학사학위 청구논문
지도교수 이진용

Discovery of Candidate Inhibitors for LRRK2 through Virtual Screening

성균관대학교 자연과학대학
화학과
이가영

학사학위 청구논문
지도교수 이진용

Discovery of Candidate Inhibitors for LRRK2 through Virtual Screening

이 논문을 이학 학사학위 청구논문으로 제출합니다.
2024 년 12 월 16 일

성균관대학교 자연과학대학
화학과
이가영

이 논문을 이가영의 이학 학사학위 논문으로 인정함.

2024 년 12 월 16 일

지도교수 이진용 (인)

심사위원 팽기욱 (인)

Abstract

Parkinson's disease (PD), the second most common neurodegenerative disorder, affects over 10 million people worldwide. Leucine-rich repeat kinase 2 (LRRK2), a critical regulator of cellular pathways such as ciliogenesis, mitophagy, and mitochondrial homeostasis, has emerged as a promising drug target for PD treatment. This study employed a two-step virtual screening approach—ligand-based and structure-based methods—to identify potential type I inhibitors targeting LRRK2's ATP binding site. A QSAR model trained on bioactivity data from the ChEMBL database was used to predict the pIC50 values of druggable compounds from the ZINC15 database, resulting in the selection of 10 candidates that satisfied criteria for drug-likeness and synthetic accessibility. The selected compounds were further analyzed using AutoDock Vina against both wild-type LRRK2 (8FO7) and its G2019S-mutated form (7LI3). Candidate 3 (ZINC00029986675) exhibited the strongest binding affinity among the candidates, with -9.8 kcal/mol for 8FO7 and -9.0 kcal/mol for 7LI3, and demonstrated favorable overlap with the reference inhibitor LRRK2-IN-1. Docking results revealed that stereoisomer pairs adopted similar poses on 8FO7 but diverged significantly on 7LI3, emphasizing the mutation's impact on ligand binding. Despite Candidate 3's promising binding properties, further optimization is required to enhance its pharmacokinetic profile and efficacy. This study highlights the effectiveness of integrating ligand-based and structure-based approaches in drug discovery and proposes Candidate 3 as a promising lead for developing selective LRRK2 inhibitors to address unmet therapeutic needs in PD.

The GitHub repository (https://github.com/ella0000/Undergrad-Thesis.git) provides all computational workflows used in this study, including data preprocessing, QSAR model development, virtual screening processes, and visualization scripts for structure-based virtual screening.

1. Introduction

Parkinson's disease(PD) is one of the most prevalent neurodegenerative diseases globally, second only to Alzheimer's disease, affecting over 10 million people worldwide [1][2]. Leucine-rich repeat kinase 2 (LRRK2) has emerged as a promising drug target for PD treatment due to its pivotal role in various cellular signaling pathways, including ciliogenesis, mitophagy, autophagy, and mitochondrial homeostasis. LRRK2 is a 286-kDa protein with seven domains (ARM, ANK, LRR, ROC, COR, KIN, and WD40) exhibiting kinase and GTPase activities [3][4]. Autosomal missense mutations in LRRK2 are a significant cause of familial PD and are also linked to sporadic cases [5]. Notably, all PD-associated mutations increase LRRK2 kinase activity, and even elevated kinase activity in wild-type (WT) LRRK2 has been implicated in sporadic PD [6]. The GTPase and kinase functions of LRRK2 are essential for regulating its cellular signaling pathways and may reciprocally influence each other to determine its function [7][8]. As a result, developing

selective LRRK2 inhibitors has become a key strategy for addressing LRRK2-related PD. Although current approaches, including ATP-competitive type I and non-competitive type II inhibitors, show promise [9], no FDA-approved LRRK2 inhibitors are commercially available. Only a few candidates, such as DNL201 and DNL151, are undergoing clinical trials, highlighting the urgent need for novel LRRK2 inhibitors [10][11][12][13].

Among the various mutations, the glycine-to-serine substitution (G2019S) within the kinase domain encoded by exon 41 [14], is the most prevalent and clinically significant. This mutation leads to an aberrant increase in kinase function [15] and accounts for 5–6% of autosomal-dominant PD cases and approximately 1% of sporadic late-onset PD cases. Patients with the G2019S mutation often exhibit Lewy bodies, suggesting that these mutations may contribute to disease through distinct pathogenic mechanisms [7]. Given that most PD patients with the LRRK2-G2019S variant are heterozygous, a precision medicine approach that selectively inhibits the pathogenic activity while preserving normal physiological LRRK2 function could provide both efficacy and safety benefits for this patient population [16].

To accelerate the discovery of selective LRRK2 inhibitors, computational methods such as Computer-Aided Drug Design (CADD) have become essential tools in modern drug discovery. CADD integrates a variety of computational tools to identify and optimize potential drug candidates [17]. Among these, Virtual Screening (VS) is a widely used technique to select suitable compounds for testing from large compound libraries. This process involves screening compounds either against a specific drug target (structure-based virtual screening) or a pharmacophore model (ligand-based virtual screening) and ranking them based on binding energy or match scores. The combination of both approaches has demonstrated good accuracy in identifying lead molecules [18]. Top-ranked compounds are subsequently analyzed and selected for experimental validation. By streamlining the identification of hit and lead compounds, virtual screening significantly reduces the time and resources required in the drug discovery process [19].

Ligand-based methods are commonly used when the X-ray structure of the target receptor is unavailable. These methods use a set of compounds known to bind to the target or to be active in functional assays as templates to identify similar compounds within a large virtual library. The similarity is generally assessed based on 2D or 3D molecular representations [20][21], and the approach requires two critical compounds: an efficient similarity measure and a reliable scoring method. Furthermore, the computational framework must efficiently screen potential ligands with reasonable accuracy and speed [22]. These methods identify patterns of distance between specific molecular features, such as aromatic systems or hydrogen bond acceptors/donors [23], and evaluate similarity by comparing these patterns. Among the virtual screening approaches, the quantitative structure-activity relationships (QSAR) analysis is particularly effective due to its high throughput, speed, and good hit rate. The QSAR model development begins with the collection of chemogenomic data from databases and literature, followed by the calculation of chemical descriptors that represent different levels of molecular structure. These descriptors are then correlated with the biological activity using machine learning techniques. Once the models are validated, QSAR models can be used to predict the

biological properties of novel compounds [24].

Structure-based approaches, primarily classical docking methods, are typically employed when the 3D structure of the target is available [25]. These methods utilize 3D structural information obtained from X-ray crystallography, NMR spectroscopy, or computational modeling to dock a collection of chemical compounds into the desired binding site of the target. Based on the predicted binding scores, a subset of these compounds is selected for further biological evaluation [26]. Molecular docking is a computational procedure designed to predict noncovalent interactions between a macromolecule (receptor) and a small molecule (ligand). It starts with the unbound structures of the interacting molecules, which are obtained from experimental data or homology modeling. The primary objectives of molecular docking are to predict the bound conformations of the ligand and to estimate its binding affinity [27]. AutoDock Vina is a widely used open-source software for molecular docking [28]. The program employs an Iterated Local Search Global optimizer and dynamically determines the number of steps required during the run time. It uses the PDBQT format, which includes information on the torsional degrees of freedom [29]. Based on the input data, AutoDock Vina explores the conformational space of the ligand, aiming to identify the most stable binding pose within the receptor's binding site [30].

2. Background and Literature Review

Type I inhibitors bind to the active form of LRRK2 by competing with ATP for the ATP-binding pocket, interacting with specific amino acid residues [31]. The first selective LRRK2 inhibitor was LRRK2-IN-1 [32], a type I inhibitor with IC50 values of 6nM for G2019-mutated LRRK2 and 13nM for wild-type LRRK2 [32][33], corresponding to pIC50 values of approximately –20.7 and –18.2, respectively. However, its inability to penetrate the blood-brain barrier has limited its use in vivo studies. HG-10-102-01, another type I inhibitor was designed to overcome this limitation as a brain-penetrant compound [34]. In 2015, MLi-2 was introduced as a highly selective and high-affinity LRRK2 inhibitor and has since been regarded as the gold standard among LRRK2 inhibitors. It has been widely used in both preclinical and clinical studies. PFE-360, another type I inhibitor, also demonstrates excellent affinity and selectivity for LRRK2. Additionally, DNL201 (previously GNE0877), which is currently in clinical trials [35], has shown effective LRRK2 inhibition in PD patients. A related compound, DNL151 (now BIIB122) [36], has advanced to phase 3 clinical trials, marking significant progress in the development of type I inhibitors.

Type II inhibitors are non-competitive with ATP and stabilize the kinase in an open, inactive conformation [37] by extending into the back pocket of the kinase active site and disrupting a critical salt bridge, which releases the regulatory αC helix [38]. Several broad-spectrum kinase inhibitors, such as ponatinib, GZD-824, and rebastinib, have demonstrated high-affinity binding to LRRK2's kinase domain [9]. However, until recently, there were no highly selective type II LRRK2 inhibitors available for studying the effects of targeting LRRK2's inactive conformation. In September 2024, Nicolai D. Raig and colleagues reported the discovery of LRRK2-selective type II kinase inhibitors, named RN277 and RN341, which act as cellular tools

targeting the inactive state of LRRK2 [15]. This study has not been certified by peer review yet.

