Journal Pre-proof

An extensive computational approach to analyze and characterize the functional mutations in the galactose-1-phosphate uridyl transferase (GALT) protein responsible for classical galactosemia

Udhaya Kumar S, Thirumal Kumar D, Siva R, George Priya Doss C, Hatem Zayed

PII: S0010-4825(19)30436-6

DOI: https://doi.org/10.1016/j.compbiomed.2019.103583

Reference: CBM 103583

To appear in: *Computers in Biology and Medicine*

Received Date: 4 November 2019

Revised Date: 12 December 2019

Accepted Date: 12 December 2019

Please cite this article as: U. Kumar S, T. Kumar D, S. R, G.P. Doss C, H. Zayed, An extensive computational approach to analyze and characterize the functional mutations in the galactose-1-phosphate uridyl transferase (GALT) protein responsible for classical galactosemia, *Computers in Biology and Medicine* (2020), doi: https://doi.org/10.1016/j.compbiomed.2019.103583.

An extensive computational approach to analyze and characterize the functional mutations in the galactose-1-phosphate uridyl transferase (GALT) protein responsible for classical galactosemia

Udhaya Kumar S, Thirumal Kumar D, Siva R, George Priya Doss C*, Hatem Zayed**

[1]School of BioSciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

[2]Department of Biomedical Sciences, College of Health and Sciences, Qatar University, Doha, Qatar

[3]Department of Neuroscience Technology, College of Applied Medical Sciences, Imam Abdulrahman Bin Faisal University, Jubail 35816, Saudi Arabia.

***Correspondence to**:

Hatem Zayed, Department of Biomedical Sciences, College of Health and Sciences, Qatar University, Doha, Qatar, Email: hatem.zayed@qu.edu.qa

George Priya Doss C., Department of Integrative Biology, School of BioSciences and Technology, Vellore Institute of Technology, Vellore, India, Email: georgepriyadoss@vit.ac.in

Abstract

Type I galactosemia is a very rare autosomal recessive genetic metabolic disorder that occurs because of the mutations present in the galactose-1-phosphate uridyl transferase (*GALT*) gene, resulting in a deficiency of the GALT enzyme. The action of the GALT enzyme is to convert galactose-1-phosphate and uridine diphosphate glucose into glucose-1-phosphate (G1P) and uridine diphosphate-galactose, a crucial second step of the Leloir pathway. A missense mutation in the GALT enzyme leads to variable galactosemia's clinical presentations, ranging from mild to severe. Our study aimed to employ a comprehensive computational pipeline to analyze the most prevalent missense mutations (p.S135L, p.K285N, p.Q188R, and p.N314D) responsible for galactosemia; these genes could serve as potential targets for chaperone therapy. We analyzed the four mutations through different computational analyses, including amino acid conservation, *in silico* pathogenicity and stability predictions, and macromolecular simulations (MMS) at 50 ns. The stability and pathogenicity predictors showed that the p.Q188R and p.S135L mutants are the most pathogenic and destabilizing. In agreement with these results, MMS analysis demonstrated that the p.Q188R and p.S135L mutants possess higher deviation patterns, reduced compactness, and intramolecular H-bonds of the protein. This could be due to the physicochemical modifications that occurred in the mutants p.S135L and p.Q188R compared to the native. Evolutionary conservation analysis revealed that the most prevalent mutations positions were conserved among different species except N314. The proposed research study is intended to provide a basis for the therapeutic development of drugs and future treatment of classical galactosemia and possibly other genetic diseases using chaperone therapy.

Key Words: classical galactosemia, GALT, prevalent mutations, molecular dynamics, misfolding

Introduction:

Classical galactosemia or Type I galactosemia (OMIM: 606999) is a rare genetic metabolic disorder that is passed down in an autosomal recessive manner [1]. The estimated frequency of galactosemia has been determined to be 1: 105,000 in France, 1: 23,500 in Ireland, 1: 45,000 in West Germany and 1: 20,000 in Brazil [2–6]. Incidence and prevalence in other countries, especially in India, have yet to be estimated [7–9]. In most organisms, the Leloir pathway is required for galactose metabolism, which is involved in the biological interconversion of glucose and galactose using three specific enzymes, namely, galactokinase (GALK1, EC 2.7.1.6, OMIM: 604313), uridine diphosphate glucose 4-epimerase (GALE, EC 5.1.3.2, OMIM: 606953), and galactose-1-phosphate uridylyltransferase (GALT, EC 2.7.7.12, OMIM: 606999). The deficient enzymes that exist in the Leloir pathway are responsible for classical galactosemia, galactokinase deficiency, and galactose epimerase deficiency [10–12]. In the Leloir pathway's second step, the GALT enzyme is necessary for the breakdown and transformation of UDP-glucose and gal-1P into G1P and UDP-galactose.

Classical galactosemia mainly occurs due to reduced activity of the enzyme GALT owing to a single nucleotide change in the galactose-1-phosphate uridylyltransferase gene (*GALT*; 9p13.3). Low activity of the GALT enzyme results in increased accumulation of galactose-1-phosphate in a variety of tissues resulting in severe clinical representation [13], including liver dysfunction, cataract formation, vomiting, hepatomegaly, hypotonia, septicemia, weight loss, diarrhea, jaundice, lethargy, bleeding tendencies, and ovarian failure, which vary from patient to patient. Intellectual retardation, and developmental and cognitive defects are also observed in patients with galactosemia [13]. So far, galactosemia has no cure; however, some treatments such as liver transplantation and gene therapy are being pursued; however, these treatments are associated with failure risk inherent in each technique. In some *GALT* gene mutations, RNA-based

therapies, stop codon read-through and antisense-mediated exon skipping could be achieved, yet not in the most prevalent mutations that cause disease (p.S135L, p.K285N, p.Q188R, and p.L195P) [14]. Patients with galactosemia are prescribed to consume diets that are lactose and galactose-free to avoid severe neonatal and long-term complications, including premature ovarian insufficiency (POI) [15].

Recently, the Yue et al. (2016) group crystallized the structure of the GALT (43 kDa) enzyme (PDB code: 5IN3 and 6GQD), which shows the zinc-binding residues Asn72, Cys75, Ser135, His184, uridylylation site or active site His186, surface-exposed loop from residues 49-63 tangled in uridine diphosphate-hexose binding. The binding site of Glu-1-P is formed by the residues Lys334, Phe335, Val337, Tyr339, Glu340, and Gln346 from one chain, accompanied by residues Asn173 and Gln188 from another chain [16]. As of January 2018, nearly 335 different mutations had been identified and disclosed in the Human Gene Mutation Database (HGMD®) and *GALT* database (GALTdb), with 214 missense, 18 nonsense, 28 silent, 6 insertion, 22 deletion, and 22 splice site mutations [13,17]. Missense mutations are reported to form the largest group worldwide, consisting of the most common mutations among classical galactosemia. Previous reports suggested that p.Q188R is the most common mutation in Caucasians (60% worldwide), p.L195P in Spanish, p.K285N in Europeans and is considered to be the second most repeated mutation, p.S135L in South Africans, and p.N314D is assumed to be the most common mutation in the Indian population [7,15]. In the current study, we have classified the most prevalent mutations that are responsible for classical galactosemia and demonstrated the potential use of a bioinformatics pipeline to observe the potential correlation of genotype and phenotype in patients with classical galactosemia.

Methodology

Datasets

We retrieved the reported missense mutation information from the UniProt (https://www.uniprot.org/uniprot/P07902) and Human Gene Mutation Database (HGMD, http://www.hgmd.cf.ac.uk/ac/index.php) databases [17–20]. Protein sequence data (accession no. P07902) were collected from the Swiss-Prot database [21]. The GALT human protein structure was acquired from the RCSB Protein Data Bank (https://www.rcsb.org/) with PDB code: 6GQD [16,22]. For further investigation, we selected the mutations namely p.S135L, p.K285N, p.Q188R, and p.N314D, which are most prevalent in this disorder for our analysis.

Evolutionary conservation analysis

We performed multiple sequence alignment (MSA) by implementing the Clustal Omega from EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/) to align sequences of GALT protein from different species to understand the evolutionary relationship between these sequences [23]. The result of MSA from Clustal Omega was manually plotted onto ESPript3.0 (Easy Sequencing in PostScript, http://espript.ibcp.fr), which helps to render multiple sequence alignments for easy analysis. In an ESPript rendering, it is possible to show the aligned sequences, similarities, hydropathy, secondary structure elements, and intermolecular contacts [24,25]. The secondary structure information of GALT was inferred from the PDB coordinates (PDB ID: 6GQD). Then, we calculated the hydropathy of amino acids as stated by the Kyte and Doolittle algorithm with an average window of 3-residue [26]. The GenBank accession numbers and the vertebrates of the GALT sequences used are *Homo sapiens*- NP_000146.2, *Gorilla gorilla*- XP_004048011.1 (predicted), *Pan troglodytes*- XP_001163419.1, *Canis lupus*- XP_852579.1, *Rattus norvegicus*- NP_001013107.1, and *Mus musculus*- NP_057867.2. We also subjected the protein sequence to the ConSurf webserver to predict the level of conservation at a defined amino acid position [27].

