Heliyon 8 (2022) e11232

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Molecular modeling, molecular dynamics simulation, and essential dynamics analysis of grancalcin: An upregulated biomarker in experimental autoimmune encephalomyelitis mice

Shamrat Kumar Paul [a], Md. Saddam [a], Khandoker Asiqur Rahaman [b], Jong-Gu Choi [c], Sang-Suk Lee [c,**], Mahbub Hasan [a,c,*]

[a] Department of Biochemistry and Molecular Biology, Life Science Faculty, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj 8100, Bangladesh
[b] Division of Biomedical Science and Technology, KIST-School, Korea University of Science and Technology, Seoul 02792, South Korea
[c] Department of Oriental Biomedical Engineering, Sangji University, Wonju 26339, South Korea

ARTICLE INFO

Keywords:
Molecular dynamics simulation
Principal component analysis
EAE
Multiple sclerosis
Grancalcin
GROMACS

ABSTRACT

The experimental autoimmune encephalomyelitis mouse model is the most commonly used animal model, and it best represents multiple sclerosis. Grancalcin (GCA) was discovered to be upregulated in EAE mice. GCA comprises 220 amino acids that have been assigned the UniprotKB ID Q8VC88. It is a calcium-binding protein that helps neutrophils adhere to fibronectin and the formation of focal adhesions. However, the protein data bank does not contain the crystal structure of mouse GCA. The current study aims to analyze the structural and physicochemical properties of GCA. Mouse GCA showed a high percentage identity (87%) with the crystal structure of des (1–52) grancalcin with bound calcium (chain A) from *Homo sapiens* identified by its PDB id 1k94_A. Using the SWISS-MODEL server, we used 1k94_A as a template protein to model the mouse GCA protein. Compared to the template structure 1K94, three potential binding sites for calcium-binding have been proposed, ranging from 13 to 20, 80 to 91, and 109 to 120 amino acids. On an i5 personal computer with 8GB of RAM, GROMACS 2020.1 was utilized to run a 100 ns molecular dynamics (MD) simulation. RMSD, Rg, and RMSF analysis of an MD simulation trajectory indicate a stable and compact state throughout the simulation period of modeled proteins. We found that GCA is primarily alpha helical (Class 1), with eight alpha helices. The essential dynamics analysis captures PCA and SASA, culminating in the biological motions that correspond to the last 1000 frames. These findings will aid the development of potential inhibitors as well as the determination of binding pockets and residues for drug-like molecules.

1. Introduction

Multiple sclerosis (MS) is an autoimmune, inflammatory disease in which the myelin sheaths around the axons of the central nervous system (CNS) are depreciated/damaged, resulting in demyelination and axonal lesion in the brain and spinal cord; the peripheral nervous system is hardly involved. It is influenced by principal cells of the immune system, CD4+ T cells: especially of TH1 and TH17 subsets, macrophages, and soluble inflammatory mediators [1, 2]. MS is divided into several subtypes: relapsing-remitting MS (RRMS), progressive MS (PMS), and -relapsing MS (PRMS); progressive MS is also subdivided into primary progressive MS (PPMS) and secondary progressive MS (SPMS). RRMS represents the stage of disease with acute exacerbations (relapse/attack). PPMS affects approximately 15% of people with MS where the neurological deterioration is present from the onset, most frequently without superimposed relapses. SPMS represents the phase of the disease where a gradual neurological deterioration (progression) follows a period of RR disease. PRMS refers to the rare variant of MS in which a few acute exacerbations are superimposed on the gradual PPMS-like course [3, 4, 5].

Experimental autoimmune encephalomyelitis (EAE) is an autoimmune disorder of the CNS having similar pathological and clinical features to MS and is most commonly used as an animal model of MS. Several antigens or proteins, including myelin basic protein, proteolipid protein, and myelin oligodendrocyte glycoprotein, play critical roles in

* Corresponding author.
** Corresponding author.
E-mail addresses: sslee@sangji.ac.kr (S.-S. Lee), mahbub.bmb@gmail.com (M. Hasan).

https://doi.org/10.1016/j.heliyon.2022.e11232
Received 17 February 2022; Received in revised form 30 May 2022; Accepted 20 October 2022

inducing EAE. Most importantly, it is initiated by entering myelin-specific encephalitogenic CD4+ T helper (Th) cells into the CNS [6]. Next, these myelin-specific T-cells are activated in the periphery and enter the CNS by binding with integrins adhesion molecules, followed by permeabilization of the blood-brain barrier (BBB) [7, 8].

The protein GCA encoded by the gene *gca*, is abundant in neutrophils and monocytes/macrophages [9]. This protein has 217 amino acids and a molecular weight of 22.4 kDa [10]. It belongs to a group of EF-hand Ca^{2+} binding proteins, including Alg-2, calpain, sorcin, and GCA. These have structural similarities, but their functions are quite different [11]. Alg-2 has been subjected to the induction of apoptosis, and calpain has proteolytic activity with effects on platelet function and cell migration through the organization of the cytoskeleton. Sorcin is associated with the cardiac ryanodine receptor and is liable for the modulation of Ca^{2+} channel activity [12]. These EF-hand Ca^{2+}-binding proteins have the particular characteristic of dimerization. Sorcin and GCA exist as homodimers, and calpains exist as heterodimers of the large catalytic and small regulatory subunits [13]. The primary structure of GCA includes an N-terminal domain of approximately 50 amino acids that are acetylated, of which 14 amino acids are removed post-translationally in the mature protein [14, 15]. The remaining 170 amino acids constitute five motifs with EF-hand features [16].

GCA is involved in various Ca^{2+}-dependent induced cellular functions by binding to specific target proteins. L-plastin/p65 has been shown as a binding partner of GCA. It belongs to the plastin group of proteins characterized by a series of Ca^{2+}, calmodulin, and actin-binding motifs. This isoform is expressed in hemopoietic cell lineages, including macrophages, neutrophils, and lymphocytes, and is phosphorylated in response to extracellular stimuli [9, 14]. GCA also interacts and binds with toll-like receptors (TLRs). It regulates type-I IFN production, cytokine/chemokine production through nuclear localization of interferon regulatory factor 7 (IRF7), NF-κB activation, and mitogen-activated protein kinase (MAPK) activation in plasmacytoid dendritic cells [17].

Grancalcin helps fibronectin-binding integrins with affinity maturation, surface expression, and surface distribution. Its effect on neutrophil adherence could be mediated by its interaction with L-plastin, which is involved in the control of integrin activity [1]. When integrins have been attached to the proteins of the extracellular matrix, they quickly develop associations with the actin cytoskeleton and cluster together to generate focal adhesions. The formation of focal adhesion complexes results in a 30% increase in the adhesion strength [2, 3, 18]. Myelin-Specific Encephalitogenic CD4+ T helper (Th) cells enter the CNS by binding with integrins and adhesion molecules, followed by permeabilization of the blood-brain barrier (BBB) [5, 7]. These T cells can produce inflammatory cytokines and have the ability to differentiate into Th1 (releasing IFN-gamma) or Th17 cells in response to stimulation (IL-17, IL-22, IL-21) [6]. Antigen-presenting cells (APCs) enter the CNS following the translocation of myelin-specific T cells, and this is critical for lymphocyte reactivation and the commencement of the inflammatory cascade in the development of EAE. The CNS parenchyma is invaded by granulocytes and macrophages, responsible for tissue inflammation and destruction [7]. Dendritic cells (DCs) have an efficient role in antigen presentation and chemokine/cytokine redemption [8]. Through the permeabilized BBB, granulocytes and macrophages are attracted by chemokines released and migrate into the CNS [9]. When autoreactive T cells interact with an antigen, they reactivate and develop, producing cytokines that stimulate nearby immune or neural cells and attract more inflammatory cells into the CNS. It is thought that activated macrophages, in particular, cause damage both indirectly and directly. These inflammatory and cytokine products destroy the myelin and axons of the CNS [10]. The formation of focal adhesion plays a critical role in the migration of leukocytes. Ping Xu et al. concluded that focal adhesion complex formation is reduced by 89% in grancalcin deficient mice [11]. A recent study by Hasan et al. reported an increased expression of *gca* in the brain tissue of EAE-induced mice [19].

A protein's amino acid sequences cannot provide its structural properties on their own, and the tertiary structure is critical to comprehending its biophysical nature. As a result, homology modeling becomes an essential tool in predicting protein structure based on the template with which the query proteins share the highest identity. This method becomes a viable way to investigate protein structures that experimental technologies have not yet solved. This in-silico study looked at mouse GCA protein homology modeling and Molecular Dynamics (MD) simulations. Due to the absence of the query protein's experimental structure in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) [20], a structural and functional analysis was performed to construct the in-silico three-dimensional structure and predict active residues. The workflow of current investigation depicted in Figure 1. This modeling and simulation study will pave the way for researchers to gain insight into the biophysical nature of GCA, its secondary structure components, and the prediction of active residues, which will be used to screen potential drug-like molecules and inhibitors.

2. Materials and methods

2.1. Primary sequence analysis

The primary sequence of GCA protein encoded by the gene *gca* from the organism *Mus musculus* was retrieved in FASTA file format from the UniProtKB [33] database using its UniProtKB identifier Q8VC88. Furthermore, the physicochemical properties of GCA protein were retrieved by submitting its FASTA file format of primary sequence to Expasy's Protparam server [34]. Subsequently, the number of alpha-helices present in the targeted GCA protein sequence, associated amino acid residues determination, presence of disordered amino acids in the whole protein sequence was analyzed by the DISOPRED tool of PSIPRED [35]. The secondary structure elements were predicted by both SOPMA [36] and Phyre2 [37] servers.

2.2. Homology modeling of GCA

The homology modeling approach is an essential bioinformatics pipeline for predicting the three-dimension structure of a target protein based on the target primary sequence information and template protein sequence alignment. Expasy's SWISS-MODEL [38] server was used to predict the 3-D structure of GCA. SWISS-MODEL performed HHblits against the SWISS-MODEL template library (SMTL) [39] and ProMod3 [40] program. Clustal Omega [41] server was used to align protein sequence between targeted GCA selected template protein. Protein structure visualization and missing residues identification in template protein structure were carried out by PyMol and UCSF Chimera [42] accordingly.

2.3. Structure validation of GCA

Structural validation is the following step of protein three-dimensional structure prediction. The predicted structure of GCA protein obtained through the homology modeling approach was validated using several servers. The PDB file of the expected structure was submitted to the structural analysis and verification server (SAVES) v6.0. The Ramachandran plot of the protein backbone structure was built using PROCHECK [43] to assess the stereochemical quality of protein. The ERRAT [24] server under saves v6.0 was used to evaluate the reliability of the predicted structure. The PROCHECK and ERRAT were also applied for the template protein to generate the Ramachandran plot and ERRAT value. The inspection of the overall protein structure is accomplished by determining of Z-score from the ProSA-web [44].

