

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

Study of the structure-activity relationship of LRRK2 enzyme inhibitors to find alternative treatments for Parkinson's disease

María Caridad García Paz y Miño

Ingeniería Química

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero químico

Quito, 02 de mayo de 2023

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

**Study of the structure-activity relationship of LRRK2 enzyme inhibitors to
find alternative treatments for Parkinson's disease**

María Caridad García Paz y Miño

Nombre del profesor, Título académico

José Ramón Mora, Ph.D.

Nombre del profesor, Título académico

Juan Diego Fonseca, Ph.D.

Quito, 02 de mayo de 2023

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: María Caridad García Paz y Miño

Código: 00205948

Cédula de identidad: 1722218417

Lugar y fecha: Quito, 02 de mayo de 2023

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

RESUMEN

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo complejo que involucra varios neurotransmisores y cuya causa subyacente aún se desconoce. Los tratamientos actuales no ofrecen un perfil farmacológico óptimo, lo que impacta en la calidad de vida de los pacientes. Además, el proceso para desarrollar medicamentos es largo y costoso, lo que dificulta el descubrimiento de nuevas opciones. En este sentido, los investigadores han recurrido a enfoques *in silico* como un camino hacia nuevos tratamientos. El presente trabajo propone la creación de una base de datos que se utilizó para construir un modelo de predicción robusto sobre la relación estructura-actividad cuantitativa (QSAR) mediante la aplicación de diferentes algoritmos de aprendizaje de máquina (machine learning). Este modelo se utilizó para el cribado de la base de datos de Drug Bank para encontrar la posible aplicación de medicamentos ya existentes como alternativas para tratar la EP. El estudio se complementó con cálculos de acoplamiento molecular que proporcionan una comprensión más profunda de las interacciones entre la enzima y los inhibidores. A partir de la base de datos, la kinasa LRRK2 y los valores de pIC_{50} fueron seleccionados respectivamente como enzima y parámetro de medición de actividad y se utilizaron para el estudio QSAR. El modelo propuesto tenía 7 descriptores y mostró una fuerte capacidad de predicción con una validación cruzada y externa superiores a 0.79. El modelo también aprobó todos los requisitos de la prueba de validación de Tropsha. El cribado de la base de datos de Drug Bank llevó a la sugerencia de tres medicamentos para estudiarlos como posibles nuevos tratamientos. Los cálculos de acoplamiento molecular ayudaron a examinar las interacciones inhibidor-enzima, pero se requiere mayor investigación para un análisis más profundo.

Palabras clave: Enfermedad de Párkinson, LRRK2, pIC_{50} , QSAR, acoplamiento molecular, cribado, Drug Bank.

ABSTRACT

Parkinson's disease (PD) is a complex neurodegenerative disorder that involves several neurotransmitters and whose underlying cause is yet unknown. Current treatments do not offer an optimal pharmacological profile, which impacts the patients' quality of life. Besides, the process to develop drugs is long and expensive, making it harder to discover new options. In this regard, researchers have turned to *in silico* approaches as a leading path towards novel medicines. The present work proposes the creation of a dataset that is used to build a robust prediction model on the quantitative structure-activity relationship (QSAR) by applying different machine learning algorithms. This model is used for the screening of the Drug Bank database to find the possible application of already existing medicines as alternative treatments for PD. Furthermore, the study is complemented with molecular docking calculations that provide a deeper understanding of the interactions between the enzyme and the inhibitors. From the dataset, leucine-rich-repeat-kinase II (LRRK2) and pIC_{50} values were respectively selected as enzyme and activity measurement parameter and were used for the QSAR study. The proposed model had 7 descriptors and exhibited a strong prediction capability with a fivefold cross and external validations greater than 0.79. The model also approved all the requirements of the Tropsha's validation test. The screening of the Drug Bank dataset led to the suggestion of three drugs to be studied as possible new treatments. Molecular docking calculations helped examine inhibitor-enzyme interactions, but more research is required for in-depth analysis.

Key words: Parkinson's disease, LRRK2, pIC_{50} , QSAR, molecular docking, screening, Drug Bank.

TABLE OF CONTENTS

Introduction	10
Methodology.....	13
Construction of the data set and enzyme selection for the computational studies	13
Quantitative structure-activity relationship (QSAR) studies.....	14
Drug Bank screening	15
Molecular docking studies.....	16
Results and discussions.....	17
Construction of the data set and enzyme selection for the computational studies	17
QSAR Studies.....	18
Drug Bank screening	23
Molecular docking studies.....	24
Structural analysis of the inhibitors selected from the Drug Bank database	25
Conclusions	27
References.....	29
Appendix A: QSAR studies	33
Appendix B: Drug Bank screening	37
Appendix C: Molecular docking studies	39

INDEX OF TABLES

Table 1. Characteristics of M1 and M10 models.....	19
Table 2. Validation based on the Tropsha's test for QSAR modeling.....	21
Table 3. Description of the selected medicines from the Drug Bank.....	23
Table 4. pIC ₅₀ values and docking affinities of the selected drugs	25
Table A1. Best 15 models from the QSAR study	33
Table A2. Q _{CV} ² , Q _{Ext} ² and corresponding MAE of the best 15 models.....	34
Table A3. Descriptors of models M1 and M10, their abbreviations and number of appearances in the best 2 and best 15 models.....	35
Table A4. Pearson coefficients of M1 model's descriptors	36
Table B1. Characteristics and applications of the Drug Bank drugs with predicted pIC ₅₀ values greater than 9	37
Table C1. pIC ₅₀ values and docking scores of the dataset molecules	39

INDEX OF FIGURES

Figure 1. Block diagram of the methodology process.....	13
Figure 2. Number of inhibitors found per enzyme and per biological activity indicator	17
Figure 3. Structure of Sunitinib	18
Figure 4. a) External and CV of M1 model, b) External and CV of M10 model.....	19
Figure 5. a) Prediction by model's equation, b) Prediction by LOO method, c) Prediction by model's equation when all molecules are included in the same set.....	20
Figure 6. Pearson correlation coefficients for descriptors of model M1	22
Figure 7. Comparison between 4K5 experimental (turquoise) and docked (violet) conformations	25
Figure 8. Chemical structure of a) Triamterene, b) Phenazopyridine, c) Compound 1_31, d) Cannabigerol, and e) 4K5	26
Figure C1. Correlation between pIC ₅₀ values and docking scores of the dataset inhibitors	41

INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder that affects approximately 10 million people worldwide and is characterized by the loss of dopaminergic neurons in the substantia nigra. It is a multifactorial disease caused by a combination of genetic and environmental factors like head injury and exposure to certain chemicals¹. However, its underlying cause, as well as the mechanisms by which it progresses, remain unknown^{2,3}. This is because PD involves several neurotransmitters³ including dopaminergic and non-dopaminergic agents³⁻⁵, which makes it a very complex disorder. In addition, the knowledge of molecular targets to treat PD is narrow and incomplete². Consequently, it is difficult to develop treatments that offer optimal results. In this context, there is an unmet need to discover new compounds that offer an effective pharmacological profile to stop Parkinson's rapid progression.

Current medicines for PD are limited, fail to stop its evolution and cause unpleasant side effects. In fact, most of them focus only on alleviating symptoms and slowing down the disease², but they don't address a solution for the root problem⁶. One of the most used drugs is levodopa (L-DOPA) in combination with monoamine oxidase B (MAO-B) inhibitors and other one-target drugs⁷. However, L-DOPA gradually loses its efficacy, causes side effects like dyskinesia⁴, and may worsen symptoms like hallucinations and the dysregulation syndrome⁵. In the same way, MAO-B inhibitors cause hepatotoxicity and cheese reaction⁸. Moreover, the mixture of compounds leads to drug-drug interactions that produce severe side effects⁹. Therefore, patients end up with a bad quality of life. The difficulty is that the process to develop new medicines is long and expensive. First, because in vivo tests are costly and show variability². In addition, in vitro tests have limitations due to the existing struggle to recreate an accurate environment since models used for the tests were originally designed to study brain cancer².

In the last two decades some *in silico*, *in vitro*, and *in vivo* studies have been conducted to target different enzymes involved in PD. In this extent, several molecular targets have been of interest as antiparkinsonian agents. One of them is adenosine A2A receptor (AA2A), which belongs to the A-family of the G-protein-coupled receptors (GPCRs). This enzyme is involved in the control of motor functions and dopamine receptor activation. Therefore, its inhibition increases the level of dopamine and enhances signaling transmission^{2,7,10}. Different compounds, such as 1,3,5-triazine-thiadiazole hybrids^{10,11}, 2-benzylidene-1-indanone, and -tetralone derivatives^{12,13}, have been investigated as potent inhibitors of this enzyme.

