Research article

Molecular dynamics simulation of wild and mutant proteasome subunit beta type 8 (PSMB8) protein: Implications for restoration of inflammation in experimental autoimmune encephalomyelitis pathogenesis

Shamrat Kumar Paul [a], Md Saddam [a], Nisat Tabassum [b], Mahbub Hasan [a],*

[a] Department of Biochemistry and Molecular Biology, Life Science Faculty, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, 8100, Bangladesh
[b] Department of Biotechnology and Genetic Engineering, Life Science Faculty, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, 8100, Bangladesh

ARTICLE INFO

ABSTRACT

Multiple Sclerosis (MS) is an autoimmune and chronic disease in the brain and spinal cord. MS has inflammatory progression characterized by its hallmark inflammatory plaques. The histological and clinical characteristics of MS are shared by Experimental Autoimmune Encephalomyelitis (EAE). Genetic and environmental factors contribute to the development of MS. In EAE-MS disease, the level of proteasome subunit beta type-8 (PSMB8), encoded by the PSMB8 gene, is increased and regulates the inflammatory response in this disease. In humans, the Nakajo-Nishimura Syndrome is caused by a mutation in the gene PSMB8, a part of the immunoproteasome subunit. Therefore, special attention to wild and mutant (G210V) PSMB8 protein is imperative. In this study, we performed a 100 ns molecular dynamics (MD) simulation for wild-type PSMB8 and the mutant G210V. Then, we analyzed the fundamental and essential simulation results using another Google Colab system. The energy analysis ensures the structural deviation due to point mutation. The trajectory of the fundamental simulation (RMSD, RMSF, and Rg) describes that the G210V mutated protein is more flexible and less stable than the wild type. We observed the conformational changes due to mutation by analyzing the RMSD average linkage hierarchical clustering, total SASA, and SASA autocorrelation. The differences in the protein's overall motion and the atoms' precise location are identified by the principal component analysis, showing that the overall motion and location of the atoms are different. Our study provides valuable insights into the dynamics and structure of this protein, which can aid in further understanding its biological functions and potential implications for disease.

1. Introduction

Multiple sclerosis (MS) is one of the most prevalent neurological disorders, which is an inflammatory autoimmune disease leading to macrophage and lymphocyte invasion in the central nervous system (CNS) [1–4]. Damages in the myelin sheath, the insulator

---

covering neurons and the spinal cord, initiate symptoms like diplopia, myopia, coordination disorder, myasthenia, mental problems, etc. [5,6]. The leading underlying cause of MS is unknown. However, susceptibility toward disease has been identified that links some genomic and environmental factors [7]. Environmental factors include smoking, ultraviolet B exposure, obesity, lower vitamin D level, and EVB infection [8,9]. Migration study also steered towards the environmental factor as a causal agent for MS development [10]. The study also pointed toward a higher susceptibility of children than adults for disease development during migration in a high-risk country [11] with high latitude. Less outdoor activity and decreased vitamin D strongly correlate with multiple sclerosis development [12]. EBV infection leads to B cell immortalization and contributes to a 2-fold increase in disease progression than in normal conditions [13]. Increased tobacco abuse and smoking in women have elevated disease susceptibility more than in men, resulting in a ratio of 3:1 (F: M) [14]. MS is also induced by immune system cells such as CD4$^+$ T cells, particularly those of the TH1 and TH17 subsets, macrophages, and soluble inflammatory mediators [15,16]. TH17 cells are responsible for the production of several cytokines, including IL-17, IL-21, and IL-22, all of which contribute to the development of MS by inducing the inflammation cascade, clonal expansion of T cells and microglia activation; IL-17 and IL-22 both contribute to an increase in the permeability of the blood-brain barrier, while IL-17 is also responsible for the synthesis of CXC chemokines, which entice neutrophils to cross the BBB (Fig. 1) [17–19].

Different patterns have been recognized in the progression of MS disease. About 15 % of the global population is affected by primary progressive MS (PPMS), which starts with the onset of disease accompanied by evidence of neurological impairment and the absence of recurrent overlapping relapses. Relapsing-remitting disease with progressive neurological impairment is a sign of secondary progressive MS (SPMS). Relapsing-remitting MS (RRMS) shows interspersed disease relapse phases and the absence of disease advancement. Lastly, the novel form, progressive relapsing MS (PRMS), manifests severe neurological impairment with no overlapping relapses [20].

A genomic study discovered an MS-associated correlation with over 150 single nucleotide polymorphisms [21]. Notably, TNFR1 [22], IL7R [23], CYP2R1 [24], IL2RA [25], and BAFF [26], are closely associated with immune function-related genes, mainly in their regulatory region. The prevailing risk of MS is with the HLA-DRB1*15 allele, the imbalance between the allele with other loci [27], and the hypothesis linking the allele having antigen presentation function [28]. T-helper (Th) cell-intervened autoimmune illness,
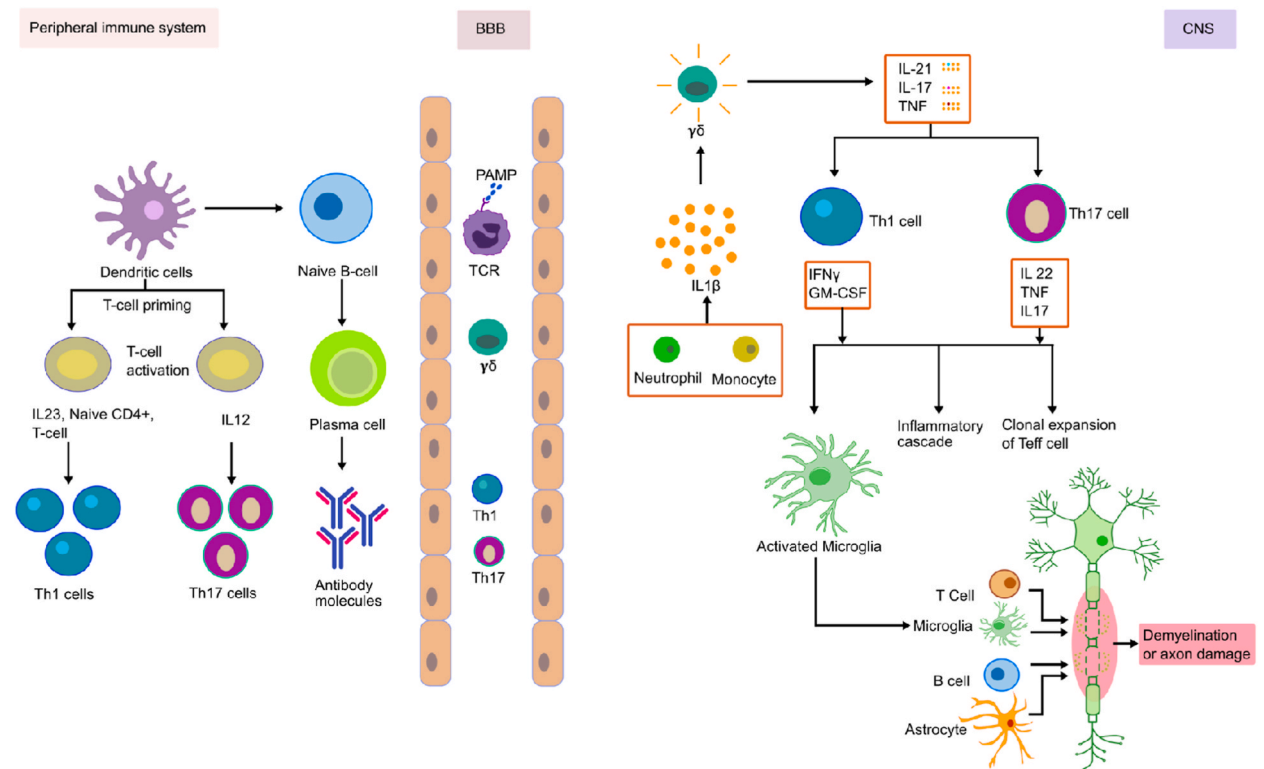

Fig. 1. Mechanism of myelin and axon damage. Antigen-presenting Cells, APCs (dendritic cells, B cells, and macrophages) in the peripheral immune system release IL-12 and IL-23. The engagement of the T cell receptor with the corresponding MHC Class II expressed antigen complexes with primary naive CD4$^+$ T and activates Th-1 and Th-17 cells. The DAMP/PAMP activates T cells and inflammatory myeloid after they are discharged into Peripheral blood circulation. Later, it migrates towards the BBB and infiltrates into the CNS. IL-1β, released from neutrophils and monocytes, activates γδ and eventually removes IL-17, IL-21, and TNF. At the same time, it heightens the Th17 activity. Th-17 releases TNF, IL-17, and IL2; at the same time, Th-1 cells discharge the GM-CSF and IFNγ. Th-17 and Th-1 both trigger microglia. Recognizing the myelin antigens on APC reactivates the T cell and stimulates the propagation of effector T cells (Teff). Cytokines, Th-1, and Th-17 released inflammatory mediators, reactive oxygen, NO, and protease activate myelin damage. Oligodendrocytes cause re-myelination of axons, starting a cascade of inflammation, neuron death, and neurological deficit.

accompanied by T-cell and monocyte invasion in the CNS, is known as experimental autoimmune encephalomyelitis (EAE). The CNS is linked to local inflammation. EAE is frequently used as a model for MS and has shown to be a valuable tool for disease pathophysiology and potential treatment approaches [29]. Proteins generated by myelin-producing oligodendrocytes in the CNS were discovered and used as autoimmune molecular targets. Primary axonal demyelination slowed down the conduction of axons throughout the CNS, resulting in gradual hind-limb immobility [30].