Determination of pathogenicity and change in protein stability

The retrieved prevalent GALT missense mutations were examined for their consequence on protein's tertiary structure and function. Various *in silico* approaches were exploited to check the

detrimental nature of the missense mutations. For the analysis of protein stability and pathogenicity, we employed the iStable and PredictSNP algorithms, respectively [28][29]. For prediction, we provided a FASTA sequence of GALT protein and mutations as the input.

Analysis of amino acid physicochemical properties

To identify the biophysical characterization of the mutated amino acid, we used the Align-GVGD tool, which integrates Grantham deviation (GD) and Grantham variation (GV) to check that the missense mutations in GALT ordered between least risk to highest risk [30,31]. Furthermore, the physicochemical characteristics of the variants were distinguished using the Project 'Have Your Protein Explained' webserver (HOPE, http://www.cmbi.ru.nl/hope/). The tool utilizes several web servers, such as BLAST, FASTA, PDB (3D structure), DSSP, UniProt, YASARA, HSSP, DAS server, and ClustalW to identify the consequences of the missense mutation on the structure of a protein [32]. The protein sequence in FASTA format was used as an input to check the physicochemical properties across the amino acids of native and mutant, such as a change in charge, size, hydrophilicity, hydrophobicity, and loss of interactions between amino acids.

Preparation of mutant protein models

The wild-type protein structure was retrieved from PDB with the code 6GQD, and missense mutations were mapped to their corresponding amino acid positions using SPDB Viewer [33]. The PDB structure possessed A21Y, A22T, T23P, and R25L mutations during crystallization. These were reversed to their wild-type form before initiating the molecular dynamics simulation. The resulting mutant model was then energy minimized by applying the inbuilt force field GROMOS96 in SPDB Viewer [34]. The interactions, such as hydrogen bonds, hydrophobic, and other nearby interactions, between the native and mutant models were studied and visualized using the UCSF Chimera Molecular Modeling System [35,36].

Macromolecular simulation (MMS)

In this study, we used GROMACS version 4.6.5 to perform MMS for the mutant and native GALT proteins [37,38]. We used the AMBER99SB-ILDN force field parameter for energy minimizations, and it includes every atom explicitly [39]. To remove the local strains, we carried out energy minimization for native and mutant proteins present in the system. The solvation of native and mutant models were placed in a box, and the protein was centered in the box at least 1.2 nm from the box's edge. The water molecules were filled using the water models of SPC in the simulation boxes [40]. Our simulation system used the periodic boundary condition (PBC) to disregard the surface effect, and the cubic simulation box was replicated in space to create an infinite lattice. To neutralize our system, the GENION procedure was used from the GROMACS package to add required ions to the system. The energy was minimized for all atoms in native and mutants separately; then, the energy-minimized structure was obtained. To ensure that the energy minimization was successful, we checked whether the potential energy (Epot) had a negative value on the order of 10^5-10^6. Following the steepest descent minimization, equilibration of pressure was carried out under an NPT ensemble for 100 ps at 300 K. After NPT, the isothermal-isochoric or canonical ensemble was carried out under NVT for 100 ps at 300 K by implementing the protein's position restraints. The temperature was maintained at 300 K in all MM simulations. The position restraints were released from the equilibrated system and a 50-ns MM simulation was run to produce data. For every 2.0 ps of the entire simulation, the velocities, coordinates, and energies were saved.

MMS trajectory analysis

We focused on predicting differences in the protein flexibility, differences in the compactness of protein, and stability lost by comparing models of native and mutant using files from the MMS trajectory. GROMACS utility commands namely g_hbond, g_gyrate, and g_rms were applied to investigate the MMS production from the trajectory files to examine the intermolecular hydrogen

bonds, the deviation in root mean square, fluctuation in root mean square and radius of gyration of the models of native and mutants. The resultant values were plotted for the native and mutant models with the help of Xmgrace[41] to show the structural deviations between them.

Results

SNP dataset

The majority of nsSNPs are missense mutations, with approximately 211 mutations present in the GALT. Up-to-date information on human GALT mutations is represented in the http://www.arup.utah.edu/database/GALT/GALT_display.php database [13]. The most common mutations, including p.S135L, p.N314D, p.Q188R, and p.K285N from different ethnicities, were the focused of this computational study for further investigation.

Screening of missense mutations by *in silico* tools

We retrieved the four most frequent missense mutations that are responsible for classical galactosemia in the *GALT* gene from the UniProt and HGMD databases. To predict their effects on the protein's structural pattern and biological function, the retrieved missense mutations were incorporated into 10 *in silico* tools, which analyze their pathogenicity and structural significance. From these four missense mutations, PredictSNP, MAPP, PhD-SNP, and SIFT classified p.S135L and p.Q188R to be deleterious; PolyPhen-1, SNAP, and PolyPhen-2 predicted p.S135L to be neutral and p.Q188R to be deleterious. Mutations such as p.K285N and p.N314D were classified as neutral by PredictSNP, MAPP, PolyPhen-1, SIFT, PhD-SNP, PolyPhen-2, and SNAP. Furthermore, these missense mutations were subjected to destabilizing activity by using iStable, MUpro, and i-Mutant2.0, which predicted a deterioration in protein stabilization in the case of the mutations p.Q188R and p.K185N. Stability tools such as iStable and i-Mutant2.0 projected a reduction in the protein's stabilization for the mutation p.S135L, whereas MUpro showed an increase in stability. In the case of p.N314D, MUpro and iStable

showed a decrease in stability, while it predicted an increase in stability upon mutation by iMutant2.0 (Table 1). Collectively, two missense mutations (p.S135L and p.Q188R) were projected to be the most pathogenic, and all variants have negative consequences on the protein's stability. Consequently, Align-GVGD and ConSurf were employed for all variants to check the biophysical characteristics and evolutionarily conservation of the amino acids. As a result, we predicted that mutations p.S135L, and p.Q188R were highly conserved, with a confidence score of nine, and p.K285N and p.N314D were the comparatively less conserved with confidence scores of 6 and 1, respectively. The Align-GVGD analysis revealed that mutations p.S135L and p.K285N impede protein function by falling in the class C65, whereas the lesser hindering effect was exhibited for the mutations p.Q188R and p.N314D with classes C35 and C15, respectively (Table 2).

Physiochemical properties of amino acids

We employed the HOPE web server to identify the physicochemical characteristics of amino acids for the mutations, which exhibited both destabilizing and pathogenic effects on the native GALT protein. Hydrophobicity, size, polarity, and charges are the amino acid's unique properties, which play a crucial part in the protein's primary structure and function due to disease-causing mutations. We included the prevalent mutations to predict their physiochemical properties of amino acids (Table 3). From this analysis, we predicted the amino acid residue p.K285N changed its charge from positive to neutral, followed by p.Q188R, and p.N314D had adapted from neutral to positive upon single amino acid substitution. It was found that serine amino acids, which are hydrophobic, were mutated to a hydrophilic amino acid, leucine; therefore, there was an increase in size, which leads to the loss of H-bonds in the protein's core and interferes with the folding pattern of the native protein. For p.Q188R and p.K285N, a size increase and size reduction were observed, respectively. Overall, modifications that occurred due to mutations might result in critical differences in the structure of the protein.

Evolutionary conservation analysis

Evolutionarily, conservation analysis was carried out using Clustal Omega and ESPript3.0 tools. We selected GALT vertebrate species for this analysis to show strict sequence conservation of amino acid residues between them. From this interpretation, it is evident that the most prevalent mutations are conserved across all organisms, except N314; at this residue p.N314D is the most frequent mutation in the Indian population at a rate of approximately 40% [7]. UDP-hexose binding sites (AA 49-63), Glu-1-phosphate binding sites (amino acids N173, Q188, K334, F335, V337, Y339, E340, Q346) and active site residues (H186) are highly conserved among vertebrates and are depicted in Fig. 1 [16]. Fig. 2(a) shows the graphical representation of the ConSurf analysis of galactose-1-phosphate uridylyltransferase protein. These conserved binding regions in proteins are crucial for proper functioning; any variations in the amino acid sequence in these positions would lead to disease phenotype in the patient.