In the same way, monoamine oxidase B is a target of interest as it plays a key role in the deamination of dopamine, which initiates a series of events that cause the development of PD¹⁴. Additionally, MAO-B inhibitors have shown neuroprotective properties³, so they are promising candidates for the treatment of this disorder. Scientists are looking for novel selective and reversible inhibitors of this target enzyme. Recent investigations have focused on acacetin 7-methyl ether¹⁵, rutamarin¹⁶, coumarins¹⁷ and derivatives from isocarboxazid⁸, 4-(3-Nitrophenyl)thiazol-2-ylhydrazone¹⁸, indole-substituted benzothiazoles and benzoxazoles¹⁹, (S)-2-(benzylamino)-propanamide²⁰, and eugenol²¹, between others.

Researchers have also focused on leucine-rich-repeat-kinase II (LRRK2) as a target in the treatment of PD. It has been shown that mutations in the LRRK2 gene are related with familial PD and are a major genetic risk factor for sporadic PD because they damage dopaminergic neurons^{6,22,23}. Therefore, it has been hypothesized that the inhibition of this enzyme may target a ground cause of the disorder and slow its progression by inducing adult neurogenesis^{6,23}. Type II kinase inhibitors, as well as derivatives from 5-azaindazole⁶ and indolinone²³, have been studied in this regard.

The searching of possible drugs assisted by *in silico* studies has attracted the attention of the research community, and the use of the quantitative structure-activity relationship (QSAR) approach in combination with other techniques as pharmacophore analysis, molecular docking, and molecular dynamics has demonstrated to be adequate for the case of different diseases, such as type 2 diabetes²⁴, primary hyperoxaluria type 1 (PHT1)²⁵, tuberculosis²⁶, drug-induced liver injury²⁷ and SARS-CoV-2²⁸. QSAR studies seek to predict the biological activity of molecules based on their structural and physicochemical properties^{25,29}. Meanwhile, molecular docking and molecular dynamic studies are used to explore the ligand-receptor interaction and predict the formation of stable complexes considering external conditions^{25,30}.

From the state of the art about the possible searching of drugs for this disease, it is extremely difficult to find a concise, systematic, and diverse dataset for the possible development of a robust and predictive model based on the quantitative structure-activity relationship (QSAR) approach. In this respect, the present work aims to propose a dataset for the construction of a model focused on different machine learning algorithms, and the model is used for the screening of the Drug Bank database, searching the second use of well-known drugs for the treatment of PD. The results obtained in the previous described modelling are complemented with the molecular docking calculation for the selected enzyme as target. It is expected that the project has academic, social and economic impacts since computational studies are cheaper than *in vitro* ones²⁷ and help reduce the number of experimental tests.

METHODOLOGY

The methodology followed in this study is based on three different approaches of computer aided drug design (CADD): QSAR, molecular docking, and virtual screening. First, a dataset was constructed based on available information in the literature. Then, the dataset was used as input for the QSAR studies. In the same way, the results of the QSAR procedure were used to perform the virtual screening. Simultaneously, molecular docking analyses were executed using the compounds from the data set in addition to the drugs that excelled in the virtual screening. The process is summarized in Figure 1.

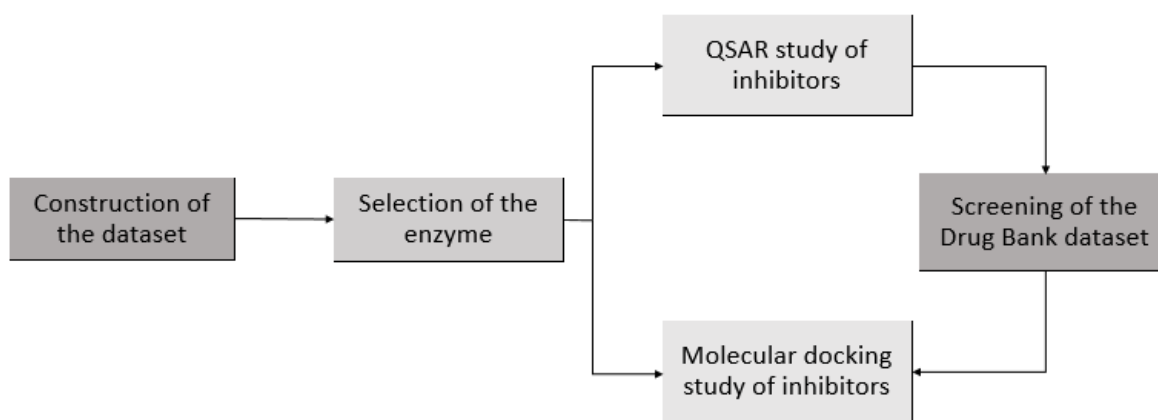


Figure 1. Block diagram of the methodology process

Construction of the data set and enzyme selection for the computational studies

For the construction of the dataset, publications from the last 15 years about the enzymes being targeted in the treatment of Parkinson's disease and current inhibitors of those enzymes were revised. Then, the parameters used to measure the inhibition of the enzymes in each study were identified. The collected information was classified according to the enzyme being studied in each paper and the parameter used to measure inhibition. The information was registered in a table in Microsoft Excel that distinguished between the following categories: name and structure of the inhibitor, name of the enzyme being targeted, the parameter used to measure the ligand-enzyme interaction and its

corresponding value, units of the quantity measured, and citation of the paper from which the information was taken. This table was used for the enzyme selection.

Based on the number of inhibitors found for each enzyme and the efficiency of the process, only one target with its corresponding activity measurement parameter was selected for the computational studies. In this case, LRRK2 enzyme and the IC₅₀ values for the inhibitors were chosen. IC₅₀ values indicate the concentration of drug required to inhibit 50% of the enzymes³¹. However, some compounds from the selected group were discarded based on the following criteria:

- values reported were inexact, meaning they used > or < to indicate concentrations above or below a certain number, and
- molecules came from a dataset with less than 4 candidates.

After polishing the list of LRRK2 inhibitors, a common control drug used for the biological assay was used as reference for the final construction of the dataset. Then, the molecules were drawn by using GaussView software.

Quantitative structure-activity relationship (QSAR) studies

The molecules' files created on the previous stage of the process, were introduced in QuBiLS-MIDAS software²⁷ and all of them were merged in a single file. Next, the structures were transformed to their Kekulé configurations to remove the resonance. Then, the molecules were optimized in 3D structures using the same Universal Force Field (UFF) parameter²⁵. The structures were merged into a single set using Open Babel software and optimized with RDKit software so that they had the same optimization parameters. Afterwards, ToMoCoMD MIDAS software was used to calculate the number of 3D topographic descriptors in a high-performance computer (HPC)²⁵ and the best descriptors were selected based on the Shannon entropy value of 0.7 and the Pearson

coefficient of 0.7. Then, the IC_{50} quantities were transformed to their corresponding pIC_{50} values using the formula $pIC_{50} = -\log(IC_{50}, M)$.

Prediction models that correlate the descriptors with the pIC_{50} values were built using different machine learning algorithms of Weka software. Three search methods—Genetic Search (GenSe), Greedy Stepwise (GreedSt), and Best First (BF)—were used with each of the following classifiers of the Wrapper Subset Evaluator: Gaussian Processes (GP), Linear Regression (LR), SMOreg, IBK, and Random Forest (RF). The process was repeated two consecutive times. Subsequently, models with less than 9 descriptors were selected and for each one of them, inhibitors were separated in a training set and a test set. The training and test sets were used to evaluate the performance and predictability of the models³². Models were evaluated and validated applying the external (Ext) and fivefold cross-validation (CV) methods.

Furthermore, the applicability domain (AD) of the test set on the training of the best model was analyzed. The AD is the theoretical range in which QSAR predictions are considered reliable and accurate, and it is defined by the training set of a model²⁵. It was determined using AMBIT software through the defined consensus by default of 4 methods: Range, Euclidean distance, City-block and probability density²⁵.

Drug Bank screening

The Drug Bank database was screened using the best model from the QSAR study and the pIC_{50} value of each compound was predicted. The compounds with a pIC_{50} value higher than 9 were selected and the current applications of each one of them were investigated to determine the most suitable drugs to be potential treatments for Parkinson's disease. The selected drugs were subjected to molecular docking studies.

Molecular docking studies

To perform the molecular docking studies, the X-ray diffraction crystal structure of human LRRK2 (PDB:4YZN) was downloaded from the Protein Data Bank (PDB). Then, PyMOL software was used to prepare the enzyme and the inhibitors for the molecular docking calculations²⁵. Waters were removed and the natural ligand was separated from the enzyme²⁵. Afterwards, all LRRK2 inhibitors from the dataset, as well as the selected compounds from the Drug Bank and the natural ligand, were docked against the enzyme using AutoDock Vina software. Before performing the docking calculations, AutodockTools was used to add polar hydrogens and obtain the structures in .PDBQT format²⁵. The coordinates used for the calculations were based on the active site of the enzyme (x=-8.308, y=16.203, z=18.565) and the grid box size was 12, 14 and 14 Angstrom for x, y, and z axes respectively.