Nakajo-Nishimura Syndrome (NNS) is an autosomal recessive disorder that causes periodic fever, skin rash, partial muscular atrophy, and joint contracture. Patients with neurodegenerative disorder experience disease progression due to a mutation (G201V) in the PSMB8 gene, which encodes the proteasome subunit beta type 8 protein. This mutation explicitly affects the immunoproteasome subunit β5i [31,32]. The mutation G201V disrupts the β-sheet structure and reduces proteasome activity, accumulating ubiquitinated and oxidized proteins and increasing levels of inflammatory markers. These findings demonstrate that a mutation within a proteasome subunit can directly cause human disease and suggest that decreased proteasomal activity may contribute to inflammation [31]. In a previous study, we analyzed a potential small molecule inhibitor-binding site by molecular modeling and dynamics simulation of Grancalcin [33], a calcium-binding protein identified as a novel biomarker upregulated in EAE mice [34]. In this work, we conduct simulation experiments on both the wild-type and mutant mouse PSMB-8 proteins to investigate the significance of mutations in protein structure and dynamics. Based on our previous work, we are developing an improved understanding of autoimmune disease pathways by merging our findings with our current work on PSMB8.

2. Result and discussion

2.1. Structural analysis of psmb8 and G210V

According to the ProtParam results, PSMB8 and its corresponding mutant protein demonstrated that the total number of atoms, molecular weight, instability index, aliphatic index, and GRAVY of the mutant (G210V) are higher than those of its WT (3UNF). However, there was no discernible difference between these proteins regarding their isoelectric point and extinction coefficient (Table 1). The mutant protein (G210V) is less stable than the WT (3UNF), as the instability index of the mutant is higher than the WT (3UNF) version of PSMB8. According to the Cast P server results, the amino acid sequences 32–33, 48–52, 56, 63, 75–80, 164, and 197–199 reflect the potential active sites for the WT (3UNF) protein (Fig. 2A). The mutant system (Fig. 2B) also has the same active residues.

2.2. GROMACS energies analysis

To analyze the conformation and the stability of the wild type (3UNF), and the mutant (G210V), we performed the energy simulation for 1000 ps and revealed the energy, temperature, pressure, and density profiles of both proteins, depicted in Fig. 3. The potential energy profiles of the 3UNF and G210V are shown in Fig. 3 A. Potential energy represents the interactions between the atoms within the system, and a comparison between the two profiles can reveal any differences in stability or conformational changes induced by the mutation. The temperature profiles of the WT and mutant systems are displayed in Fig. 3B. Monitoring temperature is crucial to ensure proper equilibration and stability. By examining the temperature fluctuations, we can assess the impact of the mutation on the protein's dynamics and identify any variations in behaviors between the two systems. The pressure profiles during simulation are given in Fig. 3C. Pressure is an essential thermodynamic parameter that characterizes the system's response to compression or expansion. By analyzing the pressure fluctuations, we can assess the stability of the systems and investigate any effects of the mutation on the protein's conformation or dynamics. The density profiles of the WT and mutant proteins are presented in Fig. 3D. Density reflects the mass distribution within the system and can provide insights into packing and solvent effects. Comparing the density profiles allows us to identify any differences in the mass distribution or packing density resulting from the mutation. This four-panel image comprehensively explains the energy landscape, temperature variations, pressure fluctuations, and density changes between the WT (3UNF) and mutant (G210V) PSMB8 proteins. It enables a detailed analysis of the effects of the G210V mutation on the protein's behavior, stability, and structural properties.

Table 1
Physiochemical characteristics of the 3UNF and G210V.

| Characteristics	Wild (3UNF)	Mutant (G210V)
Number of amino acids	230	230
Total number of atoms	3600	3609
Molecular weight (Da)	25539.15	25581.23
Theoretical pI	8.42	8.42
Extinction coefficients (M^{-1} cm^{-1}) at 280 nm in water, assuming all pairs of Cys residues form cystines	31985	31985
Extinction coefficients (M^{-1} cm^{-1}) at 280 nm in water, assuming all Cys residues are reduced	31860	31860
Instability index (II)	45.37	46.11
Aliphatic index	85.65	86.91
Grand average of hydropathicity (GRAVY)	−0.237	−0.217

A

B
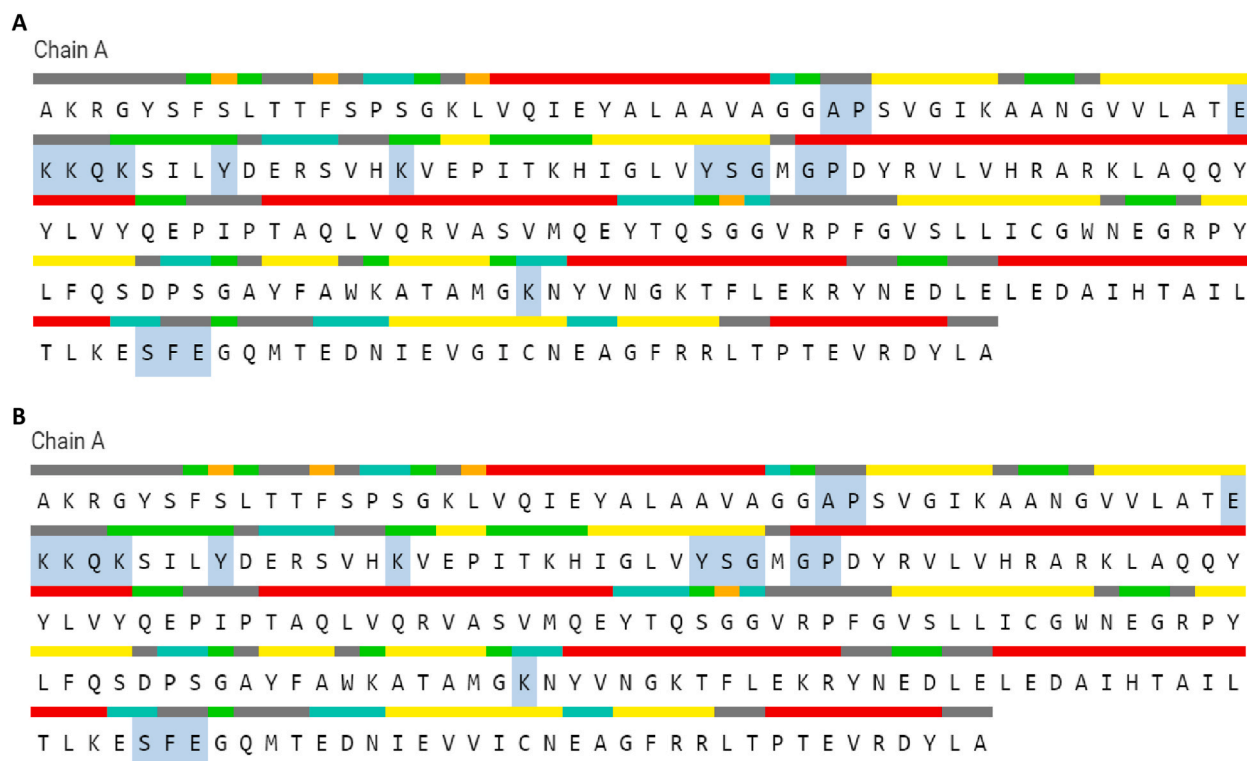
Fig. 2. Active Site Prediction. Active sites are predicted using CastP. The amino acid sequences in 32–33, 48–52, 56, 63, 75–80, 164, 197–199 residues may serve as potential therapeutic binding sites. The WT (A) and the mutant (B) proteins belong to the same active sites.

2.3. Molecular dynamics (MD) analysis

MD simulation is the most common way to study the structure and movement of proteins to find the molecular basis of how they work. Using the simulation data, we analyzed the RMSD, RMSF, Rg, and total SASA to find the stability of the proteins. We also calculated the mean value of these characters (Table 2). Using the MD trajectories of both the native and the mutant protein structures, we determined the root-mean-square deviation (RMSD) to evaluate the protein backbone's stability. The highest and lowest RMSD values for the WT and mutant-type trajectories were visualized in Fig. 4A, respectively. The WT trajectory is steady up to 35 Å and then shows a significant fluctuation between 35 Å and 65 Å before returning to a stable state up to 100 ns. In the case of mutant trajectory, it remains stable from 45 ns. From 18 ns to 35 ns and 55 ns–80 ns, the mutant structure varies from the wild-type, and the mutant trajectory intersects the wild-type seven times. An immense RMSD value for the wild-type protein is 34 Å, and the mutant protein is 38 Å. The mean RMSD value for the WT is 16.84 Å, which is lower than the mutant value (25.98 Å). These data suggest that the mutant protein is less stable than the wild-type protein.

The Root mean square fluctuation (RMSF) measures the residues' flexibility and shows how the structure moves over time. In Fig. 4B, we compare the mobility of amino acids in the wild-type protein structure to that of the mutant protein structure. According to the RMSF data, five amino acid residue ranges significantly deviate from the wild-type trajectory. The first variation occurs between the 65th and 75th residues, where the deviation ranges from 11 to 32 Å; the second variation occurs between the 82 and 125th residues, where the deviation ranges from 23 to 33 Å; the third variation occurs between the 135th and 145th residues, where the deviation ranges from 23 to 35 Å; the fourth variation occurs between the 148th and 155th residues, and finally, the fluctuation occurs between the 210-225th residues, where the deviation ranges from 10 to 30 Å. The mean RMSF values of the 3UNF and G210V are 22.80 Å and 30.51 Å, respectively. These data support the structural deviation following mutation. RMSF analysis demonstrated that the G210V mutant system was more flexible, with six highly flexible regions compared to the wild-type system.

