The Alzheimer's Code - Predictive Modelling of Protein Structures in Alzheimer's Disease

Dr. Yuvaraj Ravikumar
Dept. of Biotechnology
Acharya Institute of Technology
Karnataka, India
yuvaraj3020@acharya.ac.in

Mr. M Gowtham Raj
Dept. of CSE
Acharya Institute of Technology
Karnataka, India
gowtham2705@acharya.ac.in

Aditi Das
Dept. of CSE
Acharya Institute of Technology
Karnataka, India
aditir.21.becs@acharya.ac.in

Anjali K S
Dept. of CSE
Acharya Institute of Technology
Karnataka, India
anjalis.21.becs@acharya.ac.in

Bhavana B Hallikeri
Dept. of CSE
Acharya Institute of Technology
Karnataka, India
bhavanab.21.becs@acharya.ac.in

Deepa Tendulkar
Dept. of CSE
Acharya Institute of Technology
Karnataka, India
deepav.21.becs@acharya.ac.in

***Abstract*—Alzheimer's disease, among other neurodegenerative conditions, is closely associated with the misfolding of specific proteins, such as amyloid-beta and tau. Proposed system aims to use amino acid (AA) sequence data and machine learning (ML) techniques to predict disease related protein structures, offering insights into potential therapeutic strategies. By collecting disease-related amino acid sequences from publicly available datasets, a machine learning model is developed to detect disease presence based on these sequences, identifying key features linked to misfolding or aggregation. An interface is built to allow users to input amino acid sequences, providing a 3D structural output using predictive models. By analyzing these predicted structures, the system identifies disease-related characteristics within the proteins. The final stage involves comparing the affected protein structure with the original, non-disease form to pinpoint structural differences that could serve as targets for treatment development. This approach has the potential to accelerate the identification of disease mechanisms and aid in the discovery of therapeutic interventions, including vaccines and drugs.**

Keywords—Alzheimer's, Amino Acid, Docking

I. Introduction

Alzheimer's disease, a leading cause of cognitive decline in older adults, is characterized by the accumulation of misfolded proteins, particularly amyloid-beta (Aβ) and tau. These proteins form neurotoxic aggregates that disrupt neuronal function and eventually lead to cell death, hallmark features of Alzheimer's pathology. Despite significant progress in understanding the disease, effective therapeutic interventions remain elusive, highlighting the need for innovative approaches to address the molecular mechanisms underlying protein misfolding and aggregation.

Research in this domain often relies on computational and experimental techniques to analyze the structural properties of Aβ and tau proteins and their interactions with various ligands. Molecular docking has emerged as a powerful tool in this context, enabling the simulation of ligand-protein interactions to predict binding affinities and identify active sites. However, existing systems predominantly focus on FDA-approved ligands, limiting the exploration of novel compounds that may offer new therapeutic insights. Furthermore, these systems often lack comprehensive analysis of active sites and structural features critical to designing targeted interventions.

The proposed study aims to address these limitations by systematically analyzing natural, non-FDA-approved ligands and their interactions with amyloid-beta proteins. By leveraging publicly available databases such as UniProt, Protein Data Bank (PDB), and PubChem, the study focuses on data collection, protein structure analysis, and ligand docking simulations. Bioinformatics tools and visualization software, such as PyMOL and AutoDock, are employed to generate 3D structural representations, identify active sites, and simulate ligand-protein interactions. This approach provides a deeper understanding of the molecular interactions and identifies potential therapeutic targets.

Ultimately, this research seeks to advance the development of innovative strategies for Alzheimer's disease by providing a comprehensive framework for exploring new therapeutic compounds. The integration of advanced computational methods with experimental data underscores the potential of molecular docking in paving the way for effective treatments, improving the quality of life for individuals affected by this debilitating disease.

II. Background

Alzheimer's disease is primarily characterized by the accumulation of misfolded proteins, particularly amyloid-beta (Aβ) and tau, leading to neurotoxic aggregates that disrupt cellular functions and cause neuronal death. Understanding the interactions between these proteins and therapeutic ligands is a key area of research, with molecular docking and simulation techniques playing a pivotal role. Several studies have explored these interactions, offering valuable insights but also highlighting critical limitations in existing methodologies.

