Ministry of Higher Education and Scientific Research

Hassiba Benbouali University of Chlef

Faculty of Technology

Process Engineering Department

THESIS

PRESENTED FOR A MASTER DEGREE

FIELD: Process Engineering

SPECIALTY: Pharmaceutical Process Engineering

Presented by:

MEKHANE Selma

Theme

Targeting Central Nervous System (CNS) diseases with bioactive compounds using network pharmacology and screening tools

Proposed jury members:

MILOUDI Khaled	MCB	University of Chlef	President
OTMANINE Khaled	MCB	University of Chlef	Supervisor
AYADI Khadidja	MCB	University of Chlef	Examiner

Academic year: 2024 /2025

GRATITUDE

First and foremost, all praise and gratitude belong to Allah, the Most Merciful, for bestowing upon us the courage, perseverance, and unwavering patience to complete this journey with joy and determination. Without His blessings, this accomplishment would not have been possible.

We extend our deepest appreciation to our esteemed supervisor, Mr. Otmanine Khaled, whose invaluable guidance, unwavering support, and steadfast belief in our abilities illuminated our path. His wisdom and encouragement were the cornerstones that allowed us to navigate this journey with prudence and confidence.

Our sincere thanks to Mr. Hentabli, whose patience, time, and insightful guidance enriched every step of this process. To Dr. Hammoudi Mounir, we are grateful for his wisdom and mentorship, which profoundly shaped our work. We also express our heartfelt gratitude to Mr. H. Bendriss, whose dedication and encouragement inspired us throughout the realization of this endeavor.

To the distinguished members of the jury, Mr Miloudi and Ms Ayadi we are honored by your presence and humbled by the opportunity to present this modest work for your consideration. Your time and expertise are deeply appreciated.

Finally, to all those who supported us near and far with kindness, encouragement, or a helping hand, we offer our warmest gratitude. Your presence, whether seen or unseen, has been a beacon of strength throughout this journey.

This achievement is as much yours as it is ours.

DEDICATION

To myself,

This journey has been one of perseverance, growth, and resilience. Through late nights, challenges, and moments of doubt, I kept pushing forward proving to myself that dedication and hard work truly pay off.

I dedicate this achievement to the version of me and my dear family who never gave up, who believed in the dream even when the path was unclear. May this milestone be a reminder that I am capable of greatness, and may it inspire me to keep reaching for new heights.

Here's to me for earning this degree, for overcoming every obstacle, and for the incredible journey ahead.

To My Beloved Family,

This milestone is not mine alone it is woven from your endless love, sacrifices, and unwavering belief in me. You were my strength when I doubted, my comfort when I struggled, and my joy when I triumphed.

No words can truly capture my gratitude, but this achievement stands as a testament to the love and lessons you've poured into me. May I continue to make you proud, as you have made me eternally grateful to call you mine.

With all my heart,
SELMA

SUMMARY

ABRIVIATION

A2AA-R: The adenosine A2A receptor

AI : Artificial intelligence

ANNs: Artificial neural networks

BBB: Blood-brain barrier

CoMFA: Comparative molecular field analysis

DFT: Density functional theory

GABA-A: Gamma-aminobutyric acid type A receptors

HF: Hartree-Fock

HIA: Human intestinal absorption

HOMO: Highest Occupied MO

KI : Inhibition constant

LBs : Lewy bodies

LD50: Lethal dose 50

LNs: Lewy neurites

LUMO: Lowest Unoccupied MO

MLR: Multiple linear regression

MOE: Molecular Operating Environment

MPTP: 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine

MW: Molecular weight

NrF2: Nuclear factor erythroid 2–related factor 2

PD: Parkinson disease

PDB: Protein data base

PIC50: Negative logarithm of the half-maximal inhibitory concentration (IC50)**.**

PLIP : Protein ligand interaction profiler

QSAR: Quantitive Structure-Activity Relationship

QSPR: Quantitive Structure- Property Relationships

RMSE: Root mean square error

STRING: Search Tool for the Retrieval of Interacting Genes/Proteins

SVM: Support vector machine

SVR: Support vector regression

List of tables

List of figures

Abstract

Today we hear and see the increase in Parkinson disease, which have become among the most common diseases in the world. Therefore, we targeted some proteins that are related to the disease. Our goal in this study is to inhibit all of the A2AA-R, GABA-A, NrF2, and α-synuclein. After examining various plants from Chlef in Algeria, we selected *salvia officinalis* and *curcuma longa*. After that, we made the extraction of *curcuma longa* with ethanol, and we had obtained good yield.

Our In Silico study demonstrated that there is an inhibitory combination of the bisdemethoxycurcumin compound with A2AA-R,with a binding capacity estimated at ΔG = -9.1 kcal/mol, and the ar-turmerone compound with the A2AA-R with a binding capacity estimated at ΔG = -8.6 kcal/mol. α -turmerone had make a cohesion with NrF2 with energy estimated at ΔG=-7.4kcal/mol. Consequently, makes the bisdemethoxycurcumin-A2AA-R interaction (ΔG = -9.1 kcal/mol) demonstrates the strongest binding affinity among the tested compounds.

Our QSAR study approach was based on using two methods, multiple linear regression (MLR), support vector regression (SVR). SVR demonstrated superior predictive accuracy (Q^2 = 0.992, R^2 train = 0.989) compared to MLR (Q^2 = 0.559, R^2 train = 0.766).

After the comparison between bisdemethoxycurcumin and l-dopa , our findings suggest bisdemethoxycurcumin (BDMC) offers dual benefits comparable binding affinity to key targets (ΔG = -9.1 kcal/mol) while additionally addressing PD's core pathologies through antioxidant and anti-inflammatory mechanisms. This multifunctional activity positions BDMC as a potential disease-modifying adjunct to conventional therapy.

Key word: *curcuma longa*, *salvia officinalis*, A2AA-R, GABA-A, NrF2, α-synuclein.

ملخص

نرى اليوم زيادة في مرض الباركنسون الذي اصبح من أكثر الأمراض شيوعا في العالم , لذلك إستهدفنا بعض البروتينات المرتبطة بهذا المرض . هدفنا من هذه الدراسة هو تثبيط جميع البروتينات المتعلقة بالمرض ,NrF2 ,GABA-A A2AAR, alpha-synuclein . بعد فحص نباتات مختلفة في مدينة الشلف – الجزائر , إخترنا نبتة المريمية و الكركم و بعدها أجرينا إستخلاص مكونات الكركم باستخدام الإيثانول و حصلنا على مردود جيد .

أظهرت دراسة insilico وجود مزيج مثبط لمركب bisdemethoxycurcumin مع A2AARبسعة إرتباط تقدر ب -9.1 kcal/mol Δ G و مركب ar-turmerone مع A2AAR بسعة Δ 8.6 kcal/mol- = G, مركب alpha-turmerone مع NrF2 بسعة Δ- = G 7.4 kcal/mol . أظهرت SVR دقة تنبؤية متفوقة (Q^2 = 0.992 , R^2 train = 0.989) .

بعد المقارنة تشير نتائجنا إلى أن البيسديمتوكسي كركمين (BDMC) يقدم فوائد مزدوجة , حيث يقارب إرتباطه بالأهداف الرئيسية (ΔG = -9.1 كيلو كالور /مول) , بالإضافة إلى معالجة الأمراض الأساسية لمرض الباركنسون من خلال الاليات مضلدة للأكسدة و الإلتهابات . هذا النشاط متعدد الوظائف يمكن البيسديميتوكسي كركمين من أن يكون مكملا محتملا للعلاج التقليدي معدلا للمرض .

الكلمات المفتاحية: المريمية , الكركم، A2AAR, NrF2 ,Alpha-synuclein ,GABA-A

RÉSUMÉ

Aujourd'hui, nous entendons et voyons l'augmentation de la maladie de Parkinson, qui est devenue l'une des maladies les plus courantes au monde. Par conséquent, nous avons ciblé certaines protéines liées à la maladie. Notre objectif dans cette étude est d'inhiber toutes les A2AA-R, GABA-A, NrF2 et α-synucléine. Après avoir examiné diverses plantes de Chlef en Algérie, nous avons sélectionné la sauge officinale et le curcuma longa. Ensuite, nous avons procédé à l'extraction du curcuma longa avec de l'éthanol, et nous avons obtenu un bon rendement.

Notre étude in silico a montré qu'il existe une combinaison inhibitrice du composé bisdéméthoxycurcumine avec le récepteur A2AA-R, avec une capacité de liaison estimée à ΔG = -9,1 kcal/mol, et du composé ar-turmerone avec le récepteur A2AA-R avec une capacité de liaison estimée à ΔG = -8,6 kcal/mol. L'α-turmerone a créé une cohésion avec NrF2 avec une énergie estimée à ΔG = -7,4 kcal/mol. En conséquence, l'interaction bisdéméthoxycurcumine-A2AA-R (ΔG = -9,1 kcal/mol) démontre la plus forte affinité de liaison parmi les composés testés.

Notre approche d'étude QSAR était basée sur l'utilisation de deux méthodes, la régression linéaire multiple (RLM) et la régression par vecteurs de support (RVS). La RVS a montré une précision prédictive supérieure (Q^2 = 0,992, R^2 train = 0,989) par rapport à la RLM (Q^2 = 0,559, R^2 train = 0,766).

Après la comparaison entre le bisdéméthoxycurcumin et la l-dopa, nos résultats suggèrent que le bisdéméthoxycurcumin (BDMC) offre des avantages doubles : une affinité de liaison comparable aux cibles clés (ΔG = -9,1 kcal/mol) tout en s'attaquant également aux pathologies fondamentales de la maladie de Parkinson par des mécanismes antioxydants et anti-inflammatoires. Cette activité multifonctionnelle positionne le BDMC comme un complément potentiel modifiant la maladie à la thérapie conventionnelle.

Mot-clé : *curcuma longa, salvia officinalis*, A2AA-R, GABA-A, NrF2, α-synucléine.

GENERAL INTRODUCTION

It took almost 100 years before a meaningful advance occurred in any basic science understanding of Parkinson disease (PD) following James Parkinson's description in 1817. In current literature, Parkinson's disease is a progressive neurodegenerative disorder characterized by the loss of dopaminergic neurons in the substantia nigra, leading to motor dysfunction, tremors, and cognitive decline. The disease is associated with the accumulation of alpha-synuclein (α-syn) aggregates, neuroinflammation, and oxidative stress, which contribute to neuronal death. Despite advances in understanding PD pathogenesis, existing treatments such as levodopa and dopamine agonists only alleviate symptoms temporarily and are often accompanied by severe side effects, including dyskinesias and drug resistance.

This study employs a bifurcated research design comprising theoretical and computational components. The initial theoretical investigation focuses on elucidating the pathophysiological mechanisms underlying Parkinson's disease, with particular emphasis on key molecular targets including adenosine A2A receptors (A2AAR), γ-aminobutyric acid type A receptors (GABA-AR), nuclear factor erythroid 2-related factor 2 (Nrf2), and α-synuclein aggregates. This analysis will systematically examine the structural and functional roles of these biomarkers in disease progression and their potential as therapeutic targets.

Concurrently, we will perform a comprehensive phytochemical evaluation of two medicinal plants *Curcuma longa* (turmeric) and *Salvia officinalis* (sage) selected for their documented neuroprotective properties and putative receptor inhibition capabilities. The methodological approach incorporates advanced computational techniques, including molecular docking simulations to characterize ligand-receptor interactions, complemented by quantitative structure-activity relationship (QSAR) modeling.

In the experimental phase of this investigation, we implemented an integrated computational approach combining advanced molecular modeling techniques with artificial intelligence-driven algorithms to systematically evaluate the therapeutic potential of selected phytochemical compounds. Our methodology centered on three principal computational strategies: molecular docking simulations to predict ligand-protein binding conformations and affinities, density functional theory (DFT) calculations at the B3LYP/6-311G++(d,p) level using Gaussian 09W software for quantum chemical analysis of electronic properties (including HOMO-LUMO energies and molecular electrostatic potentials), and machine learning-enhanced quantitative structure-activity relationship (QSAR) modeling to predict biological

activity and optimize pharmacokinetic properties. These computational analyses were complemented by comprehensive ADMET (absorption, distribution, metabolism, excretion, and toxicity) profiling to assess drug-likeness and safety parameters. All simulations were conducted on high-performance computing platforms, enabling efficient and accurate virtual screening while significantly reducing the temporal and financial expenditures associated with conventional experimental approaches. This multidisciplinary computational framework demonstrates an effective strategy for accelerating the identification and optimization of novel therapeutic candidates while maintaining rigorous scientific standards.

In this study we aim to identify a process and a selective inhibitor targeting A2AA- R, GABA-A, NrF2, alpha-synuclein using molecular modeling and screening tools to prevent the disease.

Our research is divided into two parts:

a- Modeling
b- Simulation

After the research we identified two plants used to exhibit the highest inhibition rates against the selected receptors: curcuma longa, salvia officinalis

We incorporate artificial intelligence (AI) and pharmacoinformatic as innovative screening approaches, facilitating the development and the discovery of new drugs. any process that contributes to drug development or discovery falls under the domain of pharmaceutical process engineering

The primary objective of this research is to explore the efficacy of medicinal plants found in Chlef in the treatment of Parkinson’s disease, furthermore, Assessing the pharmacological potential of botanicals using AI-driven analysis and high-throughput screening tools computational screening platforms, and lastly Investigating the potential of plant-derived compounds as a therapeutic alternative to L-dopa medication.

Can targeted inhibition of key proteins including adenosine A2A receptors (A2AAR), nuclear factor erythroid 2-related factor 2 (Nrf2), gamma-aminobutyric acid type A receptors (GABA-A), and alpha-synuclein constitute a viable therapeutic strategy for the prevention and treatment of Parkinson's disease (PD)? Furthermore, can artificial intelligence (AI)-driven drug discovery platforms overcome current limitations in PD therapeutics to identify novel, disease-modifying agents?

CHAPTER I
BIBLIOGRAPHIC STUDY

Introduction

Parkinson’s Disease (PD) is a neurological condition that leads to involuntary or uncontrollable movements, including tremors, muscle rigidity, and challenges with balance and coordination. Parkinson's disease happens when the nerve cells in the brain that produce dopamine, a neurotransmitter essential for smooth and coordinated muscle movements, become damaged. The resulting shortage of dopamine causes the characteristic symptoms of the disease [1].

Parkinson's disease is a progressive condition, with the average age of onset around 55. The incidence of PD increases significantly with age, rising from 20 per 100,000 in the general population to 120 per 100,000 by age 70. In approximately 95% of PD cases, there is no clear genetic link, referred to as "sporadic" PD. In the remaining cases, the disease is inherited. As the disease progresses, symptoms become more severe. Before the introduction of levodopa, the mortality rate for individuals with PD was three times higher than that of age-matched individuals without the disease [2].

Although Parkinson's disease is widely recognized for its motor symptoms, it also presents a variety of non-motor features, such as cognitive decline, depression, sleep disturbances, and autonomic dysfunction, which can significantly impact an individual’s quality of life. While the exact cause of Parkinson’s disease remains unknown, both genetic and environmental factors are believed to contribute to its development [3].

Parkinson's disease is typically diagnosed based on clinical symptoms and medical history, as there are no definitive tests for early stages. Treatment options focus on managing symptoms and improving quality of life, but there is currently no cure or treatment that can stop the disease’s progression. Ongoing research is aimed at better understanding its causes, developing more effective treatments, and ultimately finding a cure [3].

I-1- Symptoms of parkinson disease

Parkinson's disease symptoms can vary from person to person, and early signs may be subtle and go unnoticed. Typically, symptoms start on one side of the body and gradually affect both sides, with one side often being more severely affected. Some symptoms of Parkinson's can resemble those of other conditions [4]. (see figure 1)

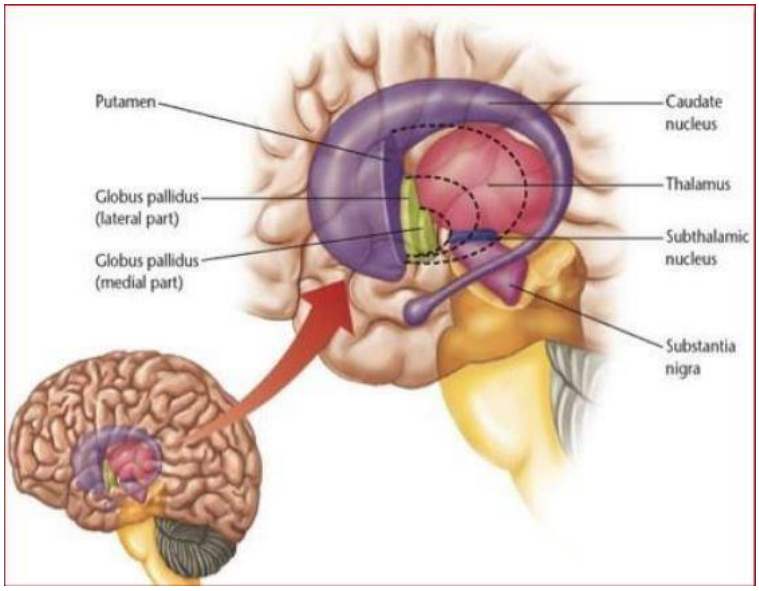


Figure 1 : Anatomy of the basal ganglia affected in PD [3]

Common symptoms of Parkinson's disease include :

- **Tremor:** This rhythmic shaking usually begins in the hands or fingers but may also start in the foot or jaw. A common type of tremor is the "pill-rolling" tremor, where you rub your thumb and forefinger back and forth. Tremors may occur when your hand is at rest or during stress and tend to lessen during activity.
- **Slowed Movement (Bradykinesia):** Parkinson’s can slow your movements, making everyday tasks more challenging, such as getting up from a chair, showering, or dressing. You may also notice reduced facial expressions and difficulty blinking.
- **Rigid Muscles:** Muscles can become stiff and painful, causing limited range of motion. This stiffness may result in short, jerky movements in your arms or other parts of the body.
- **Postural Instability and Balance Problems:** You may develop a stooped posture and experience frequent falls or difficulty maintaining balance.
- **Loss of Automatic Movements:** Movements that you normally do automatically, such as blinking, smiling, or swinging your arms while walking, may become less frequent or difficult.
- **Speech Changes:** Your voice may become softer, more rapid, slurred, or monotone. You may also pause before speaking, or your speech may lose its usual rhythm and inflection.

- **Writing Changes:** Writing may become smaller, cramped, or harder to read, a condition known as micrographia.
- **Non-Motor Symptoms:** These may include depression, anxiety, constipation, sleep disturbances, and other issues such as acting out dreams, frequent urination, a reduced sense of smell, memory problems, and excessive fatigue [4].

I-2- Treatment of PD

L-DOPA (levodopa) (see figure 2) remains the gold-standard pharmacological treatment for Parkinson's disease (PD), primarily addressing motor symptoms by restoring dopamine levels in the striatum. However, its effects on cognition are complex and often paradoxical. While L-DOPA has been demonstrated to enhance certain cognitive functions particularly those reliant on dopamine-dependent frontostriatal circuits, such as working memory, task-switching, and reward-based learning it can also induce detrimental cognitive and psychiatric side effects. These range from subtle executive deficits to severe, disabling complications like psychosis, impulse control disorders (e.g., pathological gambling or hypersexuality), and even addiction-like behaviors [5].

The apparent unpredictability of L-DOPA's cognitive effects stems from several factors. First, dopamine's role in cognition is highly region-specific: optimal levels in the prefrontal cortex (PFC) enhance executive functions, whereas excessive striatal dopamine may impair procedural learning or promote compulsive behaviors. Second, disease progression alters baseline neurochemistry; early-stage PD patients with relatively intact PFC circuitry may benefit cognitively from L-DOPA, whereas advanced-stage patients, whose PFC dopamine receptors are already saturated or whose pathology extends beyond dopaminergic systems, may experience diminishing returns or worsening deficits. Third, individual differences in drug metabolism, receptor sensitivity, and comorbid neuropathology (e.g., cholinergic or noradrenergic degeneration) further modulate responses [5],[6].

Figure 2: l-dopa medication [7]

I-3- Causes of PD

In Parkinson's disease, nerve cells in the brain, known as neurons, gradually deteriorate or die. Many of the symptoms of Parkinson's are caused by the loss of neurons that produce dopamine, a crucial chemical messenger in the brain [8].

The decrease in dopamine leads to irregular brain activity, which results in movement problems and other symptoms associated with Parkinson's disease. People with Parkinson's also experience a loss of norepinephrine, another chemical messenger that regulates essential body functions, such as blood pressure [8].

The exact cause of Parkinson's disease remains unknown, but several factors appear to contribute, including:

- **Genetic Factors:** Specific genetic changes have been linked to Parkinson’s, although these are rare unless multiple family members are affected by the disease.
- **Environmental Factors:** Exposure to certain toxins or environmental factors may increase the risk of developing Parkinson's disease later in life. For example, MPTP, a substance found in some illegal drugs marketed as "synthetic heroin," has been associated with Parkinson’s-like symptoms. Other potential environmental risks include exposure to pesticides and drinking well water. However, no environmental factor has been proven to directly cause Parkinson's.

Several changes occur in the brains of people with Parkinson's disease, and researchers are actively investigating why these changes happen and their roles in the disease. These changes include:

- **Lewy Bodies:** Clumps of abnormal proteins found in the brain, called Lewy bodies, are associated with Parkinson's disease. Researchers believe these protein clumps may hold critical insights into the cause of Parkinson’s.
- **Alpha-synuclein in Lewy Bodies:** Alpha-synuclein, a protein found in all Lewy bodies, forms in clumps that cells are unable to break down. This protein is a focal point in current Parkinson’s research, as it has been detected in the spinal fluid of individuals who later develop Parkinson's disease.
- **Mitochondrial Changes:** Mitochondria, the energy-producing structures within cells, are altered in people with Parkinson's disease. These changes can lead to cell damage, and similar mitochondrial abnormalities have been found in the brains of individuals with the disease.

I-4- Mechanisms and pathophysiology

a. Neuropathology

The hallmark pathological features of PD include selective neuronal loss in specific regions of the substantia nigra and widespread intracellular accumulation of α-synuclein protein. While neither the degeneration of pigmented dopaminergic neurons in the substantia nigra nor the presence of α-synuclein aggregates is exclusive to PD, the combination of these two neuropathological changes is diagnostic for idiopathic PD [9].

Unlike conditions marked by generalized brain atrophy, PD involves neurodegeneration in specific neuronal populations within distinct brain regions. In early-stage PD, neuronal loss is primarily localized to the ventrolateral substantia nigra, sparing other midbrain dopaminergic neurons. However, by end-stage disease, the degeneration becomes more widespread. The significant depletion of dopaminergic neurons even in early PD suggests that pathological changes precede motor symptom onset—a finding supported by recent clinicopathological studies [9].

The second key neuropathological feature is the abnormal accumulation of α-synuclein in neuronal cytoplasm across multiple brain regions. Lewy bodies, the most well-

known α-synuclein aggregates, were first identified over a century ago. Advances in histopathological techniques have since revealed a broader spectrum of α-synuclein inclusions [9].

Lewy pathology typically begins in cholinergic and monoaminergic brainstem neurons, as well as olfactory system neurons, before spreading to limbic and neocortical regions as the disease progresses. Notably, in cases with concurrent Alzheimer's pathology, α-synuclein deposition follows a distinct pattern, predominantly affecting limbic areas [9].

I-5- The definition of receptors

1 A2AAR (A2A adenosine receptor)

PDB ID : 2YDO

The A2A adenosine receptor (A2AAR) gene is located on chromosome 22q11.23 and consists of multiple exons that encode alternative transcripts. which are highly responsive to changes in the extracellular environment... The A2A adenosine receptor (A2AAR) plays a significant role in Parkinson's disease (PD), particularly in the regulation of motor function and neuroinflammation. A2AAR is highly expressed in the striatum, where it modulates dopaminergic neurotransmission by interacting with dopamine D2 receptors [10] .(see figure 3)

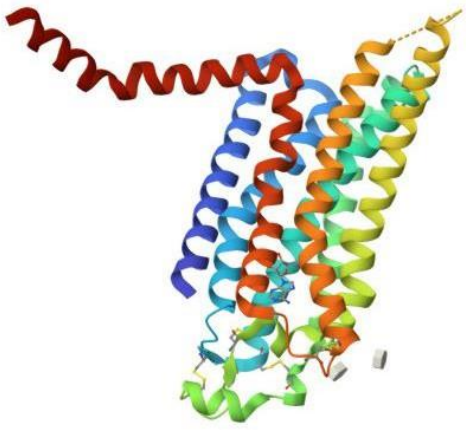


Figure 3: structure of A2AA-R

- **α-synuclein**

  PDB ID : 1XQ8

α-synuclein is a key component of the histopathological hallmarks found in various neurodegenerative disorders, collectively known as synucleinopathies. Alpha-synuclein has been detected in brainstem and cortical Lewy bodies (LBs) and Lewy neurites (LNs) using antibodies targeting the N-terminal, C-terminal, and central domains of the protein [11]. (see figure 4)

Figure 4: structure of α-synuclein

- **NrF2**

  PDB ID : 5WG1

Nuclear factor-erythroid 2-related factor 2 (Nrf2) is a 605-amino acid protein that was first cloned in 1994. It belongs to the Cap ‘n’ Collar (CNC) family of transcription factors and features a basic leucine zipper (bZIP) domain. Nrf2 consists of six highly conserved structural domains, including the Neh1 domain, which plays a crucial role in dimerization with DNA and facilitates interactions with broad-complex, tram track, and bric-a-brac (BTB) and CNC homolog (Bach) proteins. The Neh2 domain, located at the N-terminus, contains multiple lysine

residues and features a high-affinity ETGE motif along with a lower-affinity DLG motif [12]. (see figure 5)

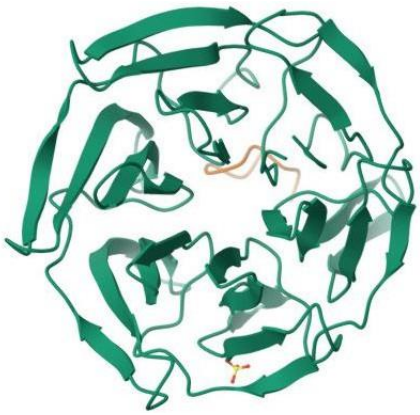


Figure 5 : structure of NrF2

- **GABA-A receptor**

  PDB ID : 6DW1

The GABA_A receptor, a key element of the inhibitory neurotransmission system, plays a significant role in the pathophysiology of Parkinson's disease (PD). Upon activation, this receptor facilitates the influx of chloride ions into neurons, resulting a decrease in neuronal excitability. The GABA_A receptor is crucial in the neurobiological processes involved in Parkinson's disease, and its modulation offers a promising strategy for therapeutic intervention [13]. (see figure 6)

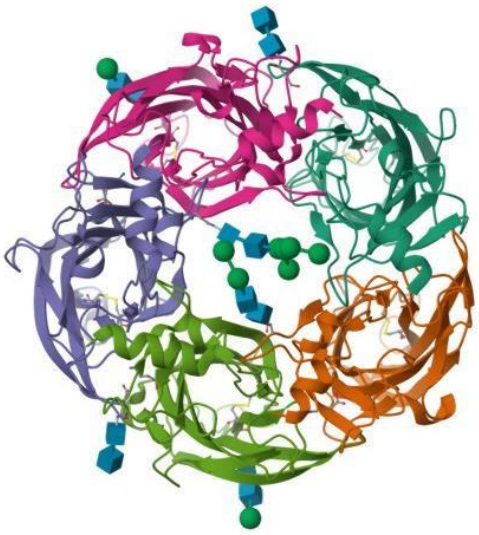


Figure 6: stucture of GABA-A receptor

I-9- The definition of plants

1 *Salvia officinalis :* مريمية

Salvia officinalis, commonly known as sage, has long been valued for its medicinal properties, particularly its antioxidant and anti-inflammatory effects. Recent research has investigated its potential neuroprotective role, especially in neurodegenerative disorders like Parkinson's disease (PD). The therapeutic benefits of Salvia officinalis in neurological conditions are largely attributed to its abundance of bioactive compounds, including polyphenols such as caffeic acid and rosmarinic acid. These compounds possess strong antioxidant properties that may help reduce oxidative stress, a key factor in the progression of PD [14]. (see figure 7)

Figure 7: a representative picture of salvia officinalis [14]

2 *Curcuma longa :* الكركم

Curcuma longa, widely known as turmeric, contains curcumin as its primary active compound, which has been extensively researched for its therapeutic potential, particularly in neuroprotection for Parkinson's disease (PD). Curcumin possesses strong antioxidant and anti-inflammatory properties, making it valuable in counteracting oxidative stress and inflammation linked to PD progression. Notably, its ability to cross the blood-brain barrier enhances its potential effectiveness in treating neurological disorders [15]. (see figure 8)

Figure 8 : a representative picture of curcuma longa [15]

I-10- Major compounds of the plants used

The major compounds of salvia officinalis and curcuma longa are represented in the following figures (see figure 9,10)

alpha,beta-thujone 1,8-cineole camphor

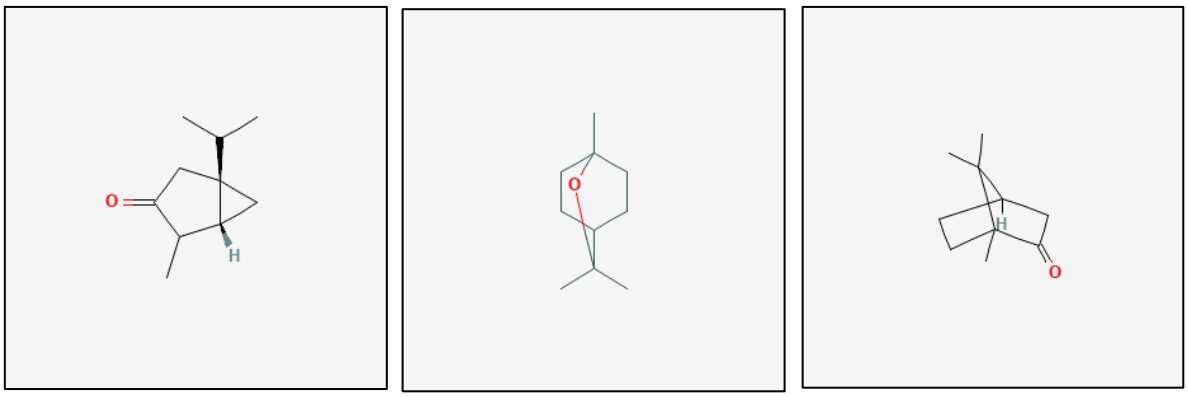


Figure 9 : chemical structure of salvia officinalis compounds studied

bisdemethoxycurcumin ar-turmerone Alpha-turmerone

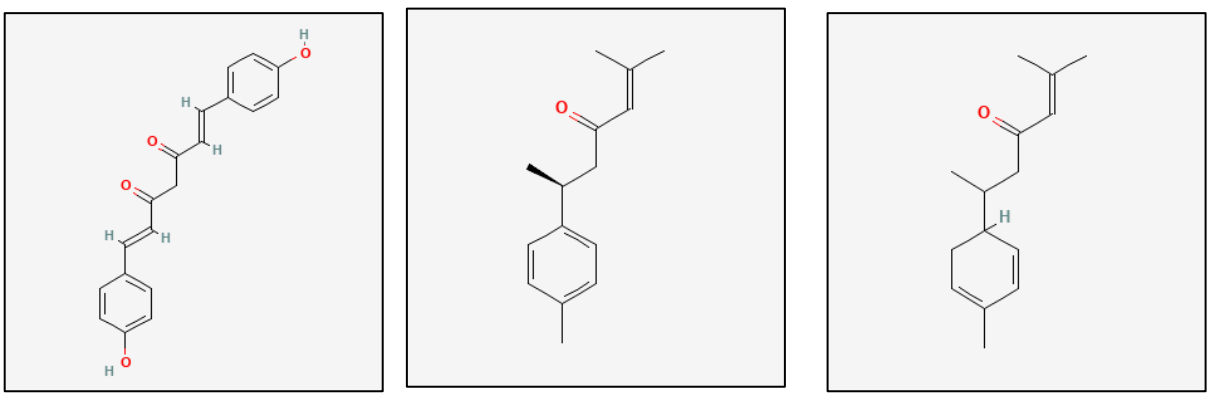


Figure 10: chemical structure of curcuma longa compounds studied

CHAPTER II

Molecular docking

II-1- Principle of molecular docking

The principles of molecular docking involve the process of predicting how a ligand molecule interacts with a target protein's active site in three-dimensional (3D) space to which we will add a very important step, which is optimization in DFT method (see figure11). Here's an overview: Docking Process, Affinity Prediction, Precise Location, Prediction, Impact of Ligand Diversity, and Orientation Variability. Overall, molecular docking is a powerful computational technique used to predict ligand-protein interactions, assess binding affinity, and elucidate the structural basis of molecular recognition. It plays a crucial role in drug discovery, virtual screening, and structure-based design of bioactive compounds [16,17].