The radius of gyration (Rg) indicates protein structural compactness. It concerns how regular secondary structures can be compactly packed into a protein's 3D structure. Increased Rg values indicate decreased protein structural compactness, implying greater flexibility and less stability. The Rg data (Fig. 4C) for the WT (3UNF) and mutant (G210V) demonstrates that the high Rg value for the wild-type protein is 38 Å at 81ns, but the mutant system's Rg is 42 Å at 85 ns. The mutant protein has a larger Rg value than the wild-type system, and there are multiple crossings between the two. Most interestingly, the mutant's Rg value drops at 90 ns. These findings imply that the mutant protein has low compactness and reduced stability up to 0–90 ns. The mean Rg value for the WT protein is 23.94 Å, while the mutant protein's value is 29.81 Å. Increasing the mean Rg value along the 100 ns simulation in the mutant system indicates its conformational changes, and we can conclude that the mutant protein is going into a slightly unfolding state several times
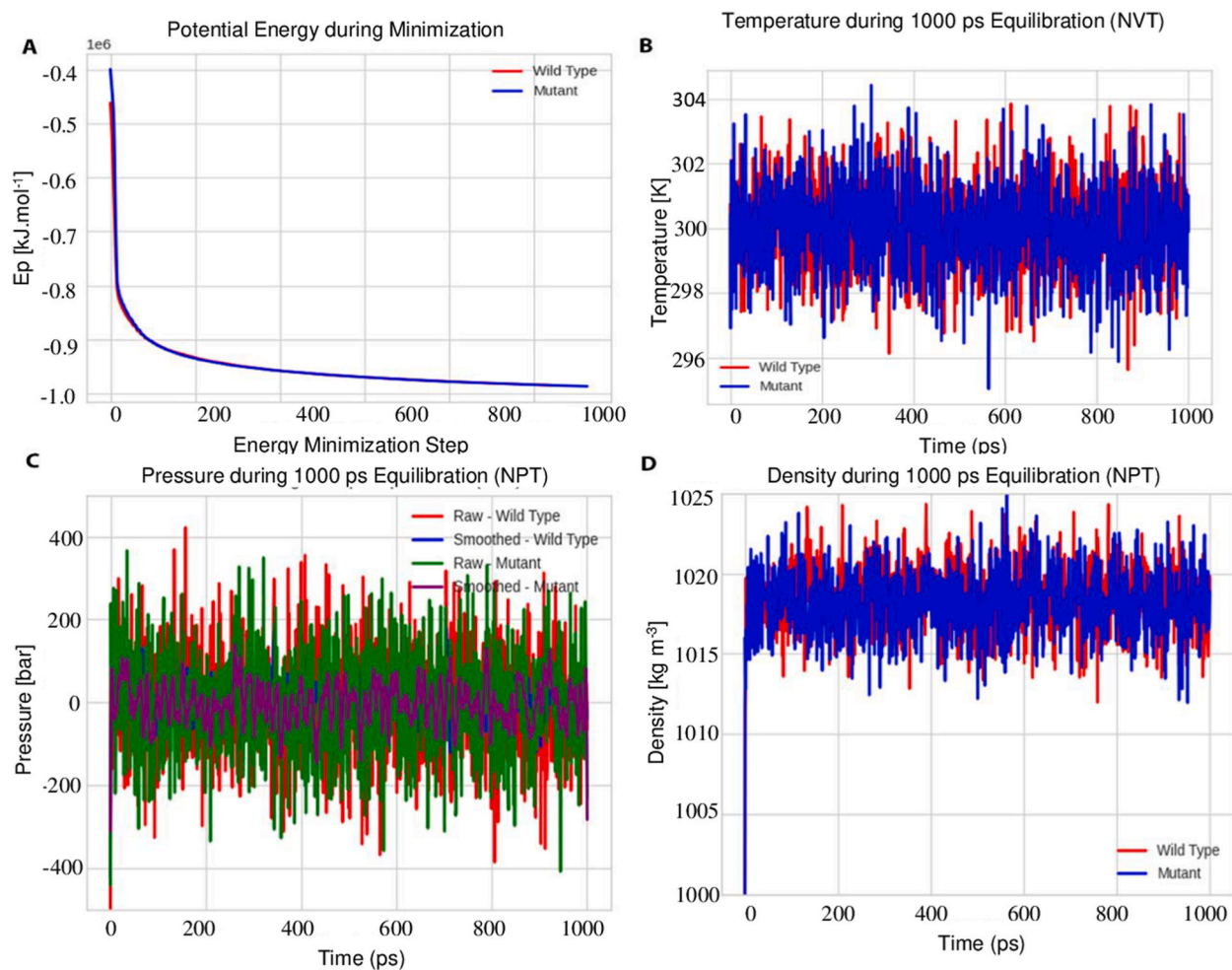
Fig. 3. Comparison of Energy, Temperature, Pressure, and Density Profiles for Wild Type and G210V Mutant of PSMB8 Protein. The image comprehensively analyzes the energy landscape, temperature variations, pressure fluctuations, and density changes for both the wild-type and mutant (G210V) PSMB8 proteins. The simulations were carried out for 1000 ps (ps) to observe the behavior and properties of the proteins. (**A**) The potential energy profile shows the conformational changes due to mutation. The wild type and the mutant are depicted in red and blue color respectively, (**B**) the temperature profiles show the equilibration and stability, (**C**) the pressure profiles show the thermodynamics parameter, (**D**)the density profiles show the packing density of the wild type and mutant protein.

Table 2
Mean value of RMSD, RMSF, Rg and total SASA.

Characters	WT (3UNF)	Mutant (G210V)
RMSD	1.684 nm	2.598 nm
RMSF	2.280 nm	3.051 nm
Rg	2.394 nm	2.981 nm
SASA	139.335 nm^2	143.841 nm^2

more than the WT.

2.4. Essential dynamics analysis

In this study, we also executed the RMSD average linkage hierarchical clustering. Besides, Cartesian coordinate PCA, pairwise distance PCA, SASA, and SASA autocorrelation of 100 ns simulation trajectory to understand the conformational behavior and evaluate the wild-type protein compared to the mutant. In the hierarchical clustering method, the clusters are arranged hierarchically according to similarities or dissimilarities between various clusters. The clustering was carried out using Ward's [35] method [35], which was found to be the most appropriate because it produces homogeneous and understandable clusters more effectively than other
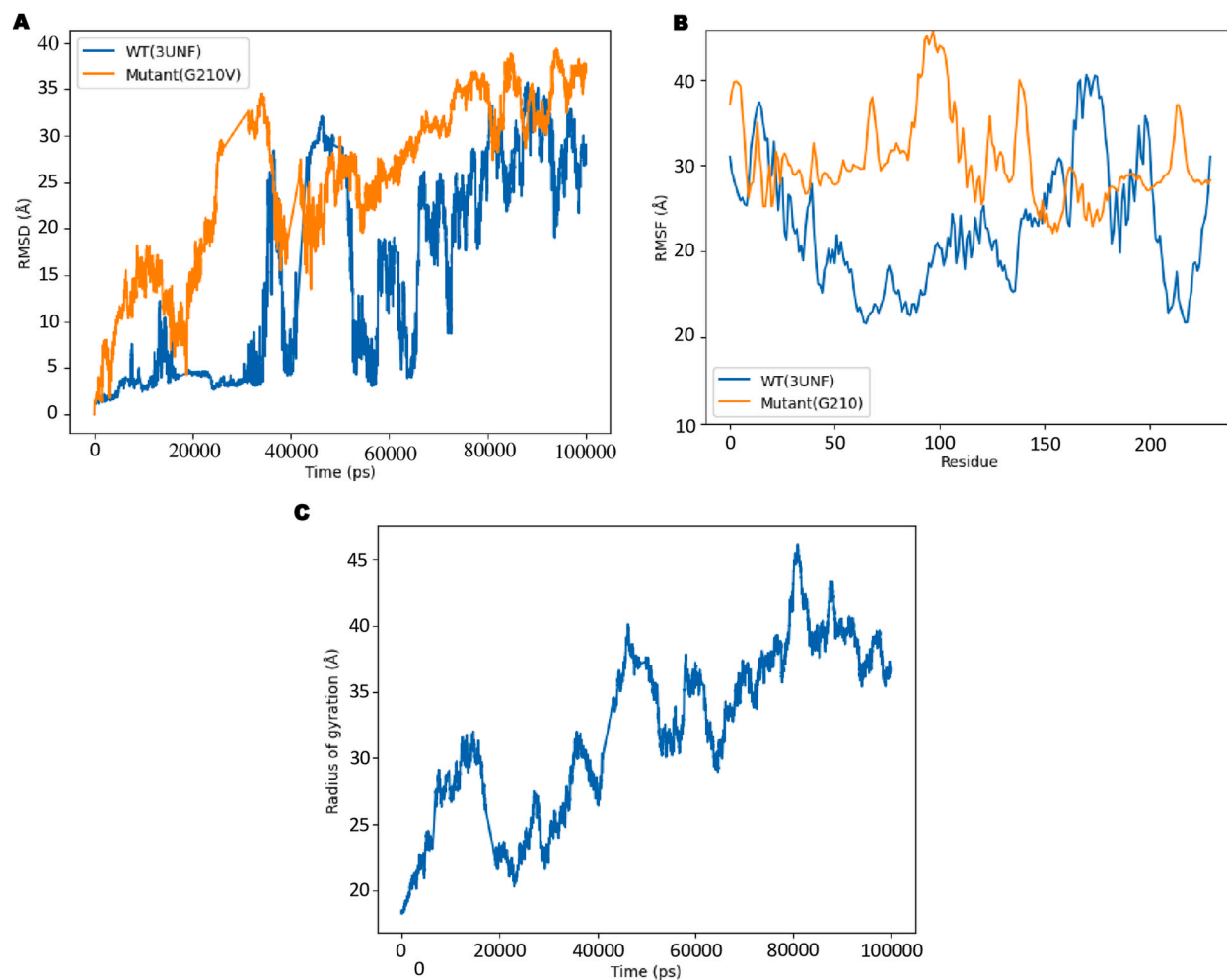
Fig. 4. RMSD, RMSF, and Rg analysis of the 3UNF and G210V mutant of PSMB8. Panel A shows the root-mean-square deviation (RMSD) of the wild type in blue and the mutant in orange over the entire 100 ns simulation period. The mutant's RMSD value is greater than the wild-type protein's, indicating that the mutant is less stable than the wild-type protein. Panel B presents the root-mean-square fluctuation (RMSF) of PSMB8 in blue and the mutant structure in orange, highlighting six flexible regions. In panel C, this reveals that the mutant protein (orange) has a higher Rg value than the wild type up to 90 ns. This indicates that the mutant protein is less compact and has less stability.

hierarchical methods [36]. In protein MD simulations, hierarchical clustering analysis based on RMSD was performed for the wild type 3UNF and the mutant, and this operation yields a trajectory dendrogram (Fig. 5). The dendrogram shows how the conformations are organized in a hierarchy and how similar they are to each other. We looked at the entire 100 ns of both the wild-type and mutant proteins' alpha carbon chains. The conformations are orange, green, red, purple, violet, and many colors. In both wild-type and mutant data, each cluster contains many subclusters. The wild-type protein (Fig. 5A) has 18 distinct clusters, whereas the mutant protein shows 24 clusters (Fig. 5B). These data suggest that structural changes occur following mutation.

