.00%	100.00%		

5. Discussion

In light of the limited information available on tafazzin domains and *TAZ* variants in context of BTHS, we have developed a comprehensive *in silico* approach to provide insights into the pathogenicity of *TAZ* variants and how they functionally correlate with protein domains.

Our results combined from databases and literature show that tafazzin is comprised of six domains: TM, AT, HX4D, MT1, MA, and MT2. However, it is important to highlight that the current information on tafazzin domains available in databases UniProt and InterPro partially aligns with the domains we curated from the literature. These databases lack a significant portion of the information we compiled in our study. Our curated data provides a more detailed understanding of these protein domains, enabling us to gain a deeper insight into their functional characteristics. The six domains we identified make up approximately 68% of the protein, while the rest of the regions remain non-assigned. The highest amount of overlap between domains can be found inside the AT domain, the largest domain in the protein that contains an active acyltransferase segment responsible for the transfer of acyl side chains in CL remodeling – the process that is defective in BTHS. While we primarily focused on analyzing each domain separately, it is important to consider that changes occurring within domains located inside the AT domain can also lead to defects in the functionality of the AT domain itself, rather than solely affecting the other domains. Therefore, alterations in the AT domain may have broader implications for the overall function and integrity of the protein.

The 3D structure analysis of our study suggests that tafazzin displays a predominantly globular structure, which implies that it may have fewer interacting partners, as globular proteins typically have fewer accessible interaction sites. Additionally, the lack of significant disordered regions in tafazzin further supports the notion that tafazzin likely functions through conformational selection, rather than induced-fit interactions typically

associated with linear proteins. Past studies have, however, suggested that CL remodeling is not induced via acyl substrate specificity and that tafazzin is a relatively non-specific transacylase with the ability to react with different phospholipids [6,153]. Taking this into consideration, it is possible that tafazzin may have conformational flexibility, allowing it to interact with various phospholipids without undergoing major conformational changes induced by ligand binding. The microenvironment, such as membrane composition and surroudning molecules, could be the factor influencing the existing conformational equilibrium of the protein. It has indeed been suggested that the specificity for acyl groups arises from self-organization, resulting from specific interactions among individual molecules within the membrane, implying that point mutations in tafazzin could potentially alter the remodeling specificity by causing tafazzin to mislocalize [6,15]. As MT1 and MT2 domains responsible for mitochondrial localization are within the AT domain, their interconnectedness in context of point mutations presents an interesting research objective.

It is noteworthy that in our study, the 3D model of tafazzin is derived from a sequence lacking exon 5, and exhibits specific regions corresponding to a segment of the AT domain and non-assigned areas near the end of the sequence in low to very low confidence scores. Intriguingly, our disordered regions analysis results show that segments from exon 5 and non-assigned areas near the end of the sequence are predicted to be disordered. This indicates potential areas of linear structure where interactions with other proteins occur, providing potential insights into additional cellular functions of tafazzin. Notably, previous research in yeast has suggested that another enzyme, Ygr110wp, genetically and biochemically interacts with tafazzin in the pathway of CL remodeling [60]. Other proteins could not only have an impact on the previously discussed acyl specificity of the process, but could also shed

light on the correlation between tafazzin's role in BTHS, apoptosis inhibition and promotion of tumorigenicity.

The Human Tafazzin Gene Variants Database from the BSF serves as a reliable framework for investigating *TAZ* mutations and tafazzin protein changes. By combining data from this database with information from UniProt, ClinVar and OMIM, we present a total of 139 pathogenic *TAZ* variants. It is important to note that the coverage of variants in UniProt, ClinVar, and OMIM is relatively small compared to the BSF database. Among the 139 pathogenic variants, approximately 45% exhibit an aberrant MLCL/CL profile. Even though it is worth considering that this percentage could be influenced by some patients not having their MLCL/CL profile analyzed, we focused on a current refinement of variants until more complete data becomes available.

Through mapping pathogenic tafazzin protein changes onto its domains and presenting a statistical enrichment calculation method, we show that the HX4D and MT1 domains demonstrate significant enrichment in protein change variants, with scores of 1.82 and 1.52, respectively. The AT and MA domains also exhibit some enrichment, scoring 1.01 and 1.16, respectively. Here, it is also important to take into account that both the HX4D and MT1 domains are part of the AT domain, so even though the AT domain has a modest enrichment, the enrichment we suggest being present in the HX4D and MT1 domains is caused by mutations that affect the AT domain as well, and should in theory contribute to the overall enrichment observed in the AT domain.

Given that various *TAZ* variants have also been observed in healthy individuals, we also present a compiled list of benign/likely benign tafazzin variants. Once again, the BSF database serves as a framework, as it includes information on benign/likely benign variants as well. Additionally, gnomAD also covers such variants; however, as with UniProt, OMIM, and ClinVar, the coverage in gnomAD is relatively limited compared to the data in the BSF database. Our

compilation comprises a total of 92 benign/likely benign *TAZ* mutation variants. Among these variants, approximately 61% are synonymous mutations, indicating that they are indeed less likely to exert any impact on the phenotype.

In the context of benign/likely benign variants, we demonstrate that the TM domain and non-assigned regions show the highest enrichment, with scores of 1.89 and 1.48, respectively. The MT1 domain also exhibits some enrichment, though less significant, with a score of 1.13. This pattern suggests a mutual exclusivity between domains enriched in pathogenic variants and those enriched in benign/likely benign variants, except for the MT1 domain, which shows enrichment in both categories. However, it is important to note that most benign/likely benign changes in the MT1 domain, as well as in other domains, are synonymous mutations. Therefore, further investigation at the RNA level is needed to gain a deeper understanding of their significance.

In further refinement of significant pathogenic variants, we present an *in silico* pathogenicity prediction analysis, using seven different programs. We demonstrate that 34 missense variants are consistently predicted as pathogenic by all seven programs, representing a substantial 55% consistency rate in their pathogenicity predictions. Among 28 variants with documented defective MLCL/CL profiles, 16 were predicted pathogenic by all seven programs, showing a consistency rate of approximately 57%. It is important to acknowledge that our analysis was limited to missense variants, which means the pool of refined pathogenic variants would likely be significantly larger if nonsense variants, frameshift variants, and deletions were included. However, given the tools available, we prioritized missense variants for this and further investigations.

