



## Research article

# Deciphering the conformational changes induced by high-risk nsSNPs in $\beta$ -lactoglobulin

Sthitaprajna Sahoo<sup>a</sup>, Vijayakumar Gosu<sup>b</sup>, Hak-Kyo Lee<sup>a,b,\*</sup>, Donghyun Shin<sup>a,\*\*</sup><sup>a</sup> Department of Agricultural Convergence Technology, Jeonbuk National University, Jeonju, 54896, Republic of Korea<sup>b</sup> Department of Animal Biotechnology, Jeonbuk National University, Jeonju, 54896, Republic of Korea

## A B S T R A C T

Whey protein from bovine milk is highly valued in the food and pharmaceutical industries because of its high protein content and abundance of essential amino acids. The relationship between whey protein and the  $\beta$ -lactoglobulin (BLG) gene has been extensively discussed because BLG is the most abundant whey protein, making up approximately 50 % of the total whey protein in bovine milk. In recent years, researchers have been interested in this gene because of its critical role in healthy milk production, and any genetic polymorphism in this gene may deteriorate the milk quality. In the current study, we identified several deleterious and damaging non-synonymous single nucleotide polymorphisms (nsSNPs) in BLG and analyzed their destabilizing effects using different computational algorithms. Cumulative results from all tools and evolutionary conservation profiles of BLG suggested that four nsSNPs, G17A, W19C, F136S, and C119R, were the most deleterious and could affect the structural integrity of the protein. Detailed molecular dynamics simulation analysis revealed that all variants induced major structural alterations, that affected the ability of the protein to interact with natural and synthetic ligands. Particularly, the G17A, F136S, and C119R variants induced large conformational changes in the EF loop and main  $\alpha$ -helix of BLG, which may affect the access of natural and synthetic ligands to the central calyx of BLG. We hope that the suggested nsSNPs will guide future studies and assist researchers in improving the quality of bovine milk.

## 1. Introduction

The  $\beta$ -lactoglobulin (BLG) is a predominant whey protein found in the milk of most mammals, with notable exception of humans [1, 2]. It belongs to the family of lipocalins, a group of transport proteins, and plays a significant role in transporting small hydrophobic molecules, including vitamins and fatty acids [3–5]. Because of its remarkable functional and nutritional value, it is widely used in the food and pharmaceutical industries. It contains essential amino acids required for human health, making it a valuable component of sports nutrition products, infant formula (although care must be taken due to its allergenic potential), and meal replacements [6]. In addition, the ability of BLG to bind and transport various bioactive compounds makes it a suitable candidate for developing drug delivery systems, particularly for hydrophobic drugs [7].

This small globular protein mainly comprises 162 amino acid residues with a molecular weight of approximately 18.3 kDa [4]. The secondary structure of BLG is characterized by nine antiparallel  $\beta$ -strands (A-I) and an external  $\alpha$ -helix at the external surface. Eight of these nine flattened  $\beta$ -strands fold up to form a well-suited internal hydrophobic cavity called calyx, which accommodates a range of ligands. This canonical calyx holds two surfaces, i.e.,  $\beta$ -strands A to D form one surface, and  $\beta$ -strands E to H form the other one. The ninth  $\beta$ -strand (I) of BLG plays a significant role in dimer interactions. These eight  $\beta$ -strands are connected by seven loops, i.e., AB, BC, CD, DE, EF, FG, and GH loops [7]. Out of these seven loops, three loops (BC, DE, and FG) are short and present at the closed end of the

\* Corresponding author. Department of Agricultural Convergence Technology, Jeonbuk National University, Jeonju, 54896, Republic of Korea.

\*\* Corresponding author.

E-mail addresses: [breedlee@jbnu.ac.kr](mailto:breedlee@jbnu.ac.kr) (H.-K. Lee), [sdh1214@gmail.com](mailto:sdh1214@gmail.com) (D. Shin).

<https://doi.org/10.1016/j.heliyon.2024.e40040>

Received 14 August 2024; Received in revised form 30 October 2024; Accepted 30 October 2024

Available online 1 November 2024

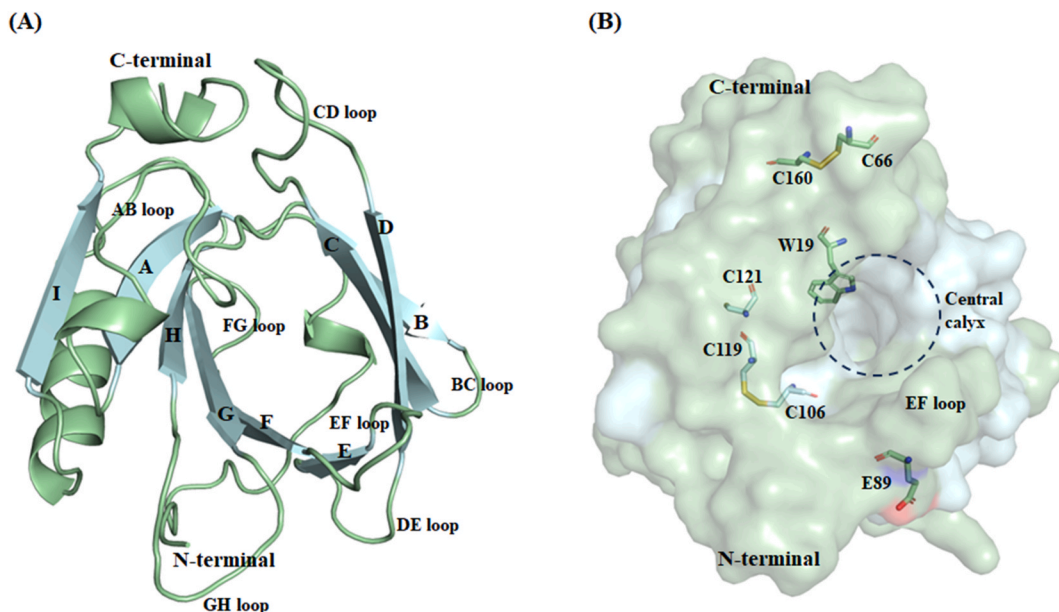
2405-8440/© 2024 Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

$\beta$ -barrel. In contrast, the other four loops (AB, CD, EF, and GH) are long, highly flexible, and present at the opened end of the  $\beta$ -barrel (Fig 1A). Although the protein retains its overall compact structure, its sensitivity to pH has been studied in greater detail. At low pH, this protein dissociates into monomers; however, at alkaline pH, the dimeric form is prevalent [7–9]. The “Tanford transition” is of particular significance at a neutral pH because it impacts the protein’s capacity to interact with synthetic and natural ligands. Briefly, the protonation state of the carboxyl group in the side chain of E89 at a pH value  $< 6$  causes the EF loop of the  $\beta$ -barrel to close, which prevents the ligand access to the central calyx of BLG. However, at pH  $> 7$ , the side chain of E89 is deprotonated, and the EF loop is exposed to the solvent, opening the EF loop (Fig 1B) [10]. Owing to its unique features, BLG has been considered a model protein in numerous studies aimed at explaining the biological significance of protein carrier mechanisms. In their native structures, ligands bind to different protein regions based on their physicochemical properties. Most small molecules can be hosted in the central calyx, and compete with palmitate (the primary natural ligand binding to the inner cavity of BLG) to stabilize BLG against physical and chemical denaturation [5]. Some molecules bind to another region of the BLG, i.e., between residues near  $\beta$ -strand B and C terminus (involving residues W19, Y20, Y42, Q44, Q59, Q68, L156, E157, E158, H161) or between the main  $\alpha$ -helix and  $\beta$ -strand G (involving residues Y102, L104, D129). These compounds do not displace palmitate or stabilize BLG during thermal denaturation [7]. Additionally, BLG contains two tryptophan residues (W19 and W61), with W19 accounting for most of the intrinsic fluorescence intensity of the protein. Hence, it is thought to be the best “reporter” of structural alterations [11–13]. In addition, BLG has five cysteine residues, four of which form disulfide bonds (C66-C160 and C106-C119), whereas the free thiol group of C121, with pH-dependent activity, participates in protein aggregation and denaturation (Fig 1B) [9,14].

The BLG gene, which encodes the prevalent whey protein in ruminant milk, is a significant factor in determining various aspects of milk composition, such as protein, fat, and lactose content. Polymorphisms in BLG may serve as potent molecular markers for evaluating these traits [15]. The distinct calyx structure of BLG, which has hydrophobic walls and several binding sites, affects protein stability, fat-soluble vitamin stability, and other functions. Given the significant impact of milk protein polymorphisms on a range of milk quality factors, it is crucial to investigate genetic variations across this protein. Single nucleotide polymorphisms (SNPs) are the most common single-base pair alterations in DNA sequences, that can be present in both the coding and non-coding regions of a gene. Assessing an individual’s SNP genotype may offer a framework for identifying disease susceptibility and the best treatment [16,17]. The non-synonymous SNPs (nsSNPs), also called missense or deleterious SNPs, are of great significance because they cause functional diversity in proteins by substituting an amino acid residue with a single base change in the coding region. These nsSNPs may significantly affect protein function by disturbing the physicochemical characteristics, structural stability, and solubility [18]. Despite playing a crucial regulatory role, non-coding SNPs have received substantially less attention from the functional analysis of genetic variants than coding region SNPs. Thus, in order to gain a better understanding of the clinical significance of these variations, it is necessary to prioritize them based on their functional impact [19–21]. However, demonstrating the functional impact of each variation through experimentation is challenging, expensive, and time-consuming in case-control association studies. Recent advancements in bioinformatics paved the way for analyzing, interpreting and deriving meaningful insights from the vast amounts of genomic and transcriptome data generated through high-throughput sequencing technologies [22–24]. Hence, the current study was designed to



**Fig. 1.** Features of the mature  $\beta$ -lactoglobulin (BLG) structure. (A) Cartoon representation of BLG and its nine  $\beta$ -strands, i.e., A to I, connected by seven significant loops. (B) Crucial residues of BLG; protonation of E89 triggers the Tanford transition; W19, the “reporter” for structural alterations; C119, C106, C160, and C66 forming disulfide bonds; C121 hidden beneath the main  $\alpha$ -helix. The internal hydrophobic cavity, i.e., the central calyx, is represented with a dashed black circle.

explore the most deleterious and damaging nsSNPs in the bovine BLG gene, and how they will affect the structural and functional integrity of the protein using a comprehensive in silico approach.