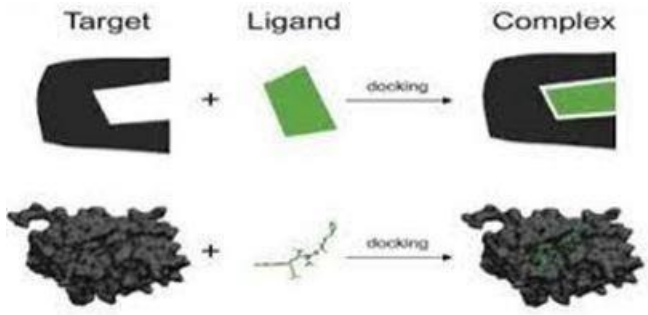


Figure 11: Principle of molecular docking [16]

II-2- The role of molecular optimization

Geometry optimization using density functional theory (DFT) is a computational method used to determine the most stable configuration of atoms within a molecule or a material (see figure2). The goal is to find the arrangement of atoms that corresponds to the lowest possible energy state, often referred to as the ground state. In DFT-based geometry optimization, the positions of atoms within the molecular or material structure are adjusted iteratively until a minimum in the total energy of the system is reached. This process involves solving the Schrödinger equation, which describes the behavior of electrons in the system, using a density functional approach. During the optimization process, the positions of the atoms are adjusted in small increments, and at each step, the total energy of the system is calculated. Optimization algorithms, such as the conjugate gradient method or the quasi-Newton method, are typically employed to efficiently search for the minimum energy configuration. Once the optimization converges to a stable configuration, the resulting geometry represents a local minimum on the potential energy surface, indicating the most stable arrangement of atoms

under the given computational conditions. This optimized geometry can then be used for further analysis, such as predicting molecular properties, simulating spectroscopic properties, or exploring chemical reactivity [16][17]. (see figure 12)

Figure 12 : gaussian view optimization method DFT

- Gaussian 09W is a powerful computational chemistry software widely used for molecular optimization, employing quantum mechanical methods to determine the most stable molecular configurations through energy minimization. It utilizes various theoretical approaches, including Density Functional Theory (DFT) with functionals like B3LYP for balanced accuracy and efficiency, Hartree-Fock (HF) theory for fundamental wavefunction-based optimization, and higher-level post-Hartree-Fock methods such as MP2 and CCSD(T) for increased precision. Additionally, semi-empirical methods like PM6 and AM1 provide faster approximations for larger molecular systems. Gaussian 09W enables key optimization tasks such as geometry refinement to achieve minimum-energy structures, transition state identification for reaction pathways, and conformational analysis to determine the most stable 3D arrangements. Frequency calculations further validate optimized structures by confirming the absence of imaginary frequencies for stable conformers or detecting transition states with a single imaginary frequency. The typical workflow involves preparing input files with specified methods and basis sets (e.g., B3LYP/6-31G*), executing optimization jobs, and analyzing output data for convergence and thermodynamic properties. Gaussian 09W's robust algorithms and wide applicability make it indispensable in computational chemistry for drug design, materials science, and reaction mechanism studies. The software's integration of advanced theoretical models with practical optimization techniques bridges the gap between experimental observations and virtual simulations, enhancing predictive accuracy in molecular research [18].

II-3- The basis of simulation methods: integrating experimental and virtual domains

Simulation techniques form a critical link between experimental (physical) and virtual (computational) approaches, enabling researchers and engineers to validate, optimize, and predict system behavior with greater efficiency and accuracy. The experimental domain provides essential real-world data, empirical validation, and physical constraints, ensuring that simulations remain grounded in observable reality. Meanwhile, the virtual domain offers cost-effective, scalable, and highly controlled environments where complex scenarios can be tested without the limitations of physical prototypes. Key methodologies facilitating this integration include hybrid simulation (e.g., Hardware-in-the-Loop and Digital Twins), which combines physical components with computational models for real-time testing ; model validation and calibration**,** where experimental data fine-tunes virtual models to enhance predictive accuracy; and high-fidelity computational modeling (e.g., Finite Element Analysis or Computational Fluid Dynamics), which replicates real-world physics with precision . This synergy not only accelerates innovation but also reduces development costs and risks, making it indispensable in fields like aerospace, robotics, and biomedical engineering. For instance, Digital Twins—virtual replicas of physical systems updated with real-time data—exemplify how tightly coupled experimental and virtual domains can optimize performance and enable predictive maintenance. Thus, the integration of these domains represents a paradigm shift in how complex systems are designed, tested, and deployed [19][20][21].

II-4- Tools

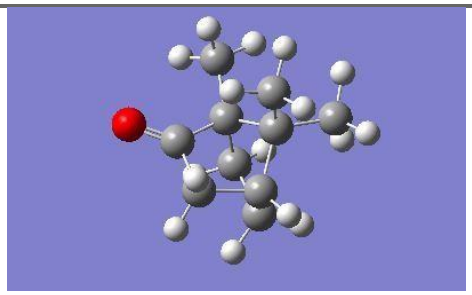
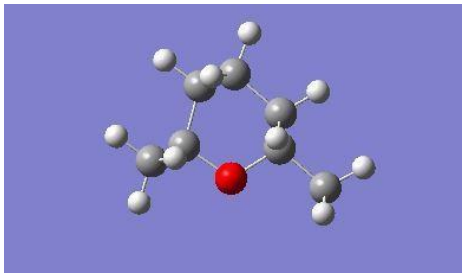
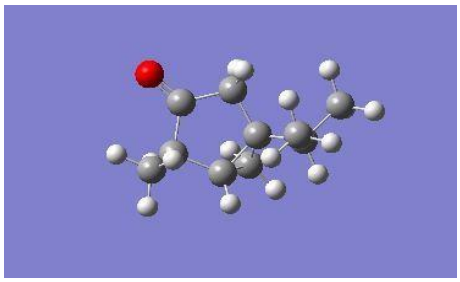
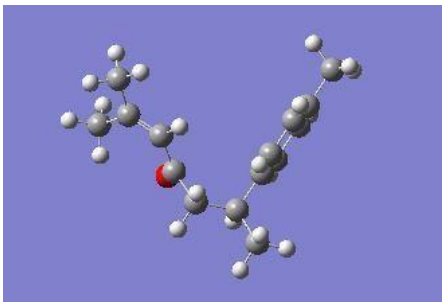
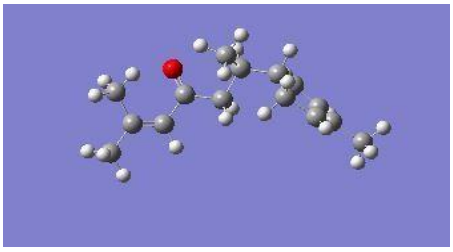
a) PubChem

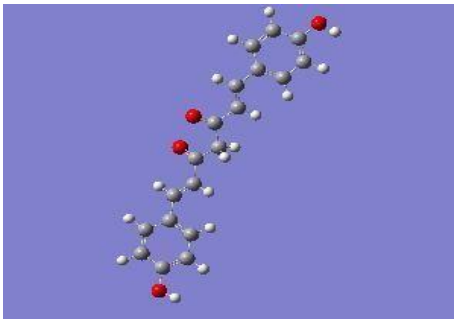
after the bibliographic search of the two plants and their major molecules we took 6 molecules from PubChem form sdf (see figure13). Particles obtained from PubChem are represented in the table (see table1)

(1) (2)

Figure 13: (1) screenshot of PubChem (2) download molecule format SDF 3D

Table 1 : molecules obtained from PubChem

Plants	Ligand	Sdf form
Salvia officinalis	Camphor	
	1,8-cineole	
	Alpha,beta-thujone	
Curcuma longa	Ar-turmerone	
	Alpha-turmerone	

	Bisdemethoxycurcumin	

b) Preparation of ligand

In this step, we will enter the molecules into the Gaussian for optimization. Follow the steps in the pictures (see figure 14 and 15)

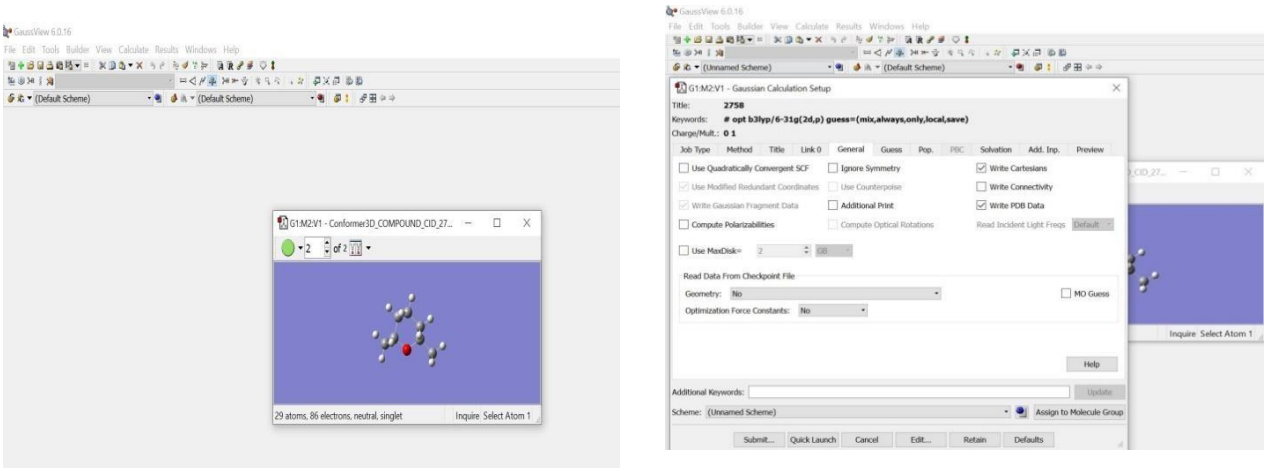
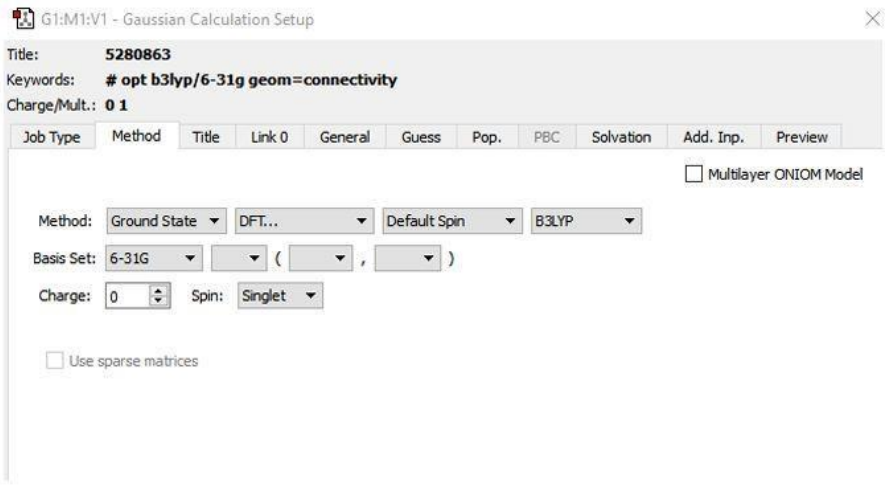
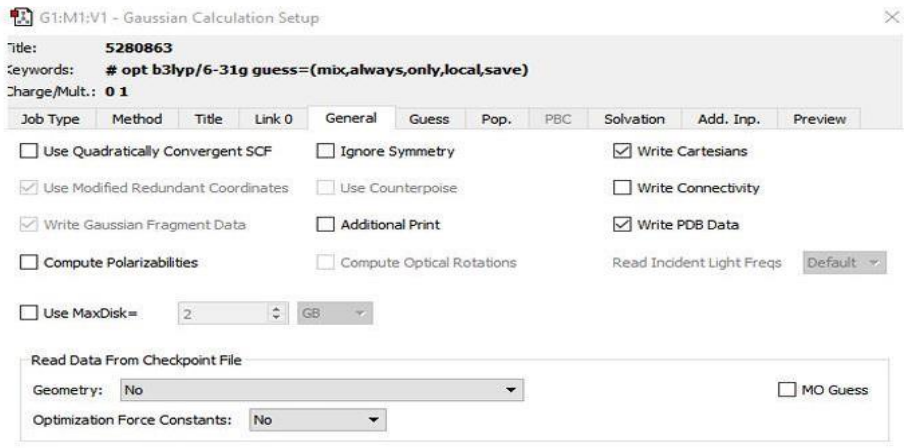


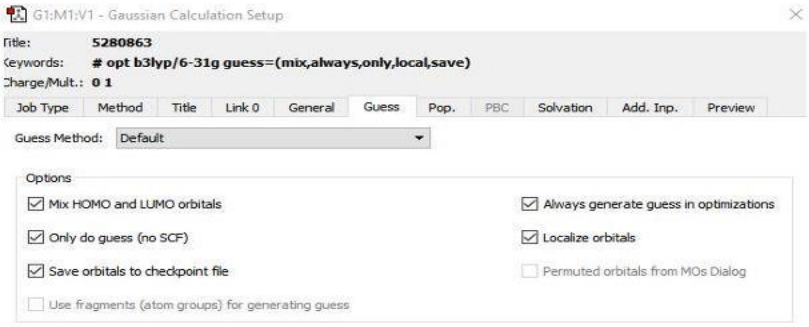
Figure 14: optimization step with gaussian view



(a)



(b)



(c)

Figure 15 : (a) DFT method ,(b) unhook write connectivity, (c) click the five lockers

- After starting the calculation process for Optimization, we wait for the calculation period (see figure 16), after which the account comes out in the form of GJF, after which we convert it to PDB (see figure 17)

```
         --  PICKER X-RAY CORP.    DIGITAL PRINTER MANUAL CA. 1964
Job cpu time:       0 days  0 hours  5 minutes 50.0 seconds.
File lengths (MBytes):  RWF=     10 Int=      0 D2E=      0 Chk=      2 Scr=      1
Normal termination of Gaussian 09 at Thu Feb 06 00:21:59 2025.
```

Figure 16 : Calculation period for optimization

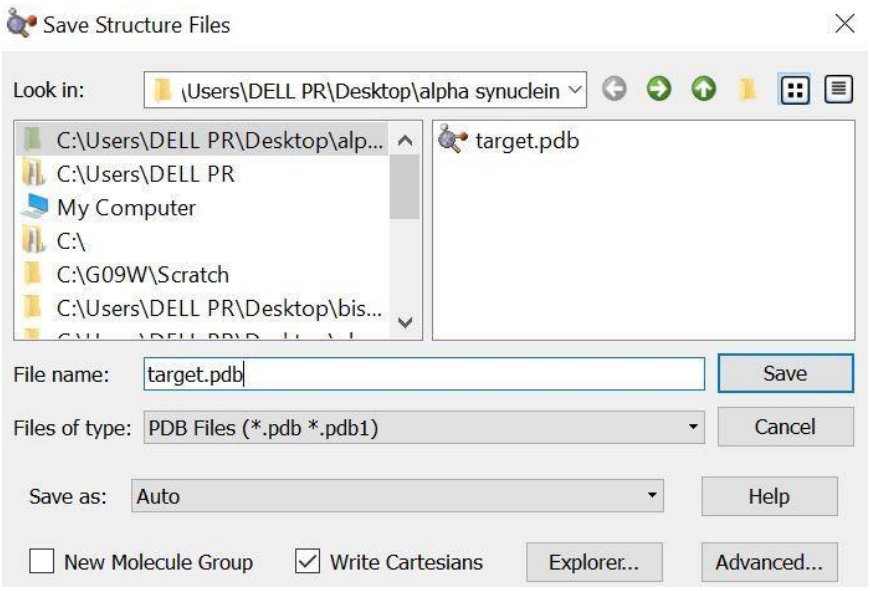


Figure 17 : form GJF converted to PDB

c) Preparation of target

In this step we will download the targets(proteins/receptors) A2A adenosine receptor, alpha-synuclein, GABA-A receptor, NrF2 to the rcsb pdb site form PDB homo sapiens (see figure 18)

- A2A adenosine receptor: 2YDO
- Alpha-synuclein: 1XQ8
- GABA-A receptor: 6DW1
- NrF2: 5WG1

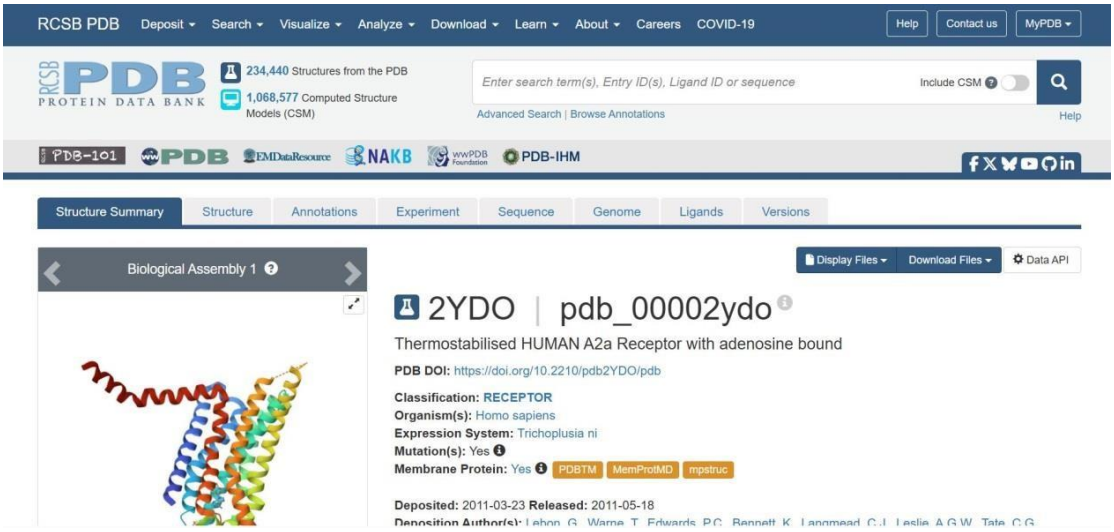


Figure 18 : protein data base

d) AutoDock Vina tools

AutoDock (see figure 19) and its associated tools play a crucial role in structure-based drug design, virtual screening, and understanding molecular interactions, offering researchers powerful computational tools for studying ligand-receptor interactions and predicting potential drug binding modes [22].

- In this step we will convert the ligand and target format PDB to format PDBQT

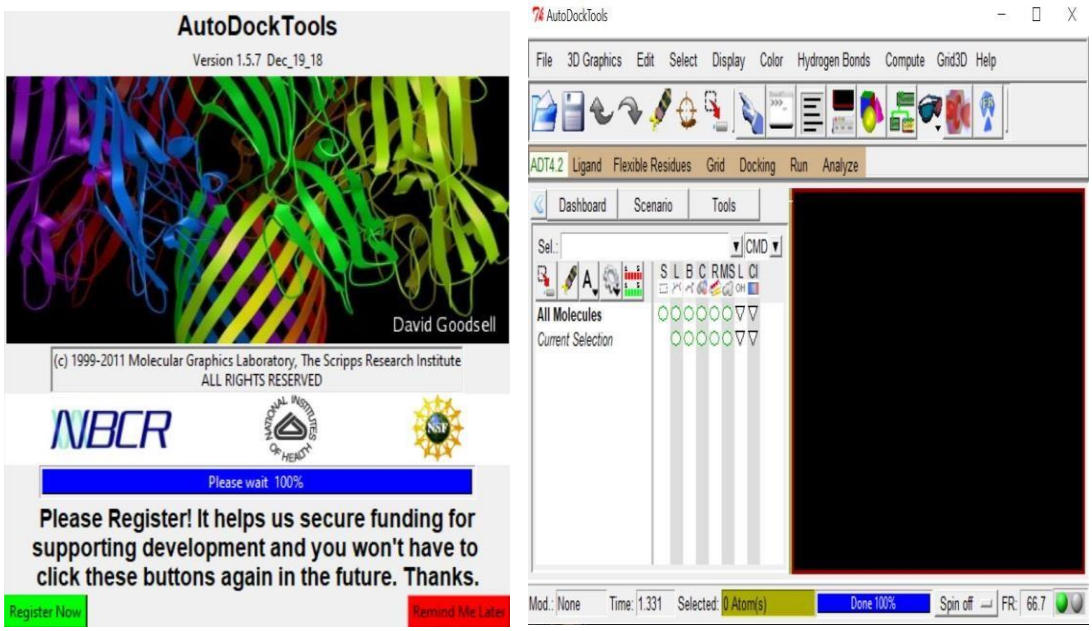


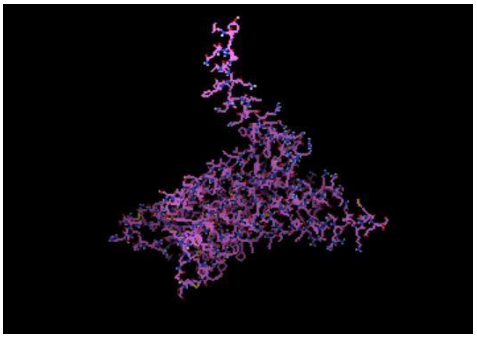
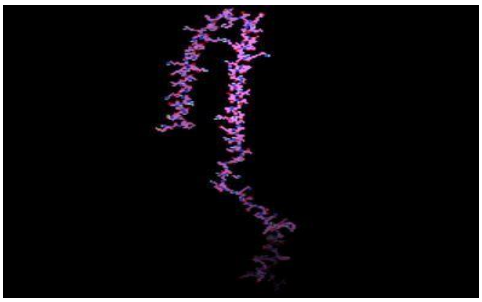
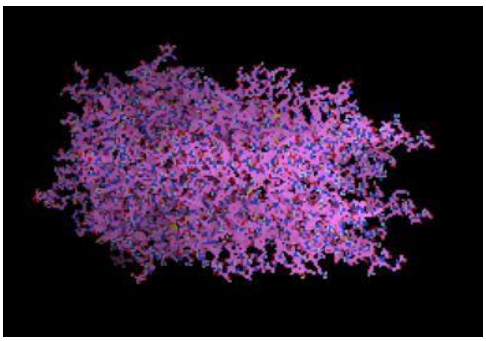
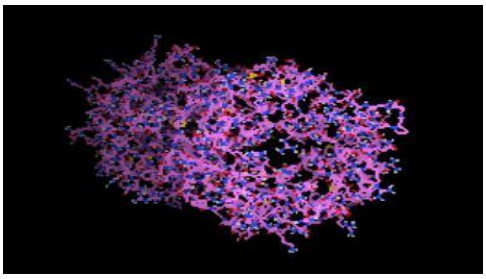
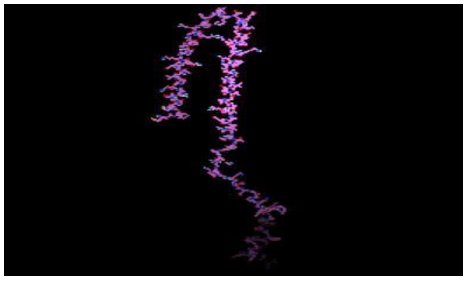
Figure 19 : screenshot of autoDock tools

e) InSilico (molecular docking)

1- presentation of proteins

Receptors (protein\target): the proteins used in this study include A2A adenosine receptor, alpha-synuclein, GABA-A receptor, NrF2, obtained from the protein data bank (PDB) in their three-dimentional (3D) structures. These proteins were prepared according to the previously described procedures and converted into the PDBQT format, prepared using the AutoDock vina tool (see table 2)

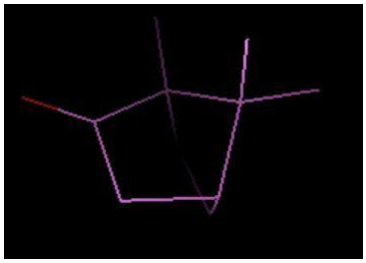
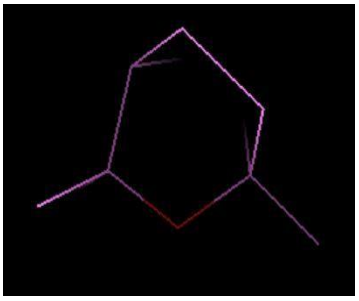
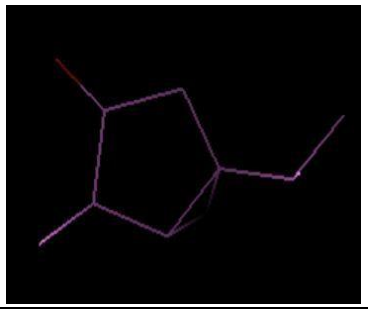
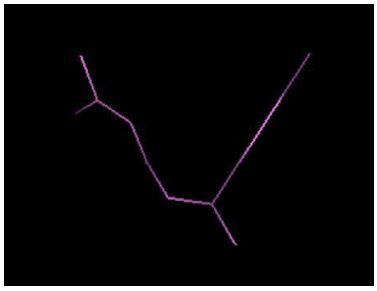
Table 2 : PDBQT form of proteins

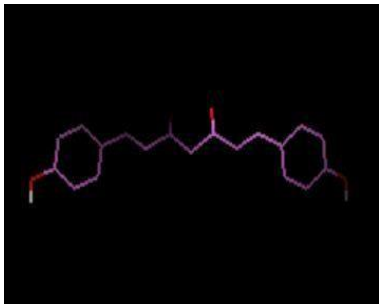
Target	Format PDBQT
A2AAR	
Alpha-synuclein	
GABA-A	
NrF2	
Alpha-synuclein	

2 -Ligand presentation

we prepared the molecules (camphor,1,8-cineole, alpha,beta-thujone, Ar-turmerone, Alpha-turmerone, Bisdemethoxycurcumin) in Gaussian / Gaussview and AutoDock Tools by converting them from PDB to PDBQT as described before, and got the results in (table 3)

Table 3 : presentation of ligands and proteins

Plants	Molecule	PDBQT format
Salvia officinalis	Camphor	
	1,8-cineole	
	Alpha,beta-thujone	
Curcuma longa	Ar-turmerone	

	Alpha-turmerone	
	Bisdemethoxycurcumin	

f) InSilico interaction between ligand pdbqt & target pdbqt

In this step, we will perform molecular docking and extract the grid box (see table 4)

Table 4 : molecular docking (InSilico) and the grid box

Target	Molecule	Grid box
A2AAR	Camphor	X= -23.905 Y= 20.179 Z= -25.400
	Ar-turmerone	X= -31 Y= 7 Z= -21
	Bisdemethoxycurcumin	X= -31 Y= 7 Z= -21
Alpha-synuclein	1,8-cineole	X= 233.803 Y= 31.617 Z= -15.164
GABA-A	Alpha,beta-thujone	X= 136 Y= 143 Z= 157
NrF2	Alpha-turmerone	X= -51 Y= 28 Z= 40

g) Vina autoDock

AutoDock Vina, introduced in 2010 by Dr. Oleg Trott at The Scripps Research Institute's Molecular Graphics Lab, represents a significant advancement in the AutoDock family. Represent a significant advancement in molecular docking software, offering improved efficiency, simplicity, and accuracy in predicting ligand-receptor interactions. Its user-friendly interface and advanced optimization algorithms make it a valuable tool for structure-based drug design and virtual screening studies [23]. In this step we will create a folder which contains the pdbqt ligand, the pdbqt target, conf file and the Vina file (see figure 20)

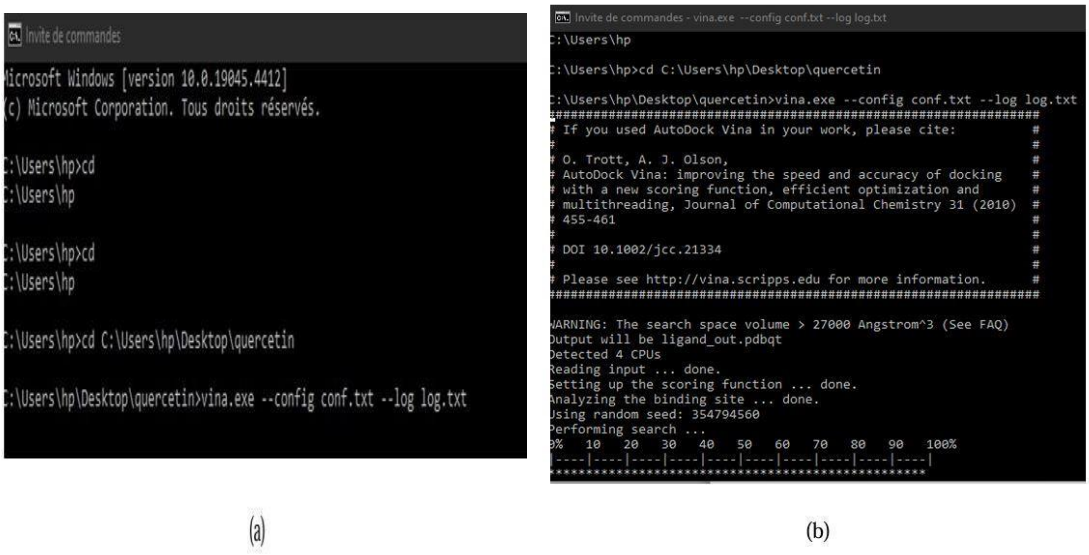


Figure 20: (a) structure of command Vina, (b) affinity calculate by Vina

h) BIOVIA Discovery Studio

It serves as a valuable platform for researchers in the pharmaceutical, biotechnology, and materials science industries, as well as academic institutions, to accelerate the process of drug discovery and development through computational methods. In this step we will enter the target pdbqt & the ligand out pdbqt to show 2D, 3D & bonds of amino acids (see figure21) [24]
.