2.4. Active site prediction

The protein's binding site and the associated amino acids in the specific region give an idea of amino acids in the entire protein sequences

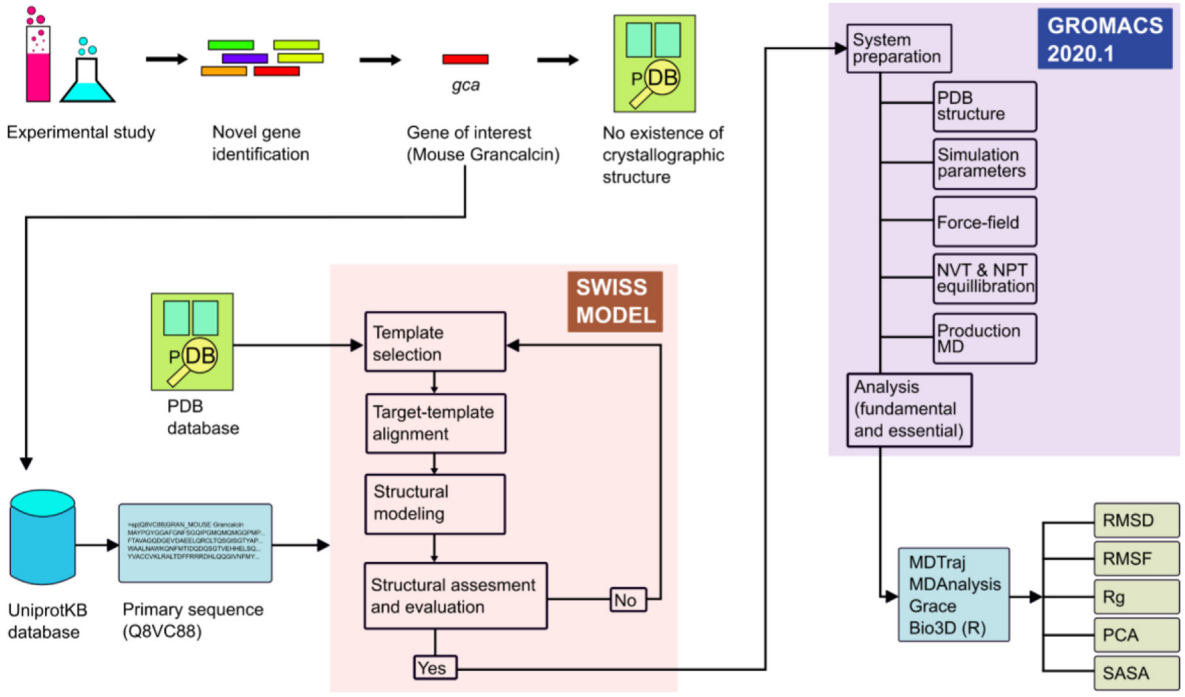

Figure 1. Study design and execution strategy: A schematic workflow, 3D structure building, and Molecular Dynamics Simulation of Mouse Grancalcin, which is encoded by the gca gene. The gene gca was selected for structural analysis by studying Hasan et al. 2017. The HHblits [21] program of SWISS-MODEL [22] determined the Grancalcin structure of *Homo sapiens*, 1k94_A as the template as it shares 87.27% identity with Mouse GCA. Following template selection. The 3D structure was predicted and then visualized by UCSF Chimera program. Predicted structure was validated by PROCHECK [23] and ERRAT [24]. The protein's binding sites were predicted using the CASTp [25] Server as well as compared with the Ca2+ binding site of the template structure 1K94_A. A 100 ns Molecular Dynamics Simulation was performed using GROMACS 2020.1 [26] to evaluate the stability of the protein. In the simulation, a 1.50 nm simulation box was created and the SPC water model was used to solvate the structure of the modeled protein. Then, the Na+ was added to the system to neutralize the negative charges, and the OPLS-AA was selected as the force field. System minimization was carried out following the Steepest Descent Algorithm, and then the system was equilibrated at Canonical Ensemble (NVT) and Isothermal Isoberic Ensemble (NPT) for 100 ps. For the development of the 100ns MD run, the LINCS algorithm [27], a 2 fs (fs) integration step, and the PME method [28] were used. We considered the last 1000 frames of the Molecular Dynamics trajectory to analyze the Essential Dynamics to capture PCA for understanding relevant conformational changes using programs namely Grace, MDAnalysis, MDTraj and Bio3D package of R [29, 30, 31]. Finally, CATH [32] analysis was performed to identify the protein's class, architecture, topology, and homologous superfamily.

participating in the protein-ligand interaction events. Furthermore, it also exerts the catalytic function facilitating the drug designing approach to design specific ligands for specific macromolecule targets. The active site prediction of the protein GCA was carried out by the CASTp server [25] and made a comparison between the protein data bank reports of Ca2+ ion binding site of template protein 1K94_A.

2.5. Molecular dynamics simulation

For the evaluation of the stability and consistency of the predicted structure of GCA, a 100 ns (100000 ps) of molecular dynamics (MD) simulation study on GROMACS 5.2 (2020.1) [26] has been conducted on an i5 computer with Ubuntu 20.04 Linux distribution. MD simulation study of template protein used for the structure prediction of GCA was also carried out to compare the trajectory data between modeled structure and predicted structure to understand how both forms (model and template) share their similarities in trajectory analysis. The protein topology was generated all-atom Optimized Potentials for Liquid Simulations (OPLS-AA) [45] force field.

Since introducing OPLS-AA and OPLS-AA/L force fields about a decade and a half ago, they have been widely used for protein and protein modification simulation [46, 47]. Notably, OPLS-AA and OPLS-AA/L produce promising results when calculating nonbonded interaction parameters [48, 49]. In addition, the following five amino acids are considered for enhancement and alignment with experimental rotamer distribution: serine, cysteine, aspartate, glutamate, and tryptophan. Furthermore, the development of peptide dihedral parameters significantly improved the ability of the OPLS-AA force field to reproduce both gas-phase conformer energies for longer peptides and aqueous phase experimental properties [50].

According to the secondary structure analysis, our query protein GCA contains a high proportion of the following amino acids: Asp (6.4%), Cys (1.8%), Glu (4.1%), Ser (5.9%), and Trp (1.4%), with Gly (10.9%) having the highest proportion. Similarly, Asp (7.3%), Cys (2.4%), Glu (4.8%), Ser (5.5%), and Trp (1.2%) are abundant in the template structure 1K94, while Leu (8.5%) is the most abundant. This is because new peptide dihedral parameters for the OPLS-AA force field have been developed and outperformed previous parameters. Therefore, we choose this force-field to study the molecular dynamics simulation of our query protein and template 1K94_A because they are abundant in these five amino acids, Asp, Cys, Glu, Ser, and Trp.

Moreover, simple point charge (SPC) water model spc216, a generic equilibrated 3-point solvent model, was used to solvate the modeled protein structure. A cubic simulation box size of 1.50 nm was generated around the predicted model and solvated with an SPC water model. The neutralization of the negative charge was carried out by adding 4 Na + ions. To get rid of the edge effect, periodic boundary conditions were implied in all directions.

The energy minimization of the system was carried out using the steepest descent algorithm with maximum steps of 50,000 both for the predicted protein structure and the template protein structure, where the consideration of tolerance was 1000 kJ mol $^{-1}$ nm $^{-1}$. After system minimization, it was equilibrated at both canonical ensemble (NVT) and isothermal–isobaric ensemble (NPT) for 100 ps. The Particle Mesh Ewald (PME) [28] summation was used for long-range electrostatic interactions with a PME order. The Linear Constraint Solver (LINCS) algorithm [27]

was applied to constrain the bonds containing hydrogen atoms, and the SETTLE algorithm [51] was used to constrain the geometry of water molecules. The application of the V-rescale weak coupling method regulated the temperature at 300K, whereas continuous maintenance of pressure at 1 atm (1.01325 bar) was modulated by Parrinello- Rahman method [52]. The LINCS [27] algorithm and a 2 fs (fs) integration step was used for the 100 ns MD run production with no restraints. Also, for the Coulombic and Lennard-Jones interactions [53], the PME method [28] was implemented.

2.6. Essential dynamics analysis

The principal component analysis is a prominent method for extracting functionally significant collective motions from a molecular dynamics (MD) trajectory. We followed the analysis of the essential dynamics from previously described methods [54, 55, 56]. In this study, initially, we considered the last 1000 frames for analyzing principal components using the Bio3D package of R [29]. After excluding the rotational and translational movements, a 3N*3N covariance matrix was constructed by superimposing coordinates onto a reference structure. The resulting symmetric matrix was then diagonalized and yielded 3N eigenvectors. Theoretically, the PCs can be calculated through the diagonalization of the data covariance (or correlation) matrix C [54]. The diagonal matrix Λ represents the eigenvalues as diagonal entries, and the matrix **V** holds the associated eigenvectors.

$$C = V \Lambda VT \quad (1)$$

The eigenvectors are related to the direction of motion. Each has a corresponding eigenvalue, reflecting the entire mean-square variation along that eigenvector of the system [57]. The eigenvector with the most significant eigenvalue accounts for the highest variance within the data. The second component is orthogonal to the first and accounts for the second-highest proportion of variance, and so on. We used the k-means and hierarchical clustering methods to cluster the trajectories in the PC space. By reducing the mean squared distance between each observation and its nearest cluster center, k-means splits the observations into k clusters. A hierarchical clustering constructs a hierarchy of the groups by calculating the distance between the observations.

To analyze C-alpha backbone trajectory as well as slicing the trajectory for last 1000 frames we applied the atom selection and related functions of MDAnalysis and MDTraj [58, 30]. Additionally, applying the MDTraj python package [31], The RMSD distance matric and hierarchical clustering method were used to cluster the MD simulation trajectory of GCA and the template structure 1K94_A. Computation of all pairwise RMSDs between conformations results in RMSD average linkage hierarchical clustering dendrogram. Finally, we 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 down into the reduced dimensional space of both modeled and template structures. Cartesian coordinates act 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 modeled and template structure.

2.7. Class, architecture, topology, and homologous superfamily (CATH) analysis

At the class level, the domains are divided into four groups based on their secondary structure content: (a) mainly alpha, (b) mainly beta, (c) mixed alpha-beta, and (d) domains with few secondary structures [59]. Then, the domains within each class are ordered to define the three-dimensional layout of secondary structure elements, regardless of how those parts are related (architecture level). The level of topology describes the structural similarity between secondary components. The structural and functional characteristics are examined at the level of the Homologous superfamily [60, 32].