Recognizing the need for further investigation into LRRK2 inhibitors, I conducted a two-step virtual screening process to identify potential candidates. This included ligand-based virtual screening using a QSAR model and structure-based virtual screening with AutoDock Vina, focusing on type I inhibition. Due to the challenges in assessing type II inhibition, the study prioritized type I inhibitors, which competitively block ATP binding at the ATP-binding site. Existing studies have primarily focused on experimentally validated inhibitors or broad computational approaches without systematically integrating ligand-based and structure-based techniques. By combining these complementary methods, my research aims to provide a more comprehensive framework for identifying selective and effective LRRK2 inhibitors. The QSAR model was trained using LRRK2 target data sourced from the ChEMBL database. Once the model was developed, it was applied to screen druggable compounds from the ZINC database by predicting their pIC50 values. From this screening, 10 final candidates were selected based on specific criteria. These candidates were then subjected to structure-based virtual screening using AutoDock Vina. Their binding affinities were compared to those of a known type I inhibitor, LRRK2-IN-1, on both wild-type LRRK2 (PDB ID: 8FO7) and G2019S-mutated LRRK2 (PDB ID: 7LI3).

3. Materials and Methods

3-1. Ligand-Based Virtual Screening

To build the QSAR model, bioactivity data targeting LRRK2 (ChEMBL ID: 1075104) was downloaded from the ChEMBL database. Compounds with IC50 values were selected to ensure quantitative bioactivity data for model training, and the pIC50 values were labeled as pChEMBL Values. A total of 2,638 compounds were initially selected. Using Jupyter Notebook, only the "SMILES" and "pChEMBL Value" columns were retained. Entries with missing pChEMBL Values or duplicate SMILES-pChEMBL combinations were removed using the df.drop_duplicates() function. For compounds with identical SMILES but differing pChEMBL Values, the values were averaged to create a single entry. Finally, the pChEMBL Value column was renamed to pIC50, resulting in a dataset of 1,426 entries prepared for training the QSAR model.

A Random Forest Regression model was selected due to its balance between computational efficiency and predictive accuracy. Molecular descriptors were calculated from the SMILES strings using RDKit. The selected descriptors included Molecular Weight (MW), Hydrophobicity (LogP), Hydrogen Bond Acceptors (HBA), Hydrogen Bond Donors (HBD), Fraction of sp3 Carbon (CSP3), Number of Rotatable Bonds (NumRotBond), Number of Rings (NumRings), Topological Polar Surface Area (TPSA), and Number of Aromatic Rings (NumAromaticRings). These descriptors served as features, while the target variable was the pIC50 value. The distribution of pIC50 values is shown in Figure 1(Fig 1).

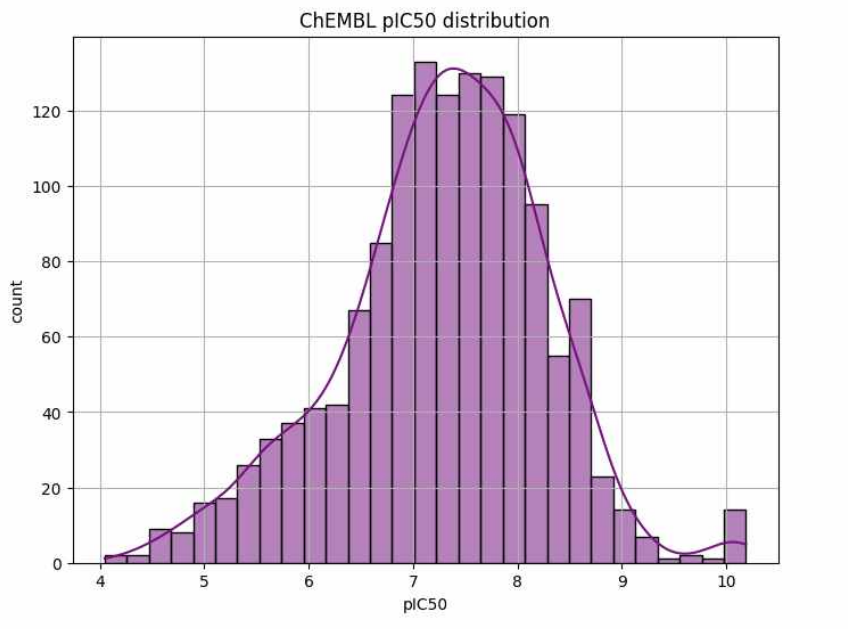

Fig 1. Distribution of pIC50 Value in the Training Dataset

The dataset was split into training and test sets, with 10% of the data (143 entries) allocated for testing and the remaining 1,283 entries used for training. The model was implemented using scikit-learn's RandomForestRegressor with default parameters, as hyperparameter tuning showed no significant improvement in performance. Mean Squared Error (MSE) was used as the loss function, and the model was trained to minimize this metric.

Subsequently, a total of 46,542 druggable compounds from the ZINC15 database were screened using the trained QSAR model. The ZINC data, provided as SMILES strings, were processed into molecular descriptors using the same RDKit functions applied during model training. The model then predicted the pIC50 values for these compounds. Based on these predictions, the compounds were ranked by their pIC50 values, and the top 100 candidates were selected for further evaluation. To ensure the novelty of the selected compounds, structural similarity to the training data was assessed using Tanimoto similarity. Compounds with a Tanimoto similarity score above 0.9 to any entry in the training dataset were excluded. Additionally, Quantitative Estimate of Drug-likeness (QED) and Synthetic Accessibility (SA) Scores were calculated for each compound. Candidates with a QED score greater than 0.4 and an SA Score below 4.0 were retained. Ultimately, 10 compounds that satisfied all these criteria were selected as final candidates for further structure-based virtual screening. Furthermore, the feature importance of the trained QSAR model was calculated and compared to the feature distribution of the top 100 candidates to analyze the contributions of individual molecular descriptors in determining pIC50 predictions.

3-2. Structure-Based Virtual Screening

Structure-based virtual screening was performed using AutoDock Vina. Input files

required for docking, in PDBQT format, were prepared using RDKit and the AutoDock Tool. Two receptors were selected from the Protein Data Bank (PDB) for the docking process. The first receptor was the wild-type LRRK2 (PDB ID: 8FO7), a Cryo-EM structure of LRRK2 bound to the type I inhibitor LRRK2-IN-1 (Fig 2). This protein was selected to evaluate the binding affinity of candidate compounds in comparison to the known type I inhibitor, LRRK2-IN-1.

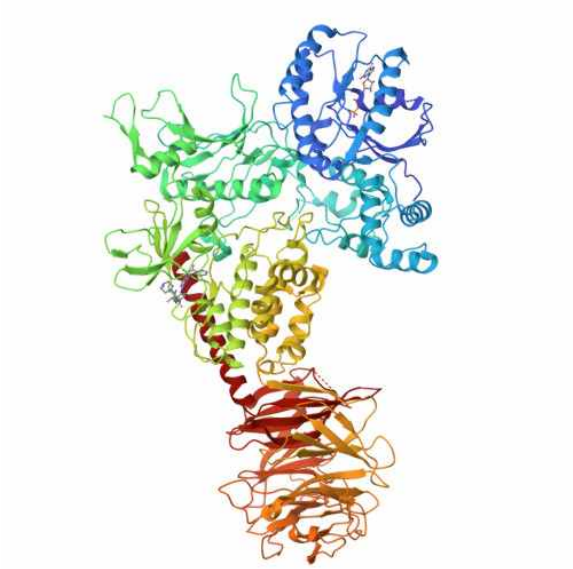

Fig 2. Structure of 8FO7

The second receptor was the G2019S-mutated LRRK2 structure (PDB ID: 7LI3, Fig 3), which represents the most prevalent LRRK2 mutation associated with Parkinson's disease. This receptor was chosen to assess the binding affinity of candidates and LRRK2-IN-1 against the mutant protein.

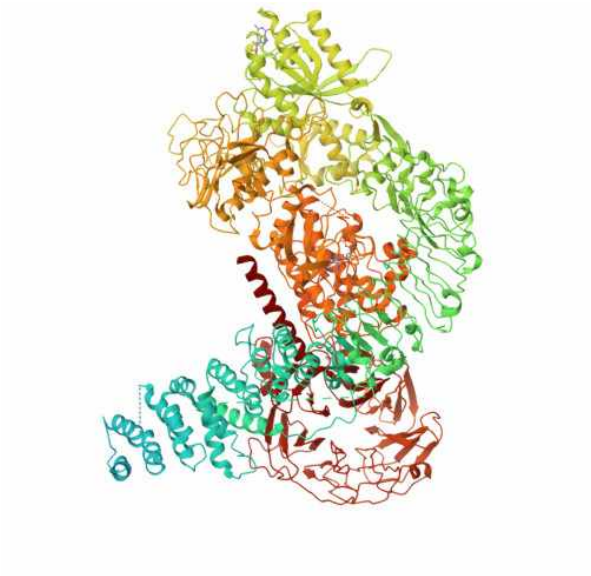

Fig 3. Structure of 7LI3

The PDB files for both receptors were downloaded from the Protein Data Bank. Using Jupyter Notebook, the files were processed to retain only the main protein atoms, with all other ligands and non-essential molecules removed. The resulting files were saved as "8FO7_prot.pdb" and "7LI3_prot.pdb," respectively. Subsequently, using the AutoDock tools, water molecules were removed, polar hydrogens were added to account for potential hydrogen bonding in docking simulations, and Kollman charges were calculated and applied to facilitate electrostatic interaction calculations. The processed files were then converted to PDBQT format.