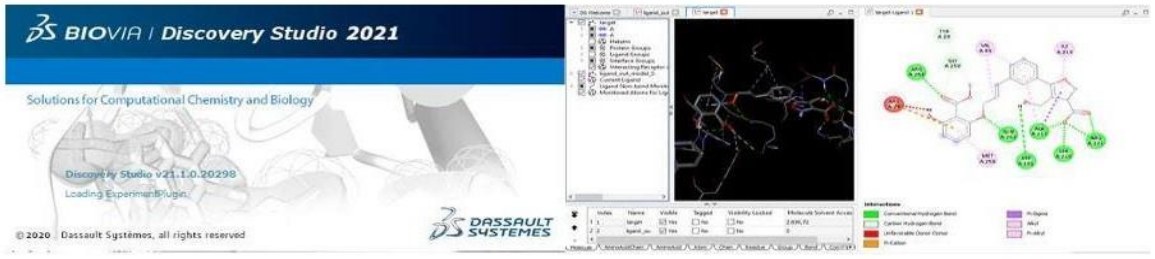


Figure 21: (a) discovery screenshot , (b) show 2D by (a)

i) Databases

Swiss ADME: http://www.swissadme.ch/ The graphical output of Swissadme consists of the BOILED-Egg directly implemented from the original publication to predict passive diffusion through HIA (human intestinal absorption) and BBB (blood-braib barrier) by position in a WLOGP-TPSA physicochemical space. In this step we copy the smile of the molecules on PubChem, and we paste in Swiss adme for the knowledge of this hydrophilic or li&pophilic molecules in boiled egg (see figure22) [25].

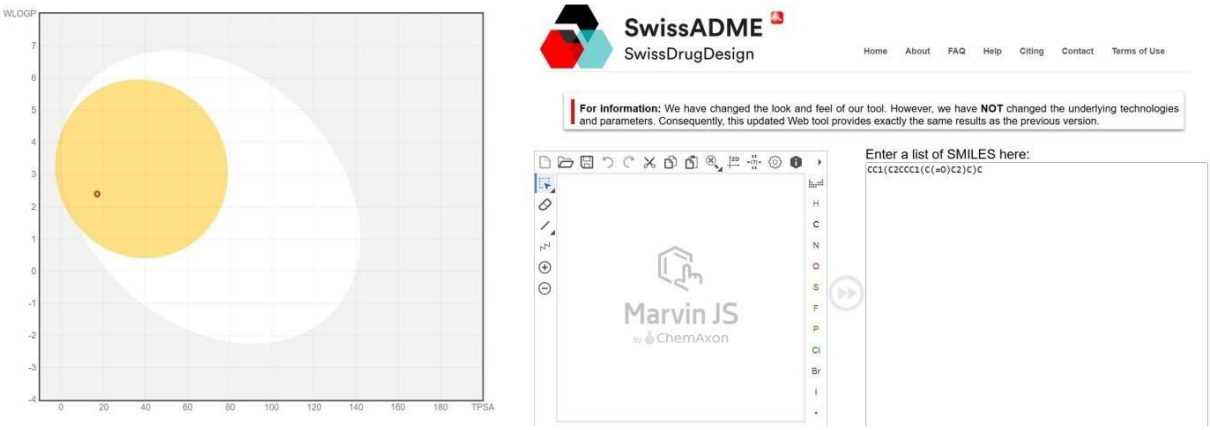


Figure 22: Swiss adme & presenting boiled egg

j) Swiss target prediction

http://www.swisstargetprediction.ch/ Swiss Target Prediction is an online tool to predict the targets of bioactive small molecules in human and other vertebrates. This is useful to understand the molecular mechanisms underlying a given phenotype or bioactivity, to rationalize possible side-effects or to predict off-targets of known molecules [26]. In this step we will copy the smile of the molecules on PubChem and paste in Swiss target prediction to see the probabilities of our molecules with our target (see figure 23)

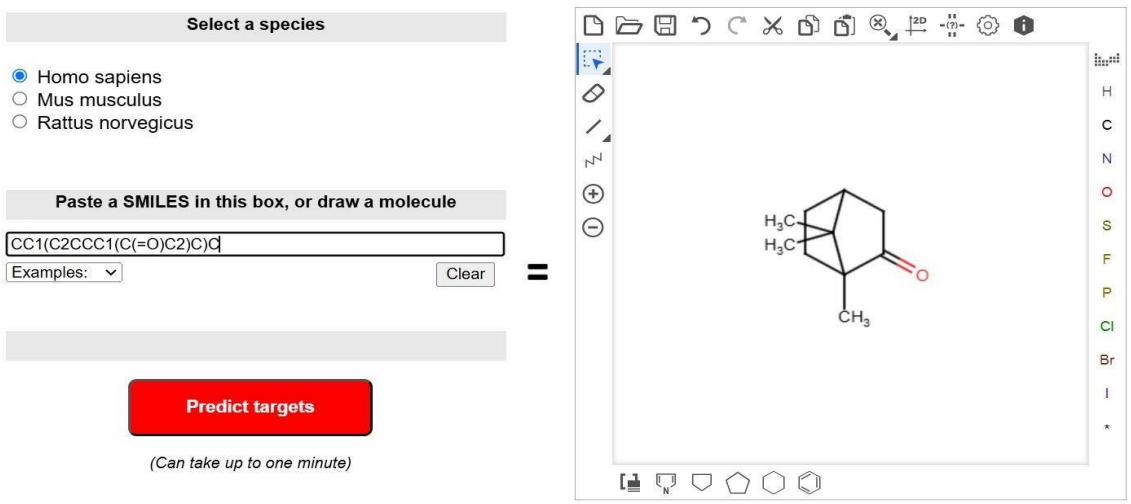


Figure 23 : probabilities of molecules target

k) ProTox 3.0

https://tox.charite.de/protox3/ is a virtual toxicity lab enabled to academic and non-commercial users via a web server, for the prediction of multiple toxicological endpoints related with a chemical structure. ProTox 3.0 contains computer-based models trained on real data (in vitro or in vivo) to predict the toxic potential of the existing and virtual compounds. The acute toxicity class as well as different endpoints is calculated for an input compound based on chemical similarities to toxic compounds and trained machine learning models. ProTox 3.0 envisages itself as a freely available complete computational platform for InSilico toxicity prediction for toxicologist, regulatory agencies, computational chemist and medicinal chemist. In this step we will copy the smile of the molecules on PubChem and paste in ProTox to have the toxicity of the molecules and these classes (see figure 24) [27].

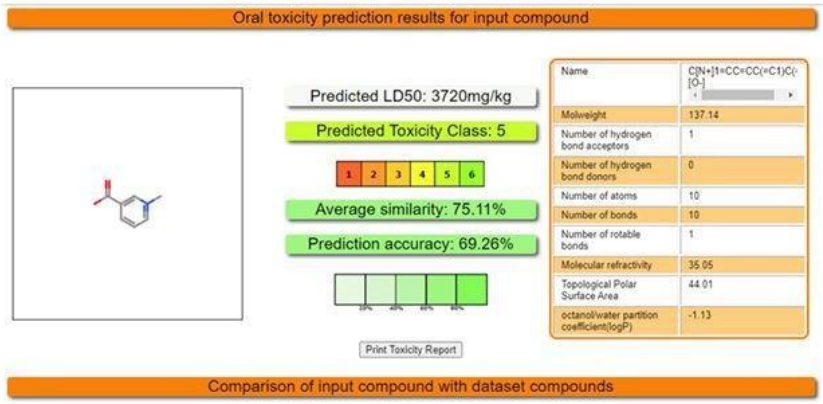


Figure 24: ProTox databases

l) Plip

https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index The Protein-Ligand Interaction Profiler (PLIP) is a web-based tool designed to automatically detect and visualize non-covalent interactions between proteins and ligands in 3D structures. is an automated computational tool used to analyze and visualize non-covalent interactions between proteins and small-molecule ligands in 3D structures [28] .(e.g., from PDB files or docking results). (see figure 25)

Figure 25: plip databases

m) CB dock

https://clab.labshare.cn/cb-dock/ For protein-ligand blind docking, is a computational method used to predict and analyze the binding interactions between small molecules (ligands) and target proteins or biomolecular structures. It combines molecular docking, scoring functions, and sometimes machine learning to estimate binding affinities, identify potential drug candidates, and optimize molecular interactions [29] [30]. (See figure 26)

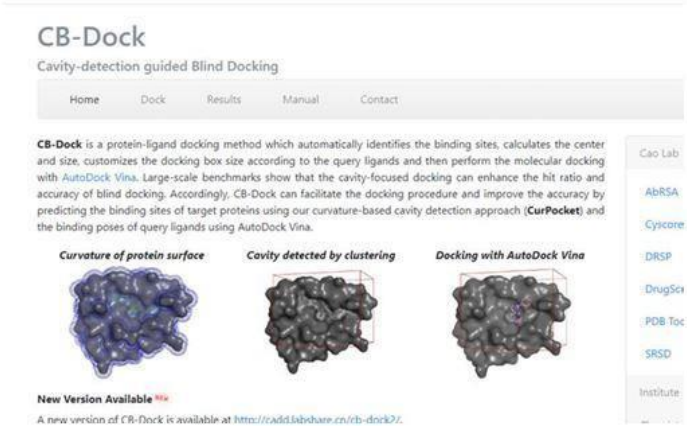


Figure 26: CB dock databases

n) String

https://string-db.org/ STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a biological database and web resource that provides comprehensive information on protein-protein interactions (PPIs), including both physical and functional associations. It integrates data from multiple sources to build a global network of protein interactions across thousands of organisms [31]. (see figure 27)

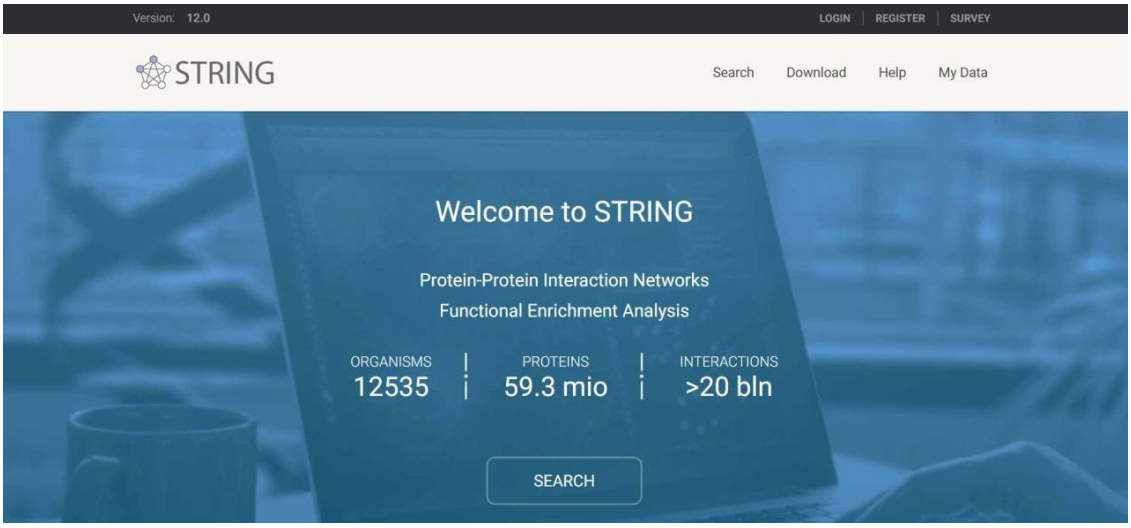


Figure 27: string protein-protein interaction networks

Conclusion

This study focuses on the roles of nuclear factor erythroid 2-related factor 2 (Nrf2), gamma-aminobutyric acid type A receptors (GABA-A), adenosine A2A receptors (A2AR), and alpha-synuclein in disease pathogenesis, with particular emphasis on their implications in Parkinson's disease (PD). Dysregulation of these proteins has been linked to multiple neurodegenerative and neuropsychiatric disorders, making them critical therapeutic targets.

Additionally, we investigated natural plant-derived compounds capable of modulating these molecular pathways. Our research highlights their potential to inhibit pathogenic protein aggregation, restore redox balance, and regulate neurotransmitter signaling, thereby offering promising avenues for symptom alleviation and disease modification in PD and related conditions.

CHAPTER III

Quantitative Structure-Activity Relationships

Introduction

Structure-Activity Relationship (SAR) modeling, first introduced by Hansch and Fujita in the 1960s, is a computational method used to establish mathematical relationships between a compound's structural characteristics and its biological activity. In a typical QSAR-based drug design approach, the process begins with the selection of a series of both active and inactive molecules, from which relevant chemical descriptors are derived. These descriptors then serve as the foundation for constructing a predictive model that links molecular features to biological activity. Once developed, the model undergoes thorough validation to ensure its accuracy and reliability. Finally, the validated model is applied to predict the activity of new compounds, providing valuable insights for drug discovery and optimization. This systematic framework bridges the gap between molecular structure and biological function, enhancing the efficiency of pharmaceutical research [32].

III- QSAR

III-1- principle of QSAR

Quantitative Structure-Activity/Property Relationships (QSARs/QSPRs) are mathematical models that correlate a drug's structural properties with its biological activity. Due to the complexity of these relationships, artificial neural networks (ANNs) are well-suited for QSAR development. While comparative molecular field analysis (CoMFA) was previously widely used, ANNs have proven more effective due to their superior nonlinear pattern recognition capabilities. For instance, in predicting pesticide retention times, an ANN-based QSPR model outperformed multiple linear regression (MLR), yielding a mean relative error of 9.04% compared to MLR's 13.8% [33].

ANN predictive performance has also been benchmarked against partial least squares (PLS) analysis. In a study by Vucicevic et al., QSPR models for α2-adrenoreceptor and imidazoline receptor activity in blood-brain barrier (BBB) permeability were developed using PLS, MLR, and ANN approaches. The ANN model demonstrated the highest statistical relevance among the three [34].

This approach establishes mathematical models that correlate molecular descriptors (quantitative representations of chemical structures) with biological activities or physicochemical properties. In pharmaceutical applications, QSAR enables researchers to : [35]

- **Predict Bioactivity**: Forecast biological effects (e.g., receptor binding affinity, enzyme inhibition) for novel compounds before synthesis
- **Optimize Lead Compounds**: Identify which structural modifications will enhance desired pharmacological properties
- **Reduce Experimental Costs**: Minimize laboratory testing by prioritizing the most promising molecular candidates

III-2- The QSAR workflow typically involves

- Calculating molecular descriptors (topological, electronic, steric, or hydrophobic parameters)
- Selecting statistically significant variables using feature reduction techniques
- Building predictive models using machine learning (e.g., random forests, neural networks) or traditional regression methods
- Validating models through rigorous statistical testing (e.g., cross-validation, external test sets)

III-3- Selection of Active Compounds and Molecular Descriptors in QSAR Modeling

The success of QSAR-based drug discovery depends on several critical factors, beginning with the identification of active compounds and the selection of appropriate molecular descriptors. First, the chemical diversity of compounds used in training and test sets significantly impacts model performance. Public databases such as ChEMBL, PubChem, ZINC, and the Accelrys Available Chemicals Directory provide extensive libraries of biologically tested compounds, enabling researchers to assemble diverse datasets. Ensuring structural variety in the training set is essential insufficient diversity may lead to biased models that fail to identify potential hits [36].

Molecular descriptors, which encode chemical information into numerical form, play a pivotal role in QSAR model development. These descriptors can be broadly categorized into experimental (e.g., logP, polarizability) and theoretical (e.g., hydrogen bond donors/acceptors, electronic properties). While experimental descriptors may carry measurement errors, theoretical descriptors are derived from computational methods and are free from experimental noise. Given the vast number of available descriptors, careful selection is crucial, as it directly influences model accuracy and predictive power. A well-optimized

descriptor set enhances the model's ability to establish meaningful structure-activity relationships, ultimately improving drug discovery outcomes [36].

III-4- Essential steps in QSAR

- Selection of data set and PIC50
- Calculation of descriptors
- Model construction
- Validation of model

III-5- Descriptors

Descriptors are mathematically derived or experimentally measured quantities that encode the structural, physicochemical, or electronic properties of chemical compounds. These descriptors serve as the *input variables (features)* in predictive modeling techniques such as *Multiple Linear Regression (MLR)* and *Support Vector Regression (SVR)*, where they correlate molecular characteristics with a biological or chemical endpoint (e.g., toxicity, binding affinity, solubility) [37].

The number of descriptors in a QSAR model depends significantly on the quality of the dataset, including the quantity and diversity of compounds as well as the nature of the modeled endpoint. Generally, biological endpoints are more complex than physicochemical properties and may require a larger number of descriptors for accurate modeling. However, it is essential to keep the number of descriptors as low as possible to ensure model robustness. Additionally, strict limitations must be imposed on the number of variables relative to the number of observations to maintain statistical reliability and avoid overfitting [38].

III-6- Classification of descriptors

Molecular descriptors can be classified in several ways. The majority of descriptors are atom-based rather than field-based. Atom-based descriptors are typically generated by analyzing 2D or 3D connectivity tables and may include 1D, 2D, or 3D information about the molecule. These descriptors encompass individual atoms, feature counts, substructural fragments, topological indices, atomic properties, pharmacophores, and computed physicochemical properties [39].

- **Constitutional descriptors (1D):** are derived directly from a molecule's molecular formula (e.g., C_6H_6O for phenol) and represent fundamental molecular properties. These include atom counts (C, H, N, etc.), functional group counts (NO_2, COOH, OH), bond type counts (single, double, aromatic), ring counts (both aromatic and aliphatic), and molecular weight. As the simplest class of molecular descriptors, they provide a basic characterization of a compound's composition without considering structural connectivity or three-dimensional arrangement.

- **2D Descriptors:** These two-dimensional descriptors effectively capture physical properties in most cases, but prove insufficient for adequately explaining certain characteristics, particularly biological activities. Derived from molecular structural formulas, this class includes topological indices that can be obtained from a molecule's 2D structure. The computation of hundreds of topological descriptors remains efficient even for large datasets.

- **Three-dimensional (3D) descriptors:** account for the spatial arrangement of atoms within a molecule. Their calculation requires **molecular modeling** using computational methods such as **semi-empirical** and **ab initio** approaches. While these descriptors are computationally intensive, they provide more detailed information and are essential for modeling **3D-dependent properties** (e.g., biological activity, binding affinity).

III-7- Multiple linear regression MLR

Multiple Linear Regression (MLR) is an extension of simple linear regression that models the relationship between a dependent variable and multiple independent variables. Unlike simple linear regression, which uses only one predictor, MLR incorporates several explanatory variables to improve prediction accuracy [40].

III-8- Applications Across Industries

MLR has broad applications in finance, marketing, and economics. Financial analysts use it to forecast stock prices and interest rates, while marketers apply it to predict

consumer behavior and sales performance. Economists leverage MLR to analyze economic indicators such as GDP growth and inflation [40].

8-1- Mathematical Foundation

The model assumes a linear relationship between the dependent variable (y) and independent variables (x_1, x_2, ..., x_n). The general MLR equation is: [40]

$$y=b_0+b_1x_1$$

where:

- y = Dependent variable
- b_0 = Intercept (baseline value)
- b_1 = Regression coefficients (slopes)
- x_1 = Independent variables

In a multiple linear regression model, there are multiple independent variables. The equation for a multiple linear regression model is:

$$y = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n$$

where:

* y is the dependent variable
* b_0 is the intercept
* b_1, b_2, ..., b_n are the slopes
* x_1, x_2, ..., x_n are the independent variables

8-2- Practical Use Cases

MLR can predict:

- Real estate prices using features like square footage, bedrooms, and location.
- Product sales based on pricing, advertising spend, and distribution channels.
- Business growth by analyzing revenue, expenses, and asset trends.

While MLR is a powerful predictive tool, its effectiveness depends on linearity assumptions. In cases of complex, nonlinear relationships, alternative modeling techniques (e.g., machine learning algorithms) may be more suitable. Proper validation is essential to ensure model reliability [40].

Data base:

To ensure the reliability of our QSAR model we adopted a data base contained of 22 molecules, the IC50 values were normalized to PIC50 using scale of logarithm

$$\mathbf{pIC50 = - \log 10(IC50*10\text{-}6) \ldots\ldots. (1)}$$

III-9- Validation of QSAR model

validation is a crucial and essential aspect for determining the reliability of models. There are several validation approaches, including internal and external validation. Recent studies have indicated that internal validation is considered necessary for model validation. Furthermore, external validation is an important and required validation method used to assess both the generalizability and true predictive ability of QSAR models. Recently, Roy et al. proposed additional validation parameters as supplementary tools [41],[42],[43].

2- Internal validation

In addition to the root mean square error (RMSE) and the coefficient of determination (R^2), the key statistical parameters used in this work to evaluate model performance include the following:

- **the cross-validation coefficient:** Q^2 appr that must be > 0,5
- $\overline{\mathbf{rm^2}}$ and $\boldsymbol{\Delta rm^2}$**:** should preferably be less than 0.2

3- External validation

The most commonly used external validation parameters and criteria are given below:

- RMSE and R^2
- Q^2 test coefficient or (R^2pred) which reflects the degree of correlation between the observed and predicted activity data from the test set. Models with R^2pred values greater than the stipulated value of 0.5 are considered well predictive [44].

- Q^2 train> 0.5, Q^2 test> 0.6
- $\frac{r^2-r_0^2}{r^2}$ < 0.1 and 0.85≤ k ≤ 1.15
- The CCC parameter can be calculated to check the reliability of the model, by detecting the distance of the observations from the fitting line and the degree of deviation of the regression line from the one passing through the origin, respectively. Any deviation of the regression line from the line of agreement (line passing through the origin) gives a CCC value less than 1 [45],[46].

The equations for these different internal and external validation parameters and the meaning of the different terms are given below.

$$RMSE = \sqrt{\frac{\sum\left(Y_{exp} - Y_{pred}\right)^2}{n}} \quad \text{..... (2)}$$

$$R^2 = 1 - \frac{\sum\left(Y_{\exp} - Y_{pred}\right)^2}{\sum\left(Y_{\exp} - \bar{Y}_{\exp}\right)^2} \quad \text{..... (3)}$$

$$Q^2_{app} = 1 - \frac{\sum\left(Y_{\exp(appl)} - Y_{pred\,(appl)}\right)^2}{\sum\left(Y_{\exp(appl)} - \bar{Y}_{\exp(appr)}\right)^2} \quad \text{.....(4)}$$

$$Q^2_{test} = 1 - \frac{\sum\left(Y_{\exp\,(test)} - Y_{pred\,(test)}\right)^2}{\sum\left(Y_{\exp\,(test)} - \bar{Y}_{\exp(appl)}\right)^2} \quad \text{..... (5)}$$

$$r_m^2 = r^2\left(1 - \sqrt{r^2 - r_0^2}\right) \quad \text{..... (6)}$$

$$r_m'^2 = r^2\left(1 - \sqrt{r^2 - r_0'^2}\right) \quad \text{.....(7)}$$

$$\bar{r}^2_m = \frac{\left(r_m^2 + r_m'^2\right)}{2} \quad \text{.....(8)}$$

$$\Delta r_m^2 = \left| r_m^2 - r_m'^2 \right| \quad \text{..... (9)}$$

$$CCC = \frac{2\sum\left(Y_{exp(test)} - \bar{Y}_{exp(test)}\right)\left(Y_{pred(test)} - \bar{Y}_{pred(test)}\right)}{\sum\left(Y_{exp(test)} - \bar{Y}_{exp(test)}\right)^2 + \sum\left(Y_{pred(test)} - \bar{Y}_{pred(test)}\right)^2 + n\sum\left(\left(\bar{Y}_{exp(test)} - \bar{Y}_{pred(test)}\right)\right)} \quad \text{..... (10)}$$

Where Yappr and Y test are the experimental values of the studied property for the training and test sets and $\bar{y}$appr and $\bar{Y}$test are the predicted values of the studied property for the training and test sets, n is the number of compounds in the database.

Conclusion

The Quantitative Structure-Activity Relationship (QSAR) approach is a powerful computational methodology in drug discovery and chemical research that correlates molecular structures with biological activity using quantitative molecular descriptors. This process involves several critical steps like the selection of compounds with experimentally determined activity data, calculation of molecular descriptors, feature selection to identify relevant chemical properties, model construction, and rigorous validation.

QSAR models span various dimensional approaches from simple 1D (constitutional descriptors) to higher-dimensional (3D, 4D) methods that incorporate spatial, electronic, and dynamic molecular properties to improve predictive accuracy. Common modeling techniques include multiple linear regression (MLR), partial least squares (PLS) regression, and support **vector** machines (SVM), with model reliability assessed through both internal (e.g., cross-validation) and external validation metrics.

While structure-based drug design has advanced significantly, QSAR remains indispensable in pharmaceutical research due to its computational efficiency, interpretability, and robust predictive power, particularly in early-stage drug discovery and chemical property optimization [47],[48],[49].