RESULTS AND DISCUSSIONS

Construction of the data set and enzyme selection for the computational studies

During the literature revision, three enzymes outstood as targets of interest in the treatment of PD (AA2A, MAO-B and LRRK2) have been primarily studied. In addition, it was determined that K_i and IC_{50} values are the main parameters used to measure the biological activity of the enzymes' inhibitors. However, the reported values for some of the molecules were inexact and, therefore, those compounds were discarded from the dataset. Based on this information, an histogram was constructed (Figure 2).

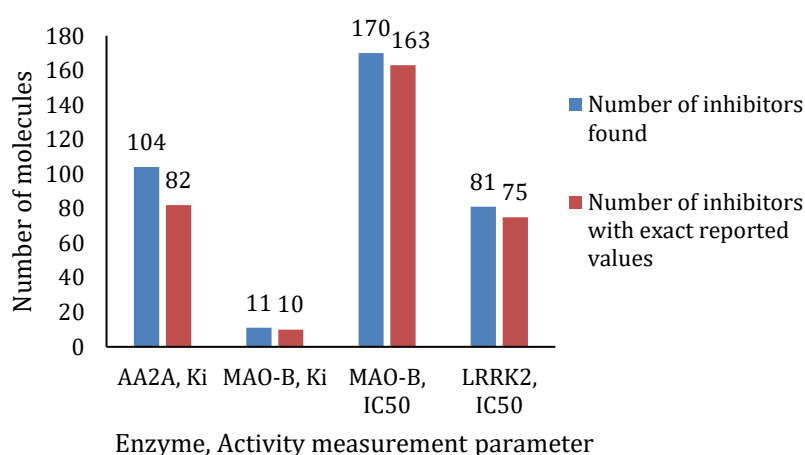


Figure 2. Number of inhibitors found per enzyme and per biological activity indicator

Figure 2 indicates the number of inhibitors found for the enzymes of interest and the corresponding parameter used to measure their activity. It can be seen that the category with the greatest number of molecules is the one that measures the inhibition of MAO-B in terms of IC_{50} values. In contrast, few studies used K_i to measure the inhibition of this enzyme. Only 11 molecules entered the previously mentioned category. This is probably related to the fact that experimental IC_{50} values are obtain at lower costs of time, materials and effort than K_i ones³³. Regarding the number of compounds studied as LRRK2 inhibitors, it is slightly inferior to the one of AA2A. However, it can still be considered a strong and reliable dataset. LRRK2 was preferred over MAO-B and AA2A

because of structural diversity of the inhibitors found and based on the fact that they have the same control compound: Sunitinib^{6,23,34}, which is a multi-specific tyrosine kinase inhibitor³⁵. The 2D structure of this molecule is depicted in Figure 3. The two rings at the left bottom were used as scaffold of the other derivative compounds, which may suggest they play an important role in the inhibition of the enzyme.

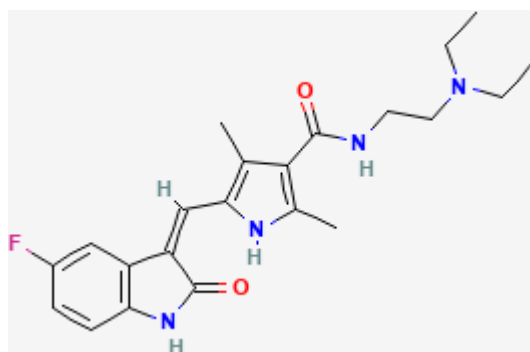


Figure 3. Structure of Sunitinib

QSAR Studies

More than forty subset of descriptors were found with the different combination of machine learning techniques and searching methods. Then, fifteen individual models with less than 9 descriptors were selected for further studies. All of them satisfy the criteria of $Q_{CV}^2 > 0.7$ when they are trained to predict the pIC_{50} values. The models were named from M1 through M15 as indicated in Table A1 from the supplementary information. After dividing the molecules in the corresponding training (75%) and test (25%) sets and performing the external and cross-validations (Table A2), the best two models were identified. They were selected on the basis that both of their validation coefficients, Q_{CV}^2 and Q_{Ext}^2 , are greater than 0.79. The exact values are shown on Table 1, as well as the corresponding MAEs. Model M1 was preferred over M10 because it has less descriptors and smaller external and cross-validation MAE.

Table 1. Characteristics of M1 and M10 models

Name of the Model	Machine Learning Algorithm	Number of descriptors	Q_{CV}^2	MAE_{CV}	Q_{Ext}^2	MAE_{Ext}
M1	LR	7	0.798	0.354	0.795	0.363
M10	SMOreg	9	0.797	0.390	0.803	0.373

The AD of model M1 was analyzed as described in the methodology section. Fortunately, all molecules from the test set entered the domain and no recalculation of the statistical parameters had to be done. The performance of model M1 was analyzed by plotting the experimental pIC_{50} values of each inhibitor against the ones predicted computationally through the model's equation, the fivefold cross-validation and leave-one-out (LOO) methods for the training set and the external evaluation for the test set. Similarly, model M10 was assessed by relating the experimental and predicted pIC_{50} values using the external and cross-validation approaches.

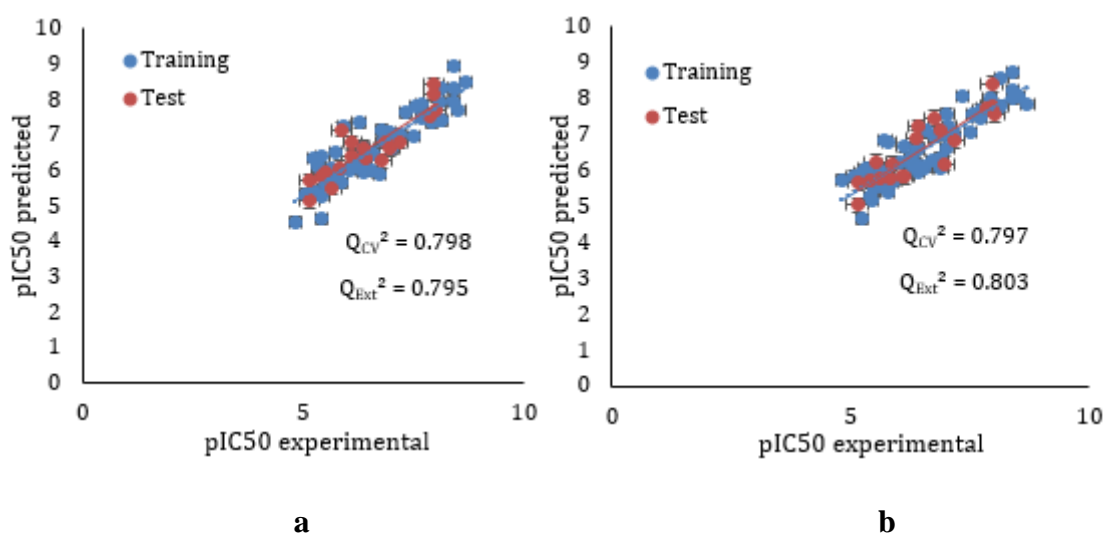


Figure 4. a) External and cross validation (CV) of M1 model, b) External and cross validation (CV) of M10 model

As it can be seen in Figure 4, the two models, M1 and M10, present a good linearity (both $Q^2s > 0.79$) which indicates their robustness to predict LRRK2 inhibition within the AD. In addition, model M1 showed a good fitting when values were plotted using the models' equation and the LOO prediction method as shown in Figure 5.