We have analyzed the protein domains based on evolutionary similarity using the Protein Data Bank (PDB) structure data and assigned them to the same superfamily within the Class, Architecture, Topology, and Homologous Superfamily (CATH) classification hierarchy. CATH stands for the four levels of the structural hierarchy [61]. Each protein structure is broken down into one or more chains, which are then subdivided into one or more domains, before classification into homologous superfamilies based on structure and function [62].

3. Results and discussion

3.1. Primary and secondary structure analysis

Retrieving the complete amino acid sequence of the Mouse GCA from UniProtKB [33], the BLASTp search was performed against the RCSB-PDB database in quest of template protein which is used as template structure in SWISS-MODEL for constructing homology modeled protein. However, HHblits search of the SWISS-MODEL identified several hits of template proteins [63, 64]. Among these hits, the Crystal structure of des (1–52) GCA with bound calcium from the organism *Homo sapiens* identified as the PDB code 1k94_A shares the identity of 87.27% with our query protein (GCA), and it has been selected as the template crystal structure. Generally, more than 40% or higher identity between the query protein and the template protein structures are acceptable for the comparative homology modeling structure construction [65]. Multiple sequence alignment (MSA) shows the conserved region and identity illustration depicted in Figure 2 [66]. Most amino acid residues are under the cut-off value (0.5), except the amino acid residues from 1-52 (Figure 3A), whereas the secondary structure elements are depicted in Figure 3B. PSIPRED webserver accomplished the disorder prediction task and the secondary structure element prediction. The predicted structure of GCA reveals an almost similar pattern of active site residues and binding pocket compared to its template 1k94_A (Figure 4).

ProtParam tool of Expasy was used to compute the different parameters of physicochemical properties and represented in Table 1. The isoelectronic point or isotonic point is the pH at which the amino acid does not migrate in an electric field. Therefore, the amino acid is neutral, i.e., the zwitterion is a dominant form in solution. Calculating isoelectric point (pI) is beneficial because at pI, solubility is negligible and mobility in an electro-focusing system becomes zero. As a result, at pI, proteins are available more stable position and compactable state. The computed pI value of our studied protein was 4.89, indicating that our protein of interest to construct a homology model will be acidic. The Isoelectric focusing method is applied for recombinant protein purification. The parameter for Expasy's ProtParam tool to calculate the extinction coefficient is 276, 278, 279, 280, and 282 nm wavelengths. However, the 280 nm is the most suitable value in this wavelength proteins absorb light strongly. The extinction coefficient of our query protein at 280 nm is 36120 M^{-1} cm^{-1} concerning the concentration of Cys, Trp, and Tyr. It depicts that there will be a higher possibility of available high concentration of Cys, Trp, and Tyr. The computed extinction coefficients will benefit the quantitative analysis of protein-protein and protein-ligand interactions in solution [68].

Whether a protein will be stable or not in a test tube can be estimated by tracking the Instability Index (II) postulated by Guruprasad et al. [69]. This Instability Index (II) evaluates the regional instability by calculating the weighted sum of dipeptides in a particular protein. In unstable proteins, this regional instability due to instability index becomes larger in value compared to stable proteins. An instability index below the value of 40 ensures a stable protein. On the other hand, values more than 40 refer to a sign of an unstable state of a protein [70]. The instability index value for our query protein was found to be 38.11. The value of II of our considered protein is less than 40, therefore thought a stable protein. The relative volume of a protein occupied by the aliphatic side chains (A, V, I,

Figure 2. Multiple Sequence Alignment (MSA) between GCA and 1k94_A. MSA between GCA and 1k94_A reveals that at the beginning of the sequence (1-55aa) 1k94_A, the template protein structure does not share any similarity or identity with the query sequence GCA. However, from amino acid sequence number 56, both sequences share a dynamic range of resemblance and identity.

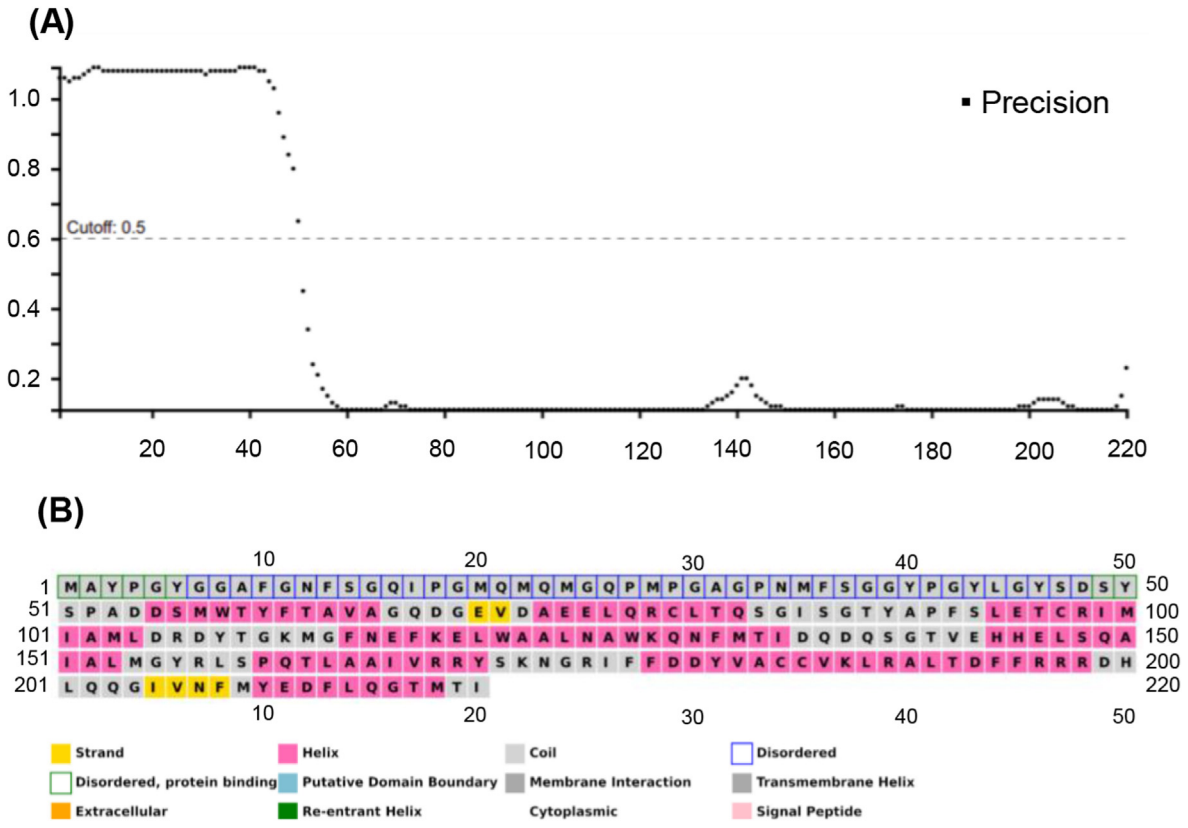

Figure 3. Prediction of disordered amino acids and secondary structure elements (PSIPRED) (A) Most amino acid residues remain below the cut of value (0.5), while 52 residues from regions 1–52 belong beyond the cut-off value. (B) The presence of secondary structure elements is shown. Different color codes beneath the residue map show the secondary structure elements. The disorder amino acid residues and secondary structure elements were predicted through the PSIPRED webserver.

and L) in that protein is the primary determinant of the aliphatic index (AI). The rising thermal stability of globular proteins is considered a good factor [71].

The AI for our query protein sequence is 61.73. This high AI value suggests that our query protein will be stable for a wide temperature range [72]. The grand average hydropathy (GRAVY) value for a particular peptide or protein is the sum of hydropathy values of all the amino acids in a protein divided by the number of residues in the sequence [73]. GRAVY calculation of our query protein is −0.268. The lower the GRAVY value, the better the interaction with water. After the synthesis, the time needed to disappear the half amount of the original protein is known as the half-life value [34]. Depending on the N-end rule, ProtParam linked up the half-life of a protein to the identity of its N-terminal residue. The half Half-life of our considered protein was found 30 h.

All the protein polypeptide chains were prearranged with 20 amino acids. Specific function of the protein depends on the characteristics of every constituent amino acid, and their location and roles determine the rate of polarity, charge, aliphatic and aromatic properties. Proteins are the best signal transducer through phosphorylation. Due to containing the hydroxyl side chain, Ser, Thr, and Tyr are three primary amino acid residues phosphorylated on a large scale by binding phosphate group [74]. Using Expasy's ProtParam tool, the highest amount of amino acid is