III-10- data set and methods

10-1- data set

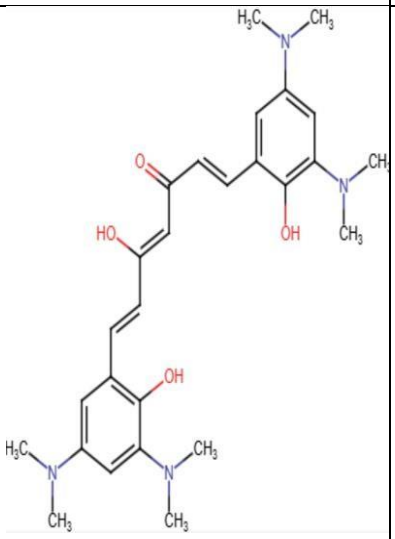
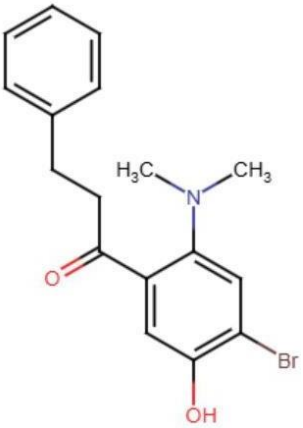
In this study, a data set of 55 derivatives of 6-Hydroxybenzothiazole-2-carboxamide compounds. The chemical structures and anti-parkinson activity (IC50) of these 55 molecules obtained from an article [50] are presented Table 5. the data was selected based on their experimentally determined anti-Parkinson activity (IC_{50} values). The compounds were prioritized according to their inhibitory potency against Parkinson's disease (PD)-relevant targets, ensuring structure-activity relevance for computational modeling.

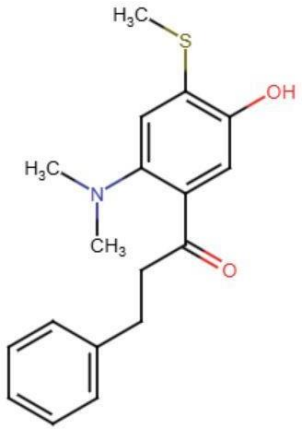
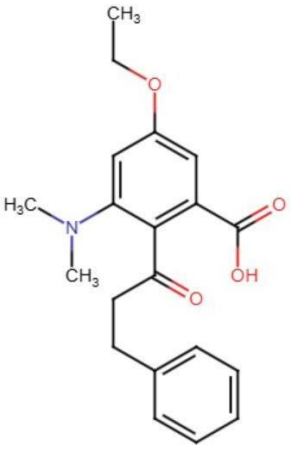
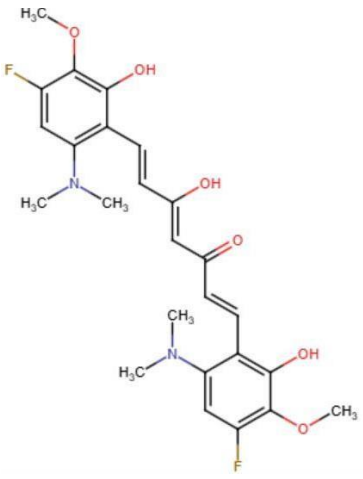
In our research, we relied specifically on QSAR with reference to some chemical characteristics that fall in particular within the QSAR research, because the QSAR database depends on biological activity, unlike QSPR which depends on the physiochemical properties of the molecule

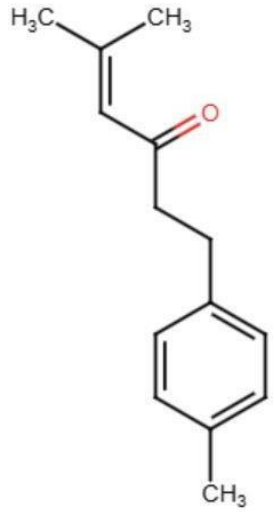
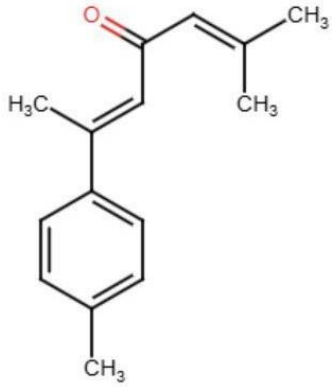
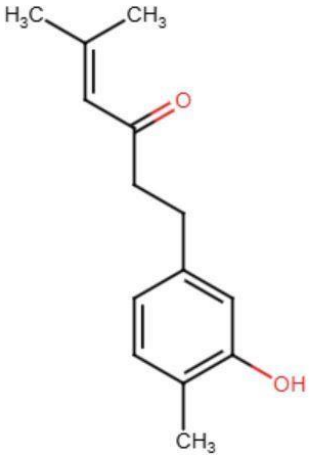
We tried to start from molecular docking to inhibit the proteins responsible for the disease, but we encountered the obstacle of the difference in the receptor (target protein). Although we succeeded in inhibiting the protein, we could not create a database that shares the same target protein. Therefore, we set up committees for QSAR modeling, as a huge database can be obtained by targeting the biological activity.

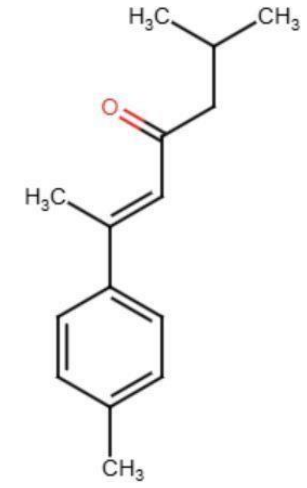
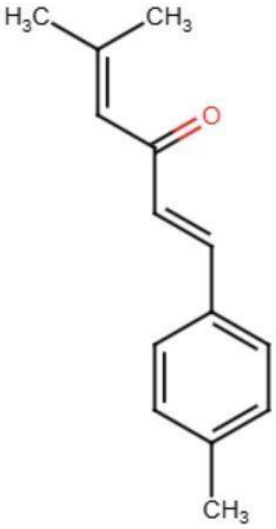
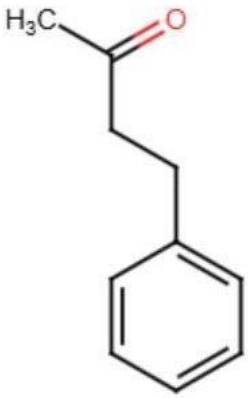
The IC50 values were converted into its logarithmic scale pIC50 = −log (IC50), to reduce the skewness of the data set, which was then used for subsequent QSAR analysis as the response variable.

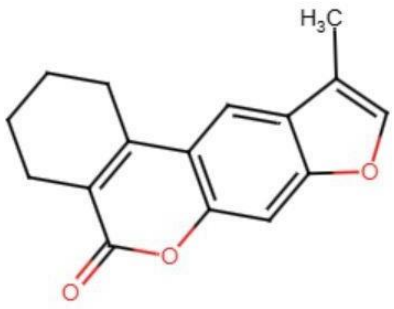
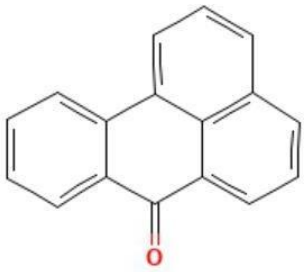
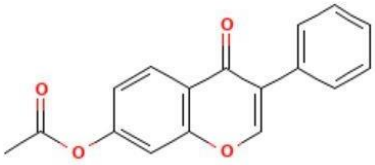
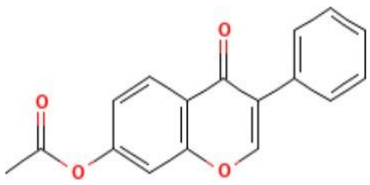
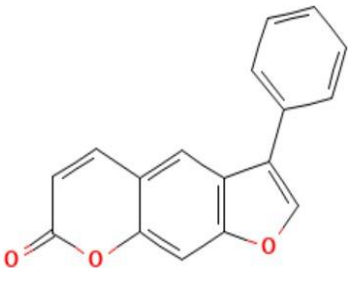
Table 5: Data base of 55 derivatives from QSAR

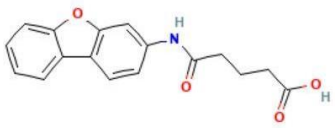
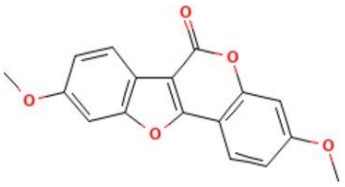
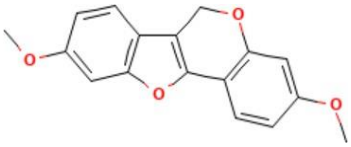
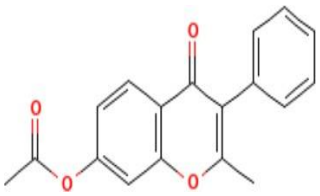
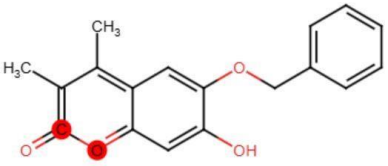
N	Smile	Chemical structure	PIC50
1	CN(C)C1=CC(N(C)C)=C(O)C(\C=C\C(=O)\C=C(/O)\C=C\C2=C(O)C(=CC(=C2)N(C)C)N(C)C)=C1		4.37
2	CN(C)C1=CC(Br)=C(O)C=C1C(=O)CCC1=CC=CC=C1		4.72
3	CCCC(=O)C1=CC(O)=CC=C1N(C)C		19.40

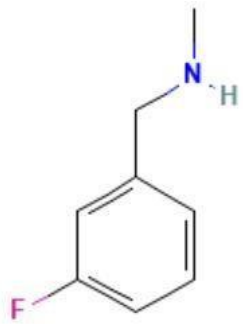
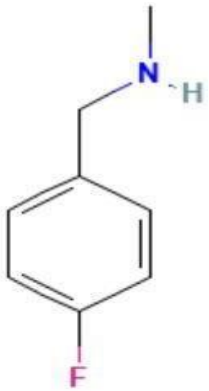
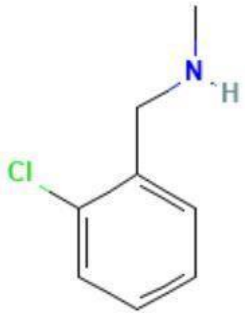
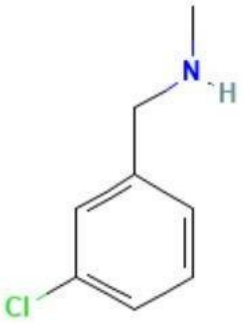
4	CSC1=C(O)C=C(C(=O)CCC2=CC=CC=C2)C(=C1)N(C)C		4.87
5	CCOC1=CC(C(O)=O)=C(C(=O)CCC2=CC=CC=C2)C(=C1)N(C)C		4.24
6	COC1=C(O)C(\C=C\C(\O)=C\C(=O)\C=C\C2=C(O)C(OC)=C(F)C=C2N(C)C)=C(C=C1F)N(C)C		4.84

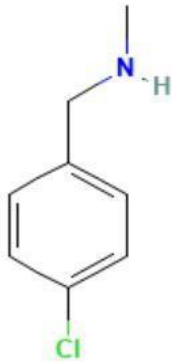
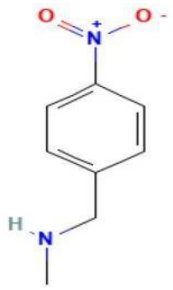
7	CC(C)=CC(=O)CCC1=CC=C(C)C=C1		11.1
8	CC(C)=CC(=O)\C=C(/C)C1=CC=C(C)C=C1		16.3
9	CC(C)=CC(=O)CCC1=CC(O)=C(C)C=C1		4.72

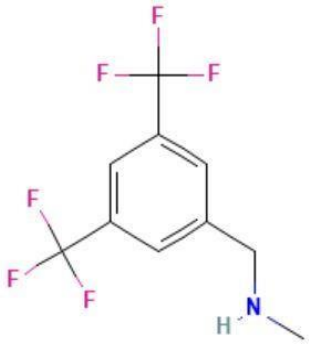
10	CC(C)CC(=O)\C=C(/C)C1=CC=C(C)C=C1		20.4
11	CC(C)=CC(=O)\C=C\C1=CC=C(C)C=C1		7.26
12	CC(=O)CCC1=CC=CC=C1		5.17

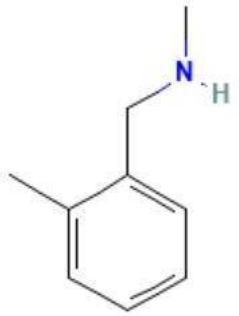
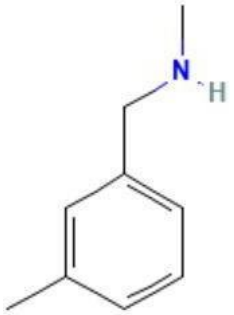
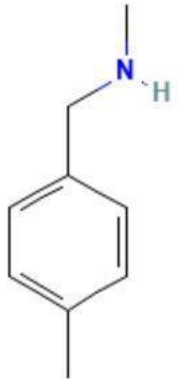
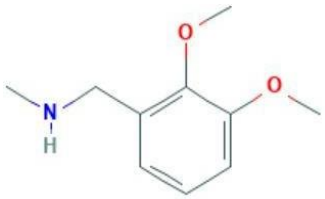
13	CC1=COC2=CC3=C(C=C12)C4=C(CCCC4)C(=O)O3		5.86
14	C1=CC=C2C(=C1)C3=CC=CC4=C3C(=CC=C4)C2=O		8.36
15	CC(=O)OC1=CC2=C(C=C1)C(=O)C(=CO2)C3=CC=CC=C3		8.25
16	C1=CC=C(C=C1)C2=COC3=C2C=C4C=CC(=O)OC4=C3		8.46
17	CC1=CC(=O)OC2=CC3=C(C=C12)C4=CC=CC=C4O3		7.3

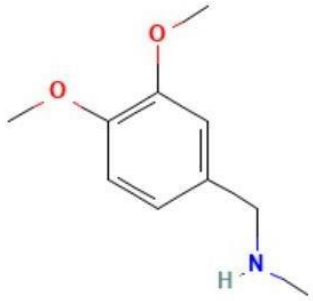
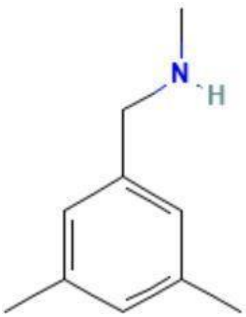
18	C1=CC=C2C(=C1)C3=C(O2)C=C(C=C3)NC(=O)CCCC(=O)O		6.72
19	COC1=CC2=C(C=C1)C3=C(O2)C4=C(C=C(C=C4)OC)OC3=O		5.28
20	COC1=CC2=C(C=C1)C3=C(O2)C4=C(C=C(C=C4)OC)OC3		4.77
21	CC1=C(C(=O)C2=C(O1)C=C(C=C2)OC(=O)C)C3=CC=CC=C3		7.55
22	CC1=C2C=C(OCC3=CC=CC=C3)C(O)=CC2=[O][C](=O)=C1C		5.51
23	CNCC1=CC=CC=C1F		0.279

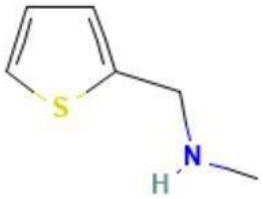
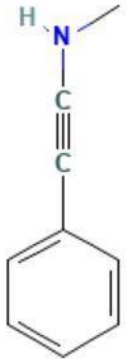
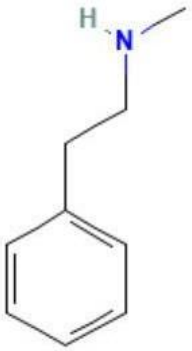
24	CNCC1=CC(=CC=C1)F		0.622
25	NCC1=CC=C(C=C1)F		0.397
26	CNCC1=CC=CC=C1Cl		0.146
27	CNCC1=CC(=CC=C1)Cl		0.437

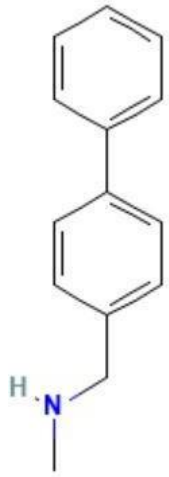
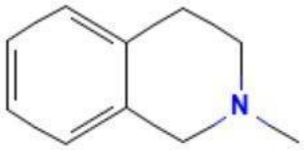
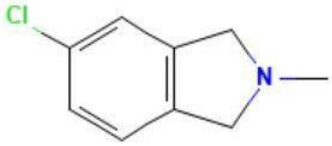
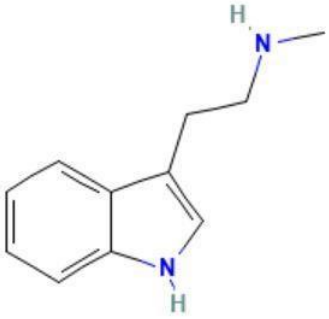
28	CNCC1=CC=C(C=C1)Cl		2.97
29	CNCC1=CC=C(C=C1)[N+](=O)[O-]		1.04
30	CNCC1=C(C(=CC=C1)F)F		0.025
31	CNCC1=CC(=C(C=C1)F)F		0.033

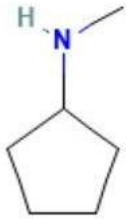
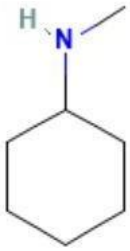
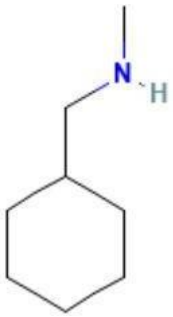
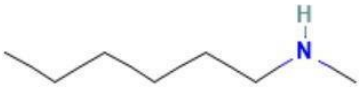
32	CNCC1=C(C=C(C=C1)F)F		0.100
33	CNCC1=CC(=CC(=C1)C(F)(F)F)C(F)(F)F		0.129
34	CNCC1=CC=CC=C1OC		0.258
35	CNCC1=CC(=CC=C1)OC		1.94

36	CC1=CC=CC=C1CNC		0.147
37	CC1=CC(=CC=C1)CNC		0.282
38	CC1=CC=C(C=C1)CNC		2.16
39	CNCC1=C(C(=CC=C1)OC)OC		1.22

40	CNCC1=CC(=C(C=C1)OC)OC		3.21
41	CNCC1=C(C=C(C=C1)OC)OC		0.971
42	CNCC1=CC(=CC(=C1)OC)OC		1.80
43	CC1=CC(=CC(=C1)CNC)C		0.230

44	CNCC1=CC(=C(C=C1)OC)F		0.851
45	CNCC1=CC=CS1		0.250
46	CNC#CC1=CC=CC=C1		0.944
47	CNCCC1=CC=CC=C1		0.041

48	CNCC1=CC=C(C=C1)C2=CC=CC=C2		2.09
49	CN1CCC2=CC=CC=C2C1		1.18
50	CN1CC2=C(C1)C=C(C=C2) Cl		0.140
51	CNCCC1=CNC2=CC=CC=C21		0.140

52	CNC1CCCC1		0.842
53	CNC1CCCCC1		0.011
54	CNCC1CCCCC1		0.118
55	CCCCCCNC		0.072

In this study, we use chemometric methods to analyze the data. Quantitative Structure-Activity Relationship (QSAR) approaches are generally classified into two main types:

1. **Linear methods**: These are based on the assumption of a linear relationship between the independent variables (molecular descriptors) and the dependent variable (biological or chemical activity). One of the most common linear techniques is Multiple Linear Regression (MLR) [51].

2. **Non-linear methods**: These are applied when the relationship between variables is more complex and cannot be accurately described using linear models. Examples of non-linear methods include, Support Vector Machines (SVM) [52].

11- MLR

We start by opening the Spyder environment, then we enter and run the code (see figure 28). After execution, we obtain the descriptor, which is a numerical representation of the characteristics of the analyzed data or signals. This descriptor is usually displayed along with a curve that visually illustrates the behavior of these features (see figure 28)

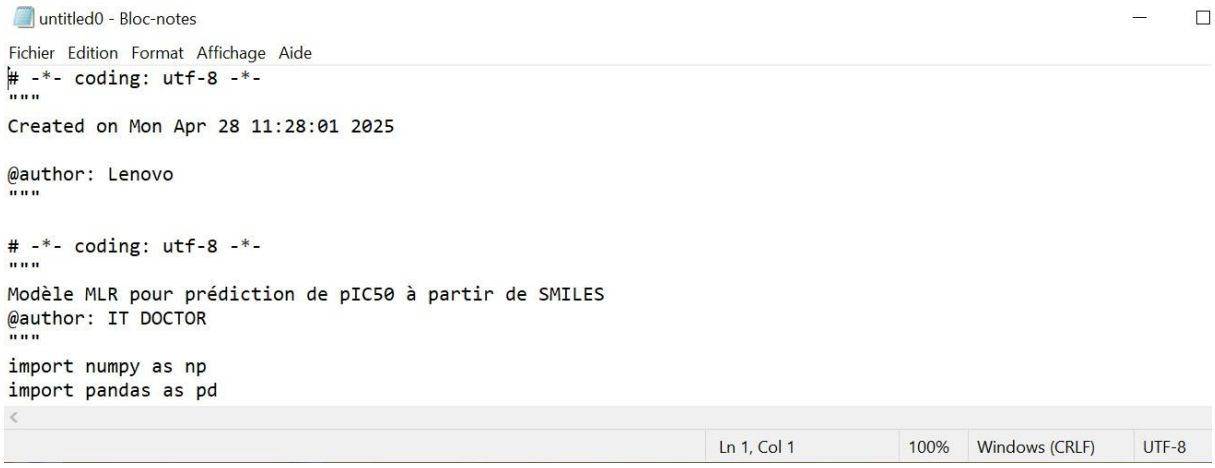


Figure 28: Python Script for PIC50 Prediction using MLR model

12- SVR

Implementing Support Vector Regression (SVR) in Spyder

SVR is implemented in a similar way in terms of steps opening the environment, running the code (see figure 29), and obtaining the results (see figure 29). However, the actual code differs because it is based on the Support Vector Regression algorithm.

Figure 29*:* Python Script for PIC50 Prediction using SVR model

Conclusion

The 9-feature model strikes an optimal balance between interpretability, predictive accuracy, and generalizability. It captures sufficient complexity to explain the variance in PIC50 without overfitting, as evidenced by the statistical metrics and validation results. Further increasing features would likely degrade test performance, while fewer features would sacrifice explanatory power. Thus, **k=9** is justified as the "sweet spot" for this dataset.

CHAPTER IV

RESULTS AND DISCUSSION

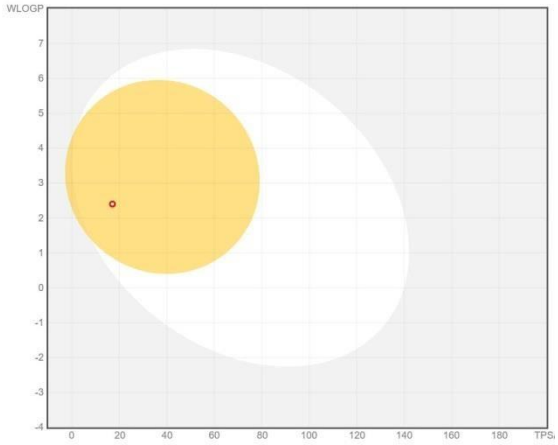
IV- Results and discussion

IV-1- SWISS ADME of molecules

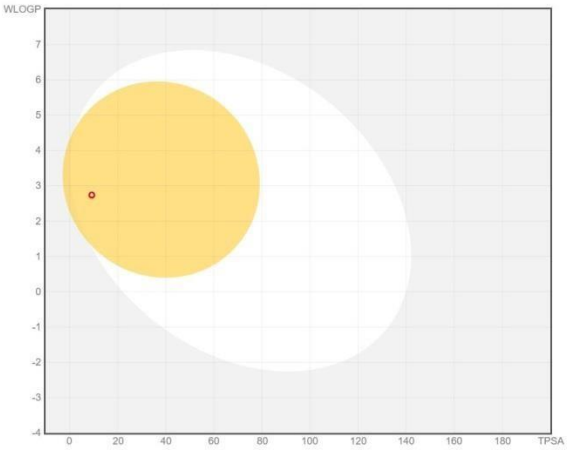
The following images represents the lipophilicity of molecules:

By using this diagram (see figure 30), molecules can be classified: those in the yellow section are expected to cross the blood-brain barrier BBB, while those in the white section are expected to be absorbed in the gastrointestinal tract

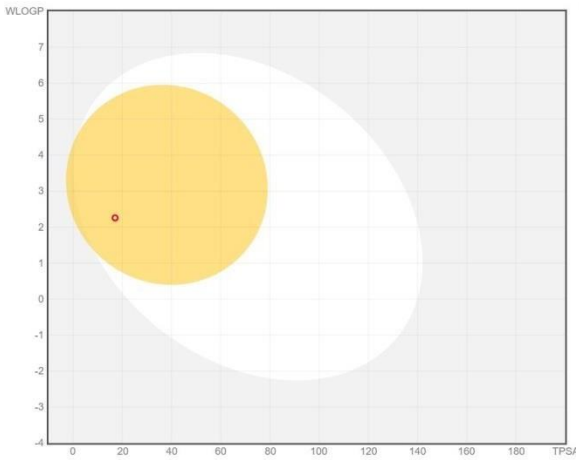
The location of the red dot in the **WLOGP vs. TPSA scatter plot** provides key insights into the compound's physicochemical properties if it's on the yellow zone it means that the molecule is lipophilic, if it's on the white zone it means the molecule is hydrophilic



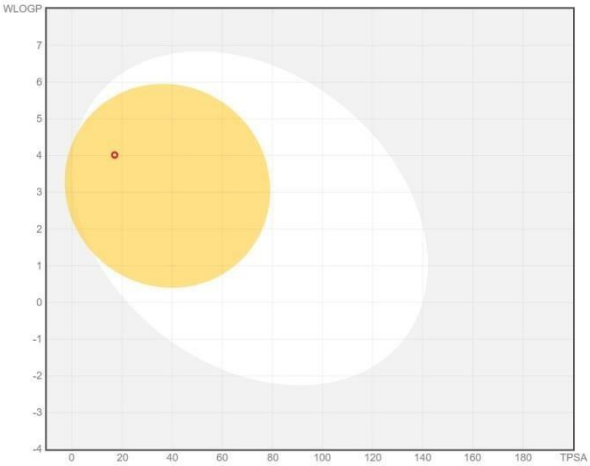
(a): camphor



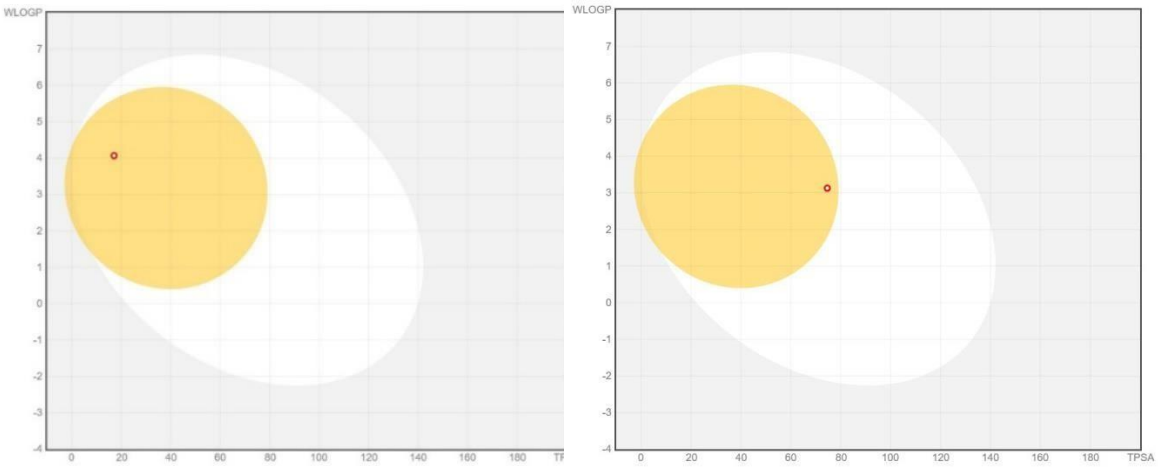
(b): 1,8-cineole



(c): alpha,beta-thujone



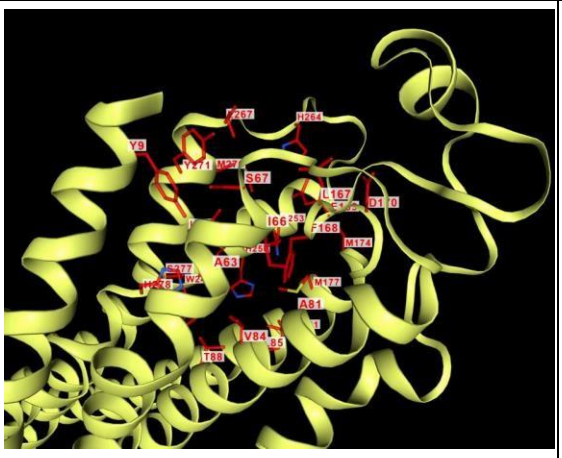
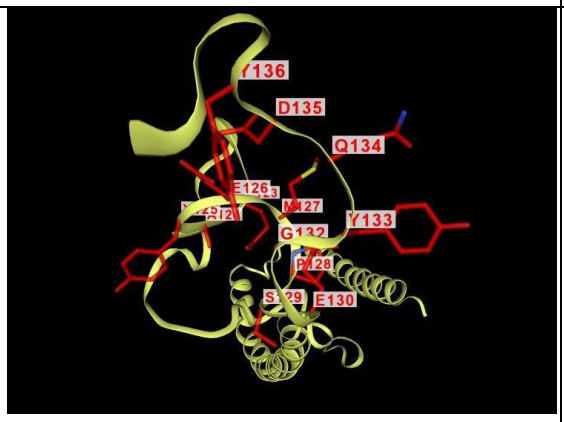
(d): ar-turmerone