## 2. Materials and methods

### 2.1. Collection of datasets

All SNPs from the bovine BLG gene were assessed and retrieved from the Ensemble (<https://ensembl.org/>) database (Ensemble gene ID: ENSBTAG00000014678, accessed on Oct 16, 2023), which corresponds to UniProt ID: P02754. With a mass of 19,883 Da, the entire protein is composed of 178 amino acids. Residues 1–16 construct the signaling protein, while residues 17–178 encode the primary BLG protein. A total of 1203 SNPs were found throughout the BLG gene (ensemble transcript ID: ENSBTAT00000019538.6), of which 68 were determined to be non-synonymous SNPs (nsSNPs) or missense variants. All information about these 68 nsSNPs (such as variant ID, position, chromosomal location, alleles, and amino acid changes) was retrieved from the Ensemble database and provided in [S1 Table](#).

BLG was initially synthesized as a precursor protein containing a signal peptide at the N-terminus. This signaling peptide is a short sequence (residue 1–16) that directs the nascent protein to the secretory pathway, where it is subsequently cleaved to produce a mature protein (17–178). Therefore, when previous structural studies refer to residue numbers, they often begin counting from the beginning of the mature protein, i.e., 1–162, which is after the signal peptide [25]. Therefore, we renumbered the selected variants ([S2 Table](#)) of the mature protein by ignoring the initial signaling peptide sequence. Of these 68 nsSNPs, seven (rs460853867 (C/W), rs458468150 (A/G), rs477086634 (L/P), rs450652849 (T/P), rs436146677 (G/R), rs448268377 (A/S), and rs448268377 (A/T)) were found in the signaling peptide.

### 2.2. Identification of most deleterious nsSNPs

Predicting the deleterious nsSNPs using in silico tools is a crucial step in understanding the potential impact of genetic variation on protein function. Hence, the most deleterious or disease-causing nsSNPs in the BLG gene were predicted by using five different bioinformatics tools, such as SIFT (sorting intolerant from tolerant) ([https://sift.bii.a-star.edu.sg/www/SIFT\\_dbSNP.html](https://sift.bii.a-star.edu.sg/www/SIFT_dbSNP.html)) [26], Phd-SNP (predictor of human deleterious single nucleotide polymorphisms) (<https://snps.biofold.org/phd-snp/phd-snp.html>) [27], PANTHER (<http://pantherdb.org/tools/csnpscoreForm.jsp>) [28], SNAP (<https://snps.biofold.org/meta-snp>) [29], and Meta-SNP [30]. By integrating multiple prediction approaches, researchers can improve the precision and reliability of their predictions, leading to better understanding of the genetic basis of diseases, identifying novel therapeutic targets and advancing personalized medicine strategies. Detailed information on identifying deleterious nsSNPs using these tools is provided in our previous communication [27], and the prediction scores for all variants are presented in ([S2 Table](#)).

### 2.3. Prediction of destabilizing nsSNPs

The stability of a protein structure can be explained by the change in Gibbs free energy upon folding. Proteins are folded into distinct three-dimensional structures that are thermodynamically favorable. Protein stability is determined by the differences in Gibbs free energy ( $\Delta G$ ) between its unfolded and folded states. A negative  $\Delta G$  suggests that the folded state is more stable, and protein is likely to remain folded. A positive  $\Delta G$  indicates that the unfolded state is more stable, which may lead to protein misfolding or degradation. In the present study, we accessed six different stability prediction tools, such as I-mutant, DUET, mCSM, SDM2, STRUM, and CUPSAT [31–36]. These tools utilize different algorithms, conservation scores, and integrated structural information to predict the effects of various mutations on protein stability. In addition, we used the MutPred2 webserver (<http://mutpred.mutdb.org/>) to predict the pathogenicity of selected nsSNPs. MutPred2 has various characteristics, including protein structure, function, and evolution. It uses several structural disorder prediction techniques, such as MARCOIL, TMHMM, and three distinct servers: PSI-BLAST, SIFT, and Pfam [26,37–40]. As a result, MutPred2 aggregates these computational server' scores before delivering the predicted outcomes.

### 2.4. Conservation profile of BLG

We used the ConSurf database (<https://consurfdb.tau.ac.il/>) to analyze the evolutionary conservation profile of each amino acid in mature BLG protein. By examining the degree of conservation at each position, ConSurf can help researchers identify functional regions of proteins, such as active sites, ligand-binding sites, or regions important for protein stability, as well as critical regulatory motifs. These regions tend to be more conserved across species than the less functionally critical regions, reflecting the evolutionary pressure to maintain their structure and function [41,42].

### 2.5. Structural preparation of wild-type (WT) and selected variants of BLG

The WT crystal structure of BLG was downloaded from the RCSB PDB database with PDB ID: 3NQ3 and preprocessed by removing water, ions, and other small molecules [43]. Furthermore, the prepared protein structure was used to model all the selected variants with possible rotamers using the PyMOL mutagenesis plugin. All modeled variants were further subjected to initial energy minimization using the AMBER-ff99SBILDN force field of GROMACS 2022.2. Finally, the stereochemical properties of the modeled complexes

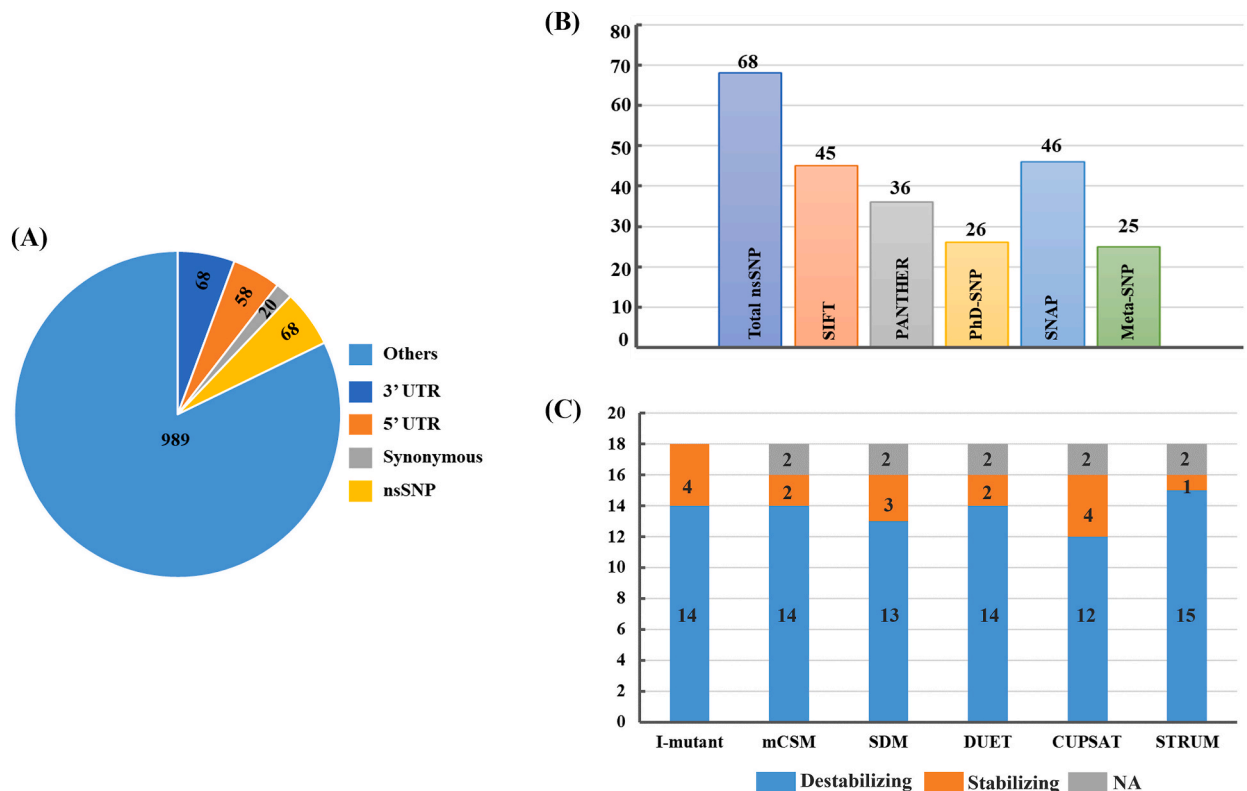
were validated using the SAVEs server (S1 Figure).