Jokar et al. utilized molecular docking algorithms such as AutoDock Vina to predict ligand-protein interactions efficiently, combined with visualization tools like UCSF Chimera and PyMOL to interpret structural data. Their work focuses primarily on FDA-approved ligands for rapid identification of potential drug candidates. However, the limitation of excluding non-FDA-approved ligands restricts the exploration of novel therapeutic options. The proposed

system aims to address this gap by incorporating a broader range of compounds for docking studies [1].

The work of Kreutzer et al. employed X-ray crystallography to analyze the high-resolution structures of Aβ oligomers, providing detailed insights into peptide assembly. While this static technique captures molecular structures at a specific point in time, it overlooks dynamic changes and transient interactions during aggregation. By integrating molecular docking and dynamic simulation techniques, the proposed system seeks to address this limitation, offering a more comprehensive understanding of ligand-protein interactions over time [2].

Xiao et al. combined experimental and computational approaches to enhance structural understanding of Aβ aggregation and identify potential therapeutic strategies. However, they noted challenges in analyzing peptide aggregation due to uncontrolled oligomerization, which complicates the identification of therapeutic targets. To overcome this, the current research employs hierarchical assembly modeling to simplify the complexity of Aβ aggregation pathways and identify key intervention points [3].

Gulia and Kumar integrated molecular docking with ADMET (absorption, distribution, metabolism, excretion, and toxicity) analysis to streamline the drug discovery process, focusing on both drug and phytocompound interactions. Their study emphasized improved safety profiling using tools like ProTox-II, reducing risks associated with adverse effects. However, their approach highlights the dependency of docking accuracy on the quality of input data. The proposed system prioritizes the use of high-resolution structural data and validated ligand conformations to enhance prediction reliability [4].

Roda and Serra-Mir explored the crosstalk between Aβ and tau proteins, analyzing FDA-approved compounds like Gantenerumab and Aducanumab. While their study provided a comprehensive view of protein interactions, they acknowledged that static docking methods might overlook transient interactions essential to understanding ligand binding. To address this, the proposed system integrates molecular dynamics simulations, offering a dynamic perspective on protein-ligand interactions and ensuring more accurate identification of active sites [5].

Collectively, these studies highlight the significance of combining experimental and computational methods to study protein misfolding and aggregation in Alzheimer's disease. The proposed research builds on this foundation, integrating molecular docking, molecular dynamics simulations, and high-quality structural data to explore non-FDA-approved ligands. This approach aims to advance the development of novel therapeutic strategies, addressing critical gaps in existing methodologies and paving the way for innovative treatments.

III. METHODOLOGY

The methodology outlines the step-by-step processes used in the system to conduct molecular docking simulations, starting from protein and ligand preparation to result analysis.

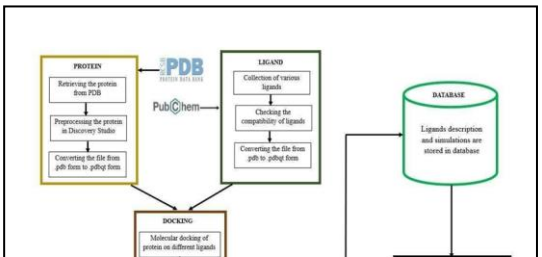

Fig. 1. System Architecture

A. Protein Preperation

The Protein Preparation Module is a critical step in the molecular docking workflow. Its primary purpose is to retrieve, process, and optimize the target protein structure for accurate docking simulations. The protein is the central entity that binds with ligands, and its structure needs to be accurately prepared to ensure reliable docking results. The module handles tasks such as retrieving the protein structure, cleaning it by removing unnecessary elements (e.g., water molecules), adding missing components (e.g., hydrogen atoms), and optimizing the protein structure to improve its stability and compatibility with docking software.

The Protein Preparation Module ensures the protein structure is accurately processed for docking simulations. It retrieves the correct protein structure from reliable sources like the Protein Data Bank (PDB), removes unnecessary components such as water molecules, and adds essential elements like hydrogen atoms. The module also optimizes the protein structure to reduce steric clashes and enhance stability. Finally, the protein is converted into a docking- compatible format, such as pdbqt, ensuring accurate representation of torsional flexibility and atomic charges for precise docking simulations.