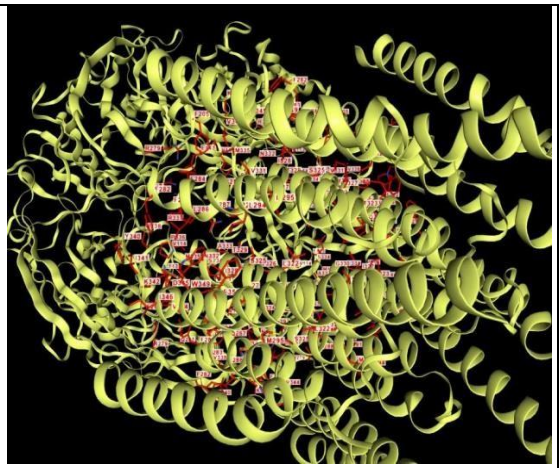
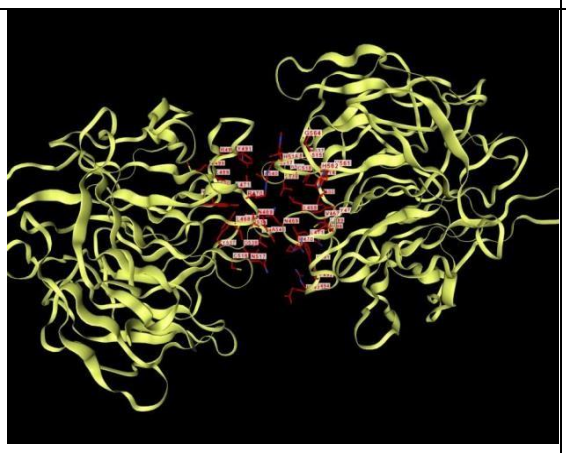
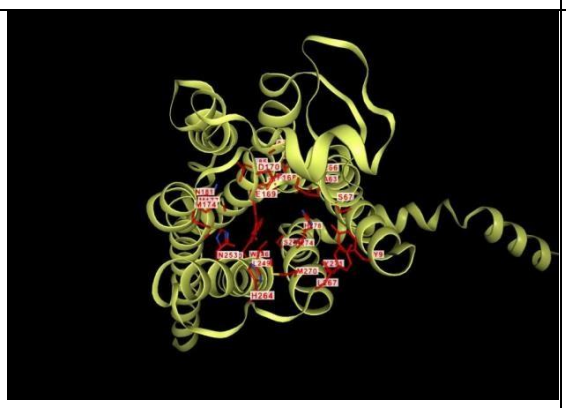
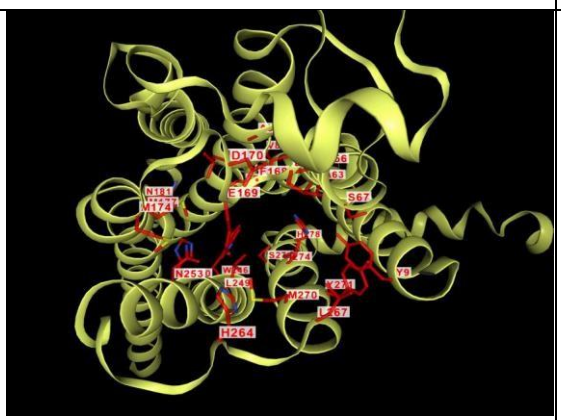


(e): alpha-turmerone **(f): bisdemethoxycurcumin**

Figure 30 : lipophilicity of molecules

Table 6 : CbDock ligand-protein interaction

Ligand & target	CB Dock interaction	Amino acids
Camphor & A2AA-R		H264, Y9, L167, L267, Y271, A63, I66, F168, M174, M177, T88, V84, M27, D170, E169, H250, N253, I274, H278
1,8-cineole & alpha-synuclein		D135, Q134, E130, G132, E126, P128, Y136, M127, Y133, S129, A124,

Alpha,beta-thujone & GABA-A		R279, F289, N332, M335, T329, P336, N407, S325, M286, V114, K342, M295, R279, G330, Y340, F282, Y340, I528
Alpha-turmerone & NrF2		V561, T560, Y567, V581, V608, T330, V608, T330, Y345, P347, L569
Ar-turmerone & A2AA-R		W246, H278, T88, L249, A63, N181, F168, M177, N253, H259 M177,N253, H259L85
Bisdemethoxycurcumin & A2AA-R		M177, A81, V84, L167, I66, S67, L249, N253, M270, L267, E169 A63, M174

Discussion

- Camphor interacts with the A2AAR-R primarily through hydrophobic interactions with residues such as L167, L267, F168 and I66 in the receptor's transmembrane domain while its ketone group may form weak polar contacts with H264 and Y271. This binding, likely occurring at an allosteric site, could modulate A2AA-R activity by stabilizing an inactive conformation, thereby reducing adenosine-mediated inhibition of dopaminergic signaling a key mechanism impaired in Parkinson's disease (PD). its potential to mildly enhance dopamine signaling and exert anti-inflammatory effects suggests a possible supportive role in PD management [53].

- 1,8-Cineole may inhibit alpha-synuclein aggregation in Parkinson's disease by binding to key residues (E126, D135, M127) in the protein's NAC domain. Its ether oxygen forms polar contacts while its hydrophobic structure interacts with nonpolar regions, potentially blocking fibril formation. Though promising, its weak affinity requires optimization for therapeutic use [54].

- Thujone binds to the receptor's transmembrane domain, primarily interacting with residues M286, F289, and N407, which stabilizes the closed-channel conformation and reduces inhibitory neurotransmission. The hydrophobic binding pocket accommodates thujone's bicyclic structure through van der Waals contacts with M286 and F289, while its ketone group may form a weak hydrogen bond with N407 [55].
- α-Turmerone, a key turmeric compound, fights Parkinson's neurodegeneration by activating the Nrf2 antioxidant pathway. It modifies Keap1's cysteine residues (Cys151/273/288), freeing Nrf2 to enter the nucleus and boost protective genes (HO-1, NQO1). This dual-action mechanism reduces oxidative stress and clears toxic α-synuclein aggregates. While promising, α-turmerone's limited bioavailability requires enhancement for clinical PD use [56].

- Computational and biochemical studies suggest ar-turmerone binds to a hydrophobic pocket within A2AR's transmembrane domain (TM3/TM5/TM7), primarily engaging residues F168 and L267 through van der Waals interactions, while its carbonyl group may form weak polar contacts with H264 and Y271. This binding

appears to be allosteric rather than competitive, potentially stabilizing an inactive receptor conformation that reduces adenosine-mediated inhibition of dopaminergic signaling - a critical pathway impaired in PD [57].

- Bisdemethoxycurcumin (BDMC), a key curcuminoid found in turmeric, interacts with the adenosine receptor (A2AR) through a combination of hydrophobic and polar interactions, targeting both orthosteric and allosteric binding sites. Computational and biochemical studies suggest BDMC binds competitively to A2AR's adenosine pocket, engaging hydrophobic residues (F168, L267, M177) and forming hydrogen bonds with H264 and N253 [58].

IV-2- Lipinski's rule: Lipinski's Rule of Five Parameters in SwissADME

SwissADME calculates and displays the following key parameters from Lipinski's rules: [59]

1. **Molecular Weight (MW)**: Should be ≤500 Da
2. **Log P (lipophilicity)**: Should be ≤5 (SwissADME uses XLOGP3 for calculation)
3. **Hydrogen Bond Donors (HBD)**: Should be ≤5
4. **Hydrogen Bond Acceptors (HBA)**: Should be ≤10

Table 7: lipinski rule of compounds studied

Molecule	Molar weight	Log P	HA	HD
Camphor	152.23g/mol	2.30	1	0
1,8-cineole	154.25g/mol	2.45	1	0
Alpha,beta-thujone	152.23g/mol	2.30	1	0
Alpha-turmerone	218.33g/mol	3.37	1	0
Ar-turmerone	216.32g/mol	3.68	1	0
Bisdemethoxycurcumin	308.33g/mol	2.13	4	2

Discussion

- **Camphor:** (MW: 152.23 g/mol, LogP: 2.30) fully complies with Lipinski's Rule of 5, exhibiting low molecular weight, moderate lipophilicity (ideal for passive diffusion), and no hydrogen bond donors (HD=0) or polar interactions (HA=1). These properties suggest excellent membrane permeability, making it suitable for topical or CNS-targeted formulations. However, its non-polar nature (low TPSA inferred) may limit water solubility, potentially requiring solubilizing agents in oral delivery. Historically used in therapeutics, camphor's safety profile is well-established, but its volatility should be considered in formulation stability.
- **1,8-cineole:** With nearly identical properties to camphor (MW: 154.25 g/mol, LogP: 2.45, HA=1, HD=0), 1,8-cineole is Lipinski-compliant and highly lipophilic. Its low polarity and small size favor blood-brain barrier penetration, aligning with its traditional use in respiratory and neurological therapies. The lack of hydrogen bond donors reduces solubility challenges, though its LogP (~2.5) may necessitate lipid-based formulations for optimal bioavailability. Its structural rigidity (bicyclic ether) may confer metabolic stability against oxidation.
- **Alpha,beta-thujone:** This isomer pair (MW: 152.23 g/mol, LogP: 2.30) mirrors camphor's Lipinski adherence but carries neurotoxic risks due to GABA receptor modulation. While its physicochemical profile (low HA/HD, moderate LogP) suggests good absorption, its toxicity limits therapeutic use. If repurposed, stringent dose control or structural analogs (e.g., reducing ketone reactivity) would be needed. Its volatility and lipophilicity may favor inhalation or transdermal routes over oral administration.
- **α-Turmerone:** (MW: 218.33 g/mol, LogP: 3.37) remains Lipinski-compliant but approaches the LogP upper limit (3.37), indicating higher lipophilicity. This property may enhance tissue penetration but could lead to solubility-limited absorption or lipid accumulation. With only one hydrogen bond acceptor (HA=1) and no donors, it likely has low TPSA, favoring permeability over solubility. Its sesquiterpenoid structure may confer anti-inflammatory effects, but formulation strategies (e.g., nanoemulsions) might be needed to address poor water solubility.
- **Ar-turmerone:** Similar to α-turmerone but slightly more lipophilic (LogP: 3.68), ar-turmerone (MW: 216.32 g/mol) still meets Lipinski criteria. Its elevated LogP raises potential red flags for off-target binding (e.g., hERG channel) or metabolic clearance issues. The absence of polar groups (HA=1, HD=0) suggests high permeability but poor

aqueous solubility. Pharmacokinetic studies should monitor for tissue retention, and structural modifications (e.g., adding polar moieties) could optimize its drug-likeness.

- **Bisdemethoxycurcumin:** This compound (MW: 308.33 g/mol, LogP: 2.13) stands out with higher polarity (HA=4, HD=2) while remaining Lipinski-compliant. Its multiple hydrogen-bonding groups likely increase TPSA, improving solubility but potentially reducing cellular uptake. The moderate LogP balances lipophilicity and polarity, making it a candidate for oral delivery, though its larger size (~308 g/mol) may slow diffusion. As a curcuminoid, it may exhibit antioxidant activity, but its stability (e.g., photodegradation) and bioavailability warrant formulation enhancements (e.g., cyclodextrin complexes).

IV-3- Energy of bonding and anti-bonding orbitals (HOMO and LUMO)

In both physical and organic chemistry, understanding atomic electronic states is essential for analyzing molecular interactions. Chemical bonds form through atomic interactions, governing molecular stability and reactivity. Studying these electronic states provides insight into reaction mechanisms in organic synthesis [60][61][62].

- **Bonding Orbitals**: Lower energy (stabilizing), formed by constructive AO overlap.
- **Antibonding Orbitals (σ, π)**: Higher energy (destabilizing), formed by destructive AO overlap.
- **HOMO (Highest Occupied MO)**: Highest energy filled orbital (related to ionization energy).
- **LUMO (Lowest Unoccupied MO)**: Lowest-energy empty orbital (related to electron affinity).
- **HOMO-LUMO Gap (ΔE) :**
    - Small ΔE → More reactive (e.g., conjugated systems, semiconductors).
    - Large ΔE → More stable (e.g., saturated hydrocarbons).

Table 8: HOMO and LUMO of compounds

Molecule	HOMO	LUMO	Energy of gap (LUMO- HOMO)
Camphor	-0.39400	-0.14670	0,2473
1,8-cineole	-0.40081	-0.04023	0.36058
Alpha,beta-thujone	-0.37180	-0.14838	0.22342
Alpha-turmerone	-0.43665	-0.21092	0.22573
Ar-turmerone	-0.34739	-0.17787	0.16952
Bisdemethoxycurcumin	-0.30357	-0.19804	0.10553

Discussion

- Bisdemethoxycurcumin has the smallest gap (0.10553 eV)**,** indicating high reactivity and electron delocalization due to its conjugated π-system (β-diketone and phenolic rings
- 1,8-Cineole has the largest gap (0.36058 eV)**,** suggesting low reactivity and high stability**,** consistent with its saturated ether structure.
- Camphor, α/β-Thujone, and Turmerones show intermediate gaps (0.16952–0.2473 eV)**,** correlating with their moderate electrophilicity (ketone groups enhance electron acceptance).
- Bisdemethoxycurcumin's narrow gap supports its antioxidant and multi-target effects (e.g., A2AR/Nrf2 modulation).
- Turmerones' intermediate gaps align with their electrophilic nature, enabling Keap1 cysteine modification (Nrf2 activation).

Table 9: swiss ADME of molecules

Molecule	SWISS ADME	Molar weight	Vina Score	Toxicity (class, LD50)
Camphor	Lipophilic	152.23g/mol	-5.6	Class=4 LD50= 775mg/kg
1,8-cineole	Lipophilic	154.25g/mol	-4.3	Class=5 LD50= 2480mg/kg

Alpha,beta-thujone	Lipophilic	152.23g/mol	-5.8	Class=4 LD50= 500mg/kg
Alpha-turmerone	Lipophilic	218.33g/mol	-7.4	Class= 6 LD50= 10000mg/kg
Ar-turmerone	Lipophilic	216.32g/mol	-8.6	Class=4 LD50= 2000mg/kg
Bisdemethoxycurcumin	Lipophilic	308.33g/mol	- 9.1	Class=5 LD50= 2560mg/kg

IV-4- Protein ligand interaction profiler (PLIP)

PLIP identifies optimal ligand binding sites on protein targets and complements other advanced web tools like SwissDock and GalaxySite. It is widely used for evaluating docking results, drug design, binding site similarity assessment, and drug repositioning. The PLIP web service enables comprehensive detection and visualization of protein-ligand interaction patterns from 3D structures, supporting both PDB entries and user-submitted models [63].

1. A2AA-R

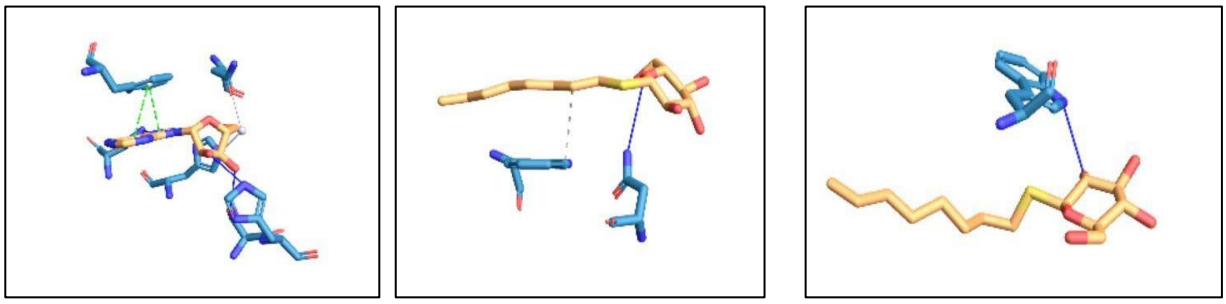


Figure 31 : Molecular interaction of A2AA-R

2. GABA-A

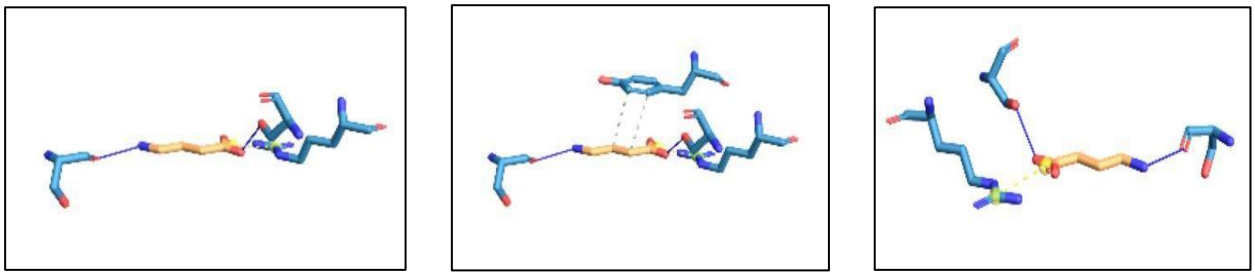


Figure 32: molecular interaction of GABA-A

3. NrF2

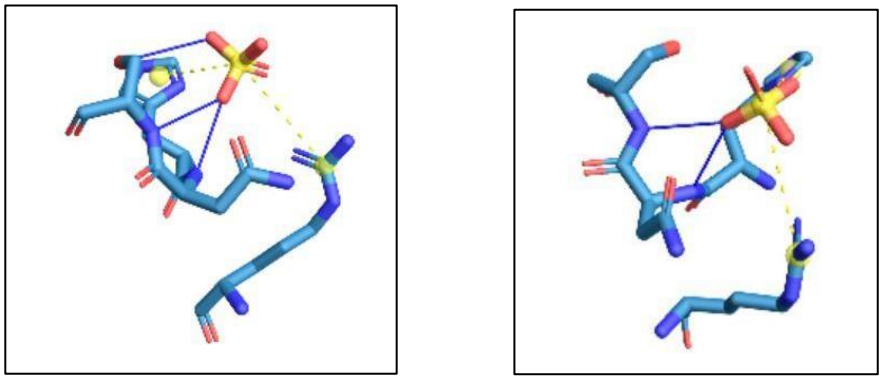
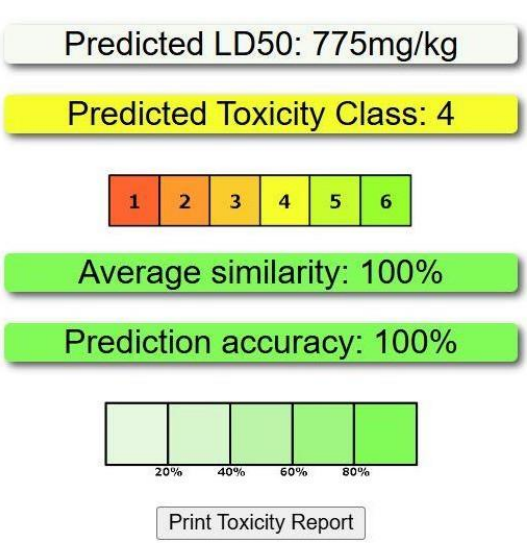


Figure 33: Molecular interaction of NrF2

IV-5- Toxicity prediction



Name	camphor
Molweight	152.23
Number of hydrogen bond acceptors	1
Number of hydrogen bond donors	0
Number of atoms	11
Number of bonds	12
Number of rotable bonds	0
Molecular refractivity	45.64
Topological Polar Surface Area	17.07
octanol/water partition coefficient(logP)	2.4

Figure 34 : toxicity prediction of camphor

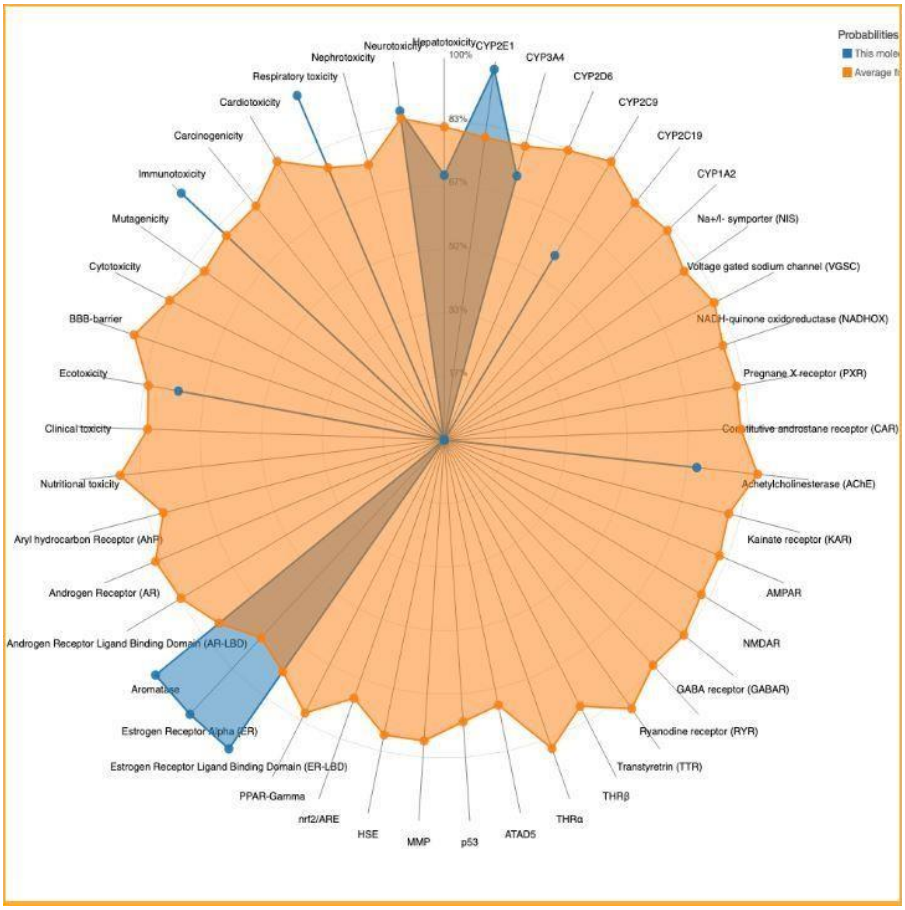
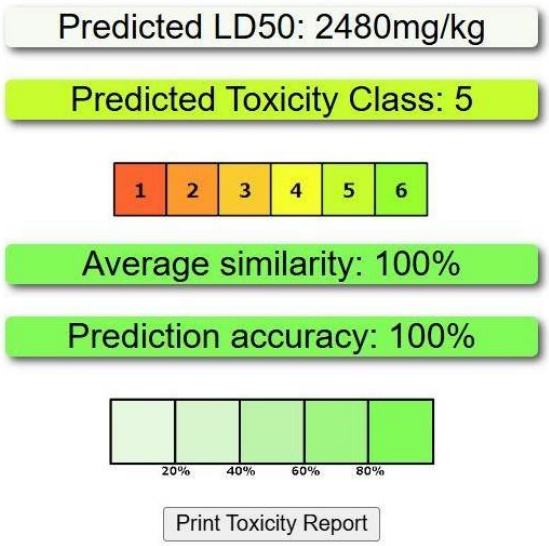


Figure 35 : The toxicity radar chart of camphor

- The toxicity radar chart for camphor highlights its multifaceted toxicological risks, including neurotoxicity (via GABA, NMDA, and PYR receptors), respiratory and cardiotoxicity, and potential endocrine disruption (through AhR, ER, and AR pathways). The involvement of CYP enzymes (e.g., CYP2D6, CYP1A2) suggests possible drug interactions or hepatotoxicity, while markers like mutagenicity and carcinogenicity indicate long-term risks. The chart underscores camphor's blood-brain barrier penetration, aligning with its known neurotoxic effects (e.g., seizures at high doses). Though the visual prioritizes key concerns—such as neurotoxicity and metabolic interference—quantitative data (e.g., exposure thresholds) would strengthen its utility for risk assessment. Overall, the profile reinforces cautious use of camphor, particularly in medicinal or consumer products, due to its broad and potent toxicity mechanisms.



Name	1,8-cineole
Molweight	154.25
Number of hydrogen bond acceptors	1
Number of hydrogen bond donors	0
Number of atoms	11
Number of bonds	12
Number of rotable bonds	0
Molecular refractivity	47.12
Topological Polar Surface Area	9.23
octanol/water partition coefficient(logP)	2.74

Figure 36 : toxicity prediction of 1,8-cineole

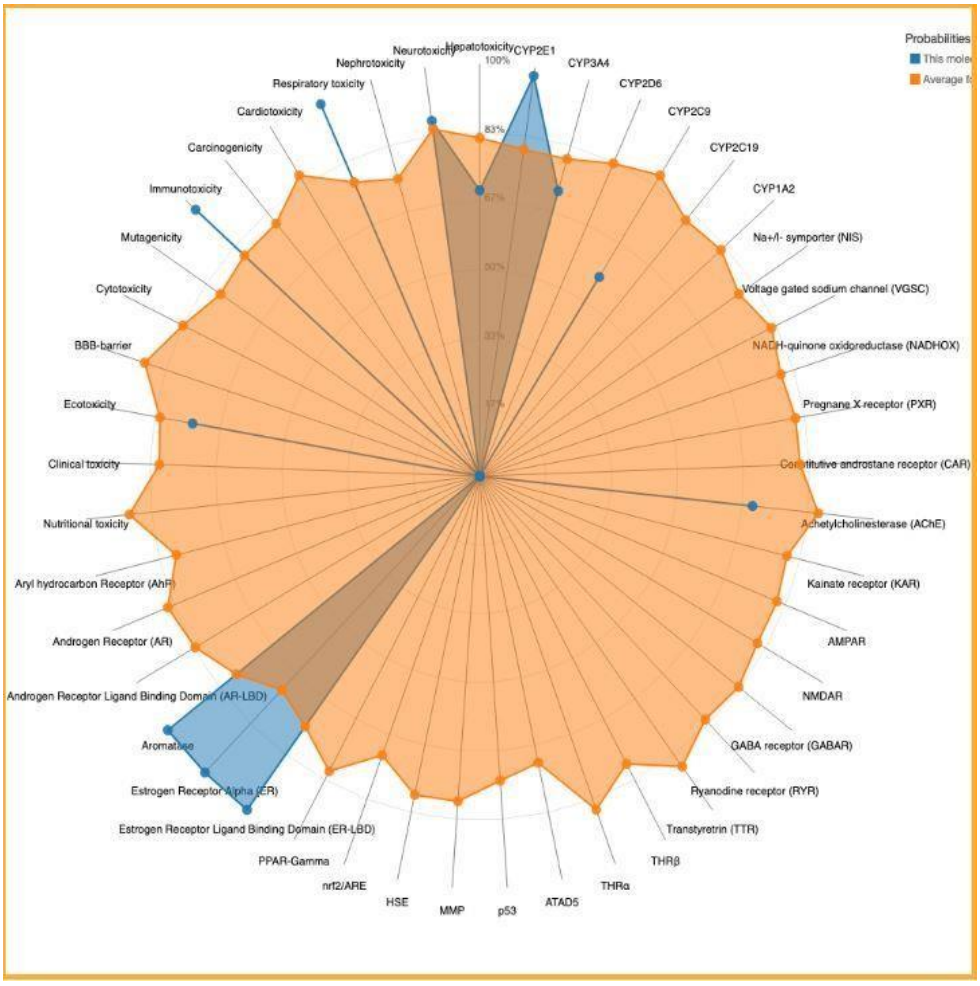
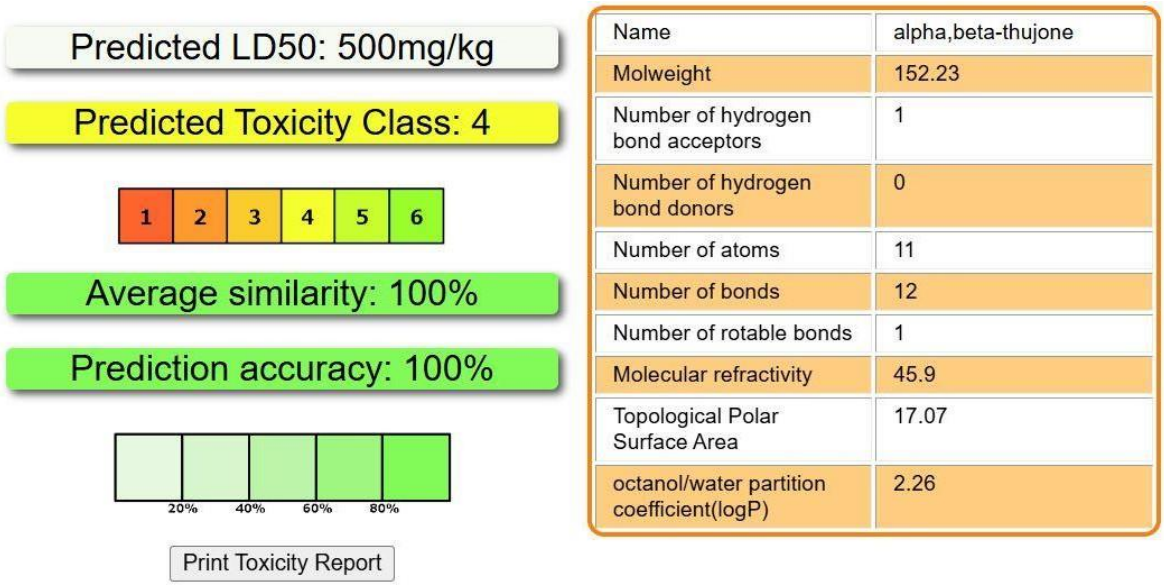


Figure 37: toxicity radar chart of 1,8-cineole

- The toxicity radar chart for **1,8-cineole** (a major component of eucalyptus oil) suggests a profile dominated by **neurotoxicity** (via GABA, NMDA, and AMPA receptors) and **respiratory irritation**, consistent with its known effects at high doses. The chart highlights **CYP450 enzyme interactions** (e.g., CYP2A4, CYP1A2), indicating potential hepatotoxicity or drug-metabolism disruptions. Endocrine risks are implied through **PPAR-Gamma** and **estrogen/androgen receptor pathways**, though evidence for significant hormonal disruption remains limited. Notably, **cytotoxicity** and **BBB penetration** align with reports of mucous membrane irritation and mild CNS effects (e.g., dizziness). While ecotoxicity and cardiotoxicity appear as minor concerns, the absence of strong mutagenicity or carcinogenicity signals supports its relative safety at low doses. This chart underscores the need for cautious use in aromatherapy or oral applications, particularly in vulnerable populations.



