Integrative Computational Framework for Dopamine D3 receptor-Targeted Drug Discovery: Bridging Genetic Variability, Structural Dynamics, and Machine Learning

Thesis submitted

in partial fulfilment of the requirements for the

degree of

MASTER OF TECHNOLOGY

in

BIOINFORMATICS

by

AKSHAY HATWAL

23/BIO/07

Under the supervision of

PROF. PRAVIR KUMAR

Professor and Dean of International Affairs

Department of Biotechnology

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, New Delhi, 110042

May, 2025

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, New Delhi, 110042

DECLARATION

I, Akshay Hatwal 23/BIO/07 hereby certify that the work which is being presented in the thesis entitled **"Integrative Computational Framework for Dopamine D3 receptor-Targeted Drug Discovery: Bridging Genetic Variability, Structural Dynamics, and Machine Learning"** in partial fulfillment of the requirements for the award of the Degree of Master of Technology, submitted in the Department of Biotechnology, Delhi Technological University is an authentic record of my own work carried out during the period from 2023 to 2025 under the supervision of Prof. Pravir Kumar.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, New Delhi, 110042

SUPERVISOR'S CERTIFICATE

Certified that Akshay Hatwal(23/BIO/07) has carried out their search work presented in this thesis entitled "**Integrative Computational Framework for Dopamine D3 receptor-Targeted Drug Discovery: Bridging Genetic Variability, Structural Dynamics, and Machine Learning"** for the award of Master of Technology from Department of Biotechnology, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Date:

Signature
Prof. Pravir Kumar
Professor and Dean IA
Department of Biotechnology
Delhi Technological University

Signature
Prof. Yasha Hasija
Head of the Department
Department of Biotechnology
Delhi Technological University

Integrative Computational Framework for Dopamine D3 receptor-Targeted Drug Discovery: Bridging Genetic Variability, Structural Dynamics, and Machine Learning

AKSHAY HATWAL

ABSTRACT

Aim: The Dopamine D3 receptor (D3R), a G-protein coupled receptor predominantly expressed in the limbic system, plays a pivotal role in modulating reward, cognition, and emotional behaviors, and is implicated in neuropsychiatric and neurodegenerative disorders such as Parkinson's disease and schizophrenia. Given its therapeutic relevance, this study presents an integrated computational framework to elucidate the structural and functional consequences of D3R mutations, their impact on ligand binding, and the application of advanced machine learning for drug discovery. Evolutionary conservation analysis using ConSurf identified functionally critical regions within D3R, while PredictSNP predicted deleterious effects for 73 out of 405 single amino acid variants, with a majority located in highly conserved regions. To further probe the dynamic consequences of mutation, molecular dynamics (MD) simulations were performed for the Wild-type and selected variants (P344T, L60P) over 100 ns. A cascade neural network-based quantitative structure-activity relationship (QSAR) model was developed and trained on a large, chemically diverse dataset. The model utilized a two-stage architecture with uncertainty estimation and active learning, leveraging molecular fingerprints and 2D/3D descriptors. Performance evaluation demonstrated robust classification of high- and low-affinity ligands, with high ROC-AUC (0.888), average precision (0.916), and strong rank correlation (Spearman 0.780). The model's precision-recall and ROC curves indicated high discriminative power, while confusion matrix analysis revealed a conservative bias, minimizing false positives at the cost of some false negatives-a trade-off often desirable in early-stage drug discovery.

Results and Conclusion: The study identified 73 deleterious D3R variants, predominantly in conserved regions, correlating with reduced dopamine binding affinity and altered receptor dynamics. Molecular docking and MD simulations confirmed that mutations in critical regions impair function, while those in variable regions are generally tolerated. The cascade neural network QSAR model achieved high accuracy (79.7%), precision (91.6%), and ROC-AUC (0.888), effectively distinguishing high- and low-affinity ligands. Misclassifications were mainly confined to borderline cases, indicating robust model calibration. This integrative computational approach provides a powerful platform for D3R-targeted drug discovery, supporting the rational design and prioritization of novel therapeutics for neuropsychiatric and neurodegenerative disorders, with future work focusing on model interpretability and experimental validation.

ACKNOWLEDGEMENT

First of all, I would like to extend my heartfelt gratitude to my supervisor, Prof. Pravir Kumar, for their constant support and encouragement throughout the course of my research. Their expertise and constant constructive feedback have always been invaluable to me in learning and understanding new yet difficult things. They have always been an inspiration to me in pursuing my research and future goals. I would also like to thank Mrs. Neetu Rani, Ms. Mehar Sahu, Mr. Rahul Tripathi, Ms. Shrutikriti Vashishth, Ms. Shefali Kardam and each and every member of the Molecular Neuroscience and Functional Genomics laboratory for always helping and guiding me through each up and down which has come along the way and giving me an environment where I can learn and grow as a person.

I am also sincerely grateful to the faculty and non-faculty staff of the Department of Biotechnology, Delhi Technological University for providing me an academic environment which is coupled with theoretical as well as practical aspects of academia and providing me every necessary help and resource to carry out my work. A special thanks to Mr. Jitender Singh, Mr. C.B. Singh, Mr. Lalit, Mr. Jaspreet, and Mr. Rajesh for their technical assistance and guidance through the project.

Finally, I would like to whole-heartedly thank my family and friends for always being my constant support and well-wishers. Their presence and belief in me have always been a source of motivation and strength, their contribution in my life can’t be put into words.

Thank you all for your valuable contributions and sacrifices which made this project a success.

Akshay Hatwal

23/BIO/07

TABLE OF CONTENTS

List of Tables

List of Figures

List of Symbols and Abbreviations

Symbol/Abbreviation	Description
D3R	Dopamine D3 Receptor
GPCR	G Protein-Coupled Receptor
CNS	Central Nervous System
MD	Molecular Dynamics
QSAR	Quantitative Structure-Activity Relationship
SMILES	Simplified Molecular Input Line Entry System
PDB	Protein Data Bank
nsSNP	Non-synonymous Single Nucleotide Polymorphism
RMSD	Root Mean Square Deviation
Rg	Radius of Gyration
AUPRC	Area Under Precision-Recall Curve
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
ADMET	Absorption, Distribution, Metabolism, Excretion, Toxicity
BBB	Blood-Brain Barrier
API	Application Programming Interface
PDB	Protein Data Bank
NVT	Constant Volume and Temperature Ensemble
NPT	Constant Pressure and Temperature Ensemble
PME	Particle Mesh Ewald
LINCS	Linear Constraint Solver
V-rescale	Velocity-rescale Thermostat
AdamW	Adaptive Moment Estimation with Weight Decay
PR	Precision-Recall

1. INTRODUCTION

The D3 receptor is found in the limbic system and concentrated in the nucleus accumbens and the islands of Calleja[1]. These distributions are relevant due to its critical role in the regulation of reward and reinforcement-related behaviors and locomotion. Encoded by the DRD3 gene, DRD3 is made up of 400 amino acid residues. As it plays an important role in cognition and behavior, D3 receptor is involved in the pathogenesis of neuropsychiatric and neurodegenerative disorders, which makes it an important therapeutic target[2]. Accili et al. in 1996, started a major study to examine the role of the D3 receptor by generating mice with a mutated DRD3 gene, by inserting a premature stop codon to eliminate receptor expression. Homozygous mice (lacking D3 receptors) showed no notable developmental abnormalities. Furthermore, they showed behavioral changes, supporting the role of D3R in modulating anxiety and depression. The findings identified that D3R is a potential therapeutic target for developing antidepressant drugs.

Further supporting this role, a study by J. H. Seo and E. V. Kuzhikandathil (2015) explored the impact of stress on D3R function. Mice that were exposed to excess restraint stress and social isolation developed anxiety and depression-like behaviors in adulthood[3]. Administering a D3 receptor antagonist, SB277011, during the stress period prevented the development of these behaviors, pointing towards the therapeutic potential of modulating D3R activity to mitigate stress-induced psychiatric conditions. Polymorphisms in the DRD3 gene have also been suspected to be linked with cognitive abnormalities and anxiety levels in Parkinson's disease (PD) patients[4]. Few genotypes were found to predispose individuals to abnormal cognitive function or higher levels of anxiety, while others appeared to reduce these symptoms. These findings emphasize the importance of understanding genetic variability in DRD3 and its influence on phenotypes, particularly in neurodegenerative disorders like PD. Structural changes in the D3 receptor significantly affect its function. For instance, D3R also participates in autophagy regulation[1].