Ligands were prepared using RDKit and OpenBabel software. The final 10 candidate compounds, initially in SMILES format, were processed using RDKit to add hydrogens and convert them sequentially to MOL, SDF, and PDB formats. Subsequently, the PDB files were converted to PDBQT format using OpenBabel. Additionally, the SDF file for the reference ligand LRRK2-IN-1 (labeled as 4K4) was downloaded from the Protein Data Bank and processed using the same workflow. The resulting ligand files were named "candid1.pdbqt" to "candid10.pdbqt," and the reference ligand as "LRRK2-IN-1.pdbqt."

The grid box for docking was identified using custom Python scripts in Jupyter Notebook. The original PDB files (8FO7 and 7LI3) were analyzed, identifying atoms labeled under 'HETATM' that correspond to the ligands 4K4 (for 8FO7) and ATP (for 7LI3). The x, y, and z coordinates of these atoms were extracted and stored in a NumPy array. The geometric center of the grid box was calculated by applying the mean function to these coordinates. The grid box size was set to 40 × 40 × 40 for 8FO7 and 30 × 30 × 30 for 7LI3. The smaller grid size for 7LI3 was chosen because its binding site is located internally within the protein, whereas the binding site in 8FO7 is exposed on the protein's surface (Fig 4).

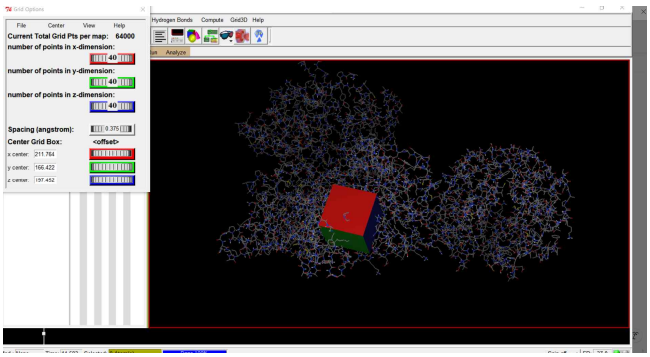

Fig 4 a. Grid box of 8FO7

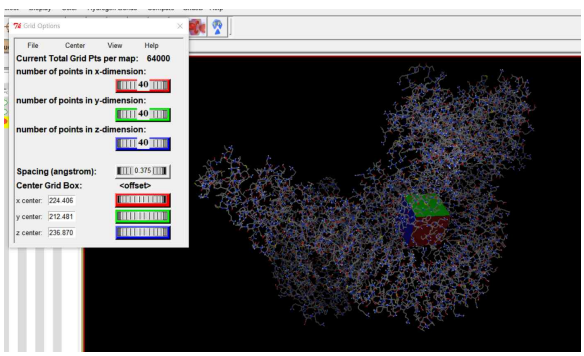

Fig 4 b. Grid box of 7LI3

The energy range for the docking simulations was set to 4, with all other parameters left at their default values. Docking was conducted using a shell script provided by the AutoDock Vina developers. For each ligand input, AutoDock Vina generated nine binding poses, each with a distinct binding energy value. The pose with the lowest (most negative) binding energy, reported in kcal/mol, was selected as the best binding pose. The binding affinities and poses of the candidate ligands were compared to each other and the reference ligand LRRK2-IN-1.

4. Results

4-1. Ligand-Based Virtual Screening Results

The trained QSAR model was evaluated against the test dataset, and the results are shown in Figure 5 (Fig 5). The model's performance was assessed using two key metrics: the coefficient of determination (R^2) and the root mean square error (RMSE). The R^2 score was 0.312, indicating that the model explains approximately 31.2% of the variance in the observed data. This result suggests that, while the model captured some underlying patterns, a substantial portion of the variance remained unexplained, indicating room for improvement in predictive accuracy. The RMSE was 0.781, reflecting the average magnitude of the model's prediction error. Considering the observed y-values, which ranged from 4.045 to 10.19 with a standard deviation of 0.980, the RMSE accounted for approximately 79.7% of the standard deviation. This result implies that the model's predictions deviated from the actual values by a margin comparable to the natural variability of the data. In summary, the results indicate that the model provided a moderate level of predictive accuracy.

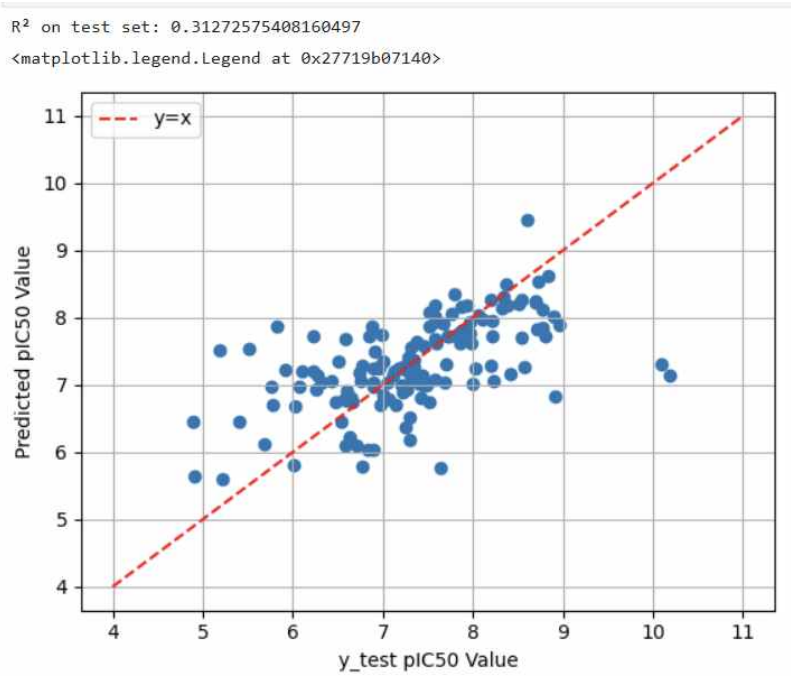

Fig 5. Correlation Between y_pred and y_test for the Test Set

The feature importance of the QSAR model was analyzed, and the results are presented in Figure 6 (Fig 6). Among the molecular descriptors, Molecular Weight (MW) demonstrated the highest importance, with a value of 0.276876, followed by the Fraction of sp3 Carbon (CSP3) at 0.236692, and LogP, which ranked third with an importance value of 0.141767. The distribution of these three key features in the ChEMBL dataset is shown in Figure 7 (Fig 7). Additionally, the distributions of MW, CSP3, and LogP for the top 100 candidate compounds identified during virtual screening are presented in Figure 8 (Fig 8).