Macromolecular Simulations (MMS)

MMS was performed for the native and mutant (p.S135L, p.N314D, p.Q188R, and p.K285N) models to understand the atomic-level changes concerning time. MMS were conducted for 50 ns, and GROMACS plugins, such as g_rms, were utilized to investigate the RMSD, g_gyrate was used to explain the change in compactness, and g_hbond was used to examine intramolecular hydrogen bond formation. Analysis of RMSD was considered a prime standard for measuring the confluence of the protein system simulation under study. The backbone RMSD values from the trajectory indicate substantial deviations in the case of mutants p.S135L and p.Q188R compared to the native protein, whereas the mutants p.K285N and p.N314D showed the least RMSD throughout the simulation and converged at ~0.28 and ~0.38 nm, respectively. The mutant (p. K285N) protein attained an average deviation range of ~0.22 to ~0.25 nm in the 32-50 ns simulation period. Among the four trajectories, the p.Q188R mutant manifested the highest deviation from ~0.35 to 0.6 nm at 33 ns, which later converged to ~0.48 nm after the 50

ns simulation. In comparison with the native complex, at 35 ns the p.S135L mutant exhibited a larger RMSD spectrum of ~0.4 to ~0.53 nm and converged at 50 ns with an RMSD range of ~0.43 nm (Fig. 3). To assess structural flexibility at the level of the residue, the RMSF value of amino acid residues was calculated from the 50 ns MMS trajectory file of the native GALT and mutant GALT proteins. From this analysis, a more significant fluctuation was observed in the N-terminal region of p.Q188R and C-terminal region of the p.S135L mutants studied against the native GALT protein (Fig. 4). Furthermore, we interpreted the radius of gyration (Rg) from the MMS trajectory to identify the change in compactness of protein upon amino acid substitutions. From the Rg graph plot, we obtained a higher radius of gyration value for the p.S135L, Q188R, and p.N314D mutants when compared to that of native GALT. However, the p.K285N mutant maintained the pattern of Rg similar to the native GALT pattern (Fig. 5). To identify the change in the formation of an intramolecular hydrogen bond between native GALT and mutant proteins, we employed the g_hbond plugin to obtain the data from the MMS trajectory. From this analysis, we noticed that an increased amount of intramolecular hydrogen bonds from 245 to 290 in the p.S135L mutant at 28 ns. In contrast, the p.Q188R and p.N314D mutants exhibited a decline in the amount of intramolecular H-bonds from 9 ns to the completion of the MMS against native GALT. The distinctions in intramolecular H-bonds are negligible between the native GALT and p.K285N proteins (Fig. 6).

Discussion

We used seven *in silico* prediction tools (PredictSNP, SIFT, MAPP, PolyPhen-1, PhD-SNP, PolyPhen-1, SNAP, PolyPhen-2 that predict the pathogenicity of the four mutants) and the stability predictors (i-Mutant2.0, iStable, and MUpro) to find the most deleterious mutations. The PredictSNP tool provides us with a prediction that is more accurate by integrating multiple algorithms of other pathogenicity testing tools [42]. We successfully used these *in silico* tools to examine the potential genotype and phenotype correlations in several genetic disorders in

combination with simulation modeling and structural bioinformatics pipelines [43,43–48]. Here, we used a combinatorial bioinformatics pipeline on the four missense mutations (Table 1 and Fig. 2(b)) and found p.S135L and p.Q188R cause the most pathogenic mutations and reduce protein stability, whereas p.K285N and p.N314D disrupted only the stability (Fig. 3).

Conservation analysis using the ConSurf server showed that S135 is a highly conserved functional residue and is exposed, and Q188 is a highly conserved structural residue and is buried, whereas K285 and N314 were moderately and least conserved, respectively (Supplementary Fig. 1). However, N314 was not conserved among the vertebrates. This suggests that this p.N314D mutation is an evolutionary remnant because of the ancestral variant D314, which was identified in the pan-ethnic proportion of nearly 10% [49,50]. The mutation effect may cause differences in the protein's physiochemical characteristics that may be disturbed by either stabilizing or destabilizing the protein [51]. In this context, it is vital to interpret the amino acid characteristics to describe their contribution to the functioning of the protein. The physicochemical features of the mutants (p.S135L, p.N314D, p.Q188R, and p.K285N) are shown in Table 3. The serine amino acid is hydrophilic, highly targetable and mutable; the replacement of serine with leucine by mutation, which is a hydrophobic small residue, might lead to depletion of H-bonds in the core of the protein and disrupt the correct folding [52]. The mutation of arginine at the site of a highly conserved glutamine residue located on the protein's surface might disturb the interactions with other proteins. This could be the possible reason for the lesser catalytic activity observed in an enzymatic study conducted on this variant [16,53].

Furthermore, we performed MMS to distinguish the effect of four missense mutations in GALT protein at the atomistic level, which will help us to better understand the detailed structural differences of the four mutants against native GALT. We applied MMS technology to assess the impact of missense mutations in different genetic diseases to elucidate the molecular mechanism of mutations and the potential genotype-phenotype correlations [44,47,54–57]. We performed a

run of MMS for 50 ns. MMS trajectories revealed that the mutation influences the local conformation at the region or site of mutations, which also alters the GALT protein's overall stability and flexibility. The MMS study carried out by Tang, M et al. (2012) through homology modeling of GALT protein suggested that the static model did not bring out any perturbation [58]. In the current study, the X-ray crystallography structure was used to perform the MMS; this enables us to unravel the structural differences compared with the existing homology models. Furthermore, we analyzed MMS trajectories using the inbuilt GROMACS package utilities, such as RMSD, the Rg, RMSF, and intramolecular H-bonds. For analysis of RMSD, we plotted a 50 ns graph and at the completion of the MMS to observe a convergence pattern of the native GALT and mutant GALT proteins. The mutants p.S135L and p.Q188R manifested greater deviation in comparison with the native protein, indicating higher structural instability of these mutants. In contrast, both the mutants p.K285N and p.N314D showed a lower deviation pattern than the native protein. At the completion of the simulation, these mutants manifested convergence at ~0.28 and ~0.38 nm, respectively, indicating that the system acquired stability (Fig. 3). Our results were consistent with Coelho et al. (2014), who suggested that p.N314D has an enzymatic activity identical to the native protein [50]. In addition, the study by Singh et al. (2012) observed that p.N314D showed milder phenotypes in the homozygous condition and intermediate phenotypes in the heterozygous condition, and GALT enzyme activity was observed ≤50% in the patients (Fig. 3) [7].

To explore the overall fluctuations, we calculated the native GALT and mutant protein's RMSF. The RMSF graph from the 50 ns simulation shows a more significant fluctuation for p.Q188R at amino acid residues 29-45 and 119-134. The beta sheets (β1, β2, and β5, β6) are located in this region (Fig. 4). These regions might be crucial for the formation of beta sheets, and the affected region would lead to GALT protein unfolding. Additionally, this mutant does not affect the metal-binding residues, as mentioned by McCorvie et al. (2016). The Rg was evaluated to recognize the difference in protein structural rigidity due to the mutation, producing a greater

radius of gyration resulting in reduced compactness [59]. Fig. 5 depicts the mutations p.S135L, p.Q188R, and p.N314D, which demonstrated higher Rg than the native and mutation p.K285N. These findings indicate that the protein's compactness was lost prior to the change in positions 135, 188, and 314 from serine to leucine, glutamine to arginine, and asparagine to aspartic acid, respectively. However, at the completion of the 50 ns simulation, the variant p.K285N achieved lower Rg, thus maintaining the protein's compactness. Additionally, we spotted the formation of hydrogen bonds between proteins to perceive their stability reduction. This is because the function of hydrogen bonds is crucial for the protein's stability [60]. As shown in Fig. 6, the amount of hydrogen bonds was lowered in the mutations such as p.Q188R and p.N314D but increased in p.S135L variant against the native and mutant p.K285N. This significant loss occurred because of limited contacts with the neighboring amino acids in mutants p.Q188R and p.N314D, thus reducing its stability (Fig. 6). As observed through our radius of gyration results, loss of surrounding amino acid interactions might be a possible reason for compactness loss in the mutant structures.

The mutation p.S135L is highly prevalent in patients of African descendant with impaired GALT activity; the biochemical analysis showed slower aggregation kinetics, which is consistent with our MMS result [61,62]. The p.Q188R is considered to be the most prevalent mutation worldwide, primarily among Caucasian patients with classical galactosemia, resulting in a faster aggregation rate and/or lower catalytic activity of the GALT enzyme [50,63], consistent with our prediction findings (Table 1 and Fig. 3). However, our bioinformatics approach was unable to predict the significant structural or functional impact of p.K285N, which is the most commonly reported mutation among European descendants. However, the patients exhibited severe phenotypes, such as learning difficulties, ovarian dysfunction, delayed milestones, speech impairment, and lack of professional training in several compound heterozygous genotypes, including p.R51L/p.K285N, p.H114P/p.K285N, p.Y209S/p.K285N, p.P325L/p.K285N, p.N314D/p.K285N, and p.G338G/p.K285N [62,63]. The enzyme kinetics study showed that

the aggregation rate of this mutation is similar to that of native GALT, and this might undergo alternative events in the cell to disrupt the function of a protein. It was reported that p.N314D was very common in the Indian population, with reduced GALT activity and slower aggregation rate; this could be associated with Duarte-2 variants [7–9]. This is well correlated with our results from intramolecular hydrogen bonds and Rg interpretation by reducing the protein's stabilization (Figs. 5 and 6). Thus, the mutations p.S135L, p.N314D, and p.Q188R indicate impeded function and/or conformational stability. Taken together, the computational findings from our study were consistent with previous experimental observations indicating that all four mutations were most deleterious, causing mild and severe phenotypes in patients [7,58].