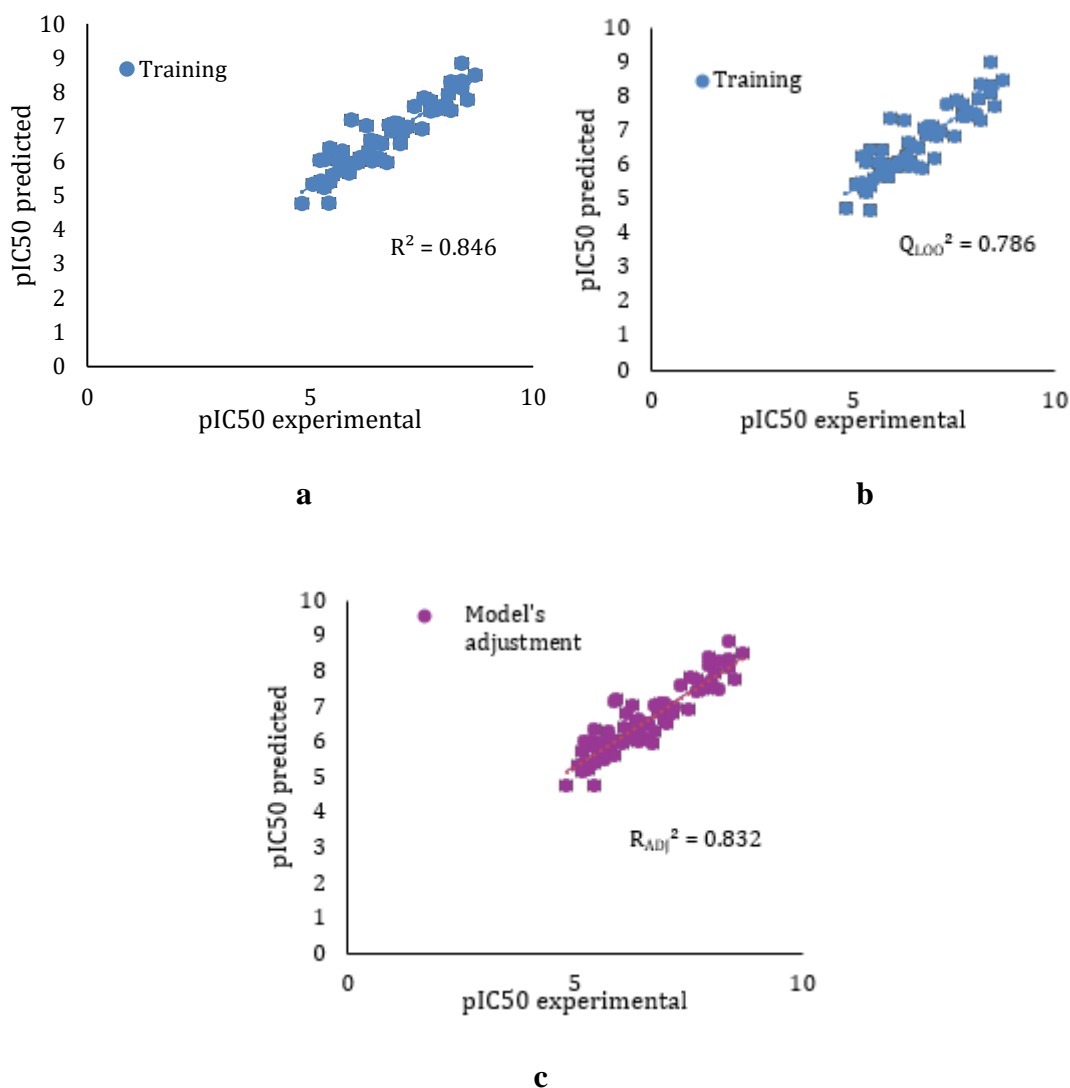


Figure 5. a) Prediction by model's equation, b) Prediction by LOO method, c) Prediction by model's equation when all molecules are included in the same set.

Model M1 has a correlation coefficient (R^2) of 0.846, meaning its equation can reliably predict the inhibitory activity of LRRK2 of this dataset. When the molecules from the test set were considered, R^2 of the adjustment only decreased by 0.014. The LOO

method of prediction presents the lowest linearity. Nonetheless, its Q^2 value varies very little in comparison to those of the external and CV. Therefore, the results are consistent between one another and M1 can be considered a robust prediction model.

Furthermore, the accuracy of model M1 was validated using the Tropsha's test as found in the literature²⁶ (Table 2). All the statistical parameters of Table 2 approved the validation test. Consequently, they corroborate the previous results on M1's robustness and reliability.

Table 2. Validation based on the Tropsha's test for QSAR modeling

Criterion	Leave-One-Out Validation		External Validation	
	Result	Assesment	Result	Assesment
$r^2 > 0.6$	0.846	Pass	0.846	Pass
$r^2_{val} > 0.5$	0.786	Pass	0.795	Pass
$(Q^2_{val} - R_0'^2)/Q^2_{val} < 0.1$	0.048	Pass	0.070	Pass
$(Q^2_{val} - R_0^2)/Q^2_{val} < 0.1$	0.002	Pass	0.000	Pass
$abs(R_0^2 - R_0'^2) < 0.1$	0.036	Pass	0.056	Pass
$0.85 < K < 1.15$	0.999	Pass	0.978	Pass
$0.85 < k' < 1.15$	0.996	Pass	1.019	Pass

Finally, the collinearity between descriptors of M1 was analyzed to guarantee there is no redundant information or overfitting in the model. A Pearson coefficient of $r < 0.7$ between the descriptors was established as baseline to consider it a strong model²⁵. No descriptor had a correlation coefficient higher than 0.6 (Figure 6), which confirms the models' strength. A table with the descriptors' full names as well as a detailed matrix of all the Pearson coefficients can be found in the supplementary information (Table A3 and Table A4).

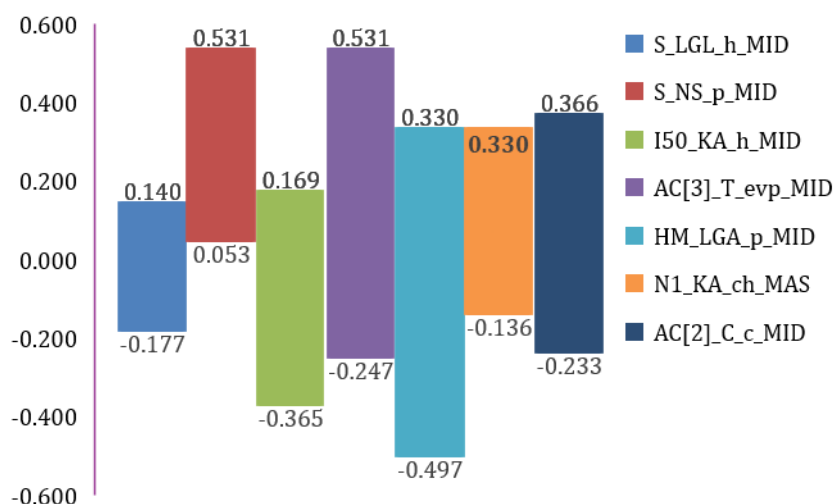


Figure 6. Pearson correlation coefficients for descriptors of model M1

The descriptors were also analyzed in terms of their number of appearances in the models. It was determined that models M1 and M10 only had one common descriptor (Table A3), which indicates there is diversity between them. This also suggests that an ensemble model could be built to significantly increase their prediction capacity while analyzing a wider range of structural and physicochemical properties. The common descriptor is present in two other of the remaining best 15 models as well. This fact advocates the importance of the descriptor in the prediction of pIC_{50} values of the dataset. Similarly, descriptor GV[5]_KA_ps_MID, which belongs to M10, is present in 40% of the best models, demonstrating its relevance in the prediction of the selected parameter.

Properties like softness (s), hardness (h), polarizability (p), electronegativity (e), Van der Waals volume (v), charge (c), and molecular weight (m) were found as descriptors. These parameters have been analyzed in the drug development of numerous diseases^{24,25,27}. It should be remarked that the property most frequently evaluated by the descriptors of M1 and M10 is polarizability. This is probably due to the presence of highly electronegative atoms like oxygen, nitrogen, and halogens in the inhibitors.

Drug Bank screening

The screening of the Drug Bank database resulted in a list of approved and experimental drugs that could be explored as potential new treatments for PD. Model M1 was used to predict the pIC_{50} values of the different drugs as described in the methodology section, of which 273 compound entered the AD. 32 out of the 273 molecules had a pIC_{50} greater than 9 and, between them, four exhibited properties that make them suitable for a second application as PD treatments. Most of the compounds that presented a $pIC_{50} > 9$ are experimental drugs (22 out of 32) and, therefore, there is no available information on their applications. For that reason, they were discarded. Similarly, antibiotics and antivirals were not considered appropriate options since antimicrobial and antiviral resistance are a major problem nowadays^{36,37}. The remaining molecules, Triamterene, Phenazopyridine, Cannabigerol, and Ademetionine, were plausible alternatives due to their anti-inflammatory and analgesic characteristics (Table 3). Nonetheless, Ademetionine was also left out because it produces unwanted side effects³⁸⁻⁴⁰. Detailed information of each of the 32 drugs is found in the supplementary information (Table B1).

Table 3. Description of the selected medicines from the Drug Bank

Name	Drug Bank ID	Description
Triamterene	DB00384	Potassium-sparing diuretic used in the treatment of edema and hypertension ⁴¹ .
Phenazopyridine	DB01438	Local anesthetic used for the relief of discomfort caused by pain, burning, and irritation ⁴² .
Cannabigerol	DB14734	Natural product found in Cannabis sativa and Helichrysum. It enhances appetite and acts as an anti-inflammatory, antioxidant and neuroprotective agent ⁴³ .
Ademetionine	DB00118	Physiologic methyl radical donor used as anti-inflammatory and applied in the treatment of depression, liver disorders, fibromyalgia, and osteoarthritis. It has also been used as dietary supplement for the support of bone and joint health, and as a mood and emotional regulator ⁴⁴ .

Molecular docking studies

The interaction between LRRK2 enzyme and its inhibitors (the compounds of the dataset and the three selected drugs) was studied in detail with molecular docking, and the obtained results were compared to the ones of the natural ligand (4K5). Two of the compounds exhibited the same docking scores as the natural ligand (-8.6 kcal/mol), and four of them presented even more negative energies (from -8.7 to -9.2 kcal/mol). For the molecules that displayed more positive scores, the range varied between -6.0 kcal/mol and -8.5 kcal/mol, of which 15 compounds had energies lower than -8 kcal/mol (Table C1). This shows that, in general, the molecules from the proposed dataset can easily interact with LRR2 enzyme. However, when taking into consideration the pIC₅₀ values, compound 1_31 is the most promising candidate to be a possible new treatment for PD since it has the highest pIC₅₀ value and almost the same binding energy as the natural ligand of LRRK2 enzyme (Table C1).