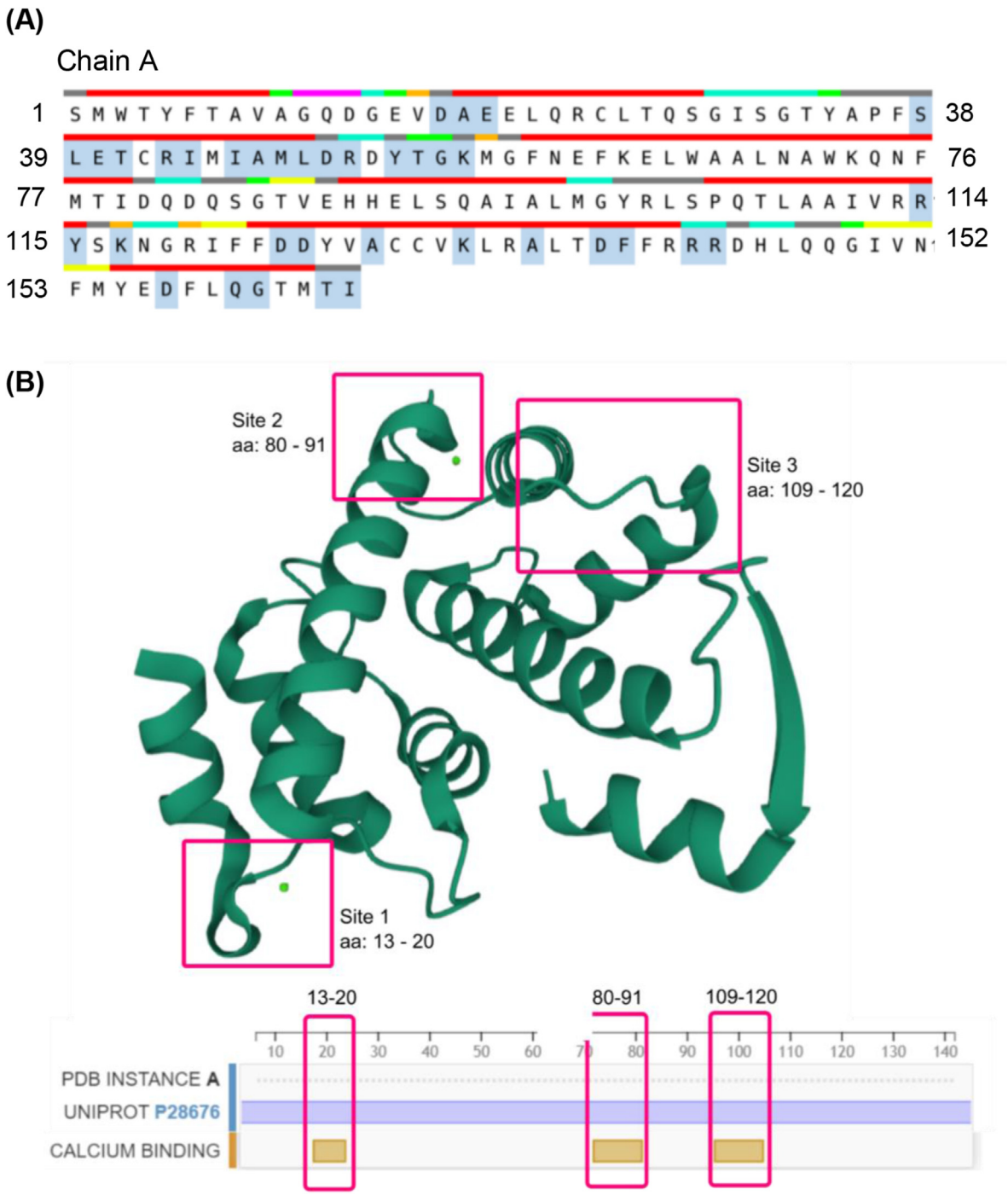

Figure 4. Binding site prediction (A) Active sites and associated amino acid residues prediction of GCA by the CastP Server. Indicated amino-acid sequence 17–19, 38–41, 43–44, 46–51, 53–56, 114–115, 117, 124–125, 128, 132, 135, 138–139, 142–143, 157, 160–161, 164–165 could be potential active sites where potential amino acid resides may show binding affinity towards drug-like molecules. (B) The protein data bank reports Ca2+ ion binding site of 1K94 [67]. According to the data, template structure 1K94 has three key binding sites that range from 13 to 20, 80 to 91, and 109 to 120.

found in Glycine (Gly) with 10.9 % and the lowest being Lysine (Lys) 2.3 %, shown in Table 2. The studied query protein was also rich in aliphatic amino acids, primarily Ala (A), Gly (G), Leu (L), and Val (A). The net percentage of hydrophobic amino acids was relatively higher in this protein, and the higher rate of hydrophobic amino acids increases the solubility of proteins in the lipid phase [75].

The prediction of the secondary structure of our query protein was completed by using the chou-Fasman method [76] and GOR tools [77]. The designed secondary structure of query protein of mouse GCA includes alpha helixes (46.36%), random coils (37.73%), beta-turn (8.72%), and extended strands (7.73%) shown in Table 3. Modeled structure and superimposed position of model protein with template protein are shown in Figure 5. For the flexibility of conformational changes and enzymatic turnover, Random coils are a considerably important factor. Data from the secondary structure of our query protein depicts that the higher amount of helix and coiling are giving the evidence of stronger and more compactable and bond. The GCA is an alpha-helices dominant protein and we found a total number of 8 alpha-helices in the predicted structure which is assessed by DISOPRED tool (Table 4).

Table 1. Physicochemical Properties of GCA calculated by Expasy's protparam tool.

Components	Values
Number of amino acids	220
Total number of atoms	3370
Molecular weight (Da)	24662.90
Theoretical pI	4.89
Extinction coefficients (M^{-1} cm^{-1}) at 280 nm in water, assuming all pairs of Cys residues form cystines	36,120
Extinction coefficients (M^{-1} $cm^{-1)}$ at 280 nm in water, assuming all Cys residues are reduced	35,870
Instability index (II)	38.11
Aliphatic index	61.73
Half-life (h)	30
Grand average of hydropathicity (GRAVY)	−0.268

Table 2. Amino acid composition of GCA calculated by Expasy's ProtParam tool.

	Query Protein (GCA)		Template (1k94_A)	
Amino Acid	Number	Percentage (%)	Number	Percentage (%)
Ala (A)	18	8.2	13	7.9
Arg (R)	11	5.0	10	6.1
Asn (N)	7	3.2	6	3.6
Asp (D)	14	6.4	12	7.3
Cys (C)	4	1.8	4	2.4
Gln (Q)	15	6.8	9	5.5
Glu (E)	9	4.1	8	4.8
Gly (G)	24	10.9	13	7.9
His (H)	3	1.4	4	2.4
Ile (I)	10	4.5	8	4.8
Leu (L)	15	6.8	14	8.5
Lys (K)	5	2.3	7	4.2
Met (M)	14	6.4	6	3.6
Phe (F)	14	6.4	11	6.7
Pro (P)	9	4.1	2	1.2
Ser (S)	13	5.9	9	5.5
Thr (T)	12	5.5	12	7.3
Trp (W)	3	1.4	2	1.2
Tyr (Y)	13	5.9	7	4.2
Val (V)	7	3.2	8	4.8

Table 3. Secondary Structure Elements of GCA determined by SOPMA.

Secondary Structure elements	Number	Percentage (%)
Alpha helix (Hh)	102	46.36
310 helix (Gg)	0	0.00
Pi helix (Ii)	0	0.00
Beta bridge (Bb)	0	0.00
Extended strand (Ee)	17	7.73
Beta turn (Tt)	18	8.18
Bend region (Ss)	0	0.00
Random coil (Cc)	83	37.73
Ambiguous states (?)	0	0.00
Other states	0	0.00

The sequence alignment of our query protein sequence and the template protein sequence was performed using the Clustal Omega web server [78]. Structure prediction using the amino acid sequence of the query protein and template from the HHblits search, accomplished by SWISS-MODEL (Schwede et al., 2003). The crystal structure of des (1–52) GCA with bound calcium from *Homo sapiens* (PDB accession: 1k94_A) shows the highest identity with our query protein. The predicted structure of GCA espoused similar structural features of the 1k94_A template (Figure 5A). The superimposition of the expected form of GCA and the template 1k94_A depicts the excellent quality of the structure (Figure 5B).

3.2. Model building and validation

SWISS-MODEL constructed a 3D structure through a comparative homology modeling approach. Unfortunately, there is a lack of experimental structures for mouse GCA in the RCSB-PDB. Therefore, the SWISS-MODEL server performed the modeling of the protein's three-dimensional structure. The Ramachandran plot's Phi and Psi distribution was calculated by observing the non-glycine, non-proline residual position illustrated in Figure 6. The crystal structure bound to calcium (1k94_A) from *Homo sapiens* is considered template protein using a SWISS-MODEL's template search program. The template selection was based on their comparatively higher query coverage, absence of missing residues, and substantial identity with query protein. As a result, the 1k94_A had an 87.27 percentage of uniqueness, and it was selected as the template protein for the model building of GCA (Table 5).

The molprobity web server reveals the stereochemical quality of the predicted GCA structure. The factors include Ramachandran plot quality, peptide bond planarity, Bad non bonded interactions, main chain hydrogen bond energy, C-alpha chirality, and overall G factor (Figure 6). The predicted GCA structure has 98.2% (160/163) of all residues were in favored regions (98%), 100% (163/163) of all residues in allowed (>99.8%) regions; and there were no outliers (phi, psi) (Figure 6A). The template structure 1k94_A has 99.4% (162/163) of all residues favored (98%) regions. 100.0% (163/163) of all residues were in allowed (>99.8%) regions, and there were no outliers (Figure 6B).

In addition, to confirm the quality of the model by measuring the Z-score, it ensures whether the predicted structure is in position in the allowable range of X-ray and NMR studies. The calculated Z-score of our predicted GCA structure was -6.14: better in structural quality (Figure 7A). The sequence position and knowledge-energy graph are depicted in Figure 7B. The PROSESS [79] webserver was used to determine the further quality check of the computationally modeled structure of GCA. The high score of covalent bond quality (6.5), covalent/packing quality (6.5), and torsional quality angle quality (7.5) indicates the better quality of the predicted structure: the overall quality factor represented in Figure 8A and the number of parameters with outlier for each residue depicted in Figure 8B.

3.3. Fundamental molecular dynamics simulation analysis

The simulation of molecular dynamics (MD) is widely used to understand the biophysical structural stability of protein structures [26]. The Groningen Machine for Chemical Simulations (GROMACS) was used for the 100 ns molecular dynamics simulation on an i5 Linux single workstation. The root-mean-square deviation (RMSD) of the protein backbone was calculated using a 100-ns MD simulation for the modeled protein and the template structure (Figure 9A). The system rises sharply and then settles into a stable state with variations ranging from 0.15 nm to 0.38 nm. A steady-state RMSD is observed for both the modeled protein and the template structure. Overall, the structural alignment of the modeled protein and the template protein (1k94 A) is good.

The fluctuation of amino acids and the regions of change in the protein is critical in explaining the activity and stability of the protein in molecular dynamics simulation. The RMSF trajectory data of molecular dynamics simulation can be used to identify amino acid movements (Figure 9B). The RMSF data shows that the fluctuations of most amino acids are between 0.06 and 0.45 nm, with the initial residues of the predicted structure diverging up to 0.4 nm. The predicted structure has the most significant variation in 50–75 residues. The resident amino acids in this fluctuating region would significantly impact its ability to

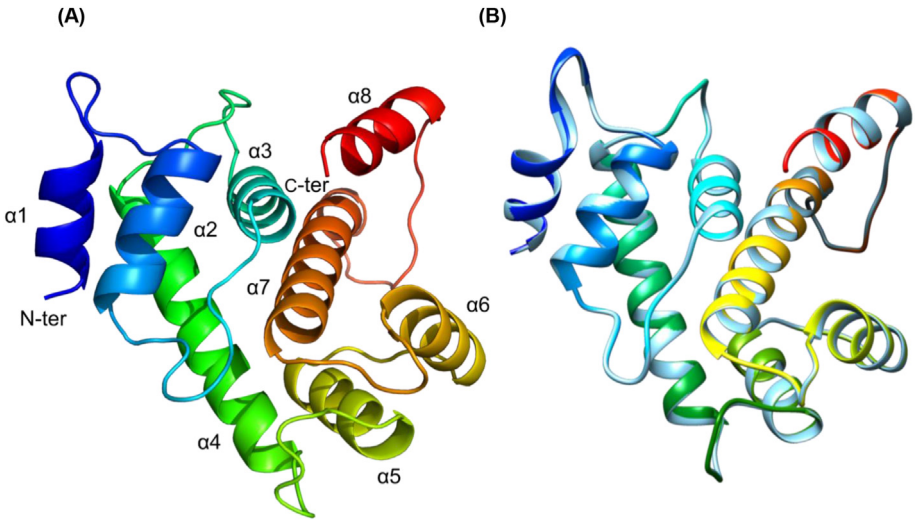

Figure 5. (A) Predicted Structure of GCA with the indicated alpha helixes. (B) Superimposition state of template and model. Gray represents the template_1k94_A, and rainbow-colored expresses the modeled protein (87.7% identity).