Conclusion

To our greatest understanding, the present study is the first attempt to use a computational pipeline to explore the crystal structure and effects due to the most prevalent mutations in the GALT protein. Among them, we observed that the p.S135L and p.Q188R mutations were extremely pathogenic and had destabilizing effects compared to the native and the other mutations. Our *in silico* assessment revealed that the mutations p.K285N and p.N314D, located in the protein's alpha helix region (α4) and loop region, respectively, are less pathogenic, whereas the p.S135L and p.Q188R pathogenic mutations are located at the loop region and the beta-sheet eight regions (β8). This location of the p.Q188R mutation could be the cause of the mutation's potential pathogenicity. Therefore, our study showed a clear understanding of the structural consequences that occurred due to mutations at specific sites by analyzing the RMSD, RMSF, and Rg plots along with intramolecular hydrogen bond formation. MMS trajectory analysis showed a more significant deviation and fluctuation in the p.Q188R mutation with a loss in compactness reduced the amount of intramolecular hydrogen bonds, while p.S135L demonstrated reduced compactness and increased intramolecular hydrogen bonds. Taken together, our MMS and computational tools results conclude that p.S135L and p.Q188R are

most deleterious among the prevalent mutations that disrupt the overall functioning of the protein. Computational pipelines are a promising tool to assist in understanding the plausible correlation of genotype and phenotype correlation in genetic disorders. Our study is anticipated to provide a platform for developing targets for chaperone therapy for genetic disorders due to missense mutations, particularly classical galactosemia.

Conflict of interest

No conflict of interest exists.

Acknowledgment

The authors thank the management of VIT for providing the facilities and encouragement to carry out this work.

References

[1] K.J. Isselbacher, E.P. Anderson, K. Kurahashi, H.M. Kalckar, Congenital Galactosemia, a Single Enzymatic Block in Galactose Metabolism, Science. 123 (1956) 635–636. https://doi.org/10.1126/science.123.3198.635.

[2] K.P. Coss, P.P. Doran, C. Owoeye, M.B. Codd, N. Hamid, P.D. Mayne, E. Crushell, I. Knerr, A.A. Monavari, E.P. Treacy, Classical Galactosaemia in Ireland: incidence, complications and outcomes of treatment, Journal of Inherited Metabolic Disease. 36 (2013) 21–27. https://doi.org/10.1007/s10545-012-9507-9.

[3] D. Mathias, H. Bickel, Follow-up study of 16 years neonatal screening for inborn errors of metabolism in West Germany, Eur J Pediatr. 145 (1986) 310–312. https://doi.org/10.1007/BF00439408.

[4] M. Murphy, B. McHugh, O. Tighe, P. Mayne, C. O'Neill, E. Naughten, D.T. Croke, Genetic basis of transferase-deficient galactosaemia in Ireland and the population history of the Irish Travellers, Eur J Hum Genet. 7 (1999) 549–554. https://doi.org/10.1038/sj.ejhg.5200327.

[5] V.E. Shih, H.L. Levy, V. Karolkewicz, S. Houghton, M.L. Efron, K.J. Isselbacher, E. Beutler, R.A. MacCready, Galactosemia Screening of Newborns in Massachusetts, New England Journal of Medicine. 284 (1971) 753–757. https://doi.org/10.1056/NEJM197104082841404.

[6] S. Schweitzer, Newborn mass screening for galactosemia, Eur J Pediatr. 154 (1995) S37–S39. https://doi.org/10.1007/BF02143801.

[7] R. Singh, B.R. Thapa, G. Kaur, R. Prasad, Biochemical and molecular characterization of GALT gene from Indian galactosemia patients: Identification of 10 novel mutations and their structural and functional implications, Clinica Chimica Acta. 414 (2012) 191–196. https://doi.org/10.1016/j.cca.2012.09.017.

[8] A. Roy, T. Samanta, R. Purkait, A. Mukherji, S. Ganguly, Etiology, Clinical Spectrum and Outcome of Metabolic Liver Diseases in Children, J Coll Physicians Surg Pak. 23 (2013) 5. https://doi.org/03.2013/JCPSP.194198.

[9] S. Bhave, A. Bavdekar, Classic Galactosemia: Indian Scenario, Indian Pediatr. 53 (2016) 2.

[10] H.M. Holden, I. Rayment, J.B. Thoden, Structure and Function of Enzymes of the Leloir Pathway for Galactose Metabolism, J. Biol. Chem. 278 (2003) 43885–43888. https://doi.org/10.1074/jbc.R300025200.

[11] J.B. Thoden, D.J. Timson, R.J. Reece, H.M. Holden, Molecular Structure of Human Galactokinase IMPLICATIONS FOR TYPE II GALACTOSEMIA, J. Biol. Chem. 280 (2005) 9662–9670. https://doi.org/10.1074/jbc.M412916200.

[12] D.J. Timson, The structural and molecular biology of type III galactosemia, IUBMB Life. 58 (2006) 83–89. https://doi.org/10.1080/15216540600644846.

[13] F.R.O. Calderon, A.R. Phansalkar, D.K. Crockett, M. Miller, R. Mao, Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene, Human Mutation. 28 (2007) 939–943. https://doi.org/10.1002/humu.20544.

[14] M. Tang, S.I. Odejinmi, H. Vankayalapati, K.J. Wierenga, K. Lai, Innovative therapy for Classic Galactosemia — Tale of two HTS, Molecular Genetics and Metabolism. 105 (2012) 44–55. https://doi.org/10.1016/j.ymgme.2011.09.028.

[15] G.T. Berry, Classic Galactosemia and Clinical Variant Galactosemia, in: M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, K. Stephens, A. Amemiya (Eds.), GeneReviews®, University of Washington, Seattle, Seattle (WA), 1993. http://www.ncbi.nlm.nih.gov/books/NBK1518/ (accessed August 7, 2019).

[16] T.J. McCorvie, J. Kopec, A.L. Pey, F. Fitzpatrick, D. Patel, R. Chalk, L. Shrestha, W.W. Yue, Molecular basis of classic galactosemia from the structure of human galactose 1-phosphate uridylyltransferase, Hum Mol Genet. 25 (2016) 2234–2244. https://doi.org/10.1093/hmg/ddw091.

[17] P.D. Stenson, M. Mort, E.V. Ball, K. Shaw, A.D. Phillips, D.N. Cooper, The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine, Hum Genet. 133 (2014) 1–9. https://doi.org/10.1007/s00439-013-1358-4.

[18] P.D. Stenson, M. Mort, E.V. Ball, K. Evans, M. Hayden, S. Heywood, M. Hussain, A.D. Phillips, D.N. Cooper, The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies, Hum Genet. 136 (2017) 665–677. https://doi.org/10.1007/s00439-017-1779-6.

[19] UniProt: a worldwide hub of protein knowledge, Nucleic Acids Res. 47 (2019) D506–D515. https://doi.org/10.1093/nar/gky1049.

[20] T. UniProt Consortium, UniProt: the universal protein knowledgebase, Nucleic Acids Res. 46 (2018) 2699–2699. https://doi.org/10.1093/nar/gky092.

[21] A. Bairoch, R. Apweiler, The SWISS-PROT Protein Sequence Data Bank and Its New Supplement TREMBL, Nucleic Acids Res. 24 (1996) 21–25. https://doi.org/10.1093/nar/24.1.21.

[22] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242. https://doi.org/10.1093/nar/28.1.235.

[23] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Molecular Systems Biology. 7 (2011) 539. https://doi.org/10.1038/msb.2011.75.

[24] P. Gouet, X. Robert, E. Courcelle, ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins, Nucleic Acids Res. 31 (2003) 3320–3323. https://doi.org/10.1093/nar/gkg556.

[25] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Res. 42 (2014) W320–W324. https://doi.org/10.1093/nar/gku316.

[26] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, Journal of Molecular Biology. 157 (1982) 105–132. https://doi.org/10.1016/0022-2836(82)90515-0.

[27] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, Nucleic Acids Res. 44 (2016) W344–W350. https://doi.org/10.1093/nar/gkw408.

[28] J. Bendl, J. Stourac, O. Salanda, A. Pavelka, E.D. Wieben, J. Zendulka, J. Brezovsky, J. Damborsky, PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations, PLOS Computational Biology. 10 (2014) e1003440. https://doi.org/10.1371/journal.pcbi.1003440.

[29] C.-W. Chen, J. Lin, Y.-W. Chu, iStable: off-the-shelf predictor integration for predicting protein stability changes, BMC Bioinformatics. 14 (2013) S5. https://doi.org/10.1186/1471-2105-14-S2-S5.