The molecular docking calculations were validated by comparing the natural ligands' experimental and docked conformations and verifying they overlap in the enzyme's active site (Figure 7). The overlaying of both structures demonstrates the viability of this procedure to be applied in the study of the binding mode of LRKK2 inhibitors. Additionally, the docking scores were plotted against the pIC₅₀ values to determine if any correlation existed between them (Figure C1). The plot shows a tendency of increasing pIC₅₀ as the docking scores become more negative. However, it was found that they cannot be used to predict the pIC₅₀ of a molecule because their correlation is extremely low (0.13).

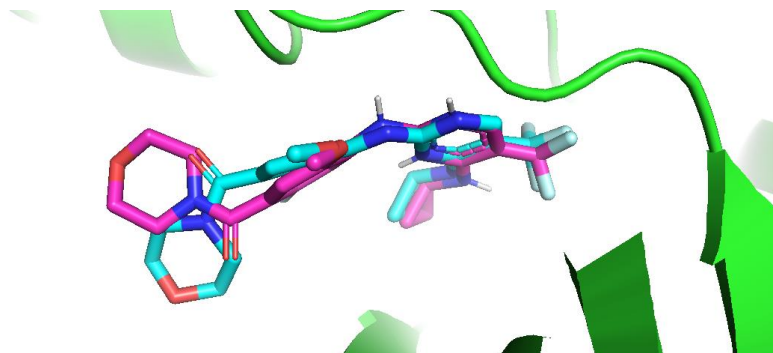


Figure 7. Comparison between 4K5 experimental (turquoise) and docked (violet) conformations

Regarding the drugs from the Drug Bank database, Triamterene was the one with the most negative docking score (Table 4), which indicates it is more easily binded to the enzyme than the other two. Moreover, Triamterene has a high pIC_{50} value compared to the molecules of the dataset. Consequently, this drugs exhibits a good interaction in terms of inhibition potential and energy needed. However, it should be taken into account that there is a significant difference (11.6%) in the docking affinity when compared to 4K5.

Table 4. pIC_{50} values and docking affinities of the selected drugs

Name	pIC_{50}	Docking score (kcal/mol)
Triamterene	9.471	-7.6
Cannabigerol	9.408	-7.1
Phenazopyridine	9.591	-6.8

Structural analysis of the inhibitors selected from the Drug Bank database

The structures of the selected medicines were also analyzed and compared to the ones of compound 1_31 (which has the highest pIC_{50} value within the molecules of the dataset) and LRRK2's natural ligand (Figure 8). Interestingly, almost all of them contain nitrogen heterocycles. The presence of resonance in these structures may affect their reactivity and interaction with the enzyme due to stabilization of the ligand and the formation of hydrogen bonds, especially in the N- and C- terminals of LRRK2 which

have shown to contain protein-protein interaction domains that regulate the enzyme's activity and localization⁴⁵. However, there is currently no concise information on the mechanisms by which LRRK2 inhibitors interact with it and, therefore, not much can be discussed in this regard.

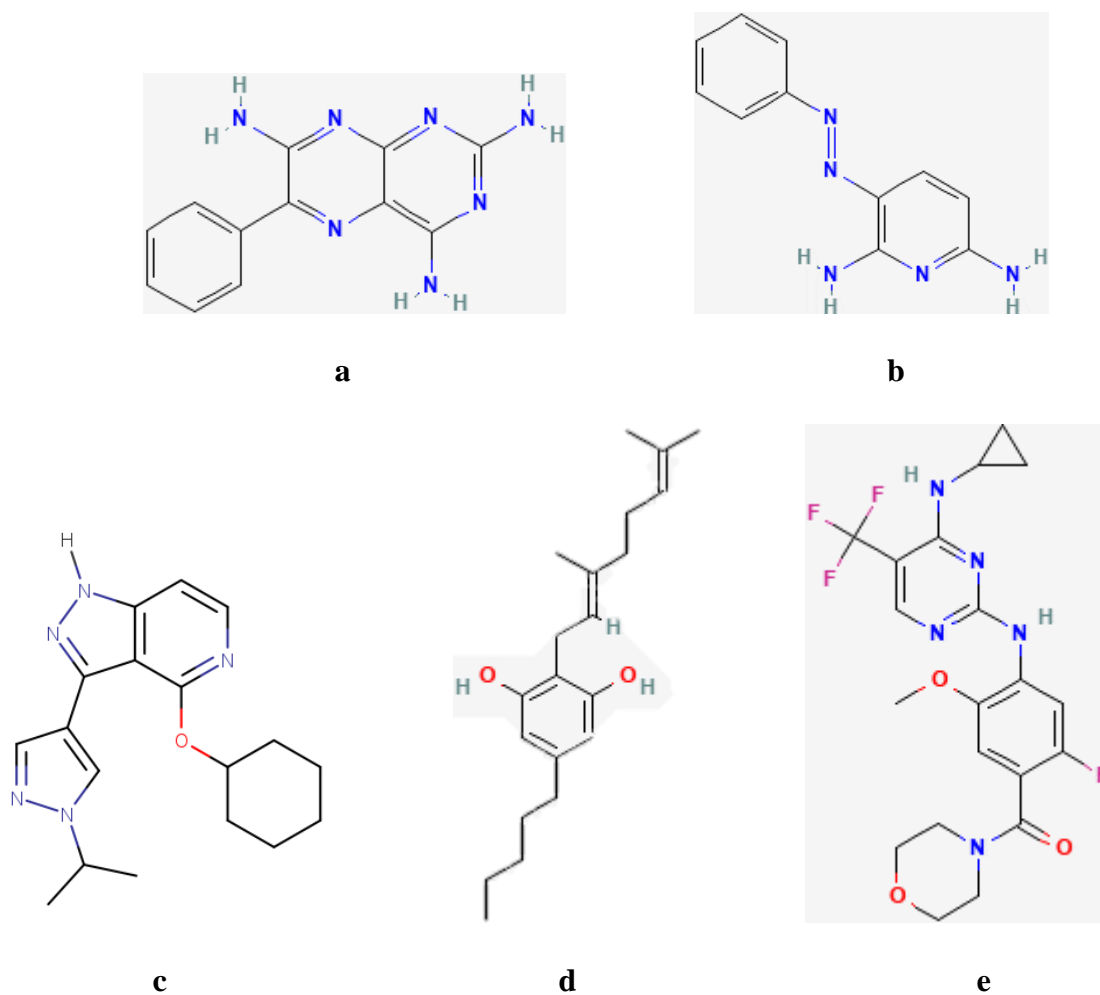


Figure 8. Chemical structure of a) Triamterene, b) Phenazopyridine, c) Compound 1_31, d) Cannabigerol, and e) 4K5

CONCLUSIONS

The dataset proposed by this work contributed to the construction of a prediction model that can be used to search the second application of already existing drugs as treatments of PD. Model M1 is a robust model with 7 descriptors that can reliably predict pIC_{50} values over its applicability domain. Predictions performed with the internal fivefold cross-validation, the LOO method and the model's equation demonstrated to have a good determination coefficient (R^2 and $Q^2 > 0.79$). Likewise, it achieved a remarkable prediction of the test set with a Q_{Ext}^2 of 0.795. Additionally, the model approved all the items analyzed in the Tropsha's test, which confirmed its consistency and strength. Regarding the descriptors of M1, there was no significant correlation between them (< 0.6), meaning there is not redundant information or overfitting. Interestingly, only one of the seven descriptors coincided with the ones of M10 (the second best model). Therefore, it is suggested to build an ensemble model in future work to enhance the prediction capacity while analyzing a wider range of properties.

The screening of the Drug Bank database led to the proposal of three already existing medicines to be investigated as potential new treatments for PD. All three compounds presented pIC_{50} values greater than those of the dataset. On the contrary, their docking scores were significantly more positive compared to the one of the natural ligand. Nevertheless, it is recommended that these molecules are tested as LRRK2 inhibitors to experimentally determine their feasibility as PD therapies or lead compounds for them. Molecule 1_31 from the dataset is also a promising candidate for a possible new treatment since it has the highest pIC_{50} value of the set and almost the same binding energy as the natural ligand of LRRK2 enzyme.

The molecular docking calculations facilitated the study of the interactions between the inhibitors and the enzyme. Still, further studies are needed for a deeper

analysis. Thus, molecular dynamic calculations are proposed as an alternative to get a more profound insight on the ligand-enzyme interactions. Even with little available information about the binding mechanisms, this procedure might be useful for a better understanding of the inhibition process. It is thought that research in this area may lead to the development of better treatments for PD⁴⁵.