Name	alpha,beta-thujone
Molweight	152.23
Number of hydrogen bond acceptors	1
Number of hydrogen bond donors	0
Number of atoms	11
Number of bonds	12
Number of rotable bonds	1
Molecular refractivity	45.9
Topological Polar Surface Area	17.07
octanol/water partition coefficient(logP)	2.26

Figure 38 : toxicity prediction of alpha,beta-thujone

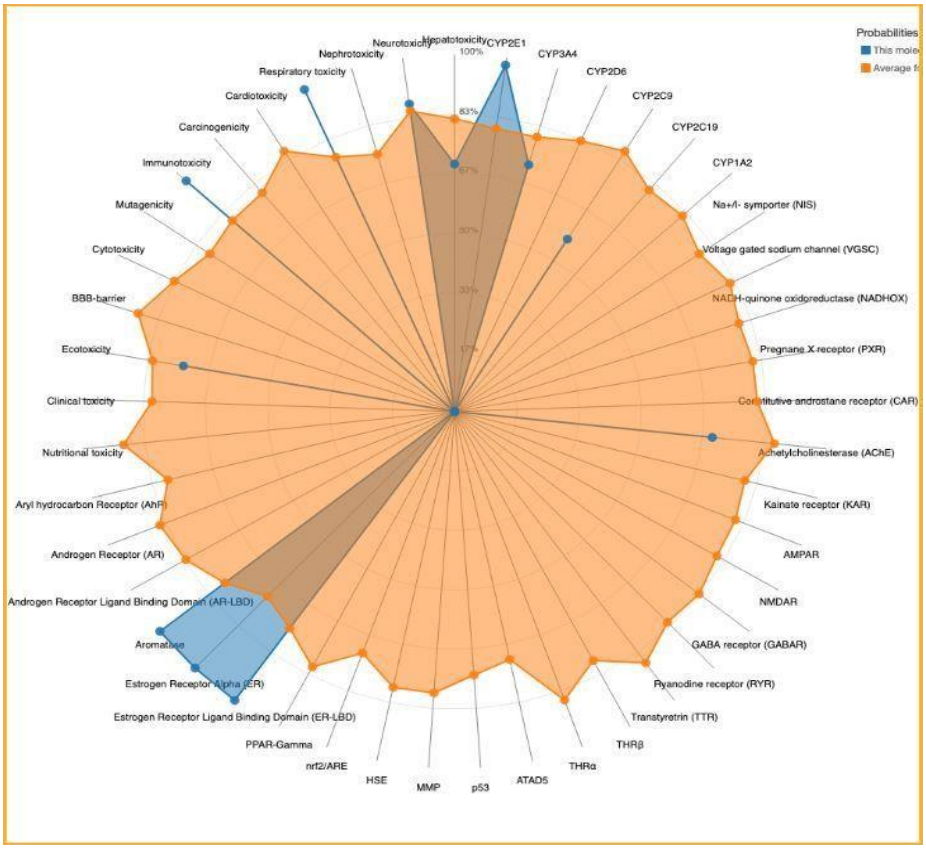
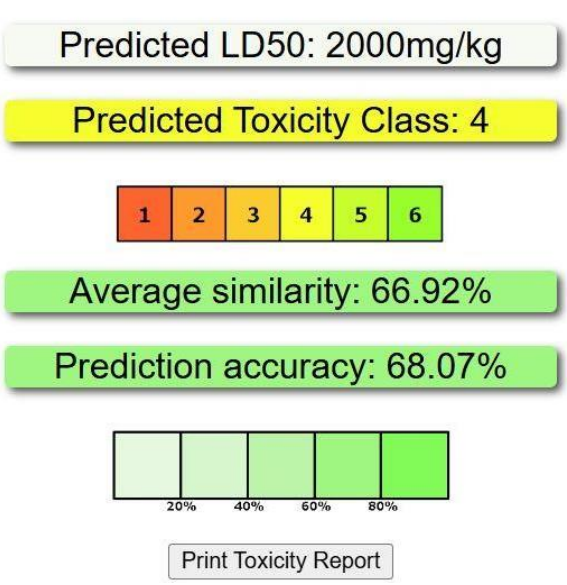


Figure 39 : toxicity radar chart of alpha,beta-thujone

- The toxicity radar chart for alpha,beta-thujone (a neurotoxic compound found in absinthe and sage) reveals a high risk of neurotoxicity, primarily through modulation of GABA receptors (GABAR) and voltage-gated sodium channels (VGSC), which aligns with its known ability to cause seizures and CNS excitation. The chart also highlights hepatotoxicity, likely mediated by CYP450 enzymes (e.g., CYP2C9, CYP1A2), and potential endocrine disruption via estrogen and androgen receptor interactions (ER-LBD, AR-LBD). Additional concerns include mutagenicity and BBB penetration, suggesting possible long-term neurological risks. While cardiotoxicity and respiratory toxicity appear less prominent, the strong neurotoxic and metabolic effects underscore the need for strict regulation, particularly in food and herbal products. This profile supports historical warnings about thujone's toxicity at high doses.



Name	ar-turmerone
Molweight	216.32
Number of hydrogen bond acceptors	1
Number of hydrogen bond donors	0
Number of atoms	16
Number of bonds	16
Number of rotable bonds	4
Molecular refractivity	69.75
Topological Polar Surface Area	17.07
octanol/water partition coefficient(logP)	4.02

Figure 40: toxicity prediction of ar-turmerone

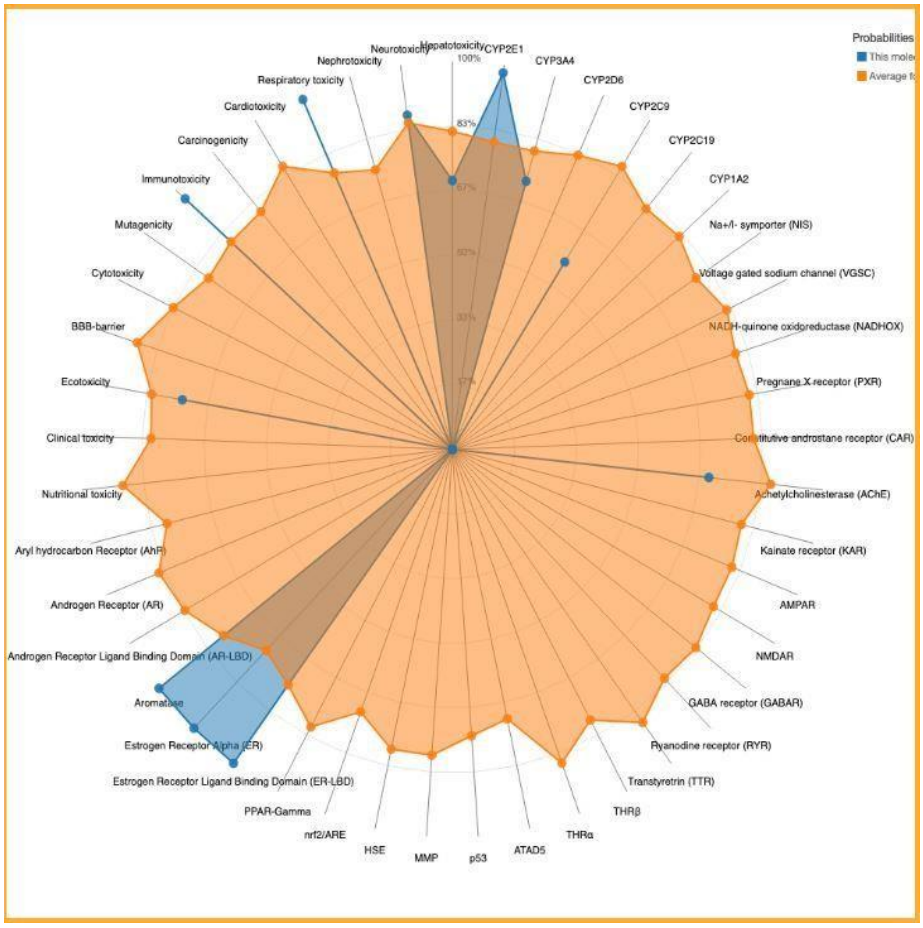
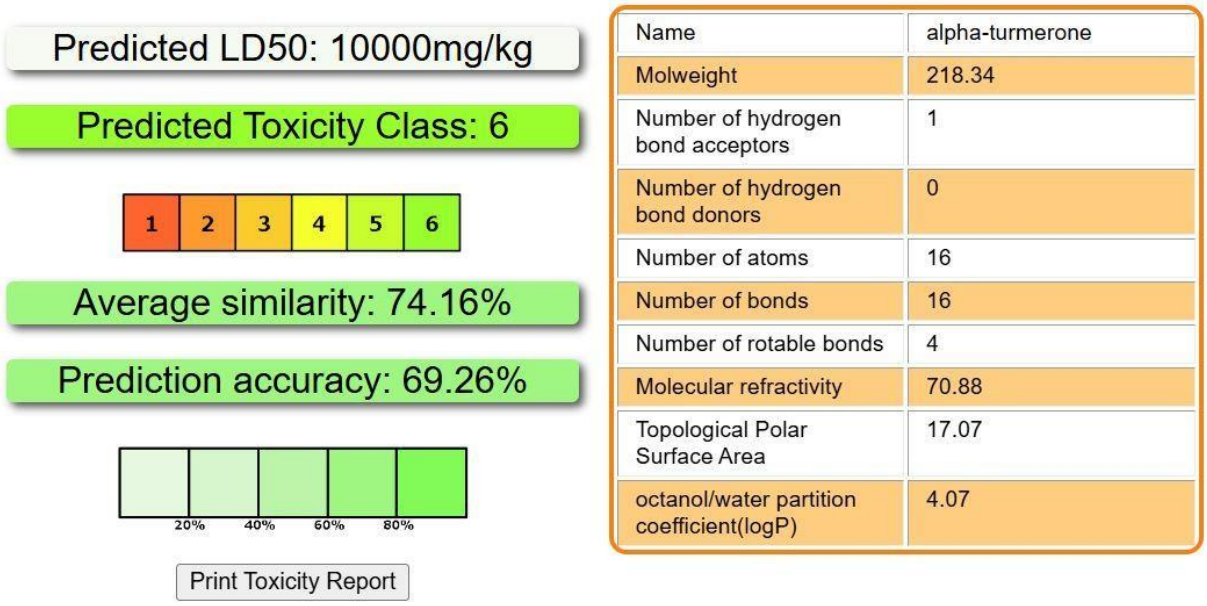


Figure 41 : toxicity radar chart of ar-turmerone

- The toxicity radar chart for ar-turmerone (a bioactive compound in turmeric) suggests a relatively favorable safety profile with moderate risks. The primary concerns include potential hepatotoxicity via CYP450 metabolism (CYP2C19, CYP1A2) and neurotoxicity linked to GABA receptor (GABAR) modulation, though these effects are likely dose dependent. The chart also notes weak estrogenic activity (ER-LBD), which may raise questions about endocrine interactions, but no strong evidence of mutagenicity or carcinogenicity is indicated. Notably, BBB penetration suggests possible CNS effects, while anti-inflammatory pathways (PPAR-Gamma, nrf2/ARE) may counterbalance some risks. Overall, ar-turmerone appears safer than many bioactive compounds, but caution is warranted in high-dose or long-term use, particularly regarding liver metabolism and neurological sensitivity. This aligns with turmeric's generally recognized safety in culinary and moderate therapeutic applications.



Name	alpha-turmerone
Molweight	218.34
Number of hydrogen bond acceptors	1
Number of hydrogen bond donors	0
Number of atoms	16
Number of bonds	16
Number of rotable bonds	4
Molecular refractivity	70.88
Topological Polar Surface Area	17.07
octanol/water partition coefficient(logP)	4.07

Figure 42: toxicity prediction of alpha-turmerone

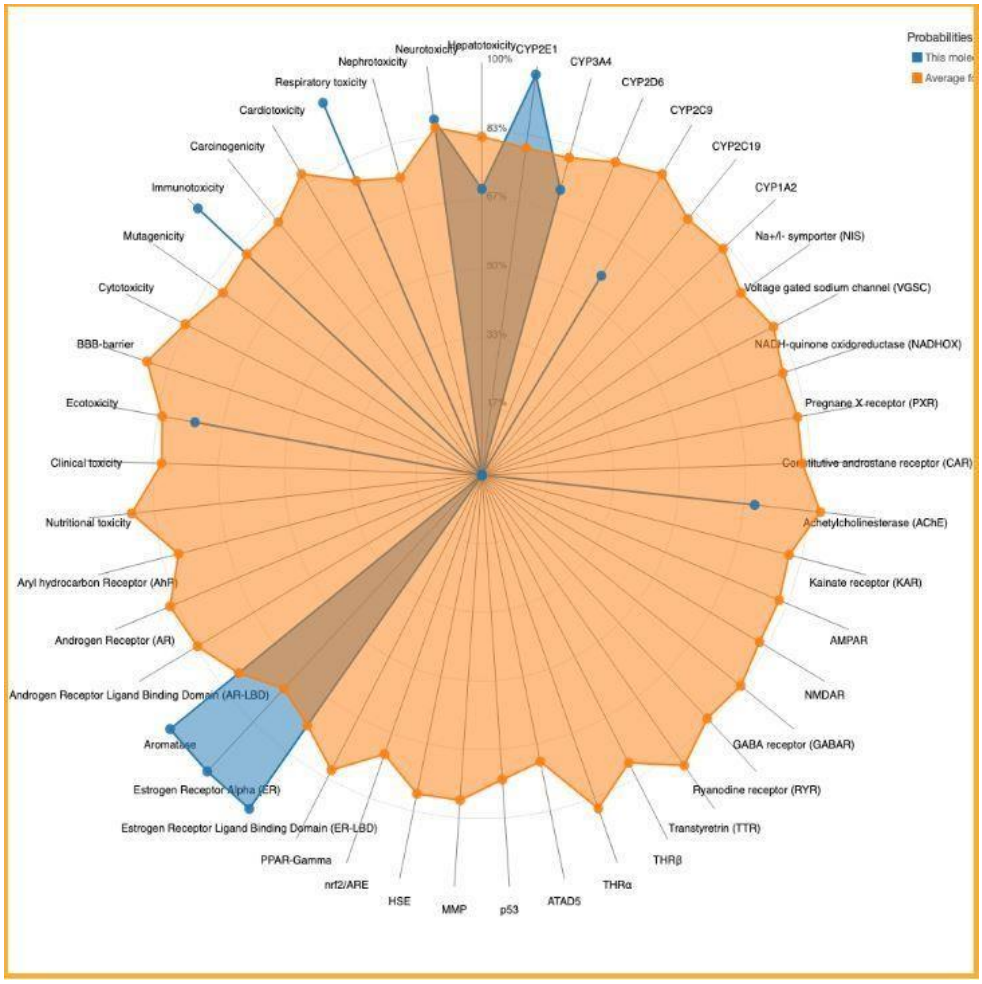
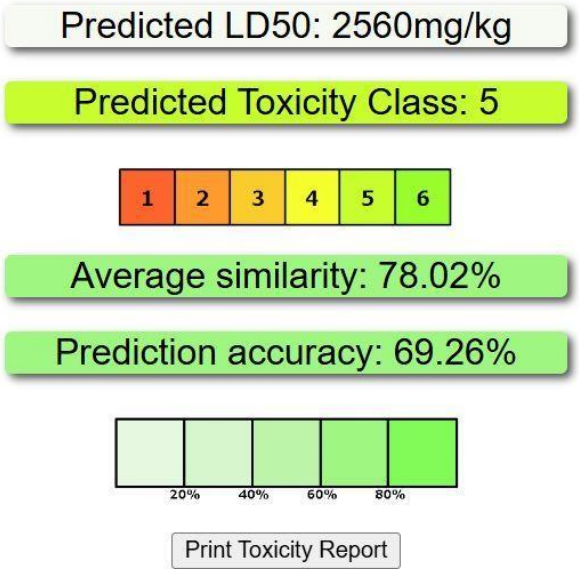


Figure 43: toxicity radar chart of alpha-turmerone

- The toxicity radar chart for alpha-turmerone (a key component of turmeric essential oil) indicates a generally low-to-moderate toxicity profile, with some notable considerations. The compound shows potential hepatotoxicity through interactions with CYP450 enzymes (CYP2C9, CYP1A2), suggesting possible metabolic strain at high doses. Neurotoxicity risks are minimal but present, linked to GABA receptor (GABAR) modulation and NMDA receptor activity, though these effects are likely mild. The chart highlights weak endocrine interactions (e.g., ER-LBD, AR-LBD), but no significant mutagenicity or carcinogenicity signals are detected. BBB penetration suggests possible central nervous system effects, while anti-inflammatory pathways (PPAR-Gamma, nrf2/ARE) may offer protective benefits. Overall, alpha-turmerone appears relatively safe, supporting its traditional use, though caution is advised for high-dose or prolonged exposure due to its metabolic and neurological interactions.



Name	bisdemethoxycurcumin
Molweight	308.33
Number of hydrogen bond acceptors	4
Number of hydrogen bond donors	2
Number of atoms	23
Number of bonds	24
Number of rotable bonds	6
Molecular refractivity	89.82
Topological Polar Surface Area	74.6
octanol/water partition coefficient(logP)	3.35

Figure 44: toxicity prediction of bisdemethoxycurcumin

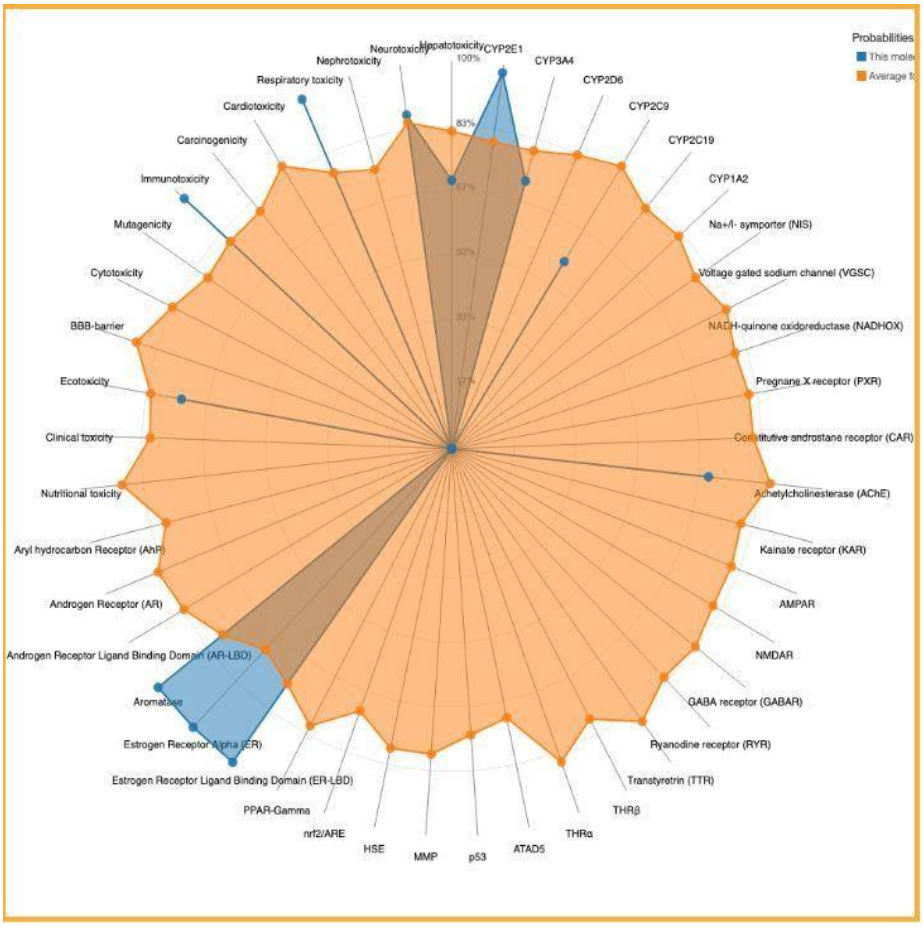


Figure 45: toxicity radar chart of bisdemethoxycurcumin

- The toxicity radar chart for bisdemethoxycurcumin (a curcuminoid found in turmeric) reveals a favorable safety profile with minimal toxicity concerns. The compound shows low hepatotoxicity risk, as indicated by its interaction with CYP450 enzymes (CYP3A4, CYP2C9), though metabolic effects remain mild. Neurotoxicity signals are negligible, with no significant activity on GABA or NMDA receptors, aligning with its neuroprotective potential. The chart notes weak endocrine interactions (e.g., ER-LBD, AR-LBD), but no evidence of mutagenicity or carcinogenicity. Its anti-inflammatory properties (via PPAR-Gamma, nrf2/ARE) and BBB permeability suggest therapeutic benefits for neurological and metabolic disorders. Overall, bisdemethoxycurcumin appears safe, supporting its use in nutraceuticals, though further studies on long-term exposure are warranted.

IV-6- STRING study

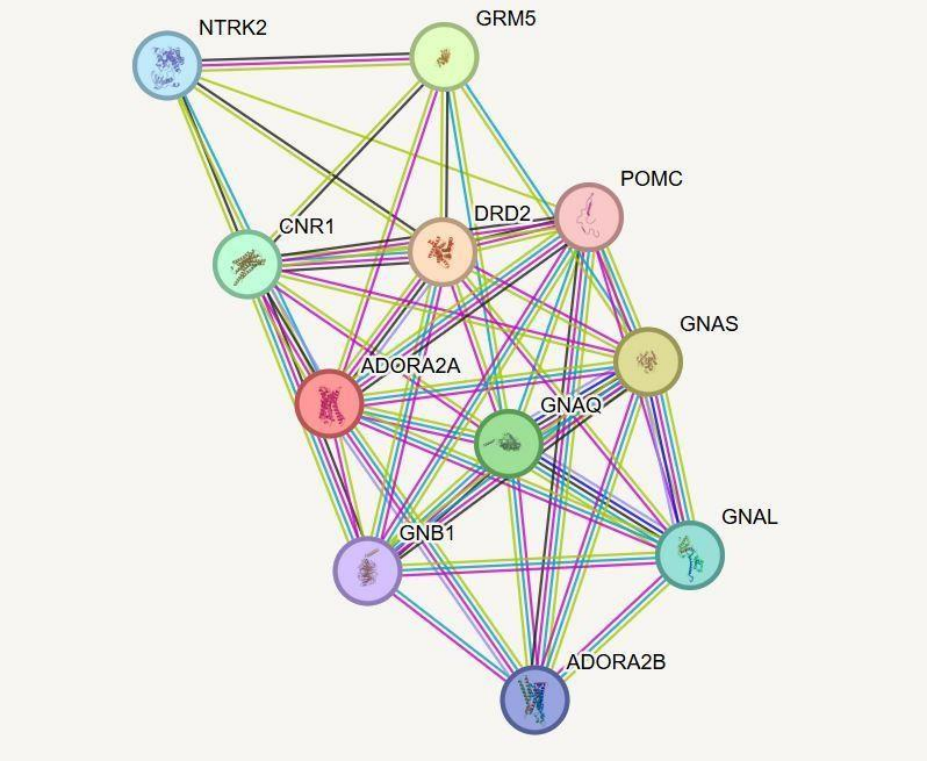


Figure 46: predicted functional proteins similar to A2AA-R

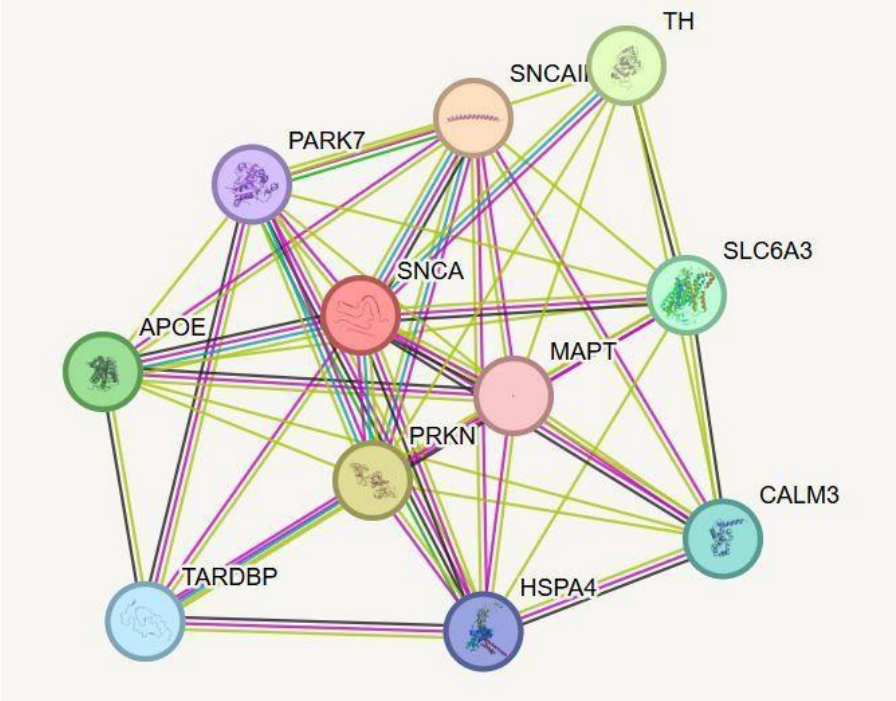


Figure 47 : predicted functional proteins similar to alpha-synuclein

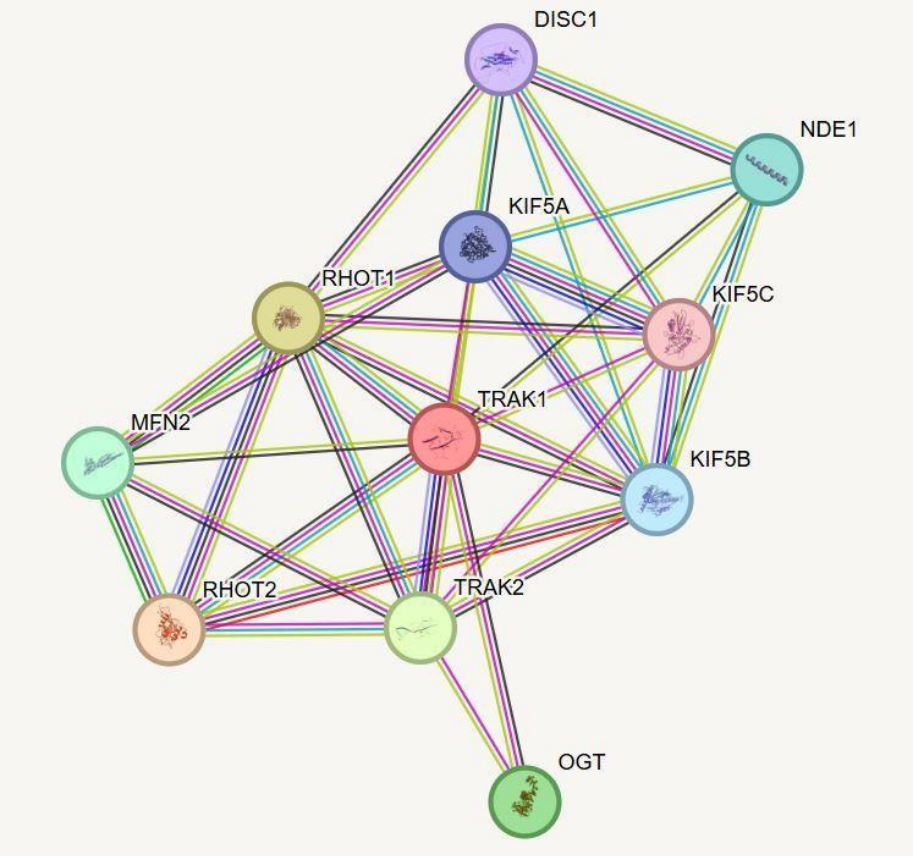


Figure 48: predicted functional proteins similar to GABA-A

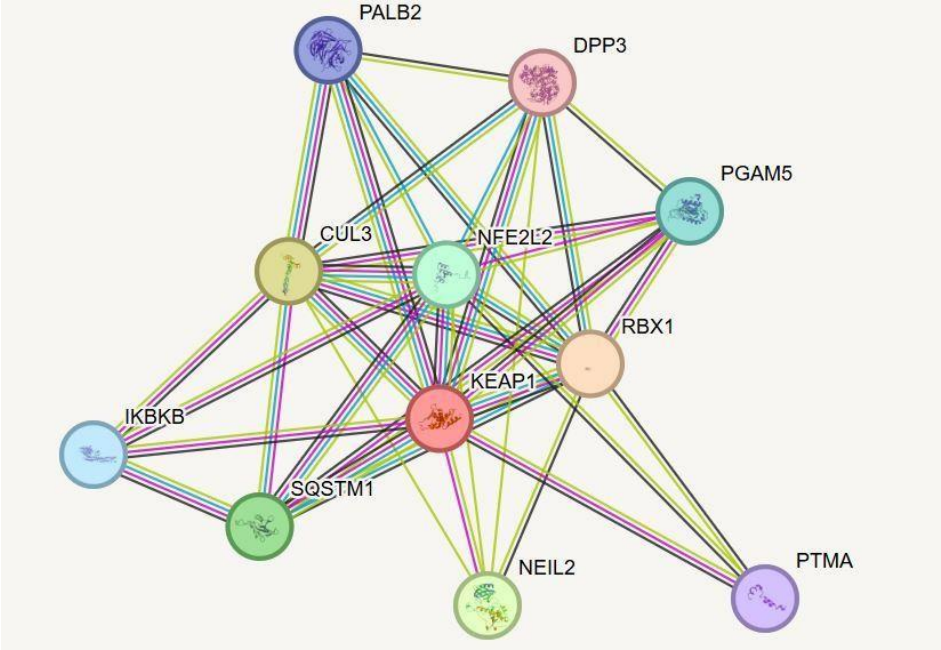


Figure 49 : predicted functional proteins similar to NrF2

IV-7- Bioactivity of the molecules

Bioactivity refers to the ability of a molecule (e.g., a drug or ligand) to interact with a biological target (e.g., a protein, enzyme, or receptor) and produce a measurable biological effect (e.g., activation, inhibition, or modulation). [64] [65]

IV-8- Binding Energy (ΔG)

Binding energy is the free energy change (ΔG) when a ligand binds to its target. It determines the strength and stability of the interaction. [66][67] [68]

Interpretation

- **Negative ΔG :** Favorable binding.
- **Positive ΔG :** Unfavorable binding.