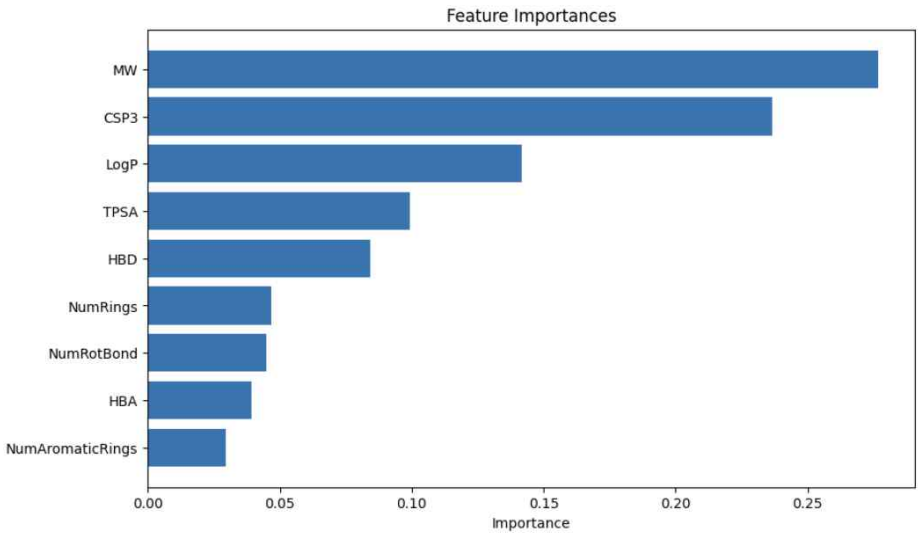

Fig 6. Feature Importance of the QSAR model

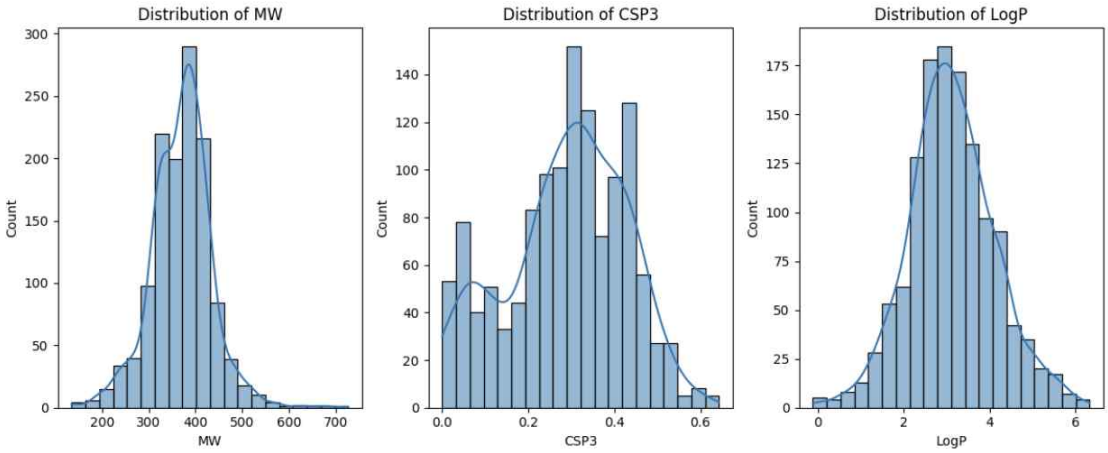

Fig 7. Distribution of MW, CSP3, and LogP for Training Dataset

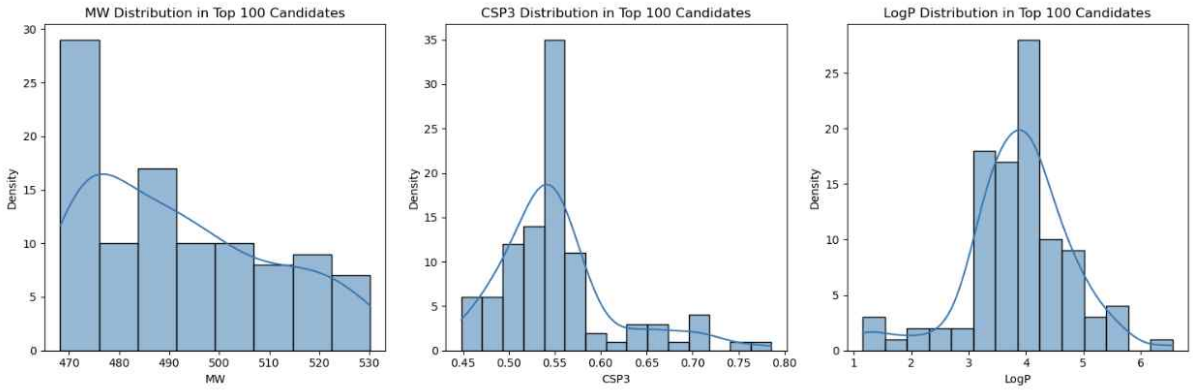

Fig 8. Distribution of MW, CSP3, and LogP for Top 100 Candidates

The final 10 candidates selected through ligand-based virtual screening are illustrated in Figure 9 (Fig 9). The ZINC IDs of the candidates, along with their predicted pIC50 values, Quantitative Estimate of Drug-likeness (QED), and Synthetic Accessibility (SA) scores, are presented in Table 1 (Table 1). Additionally, the Tanimoto similarity of the 2D structures of the candidates was analyzed, and the results are visualized as a heatmap in Figure 10 (Fig 10). As shown in Figure 10, Candidate 1 and Candidate 2, Candidate 4 and Candidate 5, and Candidate 7 and Candidate 8 are stereoisomers, differing by only one stereochemical configuration within each pair. These three pairs also exhibited approximately 0.60 structural similarity to one another.

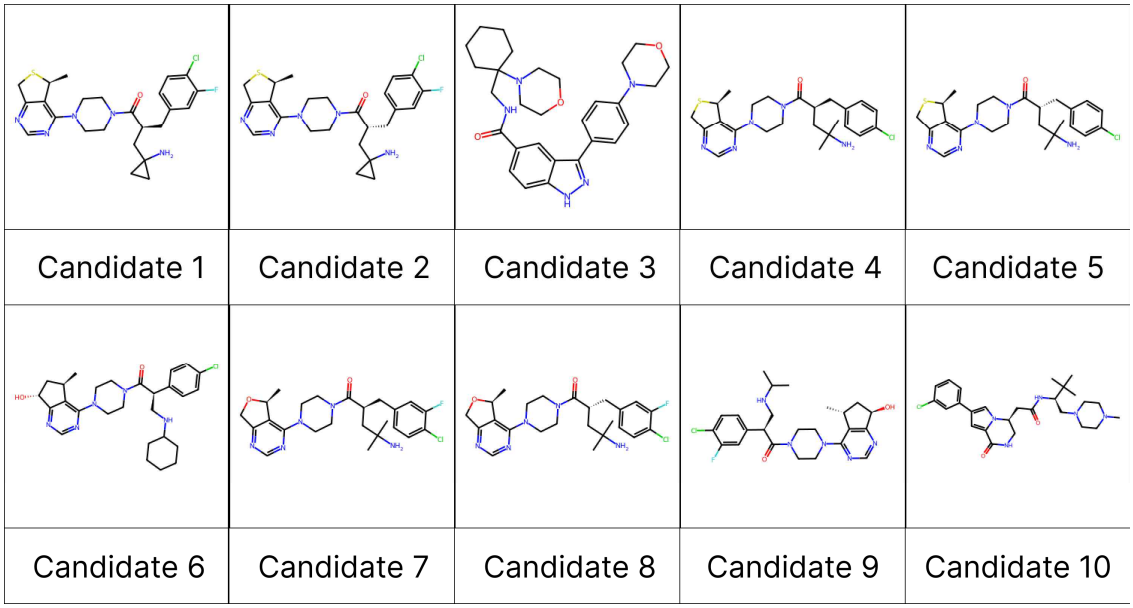

Fig 9. Structure of Top 10 Candidates

Candidates	ZINC ID	Predicted pIC50	QED	SA Score
Candidate 1	ZINC000066260034	9.88565	0.66145301	3.84582978
Candidate 2	ZINC000066260035	9.88565	0.66145301	3.84582978
Candidate 3	ZINC000299866759	9.872	0.53214022	2.73120198
Candidate 4	ZINC000066058891	9.80895	0.68104035	3.641400403
Candidate 5	ZINC000066058892	9.80895	0.68104035	3.641400403
Candidate 6	ZINC000095580355	9.8061	0.62686546	3.6475691
Candidate 7	ZINC000066261947	9.7881	0.68799066	3.779959438
Candidate 8	ZINC000066261949	9.7881	0.68799066	3.779959438
Candidate 9	ZINC000095578545	9.7659	0.6671862	3.746955381
Candidate 10	ZINC000169325511	9.7299	0.65857502	3.714633032

Table 1. ZINC IDs, Predicted pIC50 Values, QED, and SA Scores of the Final 10 Candidates

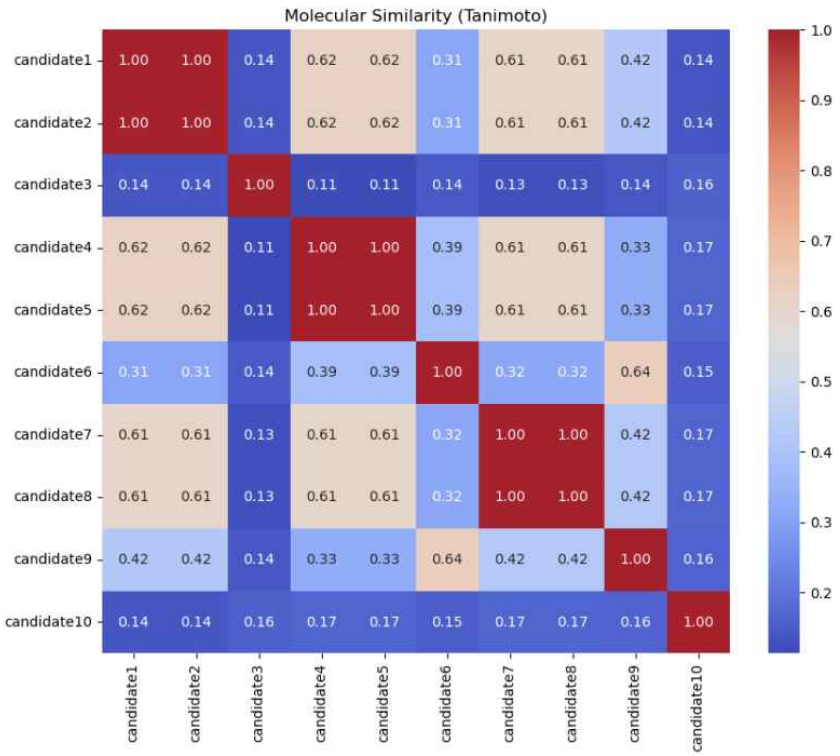

Fig 10. Heatmap of Tanimoto Similarity Among the Final 10 Candidates

The structure of LRRK2-IN-1 is illustrated in Figure 11 (Fig 11). The Tanimoto similarity values of the final 10 candidates to LRRK2-IN-1, along with Molecular Weight (MW), Fraction of sp3 Carbon (CSP3), LogP, and their bioactivities based on ChEMBL 20 data are summarized in Table 2 (Table 2).