REFERENCES

1. Parkinson's Foundation. Parkinson's Foundation. Published 2023. <https://www.parkinson.org/understanding-parkinsons/causes>
2. Koszła O, Stępnicki P, Zięba A, Grudzińska A, Matosiuk D, Kaczor AA. Current approaches and tools used in drug development against parkinson's disease. *Biomolecules*. 2021;11(6):1-16. doi:10.3390/biom11060897
3. Schapira AH V, Bezard E, Brotchie J, Calon F. Novel pharmacological targets for the treatment of Parkinson ' s disease. 2006;5:845-854. doi:10.1038/nrd2087
4. Loza MI, Carlsson J. Docking Screens for Dual Inhibitors of Disparate Drug Targets for Parkinson ' s Disease. Published online 2018. doi:10.1021/acs.jmedchem.8b00204
5. Meissner WG, Frasier M, Gasser T, et al. Priorities in Parkinson ' s disease research. 2011;10(May). doi:10.1038/nrd3430
6. Osborne J, Birchall K, Tsagris DJ, et al. Discovery of potent and selective 5-azaindazole inhibitors of leucine-rich repeat kinase 2 (LRRK2) – Part 1. *Bioorganic Med Chem Lett*. 2018;29(4):668-673. doi:10.1016/j.bmcl.2018.11.058
7. Cruz-vicente P, Silvestre S, Gallardo E. Recent Developments in New Therapeutic Agents against Alzheimer and Parkinson Diseases: In-Silico Approaches. *Molecules*. 2021;26(2193):1-28.
8. Agrawal N, Mishra P. Synthesis, monoamine oxidase inhibitory activity and computational study of novel isoxazole derivatives as potential antiparkinson agents. *Comput Biol Chem*. 2019;79(January):63-72. doi:10.1016/j.compbiolchem.2019.01.012
9. Affini A, Hagenow S, Zivkovic A, Marco-Contelles J, Stark H. Novel indanone derivatives as MAO B/H3R dual-targeting ligands for treatment of Parkinson's disease. *Eur J Med Chem*. 2018;148:487-497. doi:10.1016/j.ejmech.2018.02.015
10. Masih A, Singh S, Agnihotri AK, et al. Design and development of 1,3,5-triazine-thiadiazole hybrids as potent adenosine A2A receptor (A2AR) antagonist for benefit in Parkinson's disease. *Neurosci Lett*. 2020;735:135222. doi:10.1016/j.neulet.2020.135222
11. Masih A, Agnihotri AK, Srivastava JK, Pandey N, Bhat HR, Singh UP. Discovery of novel 1,3,5-triazine as adenosine A2A receptor antagonist for benefit in Parkinson's disease. *J Biochem Mol Toxicol*. 2021;35(3):1-7. doi:10.1002/jbt.22659
12. Van Rensburg HDJ, Legoabe LJ, Terre'Blanche G, Van Der Walt MM. 2-Benzylidene-1-Indanone Analogues as Dual Adenosine A 1 /A 2a Receptor Antagonists for the Potential Treatment of Neurological Conditions. *Drug Res (Stuttg)*. 2019;69(7):382-391. doi:10.1055/a-0808-3993

13. Janse Van Rensburg HD, Legoabe LJ, Terre'Blanche G. Synthesis and Structure Activity Relationships of Chalcone based Benzocycloalkanone Derivatives as Adenosine A 1 and/or A 2A Receptor Antagonists. *Drug Res (Stuttg)*. 2020;70(6):243-256. doi:10.1055/a-1146-2996
14. Naidoo D, Roy A, Slavětínská LP, Chukwujekwu JC, Gupta S, Van Staden J. New role for crinamine as a potent, safe and selective inhibitor of human monoamine oxidase B: In vitro and in silico pharmacology and modeling. *J Ethnopharmacol*. 2020;248:112305. doi:10.1016/j.jep.2019.112305
15. Chaurasiya ND, Zhao J, Pandey P, Doerksen RJ, Muhammad I, Tekwani BL. Selective Inhibition of Human Monoamine Oxidase B by Acacetin 7-Methyl Ether Isolated from *Turnera diffusa* (Damiana). *Molecules*. 2019;24(4):1-15. doi:10.3390/molecules24040810
16. Koziol E, Luca SV, Ağalar HG, et al. Rutamarin: Efficient liquid–liquid chromatographic isolation from *ruta graveolens* L. And evaluation of its in vitro and in silico MAO-B inhibitory activity. *Molecules*. 2020;25(11):1-12. doi:10.3390/molecules25112678
17. Rizakov AZ, Kolev MK, Velkov ZA. QSAR analysis of coumarins, flavones and their bicyclo ethers as monoamine oxidases inhibitors. *Bulg Chem Commun*. 2020;52:93-100. doi:10.34049/bcc.52.A.330
18. Secci D, Carradori S, Petzer A, et al. 4-(3-Nitrophenyl)thiazol-2-ylhydrazone derivatives as antioxidants and selective hMAO-B inhibitors: synthesis, biological activity and computational analysis. *J Enzyme Inhib Med Chem*. 2019;34(1):597-612. doi:10.1080/14756366.2019.1571272
19. Nam MH, Park M, Park H, et al. Indole-Substituted Benzothiazoles and Benzoxazoles as Selective and Reversible MAO-B Inhibitors for Treatment of Parkinson's Disease. *ACS Chem Neurosci*. 2017;8(7):1519-1529. doi:10.1021/acschemneuro.7b00050
20. Jin CF, Wang ZZ, Chen KZ, Xu TF, Hao GF. Computational Fragment-Based Design Facilitates Discovery of Potent and Selective Monoamine Oxidase-B (MAO-B) Inhibitor. *J Med Chem*. 2020;63(23):15021-15036. doi:10.1021/acs.jmedchem.0c01663
21. Dhiman P, Malik N, Khatkar A. Lead optimization for promising monoamine oxidase inhibitor from eugenol for the treatment of neurological disorder: Synthesis and in silico based study. *BMC Chem*. 2019;13(3):1-20. doi:10.1186/s13065-019-0552-4
22. Khamouli S, Belaidi S, Ouassaf M, Lanez T, Belaaouad S, Chtita S. Multi-combined 3D-QSAR, docking molecular and ADMET prediction of 5-azaindazole derivatives as LRRK2 tyrosine kinase inhibitors. *J Biomol Struct Dyn*. 2020;40(3):1285-1298. doi:10.1080/07391102.2020.1824815
23. Salado IG, Zaldivar-Diez J, Sebastian V, et al. Leucine rich repeat kinase 2 (LRRK2) inhibitors based on indolinone scaffold: Potential pro-neurogenic agents. *Eur J Med Chem*. 2017;138:328-342. doi:10.1016/j.ejmech.2017.06.060

24. Cabrera N, Cuesta SA, Mora JR, et al. In Silico Searching for Alternative Lead Compounds to Treat Type 2 Diabetes through a QSAR and Molecular Dynamics Study. *Pharmaceutics*. 2022;14(2). doi:10.3390/pharmaceutics14020232
25. Cabrera N, Cuesta SA, Mora JR, et al. Searching glycolate oxidase inhibitors based on QSAR, molecular docking, and molecular dynamic simulation approaches. *Sci Rep*. 2022;12(1):1-19. doi:10.1038/s41598-022-24196-4
26. Valencia J, Rubio V, Puerto G, et al. QSAR Studies, Molecular Docking, Molecular Dynamics, Synthesis, and Biological Evaluation of Novel Quinolinone-Based Thiosemicarbazones against Mycobacterium tuberculosis. *Antibiotics*. 2023;12(1). doi:10.3390/antibiotics12010061
27. Mora JR, Marrero-ponce Y. Ensemble Models Based on QuBiLS-MAS Features and Shallow Learning for the Prediction of Drug-Induced Liver Toxicity: Improving Deep Learning and Traditional Approaches. Published online 2020. doi:10.1021/acs.chemrestox.0c00030
28. Cuesta SA, Mora JR, Márquez EA. In silico screening of the drugbank database to search for possible drugs against sars-cov-2. *Molecules*. 2021;26(4). doi:10.3390/molecules26041100
29. Golbraikh A, Tropsha A. Beware of q²! *J Mol Graph Model*. 2002;20(4):269-276. doi:10.1016/S1093-3263(01)00123-1
30. Załuski M, Schabikowski J, Schlenk M, et al. Novel multi-target directed ligands based on annelated xanthine scaffold with aromatic substituents acting on adenosine receptor and monoamine oxidase B. Synthesis, in vitro and in silico studies. *Bioorganic Med Chem*. 2019;27(7):1195-1210. doi:10.1016/j.bmc.2019.02.004
31. Swinney DC. Molecular Mechanism of Action (MMoA) in Drug Discovery. *Annu Rep Med Chem*. 2011;46:301-317. doi:10.1016/B978-0-12-386009-5.00009-6
32. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform*. 2010;29(6-7):476-488. doi:10.1002/minf.201000061
33. Burlingham BT, Widlanski TS. An intuitive look at the relationship of Ki and IC50: A more general use for the dixon plot. *J Chem Educ*. 2003;80(2):214-218. doi:10.1021/ed080p214
34. Troxler T, Greenidge P, Zimmermann K, et al. Discovery of novel indolinone-based, potent, selective and brain penetrant inhibitors of LRRK2. *Bioorganic Med Chem Lett*. 2013;23(14):4085-4090. doi:10.1016/j.bmcl.2013.05.054
35. National Center for Biotechnology Information. PubChem Compound Summary for CID 5329102, Sunitinib. Published 2023. Accessed April 19, 2023. <https://pubchem.ncbi.nlm.nih.gov/compound/Sunitinib>
36. Kormuth KA, Lakdawala SS. Emerging antiviral resistance. *Nat Microbiol*. 2020;5(1):4-5. doi:10.1038/s41564-019-0639-7