To analyze the mobility of the protein within the framework, we also analyzed the principal components for the entire 100 ns. The pairwise distance PCA is computed by analyzing the precise location of the atoms over time. In contrast, the Cartesian coordinate PCA analysis captures a portion of the dominant overall motion. In this study, Cartesian coordinate PCA data for wild-type trajectory (Fig. 6A) reveals that the initial atoms move from the second quadrants. They rotate anticlockwise via the third, fourth, and first quadrants, and finally, they appear in the second and third quadrants. In the mutant protein, the atoms move from the second co-ordinate anticlockwise. The final atoms appear between the first and fourth quadrants (Fig. 6B). The pairwise distance trajectory shows that the initial atoms of the wild-type protein (Fig. 6C) appear in the first and second quadrants. The final atoms appear in the third and fourth quadrants. In the case of the mutant protein (Fig. 6D), the atoms are in the second and third quadrants at the initial stage, and in the final stage, the atoms are found in the second and fourth quadrants. Using PCA, we analyzed the Gibbs free energy value for the wild-type protein (Fig. 7A) and mutant protein (Fig. 7B). The WT system spanned in lower ranges than the mutant system. The WT system reflects the native and folded state of the protein, and the mutant system seems to have a slightly unfolded state. In the mutant system, we also noticed additional basins that showed the conformational alterations brought on by the mutation.

We examined the SASA to determine the solvent accessibility of the wild-type 3UNF and the mutant G210V. The highest SASA value
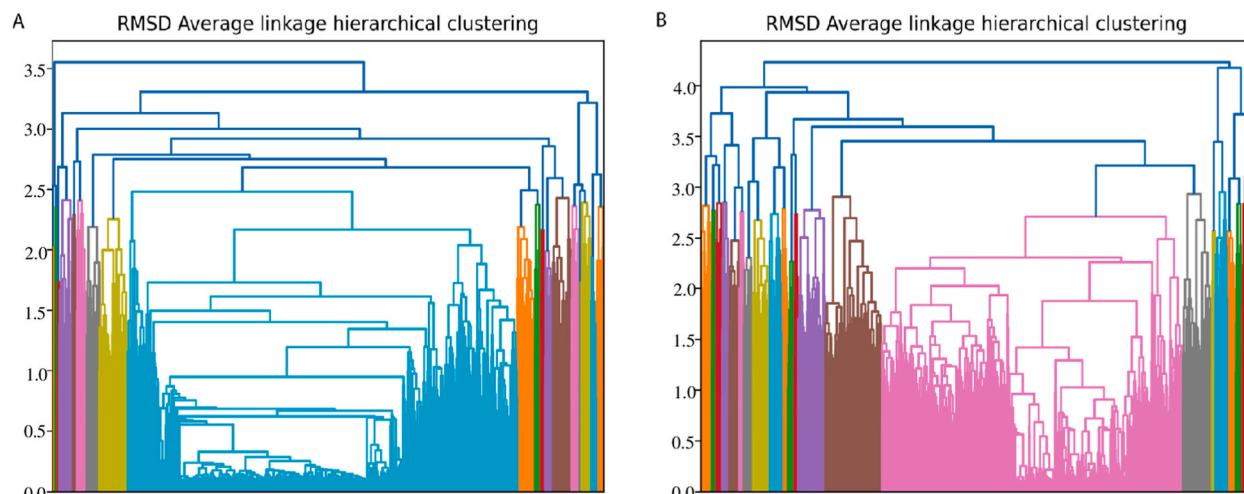
Fig. 5. RMSD Average Linkage Hierarchical clustering data reveal the conformation and the number of clusters for both the wild-type and mutant protein. (A) RMSD Hierarchical clustering data of the wild-type protein shows that there are 18 clusters depicted in red, green, orange, blue, and many other colors. (B) In the case of the mutant trajectory, it shows 24 distinct clusters.

for the mutant is 163 nm^2, while the value for the wild type is 158 nm^2. The mutant structure has a higher SASA value than the wild type (Fig. 8A and B). The mean SASA value for the WT protein (139.335 nm^2) is lower than the mutant protein (143.841 nm^2). It can be inferred that the wild-type protein is more stable than the mutant protein since a higher SASA value indicates protein expansion. The effect of amino acid substitution by changing the size of the protein surface and other properties could cause a higher variation in the SASA value.

Further, we analyzed the SASA autocorrelation analysis. The wild-type trajectory (Fig. 8C) shows that from the 0.010 ns – 1 ns time frame, it maintains a constant value of about 1 nm^2 followed by a steady decline until one peak of about 0.1 nm^2 (50–60 ns). A similar pattern has been seen in the mutant structure (Fig. 8D), except that there is no peak.

The aforementioned PCA and SASA values suggest that there are differences in stability, flexibility, and compactness between the wild-type and mutant PSMB8 proteins. These changes could be important in understanding the etiology of disorders connected to Experimental Autoimmune Encephalomyelitis (EAE). In comparison to the wild-type protein, which has a solvent-accessible surface area (SASA) value of 158 nm^2, the mutant G210V protein has a higher value, 163 nm^2, indicating increased solvent exposure and decreased structural stability. Increased protein expansion is indicated by a higher SASA score, which probably makes the mutant's structure less compact and more flexible [37]. The proteasomal activity of PSMB8, which is essential for presenting antigens to the immune system and breaking down misfolded proteins [38], may be impacted by this instability. This is a major contributing factor to autoimmune disorders like EAE.

The wild-type atoms originally moved from the second quadrant and rotated through all quadrants before settling in the second and third quadrants, highlighting structural differences even more in the PCA analysis. On the other hand, the mutant atoms started in the second quadrant and travelled counterclockwise, finishing at the region between the first and fourth quadrants. These changes in atomic motion result from the mutant protein's decreased flexibility, which may interfere with its ability to engage with immune-related signaling pathways or proteasomal complexes, aggravating autoimmune reactions [39]. Because of the mutant's enhanced solvent exposure and changed molecular motion, PSMB8's function in immunological regulation and protein homeostasis may be compromised, potentially contributing to disease processes.

Stabilizing the mutant PSMB8 protein's structure could be the primary goal of therapeutic treatments aimed at PSMB8-related disease: EAE, according to the aforementioned results of increased solvent exposure, decreased stability and, altered flexibility in the mutant protein. PSMB8's regular activity may be restored with the aid of small molecules or peptide-based therapeutics intended to decrease protein expansion, enhance folding, or improve proteasomal function. Such treatments provide a targeted approach to controlling linked inflammatory and autoimmune disorders by addressing these structural imbalances and potentially mitigating the autoimmune responses triggered by malfunctioning PSMB8.

2.5. CATH analysis

According to the CATH classification system, the CATH/Gene3D v4.3 database [40] classified our proteins into four levels. In the case of the PSMB8 protein, the domains are assigned as 3.60.20.10. The numbers 3, 60, 20, and 10 indicate the related categories. This protein belongs to the alpha-beta class, as the class level is assigned as code 3. This class has 14 distinct architectures, including Roll, Alpha-Beta Barrel, 2-layer sandwich, 3-layer aba sandwich, 4-layer sandwich, alpha-beta horseshoe, etc. The architectural score of PSMB8 is 60, which indicates that this protein has a 4-layer sandwich shape in its structure. This architecture contains 16 topologies, including deoxyribonuclease 1, chain A, metallo-beta-lactamase, chain A, glutamine phosphoribosylpyrophosphate, subunit 1, domain
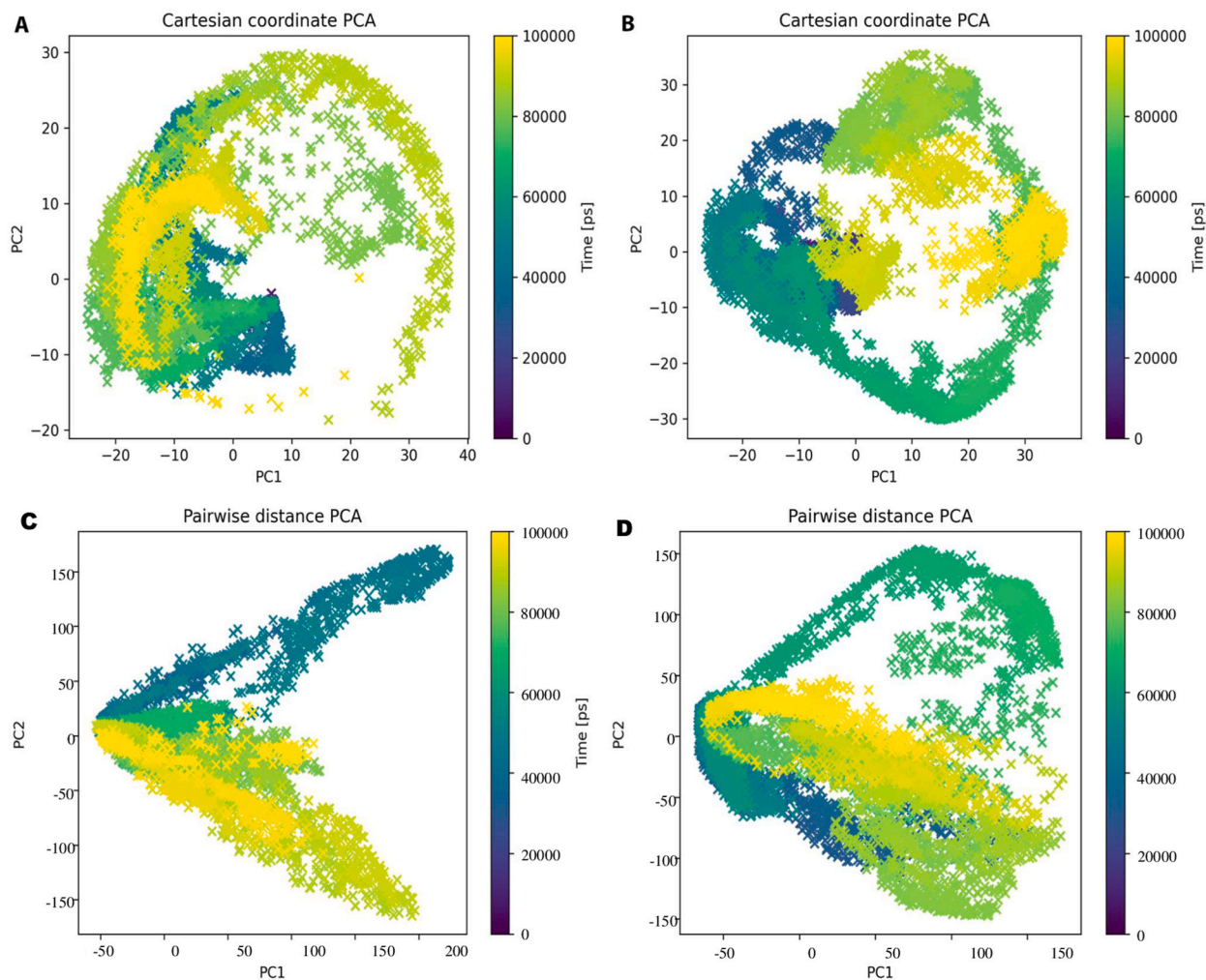
Fig. 6. Cartesian coordinate PCA analyzes the overall motion of the atoms and The Pairwise Distance PCA shows the atom's precise location during the period. (A) the wild protein trajectory shows that initial atoms rotate anticlockwise from the second coordinate, and final atoms appear at the second and third coordinates. (B) the mutant trajectory describes that the initial atoms are in the second coordinate, then rotate anticlockwise and finally appear at the third and fourth coordinates. (C) The initial atoms of the 3UNF are in the first and second quadrants, and the final ones appear in the third and fourth quadrants. (D) The mutant trajectory data suggest that the initial atoms are found in the third and fourth, and the final atoms appear in the second and fourth quadrants.

