Chapter 10
Graph Neural Network Approaches for Identifying Calpain-10 Inhibitors in Neurological Disorder Therapy

Konda Mani Saravanan, Thirunavukarasou Anand, Dhanushkodi Balamurugan, and Renganathan Senthil

10.1 Introduction

The binding process of protein ligands is essential to biological functions, rendering them a critical focus in drug design and molecular biology [1, 2]. Therefore, understanding the structural components of these interactions is crucial when formulating targeted therapies to improve drug efficacy [3]. Advancements in computational biology and artificial intelligence have enabled us to predict protein structures with greater accuracy and simulate molecular interactions more effectively [4]. Among these advancements, the DeepMind product AlphaFold has proven to be an effective technique for accurately predicting protein shapes [5, 6]. Utilizing this approach, we focused on the prediction and analysis of the 3D structure of the calpain receptor, a protein implicated in various physiological and pathological processes [7, 8]. The calpain receptor is a calcium-dependent cysteine protease that contributes to signal transduction, apoptosis, and cytoskeletal remodeling [9]. Aberrant control of calpain activity is linked to neurological illnesses, cardiovascular disorders, and

K. M. Saravanan
B Aatral Biosciences Private Limited, Bangalore, Karnataka, India

T. Anand
Faculty of Clinical Research, Sri Ramachandra Institute of Higher Education and Research, Chennai, Tamil Nadu, India

D. Balamurugan
Department of BBA, Vels Institute of Science, Technology and Advanced Studies, Chennai, Tamil Nadu, India
e-mail: dbalamurugan.sms@velsuniv.ac.in

R. Senthil (✉)
Department of Bioinformatics, Vels Institute of Science, Technology and Advanced Studies, Chennai, Tamil Nadu, India
e-mail: rsenthil.sls@vistas.ac.in

M. M. Hassan et al. (eds.), *Graph Neural Networks for Neurological Disorders*,
https://doi.org/10.1007/978-3-032-04315-3_10

cancer, rendering it a potential therapeutic target [10]. Nevertheless, information concerning the structural characteristics of calpain and its ligand-binding patterns is limited. This research addresses this deficiency by employing AlphaFold to forecast an accurate structure of the calpain receptor, hence enhancing ligand-binding predictions and virtual screening.

We employed DeepBindPoc and DeepBindGCN to delineate ligand-binding pockets and characterize protein-ligand interactions. DeepBindPoc, a deep learning framework, evaluates geometric and physicochemical features to forecast probable binding sites [11], whereas DeepBindGCN, another deep learning model, predicts binding affinities [12]. These methodologies provide swift, high-throughput screening of extensive chemical libraries, hence enhancing the rate of identifying new treatments. To assess ligand binding and protein stability, we conducted molecular dynamics simulations to evaluate the structural flexibility of the protein. The resulting RMSD and RMSF measurements provided insights into the dynamic characteristics of the ligand-bound complexes [13]. These simulations were essential for evaluating the conformational stability of receptor-ligand interactions and their impact on medication design.

This chapter delineates a comprehensive methodology that includes AI-driven modeling, deep learning pocket prediction, and molecular dynamics simulations. Our technique elucidates the structural characteristics of the calpain receptor and proposes compounds with advantageous binding constants. The findings endorse the application of computational methods to improve drug development procedures, providing a foundation for subsequent experimental validation.

The rest of this chapter is organized as Sect. 10.2, Methods, and outlines the implementation of various approaches, including the artificial intelligence-based structural modelling of the Calpain-10 receptor. It also covers the virtual screening of ligands using DeepBindGCN and AutoDock Vina and explores DeepBindPoc and molecular dynamics simulations. Section 10.3 presents the results and findings related to comparison, clustering, and classifier performance assessment. Section 10.4 evaluates the role and effectiveness of computer-aided drug design and virtual screening methodologies in contemporary drug discovery. Here, Section 10.5 concludes a framework for leveraging computational techniques to accelerate drug discovery and enhance the study of protein-ligand interactions.

10.2 Methods

10.2.1 Artificial Intelligence-Based Structural Model of the Calpain-10 Receptor

The AlphaFold Protein Structure Database was employed to acquire the structural model of the Calpain-10 receptor. AlphaFold, a deep learning-driven tool for predicting protein structures created by DeepMind, delivers high-confidence structural

predictions derived from amino acid sequences [5, 14]. The database contains precomputed structural models for several proteins, enabling rapid access to superior structural data. The amino acid sequence of the Calpain-10 receptor was obtained from a reputable protein sequence database (UniProt ID: Q9HC96), such as UniProt or NCBI GenBank, utilizing the relevant accession number or protein identifier [15]. This sequence was used to query the AlphaFold Protein Structure Database via its web interface (https://alphafold.ebi.ac.uk/) to identify the precomputed model.