[30] S.V. Tavtigian, A.M. Deffenbaugh, L. Yin, T. Judkins, T. Scholl, P.B. Samollow, D. de Silva, A. Zharkikh, A. Thomas, Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral, Journal of Medical Genetics. 43 (2006) 295–305. https://doi.org/10.1136/jmg.2005.033878.

[31] S.V. Tavtigian, G.B. Byrnes, D.E. Goldgar, A. Thomas, Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications, Human Mutation. 29 (2008) 1342–1354. https://doi.org/10.1002/humu.20896.

[32] H. Venselaar, T.A. te Beek, R.K. Kuipers, M.L. Hekkelman, G. Vriend, Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, BMC Bioinformatics. 11 (2010) 548. https://doi.org/10.1186/1471-2105-11-548.

[33] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling, ELECTROPHORESIS. 18 (1997) 2714–2723. https://doi.org/10.1002/elps.1150181505.

[34] W.F. van Gunsteren, Biomolecular Simulation: The GROMOS96 Manual and User Guide, Biomos ; Zürich, 1996.

[35] E.C. Meng, E.F. Pettersen, G.S. Couch, C.C. Huang, T.E. Ferrin, Tools for integrated sequence-structure analysis with UCSF Chimera, BMC Bioinformatics. 7 (2006) 339. https://doi.org/10.1186/1471-2105-7-339.

[36] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis, Journal of Computational Chemistry. 25 (2004) 1605–1612. https://doi.org/10.1002/jcc.20084.

[37] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. 1–2 (2015) 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

[38] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics. 29 (2013) 845–854. https://doi.org/10.1093/bioinformatics/btt055.

[39] K. Lindorff□Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field, Proteins: Structure, Function, and Bioinformatics. 78 (2010) 1950–1958. https://doi.org/10.1002/prot.22711.

[40] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935. https://doi.org/10.1063/1.445869.

[41] P.J. Turner, Grace Home, XMGRACE, Version 5.1.19. (2005). http://plasma-gate.weizmann.ac.il/Grace/ (accessed August 8, 2019).

[42] D.C. Hao, Y. Feng, R. Xiao, P. Gen Xiao, Non-neutral nonsynonymous single nucleotide polymorphisms in human ABC transporters: the first comparison of six prediction methods, Pharmacological Reports. 63 (2011) 924–934. https://doi.org/10.1016/S1734-1140(11)70608-9.

[43] C. George Priya Doss, H. Zayed, Comparative computational assessment of the pathogenicity of mutations in the Aspartoacylase enzyme, Metab Brain Dis. 32 (2017) 2105–2118. https://doi.org/10.1007/s11011-017-0090-5.

[44] A. Mosaeilhy, M.M. Mohamed, G.P.D. C, H.S.A. El Abd, R. Gamal, O.K. Zaki, H. Zayed, Genotype-phenotype correlation in 18 Egyptian patients with glutaric acidemia type I, Metab Brain Dis. 32 (2017) 1417–1426. https://doi.org/10.1007/s11011-017-0006-4.

[45] D. Thirumal Kumar, H.G. Eldous, Z.A. Mahgoub, C. George Priya Doss, H. Zayed, Computational modelling approaches as a potential platform to understand the molecular genetics association between Parkinson's and Gaucher diseases, Metab Brain Dis. 33 (2018) 1835–1847. https://doi.org/10.1007/s11011-018-0286-3.

[46] S. Udhaya Kumar, N. Priyanka, P. Sneha, C. George Priya Doss, Functional and structural characterization of missense mutations in PAX6 gene, Front. Biol. 10 (2015) 377–385. https://doi.org/10.1007/s11515-015-1346-2.

[47] O.K. Zaki, G. Priya Doss C, S.A. Ali, G.G. Murad, S.A. Elashi, M.S.A. Ebnou, T. Kumar D, O. Khalifa, R. Gamal, H.S.A. El Abd, B.N. Nasr, H. Zayed, Genotype–phenotype correlation in patients with isovaleric acidaemia: comparative structural modelling and computational analysis of novel variants, Hum Mol Genet. 26 (2017) 3105–3115. https://doi.org/10.1093/hmg/ddx195.

[48] O.K. Zaki, N. Krishnamoorthy, H.S. El Abd, S.A. Harche, R.A. Mattar, R.S. Al disi, M.Y. Nofal, R. El Bekay, K.A. Ahmed, C. George Priya Doss, H. Zayed, Two patients with Canavan disease and structural modeling of a novel mutation, Metab Brain Dis. 32 (2017) 171–177. https://doi.org/10.1007/s11011-016-9896-9.

[49] A.E. Carney, R.D. Sanders, K.R. Garza, L.A. McGaha, L.J.H. Bean, B.W. Coffee, J.W. Thomas, D.J. Cutler, N.L. Kurtkaya, J.L. Fridovich-Keil, Origins, distribution and expression of the Duarte-2 (D2) allele of galactose-1-phosphate uridylyltransferase, Hum Mol Genet. 18 (2009) 1624–1632. https://doi.org/10.1093/hmg/ddp080.

[50] A.I. Coelho, M. Trabuco, R. Ramos, M.J. Silva, I.T. de Almeida, P. Leandro, I. Rivera, J.B. Vicente, Functional and structural impact of the most prevalent missense mutations in classic galactosemia, Molecular Genetics & Genomic Medicine. 2 (2014) 484–496. https://doi.org/10.1002/mgg3.94.

[51] M. Petukh, T.G. Kucukkal, E. Alexov, On Human Disease-Causing Amino Acid Variants: Statistical Study of Sequence and Structural Patterns, Human Mutation. 36 (2015) 524–534. https://doi.org/10.1002/humu.22770.

[52] P. Creixell, E.M. Schoof, C.S.H. Tan, R. Linding, Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues, Philosophical Transactions of the Royal Society B: Biological Sciences. 367 (2012) 2584–2593. https://doi.org/10.1098/rstb.2012.0076.

[53] K. Lai, A.C. Willis, L.J. Elsas, The Biochemical Role of Glutamine 188 in Human Galactose-1-phosphate Uridyltransferase, J. Biol. Chem. 274 (1999) 6559–6566. https://doi.org/10.1074/jbc.274.10.6559.
[54] A.K. Agrahari, P. Sneha, C. George Priya Doss, R. Siva, H. Zayed, A profound computational study to prioritize the disease-causing mutations in PRPS1 gene, Metab Brain Dis. 33 (2018) 589–600. https://doi.org/10.1007/s11011-017-0121-2.
[55] S.K. Ali, P. Sneha, J.P. Christy, H. Zayed, C.G.P. Doss, Molecular dynamics-based analyses of the structural instability and secondary structure of the fibrinogen gamma chain protein with the D356V mutation, Journal of Biomolecular Structure and Dynamics. 35 (2017) 2714–2724. https://doi.org/10.1080/07391102.2016.1229634.
[56] H.J. Berendsen, S. Hayward, Collective protein dynamics in relation to function, Current Opinion in Structural Biology. 10 (2000) 165–169. https://doi.org/10.1016/S0959-440X(00)00061-0.
[57] E. Lindahl, Molecular Dynamics Simulations, in: A. Kukol (Ed.), Molecular Modeling of Proteins, Springer New York, New York, NY, 2015: pp. 3–26. https://doi.org/10.1007/978-1-4939-1465-4_1.
[58] M. Tang, A. Facchiano, R. Rachamadugu, F. Calderon, R. Mao, L. Milanesi, A. Marabotti, K. Lai, Correlation assessment among clinical phenotypes, expression analysis and molecular modeling of 14 novel variations in the human galactose-1-phosphate uridylyltransferase gene, Human Mutation. 33 (2012) 1107–1115. https://doi.org/10.1002/humu.22093.
[59] M.Yu. Lobanov, N.S. Bogatyreva, O.V. Galzitskaya, Radius of gyration as an indicator of protein structure compactness, Mol Biol. 42 (2008) 623–628. https://doi.org/10.1134/S0026893308040195.
[60] D.T. Kumar, C.G.P. Doss, P. Sneha, I.A. Tayubi, R. Siva, C. Chakraborty, R. Magesh, Influence of V54M mutation in giant muscle protein titin: a computational screening and molecular dynamics approach, Journal of Biomolecular Structure and Dynamics. 35 (2017) 917–928. https://doi.org/10.1080/07391102.2016.1166456.
[61] J.L. Fridovich-Keil, S.D. Langley, L.A. Mazur, J.C. Lennon, P.P. Dembure, J.L. Elsas, Identification and functional analysis of three distinct mutations in the human galactose-1-phosphate uridyltransferase gene associated with galactosemia in a single family., Am J Hum Genet. 56 (1995) 640–646.
[62] Y.-P. Yang, N. Corley, J. Garcia□Heras, Molecular analysis in newborns from Texas affected with galactosemia, Human Mutation. 19 (2002) 82–83. https://doi.org/10.1002/humu.9005.
[63] A. Boutron, A. Marabotti, A. Facchiano, D. Cheillan, M. Zater, C. Oliveira, C. Costa, P. Labrune, M. Brivet, Mutation spectrum in the French cohort of galactosemic patients and structural simulation of 27 novel missense variations, Molecular Genetics and Metabolism. 107 (2012) 438–447. https://doi.org/10.1016/j.ymgme.2012.07.025.