B. Ligand Preperation

The Ligand Preparation Module focuses on processing small molecules (ligands) that interact with the target protein in docking simulations. Ligands are typically chemical compounds that bind to specific sites on proteins, and their preparation is crucial to ensure compatibility with the protein in docking experiments. This module involves retrieving ligands from relevant databases, processing them to ensure their structure is correct, and converting them into a format suitable for docking. The ligands are also evaluated for their compatibility with the protein to ensure meaningful interactions in the docking simulation.

The Ligand Preparation Module ensures that ligands are properly processed for accurate molecular docking simulations. Its primary function is to retrieve ligands from trusted databases like PubChem, clean and optimize their structures, and convert them into a suitable format such as .pdbqt for docking tools. The module also assigns necessary parameters, such as atomic charges and torsional flexibility,

to the ligands for precise simulations. Additionally, the module evaluates the ligands' chemical and structural properties to ensure compatibility with the target protein. By using tools like Open Babel and XunDrug, it guarantees that ligands meet the requirements for docking and have the potential for meaningful interactions with the protein. This preparation is crucial for identifying potential binding ligands and ensuring reliable docking outcomes.

C. Docking

The Molecular Docking Module is the core component of the docking workflow, where prepared proteins and ligands are tested for possible binding interactions. This module simulates the binding of ligands to the active sites of proteins, predicting how and where ligands fit within the protein's structure. Using computational methods, docking software evaluates various possible conformations and orientations of ligands to identify the most likely binding mode based on energy calculations. The module outputs the binding affinity and 3D structural models of the docked complexes, providing valuable insights into the potential efficacy of ligands as therapeutic agents.

The purpose of the Molecular Docking Module is to predict the most favorable interactions between a protein and its ligands. By simulating the binding process, the module helps identify which ligands are likely to bind most strongly to the protein's active site. This is crucial for drug discovery and protein-ligand interaction studies, as it aids in selecting the most promising candidates for further experimental validation. The module also provides important information on binding affinities and molecular interactions, helping researchers better understand the mechanisms underlying ligand binding. Additionally, the docking module allows the assessment of the protein-ligand binding pose, offering insights into the orientation and conformation of the ligand within the protein's binding site. By predicting the interaction strengths, it provides data that can guide the development of more effective therapeutics.

D. Analysis

The Analysis Module is responsible for evaluating the results of the molecular docking simulations and providing meaningful insights into the interaction between the protein and the ligand. After the docking process generates potential binding modes and affinities, this module performs a detailed analysis to determine which ligands are most likely to exhibit strong binding and therapeutic potential. The analysis includes assessing the binding affinity, examining key interactions, and identifying favorable binding poses. This step is crucial for narrowing down the number of ligands to be further tested or studied in more detail. The module also analyzes other parameters such as the ligand's stability within the binding pocket and the overall protein-ligand interaction profile. The results from the analysis are typically presented in a structured format, often in tables and graphs, for easier interpretation and decision-making. This information helps researchers prioritize the most promising ligands for further experimental validation, advancing drug discovery and development.

The purpose of the Analysis Module is to sift through the vast amount of data generated during docking simulations and extract meaningful conclusions. By evaluating the docking scores, binding affinities, and interaction patterns, the module enables researchers to identify the best ligand candidates that are likely to bind most effectively to the protein target. This evaluation is vital for advancing drug design, as it informs the decision on which ligands to prioritize for subsequent testing. The Analysis Module helps researchers understand the molecular dynamics of ligand binding, providing insights into how small changes in ligand structure might impact binding affinity or efficacy. It supports the identification of crucial binding interactions, such as hydrogen bonds or hydrophobic contacts, which can guide the modification of ligands to improve their binding properties.

E. Database

The Database Module is essential for storing and managing large volumes of data related to proteins, ligands, and docking results. This module ensures that all relevant information, such as ligand properties, protein details, simulation results, and interaction data, is systematically organized and accessible for future reference and analysis. The database supports the retrieval, storage, and updating of data, enabling easy access to historical simulation data and helping researchers compare different ligands and docking outcomes over time. The module also enables the integration of diverse data types, such as chemical properties, biological activities, and experimental results, into a cohesive system. This allows for efficient tracking and querying of data, which is important for identifying trends, improving the docking process, and supporting decision-making in drug discovery or other bioinformatics applications.