When the Calpain-10 receptor structure was not present in the AlphaFold database, the sequence was immediately submitted to the AlphaFold pipeline for structure prediction. The prediction approach encompassed the development of numerous sequence alignments, identification of templates, and inference using a deep neural network-based model [16]. Upon computation, the highest-ranked structural model, together with corresponding confidence scores, was downloaded in PDB format for subsequent study [17]. The acquired structure was visualized and confirmed utilizing molecular visualization software, including PyMOL and ChimeraX [18, 19]. The projected local distance difference test (pLDDT) scores were analyzed to evaluate the confidence level of the structural regions, with elevated values (>70) signifying dependable predictions. Regions with diminished confidence were examined to ascertain potential structural flexibility or disordered segments. This structural model provided the basis for later computer analyses, such as docking simulations, molecular dynamics studies, and functional evaluations.

10.2.2 Virtual Screening of Ligands by DeepBindGCN and AutoDock Vina

A virtual ligand screening of the ZINC database was performed using DeepBindGCN and AutoDock Vina to identify possible binders for the Calpain-10 receptor [20]. Compounds from the ZINC database were evaluated, resulting in the selection of 1000 compounds exhibiting favorable drug-like properties according to Lipinski's Rule of Five, with a molecular weight ranging from 200 to 600 Da, an appropriate number of hydrogen bond donors and acceptors, and a suitable logP value [21]. Ligands were initially converted from their SMILES notation into three-dimensional conformations, followed by the adjustment of protonation states and the generation of tautomers [22]. The Calpain-10 receptor's 3D structure was obtained from the AlphaFold Protein Structure Database, preprocessed by removing water molecules and including polar hydrogens; Gasteiger charges were applied, and the file was saved in PDBQT format. The calpastatin (an endogenous inhibitor) is used as control for docking and molecular dynamics simulations.

10.2.3 DeepBindPoc and DeepBindGCN

DeepBindPoc is a proof of concept for a predictive deep learning network that evaluates protein-ligand interactions with exceptional accuracy and rapidity [11]. Utilizing advanced Graph Convolutional Networks (GCNs), it incorporates physicochemical properties and spatial configurations of protein-ligand complexes for precise predictions. DeepBindPoc excels in classifying binders and non-binders and assessing binding propensity, both of which are essential for drug discovery. These predictions are further corroborated by the incorporation of supplementary scoring systems, thereby yielding precise and dependable assessments of protein-ligand binding affinity [23]. Ligand-binding site prediction was conducted using DeepBindGCN, which analyzed the receptor's surface topology to identify probable ligand-binding sites [12]. The ligand screening approach utilizing DeepBindGCN predicted probable binding sites on the calpain receptor and assessed the interactions of the ligands inside these sites. DeepBindGCN, a method based on graph convolutional neural networks, utilized geometric and chemical information from the receptor's three-dimensional structure and surface to anticipate ligand-binding pockets. The receptor structure, obtained from the AlphaFold Protein Structure Database, was initially created by eliminating water molecules, incorporating hydrogens, and subsequently assigning partial charges. The anticipated binding sites were checked with established functional residues or conserved areas to ensure correctness. Compounds from the ZINC database were generated using conformation generation from 2D SMILES codes, optimization of protonation states, and tautomer generation. The ligands were subsequently subjected to virtual screening via docking into the anticipated pockets, and their binding affinities were assessed based on the docking scores. The predictions obtained by DeepBindGCN enhanced the selectivity of target binding sites, facilitating the prioritization of ligands for further screening in detailed research.

10.2.4 AutoDock Vina

The ligand screening process utilizing AutoDock Vina was succeeded by molecular docking analysis, a method for determining the interaction energies between the calpain receptor and prospective ligands [24]. The receptor structure was obtained from the AlphaFold Protein Structure Database and further modified by eliminating water molecules, including polar hydrogens, and assigning Gasteiger charges via AutoDock Tools [25]. Ligands from the ZINC database were generated by interpreting the SMILES codes, constructing three-dimensional conformations, optimizing protonation states, and producing tautomers. The receptor and ligand files were prepared for docking with AutoDock Vina software, requiring the files to be in PDBQT format. The binding site coordinates were predicted using DeepBindPoc, and the docking grid boxes were then optimized to encompass the pockets.

Molecular docking was conducted with an exhaustiveness of 8 to sample the conformational space of ligands efficiently. The binding affinities were ranked based on the lowest binding energy values, and the interactions between the ligand and receptor, including hydrogen bonding and hydrophobic contacts, were visualized using Discovery Studio Visualizer [26].

10.2.5 Molecular Dynamics Simulations

Molecular dynamics simulations were conducted using GROMACS software, version 2022, to examine the stability and dynamics of the top three Calpain-10 receptor-ligand complexes previously found using molecular docking [27]. The receptor-ligand combinations exhibiting optimal binding energies from AutoDock Vina were utilized for 50 ns simulations. The CHARMM36 force field was employed for the receptor, while the ligand topologies and parameters were obtained via the CGenFF website (https://cgenff.com/). The initial intricate structures were introduced into the simulation cube, which contains TIP3P water molecules, maintaining a minimum distance of 1.0 nm from the cube's perimeter [28]. The incorporation of appropriate counter ions equilibrated the systems, and the ionic strength was adjusted to 0.15 M NaCl to simulate physiological conditions. Geometry optimization was performed using the steepest descent method until the maximum force fell below 1000 kJ/mol/nm. Thereafter, each system underwent equilibration via two phases: Subsequent to the initial minimization, a 100 ps NVT (constant volume and temperature) equilibration was conducted, followed by a 100 ps NPT (constant pressure and temperature) equilibration, both at 300 K utilizing the V-rescale thermostat and at 1 bar pressure employing the Parrinello-Rahman barostat [29]. Hydrogen bond restrictions were implemented using the LINCS technique. At the same time, long-range electrostatics were addressed with the PME approach, with short-range van der Waals and Coulomb interactions truncated at 1.0 nm. The production molecular dynamics simulations were conducted for 50 nanoseconds and all visualizations and analyses were conducted using GROMACS tools, PyMOL, and VMD [30].