Gene knockdown studies in HeLa cells revealed that the reduced D3 receptor expression diminished autophagic flux, highlighting its importance in maintaining cellular homeostasis. This finding identified D3R as a promising therapeutic target for neurodegenerative diseases, where dysregulated autophagy is a contributing factor. In this study, we aimed to comprehensively evaluate the structural and functional implications of the D3 receptor. We analyzed the impact of deleterious single amino acid mutations in D3R identified from databases such as UniProt.

By using advanced computational tools, we investigated how these mutations influence the receptor's stability, ligand-binding affinity, and signaling capacity. These insights not only improve the understanding of these molecular interactions underlying D3 receptor dysfunction but also provide indications for identifying novel therapeutic interventions targeting D3R in psychiatric and neurodegenerative disorders.

It is important to do this structural analysis to better understand this receptor before constructing the cascade neural network (CNN). Various parameters are to be calculated to validate this model before docking. The screened drugs will be compared with the Wild-type, with dopamine, its natural ligand.

The whole process of structural investigation will validate the structure that will further be used to support the validation of the neural network[5].

If the correct structure is selected, it is expected to behave in the same way as a natural ligand-receptor interaction and show strong interaction without mutation and be unaffected.

If mutations are introduced in the structure, such testing will make the structure more reliable. The proposed CNN model classifies the provided drug SMILES dataset into two classes - 0 and 1.

Class 0 contains SMILES that have low probability of being a high affinity with the receptor. Class 1 contains molecules having a high probability of strong interaction. The model is trained with 15,000 molecules curated from PubChem using their API. A classification-type model is chosen because of the lack of a bigger and experimentally validated affinity dataset for the Dopamine D3 receptor. A larger and more niche dataset will be needed to explicitly predict the binding affinity of molecules and the D3 receptor.

This computational pipeline is adaptable to other protein-ligand systems and supports virtual screening, lead optimization, and candidate prioritization for experimental validation. The findings highlight the significance of integrating structural biology, mutational analysis, and machine learning to accelerate the discovery of selective D3R modulators. Future directions include expanding curated datasets, enhancing model interpretability through explainable AI, and bridging computational predictions with experimental validation to ensure clinical translatability. This approach advances precision medicine by elucidating the interplay between genetic variability, receptor function, and drug response in neuropsychiatric and neurodegenerative diseases.

2. LITERATURE REVIEW

2.1 Dopamine D3 receptor - structure, function, and disease relevance

2.1.1 Structural and functional characteristics

Dopamine D3 receptor is a GPCR (G-protein coupled receptor) with high affinity towards dopamine, which is approximately 420 times that of the D2 receptor. It is primarily localized in the nucleus accumbens and other regions (e.g., D1R, D2R, A2aR).
The ability of D3R to form heteromers with other receptors like D1R, D2R, and A2aR further diversifies its signaling pathways and pharmacological significance[6].

2.1.2 Role in Neurodegenerative and Neuropsychiatric Disorders

D3R involvement in neurodegenerative diseases is multidimensional. For example, its neurotrophic and neuroprotective effects help in regulating dopaminergic neuron homeostasis and preventing neurodegeneration. Abnormal D3R signaling has been involved in Parkinson's disease, essential tremors, schizophrenia, addiction, and disorders related to mood.
Knockout mice studies showed that the dysfunction in D3R due to mutation did not affect the growth or physiological characteristics of the mice; however, it made the mice more prone to anxiety and depression. This also supports the relevance of D3R in psychiatric disorders and as a therapeutic target for antidepressant drugs.

2.1.3 D3R as a drug target

D3R-selective antagonists and agonists have long been considered as therapeutic agents for treating Parkinson's disease, schizophrenia, and drug addiction. The similarity with D2R and D4R makes it harder to deal with as a therapeutic target, which is why while building the neural network, it was made sure that the model learns patterns of features specific to high binding drugs[7].

2.2 Advances in structural biology and computational modeling

2.2.1 Structure Analysis of D3R

Recent improvements in cryo-EM, crystallography, and computational modeling have provided us with atomic-level insights into D3 receptor structure and dynamics. These studies show the importance of conserved regions and effects of mutation in the receptor-ligand interaction.

2.2.2 Integrative Computational Approaches

Head compound identification and drug discovery is revolutionized by integrating structural biology, cheminformatics, and machine learning. Molecular docking, molecular dynamics, and virtual screening help researchers explore large libraries and datasets of potential drug candidates by computationally calculating and predicting interaction of target protein and drug candidate rather than

random guesswork or laborious work that has to be done in vitro and in vivo. It is a standard approach in early-stage drug discovery.

2.3 Quantitative Structure-Activity Relationship (QSAR)

2.3.1 QSAR: Principle and Evolution

QSAR models are used to predict the biological activity of compounds based on their chemical features while using statistical and machine learning algorithms to correlate molecular features and their biological outcomes. The integration of machine learning has increased the efficiency and accuracy of traditional QSAR-based drug discovery.

2.3.2 Machine Learning in QSAR

Deep neural networks, random forest, and combined model methods have become a norm in modern QSAR modeling. It enables us to extract complex, non-linear correlations from multidimensional chemical data. Integration of three-dimensional descriptors and multi-instance learning boosted the accuracy of activity prediction.

2.3.3 Cascade Neural Network QSAR Model

Cascade neural network represents a self organising and provides an incremental learning architecture that aggregates hidden neurons during training which enables the model to adapt with the complexity of the data. This approach eliminates the requirement of predefined network size reduced overfitting and accelerates convergence which makes it particularly suitable for big datasets typically found in drug Discovery (reference). Cascade models exhibit superior performance in screening and predicting drug target interaction integrated with heterogeneous molecular features (e.g., fingerprints, 2D/3D descriptors).

2.4 Virtual Screening Cascades and integrative pipelines

2.4.1 Virtual screening

Virtual screening pipelines is an aggregation of multiple computational models that helps and prioritize drug candidates experimental testing. it effectively reduces attrition rates and experimental cost. These pipelines combine molecular Docking, molecular simulation, QSAR models, ADMET prediction, and toxicity screening, enabling a thorough evaluation of candidate molecules before *in vivo* and *in vitro* experiments[8].

2.4.2 Application to D3R drug Discovery

The virtual screening pipeline both models of both Wild-type and mutant receptor enabling the identification of optimum binding affinity selectivity and safety profiles. The selection of cascade neural network QSAR models increases the predictive accuracy and efficiency of the pipeline which supports the discovery of novel modulators of Dopamine D3 receptor psychiatric and neurodegenerative diseases.

2.4.3 Importance of structure files

A holistic analysis of D3 receptor structure helps in defining specific characters that lies within the binding pocket of the target receptor. The use of experimentally determined data such as binding affinity and structure activity relationship is very important for benchmarking model performance and ensuring translational relevance.

2.4.4 Advantages and limitations of cascade neural network models

5.1 Advantages

adaptive complexity - The ability of cascade models to dynamically adjust network size which prevents overfitting or underfitting. fast convergence - Incremental learning accelerates model training and enhances robustness to local minima. Integration of Diverse Features: Capable of processing heterogeneous molecular descriptors, improving prediction accuracy. Superior Performance: Demonstrated higher accuracy in drug–target interaction prediction and virtual screening compared to traditional methods.

5.2 Limitations

Cascade neural networks, like other deep learning models, can be less interpretable than linear models or decision trees. Data Requirements: Require sufficient, high-quality data for optimal performance; performance may degrade with sparse or noisy datasets. Computational Resources: Although more efficient than some deep learning architectures, cascade models still demand significant computational power for large-scale screening34.

The convergence of structural biology, machine learning, and integrative computational pipelines is accelerating the discovery of novel therapeutics targeting the Dopamine D3 receptor. Cascade neural network QSAR models, validated using high-quality structural files and experimental data, offer a powerful framework for virtual screening and lead optimization. As data availability and computational methods continue to advance, these approaches will enable more precise, efficient, and personalized drug discovery for neuropsychiatric and neurodegenerative diseases. Expanding structural and functional datasets for D3R and its variants.