## 2.6. Molecular dynamics simulation

To understand the structural impact of these high-risk nsSNPs on the BLG protein, we implemented an all-atom molecular dynamics (MD) simulation to the WT and its selected variants. The MD simulation was performed using the GROMACS 2022.2 package, and the topology for all complexes was generated using the AMBER-ff99SBILDN force field. A dodecahedron periodic boundary with 1.0 nm was filled with TIP3P water molecules, and  $\text{Na}^+/\text{Cl}^-$  counter ions were added to neutralize the system using the *gmx genion* module of GROMACS. Next, all WT and variant complexes were minimized using the steepest-descent algorithm with a maximum tolerance of 1000 kJ/mol/nm. Subsequently, all systems were equilibrated at a temperature of 300 K using a Berendsen thermostat (V-rescale) and a pressure of 1.0 bar using the Parrinello-Rahman barostat coupling algorithm. Both the NVT and NPT ensemble processes were performed at 100 ps and 500 ps, respectively. The Particle Mesh Ewald (PME) algorithm was used to sustain long-range interactions, and a cutoff of 1 nm was applied to maintain short and long-range interactions. In addition, the LINCS algorithm was used to restrain all bonds. Finally, all systems were subjected to a 500 ns MD production run with a 2 fs time step, and 2 ps coordinates were saved throughout the MD trajectory. Trajectory information was assessed to calculate the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent-accessible surface area (SASA), and H-bonds. However, the essential dynamics of all systems were analyzed using the *gmx covar* and *gmx analog* modules of GROMACS by considering the last 400 ns trajectories. Furthermore, the Gibb's free energy landscape (FEL) was generated using the first two principal components (PCs) from principal component analysis (PCA), as reported in our previous studies [44–46].

## 2.7. Protein-protein interaction (PPI) network

We used the search tool for the retrieval of interacting genes/proteins (STRING) database web server (<https://string-db.org/>) to generate a protein-protein interaction network for BLG. The STRING database is a comprehensive resource used to predict associations between genes and proteins. It covers a vast number of organisms, making it an essential tool for researchers studying various aspects of biology, including the complexities of metabolic pathways, signaling cascades, and mechanisms underlying various diseases. The PPI network for BLG was generated with a high confidence score of 0.700 and no more than 50 indicators in the first shell of the *Bos*



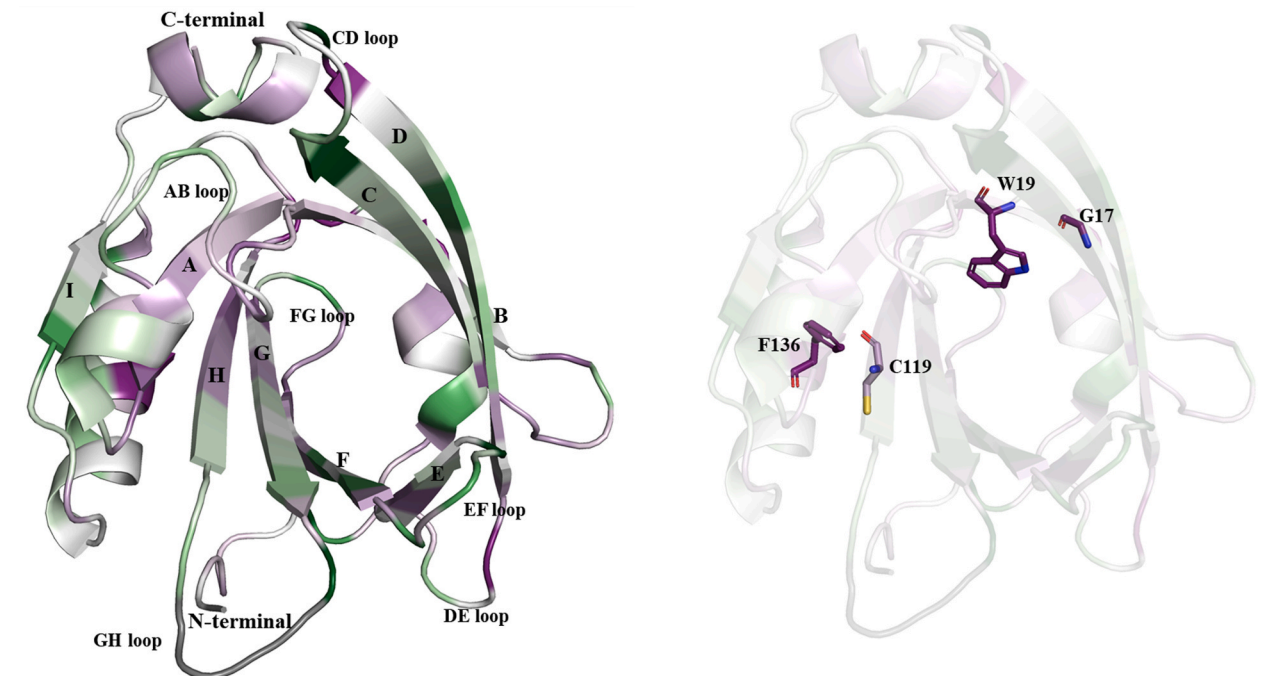
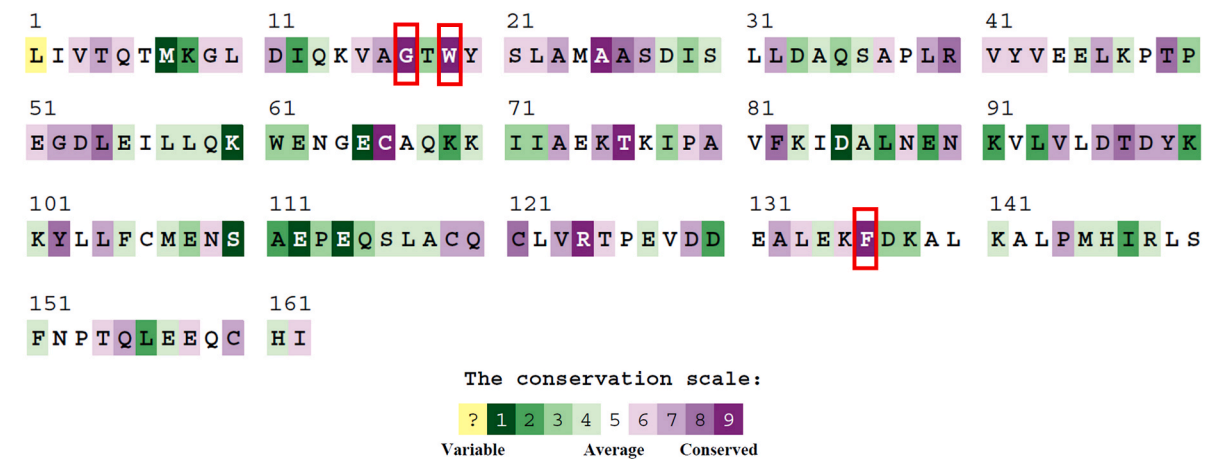
**Fig. 2.** (A) Pie chart showing the number of SNPs in BLG gene. Out of 1203 nsSNPs, 68 were identified to be non-synonymous, and 1135 were others (5' UTR: 58, 3' UTR: 68, synonymous SNPs: 20, downstream and upstream gene variants: 508, intron variants: 453, splice variants: 23, start lost and stop gained variants: 5). (B) Prediction of deleterious nsSNPs using different computational algorithms. (C) Prediction of destabilizing mutations using different computational methods.

taurus database.

### 3. Results

#### 3.1. Deleterious and destabilized nsSNPs identified in BLG

The nsSNPs, also known as missense variants, are recognized as the most significant SNPs that play a vital role in health and livestock industries by altering in vivo protein functions [47–49]. A total of 1203 SNPs were accessed from the Ensemble database, of which 68 SNPs were identified as non-synonymous (Fig 2A). All 68 nsSNPs were further subjected to five nsSNPs prediction algorithms: SIFT, PANTHER, SNAP, PhD-SNP, and Meta-SNP. The SIFT algorithm assigned a score or tolerance index (TI) to each substitution.  $TI > 0.05$  was considered tolerant (neutral), and  $TI < 0.05$  was considered intolerant (deleterious). Similarly, for the four tools, SNAP, PANTHER, PhD-SNP, and Meta-SNP, if the predicted score was  $>0.5$ , it was considered disease-causing or deleterious, and if the score was  $<0.5$ , it was considered neutral. Of the 68 nsSNPs, the distribution of deleterious nsSNPs in the BLG using SIFT, PANTHER, PhD-SNP, SNAP and Meta-SNP was 66.17 % (45), 52.94 % (36), 38.23 % (26), 67.65 % (46), and 36.76 % (25) respectively (Fig 2B). Out of the 68, common nsSNPs, 18 were predicted to be deleterious by all five tools, hence they were considered for further



**Fig. 3.** Evolutionary conservation profile of BLG gene. The red boxes on the BLG sequence indicate the most deleterious and destabilized conserved variants. The stick model represents the highly conserved variants of these residues (G17A, W19C, F136S, and C119R) that were considered for MD simulation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



analysis (S2 Table).