Table 4. The number of alpha-helices of GCA; Predicted by DISOPRED server.

Helix	Start residue	End residue	No. of residues	Sequence
1	Asp55	Ala65	11	DSMWTYFTAVA
2	Ala73	Gln82	10	AEELQRCLTQ
3	Leu94	Leu104	11	LETCRIMIAML
4	Phe114	Ile134	11	FNEFKELWAALNAWKQNFMTI
5	His144	Leu153	12	HHELSQAIAL
6	Pro160	Tyr170	11	PQTLAAIVRRY
7	Phe178	Arg198	22	FDDYVACCVKLRALTDFFRRR
8	Tyr210	Met218	9	YEDFLQGTM

bind ligands or drug-like molecules. The most intriguing aspect of this RMSF data is that both the predicted structure and the template protein exhibit the same pattern of RMS fluctuation.

The radius of gyration (Rg) is another important structural feature of the MD trajectory representing the MD simulation phenomenon. The Rg data shows that the predicted structure (showed in black) gradually increases up to around ~1.73 nm while the template structure (1k94_A showed in red) reaches around ~1.69 in the course of 5ns–10ns of simulation (out of 100 ns) (Figure 9C). However, the data represented in Figure 9C similarly depicts the structural equilibration. The compactness of the predicted model in the entire simulation period can be identified through the Rg data. The corresponding Rg data shows a relatively steady value for the predicted structure and template protein during the 100 ns MD simulation. The amino acid composition of both the predicted and template structure possesses a relatively similar number of Arg (11 *vs.* 10), Asn (7 *vs.* 6), Lys (5 *vs.* 7), Phe (14 *vs.* 11), and Thr (12 *vs.* 12). At the same time, for both cases, Gly is presented as the highest (13) in the query protein, and the predicted structure had 24 in number.

On the other hand, Trp is comparatively lower in number, with only three residues in query protein and only two residues of Pro in the template structure. And this wholesome, relatively similar number of amino acids depicts region-wise similarity in the RMS fluctuation. The RMS graph illustrates its close state in the entire simulation based on the fluctuation range for the predicted structure and the template protein (1k94_A).

3.4. Principal component analysis (PCA)

The scree plot was created by plotting the percentage of variance against the eigenvalue rank in numerical value. The highest proportion of variance has been denoted as 20.7 in percentage and goes up to 80.9%, with the eigenvalue ranking at 20 in numerical value (Figure 10A). This plot is for the last 1000 frames.

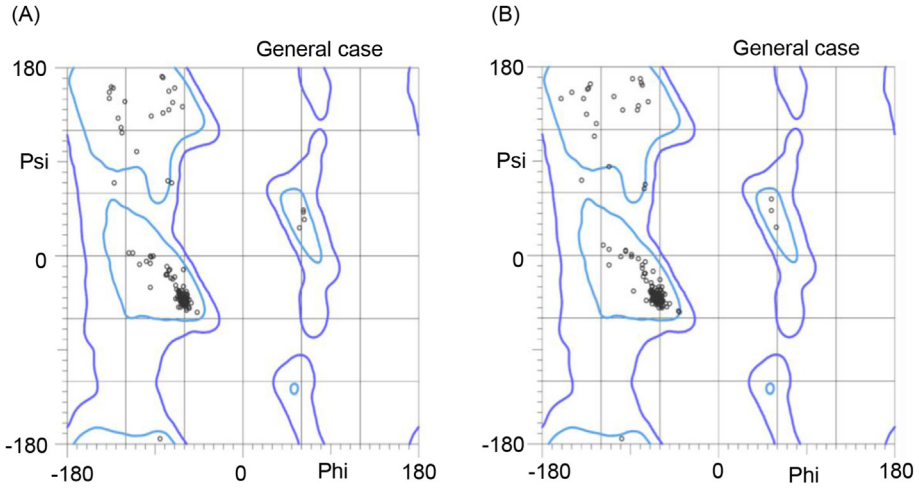

Figure 6. Ramachandran Plot analysis of modeled GCA protein (A) and template structure 1k94_A (B). Homology modeled protein (GCA) has 98.2% (160/163) of all residues were in favored regions (98%) regions, 100% (163/163) of all residues in allowed (>99.8%) regions; and there were no outliers (phi, psi) (A). The template structure 1k94_A has 99.4% (162/163) of all residues favored (98%) regions. 100.0% (163/163) of all residues were in allowed (>99.8%) regions, and there were no outliers (B).

Table 5. Template protein Selection based on the criterion for GCA structure prediction.

Description	Scientific name	Query cover	Per. identity	Accession
Crystal structure of des (1–52) GCA with bound calcium	*Homo sapiens*	75%	87.27%	1k94_A
Crystal structure of Gancalcin with Bound Calcium	*Homo sapiens*	75%	86.06%	1F4O_A
X-ray structure of calcium-free human sorcin	*Homo sapiens*	75%	65.66%	4UPG_A
The crystal structure of the natural F112L human sorcin mutant	*Homo Sapiens*	75%	65.66%	2JC2_A
Human SCBD (sorcin calcium-binding domain in complex with doxorubicin	*Homo Sapiens*	75%	65.66%	5MRA_A
The X-ray structure of the Sorcin Calcium Binding Domain (SCBD)	*Cricetulus longicaudatus*	75%	65.06%	1GJY_A

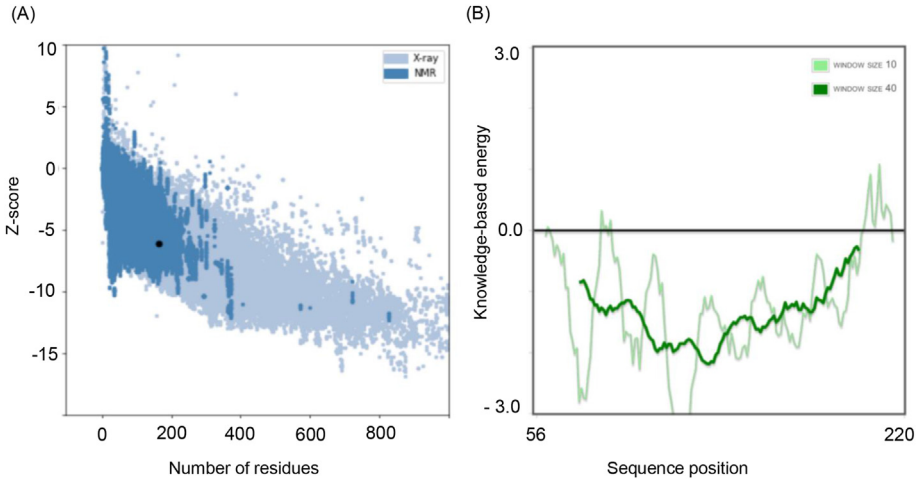

Figure 7. Validation of predicted GCA model structure by Z-score of ProSA web servers. Z-score confirms the quality of the expected structure. (A) For GCA, the curated data depicted that the structure is in the allowable range of X-ray and NMR study (overall score is −6.14). (B) The sequence position and knowledge-energy graph are in two different window sizes; window size 10 indicates light green, whereas deep green for window 40. This plot shows that the measured energy values for all of the residues of the GCA model are significantly lower than zero.

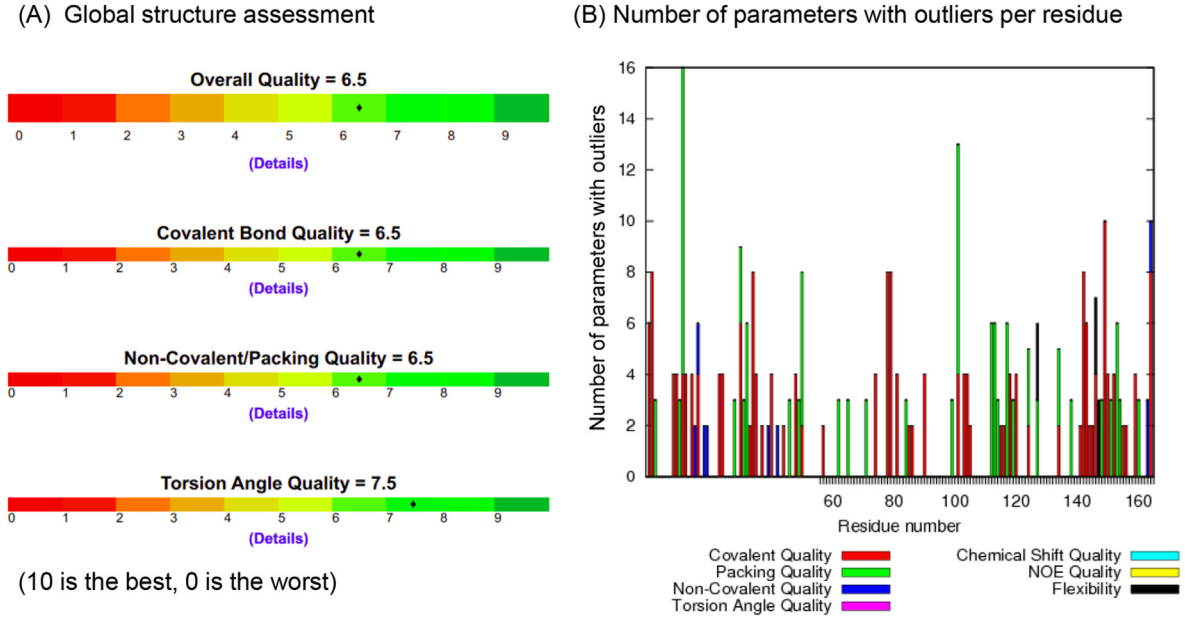

Figure 8. Quantitative evaluation of GCA by the PROSESS server. (A) Overall quality factor (B) Number of parameters with outlier for each residue. The covalent/packing quality (6.5) and torsional quality angle quality (7.5) indicate the better quality of the homology modeled structure of GCA.