With another enrichment calculation, this time specifically focused on the subset of 34 missense variants consistently predicted as pathogenic by all prediction programs, we demonstrate that not only the outcomes for all

domains align with our earlier enrichment analysis of pathogenic variants, but also that the HX4D and MT1 domains demonstrate even higher significance in this scenario, with enrichment scores of 4.30 and 2.86, respectively. These domains therefore emerge as highlighted candidates for further investigation, suggesting that their individual functions or their roles within the AT domain are critical for tafazzin's function, and that any missense mutations within them may lead to detrimental defects.

Our sequence alignment analysis reveals significant variations in tafazzin across species. Human tafazzin exhibits the highest similarity with tafazzin of the western lowland gorilla, while its lowest similarity is observed with yeast tafazzin. When focusing on the missense variants predicted as pathogenic by all programs, we show that approximately 65% of these variants fall within fully conserved regions of the tafazzin sequence. Additionally, the majority of the remaining variants are conserved across all species, except for yeast. However, it is crucial to consider the significance of yeast as a faithful model for BTHS, as CL synthesis and remodeling are conserved from yeast to humans, and yeast's high replication rate allows for quick observation of phenotypes through generations. The variants in regions that are also conserved in yeast might hold particular significance, suggesting a critical role that hasn't been evolutionarily changed even from the species with least sequence similarity to human. Furthermore, most of these fully conserved regions can be found within the HX4D and MT1 domains, making them even more intriguing research objectives.

Through one more statistical enrichment calculation, specifically targeting variants predicted pathogenic by all programs and located in evolutionarily conserved regions, we demonstrate a remarkable emphasis on the HX4D and MT1 domains, with enrichment scores of 6.65 and 4.42, respectively. We suggest that these findings emphasize the significance of further investigation

of these domains, with the goal to gain more understanding of tafazzin's pathogenicity mechanisms in BTHS.

The significance of the HX4D domain in catalyzing CL remodeling and the essential role of the MT1 domain in providing exclusive mitochondrial localization give rise to the question of how these functions might correlate to the phenotype of BTHS, and how mutations in these domains impact their respective functions. As mentioned in the Introduction section of this study, prior studies on tafazzin point mutations in yeast have shown that the R102C and R102S mutations, which correspond to the human R94C and R94S mutations within the MT1 domain, did not affect mitochondrial localization but resulted in the loss of transacylase activity [55]. Similarly, in human cells, the R94S mutation did not disrupt mitochondrial localization [13]. These findings suggest that the MT1 domain's role in mitochondrial targeting may remain intact despite certain mutations, while its function within the AT domain could be the primary target of these mutations. Considering the HX4D domain's role in catalyzing reactions within the active site, it is intriguing to speculate about possible interactions between residues in these two domains. Among the mutations identified as pathogenic through all our refinement methods, D74 and D75 mutations take place in the HX4D domain, while R94 mutations are found in the MT1 domain. These specific residues indeed raise the possibility of interactions between the two domains, as aspartate carries a negative charge, while arginine carries a positive charge, suggesting the potential for electrostatic interactions. Such interactions may implicate the presence of a potential active site, further contributing to the understanding of tafazzin's function and the effects of mutations in these crucial domains. The investigation of residue interactions between the HX4D and MT1 domains may hold promise in understanding a significant pathogenic mechanism of BTHS.

As tafazzin's crystal structure has recently been obtained from *Yarrowia lipolytica* [56], exciting opportunities for *in silico* analyses arise, including static

and molecular dynamics analyses. By examining the protein's globular structure, potential active sites and domain interactions can be predicted. Molecular dynamics would enable the examination of how these domains interact in time and how missense mutations influence the protein's dynamics. These simulations can be performed in various contexts, such as in soluble form, with the membrane, or with CL attached as a ligand. Leveraging tools such as the previously mentioned CHARMM-GUI platform and supercomputers could allow an efficient generation of these simulations, providing insights into tafazzin's behaviour and laying foundations for further *in vivo* research.

After conducting molecular dynamics analyses, experimental assays on the yeast *Saccharomyces cerevisiae* could present a promising first step in further research, as the taz1Δ model in yeast has already proven to be a valuable tool for studying *TAZ* mutations in BTHS, and pathogenic mutations in our domains of interest are conserved all the way to yeast. One such method involves using a plasmid construct called Taz1p-his [64], which allows for the incorporation of a His-tag into yeast tafazzin. This facilitates assays such as subcellular localization detection using a His-tag antibody. Additionally, growth essays on non-fermentable carbon sources can be conducted to investigate the impact of specific Taz1Δ point mutations on mitochondrial energy metabolism and CL composition. These experimental approaches hold potential for gaining conclusive insights into the functionality of *TAZ* mutations and their relevance to BTHS, as well as thorough examination of these two crucial domains.

6. Conclusion

Our bioinformatic approach for analyzing *TAZ* mutation variants in BTHS provides insights into the pathogenicity of *TAZ* variants and their functional correlation with tafazzin domains. By curating information from databases and literature, we attained a more detailed understanding of the protein's domains, identifying a total of six domains that collectively make up approximately 68% of the protein. The 3D structure analysis we conducted suggests that tafazzin has a predominantly globular structure with limited disordered regions, indicating potential interactions with fewer partners and a preference for acting via conformational selection. Through extensive refinement of pathogenic variants, including the utilization of pathogenicity prediction programs and analysis of tafazzin's evolutionarily conserved regions, we highlight the significance of the HX4D and MT1 domains and suggest their critical roles in tafazzin's function. Combining our findings with the knowledge acquired in previous research, we speculate that these domains might have a possibility of interacting with one another and may present a potential active site, with residues D74, D75 and R94 being the most relevant candidates to form interactions, owing to their complementary charges and the fact that their mutations have been continuously highlighted throughout various refinement steps of identifying pathogenic variants in BTHS. Therefore, understanding the effects of mutations in these domains may explain a significant pathogenic mechanism of BTHS. Overall, our study is aimed to lay a foundation for further research into the functional characteristics and pathogenicity of *TAZ* variants in BTHS. With the availability of tafazzin's crystal structure, *in silico* molecular dynamics, and experimental assays in yeast, the first steps in pursuing more advanced research can be made. The combination of computational and *in vivo* approaches opens opportunities for extensive continuation on this topic.