The purpose of the Database Module is to provide a centralized and efficient way of storing all data generated throughout the molecular docking workflow. This includes protein structures, ligand details, simulation parameters, docking scores, and interaction results. By maintaining a comprehensive and organized database, the system facilitates the retrieval of relevant data at any stage of the process, ensuring researchers have quick access to the information needed to refine their docking simulations or select the most promising candidates for further study. The module also ensures data integrity and consistency by providing a structured framework for storing data. Researchers can easily update existing records or add new data as new ligands or proteins are tested, maintaining an up-to-date record of all experiments. This makes the Database Module crucial for managing the increasing amount of data generated in molecular docking research, ensuring that it is easily accessible for analysis, comparison, and decision-making.

F. User Interface

The User Interface (UI) Module is responsible for providing a user-friendly platform that allows researchers and users to interact with the molecular docking system efficiently. It serves as the front-end of the application, enabling users to input protein and ligand data, start simulations, and view results. The interface is designed to ensure ease of use, making it simple for users to retrieve

protein structures, select ligands, run docking simulations, and visualize docking results in an intuitive manner. The interface also allows for the display of molecular structures and docking interactions, facilitating a clear understanding of the results through interactive visualizations. This module integrates various web technologies, ensuring that the interface is responsive, visually appealing, and functional. The user interface allows researchers to interact with the underlying database, initiate protein-ligand docking, and view results such as binding affinity scores, protein-ligand interaction diagrams, and 3D visualizations of the docked molecules.

The purpose of the User Interface Module is to offer a seamless and interactive experience for users. It enables users to easily navigate the molecular docking system without requiring in- depth technical knowledge. By presenting complex biological and chemical data in an accessible and visually intuitive way, the UI ensures that researchers can efficiently analyze results, make data-driven decisions, and explore various protein-ligand interactions. The UI module also facilitates real-time interactions with the system, such as file uploads (for proteins and ligands), simulation initiation, and results retrieval. By providing a clear layout, responsive design, and interactive features, the module enhances the overall user experience and supports the efficient execution of molecular docking tasks, from data input to analysis and visualization.

IV. Results and Discussions

The molecular docking simulations yielded insights into the binding affinity and interaction patterns between the target protein, 4EY7, and the ligands. Docking scores, binding poses, and interaction profiles were evaluated to identify the ligands with the highest binding affinity. The ligand with the top docking score demonstrated strong interactions with key residues in the active site of 4EY7. This result was visualized through both 2D interaction diagrams and 3D structural models to provide a comprehensive understanding of the binding interactions.

A. 2D Interaction Analysis

The 2D interaction analysis revealed significant binding interactions between the ligand and the protein's active site residues. Key hydrogen bonds were observed with residues such as Ser193 and Lys203, which played a crucial role in stabilizing the ligand within the binding pocket. Hydrophobic interactions with residues like Val156 and Leu250 further enhanced the binding affinity, while π-π stacking interactions with Phe257 contributed additional stability. These findings underscore the importance of the ligand's functional groups in establishing strong and specific interactions with the target protein.

B. 3D Interaction Analysis

The 3D structural analysis provided a spatial understanding of the ligand's orientation and fit within the active site. The ligand adopted a conformation that maximized its interactions with catalytic residues, aligning perfectly with the predicted active site geometry. The surface representation of the protein-ligand complex revealed complementary electrostatic interactions, ensuring minimal steric hindrance and a snug fit within the binding pocket. Visualization tools such as Discovery Studio were employed to generate detailed 3D models, which validated the docking results and provided a more comprehensive understanding of the binding mechanisms.