Developing interpretable machine learning models to enhance trust and transparency in predictions.Integrating multi-omics and systems biology data for holistic target validation and drug repurposing. Bridging the gap between computational predictions and experimental validation to ensure clinical translatability.

3. METHODS

3.1 Overview

This study is focused on understanding the structural aspects of D3 receptor and validating the available structural file to create a better platform for the validation of neural network which are used to classify between binding affinity molecules and low binding affinity molecules. While diving into the structural analysis of Dopamine D3 receptor, the effects of mutation on the binding affinity between dopamine and D3 receptor was also explored. This is a type of model that can be implemented for different target proteins and their set of ligands. the acquisition of data set required the use of APIs provided by databases such as PubChem and ZINC. these APIs were used to extract SMILES of ligands With their experimentally validated binding affinity towards the D3 receptor, in the same way these datasets can be acquired for any target protein that is present in the database. This also proposes a universally applicable cascade neural network model that can be used in drug discovery application in other diseases and disorders. we have developed a general model that classifies between high winding and low binding affinity which do not take account of the variants of the protein that are present in nature. It is proven in the earlier studies that certain nsSNP mutations can affect the symptoms by either escalating them or protecting against them, which makes it important for the researchers to integrate the data screening drugs for therapeutics.

3.2 Curation of data

Data from various databases was curated from UniProt, PubChem, ZINC databases. APIs provided by the respective database was used to extract datasets - D3 protein structure, Variant position and substitution, experimentally validated ligand SMILES and their binding affinity with D3 receptor.

(a) The structure of D3 receptor and information about 405 single amino acid substitution caused by ns-SNP was downloaded from UniProt (UniProt ID: P35462)[9].

Fig. 1. Structure of Dopamine D3 receptor

(b) A dataset of SMILES of compounds with their binding affinity was extracted from PubChem using the API provided by PubChem[10].

```python
import pubchempy as pcp
import pandas as pd

# Search for bioassays linked to the Dopamine D3 receptor
assays = pcp.get_assays('dopamine D3 receptor', 'target')

compounds_set = set()
for assay in assays:
    compounds = pcp.get_assay(assay.aid, 'aid').cids
    compounds_set.update(compounds)

# Limit to first 20 for demonstration
compound_data = []
for cid in list(compounds_set)[:20]:
    comp = pcp.Compound.from_cid(cid)
    compound_data.append({
        'CID': cid,
        'Name': comp.iupac_name,
        'SMILES': comp.isomeric_smiles
    })

# Save to CSV
df = pd.DataFrame(compound_data)
df.to_csv("pubchem_drd3_ligands.csv", index=False)
print(df)
```

Fig. 2. Snippet of pubchem API

(c) A dataset of CNS labeled drugs was also extracted using the API of ZINC[11] database as a test dataset (Fig. 3.).

```python
import pandas as pd
import requests

# ZINC15 CNS subset with SMILES (2D structures)
zinc_url = "http://files.docking.org/2D/cns/cns_2d.smi"

# Download the .smi file and read into pandas
response = requests.get(zinc_url)
if response.status_code == 200:
    with open("zinc15_cns_2d.smi", "w") as f:
        f.write(response.text)

    # Load the SMILES data into pandas
    df = pd.read_csv("zinc15_cns_2d.smi", sep="\t", names=["SMILES", "ZINC_ID"])
    print(df.head())

    # Save to CSV
    df.to_csv("zinc15_cns_drugs_smiles.csv", index=False)
    print(f"Saved {len(df)} CNS-labeled compounds from ZINC15.")
else:
    print("Failed to download the file. Check the URL or try again later.")
```

Fig. 3. Snippet of ZINC API

This dataset contained 20,000 compounds that were labeled as CNS-active drugs, these are compounds that have the ability to cross the Blood Brain Barrier(BBB). The dataset needs no filtering as the dataset contains CNS-active drugs and have all the properties like molecular weight between 200 - 450 Da, lipophilicity, BBB permeability, CYP450 inhibition, etc. making them suitable for this screening.

3.3 Predicting conserved regions of d3r structure using ConSurf

Consurf was used to predict the evolutionary conserved region of Dopamine D3 receptor. These regions are responsible for propagation of these nsSNPs to the future generations.

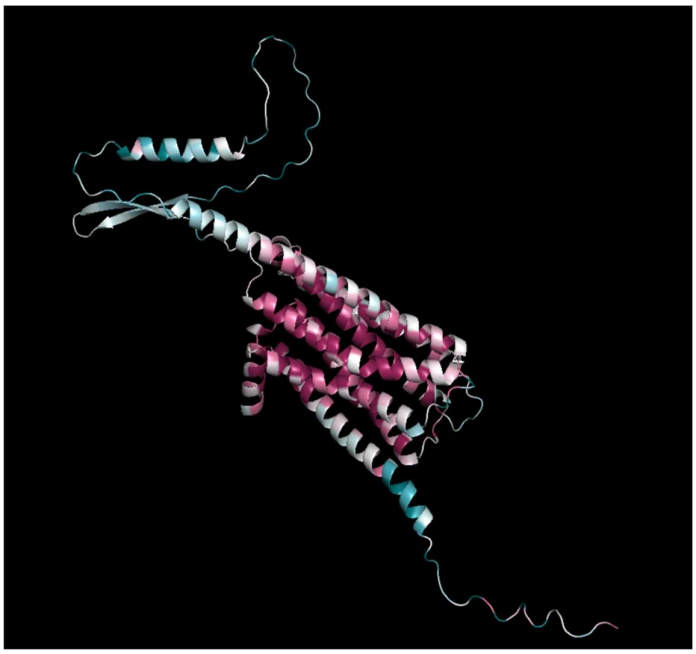

Fig. 4. Visualization of predicted conserved region (maroon) and variable region in (green)

3.4 Inducing mutations

Mutations were introduced using PyMOL software and exported in PDB format for docking analysis. When mutations were induced in the Wild-type protein, it shows small structural changes which were expected for this study. This facilitated structural and functional study of the DRD3 receptor.

3.5 Deleterious mutation prediction

PredictSNP was used to predict the impact of single amino acid mutations in the Dopamine D3 receptor (Fig. 1.). It combines results for tools such as SIFT, PhD-SNP, PolyPhen2, MAPP, PANTHER and Snap shows increased accuracy in its prediction. It provided a probability score on whether the variant is deleterious or neutral. 73 variants were screened on the basis of their highly deleterious predicted nature.

3.6 Molecular docking

AutoDock Vina was used to predict the binding affinity of Dopamine D3 receptor with its natural ligand Dopamine (DrugBank Accession Number DB00988). Mutation Wizard of PyMOL was used to induce mutations in the Wild-type Dopamine D3 receptor and prepare the 73 variants. Rigid protein docking was performed against Dopamine having five rotatable bonds. The protein being rigid may

not show the full extent of the effects of the mutations on its structure but it will give us indications on effects of mutation on the structure of the Wild-type protein. The Wild-type Dopamine D3 receptor and the 73 variants were docked against Dopamine. Docking simulations were performed with the following parameters: the grid center was set to coordinates X = 4.33, Y = 3.5, Z =-4.44, and the grid size was 22 × 30 × 26 ˚ A³ with a grid space of 0.375 ˚ A. Exhaustiveness was set to 32, and the docking was executed using 8 CPU threads. The docking was performed in a parallel fashion with a random seed of 638513881.

3.7 Molecular Dynamics Simulations

3.7.1 System Preparation & Feature Extraction

- Protein Structure Preparation:
  The initial structure (wild1.pdb) was processed using gmx pdb2gmx to generate topology and coordinate files. The OPLS-AA force field was chosen for the protein, and the SPCE water model was used for solvation. Hydrogens were added, and disulfide bonds were automatically detected and linked1.
- Box Definition:
  The processed protein was placed in a triclinic box with at least 1.0 nm distance from the protein to the box edge using gmx editconf1.
- Solvation:
  The system was solvated with SPC/E water molecules using gmx solvate, resulting in a realistic aqueous environment1.
- Ion Addition:
  The solvated system was neutralized and brought to physiological ionic strength by replacing water molecules with Cl^- ions using gmx genion.