7. References

1 Bione S, D'Adamo P, Maestrini E, Gedeon AK, Bolhuis PA, Toniolo D. A novel X-linked gene, G4.5. is responsible for Barth syndrome. *Nat Genet* 1996; **12**: 385–389.

2 Ørstavik KH, Ørstavik RE, Naumova AK, D'Adamo P, Gedeon A, Bolhuis PA *et al.* X chromosome inactivation in carriers of Barth syndrome. *Am J Hum Genet* 1998; **63**: 1457–1463.

3 Garlid AO, Schaffer CT, Kim J, Bhatt H, Guevara-Gonzalez V, Ping P. TAZ encodes tafazzin, a transacylase essential for cardiolipin formation and central to the etiology of Barth syndrome. *Gene* 2020; **726**: 144148.

4 Schlame M, Ren M, Xu Y, Greenberg ML, Haller I. Molecular symmetry in mitochondrial cardiolipins. *Chem Phys Lipids* 2005; **138**: 38–49.

5 Schlame M. Cardiolipin remodeling and the function of tafazzin. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 2013; **1831**: 582–588.

6 Xu Y, Zhang S, Malhotra A, Edelman-Novemsky I, Ma J, Kruppa A *et al.* Characterization of tafazzin splice variants from humans and fruit flies. *Journal of Biological Chemistry* 2009; **284**: 29230–29239.

7 Vaz FM, Houtkooper RH, Valianpour F, Barth PG, Wanders RJA. Only One Splice Variant of the Human TAZ Gene Encodes a Functional Protein with a Role in Cardiolipin Metabolism. *Journal of Biological Chemistry* 2003; **278**: 43089–43094.

8 Chen M, Zhang Y, Zheng PS. Tafazzin (TAZ) promotes the tumorigenicity of cervical cancer cells and inhibits apoptosis. *PLoS One* 2017; **12**. doi:10.1371/JOURNAL.PONE.0177171.

9 Pathak S, Meng WJ, Zhang H, Gnosa S, Nandy SK, Adell G *et al.* Tafazzin protein expression is associated with tumorigenesis and radiation

response in rectal cancer: a study of Swedish clinical trial on preoperative radiotherapy. *PLoS One* 2014; **9**. doi:10.1371/JOURNAL.PONE.0098317.

10 Seneviratne AK, Xu M, Schimmer AD. Tafazzin modulates cellular phospholipid composition to regulate AML stemness. *Mol Cell Oncol* 2019; **6**. doi:10.1080/23723556.2019.1620051.

11 Li X, Wu M, An D, Yuan H, Li Z, Song Y *et al.* Suppression of Tafazzin promotes thyroid cancer apoptosis via activating the JNK signaling pathway and enhancing INF2-mediated mitochondrial fission. *J Cell Physiol* 2019; **234**: 16238–16251.

12 LI D, CHEN L, ZHANG X, WANG Y, HUANG C, LI J *et al.* miR-125a-5p reverses epithelial-mesenchymal transition and restores drug sensitivity by negatively regulating TAFAZZIN signaling in breast cancer. *Mol Med Rep* 2021; **24**. doi:10.3892/MMR.2021.12452.

13 Dinca AA, Chien WM, Chin MT. Identification of novel mitochondrial localization signals in human Tafazzin, the cause of the inherited cardiomyopathic disorder Barth syndrome. *J Mol Cell Cardiol* 2018; **114**: 83–92.

14 Debnath S, Addya S. Mis-sesnse mutations in Tafazzin (TAZ) that escort to mild clinical symptoms of Barth syndrome is owed to the minimal inhibitory effect of the mutations on the enzyme function: In-silico evidence. *Interdisciplinary Sciences: Computational Life Sciences 2015 7:1* 2014; **7**: 21–35.

15 Claypool SM, McCaffery JM, Koehler CM. Mitochondrial mislocalization and altered assembly of a cluster of Barth syndrome mutant tafazzins. *J Cell Biol* 2006; **174**: 379.

16 Ye C, Shen Z, Greenberg ML. Cardiolipin remodeling: a regulatory hub for modulating cardiolipin metabolism and function. *J Bioenerg Biomembr* 2016; **48**: 113.

17 Barth PG, Scholte HR, Berden JA, Van Der Klei-Van Moorsel JM, Luyt-Houwen IEM, Van'T Veer-Korthof ET *et al.* An X-linked mitochondrial disease affecting cardiac muscle, skeletal muscle and neutrophil leucocytes. *J Neurol Sci* 1983; **62**: 327–355.

18 Pang J, Bao Y, Mitchell-Silbaugh K, Veevers J, Fang X. Barth Syndrome Cardiomyopathy: An Update. *Genes (Basel)* 2022; **13**. doi:10.3390/GENES13040656.

19 Miller PC, Ren M, Schlame M, Toth MJ, Phoon CKL. A Bayesian Analysis to Determine the Prevalence of Barth Syndrome in the Pediatric Population. *J Pediatr* 2020; **217**: 139–144.

20 Dudek J, Maack C. Barth syndrome cardiomyopathy. *Cardiovasc Res* 2017; **113**: 399–410.

21 Clarke SLN, Bowron A, Gonzalez IL, Groves SJ, Newbury-Ecob R, Clayton N *et al.* Barth syndrome. *Orphanet J Rare Dis* 2013; **8**: 23.

22 Human Tafazzin Gene Variants Database. https://barthsyndrome.org/research/taffazindatabase.html (accessed 7 Feb2023).

23 Singh HR, Yang Z, Siddiqui S, Peña LS, Westerfield BH, Fan Y *et al.* A novel Alu-mediated Xq28 microdeletion ablates TAZ and partially deletes DNL1L in a patient with Barth syndrome. *Am J Med Genet A* 2009; **149A**: 1082–1085.