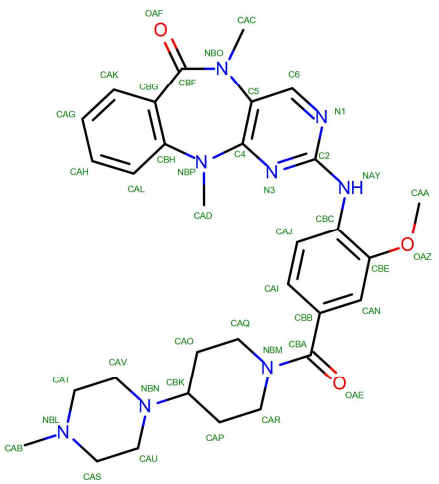

Fig 11. Structure of LRRK2-IN-1

Candidates	Tanimoto Similarity	MW	CSP3	LogP	Bioactivity (pKi)
Candidate 1	0.220588235	490.048	0.54	3.966	AKT 1 (8.52)
Candidate 2	0.220588235	490.048	0.54	3.966	AKT 1 (8.52)
Candidate 3	0.358333333	503.647	0.52	3.831	KDR (5.42), AURKA (5.62), TTK (8.11)
Candidate 4	0.209677419	474.074	0.54	4.073	AKT 1 (8.22)
Candidate 5	0.209677419	474.074	0.54	4.073	AKT 1 (8.22)
Candidate 6	0.233870968	498.071	0.59	4.025	AKT 1 (8.05), AKT 2 (7.74), AKT 3 (8.15)
Candidate 7	0.196969697	475.996	0.54	3.495	AKT 1 (7.74)
Candidate 8	0.196969697	475.996	0.54	3.495	AKT 1 (7.74)
Candidate 9	0.218045113	475.996	0.54	3.24	AKT 1 (8.52), AKT 2 (7.77), AKT 3 (8.15)
Candidate 10	0.2109375	486.06	0.54	3.261	PIM 1 (8.30), PIM 2 (6.58)

Table 2. Tanimoto Similarity to LRRK2-IN-1, Bioactivities, and Molecular Properties of the Final 10 Candidates

4-2. Structure-Based Virtual Screening Results

The calculated binding affinities of the final candidates and LRRK2-IN-1 against 8FO7 and 7LI3 using AutoDock Vina are summarized in Table 3 (Table 3). Binding affinities are considered strong if they are less than -9 kcal/mol and weak if they are greater than –7 kcal/mol [39]. For 8FO7, the strongest binding affinity was observed with the original type I inhibitor, LRRK2-IN-1 (4K4), which showed an affinity of -10.4 kcal/mol. The second strongest binding was exhibited by Candidate 3 with -9.8 kcal/mol. The remaining candidates displayed binding affinities ranging from -7 to -9 kcal/mol, indicating moderate binding.

Ligand	8FO7 (Wild-Type LRRK2)	7LI3 (G2019S-mutant LRRK2)
LRRK2-IN-1	-10.4 kcal/mol	-9.5 kcal/mol
Candidate 1	-8.4 kcal/mol	-8.0 kcal/mol
Candidate 2	-8.3 kcal/mol	-7.9 kcal/mol
Candidate 3	-9.8 kcal/mol	-9.0 kcal/mol
Candidate 4	-7.2 kcal/mol	-7.4 kcal/mol
Candidate 5	-7.3 kcal/mol	-7.7 kcal/mol
Candidate 6	-8.2 kcal/mol	-8.0 kcal/mol
Candidate 7	-8.2 kcal/mol	-7.6 kcal/mol
Candidate 8	-8.0 kcal/mol	-7.6 kcal/mol
Candidate 9	-8.1 kcal/mol	-7.8 kcal/mol
Candidate 10	-8.7 kcal/mol	-8.7 kcal/mol

Table 3. Binding Affinity Results

For 7LI3, the results followed a similar trend. LRRK2-IN-1 exhibited the strongest binding affinity, followed by Candidate 3. The other candidates demonstrated binding affinities between -7 and -9 kcal/mol; however, the binding affinities were generally weaker compared to those for 8FO7. Notably, Candidate 4 and Candidate 5 showed slight increases in binding affinity, from -7.2 to -7.4 kcal/mol and from -7.3 to -7.7 kcal/mol, respectively.

The conformations of the candidates and LRRK2-IN-1 were analyzed using py3Dmol in Jupyter Notebook. First, the differences in docking poses between stereoisomer pairs on 8FO7 were examined (Fig 12). Each pair displayed similar conformations within the 8FO7 grid box, with pyrimidine groups located on one side and benzene rings on the opposite side. However, despite Candidate 4 and Candidate 7 sharing a Tanimoto similarity of 0.61, their binding poses differed significantly, as the pyrimidine groups were located on opposite sides of the binding site (Fig 13). Additionally, the docking pose of Candidate 3 was compared with that of LRRK2-IN-1 (Fig 14). The analysis revealed an overlap between the two aromatic rings of LRRK2-IN-1 and the two benzene rings of Candidate 3. Overall, when comparing the docking poses of all 10 candidates on 8FO7, it was evident that all candidates were positioned at the ATP binding site, where LRRK2-IN-1 originally binds, competitively inhibiting ATP.