1, etc. The topology of the PSMB8 is glutamine phosphoribosylpyrophosphate, subunit 1, domain 1, which indicates that this protein has two catalytic groups: nucleophile and proton donor. This topology has 4 homologous superfamilies, and PSMB8 belongs to aminohydrolase (H = 10). Regarding the mutant protein, the CATH database analyzed the amino acid sequences and showed 121 matching domains with the wild-type protein. The most similar domain is 3UNFA. So, the mutant protein belongs to the alpha-beta class, 4-layer sandwich architecture, Glutamine Phosphoribosylpyrophosphate, subunit 1, domain 1 topology, and aminohydrolase superfamily. This analysis ensures that the mutation does not change the class, architecture, topology, and homologous superfamily of the PSMB8 protein.

3. Methods and materials

3.1. Dataset

To investigate the proteasome 20S subunit beta 8 (PSMB8) protein and its mutant G210V properties, we extracted the protein's primary sequence from the UniProtKB database using its UniProtKB identifier, P28063. We then used Expasy's Protparam [41] server to retrieve the protein's physicochemical properties. To further understand the structural characteristics of the PSMB8 protein, we analyzed its crystal structure using the PDB ID: 3UNF. We also examined the G210V mutated version of the PSMB8 protein, generated using UCSF Chimera [42]. The potential active sites of the wild-type protein and the mutant protein were analyzed using CastP
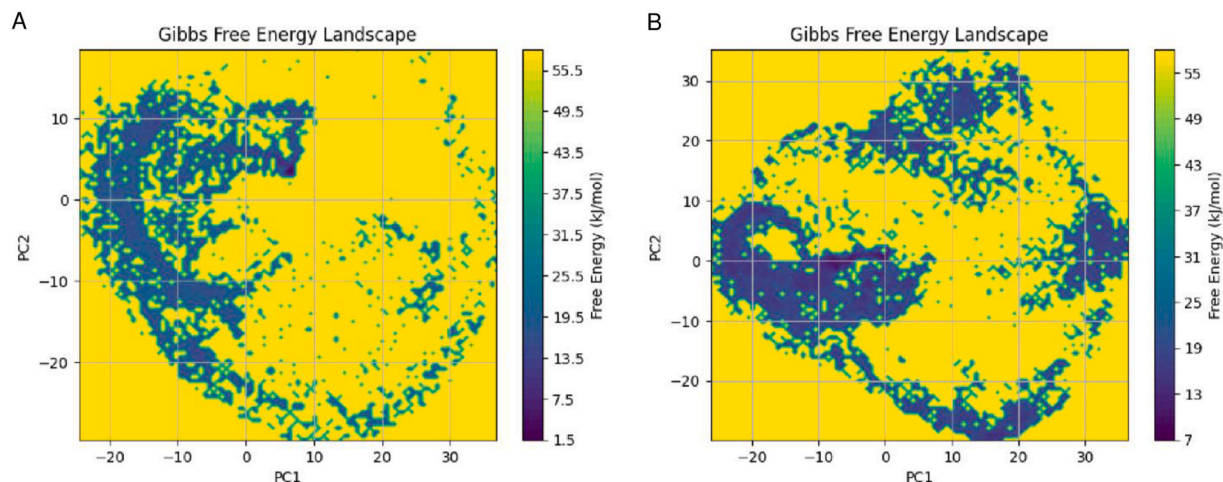
Fig. 7. Gibbs Free Energy Landscape. Free energy contour map using the PCs (PC1 and PC2) for (A) Wild type-3UNF and (B) Mutant-G210V. Here, the deep color represents the lower energy, and several basins across the G210V landscape indicate the conformational changes.
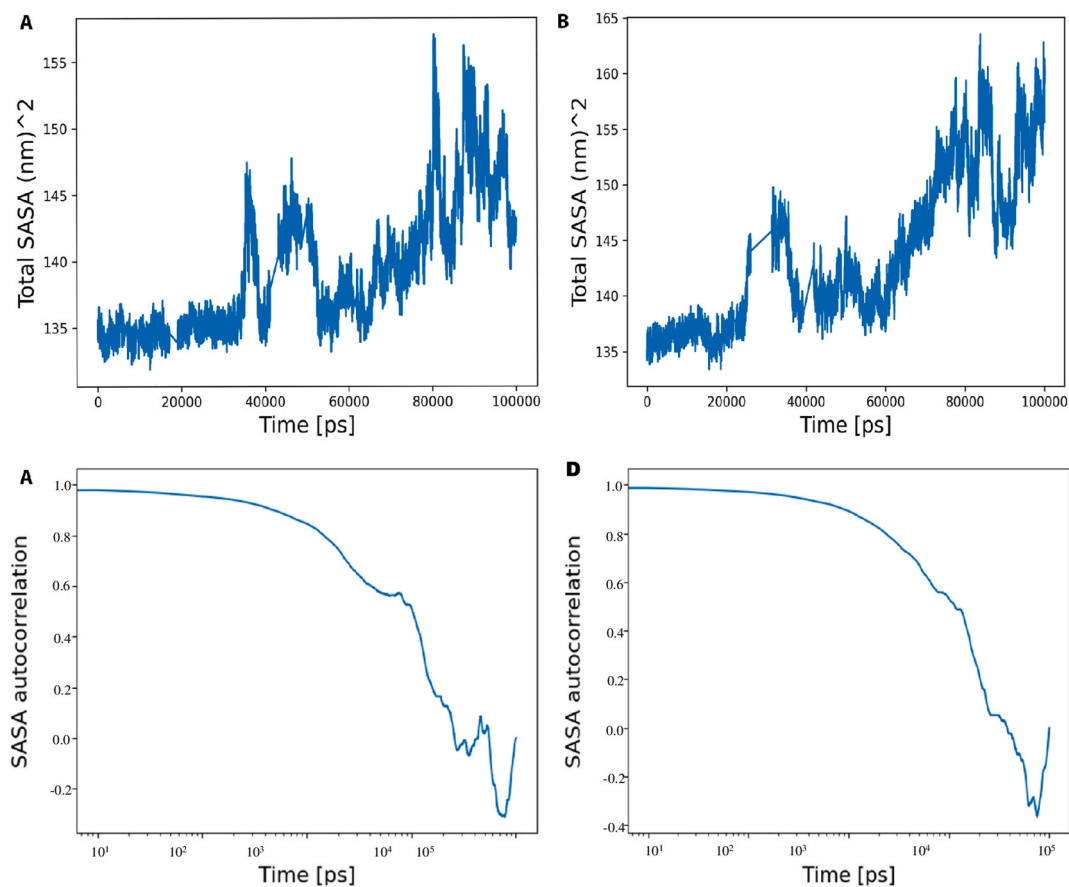

Fig. 8. Solvent Accessible Surface Area (SASA) and SASA Autocorrelation Analysis for wild and mutant protein. (A) The highest SASA value for the wild type 3UNF is 158 nm^2, and the SASA range is 132 nm2–158 nm2. (B) The mutant SASA trajectory shows that the highest value is 163 nm2, ranging from 133 to 163,163 nm2. The value of the mutant protein is larger than that of wild type one, which indicates that the mutant protein has more significant expansion and less stability. (C) the wild protein trajectory demonstrates that the protein maintains a constant value around 1 nm2 from 0.010 ns to 1 ns. Then, the downfall occurs, giving one peak of about 0.1 nm2 (50–60 ns). (D) The mutant trajectory is almost similar to the wild protein, but there is no peak in the mutant protein.

(Dundas et al., 2006). Knowing the protein's binding sites along with the amino acid residues leads to understand the protein-ligand interaction. This information will be useful in identifying potential therapeutic binding compounds.

3.2. Molecular dynamics simulations

MD simulation studies were performed on both forms (wild-type and mutant) of the protein to get insight into how both types behave in trajectory analysis. To accomplish the molecular dynamics simulation, we first followed our previous protocol described in Paul et al., 2022 [43], as well as the protocol developed by Ref. [44,44]), and then we optimized our Google Colab framework protocol both for simulation and analysis.

To assess the stability and uniformity of the structure of PSMB8 and its mutant, a 100 ns (100,000 ps) molecular dynamics (MD) simulation analysis was carried out on GROMACS 2020.6 [45] running on Google Colab for both the wild type and the mutant protein.