The root-mean-square fluctuation (RMSF) measures the average divergence of a particle from a reference position over time. The dynamic behavior of residues and protein flexibility was assessed using an RMSF plot of alpha carbon atoms. The graph depicts the stability of carbon atoms of Cα In the RMSF analysis, we considered the fluctuations greater than 1A. RMSF Vs. residue position plot (Figure 10B) shows that fluctuations occur at residue positions 50 and 125. C-terminal fluctuations are usually shared and may not require further examination for the fluctuations. Furthermore, Dynamic Cross-Correlation Maps (DCCM) were analyzed to examine the concerted motions of protein residues. The anti-correlated motions are depicted in pink, while strongly correlated motions between individual residues are highlighted in paste color (Figure 10C).

The PCA of the atomic backbone (Cα position) was performed by three configurations: PC1, PC2, and PC3. According to blue and red color

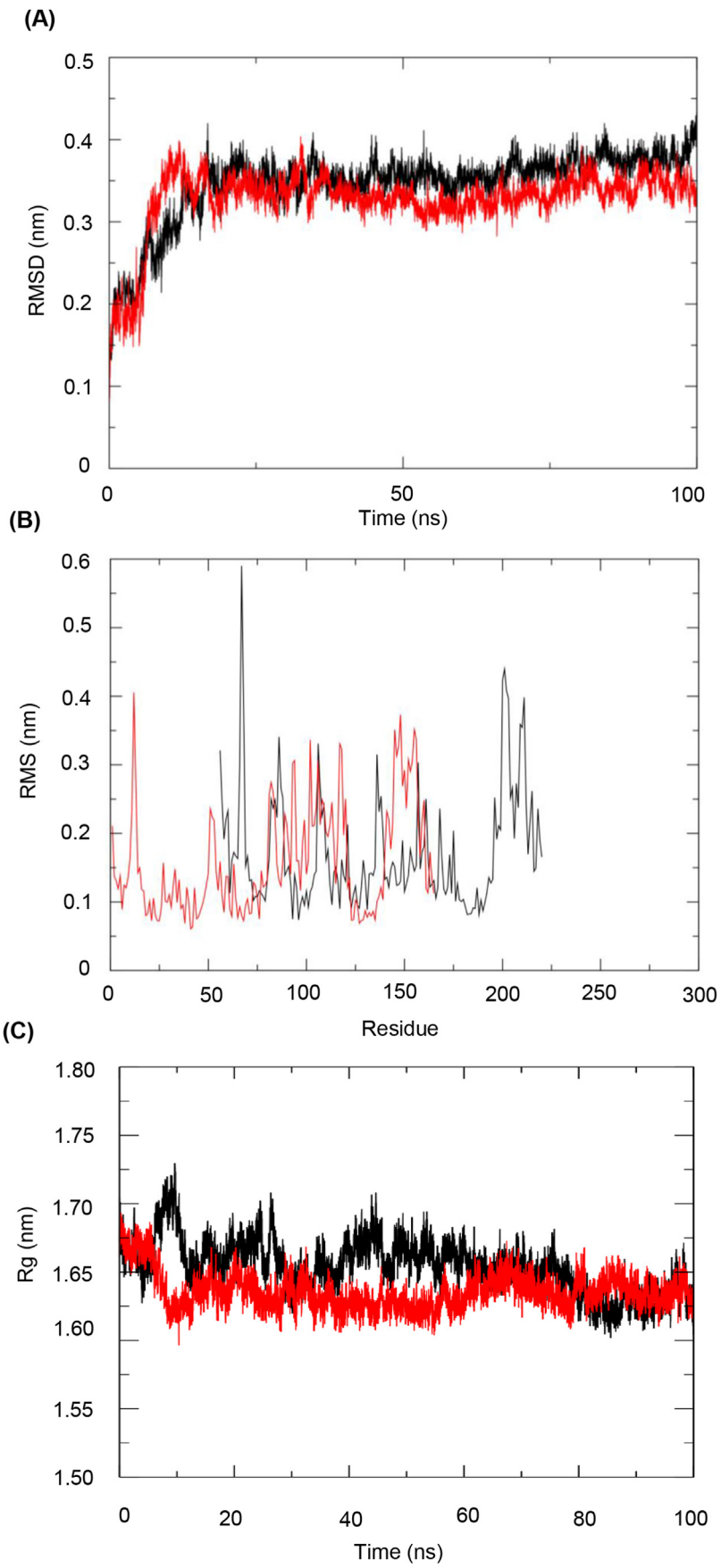

Figure 9. The results of molecular dynamics simulation. (A) RMSD of GCA (black) and template structure 1k94_A (red) during the entire 100 ns. (B) Backbone RMS fluctuation of GCA (black) and template structure 1k94_A (red) depicts two flexible regions. (C) Showing Radius of gyration (Rg) of GCA (black) and template structure 1k94_A (red). In Figure 8A, the RMSD trajectory of the template protein and the predicted structure shows that in the earlier few nanoseconds of the simulation period, the system rises sharply. The system gets to a stable state by converging the trajectory of template and query structure, and the variation follows the range of 0.15 nm–0.38 nm (up to 0.4. nm). The RMSF data (B) shows that amino acids fluctuations were in the range of 0.06 nm–0.45 nm with a vast level divergence up to 0.4 nm in the beginning. Rg data (C) shows the compactness of the predicted structure (line graph in black) with the reference of the template structure (line graph in red).

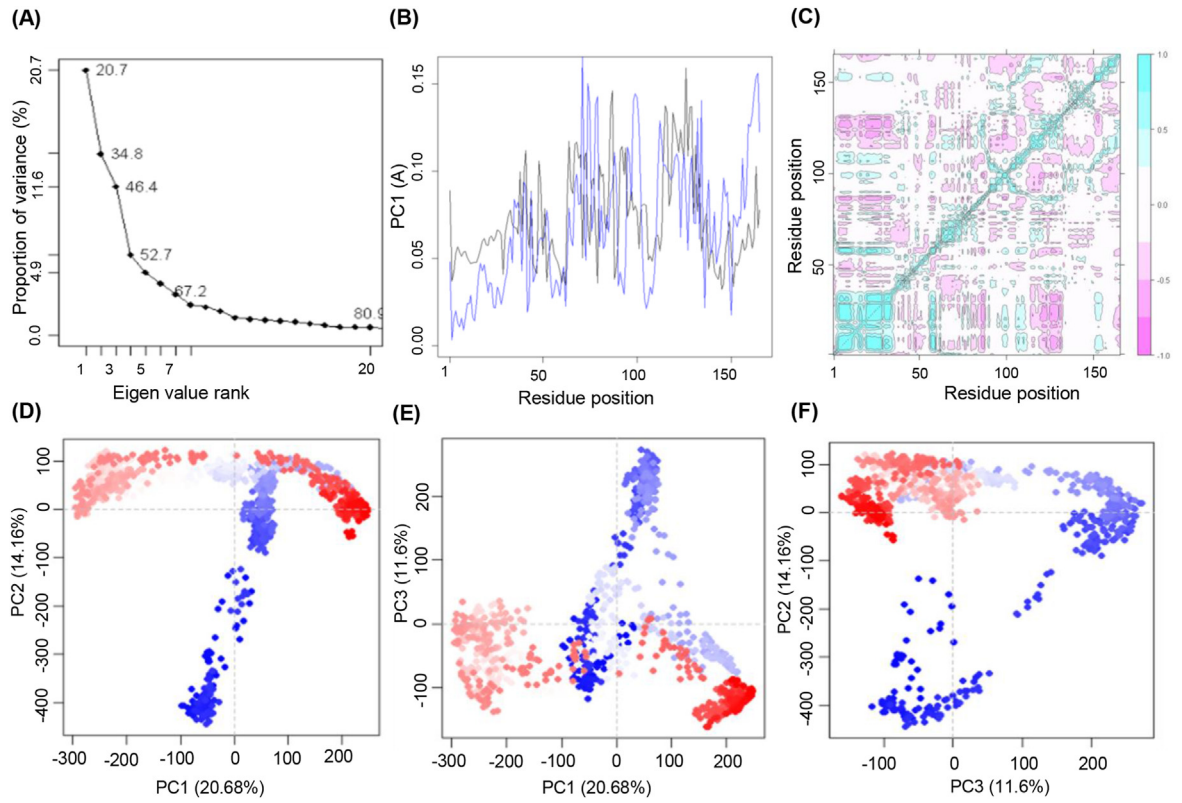

Figure 10. Principal component analysis (PCA) of Grancalcin. The principal components of the alpha carbon chain of Grancalcin (GCA) are estimated using the R-Bio3D package, and the last 1000 frames are taken into account for data interpretation. (A) The scree plot has been plotted according to the proportion of variance in percentage against its eigenvalue rank in neumerical value. (B) The root mean square fluctuation has been plotted against the residue position for PC1, which has the most components in comparison to the total components. (C) A depiction of cross-correlated residues on the Grancalcin (GCA). The pink color represents anti-correlated motion, while the paste color represents correlated motion. Depending on the complex dominant motion, the first three principle components analysis of PC2 vs. PC1, PC3 vs. PC1 and PC2 vs. PC3 are shown in figures D, E, and F, respectively. These three components account for 46.4% of the variance. The first PC accounts for more than a third of the variation (20.68%), giving it a large advantage in variance dominance. PC2 has 14.16% variability while PC3 11.6%.

clusters, the binding cluster frames were divided into two coordinates. In the analysis, the first PC is responsible for more than a third of the total variance (20.68%), which gives it a significant lead in terms of overall variance dominance. PC2 accounts for 14.16%, and PC3 possesses the lowest variability of 11.6%. The first three components account for 46.44% of the total variance (Fig. 10D–F). To understand GCA's alpha carbon chain's behavior by comparing it with its template 1K94_A, we considered analyzing RMSD average linkage hierarchical clustering, Cartesian coordinate PCA, pairwise distance PCA, and the solvent-accessible surface area (SASA) of entire 100ns simulation trajectory. We chose the last 1000 frames for clustering simulation trajectory data of the modeled and template protein.

From the aspect of modeled GCA structure, the cluster appeared as a chartreuse yellow color depicting the position of principal components in the last 1000 frames (Figure 11A). The cartesian coordinate PCA (Figure 11B) denotes the principal components of the last 1000 frames in chartreuse yellow to dark yellow colored area. Also, we can observe the pairwise distance PCA of GCA and the template structure 1K94_A in Figure 11C. On the other hand, the chartreuse yellow-colored cluster for the template structure 1K94_A can be found in the indicated chartreuse yellow-colored region (Figure 11D). At the same time, the cartesian coordinate PCA shows the chartreuse yellow-colored principal components are presented as the last 1000 frames in the fourth coordinate (Figure 11E), and Figure 11F illustrates the Pairwise distance PCA.