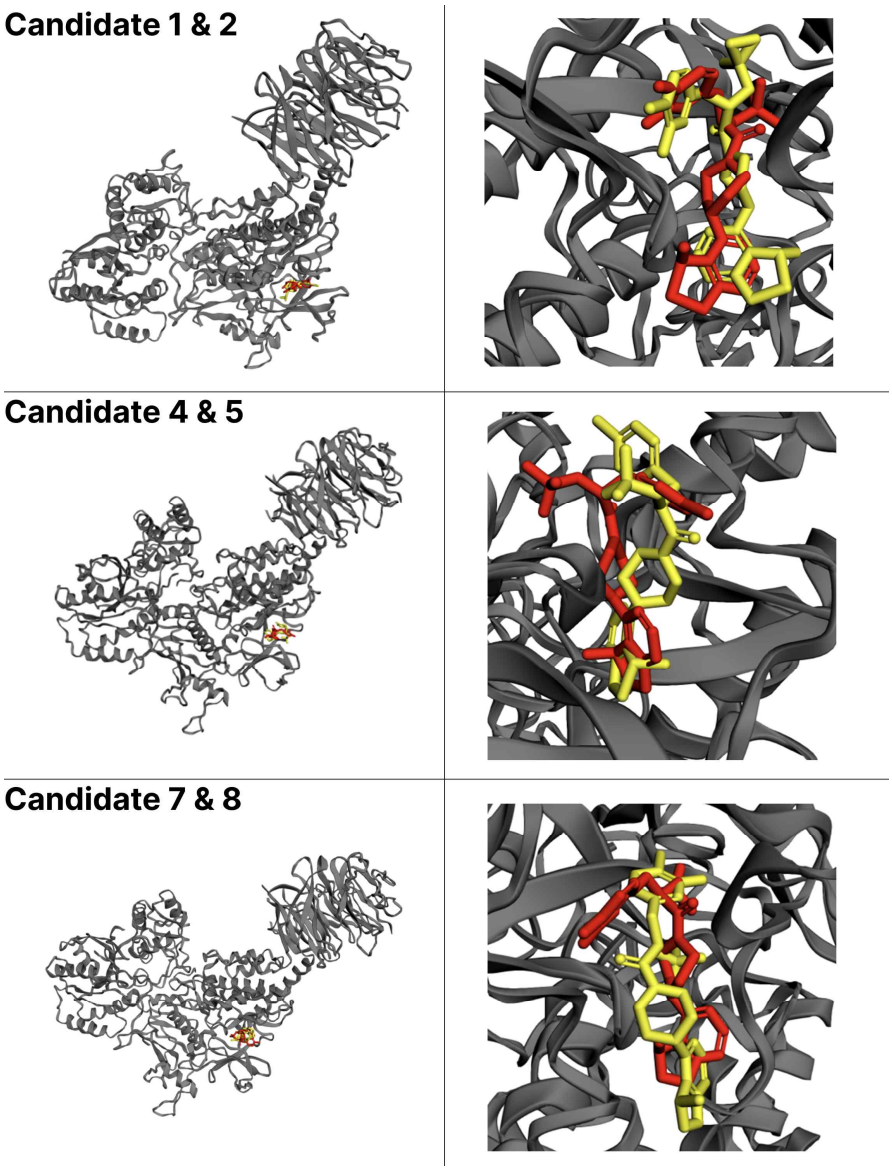

Fig 12. Comparison of Docking Poses Between Stereoisomer Pairs on 8FO7

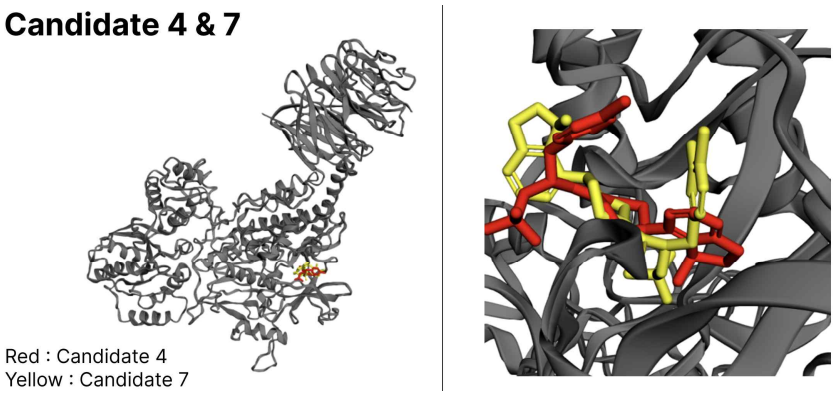

Fig 13. Comparison of Docking Poses Between Candidate 4 and Candidate 7 on 8FO7

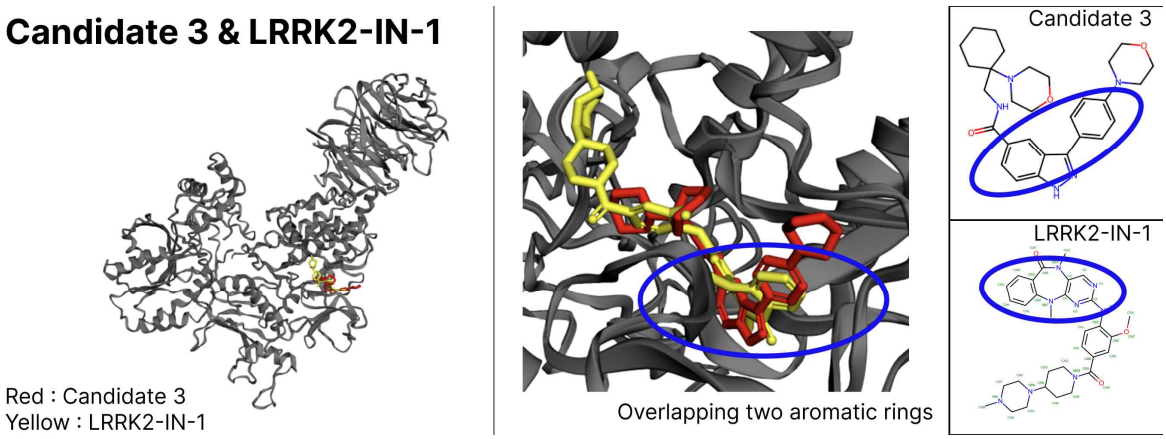

Fig 14. Comparison of Docking Poses Between Candidate 3 and LRRK2-IN-1 on 8FO7

The docking conformations of the candidates and LRRK2-IN-1 on 7LI3 were analyzed using the same approach. This time, the docking poses of stereoisomer pairs diverged significantly, despite their high 2D structural similarity (Fig 15). This suggests that some candidates found more favorable docking poses outside the ATP binding site. LRRK2-IN-1 and Candidate 3 showed a similar trend as observed on 8FO7; however, on 7LI3, the overlapping aromatic rings were different. For 8FO7, the overlapping aromatic rings are shown in blue (Fig 14), while for 7LI3, they are highlighted in red (Fig 16).

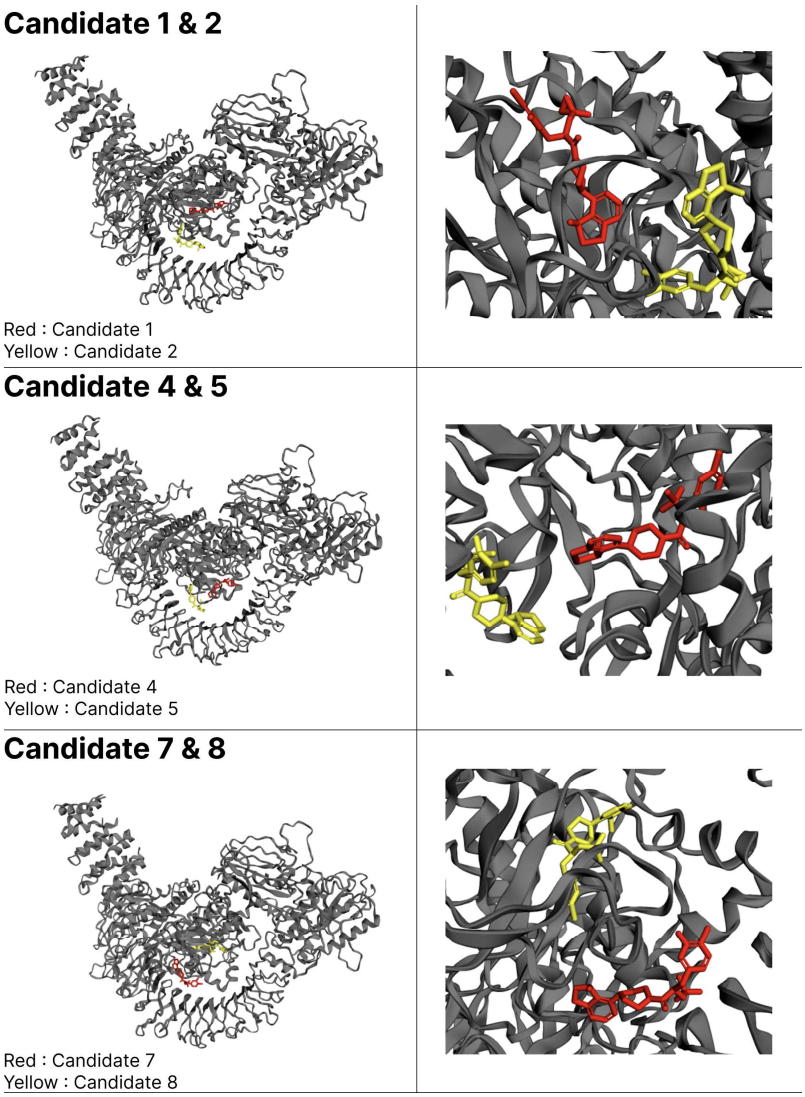

Fig 15. Comparison of Docking Poses Between Stereoisomer Pairs on 7LI3

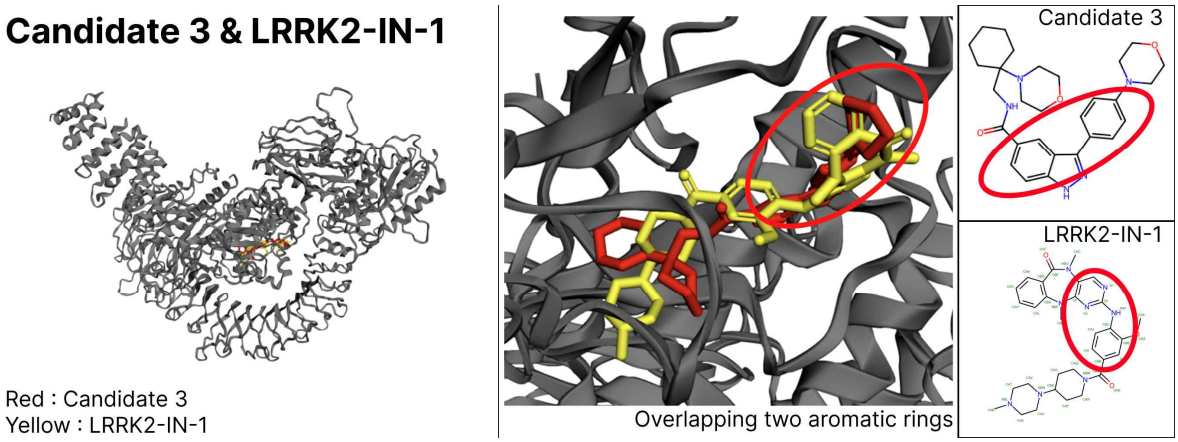

Fig 16. Comparison of Docking Poses Between Candidate 3 and LRRK2-IN-1 on 7LI3

To visualize the docking locations of all 10 candidates on 7LI3, their conformations were presented simultaneously in different colors (Fig 18). In Figure 18, the three distinct clusters were highlighted using circles: Cluster 1, which included Candidate 1, Candidate 3, and Candidate 8, was marked in red; Cluster 2, containing Candidate 4 and Candidate 6, was marked in green; and Cluster 3, comprising Candidate 2, Candidate 4, Candidate 7, Candidate 9, and Candidate 10, was marked in purple.

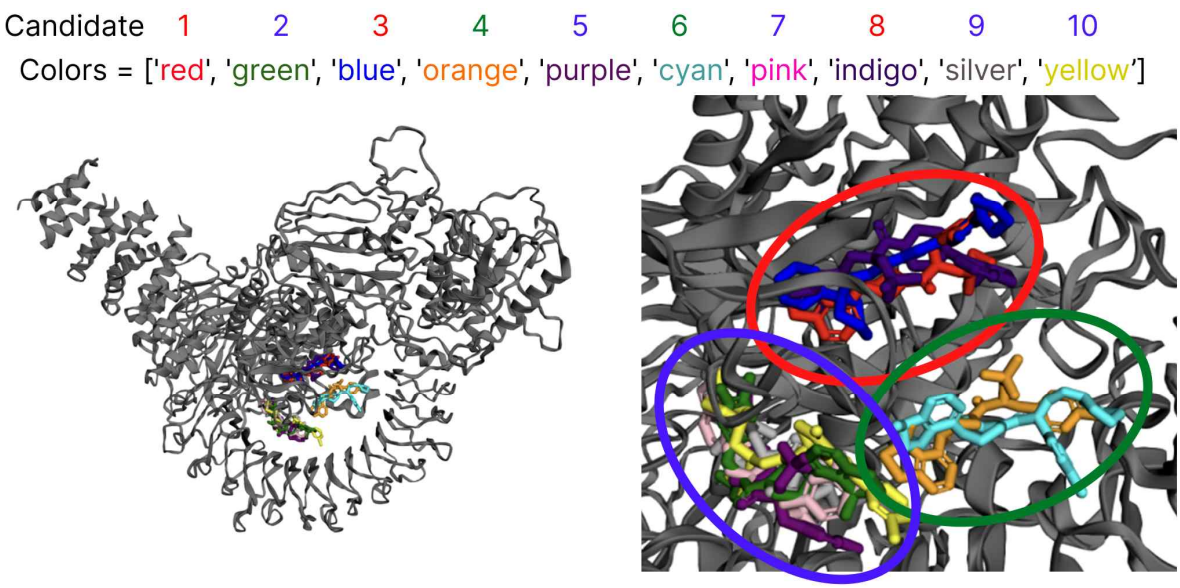

Fig 18. Visualization of Docking Locations for All 10 Candidates on 7LI3

The results revealed that to identify which cluster best competed with ATP, the docking locations were compared with the original ATP binding site. The analysis showed that Candidates 1, 3, and 8 docked properly at the ATP binding site, whereas the other clusters were hindered by other residues in 7LI3 (Fig 19).