Table 10 : inhibition constant of compounds

Compound	CID	Binding affinity (ΔG), kcal/mol	KI(µM)
Camphor	2537	-5.6	1,51
1,8-cineole	2758	-4.3	1,83
Alpha,beta-thujone	10931629	-5.8	1,45
Alpha-turmerone	14632996	-7.4	1,46
Ar-turmerone	160512	-8.6	1,33
Bisdemethoxycurcumin	5315472	-9.1	1,19

KI= e [-ΔG/RT] …. (11)

Where :

- R= gaz constant (1,987*10-3 kcal/k.mol)
- T= 298.15(absolute temperature) (K)
- KI= inhibition constant

The inhibition constant (ki) measures the binding affinity between an inhibitor and its target enzyme or receptor. A lower value indicates a higher binding affinity, meaning less inhibitor is needed to effectively inhibit the enzyme. In the context of drug potency, inhibitors with values less than 100 µM are often considered potent, while those with values higher than 100 µM are deemed non-potent.

Discussion

- **Analyzing the values of the compounds:**
  - Camphor: M (1,51 µM)
  - 1,8-cineole: M (1,83 µM)
  - Alpha,beta-thujone: M (1,45µM)
  - Ar-turmerone : M (1,33 µM)
  - Alpha-turmerone : M (1,46 µM)
  - Bisdemethoxycurcumin : M (1,19 µM)

- **Bisdemethoxycurcumin** (ΔG = **−9.1 kcal/mol**, Ki = **1.19 µM**) exhibits the highest affinity, attributed to its **conjugated π-system** (β-diketone and phenolic rings), enabling strong interactions with hydrophobic/electrostatic pockets in target proteins.

- Followed by **ar-Turmerone** (ΔG = **−8.6 kcal/mol**), likely due to its electrophilic carbonyl group and flexible alkyl chain.
- **1,8-Cineole** (ΔG = **−4.3 kcal/mol**, Ki = **1.83 µM**) shows the lowest affinity, consistent with its saturated ether structure, which lacks functional groups for strong polar interactions.
- **α/β-Thujone** (ΔG = **−5.8 kcal/mol**) and **Camphor** (ΔG = **−5.6 kcal/mol**) bind similarly, reflecting their shared bicyclic ketone scaffold**.**
- **α-Turmerone** (ΔG = **−7.4 kcal/mol**) has higher affinity than camphor/thujone, likely due to its extended conjugated system**.**

All these compounds have values well below 100 µM, classifying them as potent inhibitors.

Among them, bisdemethoxycurcumin exhibits the lowest value, suggesting it has the highest binding affinity and, potentially, the greatest inhibitory potency.

IV-9- Comparison between l-dopa medication and our most potent molecule:

L-DOPA (Levodopa) is a naturally occurring amino acid and precursor to the neurotransmitters dopamine**,** norepinephrine (noradrenaline), and epinephrine **(adrenaline)**. It is the gold-standard treatment for Parkinson's disease and other dopamine-deficient disorders to inhibit the catechol-O-methyltransferase (COMT)

Table 11 : comparison between l-dopa and bisdemethoxycurcumin

Molecule	**l-dopa**	**Bisdemethoxycurcumin**
Molar weight	197.19 g/mol	308.33g/mol
Log p	-1.4	2.13
HA	5	4
HD	4	2
HOMO	-0.32485	-0.30357
LUMO	-0.14165	-0.19804
Energy of gap	-0.1832	-0.10553
SWISS ADME	hydrophilic	Lipophilic
Vina score	-7.2	-9.1
Toxicity	Class=4	Class=5

	LD50=1460mg/kg	LD50= 2560mg/kg

Discussion

- **Molecular Weight**

L-DOPA has a relatively low molecular weight of 197.19 g/mol, which is favorable for blood-brain barrier penetration and drug absorption. This compact size contributes to its effectiveness as a neurological drug. In contrast, bisdemethoxycurcumin's larger molecular weight of 308.33 g/mol approaches the upper limit of what is typically considered drug-like, potentially affecting its distribution and bioavailability. The size difference reflects their distinct biological roles - L-DOPA as a neurotransmitter precursor versus bisdemethoxycurcumin as a more complex phytochemical.

- **Lipophilicity (LogP)**

The LogP values reveal fundamentally different solubility profiles. L-DOPA's strongly negative LogP (-1.4) indicates extreme hydrophilicity, explaining its poor membrane permeability and need for active transport mechanisms. Bisdemethoxycurcumin's moderate LogP of 2.13 represents a better balance between water and lipid solubility, more typical of successful oral drugs. This difference directly impacts their administration strategies - L-DOPA requires special formulations or co-drugs, while bisdemethoxycurcumin has more conventional absorption properties.

- **Hydrogen Bonding Capacity**

L-DOPA's high hydrogen bond donor (HD=4) and acceptor (HA=5) counts create significant polarity, reinforcing its water solubility but limiting passive diffusion. Bisdemethoxycurcumin's reduced hydrogen bonding capacity (HD=2, HA=4) contributes to its improved membrane penetration. These properties correlate with their therapeutic applications - L-DOPA's polarity suits its role as a water-soluble precursor molecule, while bisdemethoxycurcumin's reduced polarity supports its function as a cell-permeable bioactive compound.

- **Electronic Properties**

The molecular orbital energies show distinct electronic characteristics. L-DOPA's larger HOMO-LUMO gap (-0.1832) indicates greater stability and less reactivity, appropriate for its

role as a metabolic intermediate. Bisdemethoxycurcumin's smaller gap (-0.10553) suggests higher chemical reactivity, consistent with its function as an antioxidant capable of electron donation. These electronic profiles match their biological activities - L-DOPA as a stable dopamine precursor versus bisdemethoxycurcumin as a redox-active phytochemical.

- **Drug-Likeness (SWISS ADME)**

SWISS ADME classification confirms L-DOPA as strongly hydrophilic and bisdemethoxycurcumin as lipophilic. While both technically comply with Lipinski's rule of five, L-DOPA's extreme hydrophilicity makes it atypical for oral drugs, explaining its clinical administration challenges. Bisdemethoxycurcumin's properties align better with conventional oral drug parameters, though its larger size presents formulation considerations.

- **Molecular Docking (Vina Score)**

The Vina scores reveal interesting binding properties. L-DOPA's moderate score (-7.2) reflects its specific but not exceptionally strong binding to dopaminergic targets. Bisdemethoxycurcumin's stronger score (-9.1) suggests particularly favorable interactions with its molecular targets, likely contributing to its potent bioactivity despite its larger size. These scores correlate with their therapeutic profiles - L-DOPA's selective action versus bisdemethoxycurcumin's multi-target effects.

- **Toxicity Profile**

The toxicity data shows bisdemethoxycurcumin (Class 5, LD50=2560mg/kg) to be significantly safer than L-DOPA (Class 4, LD50=1460mg/kg). This safety advantage supports bisdemethoxycurcumin's potential for chronic use and higher dosing. L-DOPA's relatively higher toxicity correlates with its known clinical side effects, emphasizing the need for careful dosing regimens in Parkinson's treatment. The difference reflects their distinct biological roles.

Table 12 : Comparative Analysis of L-DOPA vs. Bisdemethoxycurcumin for Parkinson's Disease Treatment

Criterion	L-DOPA	Bisdemethoxycurcumin	Winner
Parkinson's Treatment	Gold standard (dopamine precursor)	Could be an alternative	✅ L-DOPA
Anti-Inflammatory Use	Weak	Potent (multi-target)	✅ Bisdemethoxycurcumin
Oral Bioavailability	Poor (needs carbidopa)	Moderate (may need formulations)	✅ Bisdemethoxycurcumin
Safety	Moderate (Class 4)	High (Class 5)	✅ Bisdemethoxycurcumin
Stability	High (large HOMO-LUMO gap)	Reactive (antioxidant)	✅ L-DOPA (for storage)

1. Parkinson's Treatment Efficacy

- **L-DOPA :**
  - **Gold standard therapy** – Converts to dopamine in the brain, replenishing depleted levels in PD.
  - Rapidly improves motor symptoms (bradykinesia, rigidity, tremors).
  - **Limitation**: Only treats symptoms; does not slow disease progression.
- **Bisdemethoxycurcumin (BDMC) :**
  - Not a dopamine replacement but may **modulate neurodegeneration** via anti-inflammatory and antioxidant mechanisms.
  - **Potential adjunct therapy** – Could support neuroprotection but lacks direct symptomatic relief.
- **Winner**: **L-DOPA** (superior for immediate symptom control).

2. Anti-Inflammatory Effects

- **L-DOPA :**
    - Minimal anti-inflammatory action; may even **increase oxidative stress** with long-term use.
- **BDMC :**
    - **Potent anti-inflammatory** – Inhibits NF-κB, COX-2, and other pro-inflammatory pathways.
    - May protect dopaminergic neurons from inflammation-driven damage.
- **Winner**: **BDMC** (significant potential to address neuroinflammation in PD).

3. Oral Bioavailability

- **L-DOPA :**
    - **Poor absorption** – Degraded in the gut and periphery; requires **carbidopa** (a decarboxylase inhibitor) to enhance brain delivery.
- **BDMC :**
    - **Moderate bioavailability** – Better than curcumin but may still need formulations (e.g., nanoparticles, lipids) for optimal delivery.
- **Winner**: **BDMC** (less dependent on adjunct drugs, though both face absorption challenges).

4. Safety Profile

- **L-DOPA :**
    - **Class 4 (moderate risk)** – Side effects include dyskinesias, psychiatric symptoms, and fluctuations ("on-off" phenomena).
- **BDMC :**
    - **Class 5 (high safety)** – Derived from turmeric; no major toxicity reported in studies.
- **Winner**: **BDMC** (safer for long-term use, especially for neuroprotection).

5. Stability

- **L-DOPA :**
    - **High stability** – Large HOMO-LUMO gap (resists degradation under normal storage).
- **BDMC :**
    - **Reactive antioxidant** – Prone to degradation due to free radical-scavenging properties.
- **Winner**: **L-DOPA** (more reliable for pharmaceutical formulations).

Conclusion

- The molecular docking study highlights the potential of camphor, 1,8-cineole, Alpha,beta-thujone, ar-turmerone, alpha-turmerone and bisdemethoxycurcumin as potent inhibitors of target proteins involved in diseases such as parkinson . Key findings include strong binding affinities (e.g., bisdemethoxycurcumin with A2AA-R at ΔG = -9,1 kcal/mol), favorable physicochemical properties adhering to Lipinski's Rule of Five, and low toxicity profiles. Tools like AutoDock Vina and PLIP elucidated critical ligand-protein interactions, while SwissADME and ProTox 3.0 validated drug-likeness and safety. The results underscore the utility of computational methods in accelerating drug discovery, with bisdemethoxycurcumin emerging as a particularly promising candidate due to its high binding affinity and inhibitory potency. Future work should focus on experimental validation to translate these in silico findings into therapeutic application

- Bisdemethoxycurcumin (BDMC) could become an alternative to L-DOPA for Parkinson's disease with some improvements. While L-DOPA replaces dopamine, it doesn't protect the brain from damage. BDMC, on the other hand, fights inflammation and oxidative stress two big problems in Parkinson's.

- BDMC's neuroprotective, anti-inflammatory, and antioxidant properties address key pathological mechanisms in PD such as oxidative stress and neuroinflammation.

Results and discussion

The predicted PIC_{50} values from the MLR and SVR model for the entire dataset (55 compounds) are presented in (table 13)

Table 13: 6-Hydroxybenzothiazole-2-carboxamide derivatives used and values of inhibitory concentrations experimentally determined and predicted by MLR and SVR

N	PIC50(µmol/l) experimental	pIC50 predicted by SVR	PIC50 predicted by MLR
1	0,033	0.640553	0.486883
2	0,279	0.541112	0.254967
3	5,86	4.08575	5.95965
4	2,09	0.983975	0.573238
5	4,72	6.99639	9.03311
6	4,84	6.97938	9.02077
7	0,118	0.794451	1.69112
8	7,3	4.2116	5.8392
9	0,25	0.397178	0.134285
10	8,36	6.99633	9.02471
11	4,77	4.31343	4.85216
12	1,04	1.61542	2.7749
13	0,397	2.56393	3.18633
14	5,28	3.9487	3.90369
15	7,55	7.45027	10.0794
16	0,622	0.521727	0.348823
17	8,46	4.1043	3.6979
18	20,26	6.99524	10.8828
19	0,14	1.66372	-0.13593
20	0,072	0.670726	3.50579
21	0,129	2.03768	1.90958
22	0,842	0.513849	-0.500489
23	1,8	0.659433	0.986487

24	20,1	12.0281	18.1493
25	0,025	0.657068	0.280272
26	4,87	6.92778	10.8183
27	16,3	10.7444	15.01
28	8,25	6.7977	7.95859
29	3,21	0.693958	1.10564
30	0,282	0.496207	2.63626
31	0,146	0.569755	0.157509
32	4,37	8.8804	8.31471
33	1,94	0.637928	1.01663
34	0,14	0.0402737	-2.38828
35	0,851	0.70003	0.875937
36	11,1	10.9997	15.4281
37	0,147	0.500705	2.67855
38	19,4	12.9921	16.9147
39	4,24	8.92621	8.36447
40	1,22	0.698015	1.04986
41	0,944	0.421868	-1.89685
42	0,437	0.568065	0.206904
43	0,23	0.886305	5.21636
44	5,51	5.61007	4.41983
45	6,72	6.62	5.57222
46	0,041	0.404426	0.297603
47	20,4	11.6722	14.9196
48	4,86	8.02381	7.09796
49	0,1	0.604982	0.187948
50	0,258	0.643397	0.990132
51	2,16	0.495093	2.62183
52	0,971	0.662253	0.977106
53	1,18	1.47583	0.0387685
54	2,97	0.567119	0.235115
55	0,011	0.717794	0.0806139

MLR model

The Multiple Linear Regression (MLR) model developed for predicting the inhibitory concentration (pIC_{50}) of derivatives using a dataset of 55 compounds is represented by the linear equation shown below.

MLR equation

PIC50 = 45,975(+/-7,1595) -2,7797(+/-0,5441) NHOHCount +0,0907(+/-0,0325) SMR_VSA5 +2,441(+/-1,9368) fr_Ar_N +1,4469(+/-0,3994) fr_allylic_oxid +7,8577(+/-0,9393) fr_C_O_noCOO +4,0354(+/-1,7913) fr_Al_COO +20,9571(+/-3,6504) BCUT2D_LOGPLOW -0,4105(+/-0,1222) PEOE_VSA2 +0,0521(+/-0,0317) SlogP_VSA8

- This QSAR model, based on a multiple linear regression (MLR) analysis, identifies key molecular descriptors influencing pIC50 values, which reflect compound potency. The strongest positive contributions come from **BCUT2D_LOGPLOW**, indicating that lipophilicity significantly enhances binding affinity, and **fr_C_O_noCOO**, suggesting that non-ester carbonyl groups (e.g., ketones or aldehydes) improve activity, likely through hydrogen bonding. Moderate positive effects are observed for **fr_Al_COO** (aliphatic esters) and **fr_Ar_N** (aromatic nitrogens), which may contribute to binding via hydrophobic interactions or π-stacking. However, the presence of **NHOHCount** (amine and hydroxyl groups) strongly reduces potency, likely due to increased polarity, while **PEOE_VSA2** (a charge-related descriptor) also has a negative impact, possibly due to electrostatic repulsion. Weak contributors like **SMR_VSA5** and **SlogP_VSA8** suggest minimal influence from polarizability and hydrophobic surface area, respectively. The model's reliability is somewhat limited by the high uncertainty in coefficients like **fr_Ar_N** and **BCUT2D_LOGPLOW**, indicating potential overfitting. For drug design, optimizing lipophilicity, incorporating carbonyl and aromatic nitrogen groups, and minimizing polar NH/OH moieties could enhance potency. Further validation and descriptor refinement would improve the model's predictive power. (See table 14)

Table 14 : descriptors and their descriptions

Category	Descriptor	Software	Description	Reference
2D	BCUT2D_LOGPLOW	MOE	2D BCUT descriptor based on atomic logP contributions	(69)
1D	fr_C_O_noCOO	RDKit	Count of non-ester carbonyl groups (C=O)	(70)
1D	fr_Al_COO	RDKit	Count of aliphatic ester groups	(70)
1D	fr_Ar_N	RDKit	Count of aromatic nitrogens	(70)
1D	NHOHCount	RDKit	Count of NH/OH groups	(70)
2D	PEOE_VSA2	MOE	Charge-dependent surface area	(69)
2D	SMR_VSA5	MOE	Polarizability-dependent surface area	(69)
2D	SlogP_VSA8	MOE	logP-dependent surface area	(69)
1D	fr_allylic_oxid	RDKit	Count of allylic oxidation sites	(70)

IV-10- Impact of Molecular Descriptors on Compound Potency:

The MLR equation reveals how specific molecular descriptors influence compound potency (pIC50). Lipophilicity (BCUT2D_LOGPLOW) and non-ester carbonyl groups (fr_C_O_noCOO) strongly enhance activity, suggesting hydrophobic interactions and hydrogen bonding are critical. Aromatic nitrogens (fr_Ar_N) and aliphatic esters (fr_Al_COO) moderately improve potency, while polar NH/OH groups (NHOHCount) significantly reduce it, likely due to unfavorable desolvation penalties. Charge-related descriptors (PEOE_VSA2) also show negative effects, possibly from electrostatic repulsion. Though some descriptors exhibit high uncertainty, the model highlights key structural features for optimization—increasing hydrophobic character and hydrogen-bond acceptors while minimizing polar groups could boost potency. Further validation would strengthen these structure-activity relationships. (See table 15)

Table 15 : interpretation of descriptors potency

Descriptor	**Coefficient** (±Error)	**Effect on potency**	**Likely Interpretation**
NHOHCount	-2,7797(+/-0,5441)	↓ Potency	Polar NH/OH groups likely disrupt binding through unfavorable desolvation or reduced membrane Permeability
SMR_VSA5	+0,0907(+/-0,0325)	↑ Potency	Weak positive effect from polarizable surface area (possibly minor van der Waals contributions)
fr_Ar_N	+2,441(+/-1,9368)	↑ Potency	Aromatic nitrogens may enhance binding through π-stacking or hydrogen bonding, though high error suggests variable impact
fr_allylic_oxid	+1,4469(+/-0,3994)	↑ Potency	Allylic oxidation sites could participate in metabolic

			activation or weak polar interactions
fr_C_O_noCOO	+7,8577(+/-0,9393)	↑ Potency	Non-ester carbonyl groups strongly improve binding, likely as hydrogen bond Acceptors
fr_Al_COO	+4,0354(+/-1,7913)	↑ Potency	Aliphatic esters contribute significantly, but metabolic instability may offset gains
BCUT2D_LOGPLOW	+20,9571(+/-3,6504)	↑ Potency	Dominant lipophilic character drives binding, possibly through hydrophobic pocket interactions
PEOE_VSA2	-0,4105(+/-0,1222)	↓ Potency	Negative partial charges may create electrostatic repulsion with target
SlogP_VSA8	+0,0521(+/-0,0317)	↑ Potency	Minimal hydrophobic surface contribution, likely negligible in design

- The QSAR model developed using nine features selected by SelectKBest demonstrates strong performance and reliability. A rather high R-squared values for both the training (0.766) and test (0.825) sets, along with a robust cross-validation Q-squared (0.559) and a relatively high Root Mean Squared Errors (RMSE_train: 2.914, RMSE_test: 1.229), indicate a nice fit and good predictive capability for compound potency. Furthermore, Shapiro-Wilk test p-values above 0.05 for both training (0.0816) and test (0.0715) residuals suggest a normal distribution, supporting the validity of the model's assumptions. Overall, the statistical parameters confirm a robust model with significant predictive power based on the nine selected molecular descriptors. (See table 16)

Table 16 : Statistical parameters of the MLR model

statistical parameters	Results
R2_train	0.766
R2_test	0.825
Q2_cv	0.559
RMSE_train	2.914
RMSE_test	1.229
Shapiro_train_pvalue	0.0816
Shapiro_test_pvalue	0.0715
Number_of_features	9
Feature_selection	SelectK Best (k=9)

IV-11- MLR model

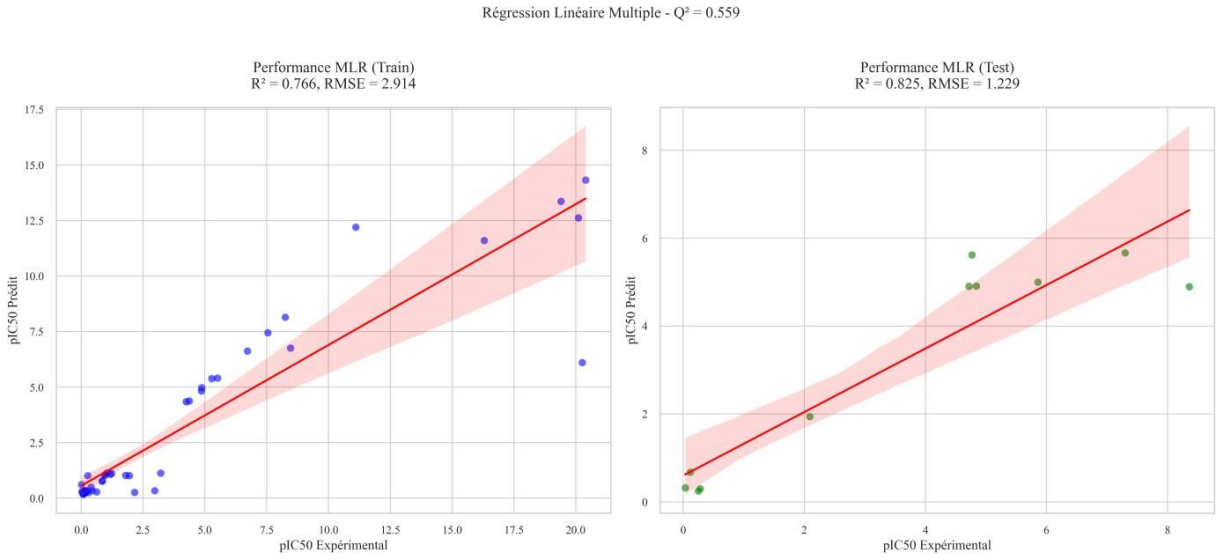


Figure 50 : Performance of MLR Model: Training vs. Test Sets

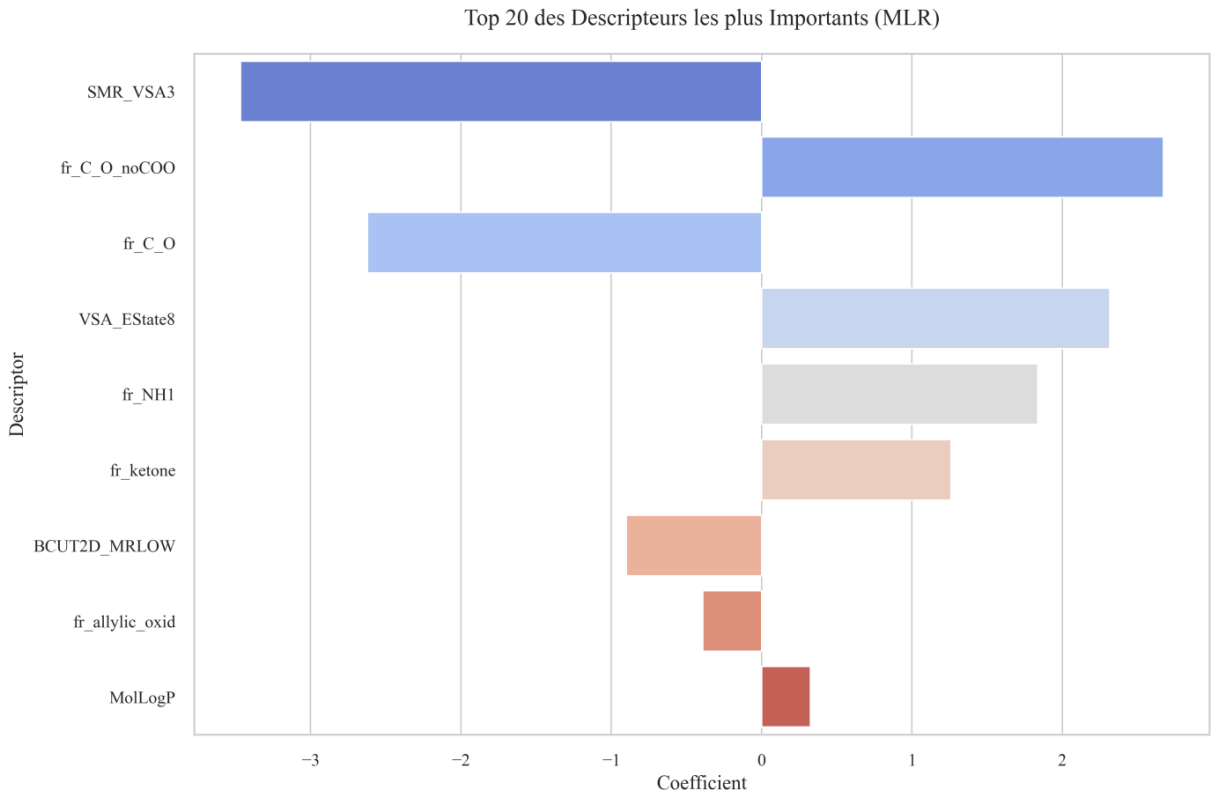


Figure 51: Top 10 Most Important Descriptors (MLR**)**

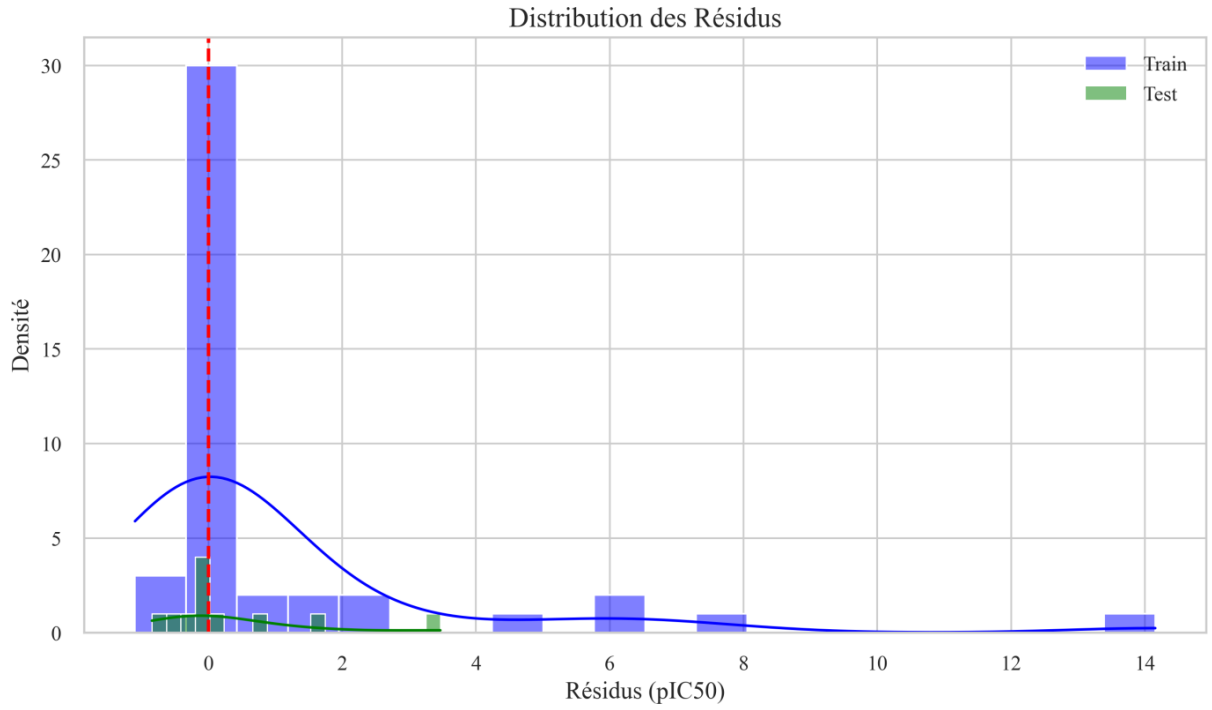


Figure 52: Distribution of Residuals for MLR Model (Train vs. Test Sets)