24 Gonzalez IL. Barth syndrome: TAZ gene mutations, mRNAs, and evolution. *Am J Med Genet A* 2005; **134**: 409–414.

25 Chang B, Momoi N, Shan L, Mitomo M, Aoyagi Y, Endo K *et al.* Gonadal mosaicism of a TAZ (G4.5) mutation in a Japanese family with Barth syndrome and left ventricular noncompaction. *Mol Genet Metab* 2010; **100**: 198–203.

26 Cosson L, Toutain A, Simard G, Kulik W, Matyas G, Guichet A *et al.* Barth syndrome in a female patient. *Mol Genet Metab* 2012; **106**: 115–120.

27 Acehan D, Malhotra A, Xu Y, Ren M, Stokes DL, Schlame M. Cardiolipin Affects the Supramolecular Organization of ATP Synthase in Mitochondria. *Biophys J* 2011; **100**: 2184.

28 Valianpour F, Mitsakos V, Schlemmer D, Towbin JA, Taylor JM, Ekert PG *et al.* Monolysocardiolipins accumulate in Barth syndrome but do not lead to enhanced apoptosis. *J Lipid Res* 2005; **46**: 1182–1195.

29 Schlame M, Towbin JA, Heerdt PM, Jehle R, DiMauro S, Blanck TJJ. Deficiency of tetralinoleoyl-cardiolipin in Barth syndrome. *Ann Neurol* 2002; **51**: 634–637.

30 Houtkooper RH, Rodenburg RJ, Thiels C, Lenthe H van, Stet F, Poll-The BT *et al.* Cardiolipin and monolysocardiolipin analysis in fibroblasts, lymphocytes, and tissues using high-performance liquid chromatography-mass spectrometry as a diagnostic test for Barth syndrome. *Anal Biochem* 2009; **387**: 230–237.

31 Kulik W, Van Lenthe H, Stet FS, Houtkooper RH, Kemp H, Stone JE *et al.* Bloodspot assay using HPLC-tandem mass spectrometry for detection of Barth syndrome. *Clin Chem* 2008; **54**: 371–378.

32 Van Werkhoven MA, Thorburn DR, Gedeon AK, Pitt JJ. Monolysocardiolipin in cultured fibroblasts is a sensitive and specific marker for Barth Syndrome. *J Lipid Res* 2006; **47**: 2346–2351.

33 Koshkin V, Greenberg ML. Cardiolipin prevents rate-dependent uncoupling and provides osmotic stability in yeast mitochondria. *Biochemical Journal* 2002; **364**: 317.

34 Klingenberg M. Cardiolipin and mitochondrial carriers. *Biochim Biophys Acta* 2009; **1788**: 2048–2058.

35 Kiebish MA, Han X, Cheng H, Chuang JH, Seyfried TN. Cardiolipin and electron transport chain abnormalities in mouse brain tumor mitochondria: lipidomic evidence supporting the Warburg theory of cancer. *J Lipid Res* 2008; **49**: 2545–2556.

36 Saric A, Andreau K, Armand AS, Møller IM, Petit PX. Barth Syndrome: From Mitochondrial Dysfunctions Associated with Aberrant Production of Reactive Oxygen Species to Pluripotent Stem Cell Studies. *Front Genet* 2016; **6**. doi:10.3389/FGENE.2015.00359.

37 Gonzalvez F, Gottlieb E. Cardiolipin: setting the beat of apoptosis. *Apoptosis* 2007; **12**: 877–885.

38 Kagan VE, Tyurin VA, Jiang J, Tyurina YY, Ritov VB, Amoscato AA *et al.* Cytochrome c acts as a cardiolipin oxygenase required for release of proapoptotic factors. *Nat Chem Biol* 2005; **1**: 223–232.

39 Hanske J, Toffey JR, Morenz AM, Bonilla AJ, Schiavoni KH, Pletneva E V. Conformational properties of cardiolipin-bound cytochrome c. *Proc Natl Acad Sci U S A* 2012; **109**: 125–130.

40 Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* 2021; **49**: D344–D354.

41 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873* 2021; **596**: 583–589.

42 Arai M. Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophysical Reviews 2018 10:2* 2018; **10**: 163–181.

43 Koshland DE. The Key–Lock Theory and the Induced Fit Theory. *Angewandte Chemie International Edition in English* 1995; **33**: 2375–2378.

44 Mészáros B, Erdös G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* 2018; **46**: W329–W337.

45 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature 2020 581:7809* 2020; **581**: 434–443.

46 Acehan D, Vaz F, Houtkooper RH, James J, Moore V, Tokunaga C *et al.* Cardiac and Skeletal Muscle Defects in a Mouse Model of Human Barth Syndrome. *J Biol Chem* 2011; **286**: 899.

47 Soustek MS, Falk DJ, Mah CS, Toth MJ, Schlame M, Lewin AS *et al.* Characterization of a Transgenic Short Hairpin RNA-Induced Murine Model of Tafazzin Deficiency. *Hum Gene Ther* 2011; **22**: 865.

48 Damschroder D, Reynolds C, Wessells R. Drosophila tafazzin mutants have impaired exercise capacity. *Physiol Rep* 2018; **6**. doi:10.14814/PHY2.13604.

49 Choy JS, O'Toole E, Schuster BM, Crisp MJ, Karpova TS, McNally JG *et al.* Genome-wide haploinsufficiency screen reveals a novel role for γ-TuSC in spindle organization and genome stability. *Mol Biol Cell* 2013; **24**: 2753.

50 Yoshikawa K, Tanaka T, Furusawa C, Nagahisa K, Hirasawa T, Shimizu H. Comprehensive phenotypic analysis for identification of genes affecting

growth under ethanol stress in Saccharomyces cerevisiae. *FEMS Yeast Res* 2009; **9**: 32–44.

51 Claypool SM, Boontheung P, McCaffery JM, Loo JA, Koehler CM. The Cardiolipin Transacylase, Tafazzin, Associates with Two Distinct Respiratory Components Providing Insight into Barth Syndrome. *Mol Biol Cell* 2008; **19**: 5143.