In the parameterization step, these proteins are parameterized using the AMBER99SB-ILDN force field [46], an improved version of the Amber ff99SB force field, which has limitations to the four residues (isoleucine, leucine, aspartate, and asparagine). AMBER99SB-ILDN force field is frequently used in MD simulations, and its parameters accurately depict the dynamics and flexibility of folded proteins. This force field is useful in simulating side-chain flexibility and backbone conformations [47], which are important for comprehending structural alterations due to mutations like PSMB8 wild-type and its mutant G210V. Although there are other force fields that are similarly widely used, such as CHARMM36m, AMBER99SB-ILDN provides an ideal compromise between precision and computing efficiency for systems where structural changes are impacted due to the mutations [48].

Then, the wild type and mutated structure were solvated using the simple point charge-extended (SPC/E) [49] water model, which represents water molecules as three-point molecules, each with a distinct charge. This model could rectify polarization energy, resulting in a greater dipole than other models. The SPC/E model appears to be near the ideal nonpolarizable water model for bulk studies. In addition, SPC/E performs better than many (simple) polarizable models when investigating large volumes of water [50,51].

Afterward, in the step of the solvation of the protein, a cubic simulation box size of 1.50 nm was defined around the protein structure, and the protein was confined at the center of the box. The negative charge was neutralized by adding four Na^+ ions. Periodic boundary conditions were applied in all directions to avoid the edge effect. The energy of the system was minimized for both the wild-type and mutant protein structures using the steepest descent technique with a maximum step of 50,000. Following the minimization, a 1000-ps equilibration period was performed in the canonical ensembles (NVT) and isothermal-isobaric (NPT) ensembles. During the NVT equilibration process at 300K temperature, the leapfrog algorithm was utilized to propagate the protein motions over time, and the position restraining method was employed to maintain the position of the heavy atoms that make up our protein. During the NPT, the Parrinello-Rahman barostat [52] method was used to maintain the pressure of the system at 1 atm.

Ultimately, the generation of the 100,000 ps (100 ns) MD run with no restrictions was carried out with the help of the LINCS algorithm and a 2-fs (fs) integration phase. The PME approach was utilized for the Coulombic and Lenard-Jones interactions. During the simulation, our system was interrupted several times, and then we utilized the appending simulation code. In the last step, the trajectory files were converted into a PDB file, and then we downloaded the file for further trajectory analysis.

3.3. Trajectory data analysis

The root-mean-square deviation (RMSD), root mean-square fluctuation (RMSF), solvent accessible surface area (SASA), radius of gyration (Rg), Principal Component Analysis (PCA), Clustering Cartesian Coordinate PCA, Pairwise Distance PCA, and Solvent Accessible Surface Area (SASA) were calculated using another Google Colab system. This system requires the installation of python, matplotlib, mdtraj, nglview, cython, pytraj, tsplot and gnuplot-x11 and biotitleGraphs were drawn with the help of the MDTraj program [53].

3.4. Principal component analysis (PCA)

Molecular dynamics (MD) simulations permit the microscopic study of the structure and dynamics of molecular systems. Principal component analysis (PCA) has become one of the most widely used techniques for analyzing the movement of proteins within this framework [54–56]. PCA is a statistical method that is utilized to decrease the number of dimensions required to explain protein dynamics in a systematic manner. This is achieved through a decomposition process that filters observed motions from the largest to the smallest spatial scales. PCA is a linear transform that uses a covariance matrix or a correlation matrix to extract the most significant data elements. This linear transform is constructed from atomic coordinates that describe the degrees of freedom (DOF) of the protein, such as the Cartesian coordinates that define atomic displacements in each conformation that makes up a trajectory [57]. PCA reduces the data by geometrically projecting them into lower dimensions, which are referred to as principal components (PCs). The primary objective of PCA is to obtain the most optimal summary of the data by utilizing a limited number of PCs [55]. In PCA, data is placed in a matrix such that each row represents a sample (observation or instance) and each column indicates a degree of freedom. After centroid subtraction, the covariance matrix of the data set is:

$$C_{ij} = \; <(x_i - <x_i>)(x_j - <x_j>)> \tag{1}$$

where $<\cdots>$ represents the average value across all samples in the data set. The correlation matrix is derived from this matrix using the formula:

$$P_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} \, C_{jj}}} \tag{2}$$

And this square symmetric matrix is diagonalized as

$$R^T \, PR = \Lambda \tag{3}$$

Using standard numerical methods, where R is an orthonormal transformation matrix (whose column vectors are the eigenvectors of P), the superscript T stands for transposition and Λ is a diagonal matrix whose elements are the eigenvalues. After arranging the columns of the eigenvector matrix R and the eigenvalue diagonal matrix Λ in decreasing order of eigenvalues, the empirical matrix is projected onto the eigenvectors to obtain the principal components [58].The principal component analysis (PCA) model comprises four fundamental elements, namely the data, the scores, the loadings, and the residuals. The actual data visualization usually depends on the nature of the data and the rules of a particular discipline. It is often possible to plot the data as curves for continuous data such as time series and spectra. It is usually plotted in other ways for discrete data, such as bar plots. Data visualization typically helps residuals. The residuals are correlated with the residual spectra for spectral data, and consequently, they provide the required chemical information [59]. Although PCA is a commonly used descriptive data analysis method, it also has several adaptations that make it effective in many circumstances and data types in many disciplines [58].

In this study, we used the principal component analysis (PCA) method to find collective movements in a molecular dynamics (MD) trajectory that were important from a functional standpoint. The entire 100000 ps were considered for the analysis of principal components in using the MDTraj program [53].

Based on the PCA, we performed the Free Energy Landscape analysis (FEL) to investigate the conformational changes of biomolecules. The Gibbs free energy is calculated using equation (4), where ΔG (X, Y) is the Gibbs free energy at coordinates, k_B is the Boltzmann constant whose value is $1.38064852 \times 10{-}23 J/K$, T is the temperature in Kelvin: in this analysis the temperature was 300k, and P(x,y) is the probability distribution of the system's structural states at the coordinates [60,61].

$$\Delta G(X) = - K_B T \, \ln P \, (X, Y) \tag{4}$$

Additionally, we used the RMSD distance matrix and hierarchical clustering method to cluster the MD simulation trajectory of both proteins (3UNF_A and the mutant G210V), resulting in an RMSD average linkage hierarchical clustering dendrogram. We then built a two-component PCA model to compute principal components of the entire 100 ns while considering the alpha carbon chain by projecting the simulation data into the reduced dimensional space of both proteins' structures. Cartesian coordinates acted as the input PCA, which is alignment-dependent. At the same time, we also computed the alignment-independent pairwise distance PCA between every atom in each frame of the alpha-carbon chain of both proteins' structures.

3.5. CATH analysis

CATH can determine the structural domains present in proteins, classify those domains into evolutionary superfamilies, and annotate those evolutionary superfamilies with structural and functional information (Sillitoe et al., 2021). Using its own structure and sequence alignment program (SSAP), it produces structural superpositions of all typical protein domains [40]. CATH database classifies the protein domain into four distinct hierarchies: (a) class, which defines the protein into six [62] major classes including mainly alpha, mainly beta, alpha/beta, few secondary structures, and special, (b) architecture, which group domains with similar general characteristics concerning the overall protein-fold shape, (c) topology, which analyze the structural similarities that exist between secondary components, and (d) homology, groups together domains that have similar structures and functions which suggests that they may have descent from the same ancestor [63]. In this study, we used the PDB ID of the wild-type protein (3UNF_A) and the amino acid sequence of the mutant protein (G210V) to determine the hierarchy of these proteins based on evolutionary relationships. We used the CATH database following the above-mentioned system.

4. Limitations and future recommendations

First, this simulation and analysis protocol is optimized for execution on Google Colab, utilizing its freely available GPU resources. This method provides a low-cost and accessible option for doing molecular dynamics (MD) simulations, which is especially useful for researchers with limited access to specialized high-performance computing resources.

However, this protocol has several limitations–Google Colab is an interactive environment, which means that if the session goes idle, the simulation is halted and must be manually resumed. Furthermore, the free edition of Google Colab limits GPU usage to about 3–4 h per day, increasing the overall simulation time. Running a 100 ns simulation of a protein consisting of ~250 amino acids can take approximately two weeks under these conditions.

Despite these constraints, the protocol offers significant potential for future research. It can be applied to explore the dynamic behavior of various proteins, enabling the identification of binding residues and active pockets. These insights could be crucial in drug discovery efforts, designing novel therapeutic agents targeting specific protein functions.

Following the modeling of wild-type (3UNF) and mutant (G210V) PSMB8 proteins, the next study could include molecular docking of drug-like libraries to select promising candidates based on binding scores and critical binding residues. The most promising compounds were then subjected to molecular dynamics (MD) simulations to investigate the stability and trajectory of the protein-ligand binding posture. Finally, experimental assays could be used to confirm the efficacy of the chosen compounds. This

comprehensive strategy will eventually pave the path for future Multiple Sclerosis (MS) treatment therapies.

5. Conclusion

In this study, the simulation and analysis were carried out using our optimized cloud computing like Google Colab instead of GROMACS software. We target an EAE disease biomarker named PSMB8 and a mutant protein (G210V) which leads to NNS in humans. We visualized temperature, pressure, and energy profile equilibrium and stability with the wild-type protein. According to this result, there were no significant fluctuations between wild- and mutant PSMB8 trajectories. We found mutant protein decreased in stability from RMSD analysis; RMSF value pointed toward a more flexible structure of mutated G210V protein, and we hypothesize valine incorporation might have aided in increased ligand binding capacity and Rg pointed toward a less compact mutated structure corresponding to wild-type, which might have increased proteins' structural activity. The RMSD clustering pointed to structural change in the mutant type. Pairwise distance PCA and Cartesian coordinate PCA clearly showed final cluster differences between both structures. SASA value depicted that the mutant structure has achieved increased solvent-accessible surface area and high variation in their structure. SASA autocorrelation trajectories also pointed toward significant similarities between structures with a single peak difference. Our CATH analysis reveals that point mutation in the 210 positions does not alter the class, architecture, topology, and homologous superfamily of the PSMB8 protein. In concisely, the mutation of Glycine to Valine at the 210th amino acid position in PSMB8 has decreased the structure's stability and compactness, which aids in gaining elevated flexibility, ligand binding capacity, and structural activity in the mutant protein. These proteins show stability during the 100 ns simulation and could be a potential therapeutic target for MS and other PSMB8 mutation-related diseases.