10.3 Results

10.3.1 AI-predicted Structural Model

The structural model is AF-Q9HC96-F1-v4, and we possess a high degree of confidence in the precision of the anticipated three-dimensional structure of the calpain receptor, as generated by DeepMind's AlphaFold algorithm. This model is derived from the amino acid sequence associated with UniProt ID Q9HC96, which was

docked to provide atomic-level structural information. AlphaFold uses deep learning and numerous sequence alignments to determine protein folding patterns accurately. The model includes confidence scores such as the projected Local Distance Difference Test (pLDDT), which assesses the stability of each residue's positioning, with scores above 70 being reliable. Secondary structural elements, binding pockets, and flexible loops are clearly delineated, facilitating functional and interaction analyses. The model is provided in Protein Data Bank (PDB) format, which is suitable for integration into molecular docking and dynamics research. AF-Q9HC96-F1-v4 aids in comprehending receptor structural alterations, ligand interactions, and potential drug binding sites for application in virtual screening for drug discovery. Figure 10.1 presents the structural and confidence metrics of the calpain receptor model as forecasted by AlphaFold. Panel A illustrates the 3D structural model, with color coding based on confidence levels: blue indicates high confidence areas, whereas red and yellow denote low or flexible regions, potentially corresponding to loops or disordered segments. The PAE matrix is displayed in the subsequent panel B, illustrating the positional uncertainty between two residues. The area of minimal positional error is indicated by a dark green hue, signifying a higher confidence

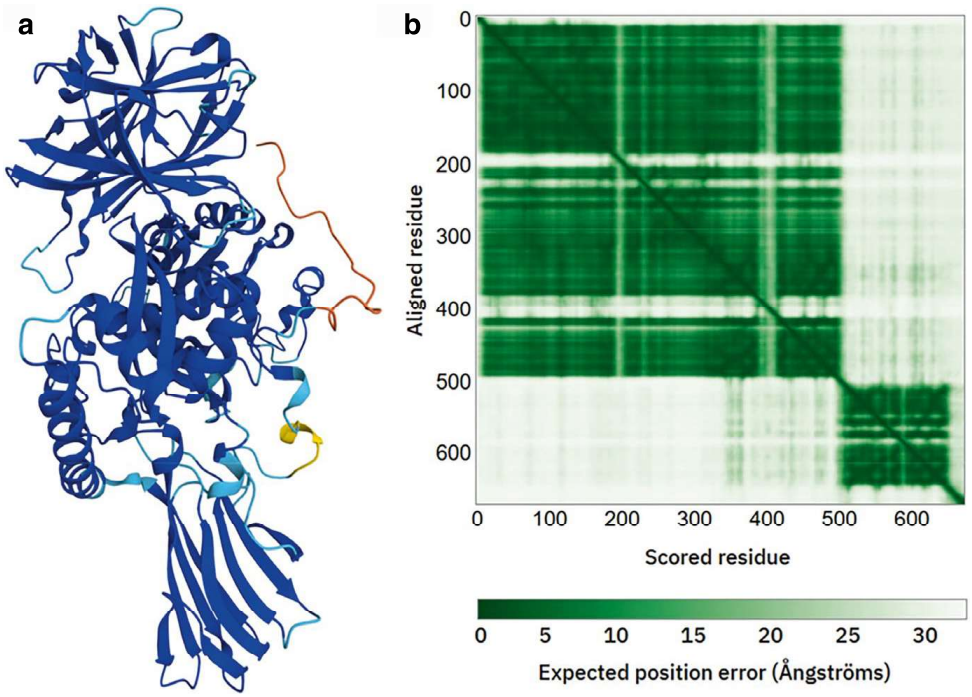

Fig. 10.1 Structural and confidence analysis of the calpain receptor model predicted by AlphaFold. Panel A shows the 3D model with color coding representing confidence levels: blue for high confidence and red/yellow for low confidence, indicating flexible or disordered regions. Panel B displays the PAE matrix, illustrating the positional uncertainty between residues, with dark green indicating minimal error and lighter green indicating more significant uncertainty. The bar at the bottom represents the predicted positional error in angstroms, which is essential for evaluating structural quality for molecular docking and simulations

level in the predictions. In contrast, the lighter green hue denotes a positional significant mistake and, thus, greater uncertainty. The bar at the bottom signifies the anticipated positional error in angstroms, providing a metric for the structural quality required for molecular docking and simulation.