52 Yadav PK, Rajasekharan R. Misregulation of a DDHD Domain-containing Lipase Causes Mitochondrial Dysfunction in Yeast. *J Biol Chem* 2016; **291**: 18562.

53 Brandner K, Mick DU, Frazier AE, Taylor RD, Meisinger C, Rehling P. Taz1, an Outer Mitochondrial Membrane Protein, Affects Stability and Assembly of Inner Membrane Protein Complexes: Implications for Barth Syndrome. *Mol Biol Cell* 2005; **16**: 5202.

54 de Tilques M de T, Lasserre JP, Godard F, Sardin E, Bouhier M, Le Guedard M *et al.* Decreasing cytosolic translation is beneficial to yeast and human Tafazzin-deficient cells. *Microbial Cell* 2018; **5**: 220.

55 Whited K, Baile MG, Currier P, Claypool SM. Seven functional classes of Barth syndrome mutation. *Hum Mol Genet* 2013; **22**: 483.

56 Schiller J, Laube E, Wittig I, Kühlbrandt W, Vonck J, Zickermann V. Insights into complex I assembly: Function of NDUFAF1 and a link with cardiolipin remodeling. *Sci Adv* 2022; **8**. doi:10.1126/SCIADV.ADD3855.

57 Ji J, Greenberg ML. Cardiolipin function in the yeast S. cerevisiae and the lessons learned for Barth syndrome. *J Inherit Metab Dis* 2022; **45**: 60.

58 Malhotra A, Edelman-Novemsky I, Xu Y, Plesken H, Ma J, Schlame M *et al.* Role of calcium-independent phospholipase A2 in the pathogenesis of Barth syndrome. *Proc Natl Acad Sci U S A* 2009; **106**: 2337–2341.

59 Mancuso DJ, Kotzbauer P, Wozniak DF, Sims HF, Jenkins CM, Guan S *et al.* Genetic ablation of calcium-independent phospholipase A2{gamma} leads to alterations in hippocampal cardiolipin content and molecular species distribution, mitochondrial degeneration, autophagy, and cognitive dysfunction. *J Biol Chem* 2009; **284**: 35632–35644.

60 Beranek A, Rechberger G, Knauer H, Wolinski H, Kohlwein SD, Leber R. Identification of a cardiolipin-specific phospholipase encoded by the gene CLD1 (YGR110W) in yeast. *J Biol Chem* 2009; **284**: 11572–11578.

61 Xu Y, Kelley RI, Blanck TJJ, Schlame M. Remodeling of cardiolipin by phospholipid transacylation. *J Biol Chem* 2003; **278**: 51380–51385.

62 Xu Y, Malhotra A, Ren M, Schlame M. The enzymatic function of tafazzin. *J Biol Chem* 2006; **281**: 39217–39224.

63 Testet E, Laroche-Traineau J, Noubhani A, Coulon D, Bunoust O, Camougrand N *et al.* Ypr140wp, 'the yeast tafazzin', displays a mitochondrial lysophosphatidylcholine (lyso-PC) acyltransferase activity related to triacylglycerol and mitochondrial lipid synthesis. *Biochem J* 2005; **387**: 617–626.

64 Ma L, Vaz FM, Gu Z, Wanders RJA, Greenberg ML. The Human TAZ Gene Complements Mitochondrial Dysfunction in the Yeast taz1Δ Mutant: IMPLICATIONS FOR BARTH SYNDROME *. *Journal of Biological Chemistry* 2004; **279**: 44394–44399.

65 Gu Z, Valianpour F, Chen S, Vaz FM, Hakkaart GA, Wanders RJA *et al.* Aberrant cardiolipin metabolism in the yeast taz1 mutant: a model for Barth syndrome. *Mol Microbiol* 2004; **51**: 149–158.

66 Claypool SM, Whited K, Srijumnong S, Han X, Koehler CM. Barth syndrome mutations that cause tafazzin complex lability. *J Cell Biol* 2011; **192**: 447.

67 Baile MG, Sathappa M, Lu YW, Pryce E, Whited K, McCaffery JM *et al.* Unremodeled and Remodeled Cardiolipin Are Functionally Indistinguishable in Yeast. *J Biol Chem* 2014; **289**: 1768.

68 Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z *et al.* IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* 2015; **31**: 3359–3361.

69 Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E *et al.* UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 2023; **51**: D523–D531.

70 Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018; **46**: D1062–D1067.

71 Home - OMIM. https://omim.org/ (accessed 24 Jun2023).

72 Chen S, Francioli LC, Goodrich JK, Collins RL, Wang Q, Alföldi J *et al.* A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. *bioRxiv* 2022; : 2022.03.20.485034.

73 Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**: 3812.

74 Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013; **Chapter 7**. doi:10.1002/0471142905.HG0720S76.

75 Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics* 2016; **32**: 2230–2232.

76 Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics* 2013; **14 Suppl 3**: 1–9.

77 Bromberg Y, Yachdav G, Rost B. SNAP predicts effect of mutations on protein function. *Bioinformatics* 2008; **24**: 2397–2398.

78 Savojardo C, Fariselli P, Martelli PL, Casadio R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. *Bioinformatics* 2016; **32**: 2542–2544.

79 Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 2013; **34**: 57–65.

80 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011; **7**: 539.

81 Schlame M, Xu Y. The Function of Tafazzin, a Mitochondrial Phospholipid-Lysophospholipid Acyltransferase. *J Mol Biol* 2020; **432**: 5043.

82 Mariani V, Biasini M, Barbato A, Schwede T. IDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* 2013; **29**: 2722–2728.

83 Zapała B, Płatek T, Wybrańska I. A novel TAZ gene mutation and mosaicism in a Polish family with Barth syndrome. *Ann Hum Genet* 2015; **79**: 218–224.

84 Mazurová S, Tesařová M, Magner M, Houšťková H, Hansíková H, Augustínová J *et al.* Novel mutations in the TAZ gene in patients with Barth syndrome. *Prague Med Rep* 2013; **114**: 139–153.