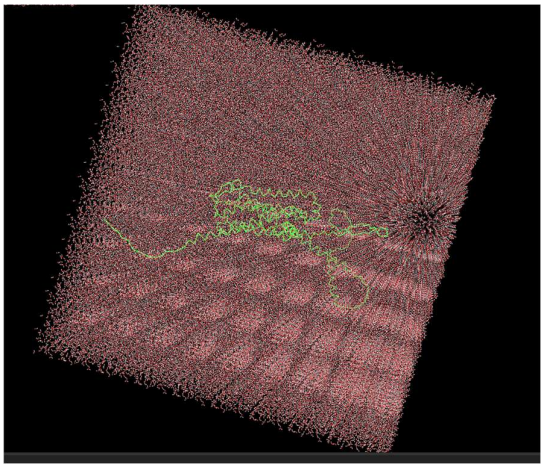

Fig. 5. Showing protein system in MD simulation using GROMACS

3.7.2 Energy Minimization, Equilibration, Production MD Simulation

- Energy minimization was performed using the steepest descent algorithm (integrator = steep) until the maximum force fell below 1000 kJ/mol/nm, as specified in the ions.mdp file3.
- NVT Equilibration (Constant Volume & Temperature) - **To stabilize temperature and relax solvent around the protein.**
    - Integrator: md (leap-frog)
    - Duration: 100 ps (or as specified)
    - Temperature coupling: V-rescale thermostat at 300 K
    - Position restraints applied to heavy atoms of the protein
    - No pressure coupling5
- NPT Equilibration (Constant Pressure & Temperature) - To stabilize system pressure and density.
    - Settings:
        - Integrator: md
        - Duration: 100 ps (or as specified)
        - Temperature coupling: V-rescale thermostat at 300 K
        - Pressure coupling: C-rescale (or Parrinello-Rahman) at 1 bar
        - Position restraints maintained
        - System allowed to adjust box dimensions to achieve target pressure4

- Production MD Simulation - **100 ns (50,000,000 steps with 2 fs timestep)**
    - Settings:
        - Integrator: md (leap-frog)
        - Periodic boundary conditions: xyz
        - Electrostatics: Particle Mesh Ewald (PME)
        - Cutoff scheme: Verlet
        - Temperature coupling: V-rescale at 300 K
        - Pressure coupling: Parrinello-Rahman or C-rescale at 1 bar
        - Constraints: LINCS for bonds involving hydrogens
        - Trajectory output: Coordinates, velocities, energies, and logs saved at defined intervals245
- Trajectory Analysis - **After 100 ns of simulation, trajectory and energy files were analyzed for:**
    - Structural Stability:
        - RMSD (Root Mean Square Deviation) to assess global conformational changes.
        - Radius of gyration for compactness.
        - Minimum distance between protein and ligand or within protein domains.
    - Intermolecular Interactions:
        - Number of hydrogen bonds over time to assess structural integrity and interactions.
    - Comparative Analysis:
        - Wild-type (yellow), P344T (green), and L60P (purple) were compared across all metrics to evaluate the effect of mutations on stability and dynamics (as shown in your previous graph analysis).

3.8 Cascade Neural Network QSAR Model

Cascade Neural Network QSAR Model Development ligand dataset were curated from PubChem and ZINC using their APIs which were focused on compounds with experimentally validated binding affinity data with D3R.The dataset included both known D3R ligands and CNS-active drugs (from ZINC), ensuring chemical diversity and relevance to blood-brain barrier permeability. each molecule was featurized using ECP4 fingerprints (1024 bits) for capturing molecular topology[16], 2D descriptors (e.g., molecular weight, logP, H-bond donors/acceptors, rotatable bonds, aromatic rings)and 3D descriptors (24 features capturing conformer statistics, surface area, TPSA, and ring system properties). Model training and architecture focuses on EnhancedBindingClassifier neural networks consisting of several fully connected layers (1030→512→256→128→1)[13], with batch normalization and dropout for regularization[14]. A deeper variant DeeperMisclassificationModel, includes additional layers and units for improved learning of complex patterns.

The model was trained on the balanced binarized dataset (high-affinity vs. low-affinity, thresholded at pChEMBL ≥ 7.0 or 8.0) using binary cross-entropy loss and the AdamW optimizer with dynamic learning rate. Early stopping was introduced based on validation AUC to prevent overfitting.

Model performance was evaluated using F1-score, ROC-AUC, balanced accuracy, and average precision. Threshold optimization was performed using youden's J statistics to maximize sensitivity and specificity. Misclassified molecules. This pipeline is versatile and can be applied to any protein ligand systems if binding data is available. If the ligand-protein interaction data is integrated in this model it will most likely increase the classification ability of the model.

The implemented model uses a two-stage cascade neural network architecture with three primary components:

- Primary Classifier: Initial classification network
- Uncertainty Estimator: Network determining prediction confidence
- Secondary Classifier: Specialized network for borderline cases

This design follows similar principles to other two-stage ensemble models in biomedical applications, like the one used for melanoma classification where multiple networks work in concert to produce higher-quality predictions.

1.Primary Classifier

The first stage consists of a neural network that processes molecular fingerprints:

```
class PrimaryClassifier(nn.Module):
    def __init__(self, input_size=1024):
        super().__init__()
        self.layers = nn.Sequential(
            nn.Linear(input_size, 512),
            nn.BatchNorm1d(512),
            nn.ReLU(),
            nn.Dropout(0.3),
            nn.Linear(512, 256),
            nn.BatchNorm1d(256),
            nn.ReLU(),
            nn.Dropout(0.3),
            nn.Linear(256, 128),
            nn.BatchNorm1d(128),
            nn.ReLU()
        )
        self.output = nn.Linear(128, 1)
```

Fig. 6. Primary Classifier

This network has four main fully-connected layers (1024→512→256→128→1) with batch normalization, ReLU activation, and dropout regularization (0.3) between layers. This architecture allows for progressively abstracting features from molecular representations while mitigating overfitting through regularization techniques.

2. Secondary Classifier

This classifier has a slightly different architecture (1024→512→256→1) and is specifically designed to handle borderline cases where the primary classifier shows uncertainty. By focusing on difficult cases, the secondary classifier can develop specialized feature recognition capabilities.

```
class SecondaryClassifier(nn.Module):
    def __init__(self, input_size=1024):
        super().__init__()
        self.fc_layers = nn.Sequential(
            nn.Linear(input_size, 512),
            nn.BatchNorm1d(512),
            nn.ReLU(),
            nn.Dropout(0.3),
            nn.Linear(512, 256),
            nn.BatchNorm1d(256),
            nn.ReLU()
        )
        self.output = nn.Linear(256, 1)
```

Fig. 7. Secondary Classifier

Uncertainty Estimation and Cascading Mechanism is a key innovation in this model is the explicit uncertainty estimation component.

```
self.uncertainty = nn.Sequential(
    nn.Linear(1024, 256),
    nn.Dropout(0.5),
    nn.Linear(256, 1)
)
```

This uncertainty estimator determines which samples should be passed to the secondary classifier, creating the cascading effect. The forward method implements this logic.
This approach resembles ensemble learning methods like those described in biomedical applications but differs in that it selectively applies the secondary classifier only to uncertain cases rather than combining all predictions. The threshold values (0.4 and 0.6) identify cases where the model is neither confident of a positive nor a negative prediction.

3. Active Learning Integration
The model incorporates active learning strategies to efficiently identify the most informative samples for labeling during training. This approach is particularly valuable in drug discovery where labeled data can be expensive and time-consuming to obtain.
The *ActiveLearningSampler* class implements two uncertainty sampling strategies:

1. Margin-based sampling: Selects samples with the smallest margin between the most likely and second most likely class predictions
2. Entropy-based sampling: Selects samples with the highest predictive entropy.

```
def query_samples(self, dataloader, n_samples=100):
    """Query the most uncertain samples"""
    # Implementation details...

    if self.strategy == 'margin':
        uncertainty = self._get_uncertainty_margin(outputs)
    elif self.strategy == 'entropy':
        uncertainty = self._get_uncertainty_entropy(outputs)
```

Fig. 9. *ActiveLearningSampler*

Custom Loss Function

The model employs a custom loss function specifically designed for pharmaceutical applications. This loss function builds upon the standard BCEWithLogitsLoss (Binary Cross Entropy with Logits) but adds an additional penalty for false negatives. In drug discovery, missing an active compound (false negative) is typically more costly than incorrectly flagging an inactive compound (false positive), as promising drug candidates should not be overlooked.