10.3.2 Deep Learning-Based Ligand Binding Pocket Prediction

The pocket prediction for the AlphaFold protein structure AF-Q9HC96-F1-v4 was conducted using DeepBindPoc, a deep learning method designed for predicting ligand-binding sites [11]. The calpain receptor structure was obtained from the AlphaFold Protein Structure Database, subsequently purified of water molecules, hydrogen atoms were included, and atomic charges were adjusted. The processed structure was subsequently uploaded to the DeepBindPoc server, which employs a graph convolutional neural network to assess the geometric and physicochemical characteristics of the protein surface. In DeepBindPoc, binding pockets were predicted through the analysis of residue spatial distribution, surface morphology, and charge distribution. Each projected pocket was allocated a confidence value that assesses the likelihood of ligand binding based on the obtained data. The highest-ranked binding sites were isolated and superimposed on conserved residues and functional domains of the protein to validate their biological relevance. The anticipated coordinates of the pockets were utilized to generate docking grids for molecular docking analysis, facilitating accurate virtual screening and ligand-binding assessment. We have used Arg202 as a central binding pocket residue for docking calculations. These predictions were essential in elucidating receptor-ligand interactions, facilitating structure-based drug design, and further computational investigations.

10.3.3 Virtual Screening Using DeepBindGCN and AutoDock Vina

DeepBindGCN is a deep learning system based on Graph Convolutional Networks (GCNs) designed for high-throughput analysis of extensive datasets [12, 31, 32]. Graph Convolutional Networks (GCNs) is a recognized methodology in deep learning, wherein nodes transmit residue-related information, and edges denote the spatial relationships among the nodes. Previous research has examined the application of Graph Convolutional Networks (GCNs) for forecasting chemical characteristics and molecular fingerprints [23, 33, 34]. Furthermore, GCNs have demonstrated considerable efficacy in predicting protein–ligand interactions in terms of both time and accuracy. DeepBindGCN comprises two distinct models: Two models were

created: DeepBindGCN_BC, a binary classifier designed to differentiate between binders and non-binders, and DeepBindGCN_RG, which forecasts the binding affinities of protein-ligand complexes. The superior performance of DeepBindGCN_BC can be attributed to the model's consideration of both the physicochemical characteristics of proteins and the spatial attributes of the ligands. Three scoring techniques were employed to assess the significance of protein–ligand complexes: The three models are DeepBindGCN_BC, DeepBindGCN_RG, and the scoring algorithm used by AutoDock Vina.

Table 10.1 presents data for 20 compounds, including the findings of DeepBindGCN_BC, DeepBindGCN_RG, and binding energy measured by AutoDock Vina in kcal/mol. All chemicals were categorized as binders using DeepBindGCN_BC, resulting in a binary classification outcome of 1. This indicates that these may serve as suitable candidates for protein-ligand interactions, provided the protein side interactions are advantageous. The DeepBindGCN_RG scores, indicating the expected binding affinity, exhibit minor variations across the compounds, all of which are elevated. Binding energy values, expressed in kcal/mol, are all negative, signifying that the compounds exhibit favorable interactions with the target proteins. Among the evaluated compounds, Samatasvir exhibits the highest DeepBindGCN_RG score of 9.52 and the most favorable binding energy of −10.60 kcal/mol. This coupling highlights its superior contact capabilities. Similarly, Orvepitant attains a DeepBindGCN_RG score of 9.44 and a marginally superior binding energy of −10.90 kcal/mol, validating an effective equilibrium between computational prediction and thermal stability. Significantly, Mk3207 exhibits a binding affinity of 9.22 and a binding energy of −10.70 kcal/mol, positioning it as a possible therapeutic candidate. Compounds such as Ditercalinium (−9.16 Kcal/mol, −10.80 Kcal/mol) and Vindesine (−9.15 Kcal/mol, −10.60 Kcal/mol) demonstrate commendable efficacy, suggesting robust interactions. The recurrence of Abamectin-component-b1a in two entries with distinct binding energies (−10.60 and − 10.50 kcal/mol) and nearly identical DeepBindGCN_RG scores (9.25 and 9.12) suggests it is a likely binder. Nonetheless, compounds with marginally lower DeepBindGCN_RG scores, such as Mergocriptine (9.00) and Metergotamine (9.01), exhibit commendable binding energy values (−10.50 kcal/mol and − 11.10 kcal/mol, respectively), indicating their merit for further evaluation. Dihydroergotamine exhibits the lowest enthalpy of production at −11.50 kcal/mol, yet its moderate binding affinity of 9.09 means it remains a viable contender for further investigation.

Figure 10.2 illustrates protein-ligand interactions in both 3D and 2D perspectives, emphasizing the binding mechanisms and interactions of specific molecules. Figures. **a**, b, and c represent the blue ligands within the active site of their target proteins, depicted as green ribbon structures, to demonstrate the requisite spatial compatibility and orientation for interaction. Furthermore, panels D, E, and F illustrate 2D interaction diagrams of specific interactions, including van der Waals (green), hydrogen bonding (blue), π-π stacking (pink), and alkyl (purple) interactions that enhance the stability of the ligand within the binding pocket. ARG, PHE, and GLN are intricately associated with ligand binding via hydrogen bonding and

Table 10.1 The results of the virtual screening of 20 compounds were analyzed using DeepBindGCN and AutoDock Vina. The table displays DeepBindGCN_BC binary classification (1 = binder), DeepBindGCN_RG predicted binding affinity scores, and AutoDock Vina binding energies (kcal/mol)