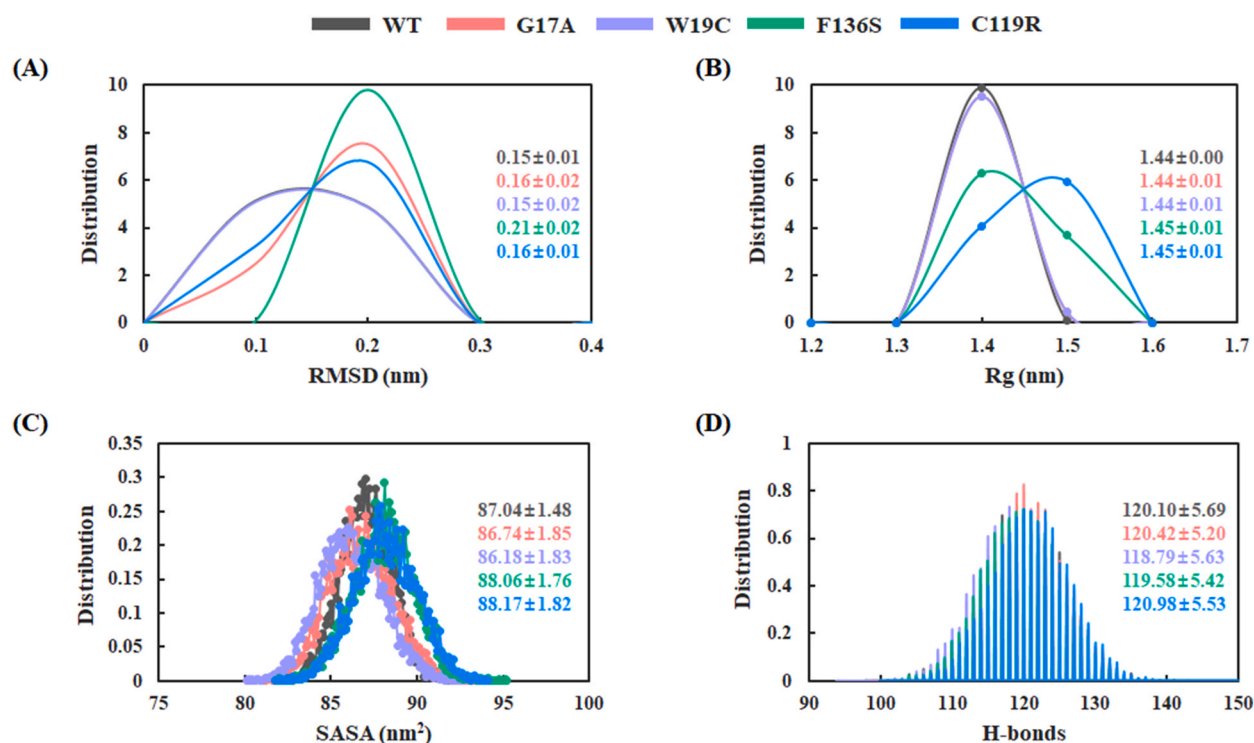
Furthermore, to identify destabilized nsSNPs, we uploaded the 18 identified deleterious variants to six bioinformatics tools. Six different tools, I-mutant, mCSM, SDM, DUET, CUPSAT, and STRUM, predicted 14, 14, 13, 14, 12, and 15 nsSNPs respectively, were destabilized (Fig 2C). Moreover, eight common nsSNPs were found to be destabilized by all six tools (S3 Table). Additionally, we used the MutPred2 tool to predict the pathogenicity of the selected variants. A MutPred2 score  $>0.5$  was considered pathogenic, and a score  $<0.5$ , benign. Nevertheless, we discovered that each of the 18 nsSNPs was pathogenic.

### 3.2. Evolutionary conservation profile of BLG

The residual evolutionary conservation profile of a protein is crucial for identifying its structural and functional impacts upon mutations. This provides a better understanding of the significance of particular amino acid residues and their restricted evolution [50]. Hence, the conservation level was assessed using the ConSurf database with the 3D structure of BLG as an input query. The ConSurf results showed that three of the eight variants selected from the protein-destabilizing analysis, G17, W19, and F136, were highly conserved, with a conservation score of 9 (Fig 3). The conservation scores of the other five variants, V15, M24, L39, L54, and C119, were calculated to be 6, 5, 7, 8, and 7, respectively. Since the C119 residue of BLG formed a disulfide bond with residue C106, any mutation at position 119 may interrupt the stability of the protein and its function. Hence, of the eight most deleterious and destabilizing nsSNPs, four variants, (i.e., three highly conserved variants, G17A, W19C, and F136S, along with the C119R variant) were considered for further analysis.

### 3.3. Structural consequences through MD simulation

All selected variants and the WT structure of BLG were subjected to MD simulations to study their structural and dynamic behaviors. As described in the methodology, a 500 ns MD simulation was performed, and the structural stability was assessed using the RMSD and Rg values of the protein backbone. The RMSD analysis suggested that after the first 100 ns of MD simulation, all variants maintained their structural stability with an RMSD value between 0.12 and 0.25 nm. In contrast to the other variants, the F136S variant showed a larger RMSD of more than 0.2 nm (S2 Figure). Moreover, we calculated the probability density distribution of RMSD, which showed an average RMSD value of  $0.15 \pm 0.01$ ,  $0.16 \pm 0.02$ ,  $0.15 \pm 0.02$ ,  $0.21 \pm 0.02$ , and  $0.16 \pm 0.01$  nm for WT, G17A, W19C, F136S, and C119R respectively. The WT and W19C variants of BLG showed similar unimodal distributions, whereas the F136S



**Fig. 4.** The Probability density distribution of (A) RMSD, (B) Rg, (C) SASA, and (D) intra H-bonds throughout the trajectories. The average and standard deviation values are represented according to their respective colors. Black, red, purple, green, and blue colors represent the WT, G17A, W19C, F136S, and C119R systems of BLG, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

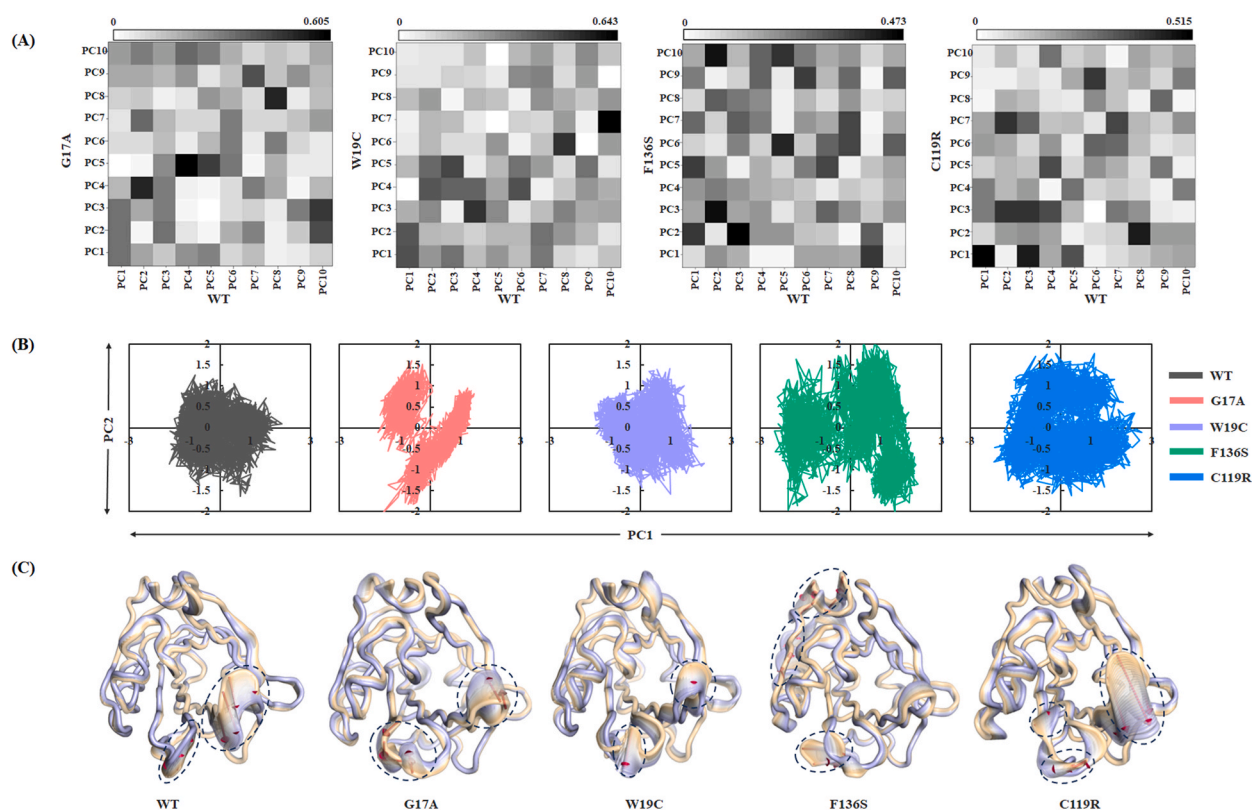
variant showed a distinct RMSD distribution (Fig 4A). Furthermore, the C119R variant showed relatively higher Rg values, indicating lower stability than other variants. The probability density distribution of Rg for WT and variants clustered at  $0.14 \pm 0.05$ ,  $0.14 \pm 0.07$ ,  $0.14 \pm 0.07$ ,  $0.15 \pm 0.07$ , and  $0.15 \pm 0.08$  nm for WT, G17A, W19C, F136S, and C119R respectively (Fig 4B). Based on the Rg distribution, no significant structural changes were observed in the G17A and W19C variants with respect to the WT BLG structure. In contrast, the F136S and C119R variants exhibited higher Rg values than those of the WT.