4. Molecular Feature Extraction

The model utilizes advanced molecular feature extraction techniques from RDKit.

```
def enhanced_3d_features(mol):
    # Generate multiple conformers
    cids = AllChem.EmbedMultipleConfs(mol, numConfs=10, pruneRmsThresh=0.5)

    # Conformer trajectory analysis
    # Shape descriptors
    pmom = rdShapeHelpers.PrincipalMoments(mol)
    pbf = rdMolDescriptors.CalcPBF(mol)

    # Quantum chemical features
    rdPartialCharges.ComputeGasteigerCharges(mol)

    # Solvent accessibility
    sasa = rdFreeSASA.CalcSASA(mol, rdFreeSASA.SASA_CLASS_NSURFACE)
```

Fig. 10. Feature extraction using Rdkit

This function extracts several types of molecular descriptors:

1. Conformer analysis: Generates and analyzes 3D conformers of molecules
2. Shape descriptors: Calculates principal moments and plane of best fit
3. Quantum chemical features: Computes Gasteiger partial charges
4. Solvent accessibility: Determines molecular surface areas accessible to solvent

The model also utilizes Morgan fingerprints, which are circular topological fingerprints that capture local molecular substructures. These binary representations encode presence or absence of specific molecular substructures and are widely used in cheminformatics for similarity searching and machine learning.

The training pipeline implements a multi-cycle approach with active learning integration. This approach follows an iterative process where:

1. The model is trained on available labeled data
2. Uncertain samples are identified using active learning
3. These samples would typically be labeled (in a real-world scenario)
4. The model is retrained with the augmented dataset

This cycle repeats, progressively improving the model's performance by focusing on the most informative samples.

4. RESULTS

4.1 Prediction of Deleterious Mutation

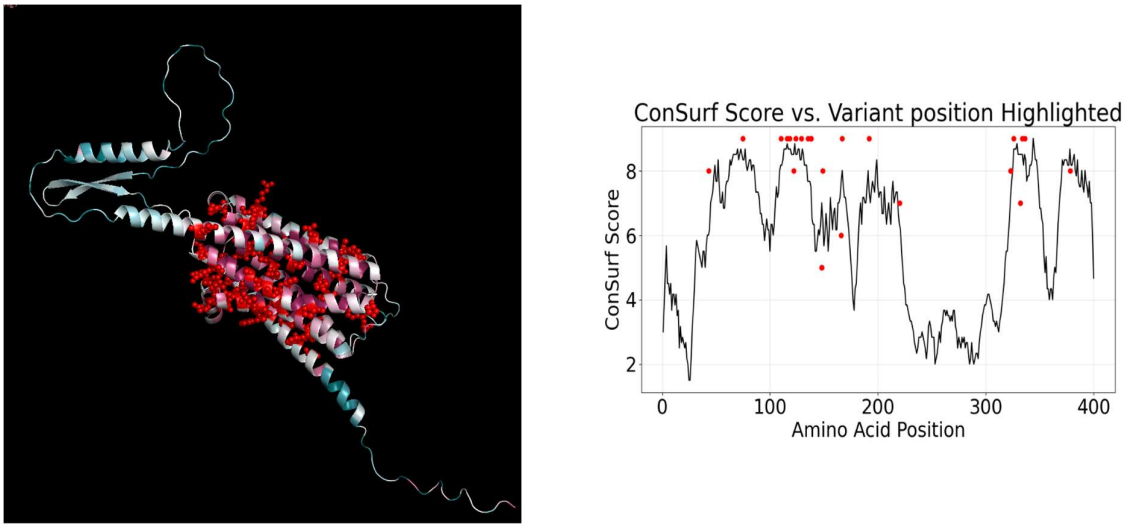

Fig. 11. (left) Dopamine D3 receptor structure colored green as highly variable and maroon as highly conserved regions with highlighted (red dots) with amino acid position of deleterious mutation predicted

Out of 405 variants, 73 variants were predicted to be deleterious (more than 86 percent probability) by PredictSNP. The relation between positions of conserved amino acid residues (conservation scores 8-9) and deleterious amino acids indicates that these sites are functionally or structurally important as the functionally relevant residues also fall in these regions. 50 of these variants fall in highly conserved regions (score 9-8), this shows that amino acid residues falling in this region are conserved so that the protein can function properly and pass down through generations.

4.2 Docking results

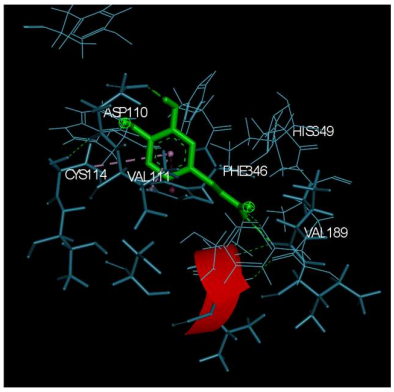

Fig. 12. Wild-type receptor docked with dopamine for structure validation and analyzed for binding with activating residues critical to D3 receptor

The best pose of Dopamine and Dopamine D3 receptor aligns with the structural and function studies conducted with Asp110, Val111, Ser192, Ser193, His349, Val189 being key players in ligand binding and activation of the receptor. This also validated the selection of grid parameters for this molecular docking. Docking studies using AutoDock Vina assessed the impact of these mutations on dopamine binding affinity. The Wild-type D3R showed a binding affinity of -5.757 kcal/mol with dopamine. Several variants, such as V334M, L109R, K326E, and L60P (TABLE I), exhibited significantly reduced binding affinity (e.g., L109R: -3.979 kcal/mol), indicating impaired receptor function. In contrast, variants like P344T, W370C, I382T, and P380S (TABLE II) maintained or even improved binding affinity, suggesting minimal structural disruption. This aligns with their location in more exposed or less functionally critical regions of the protein.

TABLE I

Mutations showing weakest binding affinity with dopamine

dbSNP ID	Variant	Affinity (kcal/mol)
	Wild-type	-5.757
rs756246784	L109R	-3.979
rs765754670	K326E	-3.980
rs747070751	L60P	-3.983
rs370774588	R323W	-3.984
rs370774588	I118N	-3.985
rs1380219014	P84A	-3.986
rs2077578658	L160P	–3.987
rs2077506873	S192F	-3.989
rs1263745324	Y129C	-3.990

TABLE II

Variants showing similar binding affinity with dopamine

dbSNP ID	Variant	Affinity (kcal/mol)
	Wild-type	-5.757
rs868709956	P344T	-5.977
rs2107825397	W370C	-5.838
TCGA novel	I382T	-5.802
rs2077399965	P380S	-5.787

These findings underscore that mutations in conserved regions often lead to functional impairment, while those in variable regions may be tolerated. The results highlight the importance of genetic variability in D3R function and its implications for drug response and precision medicine.

The change in interaction after mutation can also be featurized, which can be integrated to the ML model to further increase its accuracy in class prediction of predicting the binding affinity explicitly. Wild-type dopamine D3 and two variants were further analyzed using molecular dynamic simulations.

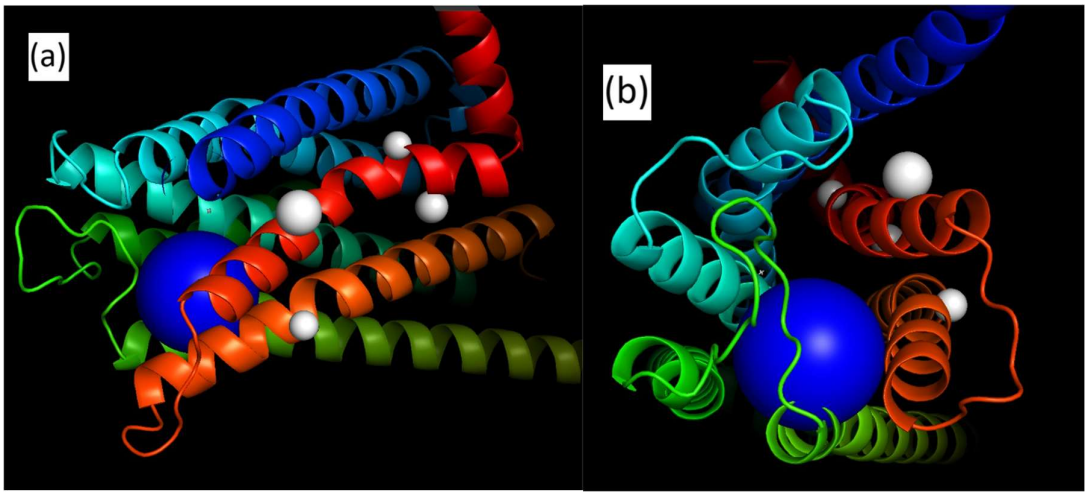

Fig. 13. (a) and (b) show the locations of variants exhibiting similar or increased binding affinity (white), with the binding site (blue sphere).