Figure Legends

Fig. 1. Multiple sequence alignment of galactose-1-phosphate uridylyltransferase. MSA of GALT of a vertebrate species including *Homo sapiens* was performed by using Clustal Omega and plotted using ESPRIPTv3.0. The bold black font represents strictly conserved amino acid

residues, while those in red have 70% or more identity. The hydropathy plot is shown in the bar below the sequence block. Hydrophobic regions are marked in pink, hydrophilic in cyan, and intermediate regions are marked in gray. A brown star represents the UDP-hexose binding site, a blue circle denotes glucose-1-phosphate binding sites, and a black triangle indicates the H186 active site. The most prevalent mutations are highlighted in green. The secondary structure presented on top of the sequence alignment is inferred from the crystal structure of human GALT (PDB ID: 6GQD).

Fig. 2. (a) ConSurf results demonstrating the conservation of the native residues based on conservation grades and conservation levels. The enlarged portion depicts the location of native residues. Residues p.S135 and p.Q188 are evolutionarily conserved and shown in maroon, whereas p.K285 and p.N314 are conserved average and least are shown in white and turquoise, respectively. (b) The prevalent mutations were mapped on the three-dimensional structure of GALT protein. A few alpha helices and beta sheets of the secondary structure of GALT protein were labeled. The red sphere represents the native amino acids p.S135, p.Q188, p.K285, and p.N314. An enlarged graphical image shows the superimposition of native and mutant models. The native amino acids are shown in green and mutants in a red ball and stick model, respectively.

Fig. 3. Backbone RMSD of native and mutant GALT protein structures. In this graph, the running averages of 50 ns MMS of native and mutants were plotted using the Xmgrace plotting tool. The colors black, red, green, blue, and violet indicate native, S135L, Q188R, K285N, and N314D, respectively. The abscissa is time (ns), and the ordinate is RMSD (nm).

Fig. 4. Backbone RMSF of native and mutant GALT protein structures. In this graph, the native was plotted with actual RMSF values along with the running averages of all mutants during 50 ns molecular dynamics simulation using the Xmgrace plotting tool. The colors black, red, green, blue, and violet indicate native, S135L, Q188R, K285N, and N314D mutant structures, respectively. The abscissa portrays the residues, and the ordinate is RMSF (nm).

Fig. 5. Analysis of the radius of gyration of native and mutant GALT protein structures for 50 ns MMS. The colors black, red, green, blue, and violet indicate native, S135L, Q188R, K285N, and N314D mutant structures, respectively. The abscissa is time (ps), and the ordinate is Rg (nm).

Fig. 6. Differences in a number of intramolecular H-bonds between native and mutant GALT protein structures for 50 ns MMS. The colors black, red, green, blue, and violet

indicate native, S135L, Q188R, K285N, and N314D mutant structures, respectively. The abscissa is time (ps), and the ordinate is a number of intramolecular H-bonds formed in the GALT protein.

Table 1 Pathogenicity and stability prediction of the retrieved GALT protein mutations using PredictSNP and the iStable server

Table 2 Biophysical characterization and evolutionarily conservation prediction using Align-GVGD and ConSurf

Table 3 Amino acid physicochemical features are tabulated by using HOPE webserver.

Table 1: Pathogenicity and stability prediction of the retrieved GALT protein mutations using PredictSNP and iStable server

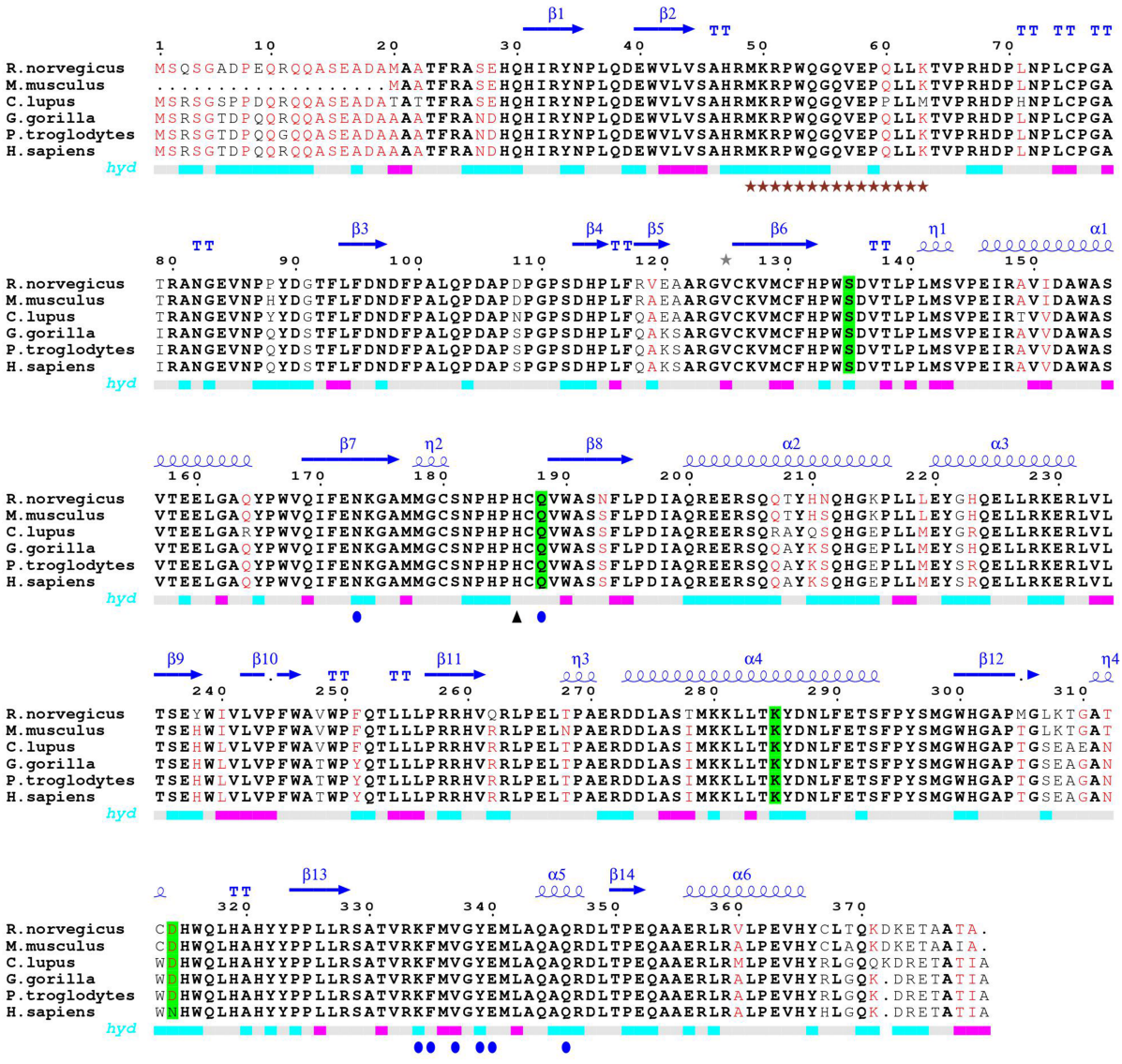
Variants	PredictSNP	MAPP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP	i□ Mutant2.0	MUpro	iStable
S135L	Deleterious	Deleterious	Deleterious	Neutral	Neutral	Deleterious	Neutral	Decrease	Increase	Decrease
Q188R	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Decrease	Decrease	Decrease
K285N	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Decrease	Decrease	Decrease
N314D	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Increase	Decrease	Decrease