CRediT authorship contribution statement

Shamrat Kumar Paul: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Md Saddam:** Writing – original draft, Methodology, Investigation, Data curation. **Nisat Tabassum:** Writing – original draft, Methodology, Formal analysis, Data curation. **Mahbub Hasan:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Data curation, Conceptualization.

Data availability statement

All trajectory data produced during the simulation are available at the open data-sharing platform Zenodo (https://zenodo.org/record/8070983; https://doi.org/10.5281/zenodo.8070983. Simulation and Analysis methods are elaborately described at the following Github repository: https://github.com/paulshamrat/ColabMDA.

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.

Acknowledgment

References

[1] S. Haase, R.A. Linker, Inflammation in multiple sclerosis, Therapeutic Advances in Neurological Disorders 14 (2021) 17562864211007688.
[2] D. Jakimovski, S. Bittner, R. Zivadinov, S.A. Morrow, R.H. Benedict, F. Zipp, B. Weinstock-Guttman, Multiple sclerosis, Lancet (London, England) 403 (10422) (2024) 183–202, https://doi.org/10.1016/S0140-6736(23)01473-3.
[3] H. Lassmann, Pathogenic mechanisms associated with different clinical courses of multiple sclerosis, Front. Immunol. 9 (2019) 3116.
[4] M.A.L. Pinheiro, G. Kooij, M.R. Mizee, A. Kamermans, G. Enzmann, R. Lyck, M. Schwaninger, B. Engelhardt, H.E. de Vries, Immune cell trafficking across the barriers of the central nervous system in multiple sclerosis and stroke, Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1862 (3) (2016) 461–471.
[5] R. Mazumder, C. Murchison, D. Bourdette, M. Cameron, Falls in people with multiple sclerosis compared with falls in healthy controls, PLoS One 9 (9) (2014) 1–7, https://doi.org/10.1371/journal.pone.0107620.
[6] S.M. Piryonesi, S. Rostampour, S.A. Piryonesi, Predicting falls and injuries in people with multiple sclerosis using machine learning algorithms, Multiple Sclerosis and Related Disorders 49 (2021) 102740, https://doi.org/10.1016/j.msard.2021.102740.
[7] J. Correale, M.I. Gaitán, Multiple sclerosis and environmental factors: the role of vitamin D, parasites, and Epstein-Barr virus infection, Acta Neurol. Scand. 132 (S199) (2015) 46–55, https://doi.org/10.1111/ane.12431.
[8] S.V. Ramagopalan, R. Dobson, U.C. Meier, G. Giovannoni, Multiple sclerosis: risk factors, prodromes, and potential causal pathways, Lancet Neurol. 9 (7) (2010) 727–739, https://doi.org/10.1016/S1474-4422(10)70094-6.
[9] M. Hasan, J.E. Seo, K.A. Rahaman, M.J. Kang, B.H. Jung, O.S. Kwon, Increased levels of brain serotonin correlated with MMP-9 activity and IL-4 levels resulted in severe experimental autoimmune encephalomyelitis (EAE) in obese mice, Neuroscience 319 (2016) 168–182, https://doi.org/10.1016/j.neuroscience.2016.01.045.
[10] J.F. Kurtzke, Epidemiology in multiple sclerosis: a pilgrim's progress, Brain 136 (9) (2013) 2904–2917, https://doi.org/10.1093/brain/awt220.

[11] N. Koch-Henriksen, P.S. Sørensen, The changing demographic pattern of multiple sclerosis epidemiology, Lancet Neurol. 9 (5) (2010) 520–532, https://doi.org/10.1016/S1474-4422(10)70064-8.

[12] M.B. Sintzel, M. Rametta, A.T. Reder, Vitamin D and multiple sclerosis: a comprehensive review, Neurology and Therapy 7 (1) (2018) 59–85, https://doi.org/10.1007/s40120-017-0086-4.

[13] S.I. Tracy, K. Kakalacheva, J.D. Lünemann, K. Luzuriaga, J. Middeldorp, D.A. Thorley-Lawson, Persistence of epstein-barr virus in self-reactive memory B cells, J. Virol. 86 (22) (2012) 12330–12340, https://doi.org/10.1128/JVI.01699-12.

[14] S.-M. Orton, B.M. Herrera, I.M. Yee, W. Valdar, S.V. Ramagopalan, A.D. Sadovnick, G.C. Ebers, Sex ratio of multiple sclerosis in Canada: a longitudinal study, Lancet Neurol. 5 (11) (2006) 932–936, https://doi.org/10.1016/S1474-4422(06)70581-6.

[15] M. Buc, Role of regulatory T cells in pathogenesis and biological therapy of multiple sclerosis, Mediat. Inflamm. 2013 (2013), https://doi.org/10.1155/2013/963748.

[16] F. Jadidi-Niaragh, A. Mirshafiey, Th17 Cell, the new player of neuroinflammatory process in multiple sclerosis, Scand. J. Immunol. 74 (1) (2011) 1–13, https://doi.org/10.1111/j.1365-3083.2011.02536.x.

[17] M.S. Maddur, P. Miossec, S.V. Kaveri, J. Bayry, Th17 cells: biology, pathogenesis of autoimmune and inflammatory diseases, and therapeutic strategies, Am. J. Pathol. 181 (1) (2012) 8–18, https://doi.org/10.1016/J.AJPATH.2012.03.044.

[18] J.S. Tzartos, M.A. Friese, M.J. Craner, J. Palace, J. Newcombe, M.M. Esiri, L. Fugger, Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis, Am. J. Pathol. 172 (1) (2008) 146–155, https://doi.org/10.2353/AJPATH.2008.070690.

[19] T. Carlson, M. Kroenke, P. Rao, T.E. Lane, B. Segal, The Th17–ELR+ CXC chemokine pathway is essential for the development of central nervous system autoimmune disease, J. Exp. Med. 205 (4) (2008) 811–823, https://doi.org/10.1084/JEM.20072404.

[20] F.D. Lublin, S.C. Reingold, J.A. Cohen, G.R. Cutter, P.S. Sørensen, A.J. Thompson, J.S. Wolinsky, L.J. Balcer, B. Banwell, F. Barkhof, B. Bebo, P.A. Calabresi, M. Clanet, G. Comi, R.J. Fox, M.S. Freedman, A.D. Goodman, M. Inglese, L. Kappos, C.H. Polman, Defining the clinical course of multiple sclerosis: the 2013 revisions, Neurology 83 (3) (2014) 278–286, https://doi.org/10.1212/WNL.0000000000000560.

[21] S. Sawcer, R.J.M. Franklin, M. Ban, Multiple sclerosis genetics, Lancet Neurol. 13 (7) (2014) 700–709, https://doi.org/10.1016/S1474-4422(14)70041-9.

[22] A.P. Gregory, C.A. Dendrou, K.E. Attfield, A. Haghikia, D.K. Xifara, F. Butter, G. Poschmann, G. Kaur, L. Lambert, O.A. Leach, S. Prömel, D. Punwani, J.H. Felce, S.J. Davis, R. Gold, F.C. Nielsen, R.M. Siegel, M. Mann, J.I. Bell, L. Fugger, TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis, Nature 488 (7412) (2012) 508–511, https://doi.org/10.1038/nature11307.

[23] F. Lundmark, K. Duvefelt, E. Iacobaeus, I. Kockum, E. Wallström, M. Khademi, A. Oturai, L.P. Ryder, J. Saarela, H.F. Harbo, E.G. Celius, H. Salter, T. Olsson, J. Hillert, Variation in interleukin 7 receptor α chain (IL7R) influences risk of multiple sclerosis, Nat. Genet. 39 (9) (2007) 1108–1113, https://doi.org/10.1038/ng2106.

[24] D. Manousaki, T. Dudding, S. Haworth, Y.-H. Hsu, C.-T. Liu, C. Medina-Gómez, T. Voortman, N. van der Velde, H. Melhus, C. Robinson-Cohen, D.L. Cousminer, M. Nethander, L. Vandenput, R. Noordam, V. Forgetta, C.M.T. Greenwood, M.L. Biggs, B.M. Psaty, J.I. Rotter, J.B. Richards, Low-frequency synonymous coding variation in CYP2R1 has large effects on vitamin D levels and risk of multiple sclerosis, Am. J. Hum. Genet. 101 (2) (2017) 227–238, https://doi.org/10.1016/j.ajhg.2017.06.014.

[25] L.M. Maier, C.E. Lowe, J. Cooper, K. Downes, D.E. Anderson, P.M. Clark, B. Healy, N. Walker, C. Aubin, J.R. Oksenberg, S.L. Hauser, A. Compston, S. Sawcer, T. International, M. Sclerosis, P. L. De Jager, L.S. Wicker, J.A. Todd, D.A. Hafler, IL2RA Genetic Heterogeneity in Multiple Sclerosis and Type 1 Diabetes Susceptibility and Soluble Interleukin-2 Receptor Production 5 (1) (2009), https://doi.org/10.1371/journal.pgen.1000322.

[26] A. Mulas, M.G. Piras, M. Lobina, S. Lai, M. Marongiu, V. Serra, M. Marongiu, G. Sole, F. Busonero, A. Maschio, M. Piga, D. Firinu, I. Kockum, I.L. Bomfim, T. Olsson, L. Alfredsson, A. Suarez, P.E. Carreira, M. Marchini, S. Sanna, new england journal (2017), https://doi.org/10.1056/NEJMoa1610528.

[27] J.A. Hollenbach, J.R. Oksenberg, The immunogenetics of multiple sclerosis: a comprehensive review, J. Autoimmun. 64 (2015) 13–25, https://doi.org/10.1016/j.jaut.2015.06.010.

[28] L. Moutsianas, L. Jostins, A.H. Beecham, A.T. Dilthey, D.K. Xifara, M. Ban, T.S. Shah, N.A. Patsopoulos, L. Alfredsson, C.A. Anderson, K.E. Attfield, S. E. Baranzini, J. Barrett, T.M.C. Binder, D. Booth, D. Buck, E.G. Celius, C. Cotsapas, S. D'Alfonso, G. McVean, Class II HLA interactions modulate genetic risk for multiple sclerosis, Nat. Genet. 47 (10) (2015) 1107–1113, https://doi.org/10.1038/ng.3395.