85 Bowron A, Honeychurch J, Williams M, Tsai-Goodman B, Clayton N, Jones L *et al.* Barth syndrome without tetralinoleoyl cardiolipin deficiency: a possible ameliorated phenotype. *J Inherit Metab Dis* 2015; **38**: 279–286.

86 Hastings R, Steward C, Tsai-Goodman B, Newbury-Ecob R. Dysmorphology of Barth syndrome. *Clin Dysmorphol* 2009; **18**: 185–187.

87 Bachou T, Giannakopoulos A, Trapali C, Vazeou A, Kattamis A. A novel mutation in the G4.5 (TAZ) gene in a Greek patient with Barth syndrome. *Blood Cells Mol Dis* 2009; **42**: 262–264.

88 Thompson WR, Decroes B, Mcclellan R, Rubens J, Vaz FM, Kristaponis K *et al.* New targets for monitoring and therapy in Barth syndrome. *Genet Med* 2016; **18**: 1001–1010.

89 Kuijpers TW, Maianski NA, Tool ATJ, Becker K, Plecko B, Valianpour F *et al.* Neutrophils in Barth syndrome (BTHS) avidly bind annexin-V in the absence of apoptosis. *Blood* 2004; **103**: 3915–3923.

90 Lu YW, Galbraith L, Herndon JD, Lu YL, Pras-Raves M, Vervaart M *et al.* Defining functional classes of Barth syndrome mutation in humans. *Hum Mol Genet* 2016; **25**: 1754–1770.

91 Houtkooper RH, Turkenburg M, Poll-The BT, Karall D, Pérez-Cerdá C, Morrone A *et al.* The enigmatic role of tafazzin in cardiolipin metabolism. *Biochim Biophys Acta* 2009; **1788**: 2003–2014.

92 Chatzispyrou IA, Guerrero-Castillo S, Held NM, Ruiter JPN, Denis SW, IJlst L *et al.* Barth syndrome cells display widespread remodeling of mitochondrial complexes without affecting metabolic flux distribution. *Biochim Biophys Acta Mol Basis Dis* 2018; **1864**: 3650–3658.

93 Suzuki-Hatano S, Sriramvenugopal M, Ramanathan M, Soustek M, Byrne BJ, Cade WT *et al.* Increased mtDNA Abundance and Improved Function

in Human Barth Syndrome Patient Fibroblasts Following AAV- TAZ Gene Delivery. *Int J Mol Sci* 2019; **20**. doi:10.3390/IJMS20143416.

94 D'Adamo P, Fassone L, Gedeon A, Janssen EAM, Bione S, Bolhuis PA *et al.* The X-linked gene G4.5 is responsible for different infantile dilated cardiomyopathies. *Am J Hum Genet* 1997; **61**: 862–867.

95 Steward CG, Newbury-Ecob RA, Hastings R, Smithson SF, Tsai-Goodman B, Quarrell OW *et al.* Barth syndrome: an X-linked cause of fetal cardiomyopathy and stillbirth. *Prenat Diagn* 2010; **30**: 970–976.

96 Abe M, Hasegawa Y, Oku M, Sawada Y, Tanaka E, Sakai Y *et al.* Mechanism for Remodeling of the Acyl Chain Composition of Cardiolipin Catalyzed by Saccharomyces cerevisiae Tafazzin. *J Biol Chem* 2016; **291**: 15491–15502.

97 Vanderniet JA, Benitez-Aguirre PZ, Broderick CR, Kelley RI, Balasubramaniam S. Barth syndrome with severe dilated cardiomyopathy and growth hormone resistance: a case report. *J Pediatr Endocrinol Metab* 2021; **34**: 951–955.

98 Johnston J, Kelley RI, Feigenbaum A, Cox GF, Iyer GS, Funanage VL *et al.* Mutation characterization and genotype-phenotype correlation in Barth syndrome. *Am J Hum Genet* 1997; **61**: 1053–1058.

99 Kirwin SM, Manolakos A, Barnett SS, Gonzalez IL. Tafazzin splice variants and mutations in Barth syndrome. *Mol Genet Metab* 2014; **111**: 26–32.

100 Sabater-Molina M, Guillén-Navarro E, García-Molina E, Ballesta-Martínez MJ, Escudero F, Ruiz-Espejo F. Barth syndrome in adulthood: a clinical case. *Rev Esp Cardiol (Engl Ed)* 2013; **66**: 68–70.

101 Borna NN, Kishita Y, Ishikawa K, Nakada K, Hayashi JI, Tokuzawa Y *et al.* A novel mutation in TAZ causes mitochondrial respiratory chain disorder without cardiomyopathy. *J Hum Genet* 2017; **62**: 539–547.

102 Imai-Okazaki A, Kishita Y, Kohda M, Mizuno Y, Fushimi T, Matsunaga A *et al.* Cardiomyopathy in children with mitochondrial disease: Prognosis and genetic background. *Int J Cardiol* 2019; **279**: 115–121.

103 Sakamoto O, Kitoh T, Ohura T, Ohya N, Iinuma K. Novel missense mutation (R94S) in the TAZ (G4.5) gene in a Japanese patient with Barth syndrome. *J Hum Genet* 2002; **47**: 229–231.

104 Rigaud C, Lebre AS, Touraine R, Beaupain B, Ottolenghi C, Chabli A *et al.* Natural history of Barth syndrome: a national cohort study of 22 patients. *Orphanet J Rare Dis* 2013; **8**. doi:10.1186/1750-1172-8-70.

105 Brady AN, Shehata BM, Fernhoff PM. X-linked fetal cardiomyopathy caused by a novel mutation in the TAZ gene. *Prenat Diagn* 2006; **26**: 462–465.

106 Claypool SM, Whited K, Srijumnong S, Han X, Koehler CM. Barth syndrome mutations that cause tafazzin complex lability. *J Cell Biol* 2011; **192**: 447–462.

107 Thiels C, Fleger M, Huemer M, Rodenburg RJ, Vaz FM, Houtkooper RH *et al.* Atypical Clinical Presentations of TAZ Mutations: An Underdiagnosed Cause of Growth Retardation? *JIMD Rep* 2016; **29**: 89–93.