S.No	Compound	DeepBindGCN_BC	DeepBindGCN_RG	Binding energy (Kcal/mol)
1	Dihydroergotamine_ZINC000003978005	1	9.09	−11.50
2	Bolazine_ZINC000008214506	1	9.10	−11.20
3	Metergotamine_ZINC000072266819	1	9.01	−11.10
4	Ergotamine_ZINC000052955754	1	9.06	−11.00
5	Orvepitant_ZINC000056898864	1	9.44	−10.90
6	Dihydroergocristine_ZINC000003947494	1	9.04	−10.90
7	Ledipasvir_ZINC000150338819	1	9.02	−10.90
8	Venetoclax_ZINC000150338755	1	9.04	−10.80
9	Velpatasvir_ZINC000504665933	1	9.10	−10.80
10	Ditercalinium_ZINC000004215707	1	9.16	−10.80
11	Mk3207_ZINC000103760984	1	9.22	−10.70
12	alpha_Ergocryptine_ZINC000222341315	1	9.13	−10.70
13	Zosuquidar_ZINC000100029945	1	9.14	−10.70
14	Samatasvir_ZINC000150588806	1	9.52	−10.60
15	Vindesine_ZINC000008214470	1	9.15	−10.60
16	Abamectin-component-b1a_ZINC000252694903	1	9.25	−10.60
17	Mergocriptine_ZINC000072266905	1	9.00	−10.50
18	Uk432097_ZINC000095539256	1	9.05	−10.50
19	Abamectin-component-b1a_ZINC000252673975	1	9.12	−10.50
20	Etoposide_ZINC000003830818	1	9.13	−10.50

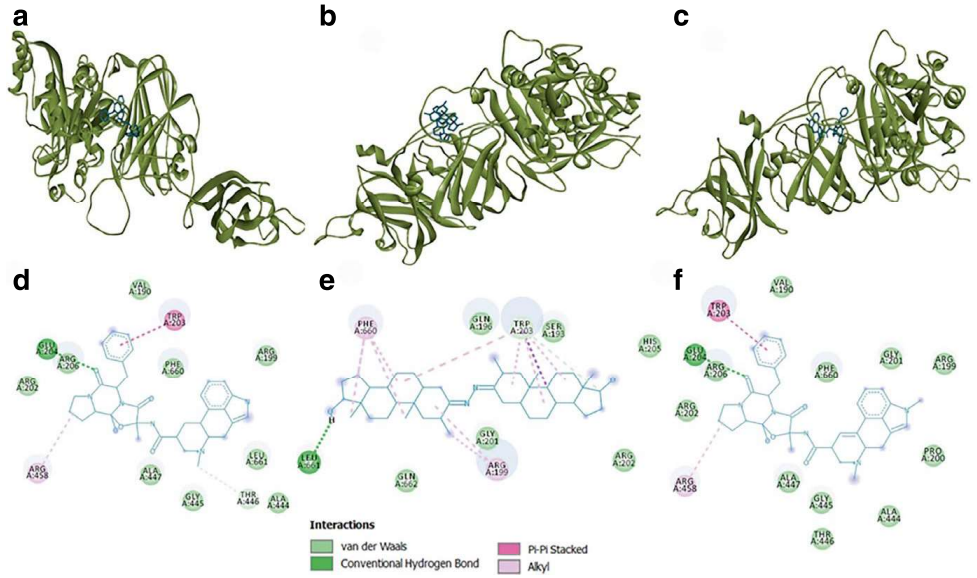

Fig. 10.2 Structural representations and interaction analyses of ligand binding within the protein. (**a–c**) 3D structures of the protein-ligand complexes for Ligand1, Ligand2, and Ligand3, respectively, showing ligand positioning within the binding pocket. (**d–f**) 2D interaction diagrams for Ligand1, Ligand2, and Ligand3 highlight key molecular interactions such as van der Waals forces, hydrogen bonds, π-π stacking, and alkyl interactions. Color coding illustrates the interaction types, emphasizing the binding stability and molecular contacts in each complex

hydrophobic interactions, hence augmenting the stability and binding affinity of the molecule. Collectively, these models provide a lucid depiction of the binding interactions, demonstrating that the ligand's orientation, position, and interactions with particular protein residues are essential to the binding affinity. This comprehensive study is highly beneficial for medicinal chemistry in the design of novel ligands with enhanced potency and selectivity.