4.3 Molecular Dynamics (MD) Simulation Analysis

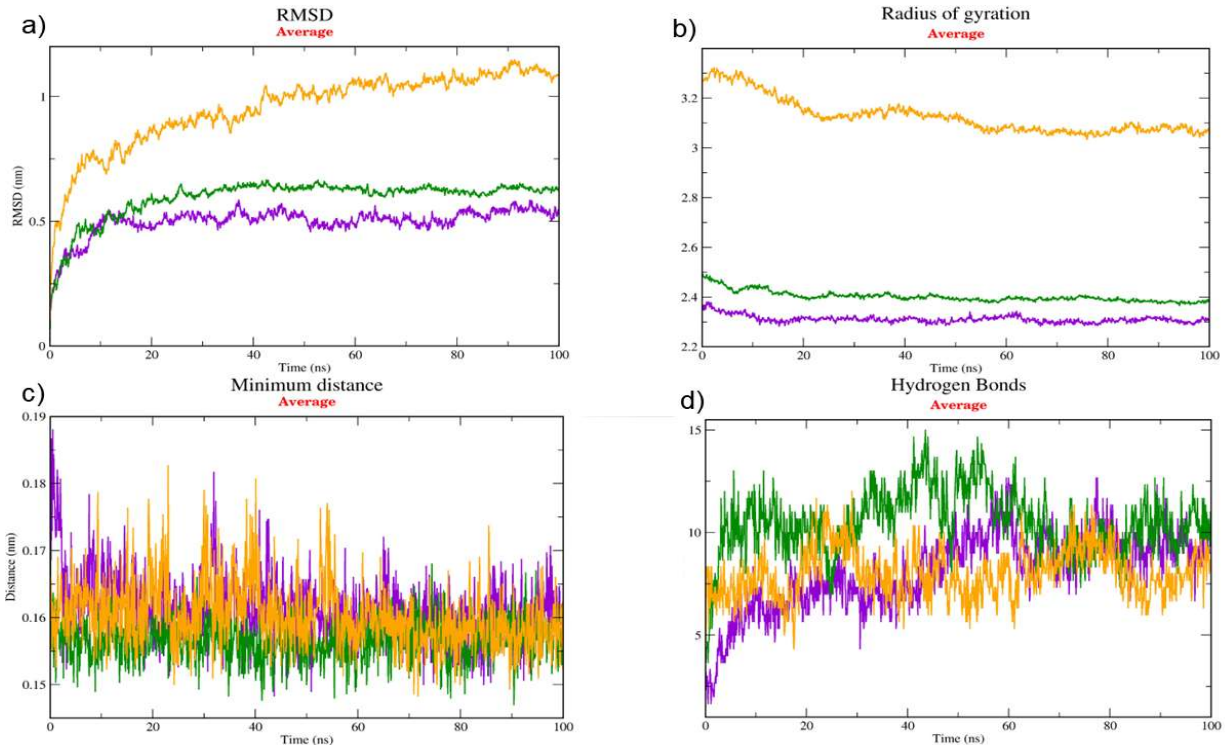

Fig. 14. The MD simulation results are visualized in the figurex , which compares wild-type (yellow) and mutant (green-P344T , purple- L60P) D3R complexes over 100 ns for several structural metrics

In these four panels, the behavior of three protein systems is compared over a 100 ns molecular dynamics (MD) simulation:

- Yellow: Wild-type protein
- Green: P344T
- Purple: L60P

a) RMSD (Root Mean Square Deviation)

- Description: RMSD measures the average deviation of atomic positions from the initial structure, indicating overall structural stability.
- Interpretation:
    - Wildtype (Yellow): Shows the highest RMSD, rising above 1.0 nm and not stabilizing, indicating significant structural fluctuations or instability.
    - P344T (Green): Intermediate RMSD (~0.6 nm), more stable than wildtype but less so than L60P.
    - L60P (Purple): Lowest RMSD (~0.5 nm), suggesting the highest structural stability among the three.

b) Radius of Gyration

- Description: This measures the compactness of the protein structure.
- Interpretation:
    - Wildtype (Yellow): Highest radius of gyration (~3.2 nm), indicating a more expanded, less compact structure.
    - P344T (Green): Intermediate compactness (~2.4 nm).
    - L60P (Purple): Most compact (~2.2 nm), suggesting L60P folds into a tighter structure.

c) Minimum Distance

- Description: The minimum distance between protein and ligand or between domains, reflecting close contacts.
- Interpretation:
    - All three systems fluctuate around similar values (0.15 -- 0.19 nm), but L60P (purple) and P344T (green) show slightly lower and more stable minimum distances, suggesting more persistent close contacts compared to wildtype.

d) Hydrogen Bonds

- Description: Number of hydrogen bonds, a key indicator of structural integrity and intermolecular interactions.

- Interpretation:
    - Wildtype (Yellow): Fewer hydrogen bonds overall, with values mostly below 10.
    - P344T (Green): Highest and most variable hydrogen bond count, peaking above 12, suggesting increased interaction potential or structural rigidity.
    - L60P (Purple): Intermediate hydrogen bond count, more stable than wildtype but less than P344T.

a) RMSD (Root Mean Square Deviation)
The Wild-type (yellow) shows the highest RMSD, indicating greater conformational flexibility or less stability over time.

Mutants (green, purple) display lower RMSD values, suggesting that certain mutations may stabilize the structure or restrict its flexibility.

b) Radius of Gyration
The Wild-type consistently exhibits a larger radius of gyration, reflecting a more expanded structure.

Mutants show lower and more stable values, pointing to increased compactness, which may affect ligand accessibility and receptor function.

c) Minimum Distance
The minimum distance between key structural elements fluctuates similarly across all variants, but the Wild-type maintains slightly higher average distances, again reflecting greater flexibility.

d) Hydrogen Bonds
The Wild-type forms more hydrogen bonds on average, which may contribute to its dynamic stability.

Mutants show fewer hydrogen bonds, possibly indicating altered intra-protein interactions due to mutation-induced structural changes.

MD simulations confirm that deleterious mutations can significantly alter the structural dynamics of D3R. While some mutations stabilize the receptor (lower RMSD and Rg), this may come at the cost of functional flexibility, potentially impairing ligand binding or signaling. Conversely, Wild-type D3R maintains a balance between stability and flexibility, essential for optimal function.

4.4 Cascade Neural Network QSAR Model

Excellent classification and ranking was observed with High AUCs for both ROC and PR curves, with well-separated prediction scores for each class. Model is well-calibrated: Most predictions are confidently assigned to the correct class.
Strong monotonic relationship between predicted and experimental rankings. Misclassifications occur primarily in the overlap region of the prediction score histogram.

The precision-recall (PR) curve is an essential tool for evaluating the performance of classification models, especially in situations where the classes are imbalanced or when the positive class is rare and of critical interest-such as in drug discovery, medical diagnostics, or fraud detection.

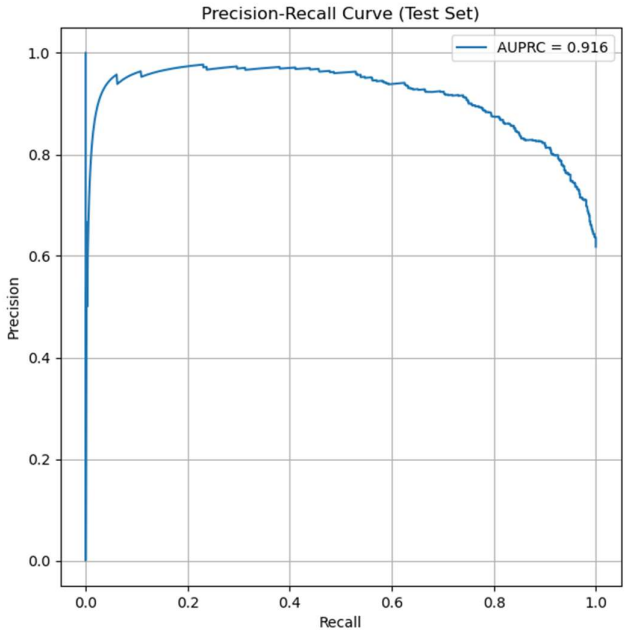

Fig 15. AUPRC (Area Under Precision-Recall Curve): 0.916.