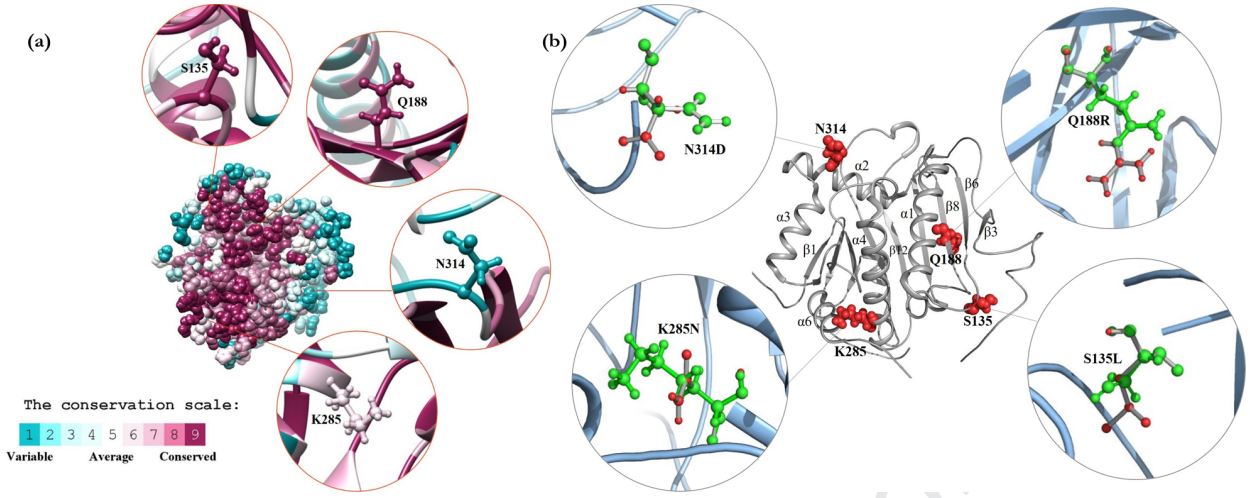
Table 2: Biophysical characterization and evolutionarily conservation prediction using Align GVGD and ConSurf

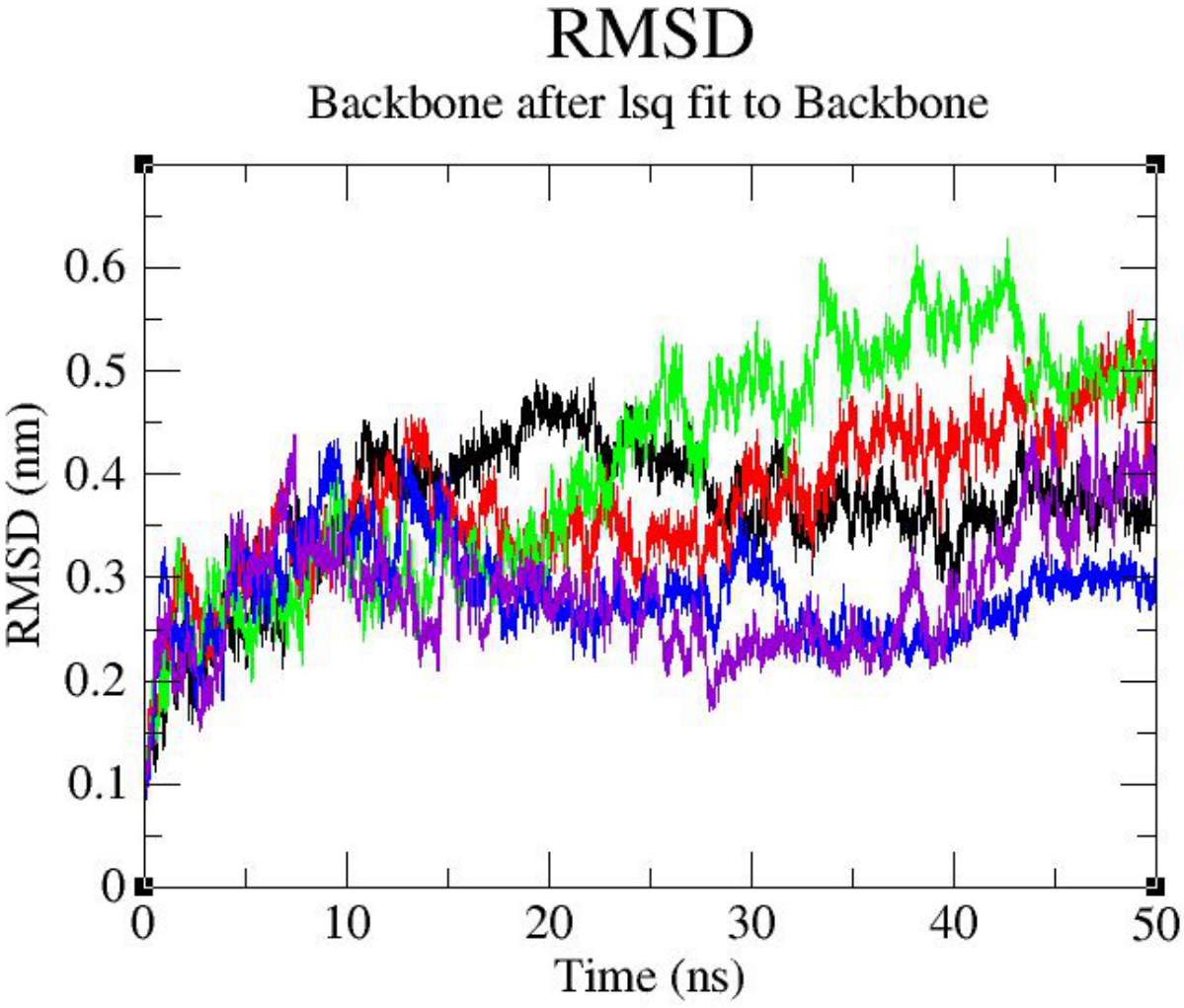
Accession Number	Variants	Align GVGD	ConSurf
VAR_002571	S135L	Class 65	Cons.score 9
VAR_002587	Q188R	Class 35	Cons.score 9
VAR_002609	K285N	Class 65	Cons.score 6
VAR_002613	N314D	Class 15	Cons.score 1

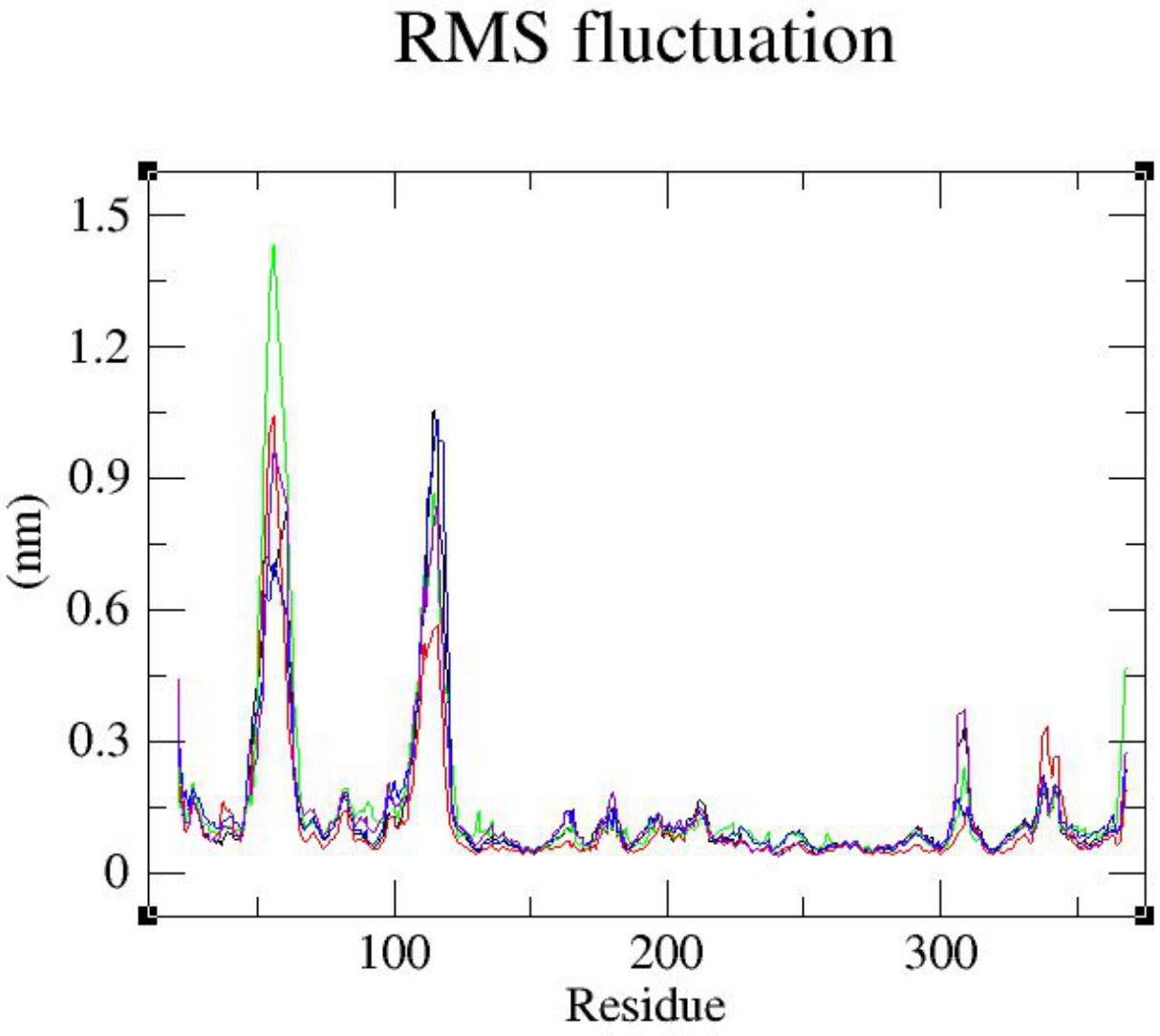
Table 3 Amino acid physicochemical features are tabulated by using HOPE webserver.