37. Ramanathan K, Antognini D, Combes A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. 2020;(January):19-21.
38. Crowell BG, Benson R, Shockley D, Charlton CG. S-adenosyl-l-methionine decreases motor activity in the rat: Similarity to Parkinson's disease-like symptoms. *Behav Neural Biol.* 1993;59(3):186-193. doi:10.1016/0163-1047(93)90950-M
39. Lamango NS, Ayuk-Takem LT, Nesby R, Zhao WQ, Charlton CG. Inhibition mechanism of S-adenosylmethionine-induced movement deficits by prenylcysteine analogs. *Pharmacol Biochem Behav.* 2003;76(3-4):433-442. doi:10.1016/J.PBB.2003.08.017
40. Lamango NS, Charlton CG. Farnesyl-l-Cysteine Analogs Block SAM-Induced Parkinson's Disease-Like Symptoms in Rats. *Pharmacol Biochem Behav.* 2000;66(4):841-849. doi:10.1016/S0091-3057(00)00274-4
41. National Center for Biotechnology Information. PubChem Compound Summary for CID 5546, Triamterene. Published 2023. Accessed April 24, 2023. <https://pubchem.ncbi.nlm.nih.gov/compound/Triamterene>
42. National Center for Biotechnology Information. PubChem Compound Summary for CID 4756, Phenazopyridine. Published 2023. Accessed April 24, 2023. <https://pubchem.ncbi.nlm.nih.gov/compound/Phenazopyridine>
43. National Center for Biotechnology Information. PubChem Compound Summary for CID 5315659, Cannabigerol. Published 2023. Accessed April 24, 2023. <https://pubchem.ncbi.nlm.nih.gov/compound/Cannabigerol>
44. National Center for Biotechnology Information. PubChem Compound Summary for CID 34755, S-adenosylmethionine. Published 2023. Accessed April 24, 2023. <https://pubchem.ncbi.nlm.nih.gov/compound/S-adenosylmethionine>
45. Rosenbusch KE, Kortholt A. Activation Mechanism of LRRK2 and Its Cellular Functions in Parkinson's Disease. *Parkinsons Dis.* 2016;2016. doi:10.1155/2016/7351985
46. The University of Alberta, OMx Personal Health Analytics Inc. Drug Bank Online. Published 2023. Accessed April 24, 2023. <https://go.drugbank.com/>

APPENDIX A: QSAR STUDIES

Table A1. Best 15 models from the QSAR study

Name of the Model	Machine learning algorithm applied	Descriptive name of the model
M1	LR	JM_LR_GA_3_7
M2	LR	JM_LR_GA_4_6
M3	LR	GP_GreedSt_50_LR_BF_10_M_8
M4	IBK	LR_GenSe_245_IBK_GreedSt_7
M5	SMOreg	GP_GreedSt_50_LR_BF_10_M_8
M6	LR	GP_BF_49_LR_BF_10_M_7
M7	SMOreg	SMOreg_GenSe_225_SMOreg_GreedSt_11_M_8
M8	LR	JM_LR_GA_2_6
M9	SMOreg	GP_BF_49_LR_BF_10_M_7
M10	SMOreg	SMOreg_GreedSt_21_IBK_BF_11_M_9
M11	LR	JM_LR_GA_1_7
M12	IBK	IBK_BF_23_IBK_GreedSt_11_M_7
M13	IBK	IBK_GreedSt_11_M_7
M14	LR	SMOreg_GenSe_225_SMOreg_GreedSt_11_M_8
M15	LR	SMOreg_GreedSt_21_IBK_BF_11_M_9

The table presents the best fifteen models from the QSAR study. It provides information about the name given to easily identify the model, the machine learning algorithm applied for the CV and the name given to indicate the different characteristics of each model. Regarding the descriptive names, the first group of letters indicates the classifier of the Wrapper Subset Evaluator used and the second one is the search method applied. The numbers correspond to the amount of descriptors of the models after each multiple linear regression was performed. The letter M at the end of the names indicates that some descriptors were manually removed. It should be noted that models M1, M2, M8, and M11 were named on a slightly different basis. However, LR and GA (Genetic algorithm) still correspond to the classifier and search method used, and their last number does indicate the amount of descriptors of that model.

Table A2. Q_{CV}^2 , Q_{Ext}^2 and corresponding MAE of the best 15 models

Name of the Model	Q_{CV}^2	MAE_{CV}	Q_{Ext}^2	MAE_{Ext}
M1	0.798	0.354	0.795	0.363
M2	0.760	0.394	0.801	0.358
M3	0.873	0.310	0.754	0.425
M4	0.678	0.470	0.819	0.359
M5	0.862	0.316	0.769	0.406
M6	0.866	0.324	0.745	0.410
M7	0.725	0.422	0.848	0.323
M8	0.835	0.326	0.714	0.429
M9	0.856	0.328	0.761	0.416
M10	0.797	0.390	0.803	0.373
M11	0.842	0.331	0.782	0.372
M12	0.799	0.382	0.730	0.383
M13	0.799	0.382	0.730	0.383
M14	0.678	0.470	0.819	0.359
M15	0.758	0.424	0.821	0.375

Table A2 presents the fivefold Q_{CV}^2 and Q_{Ext}^2 for the best fifteen models with the corresponding mean absolute error for each one.

Table A3. Descriptors of models M1 and M10, their abbreviations and number of appearances in the best 2 and best 15 models

Model	Descriptor	Abbreviated name	Number of appearance in best 2 models	Number of appearance in best 15 models
M1	S_TrC_AB_nCi_3_M22(M5)_SS1_T_LG3L[1-2]_LGL[1-2]_h_MID	S_LGL_h_MID	1	1
	I50_TrF_AB_nCi_3_M21(M11)_SS7_T_KA_h_MID	I50_KA_h_MID	1	2
	AC[3]_I50_Tr_AB_nCi_3_M20(M11)_SS4_T_LG3P[2]_LGP[2]_e-v-p_MID	AC[3]_T_evp_MID	1	4
	HM_TrC_AB_nCi_3_M25(M13)_SS3_T_LGA[1.0-2.0]_p_MID	HM_LGA_p_MID	1	2
	N1_B_AB_nCi_2_SS2_H_T_KA_c-h_MAS	N1_KA_ch_MAS	1	3
	AC[2]_K_TrC_AB_nCi_3_M20(M13)_NS1_C_KA_c_MID	AC[2]_C_c_MID	1	2
M1 & M10	S_TrC_AB_nCi_3_M22(M13)_NS3_P_KA_p_MID	S_NS_p_MID	2	4
M10	AM_TrC_AB_nCi_3_M21(M15)_SS3_T_LG3P[3]_LGP[3]_p_MID	AM_LGP_p_MID	1	2
	K_TrC_AB_nCi_3_M25(M8)_SS7_T_LGA[6.0-7.0]_p_MID	K_LGA_p_MID	1	1
	AC[1]_K_TrC_AB_nCi_3_M21(M15)_NS1_C_KA_c_MID	AC[1]_KA_c_MID	1	1
	GV[5]_K_TrB_AB_nCi_3_M25(M1)_NS4_A_KA_p-s_MID	GV[5]_KA_ps_MID	1	6
	VC_TrC_AB_nCi_3_M25(M15)_SS5_T_KA_p_MID	VC_KA_p_MID	1	1
	AC[5]_K_TrC_AB_nCi_3_M21(M13)_MP6_T_KA_m_MID	AC[5]_KA_m_MID	1	1
	P3_F_AB_nCi_2_SS2_H_T_NSRW_e_MAS	P3_KA_e_MAS	1	1
	S_B_AB_nCi_2_MP5_D_LGP[1;2;6]_c-m_MAS	S_KA_cm_MAS	1	5

Table A4. Pearson coefficients of M1 model's descriptors

Descriptors	S_LGL_h_MID	S_NS_p_MID	I50_KA_h_MID	ACI3]_T_evp_MID	HM_LGA_p_MID	N1_KA_ch_MAS	ACI2]_C_c_MID
S_LGL_h_MID	1						
S_NS_p_MID	0.053	1					
I50_KA_h_MID	0.0635	-0.3647	1				
ACI3]_T_evp_MID	-0.0618	0.5313	-0.2465	1			
HM_LGA_p_MID	-0.1774	-0.4969	0.1689	-0.2401	1		
N1_KA_ch_MAS	0.1404	-0.1356	-0.1045	-0.1152	0.3298	1	
ACI2]_C_c_MID	0.0764	0.3655	-0.0107	0.0734	-0.2327	-0.0249	1

It should be noted that Table A4 represents a symmetrical matrix and, therefore, only half of it is presented to avoid redundant information.