Computation of solvent accessible surface area (SASA) was accomplished using the algorithm from Shrake and Rupley, where the MDTraj function md. shrake_rupley has been applied. The resulting SASA array conveys that for our modeled protein from the beginning of 80 ns to around 90 ns, the high SASA value (~130 nm2) is presented (Figure 11G). A high SASA value of template structure 1K94 is between 85 ns and 100 ns, with some other peaked values around 60 ns (Figure 11H).

3.5. Structural analysis: CATH

We have examined the domains of template protein (1k94_A) using the CATH/Gene3D v4.3 database [30]. The domains are classified as 1.10.238.10, where each number represents a unique CATH code. The result shows that the template protein belongs to class 1 (C = 1), indicating that it is mainly alpha-helical. Data from Table 6. The secondary structure elements of GCA determined by SOPMA also confirm this. In class 1, there are five distinct architectures, including up-down bundle, alpha/alpha barrel, and horseshoe. 1k94 A belongs to the orthogonal bundle architectures (A = 10), indicating its more spherical structure. The orthogonal bundle architecture contains 290 distinct topologies, including DNA polymerase; domain 1, Recoverin; domain 1, Helix hairpins, Globin-like, Transferase domain 1, etc. 1k94_A is a member of the Recoverin; domain 1 topology (T = 238), indicating that our template protein can interact with Ca^{2+}. This topology has 18 distinct superfamilies, including EF-hand, formin, pheromone, etc. The 1k94_A is a part of the EF-hand superfamily (H = 10), indicating this protein has calcium-binding motifs. CATH analysis shows that 1k94_A is mainly alpha-helical in class and Orthogonal Bundle in architecture. It belongs to the EF-hand superfamily and has a Recoverin domain 1. Therefore, we can propose that the GCA protein has a similar class, architecture, topology, and homologous superfamily as the 1k94_A is 87% identical to the Grancalcin.

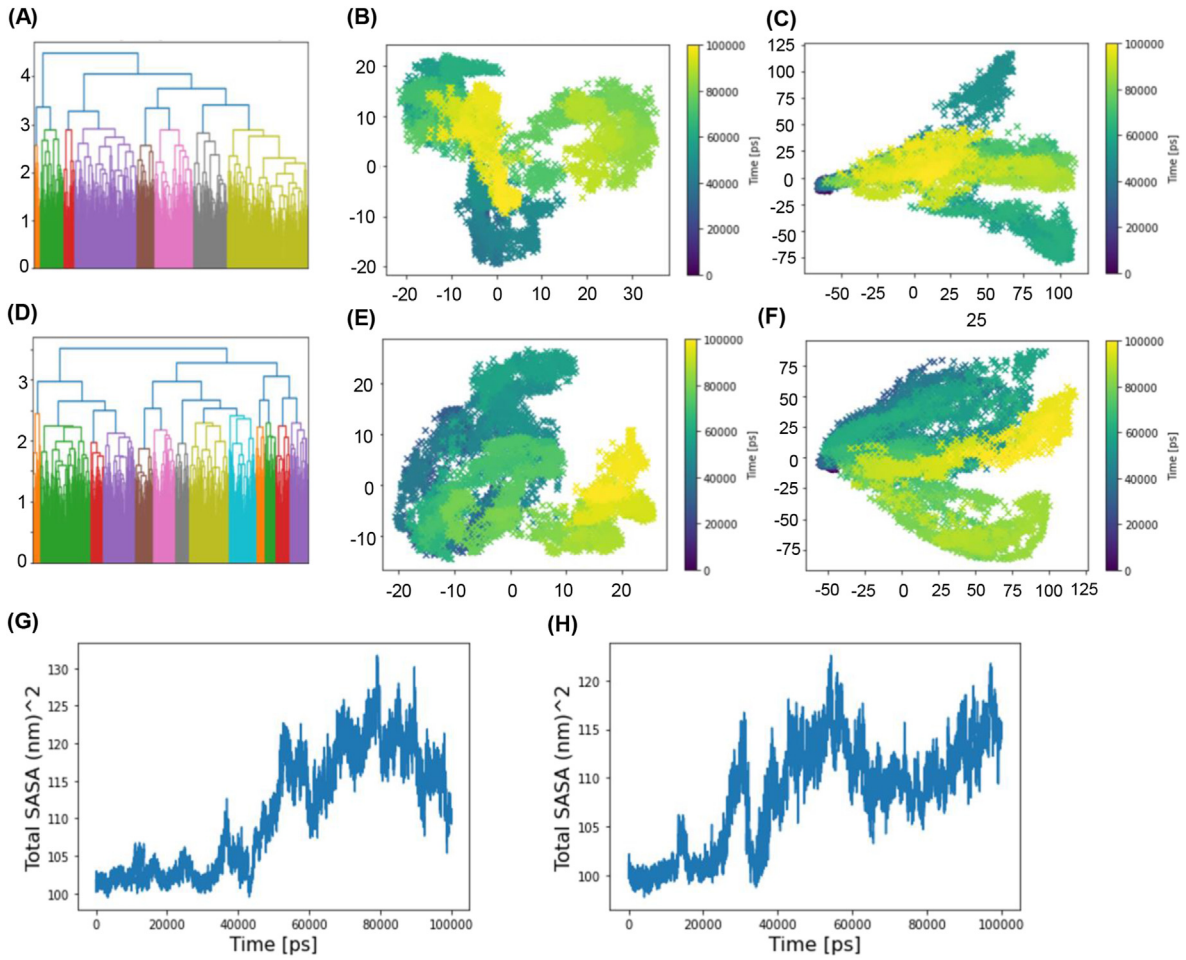

Figure 11. Essential dynamics analysis of Grancalcin and the template 1k94_A. Figure A, B, C and D depict the clustering dendrogram and PCA plot of last 1000 frames of the entire simulation period for the modeled structure (GCA) and the template structure (1k94) accordingly. (E) The states of the principal component (depicted in yellow) of 1k94_A for the last 1000 frames. (F) The pairwise distance PCA for 1k94_A. (G) SASA analysis shows that our model protein has a high SASA value (130 nm^2) in between 80 ns to 90 ns. (H) This plot shows that the template structure 1K94 has a high SASA value between 85 ns and 100 ns, with other peaks around 60 ns.

Table 6. Results of CATH analysis of 1k94_A.

Level	CATH Code	Description
C	1	Mainly Alpha
A	1.10	Orthogonal Bundle
T	1.10.238	Recoverin; domain 1
H	1.10.238.10	ER- hand

4. Conclusion

We presented the predicted tertiary structure of GCA using homology modeling and molecular dynamics simulation (100 ns), with the crystal structure of des (1–52) GCA with bound calcium of *Homo sapiens* serving as the template structure. The interaction of Ca^{2+} and GCA has revealed more information from the perspective of tertiary structure. GCA comprises 8 alpha-helices with disordered protein at the beginning of the sequence (0–55 amino acids). A 100 ns molecular dynamics simulation of the predicted structure also reveals the structural and dynamics stability of predicted structure GCA in comparison with the template structure 1K94_A. The RMSD, RMSF, and Rg trajectories converge almost entirely within the 100 ns of the MD simulation, indicating the compactness of the expected structure of mouse GCA compared to the template structure 1K94_A. The essential dynamics analysis captures PCA and SASA and culminates in the motions during MD simulation corresponding to the last 1000 frames. The CASTp server predicts GCA active sites and associated amino acid residues. The indicated amino-acid sequences 17–19, 38–41, 43–44, 46–51, 53–56, 114–115, 117, 124–125, 128, 132, 135, 138–139, 142–143, 157, 160–161, 164–165 predicted the potential active sites which is the similar region of the Ca2+ binding sites of template structure 1K94_A. These findings will aid in the development of potential inhibitors and the discovery of binding pockets and residues for drug-like molecules.

Declarations

Author contribution statement

Shamrat Kumar Paul: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Md. Saddam, Khandoker Asiqur Rahaman, Jong-Gu Choi: Contributed reagents, materials, analysis of data; Wrote the paper.

Sang-Suk Lee, Mahbub Hasan: Conceived and designed the experiment; Wrote the paper.

Funding statement

This research was supported partly by the Brain Pool program (2021H1D3A2A01099303), funded by the Ministry of Science and ICT,

and the Basic Science Research Program (2021R1I1A3054773), funded by the Ministry of Education through the National Research Foundation of Korea (NRF).

Data availability statement

Data will be made available on request.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

This project was partly supported by the Ministry of Science and Technology of Bangladesh by providing an NST fellowship project to Shamrat Kumar Paul. We are thankful to Dr. Anthony Nash for his computation support during the MD simulation study provided by Science Outreach Servers. Furthermore, we thank Chowdhury Salman Rahman from Dept. Of Computer Science and Engineering, BSMRSTU for his technical expertise in GPU and CPU computing.