108 Vreken P, Valianpour F, Nijtmans LG, Grivell LA, Plecko B, Wanders RJA *et al.* Defective remodeling of cardiolipin and phosphatidylglycerol in Barth syndrome. *Biochem Biophys Res Commun* 2000; **279**: 378–382.

109 Peake RWA, Daly KP. Dilated Cardiomyopathy in a 2-Year-Old Infant. *Clin Chem* 2017; **63**: 433–435.

110 Ichida F, Tsubata S, Bowles KR, Haneda N, Uese K, Miyawaki T *et al.* Novel gene mutations in patients with left ventricular noncompaction or Barth syndrome. *Circulation* 2001; **103**: 1256–1263.

111 Schlame M, Kelley RI, Feigenbaum A, Towbin JA, Heerdt PM, Schieble T *et al.* Phospholipid Abnormalities in Children with Barth Syndrome. *J Am Coll Cardiol* 2003; **42**: 1994–1999.

112 Al-Wakeel-Marquard N, Degener F, Herbst C, Kühnisch J, Dartsch J, Schmitt B *et al.* RIKADA Study Reveals Risk Factors in Pediatric Primary Cardiomyopathy. *J Am Heart Assoc* 2019; **8**. doi:10.1161/JAHA.119.012531.

113 Wang C, Hata Y, Hirono K, Takasaki A, Ozawa SW, Nakaoka H *et al.* A Wide and Specific Spectrum of Genetic Variants and Genotype-Phenotype Correlations Revealed by Next-Generation Sequencing in Patients with Left Ventricular Noncompaction. *J Am Heart Assoc* 2017; **6**. doi:10.1161/JAHA.117.006210.

114 Hirono K, Hata Y, Nakazawa M, Momoi N, Tsuji T, Matsuoka T *et al.* Clinical and Echocardiographic Impact of Tafazzin Variants on Dilated Cardiomyopathy Phenotype in Left Ventricular Non-Compaction Patients in Early Infancy. *Circ J* 2018; **82**: 2609–2618.

115 Wu Q, Ma X, Shi H, Kong X, Ren S, Jiao Z. [Genetic analysis of a family with recurrent hydrops fetalis and dilated cardiomyopathy]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2019; **36**: 1028–1030.

116 Cantlay AM, Shokrollahi K, Allen JT, Lunt PW, Newbury-Ecob RA, Steward CG. Genetic analysis of the G4.5 gene in families with suspected Barth syndrome. *J Pediatr* 1999; **135**: 311–315.

117 Bissler JJ, Tsoras M, Göring HHH, Hug P, Chuck G, Tombragel E *et al.* Infantile dilated X-linked cardiomyopathy, G4.5 mutations, altered lipids, and ultrastructural malformations of mitochondria in heart, liver, and skeletal muscle. *Lab Invest* 2002; **82**: 335–344.

118 Christodoulou J, McInnes RR, Jay V, Wilson G, Becker LE, Lehotay DC *et al.* Barth syndrome: clinical observations and genetic linkage studies. *Am J Med Genet* 1994; **50**: 255–264.

119 Bleyl SB, Mumford BR, Thompson V, Carey JC, Pysher TJ, Chin TK *et al.* Neonatal, lethal noncompaction of the left ventricular myocardium is allelic with Barth syndrome. *Am J Hum Genet* 1997; **61**: 868–872.

120 Donati MA, Malvagia S, Pasquini E, Morrone A, Marca G La, Garavaglia B *et al.* Barth syndrome presenting with acute metabolic decompensation in the neonatal period. *J Inherit Metab Dis* 2006; **29**: 684.

121 Pronicka E, Piekutowska-Abramczuk D, Ciara E, Trubicka J, Rokicki D, Karkucinska-Wieckowska A *et al.* New perspective in diagnostics of mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric centre. *J Transl Med* 2016; **14**. doi:10.1186/S12967-016-0930-9.

122 Valianpour F, Wanders RJA, Overmars H, Vreken P, Van Gennip AH, Baas F *et al.* Cardiolipin deficiency in X-linked cardioskeletal myopathy and neutropenia (Barth syndrome, MIM 302060): a study in cultured skin fibroblasts. *J Pediatr* 2002; **141**: 729–733.

123 Kelley RI, Cheatham JP, Clark BJ, Nigro MA, Powell BR, Sherwood GW *et al.* X-linked dilated cardiomyopathy with neutropenia, growth retardation, and 3-methylglutaconic aciduria. *J Pediatr* 1991; **119**: 738–747.

124 Gonzalvez F, D'Aurelio M, Boutant M, Moustapha A, Puech JP, Landes T *et al.* Barth syndrome: cellular compensation of mitochondrial dysfunction and apoptosis inhibition due to changes in cardiolipin remodeling linked to tafazzin (TAZ) gene mutation. *Biochim Biophys Acta* 2013; **1832**: 1194–1206.

125 Tsujii N, Hayashi T, Hayashi T, Kimura A, Nishikubo T. Barth syndrome associated with triple mutation. *Pediatr Int* 2018; **60**: 385–387.

126 Zhao Y, Feng Y, Ding X, Dong S, Zhang H, Ding J *et al.* Identification of a novel hypertrophic cardiomyopathy-associated mutation using targeted next-generation sequencing. *Int J Mol Med* 2017; **40**: 121–129.

127 Ferri L, Donati MA, Funghini S, Malvagia S, Catarzi S, Lugli L *et al.* New clinical and molecular insights on Barth syndrome. *Orphanet J Rare Dis* 2013; **8**. doi:10.1186/1750-1172-8-27.

128 Takeda A, Sudo A, Yamada M, Yamazawa H, Izumi G, Nishino I *et al.* Barth syndrome diagnosed in the subclinical stage of heart failure based on the presence of lipid storage myopathy and isolated noncompaction of the ventricular myocardium. *Eur J Pediatr* 2011; **170**: 1481–1484.