We also calculated the SASA to analyze the accessible surface area for biomolecular interactions and represented in S2 Figure. The G17A and W19C variants of BLG showed a similar type of distribution with an average SASA value of  $86.74 \pm 1.85$  nm<sup>2</sup> and  $86.18 \pm 1.83$  nm<sup>2</sup>, respectively. The F136S and C119R variants had relatively higher average SASA value than others (Fig 4D). To predict the effect of BLG variants on intra-molecular interactions, we calculated the number of intra H-bond interactions throughout the trajectory as, represented in Fig 4D. The results revealed that the average number of intra H-bond interactions for WT, G17A, W19C, F136S, and C119R was approximately 120, 120, 119, 120, and 121, respectively (S2 Figure).

### 3.4. Essential dynamic analysis of BLG variants

The collective conformational changes, particularly induced by these variants, were studied using PCA. Here, we considered the last 400 ns of trajectories for PCA and identified the principal components (PCs) with larger eigenvalues, which represent the most dominant motions of the protein backbone. The results revealed diagonalized co-variances of 2.81, 3.09, 2.60, 5.13, and 4.26 for the WT, G17A, W19C, F136S, and C119R variants, respectively. Additionally, the first 10 PCs, which accounted for more than 60 % of the total covariance, were used to define the similarities and differences between the essential subspace of the WT and variants through the root mean square inner product (RMSIP). The RMSIP calculation variants showed normalized overlaps of 0.517, 0.573, 0.415, and 0.490 for variants G17A, W19C, F136S, and C119R, respectively, with respect to the WT BLG (Fig 5A). In particular, the RMSIP of the first three PCs differed significantly between the F136S and C119R variants.

The first three PCs represented most of the dominant motions with a cumulative percentage of 31.91 %, 42.46 %, 32.91 %, 48.76 %, and 41.64 % for the WT, G17A, W19C, F136S, and C119R variants, respectively, which indicated that all variants had higher collective



**Fig. 5.** Conformation changes induced by BLG variants assessed through PCA. (A) Root Mean Square Inner Products (RMSIPs) of the first ten PCs show similarities and differences between the conformational spaces of the WT and variants. Black, red, purple, green, and blue colors represent the WT, G17A, W19C, F136S, and C119R systems of BLG, respectively. (B) Projection of first two PCs for all the systems. (C) Structural changes in all the systems with respect to PC1 are represented using extreme conformations with 30 frames. Black-dashed circles represent large conformational changes, and red arrows represent the size and direction of movements. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

flexibility compared to the WT (S3 Figure A, B). Hence, the first two PCs from the PCA were assessed to plot a 2D projection representing the spread in phase space (Fig 5B). Moreover, we assessed the most dominant conformational changes in the collective motion of all systems by considering 30 extreme conformations from the first PC. The results revealed that all systems showed a larger structural alteration in the GH loop region (Fig 5C). The W19C and C119R variants, along with the WT complex of BLG, exhibited similar kinds of conformational changes in the EF loop region by mimicking the Tanford transition. In particular, in the C119R variant, the EF loop was exposed to the solvent by opening access to the central calyx of the BLG. However, the G17A variant causes closure of the EF loop by blocking access to the central calyx. Although the F136S variant did not show any significant changes in the EF loop region, it exhibited large conformational changes in the main  $\alpha$ -helix and near C-terminal region (Fig 5C).

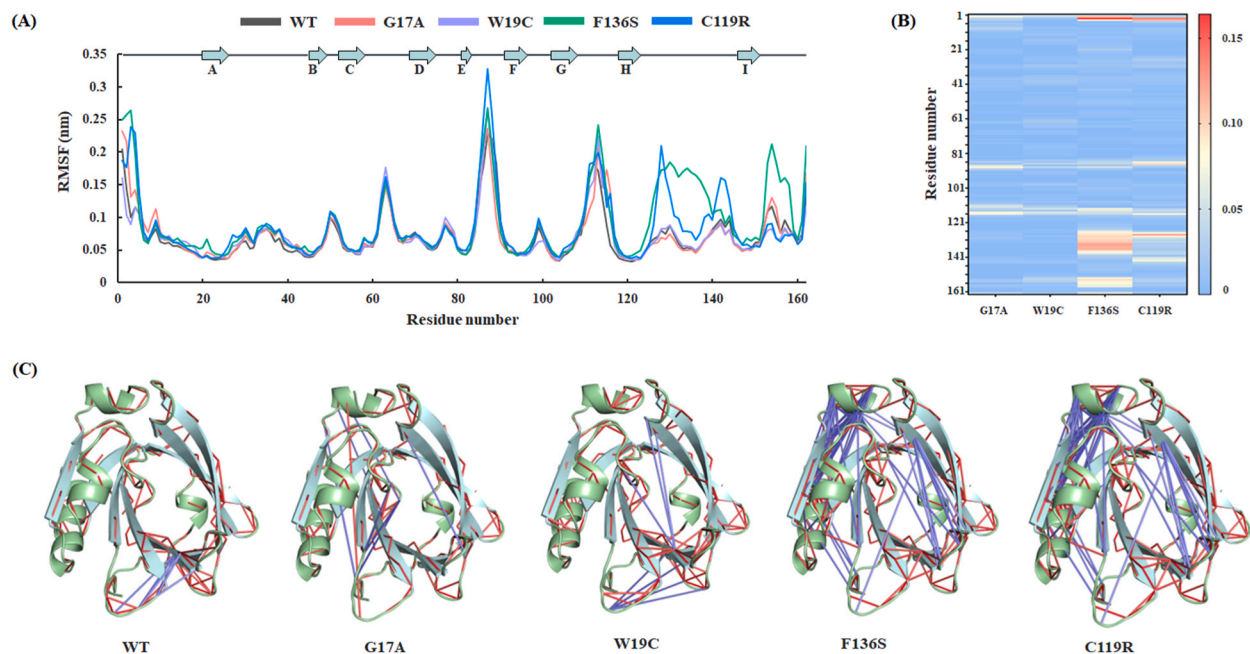
### 3.5. Changes in residual fluctuation and residual cross-correlation

RMSF analysis was performed to assess the residual fluctuation of each amino acid throughout the trajectory and to identify the region that influences structural changes and stability. Fig 6A shows that the F136S and C119R variants increased the residual fluctuation compared to the others, and these fluctuations were pronounced in the EF and GH loop regions. Unlike other variants, the F136S variant showed residual fluctuation in the main  $\alpha$ -helix region and near the C terminal region, indicating that this variant may interrupt the intramolecular signaling required for the stability of functional regions like the EF loop region. The average residual fluctuation for WT, G17A, W19C, F136S, and C119R variants was calculated to be 0.07, 0.07, 0.07, 0.09, and 0.08 nm, respectively. Additionally, the  $\Delta$ RMSF value for all the variants with respect to the WT were calculated and represented in Fig 6B. The lower  $\Delta$ RMSF value represents the rigid region with blue color, and the higher  $\Delta$ RMSF value represents the flexible region with dark red color. The heatmap of  $\Delta$ RMSF clearly indicates the residual flexibility in the F136S and C119R variants.

Moreover, the internal dynamics of the WT and other BLG variants were studied using a dynamic cross-correlation map (DCCM). The last 400 ns of MD trajectories of each system were considered to construct a DCCM map using the C $\alpha$  atoms and represented in (S3 Figure C). The cyan color in the DCCM map represents positively correlated motions (+1.0), and the pink color represents negatively correlated motions (-1.0). Furthermore, the structural residual correlation patterns in all systems were assessed and are presented in Figure: 6C. All systems showed similar types of correlated motions, however, the F136S and C119 variants showed larger anti-correlated motions, than the others. The EF loop region, which exhibited larger RMSF fluctuations, was also observed to have anti-correlated motions with the "C"  $\beta$ -sheet of BLG. In contrast, most of the anti-correlated motions were observed in the main  $\alpha$ -helix of F136S and C119R variants.