References

[1] M. Buc, Role of regulatory T cells in pathogenesis and biological therapy of multiple sclerosis, Mediat. Inflamm. 2013 (2013).
[2] F. Jadidi-Niaragh, A. Mirshafiey, Th17 Cell, the new player of neuroinflammatory process in multiple sclerosis, Scand. J. Immunol. 74 (1) (2011) 1–13.
[3] I.K. Sand, Classification, diagnosis, and differential diagnosis of multiple sclerosis, Curr. Opin. Neurol. 28 (3) (2015) 193–205.
[4] 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.
[5] R. Milo, A. Miller, Revised diagnostic criteria of multiple sclerosis, Autoimmun. Rev. 13 (4–5) (2014) 518–524.
[6] A. Amedei, D. Prisco, M.M. D'Elios, Multiple sclerosis: the role of cytokines in pathogenesis and in therapies, Int. J. Mol. Sci. 13 (10) (2012) 13438–13460.
[7] G.C. Furtado, M.C.G. Marcondes, J.-A. Latkowski, J. Tsai, A. Wensky, J.J. Lafaille, Swift entry of myelin-specific T lymphocytes into the central nervous system in spontaneous autoimmune encephalomyelitis, J. Immunol. 181 (7) (2008) 4648–4655.
[8] R.A. O'Connor, et al., Cutting edge: Th1 cells facilitate the entry of Th17 cells to the central nervous system during experimental autoimmune encephalomyelitis, J. Immunol. 181 (6) (2008) 3750–3754.
[9] F. Liu, H. Shinomiya, T. Kirikae, H. Hirata, Y. Asano, Characterization of murine grancalcin specifically expressed in leukocytes and its possible role in host defense against bacterial infection, Biosci. Biotechnol. Biochem. 68 (4) (2004) 894–902.
[10] Q. Han, J. Jia, Y. Li, K. Lollike, M. Cygler, Crystallization and preliminary X-ray analysis of human grancalcin, a novel cytosolic Ca2+-binding protein present in leukocytes, Acta Crystallogr. Sect. D Biol. Crystallogr. 56 (6) (2000) 772–774.
[11] M. Maki, Y. Kitaura, H. Satoh, S. Ohkouchi, H. Shibata, Structures, functions and molecular evolution of the penta-EF-hand Ca 2+-binding proteins, Biochim. Biophys. Acta, Proteins Proteomics 1600 (1–2) (2002) 51–60.
[12] J. Roes, B.K. Choi, D. Power, P. Xu, A.W. Segal, Granulocyte function in grancalcin-deficient mice, Mol. Cell Biol. 23 (3) (2003) 826–830.
[13] Y. Kitaura, H. Satoh, H. Takahashi, H. Shibata, M. Maki, Both ALG-2 and peflin, penta-EF-hand (PEF) proteins, are stabilized by dimerization through their fifth EF-hand regions, Arch. Biochem. Biophys. 399 (1) (2002) 12–18.
[14] K. Lollike, A.H. Johnsen, I. Durussel, N. Borregaard, J.A. Cox, Biochemical characterization of the penta-EF-hand protein grancalcin and identification of L-plastin as a binding partner, J. Biol. Chem. 276 (21) (2001) 17762–17769.
[15] A. Boyhan, C.M. Casimir, J.K. French, C.G. Teahan, A.W. Segal, Molecular cloning and characterization of grancalcin, a novel EF-hand calcium-binding protein abundant in neutrophils and monocytes, J. Biol. Chem. 267 (5) (1992) 2928–2933.
[16] J. Jia, Q. Han, N. Borregaard, K. Lollike, M. Cygler, Crystal structure of human grancalcin, a member of the penta-EF-hand protein family, J. Mol. Biol. 300 (5) (2000) 1271–1281.
[17] T.W. Kim, S. Hong, A.H. Talukder, V. Pascual, Y.J. Liu, Grancalcin (GCA) modulates Toll-like receptor 9 (TLR9) mediated signaling through its direct interaction with TLR9, Eur. J. Immunol. 46 (3) (2016) 712–724.
[18] E. Zamir, B. Geiger, Molecular complexity and dynamics of cell-matrix adhesions, J. Cell Sci. 114 (Pt 20) (2001) 3583–3590.
[19] M. Hasan, et al., 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.
[20] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide protein data bank, Nat. Struct. Biol. 10 (12) (2003) 980. United States.
[21] M. Remmert, A. Biegert, A. Hauser, J. Söding, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment, Nat. Methods 9 (2) (2012) 173–175.
[22] A. Waterhouse, et al., SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (W1) (2018) W296–W303.
[23] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (2) (Apr. 1993) 283–291.
[24] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, Protein Sci. 2 (9) (1993) 1511–1519.
[25] W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: computed atlas of surface topography of proteins, Nucleic Acids Res. 46 (W1) (2018) W363–W367.
[26] M.J. Abraham, et al., Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1 (2) (2015) 19–25.
[27] 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.
[28] H.G. Petersen, Accuracy and efficiency of the particle mesh Ewald method, J. Chem. Phys. 103 (9) (1995) 3668–3679.
[29] B.J. Grant, L. Skjaerven, X.-Q. Yao, The Bio3D packages for structural bioinformatics, Protein Sci. 30 (1) (2021) 20–30.
[30] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, MDAnalysis: a toolkit for the analysis of molecular dynamics simulations, J. Comput. Chem. 32 (10) (2011) 2319–2327.
[31] R.T. McGibbon, et al., MDTraj: a modern open library for the analysis of molecular dynamics trajectories, Biophys. J. 109 (8) (2015) 1528–1532.
[32] 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.
[33] A. Bateman, et al., UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Res. 49 (D1) (2021) D480–D489.
[34] E. Gasteiger, et al., Protein identification and analysis tools on the ExPASy server, in: The Proteomics Protocols Handbook, 2009.
[35] D.W.A. Buchan, D.T. Jones, The PSIPRED protein analysis workbench: 20 years on, Nucleic Acids Res. 47 (W1) (2019) W402–W407.
[36] G. Deléage, ALIGNSEC: viewing protein secondary structure predictions within large multiple sequence alignments, Bioinformatics 33 (24) (2017) 3991–3992.
[37] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J.E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858.
[38] A. Waterhouse, et al., SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (W1) (2018) W296–W303.
[39] S. Bienert, et al., The SWISS-MODEL Repository-new features and functionality, Nucleic Acids Res. 45 (D1) (2017) D313–D319.
[40] G. Studer, G. Tauriello, S. Bienert, M. Biasini, N. Johner, T. Schwede, ProMod3 – a versatile homology modelling toolbox, PLoS Comput. Biol. 17 (1) (2021) e1008667.
[41] F. Sievers, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539.
[42] E.F. Pettersen, et al., UCSF Chimera – a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (13) (2004) 1605–1612.
[43] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (2) (1993) 283–291.
[44] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2) (2007).
[45] S.W.I. Siu, K. Pluhackova, R.A. Böckmann, Optimization of the OPLS-AA force field for long hydrocarbons, J. Chem. Theor. Comput. 8 (4) (2012) 1459–1470.
[46] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, J. Phys. Chem. B 105 (28) (2001) 6474–6487.
[47] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (45) (1996) 11225–11236.
[48] A.T. Tzanov, M.A. Cuendet, M.E. Tuckerman, How accurately do current force fields predict experimental peptide conformations? An adiabatic free energy dynamics study, J. Phys. Chem. B 118 (24) (2014) 6539–6552.
[49] M.R. Shirts, J.W. Pitera, W.C. Swope, V.S. Pande, Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins, J. Chem. Phys. 119 (11) (2003) 5740–5761.
[50] M.J. Robertson, J. Tirado-Rives, W.L. Jorgensen, Improved peptide and protein torsional energetics with the OPLS-AA force field, J. Chem. Theor. Comput. (2015).
[51] S. Miyamoto, P.A. Kollman, Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models, J. Comput. Chem. 13 (8) (1992) 952–962.
[52] R. Martoňák, A. Laio, M. Parrinello, Predicting crystal structures: the parrinello-rahman method revisited, Phys. Rev. Lett. 90 (7) (2003) 4.
[53] F. Pirani, S. Brizi, L.F. Roncaratti, P. Casavecchia, D. Cappelletti, F. Vecchiocattivi, Beyond the Lennard-Jones model: a simple and accurate potential function probed by high resolution scattering data useful for molecular dynamics simulations, Phys. Chem. Chem. Phys. 10 (36) (2008) 5489–5503. The Royal Society of Chemistry.

[54] 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.
[55] M.A. Anwar, S. Choi, Structure-activity relationship in tlr4 mutations: atomistic molecular dynamics simulations and residue interaction network analysis, Sci. Rep. 7 (2017).
[56] N. Spellmon, X. Sun, N. Sirinupong, B. Edwards, C. Li, Z. Yang, Molecular dynamics simulation reveals correlated inter-lobe motion in protein lysine methyltransferase SMYD2, PLoS One 10 (12) (2015) 1–10.
[57] S. Haider, G.N. Parkinson, S. Neidle, Molecular dynamics and principal components analysis of human telomeric quadruplex multimers, Biophys. J. 95 (1) (2008) 296–311.
[58] N. Michaud-Agrawal, et al., MDTraj: a modern open library for the analysis of molecular dynamics trajectories, Biophys. J. 109 (8) (2015) 1528–1532.
[59] M. Knudsen, C. Wiuf, Genome databases the CATH database, Hum. Genom. 4 (3) (2010) 207–212.
[60] I. Sillitoe, N. Dawson, J. Thornton, C. Orengo, The history of the CATH structural classification of protein domains, Biochimie 119 (2015) 209–217.
[61] N.L. Dawson, et al., CATH: an expanded resource to predict protein function through structure and sequence, Nucleic Acids Res. 45 (D1) (2017) D289–D295.
[62] A.I. Cuff, et al., The CATH classification revisited-architectures reviewed and new ways to characterize structural divergence in superfamilies, Nucleic Acids Res. 37 (Suppl. 1) (2009) 310–314.
[63] A. Bateman, et al., UniProt: the universal protein knowledgebase, Nucleic Acids Res. 45 (D1) (2017) D158–D169.
[64] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (3) (1990) 403–410.
[65] A. Fiser, Template-based protein structure modeling, Methods Mol. Biol. 673 (2010) 73–94.
[66] P. Bawono, M. Dijkstra, W. Pirovano, A. Feenstra, S. Abeln, J. Heringa, Multiple sequence alignment, in: Methods in Molecular Biology, 2017.
[67] J. Jia, N. Borregaard, K. Lolliked, M. Cygler, Structure of Ca2+-loaded human grancalcin, Acta Crystallogr. Sect. D Biol. Crystallogr. 57 (12) (2001) 1843–1849.
[68] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 182 (2) (1989) 319–326.
[69] K. Guruprasad, B.V.B. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Protein Eng. Des. Sel. 4 (2) (1990) 155–161.
[70] S. Idicula-Thomas, P. V Balaji, Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli, Protein Sci. 14 (3) (2005) 582–592.
[71] A. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (6) (1980) 1895–1898.
[72] K. Sivakumar, S. Balaji, Gangaradhakrishnan, Silico Characterization of Antifreeze Proteins Using Computational Tools and Servers, 2007.
[73] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1) (1982) 105–132.
[74] F. Ardito, M. Giuliani, D. Perrone, G. Troiano, L. Lo Muzio, The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review), Int. J. Mol. Med. 40 (2) (2017) 271–280.
[75] J.F. Zayas, Functionality of Proteins in Food, 1997.
[76] H. Chen, F. Gu, Z. Huang, Improved Chou-Fasman method for protein secondary structure prediction, BMC Bioinf. 7 (S14) (2006) 2105–2107.
[77] T.Z. Sen, R.L. Jernigan, J. Garnier, A. Kloczkowski, GOR V server for protein secondary structure prediction, Bioinformatics 21 (11) (2005) 2787–2788.
[78] P. Xu, J. Roes, A.W. Segal, M. Radulovic, The role of grancalcin in adhesion of neutrophils, Cell. Immunol. 240 (2) (2006) 116–121.
[79] M. Berjanskii, et al., PROSESS: a protein structure evaluation suite and server, Nucleic Acids Res. 38 (Suppl. 2) (2010) W633–W640.