10.3.4 Molecular Dynamics Simulations

Figure 10.3 presents an RMSD plot comparing the control (calpastatin) with three ligand-bound systems (Ligand1, Ligand2, and Ligand3) over time measured in picoseconds. RMSD is a crucial metric for assessing the structural variations and overall stability of biomolecular simulations concerning molecular complexes. For the control indicated by the black line, RMSD exhibits a gradual increase in the initial phases, then oscillating within the range of 0.35–0.45 Å after 10,000 picoseconds. This indicates that the control system undergoes minor structural modifications before returning to its equilibrium state while maintaining structural stability in the simulation [35]. Like the control, Ligand1 (Dihydroergotamine), depicted in orange, exhibits a rapid early increase followed by a decline to a plateau at

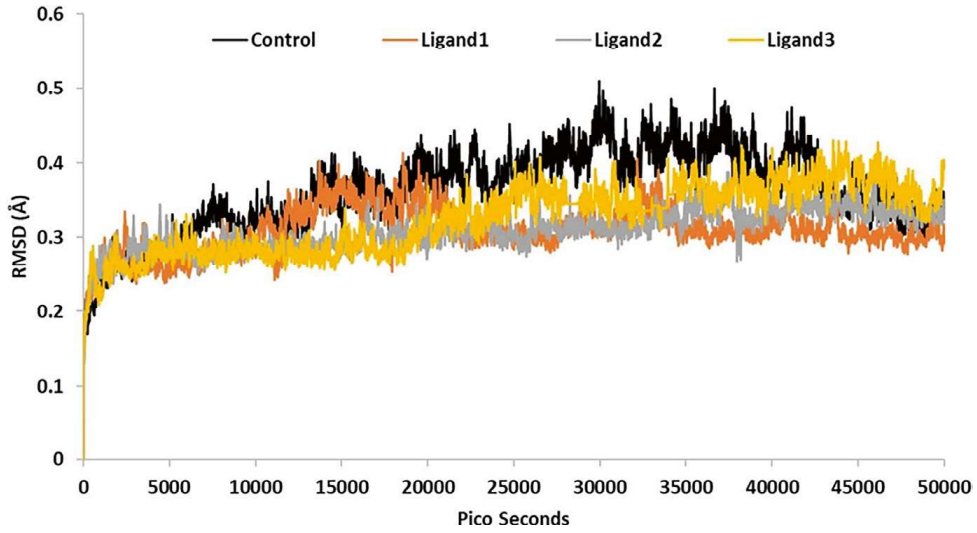

Fig. 10.3 RMSD profiles of the control and ligand-bound systems (Ligand1, Ligand2, and Ligand3) over 50,000 picoseconds. The graph shows structural deviations, indicating that Ligand2 exhibits the highest stability with the lowest RMSD, followed by Ligand1 and Ligand3, while the control displays more significant fluctuations

approximately 0.3–0.4 Å. This indicates that the binding of Ligand1 may restrict conformational alterations and enhance the stability of the complex. Ligand2 (Bolazine), depicted in gray, exhibits the lowest RMSD values (~0.25–0.35 Å) among all groups, signifying little structural variations over time. Consequently, the reduced RMSD suggests that Ligand2 may enhance the structural stability of the complex due to robust binding interactions with the site.

The yellow line illustrates that the RMSD for Ligand3 (Metergotamine) begins at approximately 0.3 to 0.45 Å, which is marginally more significant than that of Ligand1 and Ligand2, however comparable to the control level. This signifies moderate stabilizing effects, albeit with significantly greater flexibility than in Ligand2. The comparative RMSD profiles demonstrate distinct structural dynamics for each system, indicating that Ligand2 confers the most excellent stability to the protein, whilst Ligand1 and Ligand3 provide the subsequent highest stability. These discoveries may assist in guiding further research aimed at comprehending ligand binding mechanisms and their importance in molecular interactions.

Figure 10.4 illustrates the Root Mean Square Fluctuation (RMSF) for the control and three ligand-bound systems in relation to residue indices. RMSF provides insights into the flexibility of the backbone and the localized motion of individual residues; regions exhibiting more flexibility may correlate with loops, termini, or unstructured segments. The control group (black line) illustrates the variability value distributed over residues, with increased flexibility indicated by the peaks. The observed peaks likely result from loop regions or solvent-exposed residues, which often exhibit greater flexibility in molecular dynamics simulations.

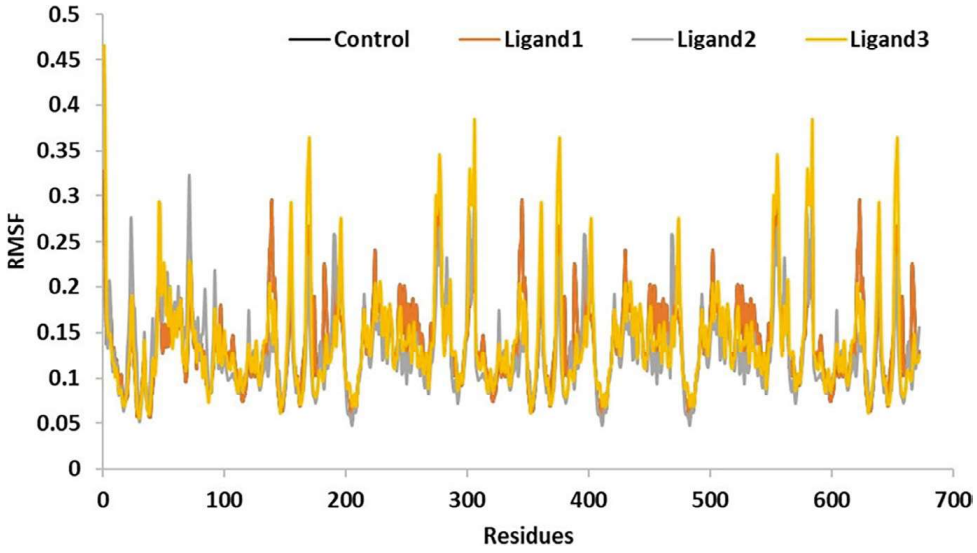

Fig. 10.4 RMSF profiles of the control and ligand-bound systems (Ligand1, Ligand2, and Ligand3) across residues. The data highlight regions of higher flexibility, with Ligand2 exhibiting slightly lower fluctuations, suggesting enhanced structural stability compared to the other systems