### 3.6. Free energy landscape (FEL) analysis of all the variants

FEL maps were generated for all complexes by assessing the first two PCs as shown in Fig 7A. FEL analysis revealed more than two



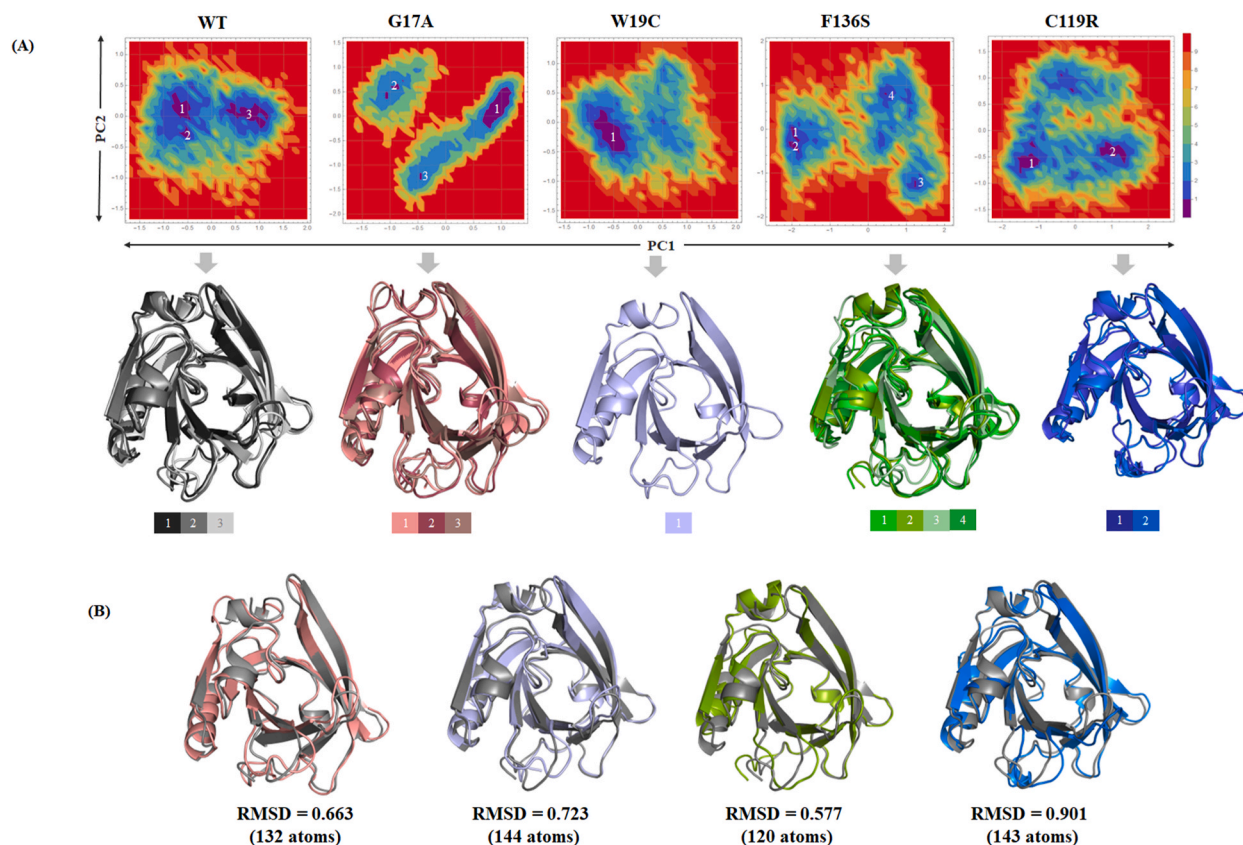
**Fig. 6.** Residual flexibility throughout the MD simulation. (A) RMSF of backbone atoms throughout the trajectories. (B) Heatmap representing  $\Delta$ RMSF of G17A, W19C, F136S, and C119R relative to WT. (C) Structural representation of the strong residual cross-correlation in all the systems. Red and blue lines denote the most correlated and anti-correlated interactions between two residues. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



global minima clusters in all variants and the WT, except for the W19C variant. The W19C variants showed a single global minima cluster represented by a deep blue color. In contrast to WT, the variants showed evidence of a substantial transition during MD simulations. In particular, the F136S and C119R variants underwent large conformational changes with widely distributed global minima. Furthermore, representative structures from each global minima were retrieved and superimposed to obtain a clear picture of the structural transition. Additionally, the representative structures of each variant were superimposed with respect to the representative WT structure of BLG (Fig 7B). The superimposition results revealed an RMSD of 0.663 Å (132 atoms), 0.723 Å (144 atoms), 0.577 Å (120 atoms), and 0.901 Å (143 atoms) for G17A, W19C, F136S, and C119R respectively, with respect to WT structure. From the superimposition, the molecular interactions of each variant with other amino acids were also analyzed with respect to the WT, as shown in (S4 Figure A). No differences were observed in the interaction pattern upon the substitution of glycine with alanine at position 17, and they formed a common interaction with the L46 residue. Similarly, the W19 residue interacts with E44 and V15; however, when substituted with cysteine, it forms only one interaction with the E44 residue. The C119 residue formed one hydrogen bond and one disulfide bond with the A25 and C106 residues, respectively, but when it mutated to arginine, it formed two hydrogen bonds with the A25 and Q5 residues. Moreover, the F136S variants exhibited large differences in the interaction pattern, suggesting that after mutating to serine, there were two more H-bond interactions with L133 and D137. The increase in the number of hydrogen bonds in the F136S variant may reflect a larger conformational transition in the main helical region of BLG. All the above analyses indicated that the F136S and C119R variants may induce conformational changes, particularly in the main helix and EF loop regions of BLG. Additionally, the G17A variant may induce structural changes in the EF-loop and block access to the central calyx.

### 3.7. Analysis of the PPI network

To assess the association between BLG and other proteins, a PPI network was generated using a high confidence score (0.700) in the STRING database (S4 Figure B). The PPI network for BLG (also known as PAEP protein) is comprised 18 nodes, 72 edges, and an average node degree of 8 with an average local clustering coefficient of 0.785. Proteins associated with BLG, and their corresponding scores are listed in S4 Table. The network predicted that the BLG protein was associated with different biological processes, such as response to 11-deoxycorticosterone (GO:1903496), response to dehydroepiandrosterone (GO:1903494), negative regulation of lactation (GO:1903488), response to progesterone (GO:0032570), response to growth hormone (GO:0060416), and response to estradiol (GO:0032355). Hence, any variation in BLG may have a significant impact on several biological processes because they may



**Fig. 7.** Free energy landscape of all the systems and their representative structures. (A) Low energy structures were retrieved from their respective global minima and superimposed. (B) Superimposition of WT and representative mutant conformations are shown with RMSD values.

act as key regulators.

#### 4. Discussion

BLG is the predominant whey protein in bovine milk, constituting about 50 % of the whey protein [9,51]. It plays a major role in the food and pharmaceutical industries due to its ability to bind hydrophobic molecules and vitamins, like retinol (Vitamin A), enhancing nutrient bioavailability [52–54]. Previous studies reported that at a neutral pH, the Tanford transition phase of BLG is of particular interest because it impacts its capacity to interact with both synthetic and natural ligands [10]. In recent years, researchers have become interested in this gene because of its critical role in healthy milk production, and genetic polymorphisms in this gene may affect milk quality. Among the different genetic polymorphisms, nsSNPs are of great significance because a single-base change in the coding region causes functional diversity in the protein by altering the amino acid residue. These nsSNPs may substantially affect the structural stability and solubility of proteins [18]. Preventing the occurrence of deleterious nsSNPs is crucial in efforts to enhance the quality of dairy products and address allergenicity concerns. Several strategies and techniques can be employed to prevent the expression of such SNPs in BLG. One of the most effective methods is to use selective breeding programs by identifying cattle with beneficial genetic variants of the BLG gene. Using Marker-Assisted Selection (MAS), breeders can select animals with favorable gene variants for reproduction, minimizing the passing of deleterious nsSNPs to future generation. Advanced gene editing technologies such as CRISPR-Cas9 have revolutionized genetic engineering. This system can be used to directly edit the harmful variants in the BLG gene, either correcting deleterious nsSNPs or introducing protective mutations to counteract potential damage. Other approaches such as maintaining genetic diversity, environmental and nutritional management can also contribute to reduce the risk and impact of deleterious nsSNPs in BLG. To improve the efficiency of BLG despite of having damaging nsSNPs requires focus on counteracting the negative effects of these mutations. Strategies like directed protein engineering, molecular chaperon-assisted folding, gene editing can enhance BLG function. Furthermore, predictive modelling and MD simulations can identify the unstable regions, enabling targeted interventions like mutagenesis or small molecule design to improve the protein stability and function. However, to understand the clinical significance of these nsSNPs, it is important to prioritize them based on their functional impact. Given the challenges in experimentally assessing the impact of these variations, the current study utilized a series of bioinformatics tools to identify the most deleterious and damaging nsSNPs and assessed their structural and functional impact on BLG.

Using five different tools SIFT, PANTHER, Phd-SNP, Meta-SNP, and SNAP, 18 nsSNPs were identified as the most damaging and deleterious. Furthermore, these 18 nsSNPs were subjected to six different bioinformatics algorithms (I-mutant, mCSM, SDM, DUET, CUPSAT, and STRUM) to identify destabilized nsSNPs in BLG. As a result, eight common nsSNPs were found to be destabilized by all six algorithms. A series of algorithms, including both sequence and structure-based approaches, were used to optimize the prediction accuracy because it is important to note that the prediction of damaging and deleterious nsSNPs using a single tool may lead to false positives. Hence, based on the prediction results of all different tools and residual conservation profiles, four high-risk nsSNPs, G17A, W19C, F136S, and C119R, along with the WT structure of BLG, were selected for MD simulation to study the conformational changes induced by these variants.

The overall MD simulation results suggested that the WT and W19C variants of BLG maintained similar structural integrity throughout the simulation, indicating higher structural stability than the other variants. Moreover, the F136S and C119R variants showed distinct unimodal distributions, indicating lower structural and conformational stability than the other variants. The disulfide bond formed between C119 and C106 preserved the structural integrity and stability of BLG. However, no disulfide bond was formed with C106 when the cysteine residue was substituted with arginine at position 119. This could explain why the C119R variant exhibited lower structural stability than the other variants. The essential dynamics results suggest that the C119R and G17A variants caused severe conformational alterations in the EF loop region, mimicking the deprotonated and protonated states of E89, respectively [10]. Unlike other variants, the F136S variant exhibited larger conformational changes in the main  $\alpha$ -helix and near C-terminal region of BLG. Residual flexibility analysis using RMSF and  $\Delta$ RMSF revealed that all variants exhibited common fluctuations in the EF loop and GH loop region. However, the F136S and C119 variants exhibited distinct residual fluctuations in the main  $\alpha$ -helix residues, which were later observed to exhibit anti-correlated motions.