The curve shows that the model maintains high precision across a wide range of recall values. This indicates strong performance, especially in identifying high-affinity molecules even if the dataset is imbalanced.

2. Prediction Score Histogram by True Class and ROC curve

- Blue (True Positive Class): Concentrated at higher predicted probabilities, showing the model assigns high scores to actual high-affinity molecules.
- Orange (True Negative Class): Concentrated at lower predicted probabilities, indicating low scores for actual negatives.
- Overlap: The middle region, where the two distributions overlap, corresponds to the area where most misclassifications (false positives and false negatives) occur.

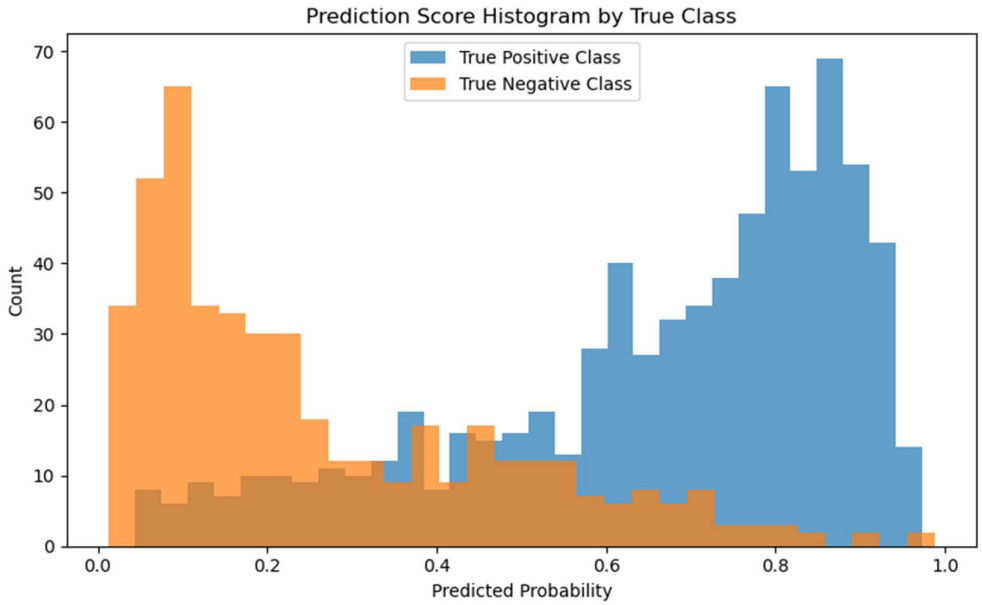

Fig. 16. Prediction Score Histogram by True Class and ROC curve

A prediction score histogram by true class is a visualization that shows the distribution of model-predicted probabilities (or scores) for each actual class label in your dataset (e.g., high vs. low binding affinity). This plot is crucial for understanding and diagnosing the behavior of classification models, especially in scientific and drug discovery contexts.

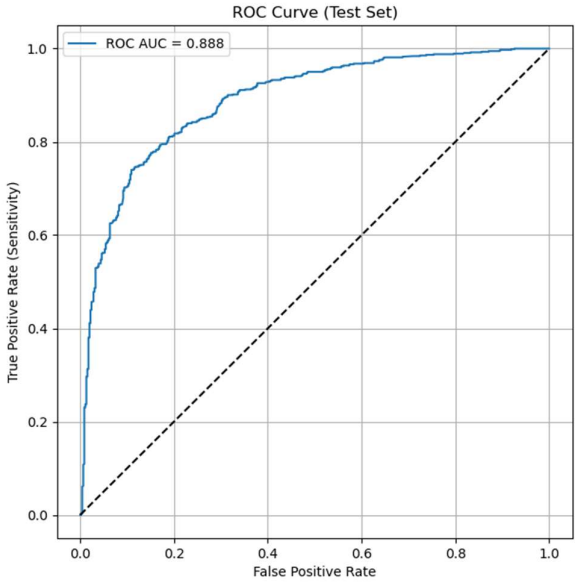

Fig. 17. ROC curve

The ROC curve is well above the diagonal, confirming strong discriminative ability. The model performs much better than random guessing, with a high true positive rate and low false positive rate across thresholds.

3. Confusion matrix

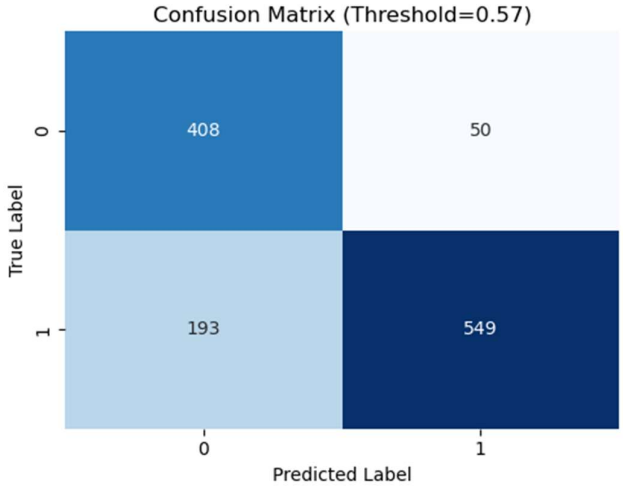

Fig. 18. Confusion Matrix

The confusion matrix at a threshold of 0.57 shows that out of 1,200 molecules, the cascade neural network model correctly classified 408 low-affinity (true negatives) and 549 high-affinity (true positives) compounds, while misclassifying 50 low-affinity compounds as high-affinity (false positives) and 193 high-affinity compounds as low-affinity (false negatives). This results in an overall accuracy of about 79.7%, with high precision (91.6%) indicating that most predicted actives are truly active, and good specificity (89.1%) showing effective filtering of inactives.
However, the recall (74%) reveals that about a quarter of actual high-affinity molecules are missed, suggesting the model is conservative in its predictions-favoring fewer false positives at the cost of more false negatives. This balance is often desirable in drug discovery to minimize wasted resources on false leads, but may require further tuning if maximizing the identification of all true actives is critical.

4. Ranking Correlation (Spearman/Kendall)

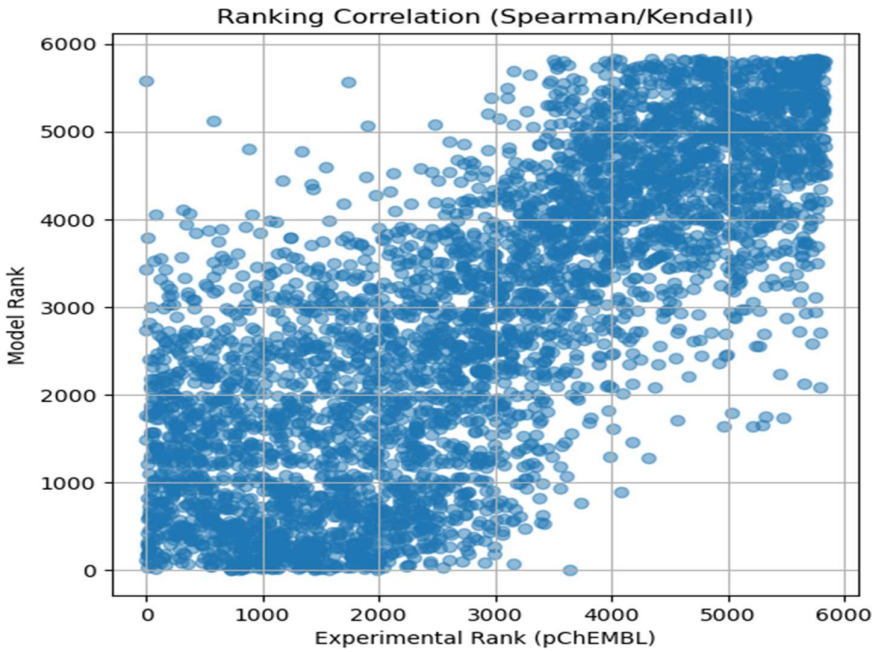

Fig. 18. Ranking Correlation

Scatter plot of experimental vs. model ranks. The positive trend indicates a strong monotonic relationship between predicted and experimental rankings. This supports the high Spearman (0.780) and Kendall (0.565) rank correlation coefficients reported, demonstrating the model's effectiveness at ranking molecules by affinity. Excellent classification and ranking: High AUCs for both ROC and PR curves, with well-separated prediction scores for each class. Model is well-calibrated: Most predictions are confidently assigned to the correct class.Strong monotonic relationship between predicted and experimental rankings. Misclassifications occur primarily in the overlap region of the prediction score histogram.