[29] C.S. Constantinescu, N. Farooqi, K. O'Brien, B. Gran, Experimental autoimmune encephalomyelitis (EAE) as a model for multiple sclerosis (MS), Br. J. Pharmacol. 164 (4) (2011) 1079–1106, https://doi.org/10.1111/j.1476-5381.2011.01302.x.

[30] A.P. Robinson, C.T. Harp, A. Noronha, S.D. Miller, The experimental autoimmune encephalomyelitis (EAE) model of MS. utility for understanding disease pathophysiology and treatment, Handb. Clin. Neurol. 122 (2014), https://doi.org/10.1016/B978-0-444-52001-2.00008-X. Elsevier B.V.

[31] K. Arima, A. Kinoshita, H. Mishima, N. Kanazawa, T. Kaneko, T. Mizushima, K. Ichinose, H. Nakamura, A. Tsujino, A. Kawakami, M. Matsunaka, S. Kasagi, S. Kawano, S. Kumagai, K. Ohmura, T. Mimori, M. Hirano, S. Ueno, K. Tanaka, K.-I. Yoshiura, Proteasome assembly defect due to a proteasome subunit beta type 8 (PSMB8) mutation causes the autoinflammatory disorder, Nakajo-Nishimura syndrome, Proceedings of the National Academy of Sciences of the United States of America 108 (36) (2011) 14914–14919, https://doi.org/10.1073/pnas.1106015108.

[32] K. Ohmura, Nakajo – Nishimura Syndrome and Related Proteasome-Associated Autoin Fl Ammatory Syndromes, 2019, pp. 259–265.

[33] S.K. Paul, M. Saddam, K.A. Rahaman, J.G. Choi, S.S. Lee, M. Hasan, Molecular modeling, molecular dynamics simulation, and essential dynamics analysis of grancalcin: an upregulated biomarker in experimental autoimmune encephalomyelitis mice, Heliyon 8 (10) (2022) e11232, https://doi.org/10.1016/j.heliyon.2022.e11232.

[34] M. Hasan, J.E. Seo, K.A. Rahaman, H. Min, K.H. Kim, J.H. Park, C. Sung, J. Son, M.J. Kang, B.H. Jung, W.S. Park, O.S. Kwon, Novel genes in brain tissues of EAE-induced normal and obese mice: upregulation of metal ion-binding protein genes in obese-EAE mice, Neuroscience 343 (2017) 322–336, https://doi.org/10.1016/j.neuroscience.2016.12.002.

[35] J.H. Ward, Hierarchical grouping to optimize an objective function, J. Am. Stat. Assoc. 58 (301) (1963) 236–244, https://doi.org/10.1080/01621459.1963.10500845.

[36] T. Reutterer, D. Dan, Cluster analysis in marketing research, Handbook of Market Research XX (May) (2021) 221–249, https://doi.org/10.1007/978-3-319-57413-4_11.

[37] M. Rahimi, M. Taghdir, F. Abasi Joozdani, Dynamozones are the most obvious sign of the evolution of conformational dynamics in HIV-1 protease, Sci. Rep. 13 (1) (2023) 14179, https://doi.org/10.1038/s41598-023-40818-x.

[38] D.A. Ferrington, D.S. Gregerson, Immunoproteasomes: structure, function, and antigen presentation, Progress in Molecular Biology and Translational Science 109 (2012) 75–112, https://doi.org/10.1016/B978-0-12-397863-9.00003-1.

[39] C.C. Goetzke, F. Ebstein, T. Kallinich, Role of proteasomes in inflammation, J. Clin. Med. 10 (8) (2021), https://doi.org/10.3390/jcm10081783.

[40] C.A. Orengo, A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, J.M. Thornton, Cath - a hierarchic classification of protein domain structures, Structure 5 (8) (1997) 1093–1109, https://doi.org/10.1016/s0969-2126(97)00260-8.

[41] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: The Proteomics Protocols Handbook, 2009, p. 571, https://doi.org/10.1385/1-59259-890-0.

[42] E.F. Pettersen, T.D. Goddard, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J.H. Morris, T.E. Ferrin, UCSF ChimeraX: structure visualization for researchers, educators, and developers, Protein Sci. : A Publication of the Protein Society 30 (1) (2021) 70–82, https://doi.org/10.1002/pro.3943.

[43] S.K. Paul, M. Saddam, K.A. Rahaman, J.-G. Choi, S.-S. Lee, M. Hasan, Molecular modeling, molecular dynamics simulation, and essential dynamics analysis of grancalcin: an upregulated biomarker in experimental autoimmune encephalomyelitis mice, Heliyon 8 (10) (2022) e11232, https://doi.org/10.1016/j.heliyon.2022.e11232.

[44] P.R. Arantes, M.D. Polêto, C. Pedebos, R. Ligabue-Braun, Making it rain: cloud-based molecular simulations for everyone, J. Chem. Inf. Model. 61 (10) (2021) 4852–4856, https://doi.org/10.1021/acs.jcim.1c00998.

[45] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindah, 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.

[46] 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 78 (8) (2010) 1950–1958, https://doi.org/10.1002/prot.22711.

[47] I.Y. Akbayrak, S.I. Caglayan, S. Durdagi, L. Kurgan, V.N. Uversky, B. Ulver, H. Dervisoğlu, M. Haklidir, O. Hasekioglu, O. Coskuner-Weber, Structures of MERS-CoV macro domain in aqueous solution with dynamics: impacts of parallel tempering simulation techniques and CHARMM36m and AMBER99SB force field parameters, Proteins 89 (10) (2021) 1289–1299, https://doi.org/10.1002/prot.26150.

[48] V.H. Man, X. He, J. Gao, J. Wang, Effects of all-atom molecular mechanics force fields on amyloid peptide assembly: the case of PHF6 peptide of tau protein, J. Chem. Theor. Comput. 17 (10) (2021) 6458–6471, https://doi.org/10.1021/acs.jctc.1c00028.

[49] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, J. Comput. Chem. 18 (12) (1997) 1463–1472, https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[50] A. Ghazouani, J. M'halla, Simple computing of the viscosity of water-dioxane mixtures, according to a fluctuating SPC/E-I(h) interstitial model, J. Comput. Chem. 38 (22) (2017) 1952–1965, https://doi.org/10.1002/jcc.24841.

[51] J.-B. Linse, J.S. Hub, Three- and four-site models for heavy water: SPC/E-HW, TIP3P-HW, and TIP4P/2005-HW, J. Chem. Phys. 154 (19) (2021) 194501, https://doi.org/10.1063/5.0050841.

[52] R. Martonák, A. Laio, M. Parrinello, Predicting crystal structures: the Parrinello-Rahman method revisited, Phys. Rev. Lett. 90 (7) (2003) 75503, https://doi.org/10.1103/PhysRevLett.90.075503.

[53] R.T. McGibbon, K.A. Beauchamp, M.P. Harrigan, C. Klein, J.M. Swails, C.X. Hernández, C.R. Schwantes, L.-P. Wang, T.J. Lane, V.S. Pande, MDTraj: a modern open library for the analysis of molecular dynamics trajectories, Biophys. J. 109 (8) (2015) 1528–1532, https://doi.org/10.1016/j.bpj.2015.08.015.

[54] M.A. Balsera, W. Wriggers, Y. Oono, K. Schulten, Principal component analysis and long time protein dynamics, The Journal of Physical Chemistry 100 (7) (1996) 2567–2572, https://doi.org/10.1021/jp9536920.

[55] J. Lever, M. Krzywinski, N. Altman, Principal component analysis, Nat. Methods 14 (7) (2017) 641–642, https://doi.org/10.1038/nmeth.4346.

[56] A. Wolf, K.N. Kirschner, Principal component and clustering analysis on molecular dynamics data of the ribosomal L11·23S subdomain, J. Mol. Model. 19 (2) (2013) 539–549, https://doi.org/10.1007/s00894-012-1563-4.

[57] F. Sittel, A. Jain, G. Stock, Principal component analysis of molecular dynamics: on the use of Cartesian vs. internal coordinates, J. Chem. Phys. 141 (1) (2014) 14111, https://doi.org/10.1063/1.4885338.

[58] I.T. Jollife, J. Cadima, Principal component analysis: a review and recent developments, Phil. Trans. Math. Phys. Eng. Sci. (2065) (2016) 374, https://doi.org/10.1098/RSTA.2015.0202.

[59] X.J. Wang, L.J. Medeiros, C.E. Bueso-Ramos, G. Tang, S. Wang, Y. Oki, P. Desai, J.D. Khoury, R.N. Miranda, Z. Tang, N. Reddy, S. Li, P53 expression correlates with poorer survival and augments the negative prognostic effect of MYC rearrangement, expression or concurrent MYC/BCL2 expression in diffuse large B-cell lymphoma, Mod. Pathol. 30 (2) (2017) 194–203, https://doi.org/10.1038/modpathol.2016.178.

[60] G.G. Maisuradze, A. Liwo, H.A. Scheraga, Relation between free energy landscapes of proteins and dynamics, J. Chem. Theor. Comput. 6 (2) (2010) 583–595, https://doi.org/10.1021/CT9005745.

[61] L. Piao, Z. Chen, Q. Li, R. Liu, W. Song, R. Kong, S. Chang, Molecular dynamics simulations of wild type and mutants of SAPAP in complexed with shank3, Int. J. Mol. Sci. 20 (1) (2019), https://doi.org/10.3390/ijms20010224.

[62] M. Knudsen, C. Wiuf, The CATH database, Hum. Genom. 4 (3) (2010) 207–212, https://doi.org/10.1186/1479-7364-4-3-207.

[63] V. Nallapareddy, N. Bordin, I. Sillitoe, M. Heinzinger, M. Littmann, V.P. Waman, N. Sen, B. Rost, C. Orengo, CATHe: detection of remote homologues for CATH superfamilies using embeddings from protein language models, Bioinformatics 39 (1) (2023), https://doi.org/10.1093/bioinformatics/btad029.