Although the current study provides detailed structural insights into the impact of high-risk nsSNPs on the BLG, further structural and biochemical comparative analyses of these variants are required to validate our results. Additionally, long-range MD simulations and extensive sampling techniques are required to be performed to come up with definitive conclusions.

#### 5. Conclusion

In the present study, a series of bioinformatics tools were used to identify four high-risk nsSNPs, G17A (rs437398769), W19C (rs474719380), F136S (rs440613020), and C119R (rs3423321025) in the BLG gene from the Ensemble database, which induced major structural changes revealed by molecular dynamic simulation. In particular, the F136S and C119R variants exhibited a larger conformational change in the main  $\alpha$ -helix and EF loop of BLG, respectively, indicating lower structural stability. Moreover, the G17A variant induces structural changes in the EF loop region, which causes the EF loop of the  $\beta$ -barrel structure to close, blocking ligand access to the central calyx of the BLG. Although additional experimental validation is needed to confirm these variants, we hope these results will guide future studies and assist in improving milk quality and the management of cattle milk production.

## CRediT authorship contribution statement

**Shhitaprajna Sahoo:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Vijayakumar Gosu:** Writing – review & editing. **Hak-Kyo Lee:** Writing – review & editing, Supervision. **Donghyun Shin:** Writing – review & editing, Supervision, Conceptualization.

## Data availability statement

The raw simulations trajectory data of this study is publicly available on FigShare (<https://figshare.com/>) with accession number (<https://doi.org/10.6084/m9.figshare.26183426.v1>).

## Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022R1A2C4002510) and Science and Technology Project Opens the Future of the Region through INNOPOLIS Foundation grant funded by the Ministry of Science and ICT of Korea (2022-DD-UP-0333).

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The authors thank the Jeonbuk National University for providing the facilities necessary for this study.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e40040>.

## References

- [1] S. Patel, Emerging trends in nutraceutical applications of whey protein and its derivatives, *J. Food Sci. Technol.* 52 (2015) 6847–6858, <https://doi.org/10.1007/s13197-015-1894-0>.
- [2] P.R. Majhi, R.R. Ganta, R.P. Vanam, E. Seyrek, K. Giger, P.L. Dubin, Electrostatically driven protein aggregation:  $\beta$ -lactoglobulin at low ionic strength, *Langmuir* 22 (2006) 9150–9159, <https://doi.org/10.1021/la053528w>.
- [3] L. Sawyer, G. Kontopidis, The core lipocalin, bovine  $\beta$ -lactoglobulin, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 1482 (2000) 136–148, [https://doi.org/10.1016/S0167-4838\(00\)00160-6](https://doi.org/10.1016/S0167-4838(00)00160-6).
- [4] S. Brownlow, J.H.M. Cabral, R. Cooper, D.R. Flower, S.J. Yewdall, I. Polikarpov, A.C. North, L. Sawyer, Bovine  $\beta$ -lactoglobulin at 1.8 Å resolution — still an enigmatic lipocalin, *Structure* 5 (1997) 481–495, [https://doi.org/10.1016/S0969-2126\(97\)00205-0](https://doi.org/10.1016/S0969-2126(97)00205-0).
- [5] G. Kontopidis, C. Holt, L. Sawyer, *Invited review:  $\beta$ -lactoglobulin: binding properties, structure, and function*, *J. Dairy Sci.* 87 (2004) 785–796, [https://doi.org/10.3168/jds.S0022-0302\(04\)73222-1](https://doi.org/10.3168/jds.S0022-0302(04)73222-1).
- [6] E.G. Varlamova, O.G. Zaripov, Beta-lactoglobulin–nutrition allergen and nanotransporter of different nature ligands therapy with therapeutic action, *Res. Vet. Sci.* 133 (2020) 17–25, <https://doi.org/10.1016/j.rvsc.2020.08.014>.
- [7] Z. Shafaei, B. Ghalandari, A. Vaseghi, A. Divsalar, T. Haertlé, A.A. Saboury, L. Sawyer,  $\beta$ -Lactoglobulin: an efficient nanocarrier for advanced delivery systems, *Nanomed. Nanotechnol. Biol. Med.* 13 (2017) 1685–1692, <https://doi.org/10.1016/j.nano.2017.03.007>.
- [8] K. Kuwata, M. Hoshino, V. Forge, S. Era, C.A. Batt, Y. Goto, Solution structure and dynamics of bovine  $\beta$ -lactoglobulin A, *Protein Sci.* 8 (1999) 2541–2545, <https://doi.org/10.1110/ps.8.11.2541>.
- [9] A. Barbiroli, S. Iametti, F. Bonomi, Beta-lactoglobulin as a model food protein: how to promote, prevent, and exploit its unfolding processes, *Molecules* 27 (2022) 1131, <https://doi.org/10.3390/molecules27031131>.
- [10] B.Y. Qin, M.C. Bewley, L.K. Creamer, H.M. Baker, E.N. Baker, G.B. Jameson, Structural basis of the Tanford transition of bovine  $\beta$ -lactoglobulin, *Biochemistry* 37 (1998) 14014–14023, <https://doi.org/10.1021/bi981016t>.
- [11] S. Geng, Z. Jiang, H. Ma, Y. Wang, B. Liu, G. Liang, Interaction mechanism of flavonoids and bovine  $\beta$ -lactoglobulin: experimental and molecular modelling studies, *Food Chem.* 312 (2020) 126066, <https://doi.org/10.1016/j.foodchem.2019.126066>.
- [12] A. Yousefi, S. Ahrari, F. Panahi, Y. Ghasemi, R. Yousefi, Binding analysis of the curcumin-based synthetic alpha-glucosidase inhibitors to beta-lactoglobulin as potential vehicle carrier for antidiabetic drugs, *J. Iran. Chem. Soc.* 19 (2022) 489–503, <https://doi.org/10.1007/s13738-021-02323-8>.
- [13] L. Liang, M. Subirade, Study of the acid and thermal stability of  $\beta$ -lactoglobulin–ligand complexes using fluorescence quenching, *Food Chem.* 132 (2012) 2023–2029, <https://doi.org/10.1016/j.foodchem.2011.12.043>.
- [14] K. Sakai, K. Sakurai, M. Sakai, M. Hoshino, Y. Goto, Conformation and stability of thiol-modified bovine  $\beta$  lactoglobulin, *Protein Sci.* 9 (2000) 1719–1729.
- [15] L. Wodas, M. Mackowski, A. Borowska, K. Puppel, B. Kuczynska, J. Cieslak, Genes encoding equine  $\beta$ -lactoglobulin (LGB1 and LGB2): polymorphism, expression, and impact on milk composition, *PLoS One* 15 (2020) e0232066, <https://doi.org/10.1371/journal.pone.0232066>.
- [16] M. Arshad, A. Bhatti, P. John, Identification and in silico analysis of functional SNPs of human TAGAP protein: a comprehensive study, *PLoS One* 13 (2018) e0188143, <https://doi.org/10.1371/journal.pone.0188143>.
- [17] E.S. Lander, The new genomics: global views of biology, *Science* 274 (1996) 536–539, <https://doi.org/10.1126/science.274.5287.536>.