5. CONCLUSION AND FUTURE ASPECTS

This research presents an integrated computational framework for investigating the structural and functional implications of the Dopamine D3 receptor (D3R), with a focus on mutation impact, ligand binding, and machine learning-based drug discovery. The study systematically combined evolutionary conservation analysis, deleterious mutation prediction, molecular docking, molecular dynamics (MD) simulations, and advanced QSAR modeling using cascade neural networks to address challenges in D3R-targeted drug discovery. Conservation analysis and mutation prediction revealed that deleterious variants are predominantly located in highly conserved regions of D3R, which are critical for receptor stability and function. Docking studies demonstrated that mutations in these regions often lead to reduced binding affinity for dopamine, while mutations in more variable regions are generally tolerated. MD simulations confirmed that such mutations can alter receptor dynamics, with some stabilizing the structure but potentially impairing functional flexibility.

The cascade neural network QSAR model, trained on a large and chemically diverse dataset, achieved robust performance in classifying high- and low-affinity ligands. The model demonstrated high accuracy, precision, recall, ROC-AUC (0.888), and average precision (AUPRC 0.916), as visualized in the included ROC and precision-recall curves. Prediction score histograms and rank correlation plots further validated the model's ability to distinguish and rank ligand affinities effectively. Misclassifications were mainly observed in molecules with borderline prediction scores, highlighting areas for further model refinement.The developed pipeline is adaptable and can be applied to other protein-ligand systems, provided sufficient binding data is available. The approach supports virtual screening, lead optimization, and the identification of candidate molecules for experimental validation, thereby accelerating the early stages of drug discovery. Given the involvement of D3R in neuropsychiatric and neurodegenerative disorders, this work advances the understanding of how genetic variability influences drug response and receptor function.

The framework supports the rational design of selective D3R modulators, with potential applications in treating conditions such as Parkinson's disease, schizophrenia, and depression.

While cascade neural networks offer high predictive power, their interpretability remains limited compared to simpler models. Future work should focus on integrating explainable AI techniques to enhance model transparency. The accuracy of machine learning models depends on the quality and quantity of available data. Expanding curated datasets, especially for rare variants and experimentally validated affinities, will further improve model reliability. Bridging computational predictions with experimental assays remains essential to ensure clinical translatability. The pipeline should be iteratively refined with feedback from in vitro and in vivo studies. Incorporating multi-omics data and systems biology approaches could provide a more holistic understanding of D3R function and its role in disease.

6. REFERENCES

[1] Wang D, Ji X, Liu J, Li Z, Zhang X. Dopamine Receptor Subtypes Differentially Regulate Autophagy. Int J Mol Sci. 2018 May 22;19(5):1540. doi: 10.3390/ijms19051540. PMID: 29786666; PMCID: PMC5983733.

[2] D. Accili, C. S. Fishburn, J. Drago, H. Steiner, J. E. Lachowicz, B. H. Park, E. B. Gauda et al., "A targeted mutation of the D3 dopamine receptor gene is associated with hyperactivity in mice," Proc. Natl. Acad. Sci. U.S.A., vol. 93, no. 5, pp. 1945–1949, 1996.

[3] J. H. Seo and E. V. Kuzhikandathil, "Dopamine D3 receptor mediates preadolescent stress-induced adult psychiatric disorders," PLoS One, vol. 10, no. 11, p. e0143908, Nov. 2015, doi: 10.1371/journal.pone.0143908.

[4] J. Gonc ¸alves, A. Mendes, J. Dam´ asio, N. Vila-Ch˜a, D. Boleixa, B. Leal, and S. Cavaco, "DRD3 predicts cognitive impairment and anxiety in Parkinson's disease: Susceptibility and protective effects," J. Parkinsons Dis., vol. 14, no. 2, pp. 313–324, 2024, doi: 10.3233/JPD-230292.

[5] Gupta A, Zhou HX. Machine Learning-Enabled Pipeline for Large-Scale Virtual Drug Screening. J Chem Inf Model. 2021 Sep 27;61(9):4236-4244. doi: 10.1021/acs.jcim.1c00710. Epub 2021 Aug 17. PMID: 34399578; PMCID: PMC8478848.

[6] Bono F, Mutti V, Fiorentini C, Missale C. Dopamine D3 receptor Heteromerization: Implications for Neuroplasticity and Neuroprotection. Biomolecules. 2020 Jul 9;10(7):1016. doi: 10.3390/biom10071016. PMID: 32659920; PMCID: PMC7407647.

[7] Maramai S, Gemma S, Brogi S, Campiani G, Butini S, Stark H, Brindisi M. Dopamine D3 receptor Antagonists as Potential Therapeutics for the Treatment of Neurological Diseases. Front Neurosci. 2016 Oct 5;10:451. doi: 10.3389/fnins.2016.00451. PMID: 27761108; PMCID: PMC5050208.

[8] De Deurwaerdère P, Chagraoui A. Dopamine D3 receptor: Contemporary Views of Its Function and Pharmacology for Neuropsychiatric Diseases. Biomolecules. 2021 May 11;11(5):713. doi: 10.3390/biom11050713. PMID: 34064602; PMCID: PMC8151522.

[9] UniProt Consortium, "D(3) dopamine receptor (DRD3), Homo sapiens (Human)," UniProt, P35462. [Online]. Available: https://www.uniprot.org/uniprotkb/P35462/entry. Accessed: May 13, 2025.

[10] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, "PubChem 2025 update," Nucleic Acids Research, vol. 53, no. D1, pp. D1516–D1525, Jan. 2025, doi: 10.1093/nar/gkae1059.

[11] T. Sterling and J. J. Irwin, "ZINC 15 – Ligand Discovery for Everyone," J. Chem. Inf. Model., vol. 55, no. 11, pp. 2324–2337, 2015, doi: 10.1021/acs.jcim.5b00559.

[12] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX, vol. 1–2, pp. 19–25, 2015, doi: 10.1016/j.softx.2015.06.001

[13] X. Li et al., "A Cascade Graph Convolutional Network for Predicting Protein–Ligand Binding Affinity," Front. Chem., vol. 9, 2021, Art. no. 8070477, doi: 10.3389/fchem.2021.852893.

[14] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," Proc. Int. Conf. Mach. Learn., pp. 448–456, 2015.

[15] N. Srivastava et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," J. Mach. Learn. Res., vol. 15, no. 1, pp. 1929–1958, 2014.

[16] S. D. Axen et al., "A simple representation of three-dimensional molecular structure," J. Med. Chem., vol. 60, no. 17, pp. 7393–7409, 2017, doi: 10.1021/acs.jmedchem.7b00696.
For conformer statistics, surface area, TPSA, and ring properties.

[18] H. Kwon et al., "Protein-ligand binding affinity prediction using multi-instance learning," Front. Pharmacol., vol. 15, 2024, Art. no. 1518875, doi: 10.3389/fphar.2024.1518875.

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, New Delhi, 110042

PLAGIARISM VERIFICATION

Title of the Thesis **"Integrative Computational Framework for Dopamine D3 receptor-Targeted Drug Discovery: Bridging Genetic Variability, Structural Dynamics, and Machine Learning"** Total Pages **66** Name of the Scholar **Akshay Hatwal (23/BIO/07).**

Supervisor

Prof. Pravir Kumar

Department of Biotechnology

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: **Turnitin,** Similarity Index: **5%**, Total Word Count: **6,361**

Date: _____

Candidate's Signature **Signature of Supervisor**

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

Match Groups

17 Not Cited or Quoted 5%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
1% Publications
3% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.