Ligand1 (orange) and Ligand3 (yellow) exhibit comparable oscillation patterns, with peak values and placements around those of the control. This observation indicates that these ligands do not influence the intrinsic plasticity of the protein structure and maintain dynamic properties akin to those of the unbound protein. Throughout the experiment, Ligand2 (gray) exhibited somewhat reduced oscillations in various places relative to the other systems. The reduction in flexibility may indicate increased rigidity, likely in specific residues participating in interactions with Ligand2, resulting in a more stable complex, as previously demonstrated.

The RMSF profiles demonstrate that all systems have a similar degree of dynamic flexibility; however, Ligand2 renders the protein structure somewhat more stable. These data suggest that Ligand2 may establish more substantial and more stable connections, perhaps enhancing structural stability while maintaining flexibility in the functional domains. Further investigation of the binding interactions may elucidate the molecular mechanisms behind these differences.

10.4 Discussion

The molecular modeling of the calpain receptor utilizing AlphaFold (AF-Q9HC96-F1-v4) has provided a robust basis for elucidating the topographical structure and functional dynamics of the receptor [5]. The model is appropriate for molecular docking and simulation investigations because of its high-confidence predictions of secondary structural features, binding pockets, and flexible loops, together with

pLDDT scores. AlphaFold's structural data enables this model to deliver precise evaluations of receptor-ligand interactions, hence enhancing drug discovery efforts. Pocket predictions utilizing DeepBindPoc identified several critical residues, including Arg202, as the primary residues interacting with the ligand [11]. Utilizing GCNs enabled the identification of ligand-binding sites based on structural and chemical characteristics. The projected pockets coincided with conserved functional domains, so affirming their biological relevance. The data facilitated the creation of docking grids, and virtual screening of the receptor demonstrated its suitability for ligand binding, thereby confirming its utility in structure-based drug discovery [36].

Consequently, we utilized DeepBindGCN and AutoDock Vina for virtual screening to augment our understanding of ligand-receptor interactions. DeepBindGCN_BC and DeepBindGCN_RG produced excellent estimates of binding affinity, corroborated by negative binding energy values from AutoDock Vina. Samatasvir, Orvepitant, and Mk3207 have been discovered to possess the highest binding affinities and the most advantageous binding energy values, making them promising therapeutic candidates. Notably, Samatasvir exhibited a DeepBindGCN_RG of 9.52 and a binding energy of −10.60 kcal/mol, whereas Orvepitant demonstrated a marginally superior binding energy of −10.90 kcal/mol, suggesting robust binding affinities. The findings indicate that Abamectin-component-b1a serves as a binder across various entry and enhances its application in medication development. The molecular dynamics simulations provided further insight into the structural stability and dynamics of the ligand-receptor complexes. The analysis of RMSD values, considering protein structure and flexibility, indicated that Ligand2 exhibited the most advantageous RMSD values (~0.25–0.35 Å), signifying the compound's superior structural stability. Nevertheless, Ligand1 and Ligand3 exhibited marginally elevated RMSD values of approximately 0.3–0.4 Å, signifying modest stability compared to the control. It is suggested that Ligand2 induces minimal conformational alteration due to its robust binding and superior compatibility within the binding region.

A comprehensive evaluation of residue-level dynamics by RMSF analysis corroborated the conclusion about the differential impact of ligand binding on protein flexibility. Ligand2 exhibited a reduction in fluctuations across multiple areas of the protein and enhanced stiffness in the interacting residues. Ligand1 and Ligand3 exhibited flexibility profiles similar to the control, suggesting they do not impair intrinsic flexibility while providing minor stability. Collectively, these findings indicate the potential for enhancing structural stability through the use of Ligand2 while preserving functional versatility—an amalgamation that positions Ligand2 for further advancement. A synthesis of deep learning models and molecular dynamics simulations has been employed to assess the viability of therapeutic candidates and characterize their binding properties. The discovered compounds exhibited favorable docking scores and dynamic stability with the target protein, indicating their potential for therapeutic application. Future research will investigate these interactions in further depth through experimentation, improvement of lead compound structures, and the identification of additional ligands to enhance the chemical

library. This study emphasizes the efficacy of computer-aided drug design and virtual screening techniques in contemporary drug development processes.

10.5 Conclusion

This study employed computational techniques, such as AlphaFold, DeepBindPoc, DeepBindGCN, and molecular dynamics simulations, to examine the structural characteristics and ligand-binding potential of the calpain receptor. The AI-predicted structural model demonstrated high confidence and accuracy for subsequent analysis. Pocket docking and virtual screening facilitated the identification of several potent ligand compounds exhibiting favorable binding affinities and stability. MD simulations validated that the configurations of ligand-receptor complexes were stable, with Ligand2 exhibiting the highest stability. Based on our discoveries, we firmly endorse the application of AI and deep learning algorithms in structural biology and drug discovery. These computational methods enable high-throughput screening and precise modeling, significantly reducing the time and cost associated with experimentation. Subsequent research will focus on the experimental validation of the specified ligands and the optimization of lead compounds for therapeutic agent development. This paper establishes a framework for employing computational methods to accelerate drug discovery and promote the investigation of protein-ligand interactions.