- [18] N. Tokuriki, D.S. Tawfik, Stability effects of mutations and protein evolvability, *Curr. Opin. Struct. Biol.* 19 (2009) 596–604, <https://doi.org/10.1016/j.sbi.2009.08.003>.
- [19] R. Kaur, J. Singh, M. Kaur, Structural and functional impact of SNPs in P-selectin gene: a comprehensive in silico analysis, *Open Life Sci.* 12 (2017) 19–33, <https://doi.org/10.1515/biol-2017-0003>.
- [20] M. Krawczak, E.V. Ball, I. Fenton, P.D. Stenson, S. Abeyasinghe, N. Thomas, D.N. Cooper, Human gene mutation database-a biomedical information and research resource, *Hum. Mutat.* 15 (2000) 45–51, [https://doi.org/10.1002/\(SICI\)1098-1004\(200001\)15:1<45::AID-HUMU10>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1098-1004(200001)15:1<45::AID-HUMU10>3.0.CO;2-T).
- [21] L. Prokunina, M.E. Alarcón-Riquelme, Regulatory SNPs in complex diseases: their identification and functional validation, *Exp. Rev. Mol. Med.* 6 (2004) 1–15, <https://doi.org/10.1017/S1462399404007690>.
- [22] C.O. Okoye, H. Jiang, M. Nazar, X. Tan, J. Jiang, Redefining modern food analysis: significance of omics analytical techniques integration, chemometrics and bioinformatics, *TrAC, Trends Anal. Chem.* 175 (2024) 117706, <https://doi.org/10.1016/j.trac.2024.117706>.
- [23] H. Manoochehri, S. Asadi, H. Tanzadehpahan, M. Sheykhasan, M. Ghorbani, CDC25A is strongly associated with colorectal cancer stem cells and poor clinical outcome of patients, *Gene Reports* 25 (2021) 101415, <https://doi.org/10.1016/j.genrep.2021.101415>.
- [24] M. Sheykhasan, Y. Ahmadyousefi, R. Seyedebrahimi, H. Tanzadehpahan, H. Manoochehri, P. Dama, N.F. Hosseini, M. Akbari, M.E. Farsani, DLX6-AS1: a putative lncRNA candidate in multiple human cancers, *Exp. Rev. Mol. Med.* 23 (2021) e17, <https://doi.org/10.1017/erm.2021.17>.
- [25] H. Tanzadehpahan, A. Bahmani, N. Hosseinpour Moghadam, H. Gholami, H. Mahaki, A. Farmany, M. Saidijam, Synthesis, anticancer activity, and  $\beta$ -lactoglobulin binding interactions of multitargeted kinase inhibitor sorafenib sorafenib sorafenib (SORT) using spectroscopic and molecular modelling approaches, *Luminescence* 36 (2021) 117–128, <https://doi.org/10.1002/bio.3929>.
- [26] N.-L. Sim, P. Kumar, J. Hu, S. Henikoff, G. Schneider, P.C. Ng, SIFT web server: predicting effects of amino acid substitutions on proteins, *Nucleic Acids Res.* 40 (2012) W452–W457, <https://doi.org/10.1093/nar/gks539>.
- [27] E. Capriotti, P. Pariselli, PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants, *Nucleic Acids Res.* 45 (2017) W247–W252, <https://doi.org/10.1093/nar/gkx369>.
- [28] H. Mi, D. Ebert, A. Muruganujan, C. Mills, L.-P. Albu, T. Mushayama, P.D. Thomas, PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API, *Nucleic Acids Res.* 49 (2021) D394–D403, <https://doi.org/10.1093/nar/gkaa1106>.
- [29] Y. Bromberg, B. Rost, SNAP: predict effect of non-synonymous polymorphisms on function, *Nucleic Acids Res.* 35 (2007) 3823–3835, <https://doi.org/10.1093/nar/gkm238>.
- [30] E. Capriotti, R.B. Altman, Y. Bromberg, Collective judgment predicts disease-associated single nucleotide variants, *BMC Genom.* 14 (2013) S2, <https://doi.org/10.1186/1471-2164-14-S3-S2>.
- [31] E. Capriotti, P. Pariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res.* 33 (2005) W306–W310, <https://doi.org/10.1093/nar/gki375>.
- [32] D.E.V. Pires, D.B. Ascher, T.L. Blundell, DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach, *Nucleic Acids Res.* 42 (2014) W314–W319, <https://doi.org/10.1093/nar/gku411>.
- [33] D.E.V. Pires, D.B. Ascher, T.L. Blundell, mCSM: predicting the effects of mutations in proteins using graph-based signatures, *Bioinformatics* 30 (2014) 335–342, <https://doi.org/10.1093/bioinformatics/btt691>.
- [34] A.P. Pandurangan, B. Ochoa-Montano, D.B. Ascher, T.L. Blundell, SDM: a server for predicting effects of mutations on protein stability, *Nucleic Acids Res.* 45 (2017) W229–W235, <https://doi.org/10.1093/nar/gkx439>.
- [35] L. Quan, Q. Lv, Y. Zhang, STRUM: structure-based prediction of protein stability changes upon single-point mutation, *Bioinformatics* 32 (2016) 2936–2946, <https://doi.org/10.1093/bioinformatics/btw361>.
- [36] V. Parthiban, M.M. Gromiha, D. Schomburg, CUPSAT: prediction of protein stability upon point mutations, *Nucleic Acids Res.* 34 (2006) W239–W242, <https://doi.org/10.1093/nar/gkl190>.
- [37] M. Delorenzi, T. Speed, An HMM model for coiled-coil domains and a comparison with PSSM-based predictions, *Bioinformatics* 18 (2002) 617–625, <https://doi.org/10.1093/bioinformatics/18.4.617>.
- [38] A. Krogh, B. Larsson, G. von Heijne, E.L.L. Sonnhammer, Predicting transmembrane protein topology with a hidden markov model: application to complete genomes, *J. Mol. Biol.* 305 (2001) 567–580, <https://doi.org/10.1006/jmbi.2000.4315>.
- [39] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402, <https://doi.org/10.1093/nar/25.17.3389>.
- [40] R.D. Finn, P. Coghill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: towards a more sustainable future, *Nucleic Acids Res.* 44 (2016) D279–D285, <https://doi.org/10.1093/nar/gkv1344>.
- [41] 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>.
- [42] A. Ben Chorin, G. Masrati, A. Kessel, A. Narunsky, J. Sprinzak, S. Lahav, H. Ashkenazy, N. Ben-Tal, ConSurf-DB: an accessible repository for the evolutionary conservation patterns of the majority of PDB proteins, *Protein Sci.* 29 (2020) 258–267, <https://doi.org/10.1002/pro.3779>.
- [43] J. Loch, A. Polit, A. Görecki, P. Bonarek, K. Kurpiewska, M. Dziedzicka-Wasyłewska, K. Lewiński, Two modes of fatty acid binding to bovine  $\beta$ -lactoglobulin—crystallographic and spectroscopic studies, *J. Mol. Recogn.* 24 (2011) 341–349, <https://doi.org/10.1002/jmr.1084>.
- [44] S. Sahoo, S. Son, H.-K. Lee, J.-Y. Lee, V. Gosu, D. Shin, Impact of nsSNPs in human AIM2 and IFI16 gene: a comprehensive in silico analysis, *J. Biomol. Struct. Dyn.* 0 (2023) 1–13, <https://doi.org/10.1080/07391102.2023.2206907>.
- [45] S. Sahoo, H.-K. Lee, D. Shin, Structure-based virtual screening and molecular dynamics studies to explore potential natural inhibitors against 3C protease of foot-and-mouth disease virus, *Front. Vet. Sci.* 10 (2024). <https://www.frontiersin.org/articles/10.3389/fvets.2023.1340126>. (Accessed 23 January 2024).
- [46] S. Sahoo, M. Samantaray, M. Jena, V. Gosu, P.P. Bhuyan, D. Shin, B. Pradhan, In Vitro and in silico studies to explore potent antidiabetic inhibitor against human pancreatic alpha-amylase from the methanolic extract of the green microalga *Chlorella vulgaris*, *J. Biomol. Struct. Dyn.* 42 (2024) 8089–8099, <https://doi.org/10.1080/07391102.2023.2244592>.
- [47] L.R. Pal, J. Moul, Genetic basis of common human disease: insight into the role of missense SNPs from genome-wide association studies, *J. Mol. Biol.* 427 (2015) 2271–2289, <https://doi.org/10.1016/j.jmb.2015.04.014>.
- [48] Y. Han, R. Dorajoo, X. Chang, L. Wang, C.-C. Khor, X. Sim, C.-Y. Cheng, Y. Shi, Y.C. Tham, W. Zhao, M.L. Chee, C. Sabanayagam, M.L. Chee, N. Tan, T.Y. Wong, E.-S. Tai, J. Liu, D.Y.T. Goh, J.-M. Yuan, W.-P. Koh, R.M. van Dam, A.F. Low, M.Y.-Y. Chan, Y. Friedlander, C.-K. Heng, Genome-wide association study identifies a missense variant at APOA5 for coronary artery disease in Multi-Ethnic Cohorts from Southeast Asia, *Sci. Rep.* 7 (2017) 17921, <https://doi.org/10.1038/s41598-017-18214-z>.
- [49] M.M. Kęsek-Woźniak, E. Wojtas, A.E. Zielak-Steciwo, Impact of SNPs in ACACA, SCD1, and DGAT1 genes on fatty acid profile in bovine milk with regard to lactation phases, *Animals* 10 (2020) 997, <https://doi.org/10.3390/ani10060997>.
- [50] V. Ramensky, P. Bork, S. Sunyaev, Human non-synonymous SNPs: server and survey, *Nucleic Acids Res.* 30 (2002) 3894–3900, <https://doi.org/10.1093/nar/gkf493>.
- [51] H. Deeth, N. Bansal, Chapter 1 - whey proteins: an overview, in: H.C. Deeth, N. Bansal (Eds.), *Whey Proteins*, Academic Press, 2019, pp. 1–50, <https://doi.org/10.1016/B978-0-12-812124-5.00001-1>.
- [52] M. Guo, G. Wang, Nutritional applications of whey protein, in: *Whey Protein Production, Chemistry, Functionality, and Applications*, John Wiley & Sons, Ltd, 2019, pp. 141–156, <https://doi.org/10.1002/9781119256052.ch6>.
- [53] S. Boscaini, P. Skuse, K.N. Nilaweera, J.F. Cryan, P.D. Cotter, The ‘Whey’ to good health: whey protein and its beneficial effect on metabolism, gut microbiota and mental health, *Trends Food Sci. Technol.* 133 (2023) 1–14, <https://doi.org/10.1016/j.tifs.2022.12.009>.
- [54] E. Ha, M.B. Zemel, Functional properties of whey, whey components, and essential amino acids: mechanisms underlying health benefits for active people, *J. Nutr. Biochem.* 14 (2003) 251–258, [https://doi.org/10.1016/S0955-2863\(03\)00030-5](https://doi.org/10.1016/S0955-2863(03)00030-5), submitted for publication.