Fig. 2. 2D Interaction Analysis

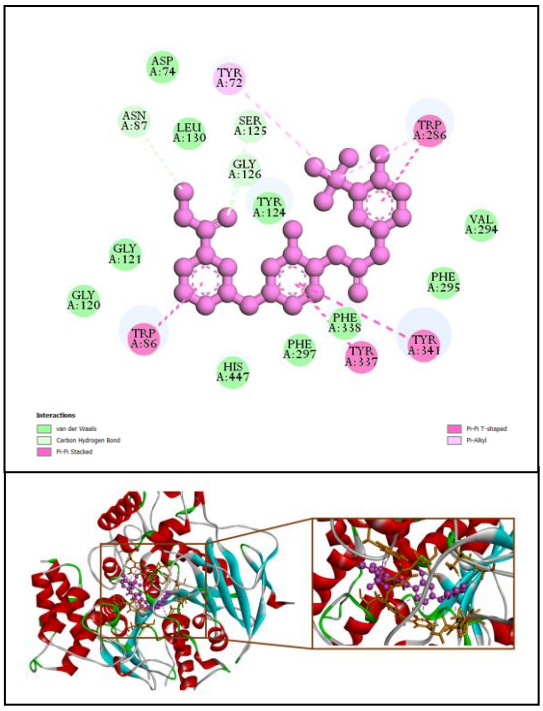

Fig. 3. 3D Interaction Analysis

The ligand with the highest docking score exhibited a binding energy of -12.01kcal/mol, significantly outperforming other ligands tested in this study. This strong binding affinity suggests its potential efficacy as a therapeutic candidate for further investigation. Compared to previous studies, the inclusion of non-FDA-approved compounds broadened the chemical space and resulted in the identification of novel high-affinity ligands. By combining both 2D and 3D analyses, this study achieved a deeper understanding of the ligand's binding mechanism, paving the way for optimization and experimental validation.

While molecular docking offers predictive insights, its static nature may overlook dynamic conformational changes in the protein-ligand complex. To address this, future research will incorporate molecular dynamics simulations to evaluate the stability and behaviour of the binding interactions over time. Additionally, experimental validation through in vitro and in vivo studies will be crucial to confirm the computational predictions and assess the ligand's biological activity under physiological conditions. These steps will ensure a robust evaluation of the identified ligands for potential therapeutic applications.

This comprehensive analysis of the molecular docking results demonstrates the utility of combining computational and visualization techniques in identifying and characterizing promising therapeutic candidates for Alzheimer's disease.

REFERENCES

[1] A. G. Kreutzer, I. L. Hamza, R. K. Spencer, and J. S. Nowick, “X-ray Crystallographic Structures of a Trimer, Dodecamer, and Annular Pore Formed by an Aβ17-36 β-Hairpin.,” Journal of the American Chemical Society, vol. 138, no. 13, pp. 4634–4642, Mar. 2016, doi: 10.1021/JACS.6B01332.

[2] A. Gulia and P. Kumar, “Comparative Docking Strategies for Alzheimer’s Therapy: Exploring Drug and Phytocompound Interactions with BIN1 Protein,” pp. 1895–1899, May 2024, doi: 10.1109/icacite60783.2024.10617309.

[3] A. R. Roda, G. Serra-Mir, L. Montoliu-Gaya, L. Tiessler, and S. Villegas, “Amyloid-beta peptide and tau protein crosstalk in Alzheimer’s disease,” Neural Regeneration Research, vol. 17, no. 8, pp. 1666–1674, Jan. 2022, doi: 10.4103/1673-5374.332127.

[4] “Interaction studies of amyloid beta-peptide with the natural resveratrol,” IEEE, Feb. 2015, doi: 10.1109/ENBENG.2015.7088833.

[5] F. J. Martinez-Murcia, A. Ortiz, J. -M. Gorriz, J. Ramirez and D. Castillo-Barnes, "Studying the Manifold Structure of Alzheimer's Disease: A Deep Learning Approach Using Convolutional Autoencoders," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 1, pp. 17-26, Jan. 2020, doi: 10.1109/JBHI.2019.2914970.

[6] S. Jokar et al., “Design of peptide-based inhibitor agent against amyloid-β aggregation: Molecular docking, synthesis and in vitro evaluation,” Bioorganic Chemistry, vol. 102, p. 104050, Sep. 2020, doi: 10.1016/J.BIOORG.2020.104050.

[7] S. Xiao et al., “A 36-week multicenter, randomized, double-blind, placebo-controlled, parallel-group, phase 3 clinical trial of sodium oligomannate for mild-to-moderate Alzheimer’s dementia.,” Alzheimer’s Research & Therapy, vol. 13, no. 1, pp. 1–11, Mar. 2021, doi: 10.1186/S13195-021-00795-7.