APPENDIX B: DRUG BANK SCREENING

Table B1. Characteristics and applications of the Drug Bank drugs with predicted pIC₅₀ values greater than 9

Drug Bank ID	pIC ₅₀	Chemical Formula	Status	Characteristics and applications
DB00426	10.058	C ₁₄ H ₁₉ N ₅ O ₄	approved; investigational	-Guanine analogue used to treat herpes virus infections ⁴⁶ . -Rapidly transforms to the active ⁴⁶ . antiviral compound penciclovir ⁴⁶ . -Hepatic metabolism ⁴⁶ . -77% absorption ⁴⁶ . -Symptoms of overdose include constipation, diarrhea, dizziness, fatigue, fever, headache, nausea, and vomiting ⁴⁶ .
DB01421	11.798	C ₂₃ H ₄₅ N ₅ O ₁₄	approved; investigational	-Antibiotic used in the treatment of acute and chronic intestinal amebiasis and for the management of hepatic coma ⁴⁶ . -Poorly absorbed after oral administration ⁴⁶ .
DB01896	15.443	C ₆ H ₈ BNO ₂	experimental	-Not Available
DB02037	10.702	C ₄ H ₆ N ₄ O ₂	experimental	-Not Available
DB02797	13.032	C ₆ H ₆ BNO ₄	experimental	-Not Available
DB03505	10.124	C ₈ H ₈ N ₄ O	experimental	-Not Available
DB04008	10.032	C ₁₇ H ₁₆ N ₈ Zn	experimental	-Not Available
DB04360	11.962	C ₈ H ₇ BO ₂ S	experimental	-Not Available
DB08993	11.306	C ₂₅ H ₄₃ N ₁₃ O ₁₀	experimental	-Antitubercular agent ⁴⁶ .
DB13673	10.932	C ₁₈ H ₃₇ N ₅ O ₁₀	experimental	-Antibiotic used in the treatment of eye infections ⁴⁶ .
DB00384	9.471	C ₁₂ H ₁₁ N ₇	approved	-Potassium-sparing diuretic used in the treatment of edema and hypertension ⁴¹ .
DB00955	9.156	C ₂₁ H ₄₁ N ₅ O ₇	approved; investigational	- Antibiotic used to treat a wide variety of infections in the body ⁴⁶ .
DB01438	9.591	C ₁₁ H ₁₁ N ₅	approved	- Local anesthetic used for the relief of discomfort caused by pain, burning, and irritation ⁴² .
DB02844	9.774	C ₁₈ H ₂₉ N ₇ O ₃ S	experimental	-Not Available
DB03016	9.324	C ₁₃ H ₁₁ N ₅ O	experimental	-Not Available
DB04790	9.043	C ₂₀ H ₂₂ N ₄ O ₄	experimental	-Not Available
DB04791	9.897	C ₂₀ H ₂₂ N ₄ O ₄	experimental	-Not Available
DB04793	9.982	C ₂₀ H ₂₂ N ₄ O ₄	experimental	-Not Available

DB06915	9.686	C10H8O5	experimental	-Not Available
DB06937	9.116	C13H9NO4	experimental	-Not Available
DB07032	9.376	C14H10O3	experimental	-Not Available
DB07649	9.383	C19H23N5OS	experimental	-Not Available
DB07698	9.553	C18H14ClN5	experimental	-Not Available
DB07731	9.908	C9H10N6O	experimental	-Not Available
DB08048	9.638	C13H10N2O3	experimental	-Not Available
DB08163	9.939	C17H27N7O4	experimental	-Not Available
DB08694	9.881	C14H11N7O	experimental	-Not Available
DB08707	9.123	C17H11ClN4O	experimental	-Not Available
DB08787	9.381	C16H11Cl2N5	experimental	-Not Available
DB14734	9.408	C21H32O2	experimental	-A natural product found in Cannabis sativa and Helichrysum species ⁴³ . -It enhances appetite and acts as an anti-inflammatory, antioxidant and neuroprotective agent ⁴³ .
DB00118	9.726	C15H22N6O5S	approved; investigational; nutraceutical	- Physiologic methyl radical donor used as anti-inflammatory and applied in the treatment of depression, liver disorders, fibromyalgia, and osteoarthritis. It has also been used as dietary supplement for the support of bone and joint health, and as a mood and emotional regulator ⁴⁴ .
DB09004	9.101	C14H22ClNO	withdrawn	-Cough suppressant that is withdrawn from the US and EU markets. Clobutinol may prolong the QT interval. In 2007, Clobutinol was determined to cause cardiac arrhythmia in some patients ⁴⁶ .

APPENDIX C: MOLECULAR DOCKING STUDIESTable C1. pIC₅₀ values and docking scores of the dataset molecules

Name	pIC₅₀	Docking scores (kcal/mol)
3_12	7.509	-9.2
1_33	8.398	-8.8
1_29	6.387	-8.8
1_30	7.022	-8.7
1_35	7.921	-8.6
1_28	6.770	-8.6
1_31	8.699	-8.5
3_13	8.155	-8.5
3_7	8.097	-8.4
1_7	6.420	-8.4
3_3	7.854	-8.3
3_11	8.398	-8.2
2_33	8.000	-8.2
1_34	7.824	-8.2
2_36	7.699	-8.2
2_34	7.699	-8.2
1_27	6.959	-8.2

1_26	6.886	-8.2
2_37	5.272	-8.2
1_32	7.959	-8.1
2_16	6.959	-8.1
2_38	6.886	-8.0
1_1	5.310	-7.9
3_6	8.523	-7.8
3_4	8.155	-7.8
3_2	7.959	-7.8
3_1	7.337	-7.8
1_3	7.013	-7.8
2_23	5.156	-7.8
2_48	5.058	-7.8
Sunitinib	7.553	-7.7
2_39	6.553	-7.7
2_42	6.310	-7.7
3_10	8.398	-7.6
1_4	6.824	-7.6
1_2	6.409	-7.6
2_43	6.119	-7.6
2_40	5.824	-7.6
2_11	4.812	-7.6
3_5	8.046	-7.5
1_10	5.796	-7.5
2_8	5.631	-7.5
2_10	5.498	-7.5
2_44	5.921	-7.4
1_20	5.721	-7.4
2_45	5.438	-7.4
2_21	5.417	-7.4
1_18	7.174	-7.3
1_9	6.409	-7.3
1_39	6.114	-7.3
1_6	6.081	-7.3
2_47	5.876	-7.3
2_19	5.873	-7.3
2_4	5.438	-7.3
2_49	5.220	-7.3
2_12	5.138	-7.3
1_23	7.066	-7.2
2_2	5.664	-7.2
2_13	5.421	-7.2
1_19	6.886	-7.1

1_17	6.721	-7.1
1_11	6.387	-7.1
2_20	5.770	-7.1
1_25	6.602	-7
2_50	6.260	-7
1_8	5.523	-7
1_12	6.276	-6.8
1_24	6.770	-6.7
1_15	5.620	-6.5
2_22	5.240	-6.5
1_16	6.523	-6.4
1_22	7.167	-6.3
1_14	6.056	-6.3
1_13	5.328	-6.3
1_21	6.357	-6

Table C1 presents the pIC₅₀ values and docking scores of all the molecules from the dataset in increasing order of affinity energy.

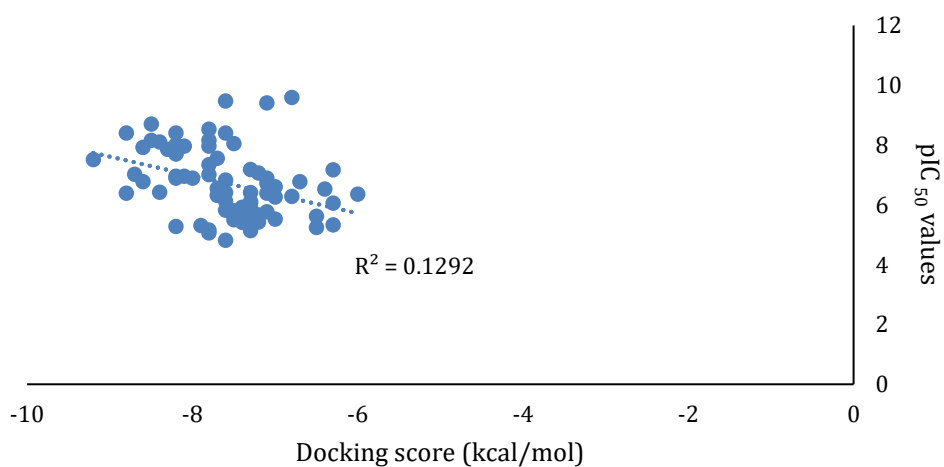


Figure C1. Correlation between pIC₅₀ values and docking scores of the dataset inhibitors