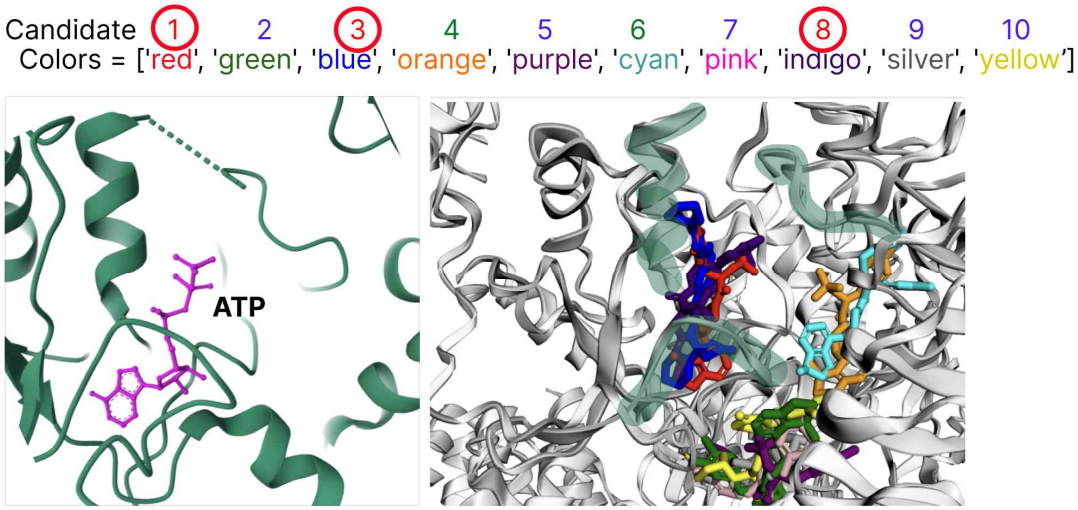

Fig 19. Comparison of Candidate Clusters with the ATP Binding Site on 7LI3

5. Conclusion

This study successfully employed ligand-based and structure-based virtual screening to identify potential inhibitors targeting LRRK2 and its G2019S-mutated form, both of which are critical in Parkinson's disease pathogenesis. A QSAR model was developed using bioactivity data from the ChEMBL database, enabling the virtual screening of a druggable compound library. Among the top 10 candidates identified, Candidate 3 (ZINC00029986675) demonstrated strong potential as an LRRK2 inhibitor. Despite not having the highest predicted pIC50 value, it ranked third with a predicted pIC50 of 9.872 and exhibited the strongest binding affinities among the candidates: -9.8 kcal/mol for wild-type LRRK2 (8FO7) and -9.0 kcal/mol for the G2019S-mutated LRRK2 (7LI3).

Docking analyses revealed important differences between wild-type and G2019S-mutated LRRK2. For wild-type LRRK2 (8FO7), all candidates docked at the ATP binding site, with stereoisomer pairs adopting similar binding poses. In contrast, on G2019S-mutated LRRK2 (7LI3), the docking poses of stereoisomer pairs diverged significantly, suggesting that the mutation influenced ligand binding. Furthermore, only Candidates 1, 3, and 8 successfully docked at the ATP binding site on 7LI3, while the other candidates were hindered by structural variations in the mutated protein. Candidate 3 displayed a favorable overlap with the aromatic rings of LRRK2-IN-1 in both 8FO7 and 7LI3, underscoring its potential as an effective inhibitor.

Although Candidate 3's binding affinity and predicted pIC50 did not surpass those of LRRK2-IN-1, Candidate 3 has the potential for optimization. Further modifications to its structure could enhance its binding affinity and reduce the dosage required for enzymatic inhibition. Additionally, LRRK2-IN-1's inability to cross the blood-brain barrier underscores the need for inhibitors with better pharmacokinetic properties. This study highlights Candidate 3 as a promising ligand for further development, offering a potential solution to address the challenges in treating Parkinson's disease patients.

6. Discussion

This study has several limitations that should be addressed in future research. One major limitation is the insufficient amount of LRRK2-targeted bioactivity data available in the ChEMBL database to train the QSAR model effectively. While the model demonstrated moderate predictive accuracy, as indicated by an RMSE within a reasonable range, the relatively low R^2 score highlights its inability to fully capture the complexity of the data. To overcome this limitation, generative models could be employed to augment the training dataset, enabling the model to learn from a more diverse range of data.

Additionally, while the Random Forest Regressor provided satisfactory results, implementing alternative machine learning architectures, such as Graph Neural Networks (GNNs) or Convolutional Neural Networks (CNNs), could potentially improve predictive performance. An initial attempt to use a deep neural network with ECFP features encountered extrapolation

issues, as the predicted pIC50 values for the ZINC library exceeded the experimental range in the ChEMBL dataset. Ensemble techniques may provide a solution to this problem by combining predictions from multiple models to improve generalization.

This study focused exclusively on type I LRRK2 inhibition due to the challenges in quantifying ligands' ability to stabilize the inactive conformation of LRRK2. Expanding this research to include type II or allosteric inhibitors could provide additional therapeutic options. By addressing these limitations and exploring new methodologies, future studies could contribute to the development of more robust and effective LRRK2 inhibitors for Parkinson's disease treatment.

7. References

[1] Statistics | Parkinson's Foundation. https://www.parkinson.org/understanding-parkinsons/statistics (accessed 2024-12-11).

[2] Zimprich, A.; Müller-Myhsok, B.; Farrer, M.; Leitner, P.; Sharma, M.; Hulihan, M.; Lockhart, P.; Strongosky, A.; Kachergus, J.; Calne, D. B.; Stoessl, J.; Uitti, R. J.; Pfeiffer, R. F.; Trenkwalder, C.; Homann, N.; Ott, E.; Wenzel, K.; Asmus, F.; Hardy, J.; Wszolek, Z.; Gasser, T. The PARK8 Locus in Autosomal Dominant Parkinsonism: Confirmation of Linkage and Further Delineation of the Disease-Containing Interval. Am J Hum Genet 2004, 74 (1), 11–19.

[3] Taylor, M.; Alessi, D. R. Advances in Elucidating the Function of Leucine-Rich Repeat Protein Kinase-2 in Normal Cells and Parkinson's Disease. Curr Opin Cell Biol 2020, 63, 102–113. https://doi.org/10.1016/j.ceb.2020.01.001.

[4] Usmani, A.; Shavarebi, F.; Hiniker, A. The Cell Biology of LRRK2 in Parkinson's Disease. Molecular and Cellular Biology 2021, 41 (5), e00660-20. https://doi.org/10.1128/MCB.00660-20.

[5] Snead, D. M.; Matyszewski, M.; Dickey, A. M.; Lin, Y. X.; Leschziner, A. E.; Reck-Peterson, S. L. Structural Basis for Parkinson's Disease-Linked LRRK2's Binding to Microtubules. Nat Struct Mol Biol 2022, 29 (12), 1196–1207. https://doi.org/10.1038/s41594-022-00863-y.

[6] West, A. B.; Moore, D. J.; Biskup, S.; Bugayenko, A.; Smith, W. W.; Ross, C. A.; Dawson, V. L.; Dawson, T. M. Parkinson's Disease-Associated Mutations in Leucine-Rich Repeat Kinase 2 Augment Kinase Activity. Proceedings of the National Academy of Sciences 2005, 102 (46), 16842–16847. https://doi.org/10.1073/pnas.0507360102.

[7] Xiong, Y.; Dawson, T. M.; Dawson, V. L. Models of LRRK2 Associated Parkinson's Disease. Adv Neurobiol 2017, 14, 163–191. https://doi.org/10.1007/978-3-319-49969-7_9.

[8] Berwick, D. C.; Harvey, K. LRRK2 Signaling Pathways: The Key to Unlocking

Neurodegeneration? Trends in Cell Biology 2011, 21 (5), 257–265. https://doi.org/10.1016/j.tcb.2011.01.001.

[9] Zhu, H.; Hixson, P.; Ma, W.; Sun, J. Pharmacology of LRRK2 with Type I and II Kinase Inhibitors Revealed by Cryo-EM. Cell Discov 2024, 10 (1), 1–12. https://doi.org/10.1038/s41421-023-00639-8.

[10] Thakur, G.; Vikas Kumar; Won, C. Structural Insights and Development of LRRK2 Inhibitors for Parkinson's Disease in the Last Decade. https://www.mdpi.com/2073-4425/13/8/1426 (accessed 2024-12-11).

[11]Best-in-Class LRRK2 Therapies for Parkinson's Disease. https://www.bioptic.io/case-study/best-in-class-lrrk2-therapies-for-parkinsons-disease (accessed 2024-12-11).

[12] DNL151 | ALZFORUM. https://www.alzforum.org/therapeutics/dnl151 (accessed 2024-12-11).

[13] DNL201 | ALZFORUM. https://www.alzforum.org/therapeutics/dnl201 (accessed 2024-12-11).