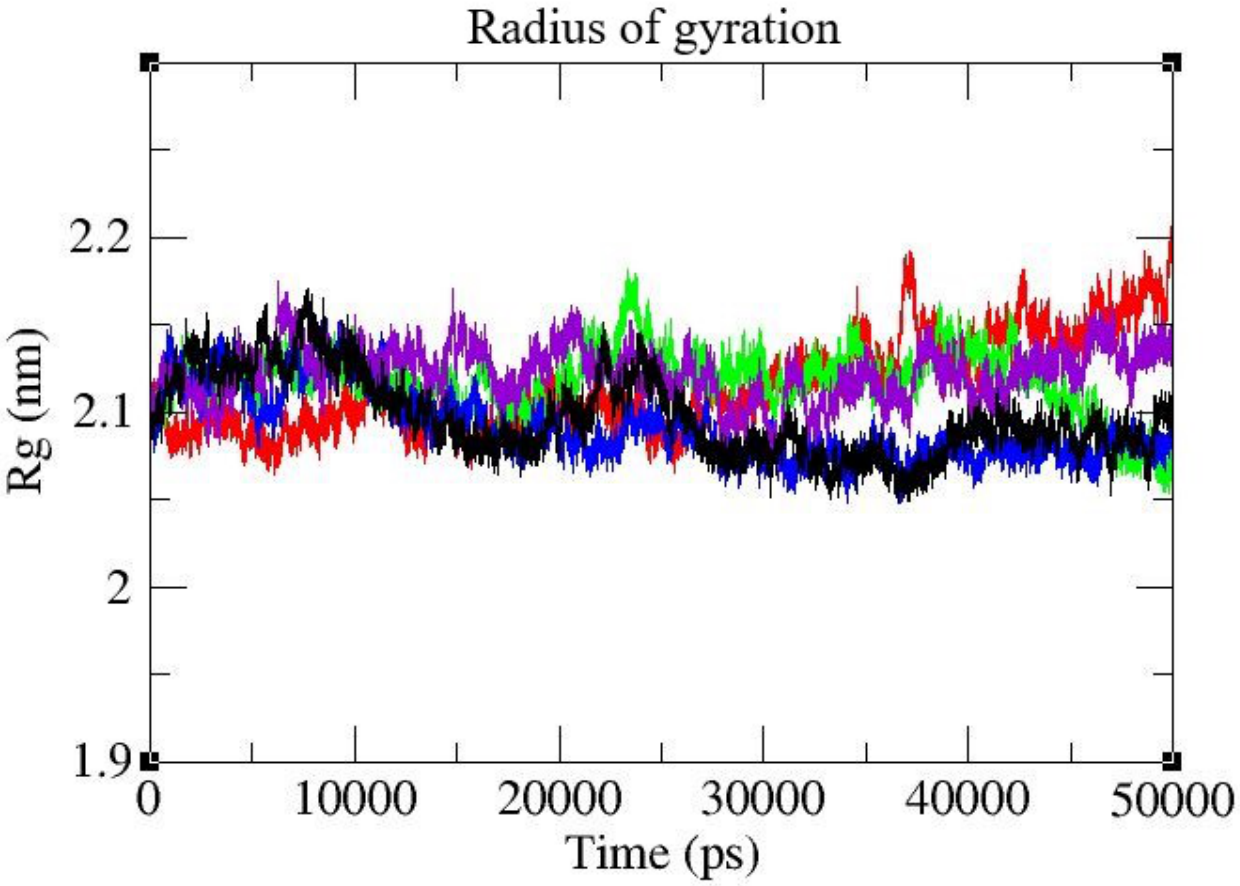
S.No	Variants	Change in charge	Change in hydrophilicity/hydrophobicity	Change in size	Mutation effect
1.	S135L	Unchanged	Hydrophilic to hydrophobic	Increase	The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding
2.	Q188R	Neutral to positive	Unchanged	Increase	The residue is located on the surface of the protein, mutation of this residue can disturb interactions with other molecules or other parts of the protein
3.	K285N	Positive to neutral	Unchanged	Decrease	The mutation will cause an empty space in the core of the protein
4.	N314D	Neutral to negative	Unchanged	Unchanged	The mutation introduces a charge at this position, this can cause repulsion between the mutant residue and neighbouring residues

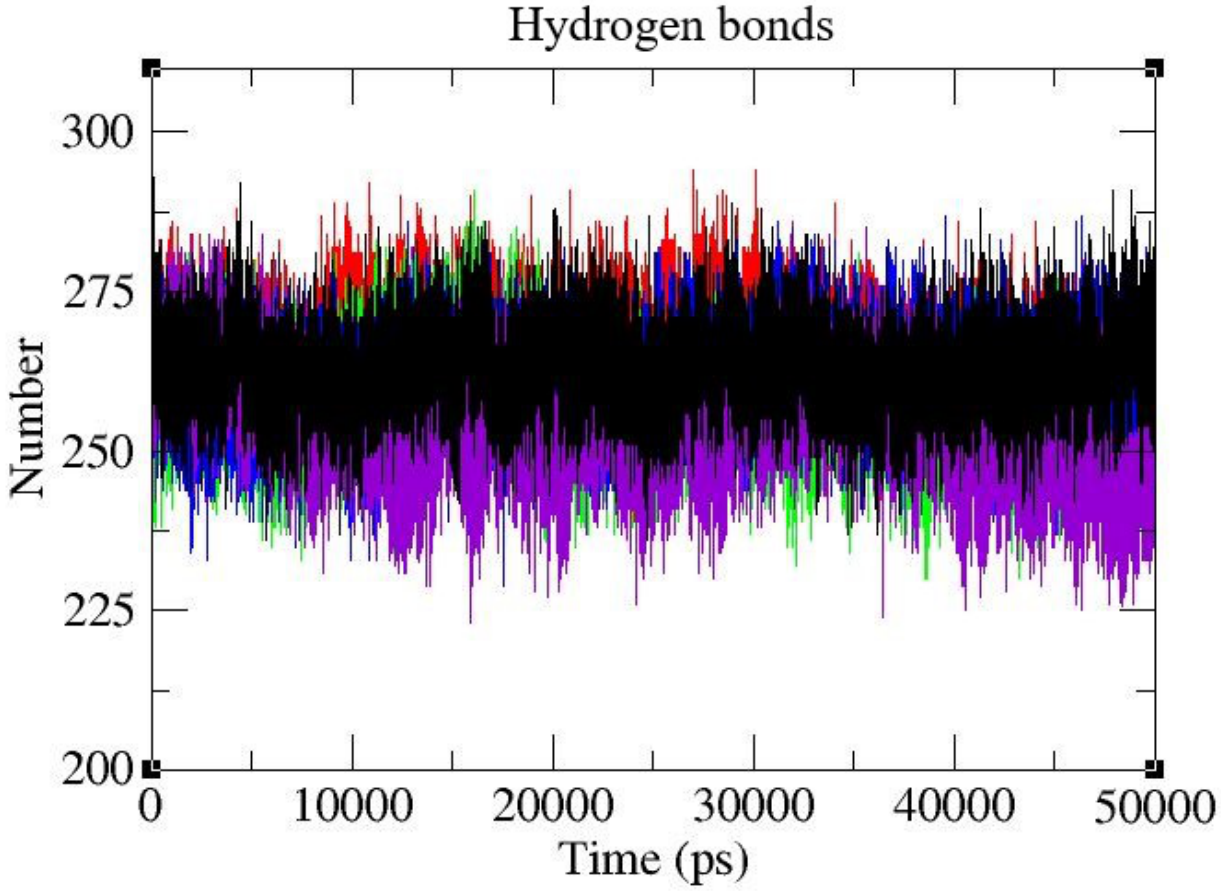

R.norvegicus
M.musculus
C.lupus
G.gorilla
P.troglodytes
H.sapiens
hyd




RMSD
Backbone after lsq fit to Backbone
RMSD (nm)
0
0.1
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40
50
Time (ns)


RMS fluctuation
1.5
1.2
0.9
0.6
0.3
0
(nm)
100
200
300
Residue


Radius of gyration
Rg (nm)
2.2
2.1
2
1.9
0
10000
20000
30000
40000
50000
Time (ps)


Hydrogen bonds
300
275
250
225
200
Number
0
10000
20000
30000
40000
50000
Time (ps)

Classical Galactosemia

p.Q188R and p.S135L most pathogenic and destabilizing mutants

Molecular Dynamics

We have no conflicts of interest to disclose