10.6 Data Availability

The data generated from the current study is presented in the paper.

References

1. Kairys V, et al. Recent advances in computational and experimental protein-ligand affinity determination techniques. Expert Opin Drug Discov. 2024;19(6):649–70.
2. Wang H. Prediction of protein–ligand binding affinity via deep learning models. Brief Bioinform. 2024;25(2):bbae081.
3. Wang DD, Chan M-T, Yan H. Structure-based protein–ligand interaction fingerprints for binding affinity prediction. Comput Struct Biotechnol J. 2021;19:6291–300.
4. Senthil R, et al. Bibliometric analysis of artificial intelligence in healthcare research: trends and future directions. Future Healthcare J. 2024;11(3):100182.
5. Borkakoti N, Thornton JM. AlphaFold2 protein structure prediction: implications for drug discovery. Curr Opin Struct Biol. 2023;78:102526.
6. Saravanan KM. The transformative potential of deep learning and alphafold in addressing brain disorders, vol. 3; 2024. p. 72–3.

7. Sazontova TG, Matskevich AA, Arkhipenko YV. Calpains: physiological and pathophysiological significance. Pathophysiology. 1999;6(2):91–102.
8. Croall DE, Ersfeld K. The calpains: modular designs and functional diversity. Genome Biol. 2007;8:1–11.
9. Metwally E, et al. Calpain signaling: from biology to therapeutic opportunities in neurodegenerative disorders. Front Vet Sci. 2023;10:1235163.
10. Zhang X, et al. Calpain: the regulatory point of cardiovascular and cerebrovascular diseases. Biomed Pharmacother. 2024;179:117272.
11. Zhang H, et al. DeepBindPoc: a deep learning method to rank ligand binding pockets using molecular vector representation. PeerJ. 2020;8:e8864.
12. Zhang H, Saravanan KM, Zhang JZH. DeepBindGCN: integrating molecular vector representation with graph convolutional neural networks for protein–ligand interaction prediction. Molecules. 2023;28(12):4691.
13. Feng Y, et al. Hybrid drug-screening strategy identifies potential SARS-CoV-2 cell-entry inhibitors targeting human transmembrane serine protease. Struct Chem. 2022;33(5):1503–15.
14. Skolnick J, et al. 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. J Chem Info Model. 2021;61:4827–31.
15. UniProt Consortium. The universal protein resource (UniProt). Nucleic Acids Res, Suppl_1. 2007;36:D190–5.
16. Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9.
17. Berman HM, et al. The protein data bank. Nucleic Acids Res. 2000;28(1):235–42.
18. Lill MA, Danielson ML. Computer-aided drug design platform using PyMOL. J Comput Aided Mol Des. 2011;25:13–9.
19. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605–12.
20. Irwin JJ, et al. ZINC: a free tool to discover chemistry for biology. J Chem Inf Model. 2012;52(7):1757–68.
21. Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol. 2004;1(4):337–41.
22. Bento AP, et al. An open source chemical structure curation pipeline using RDKit. J Cheminformatics. 2020;12:1–16.
23. Zhang H, Saravanan KM. Advances in deep learning assisted drug discovery methods: a self-review. Curr Bioinforma. 2024;19(10):891–907.
24. Oleg OT, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455–61.
25. Goodsell DS, et al. The AutoDock suite at 30. Protein Sci. 2021;30(1):31–43.
26. Baroroh U, et al. Molecular interaction analysis and visualization of protein-ligand docking using biovia discovery studio visualizer. Indonesian J Comput Biol (IJCB). 2023;2(1):22–30.
27. Pronk S, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013;29(7):845–54.
28. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. Chem A Eur J. 2001;105(43):9954–60.
29. Jorgensen WL, et al. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926–35.
30. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33–8.
31. Zhang H, et al. DeepBindRG: a deep learning based method for estimating effective protein–ligand affinity. PeerJ. 2019;7:e7362.
32. Zhang H, et al. DeepBindBC: a practical deep learning method for identifying native-like protein-ligand complexes in virtual screening. Methods. 2022;205:247–62.

33. Zhang H, et al. Artificial intelligence and computer-aided drug discovery: methods development and application. Methods (San Diego, Calif.). 2024:S1046–2023.
34. Sreeraman S, et al. Drug design and disease diagnosis: the potential of deep learning models in biology. Curr Bioinforma. 2023;18(3):208–20.
35. Zhang H, et al. An integrated deep learning and molecular dynamics simulation-based screening pipeline identifies inhibitors of a new cancer drug target TIPE2. Front Pharmacol. 2021;12:772296.
36. Zhang H, et al. Deep learning based drug screening for novel coronavirus 2019-nCov. Interdiscip Sci: Comput Life Sci. 2020;12:368–76.