[14] Bouhouche, A.; Tibar, H.; Ben El Haj, R.; El Bayad, K.; Razine, R.; Tazrout, S.; Skalli, A.; Bouslam, N.; Elouardi, L.; Benomar, A.; Yahyaoui, M.; Regragui, W. LRRK2 G2019S Mutation: Prevalence and Clinical Features in Moroccans with Parkinson's Disease. Parkinson's Disease 2017, 2017 (1), 2412486. https://doi.org/10.1155/2017/2412486.

[15] Raig, N. D.; Surridge, K. J.; Sanz-Murillo, M.; Dederer, V.; Krämer, A.; Schwalm, M. P.; Elson, L.; Chatterjee, D.; Mathea, S.; Hanke, T.; Leschziner, A. E.; Reck-Peterson, S. L.; Knapp, S. Type-II Kinase Inhibitors That Target Parkinson's Disease-Associated LRRK2. bioRxiv October 31, 2024, p 2024.09.17.613365. https://doi.org/10.1101/2024.09.17.613365.

[16] Garofalo, A. W.; Bright, J.; De Lombaert, S.; Toda, A. M. A.; Zobel, K.; Andreotti, D.; Beato, C.; Bernardi, S.; Budassi, F.; Caberlotto, L.; Gao, P.; Griffante, C.; Liu, X.; Mengatto, L.; Migliore, M.; Sabbatini, F. M.; Sava, A.; Serra, E.; Vincetti, P.; Zhang, M.; Carlisle, H. J. Selective Inhibitors of G2019S-LRRK2 Kinase Activity. J Med Chem 2020, 63 (23), 14821–14839. https://doi.org/10.1021/acs.jmedchem.0c01243.

[17] Muegge, I.; Bergner, A.; Kriegl, J. M. Computer-Aided Drug Design at Boehringer Ingelheim. J Comput Aided Mol Des 2017, 31 (3), 275–285. https://doi.org/10.1007/s10822-016-9975-3.

[18] Batool, M.; Ahmad, B.; Choi, S. A Structure-Based Drug Discovery Paradigm. International Journal of Molecular Sciences 2019, 20 (11), 2783. https://doi.org/10.3390/ijms20112783.

[19] Arya, H.; Coumar, M. S. Chapter 4 - Lead Identification and Optimization. In The Design & Development of Novel Drugs and Vaccines; Bhatt, T. K., Nimesh, S., Eds.; Academic Press, 2021; pp 31–63. https://doi.org/10.1016/B978-0-12-821471-8.00004-0.

[20] Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry. J. Med. Chem. 2014, 57 (8), 3186–3204. https://doi.org/10.1021/jm401411z.

[21] Jung, S.; Vatheuer, H.; Czodrowski, P. VSFlow: An Open-Source Ligand-Based Virtual Screening Tool. Journal of Cheminformatics 2023, 15 (1), 40. https://doi.org/10.1186/s13321-023-00703-1.

[22] Hamza, A.; Wei, N.-N.; Zhan, C.-G. Ligand-Based Virtual Screening Approach Using a New Scoring Function. J. Chem. Inf. Model. 2012, 52 (4), 963–974. https://doi.org/10.1021/ci200617d.

[23] Saeh, J. C.; Lyne, P. D.; Takasaki, B. K.; Cosgrove, D. A. Lead Hopping Using SVM and 3D Pharmacophore Fingerprints. J. Chem. Inf. Model. 2005, 45 (4), 1122–1133. https://doi.org/10.1021/ci049732r.

[24] Neves, B. J.; Braga, R. C.; Melo-Filho, C. C.; Moreira-Filho, J. T.; Muratov, E. N.; Andrade, C. H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. Front. Pharmacol. 2018, 9. https://doi.org/10.3389/fphar.2018.01275.

[25] Torres, P. H. M.; Sodero, A. C. R.; Jofily, P.; Silva-Jr, F. P. Key Topics in Molecular Docking for Drug Design. International Journal of Molecular Sciences 2019, 20 (18), 4574. https://doi.org/10.3390/ijms20184574.

[26] Li, Q.; Shah, S. Structure-Based Virtual Screening. In Protein Bioinformatics: From Protein Modifications and Networks to Proteomics; Wu, C. H., Arighi, C. N., Ross, K. E., Eds.; Springer: New York, NY, 2017; pp 111–124. https://doi.org/10.1007/978-1-4939-6783-4_5.

[27] Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. J Comput Chem 2010, 31 (2), 455–461. https://doi.org/10.1002/jcc.21334.

[28] Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J. Chem. Inf. Model. 2021, 61 (8), 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.

[29] Garrett M., M.; David S., G.; Michael E., P. User Guide AutoDock Version 4.2 : Automated Docking of Flexible Ligands to Flexible Receptors, 2014.

[30] Sharma, A.; Yennamalli, R. M. Chapter 16 - Docking Strategies. In Basic Biotechniques for Bioprocess and Bioentrepreneurship; Bhatt, A. K., Bhatia, R. K., Bhalla, T. C., Eds.; Academic Press, 2023; pp 243–258. https://doi.org/10.1016/B978-0-12-816109-8.00016-7.

[31] Services | ATP Competition Assay | International Centre for Kinase Profiling. https://www.kinase-screen.mrc.ac.uk/services/atp-competition-assay (accessed 2024-12-11).

[32] Deng, X.; Dzamko, N.; Prescott, A.; Davies, P.; Liu, Q.; Yang, Q.; Lee, J.-D.; Patricelli, M. P.; Nomanbhoy, T. K.; Alessi, D. R.; Gray, N. S. Characterization of a Selective Inhibitor of the Parkinson's Disease Kinase LRRK2. Nat Chem Biol 2011, 7 (4), 203–205. https://doi.org/10.1038/nchembio.538.

[33] LRRK2-IN-1 | LRRK2 Inhibitor | MedChemExpress. MedchemExpress.com. https://www.medchemexpress.com/LRRK2-IN-1.html (accessed 2024-12-11).

[34] Liu, X.; Kalogeropulou, A. F.; Domingos, S.; Makukhin, N.; Nirujogi, R. S.; Singh, F.; Shpiro, N.; Saalfrank, A.; Sammler, E.; Ganley, I. G.; Moreira, R.; Alessi, D. R.; Ciulli, A. Discovery of XL01126: A Potent, Fast, Cooperative, Selective, Orally Bioavailable, and Blood–Brain Barrier Penetrant PROTAC Degrader of Leucine-Rich Repeat Kinase 2. J. Am. Chem. Soc. 2022, 144 (37), 16930–16952. https://doi.org/10.1021/jacs.2c05499.

[35] Estrada, A. A.; Chan, B. K.; Baker-Glenn, C.; Beresford, A.; Burdick, D. J.; Chambers, M.; Chen, H.; Dominguez, S. L.; Dotson, J.; Drummond, J.; Flagella, M.; Fuji, R.; Gill, A.; Halladay, J.; Harris, S. F.; Heffron, T. P.; Kleinheinz, T.; Lee, D. W.; Pichon, C. E. L.; Liu, X.; Lyssikatos, J. P.; Medhurst, A. D.; Moffat, J. G.; Nash, K.; Scearce-Levie, K.; Sheng, Z.; Shore, D. G.; Wong, S.; Zhang, S.; Zhang, X.; Zhu, H.; Sweeney, Z. K. Discovery of Highly Potent, Selective, and Brain-Penetrant Aminopyrazole Leucine-Rich Repeat Kinase 2 (LRRK2) Small Molecule Inhibitors. J. Med. Chem. 2014, 57 (3), 921–936. https://doi.org/10.1021/jm401654j.

[36] Jennings, D.; Huntwork-Rodriguez, S.; Henry, A. G.; Sasaki, J. C.; Meisner, R.; Diaz, D.; Solanoy, H.; Wang, X.; Negrou, E.; Bondar, V. V.; Ghosh, R.; Maloney, M. T.; Propson, N. E.; Zhu, Y.; Maciuca, R. D.; Harris, L.; Kay, A.; LeWitt, P.; King, T. A.; Kern, D.; Ellenbogen, A.; Goodman, I.; Siderowf, A.; Aldred, J.; Omidvar, O.; Masoud, S. T.; Davis, S. S.; Arguello, A.; Estrada, A. A.; de Vicente, J.; Sweeney, Z. K.; Astarita, G.; Borin, M. T.; Wong, B. K.; Wong, H.; Nguyen, H.; Scearce-Levie, K.; Ho, C.; Troyer, M. D. Preclinical and Clinical Evaluation of the LRRK2 Inhibitor DNL201 for Parkinson's Disease. Science Translational Medicine 2022, 14 (648), eabj2658. https://doi.org/10.1126/scitranslmed.abj2658.

[37] Roskoski, R. Classification of Small Molecule Protein Kinase Inhibitors Based upon the Structures of Their Drug-Enzyme Complexes. Pharmacological Research 2016, 103, 26–48.

https://doi.org/10.1016/j.phrs.2015.10.021.

[38] Zhao, Z.; Wu, H.; Wang, L.; Liu, Y.; Knapp, S.; Liu, Q.; Gray, N. S. Exploration of Type II Binding Mode: A Privileged Approach for Kinase Inhibitor Focused Drug Discovery? ACS Chem. Biol. 2014, 9 (6), 1230–1241. https://doi.org/10.1021/cb500129t.

[39] Carlson, H. A.; Smith, R. D.; Khazanov, N. A.; Kirchhoff, P. D.; Dunbar, J. B.; Benson, M. L. Differences between High- and Low-Affinity Complexes of Enzymes and Non-Enzymes. J Med Chem 2008, 51 (20), 6432–6441. https://doi.org/10.1021/jm8006504.