129 Folsi V, Miglietti N, Lombardi A, Boccacci S, Utyatnikova T, Donati C *et al.* Cardiomyopathy in a male patient with neutropenia and growth delay. *Ital J Pediatr* 2014; **40**. doi:10.1186/1824-7288-40-45.

130 Imai-Okazaki A, Kishita Y, Kohda M, Yatsuka Y, Hirata T, Mizuno Y *et al.* Barth Syndrome: Different Approaches to Diagnosis. *J Pediatr* 2018; **193**: 256–260.

131 Baban A, Adorisio R, Corica B, Rizzo C, Calì F, Semeraro M *et al.* Delayed appearance of 3-methylglutaconic aciduria in neonates with early onset metabolic cardiomyopathies: A potential pitfall for the diagnosis. *Am J Med Genet A* 2020; **182**: 64–70.

132 Man E, Lafferty KA, Funke BH, Lun KS, Chan SY, Chau AKT *et al.* NGS identifies TAZ mutation in a family with X-linked dilated cardiomyopathy. *BMJ Case Rep* 2013; **2013**. doi:10.1136/BCR-2012-007529.

133 Lindenbaum RH, Andrews PS, Khan AS. Two cases of endocardial fibroelastosis--possible x-linked determination. *Br Heart J* 1973; **35**: 38–40.

134 Vasilescu C, Ojala TH, Brilhante V, Ojanen S, Hinterding HM, Palin E *et al.* Genetic Basis of Severe Childhood-Onset Cardiomyopathies. *J Am Coll Cardiol* 2018; **72**: 2324–2338.

135 Brion M, De Castro Lopez MJ, Santori M, Munuzuri AP, Abel BL, Souto AMB *et al.* Prospective and Retrospective Diagnosis of Barth Syndrome Aided by Next-Generation Sequencing. *Am J Clin Pathol* 2016; **145**: 507–513.

136 Sağ E, Kamaşak T, Kaya G, Çakır M. A rare clinical association: Barth syndrome and cystic fibrosis. *Turk J Pediatr* 2019; **61**: 134–138.

137 Yen TY, Hwu WL, Chien YH, Wu MH, Lin MT, Tsao LY *et al.* Acute metabolic decompensation and sudden death in Barth syndrome: report of a family and a literature review. *Eur J Pediatr* 2008; **167**: 941–944.

138 Seitz A, Hinck A, Bekeredjian R, Sechtem U. Late diagnosis of Barth syndrome in a 39-year-old patient with non-compaction cardiomyopathy and neutropenia. *ESC Heart Fail* 2020; **7**: 697–701.

139 Ronvelia D, Greenwood J, Platt J, Hakim S, Zaragoza M V. Intrafamilial variability for novel TAZ gene mutation: Barth syndrome with dilated cardiomyopathy and heart failure in an infant and left ventricular noncompaction in his great-uncle. *Mol Genet Metab* 2012; **107**: 428–432.

140 Ances BM, Sullivan J, Weigele JB, Hwang V, Messé SR, Kasner SE *et al.* Stroke associated with Barth syndrome. *J Child Neurol* 2006; **21**: 805–807.

141 Kim GB, Kwon BS, Bae EJ, Noh C Il, Seong MW, Park SS. A novel mutation of the TAZ gene in Barth syndrome: acute exacerbation after contrast-dye injection. *J Korean Med Sci* 2013; **28**: 784–787.

142 Avdjieva-Tzavella D, Kathom H, Ivanova M, Yordanova I. Barth syndrome in male and female siblings caused by a novel mutation in the taz gene. *Article in Genetic counseling* 2016.https://www.researchgate.net/publication/311707892 (accessed 8 Jul2023).

143 Nie CX, Li L, Wan CQ. [Barth syndrome in a case]. *Zhonghua Er Ke Za Zhi* 2019; **57**: 146–147.

144 Woiewodski L, Ezon D, Cooper J, Feingold B. Barth Syndrome with Late-Onset Cardiomyopathy: A Missed Opportunity for Diagnosis. *J Pediatr* 2017; **183**: 196–198.

145 Finsterer J, Stollberger C. Acquired noncompaction in Barth syndrome due to the TAZ mutation c.481_482ins20. *J Pediatr* 2017; **186**: 214.

146 Vesel S, Stopar-Obreza M, Trebušak-Podkrajšek K, Jazbec J, Podnar T, Battelino T. A novel mutation in the G4.5 (TAZ) gene in a kindred with Barth syndrome. *Eur J Hum Genet* 2003; **11**: 97–101.

147 Fan Y, Steller J, Gonzalez IL, Kulik W, Fox M, Chang R *et al.* A Novel Exonic Splicing Mutation in the TAZ (G4.5) Gene in a Case with Atypical Barth Syndrome. *JIMD Rep* 2013; **11**: 99–106.

148 Marziliano N, Mannarino S, Nespoli L, Diegoli M, Pasotti M, Malattia C *et al.* Barth syndrome associated with compound hemizygosity and heterozygosity of the TAZ and LDB3 genes. *Am J Med Genet A* 2007; **143A**: 907–915.

149 Gene symbol: TAZ. Disease: Barth syndrome - PubMed. https://pubmed.ncbi.nlm.nih.gov/18846664/ (accessed 8 Jul2023).

150 Kirwin SM, Vinette KM, Schwartz SB, Funanage VL, Gonzalez IL. Multiple transmissions of Barth syndrome through an oocyte donor with a de novo TAZ mutation. *Fertil Steril* 2007; **87**: 976.e5-976.e7.

151 Sweeney RT, Davis GJ, Noonan JA. Cardiomyopathy of unknown etiology: Barth syndrome unrecognized. *Congenit Heart Dis* 2008; **3**: 443–448.

152 Aljishi E, Ali F. Barth syndrome: an X-linked cardiomyopathy with a novel mutation. *Indian J Pediatr* 2010; **77**: 1432–1433.

153 Malhotra A, Xu Y, Ren M, Schlame M. Formation of molecular species of mitochondrial cardiolipin. 1. A novel transacylation mechanism to shuttle fatty acids between sn-1 and sn-2 positions of multiple phospholipid species. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 2009; **1791**: 314–320.