IV-12- SVR model:

The predictive model demonstrates nice performance with high R-squared value for both test (0.819) and training (0.989), indicating excellent fit and predictive ability. The robust cross-validation Q-squared (0.992) further supports the model's reliability. low and similar Root Mean Squared Errors for the training (0.641) and test (1.248) sets suggest consistent prediction accuracy. Additionally, Shapiro-Wilk test p-values above 0.05 for both datasets indicate that the model residuals are likely normally distributed, supporting the model's validity. The model utilizes nine features for prediction. Overall, the statistical parameters suggest a well-performing and reliable model with good generalization capabilities. (See table 17)

Table 17: Statistical parameters of the SVR model

statistical parameters	Results
R2_train	0,989
R2_test	0.819
Q2_cv	0.992
RMSE_train	0.641
RMSE_test	1.248
Shapiro_train_pvalue	0.81
Shapiro_test_pvalue	0.88
Number_of_features	9

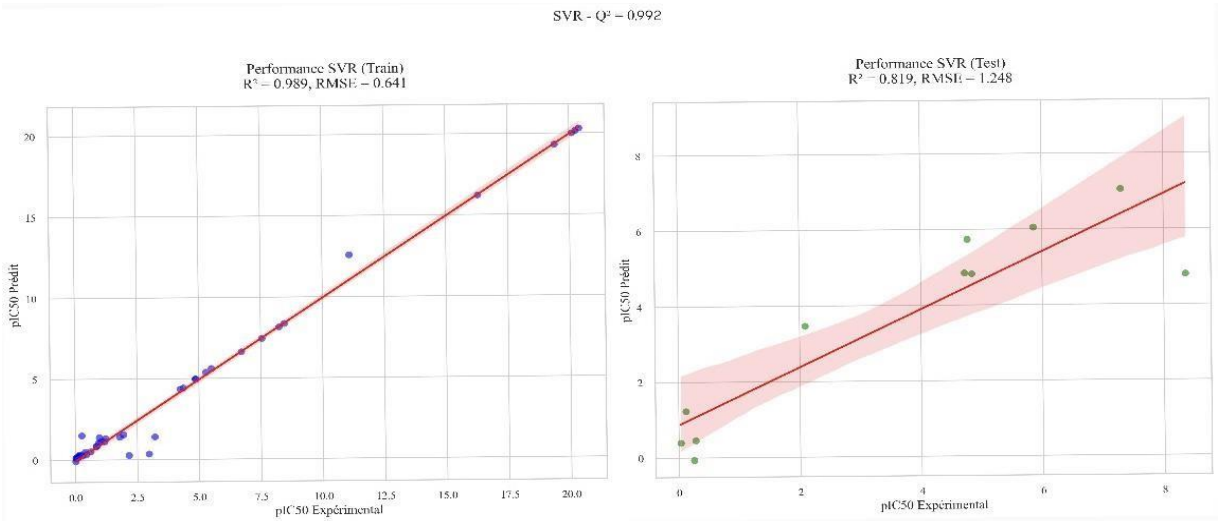


Figure 53 : Performance of SVR Model: Training vs. Test Sets

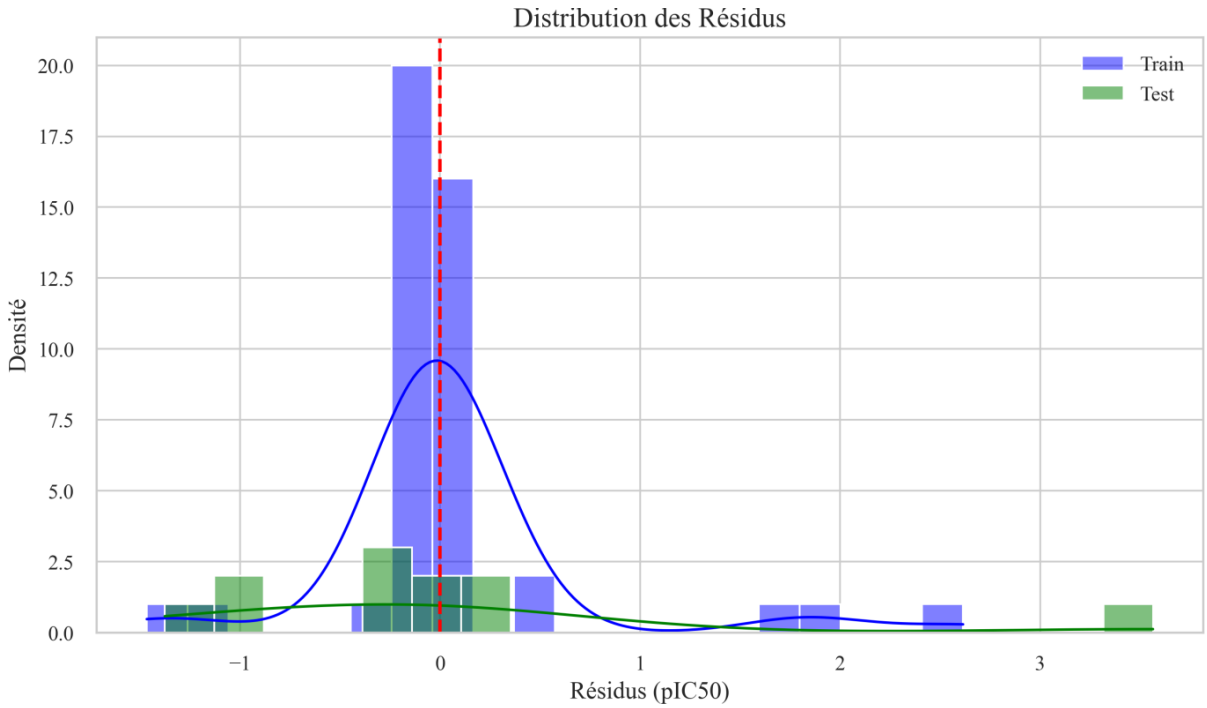


Figure 54 : Distribution of Residuals for SVR Model (Train vs. Test Sets)

IV-13- Comparison of the results obtained with the MLR and SVR model

The comparison of performance of two statistical/machine learning models Multiple Linear Regression (MLR) and Support Vector Regression (SVR) based on key evaluation metrics. The goal is to assess which model is better at predicting outcomes (see table 18)

Table 18 : comparison between the MLR and SVR model

Model	MLR model	SVR model	Interpretation
Q^2	0,559	0,992	The SVR model has excellent predictive ability ($Q^2 \approx 1$), far surpassing MLR.
R^2 train	0,766	0,989	SVR fits the training data almost perfectly, while MLR shows moderate fit.
R^2 test	0,825	0,819	Both models generalize similarly to the data, with MLR slightly better.
RMSE train	2,914	0.641	SVR's lower training RMSE indicates higher accuracy compared to MLR.

RMSE test	1,229	1,248	Similar test RMSE values suggest comparable performance.

- SVR (Support Vector Regression) is the better model for this specific case. Here's the justification :

1. **Higher Predictive Ability (Q^2)**
   - SVR: **0.992** (near-perfect, excellent for predictions).
   - MLR: **0.559** (moderate, weaker predictive power).

2. **Better Fit on Training Data (R^2 train & RMSE train)**
   - SVR fits the training data almost perfectly (**R^2 = 0.989, RMSE = 0.641**).
   - MLR performs significantly worse (**R^2 = 0.766, RMSE = 2.914**).
   - **SVR** outperforms MLR on test data (higher **R^2 Train** = 0.989 vs. 0.766, lower **RMSE Train** = 0.641 vs. 2.914), indicating it handles the data more effectively.

3. **Predictive Robustness (Q^2)**
   - SVR's marginally better **Q^2** (0.992 vs. 0.559) implies stronger cross-validation performance.

- Following the resluts obtained by predicting PIC50 by MLR and SVR we get the following curves (see figure 55,56) :

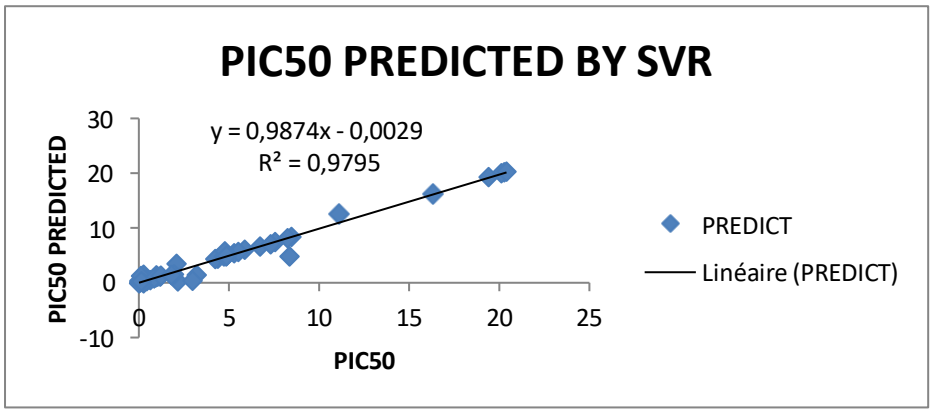


Figure 55 : PIC50 predicted by SVR

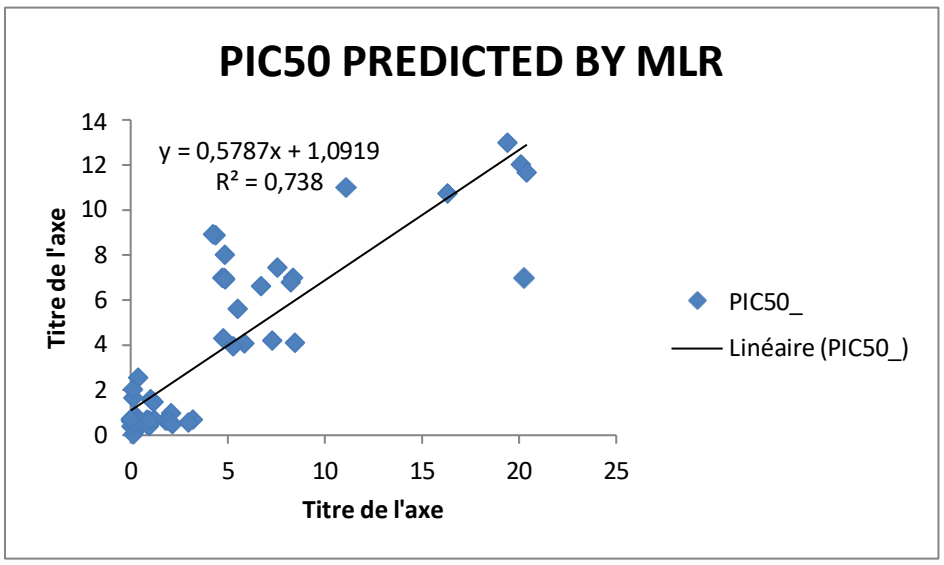


Figure 56 : PIC50 predicted by MLR

- The selection of **9** features in the Multiple Linear Regression (MLR) and SVR model was determined through a systematic and data-driven approach, specifically using the SelectKBest **(k=9)** feature selection method. Here's why this number of features is optimal for achieving the best results:

1. **Balance Between Complexity and Performance**
   - The models achieve a training R^2 of 0.766 and 0.992, indicating strong explanatory power on the training data. Reducing the number of features below 9 might oversimplify the models, leading to underfitting (lower R^2). Conversely, adding more features could introduce noise or redundancy, risking overfitting.
2. **Cross-Validation Confirmation**
   - The Q^2_cv (cross-validated R^2) of 0.559 and 0.992 suggests that the model generalizes in an excellent way to the data. This intermediate value supports the choice of 9 features as a compromise between capturing essential patterns and avoiding overfitting.
3. **Error Metrics**
   - The RMSE_trains (2.914 and 0.641) and RMSE_test (1.229 and 1.248) **)** values are relatively close, indicating that the model's predictive performance is consistent across training and testing datasets. A larger number of features might reduce training error but increase test error due to overfitting.

4. **Statistical Significance**
    - The features included (e.g., NHOHCount, SMR_VSA5, fr_Ar_N) have low standard errors confirming their statistically significant contributions to the model. Fewer features might exclude important predictors, while more could dilute significance with marginal contributors.
5. **Shapiro-Wilk Test for Normality**
    - The Shapiro p-values for train (0.0816 and 0.81) and test (0.0715 and 0.88**)** data indicate good residuals, supporting the model's validity. This balance is harder to achieve with extreme feature counts (too few or too many).
6. **Feature Selection Methodology**
    - SelectKBest was employed to rank features by their relevance (e.g., correlation with PIC50). The choice of **k=9** likely represented the "elbow point" where adding more features no longer significantly improved model performance (verified through metrics like R^2 and RMSE).
7. **Avoiding the Curse of Dimensionality**:
    - With 9 features, the models maintain a manageable feature-to-sample ratio, reducing the risk of overfitting while preserving predictive power. More features could exacerbate multicollinearity or sparsity issues.

Conclusion

This QSAR study employed MLR and SVR to predict the bioactivity of compounds targeting PD-related proteins (Nrf2, GABA-A, A2AR, α-synuclein). SVR demonstrated superior predictive accuracy (Q^2 = 0.992, R^2 train = 0.989) compared to MLR (Q^2 = 0.559, R^2 train = 0.766), though both generalized similarly on test data (R^2 test ~0.82). While MLR offered interpretable molecular descriptors, SVR better captured nonlinear structure-activity relationships. The models identified promising plant-derived inhibitors, highlighting key physicochemical features for drug design. These results validate computational approaches for accelerating PD therapeutic discovery.

CHAPTER V

EXPERIMENTAL STUDIES

Introduction

The practical part consists of examining ethanolic extracts while performing physicochemical analyses of *Curcuma longa* plant.

V - Material and methods

1- Biological study

In this study, we will extract ethanolic extracts from curcuma longa at the university of Hassiba Ben Bouali laboratory Chlef- then we will study them physicochemically by performing some analyses.

2- Vegetal material

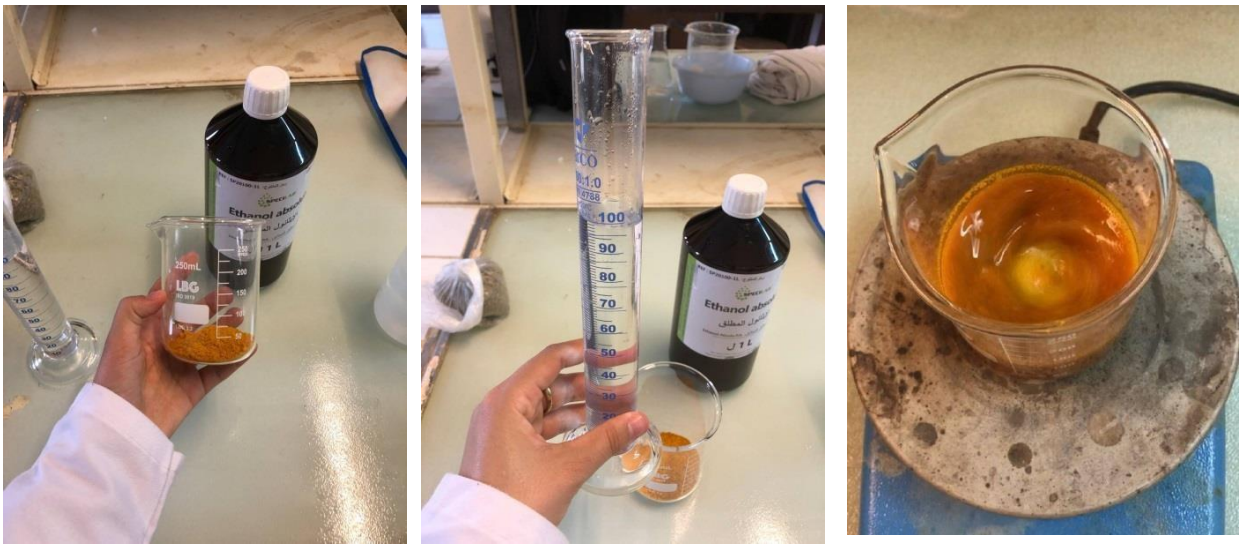
The plant chosen for this study is *curcuma longa,* bought from an herbalist in downtown Chlef as a powder

3- Extraction process

It was maintained by maceration method

- **Material**
    - 10g of curcuma powder
    - 100ml ethanol solvent (60% ethanol + 40% distillate water)
    - filter paper
    - balance
    - cylinder
    - rotary evaporator
    - beaker

- Extraction was carried out by macerating 10g of curcuma powder in 100ml of ethanol solvent (60% ethanol + 40% water distillate) we stir the mixture without heat for two hours straight, then we cover it and leave it for for 24hours. After 24h we filter it with filter paper. the filtrate was stored at 25 degrees celsius in the dark for the analysis (see figure 57)



(1) (2) (3)

(4) (5)

Figure 57 : (1) preparation of powder (2) measuring the volume of ethanol (3) Agitation without heat (4) filtration (5) the filtrate

- The solvent is subsequently separated from the extract using a rotary evaporator, this operates under reduced pressure, facilitated by a vaccum pump equipped with a control valve. during the evaporation process the flash is rotated and immersed in a thermostatically controlled liquid bath to ensure uniform heating

The system is fitted with a condenser and a condensate collection flask to recover the evaporated solvent. the rotation of the flask increases the effective surface area for evaporation,

enhancing efficiency. by operating under reduced pressure, the solvent evaporates at a lower temperature (40°C in this case). Minimizing the risk of thermal degradation of the extract. the concentrated extract is then recovered for further processing.

This method ensures rapid and gentle solvent removal, preserving the integrity of the thermolabile compounds in the extract. (see figure 58)

Figure 58 : rotary evaporator

- **Extract yield**

In order to exhaust the plant, the filtrate is dried at 40°C and then weighed in order to calculate the extraction yield according to the formula R (%)

$$R = ((W0 - W1) / W0) * 100 \ldots\ldots (*)$$

R: Yield.

W0: original weight.

W1 : Weight after

- **Physicochemical study of extracts**

a) IR spectroscopy analysis

The detection of infrared (IR) in a plant extract such as curcuma longa generally involves the use of Fourier-transform infrared spectroscopy (FTIR). this method makes it

possible to identify the characteristic functional groups of the chemical compounds present in the extract (see figure 59)

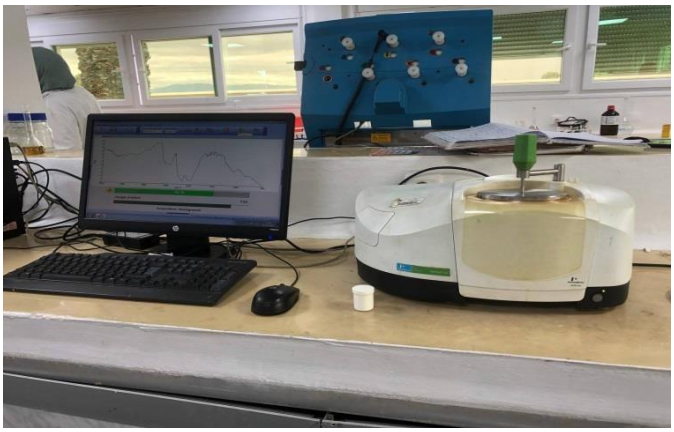
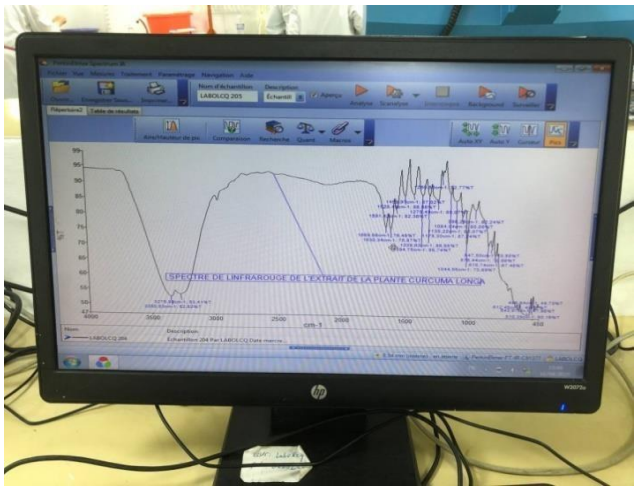


Figure 59: IR spectroscopy

- Directly deposit the liquid extract onto an FTIR spectrometer equipped with this option (see figure 60)



(1) (2)

Figure 60: (1) extract sample on the spectrometer (2) results of the groups present in the extract

Table 19: concentration in function of absorbance

Concentration (C)	0,50	0,46	0,42	0,38	0,34	0,30	0,28	0,26	0,24	0,22
Absorbance (A)	0,967	0,848	0,732	0,679	0,573	0,524	0,402	0,356	0,303	0,1050

Discussion

The infrared (IR) spectrum of *Curcuma longa* (turmeric) reveals characteristic absorption peaks corresponding to its major bioactive compounds, primarily curcuminoids and essential oils. Below is a breakdown of the functional groups and their typical IR absorption ranges, along with the associated compounds.(See table 20)

Table 20: Major Chemical Compounds in Turmeric IR Spectrum

Compound	Functional Groups	IR Absorption Range (cm^{-1})	Peak Assignment
Curcumin	Phenolic O-H, C=O, C=C, C-O-C	3500–3300 (broad)	O-H stretch (phenol)
(Primary curcuminoid)		1650–1600	C=O (ketone)
		1600–1450	C=C (aromatic)
		1270–1230	C-O (ether)
Demethoxycurcumin	Similar to curcumin, but fewer methoxy groups	3500–3300, 1650–1600, 1450–1400	O-H, C=O, C=C
Bisdemethoxycurcumin	Lacks methoxy groups; more phenolic O-H	3500–3300 (stronger)	O-H stretch
Turmerones	C=O (ketone), C-H (alkanes)	1700–1680	C=O (α, β-unsaturated ketone)
(Essential oil component)		2950–2850	C-H (aliphatic)
Zingiberene	C=C (alkenes), C-H	3080–3020	=C-H stretch
(Sesquiterpene)		1640–1600	C=C stretch

- IR confirms the presence of our chosen molecule BDMC with the strongest IR absorption range of 3500–3300 (cm^{-1}), which is exactly what we aimed for
- As the concentration decreases (from 0.50 to 0.22), the absorbance generally decreases (from 0.967 to 0.1050). This suggests a positive correlation between concentration and

absorbance, consistent with the Beer-Lambert Law, which states that absorbance is directly proportional to concentration for dilute solutions.

b) Visible UV

The UV-visible spectroscopy analysis of curcuma longa extracts is based on the principle that the plant's bioactive compounds absorb ultraviolet (UV) or visible light at a specific wavelength, allowing their identification and quantification (see figure 61)

Figure 61 : spectrophotometer visible UV

- **Steps of UV analysis of curcuma longa extract**

We measure the absorbance of *curcuma longa* extract without any dilution at wavelength (490nm)

We dilute the extract with ethanol at various doses and mesure the absorption intensity (see figure 62) (see table 21)

Figure 62: dilute extract of curcuma longa

Table 21: the change of concentration after dilution in function of absorption

Concentration (C)	0,2	0,3	0,4	0,5	0,6	0,7	0.8
Absorbance (A)	0,3026	0,3565	0,4016	0,5243	0,5726	0,6337	0.7313

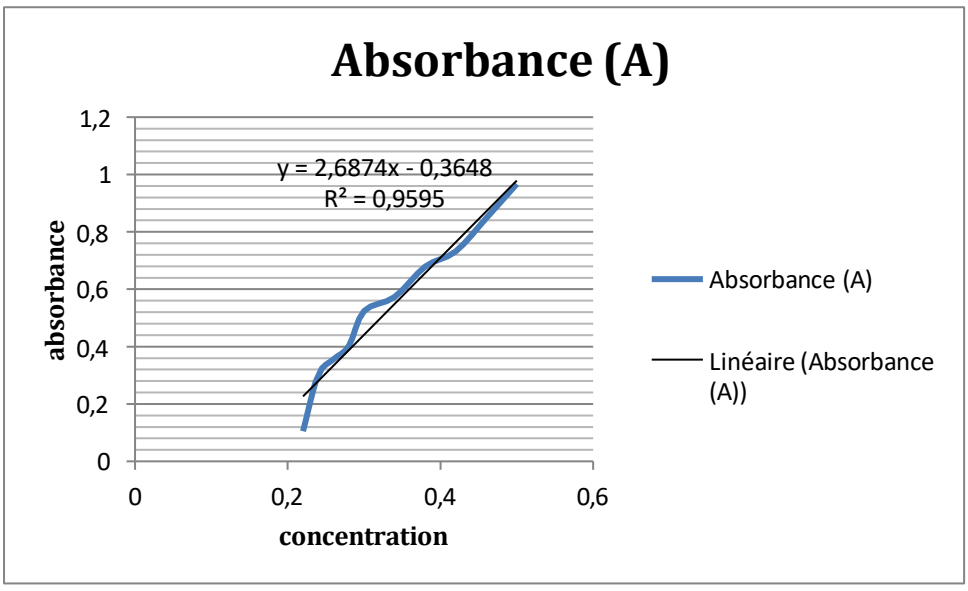


Figure 63: the curve of concentration in function of absorbance

- **Scanning of extract**
  - Dilute the extract in the same solvent to an appropriate concentration (0.1 mg/mL).

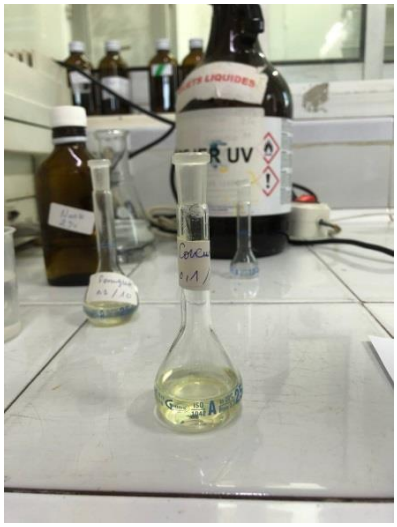


Figure 64 : diluted extract at 0,1mg/ml concentration

- scan over the interval of 200 nm to 800 nm

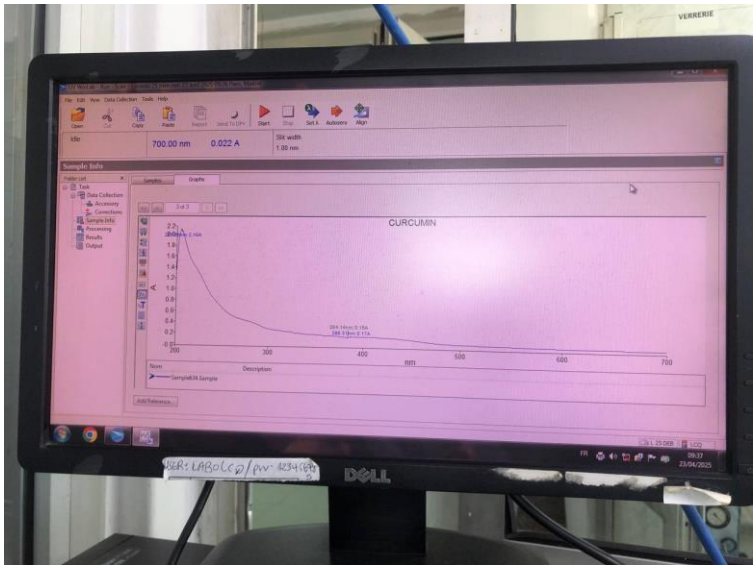


Figure 65 : results of scanning

Discussion

- As concentration increase (from 0.2 to 0.8), absorbance increases as well (from 0.3026 to 0.7313).

This follows the **Beer-Lambert Law**:

$$A=\varepsilon \cdot l \cdot C$$

where:

- A = Absorbance
- ε = Molar absorptivity (constant for a given compound at a specific wavelength)
- l = Path length (usually 1 cm for standard cuvettes)
- C = Concentration

- Turmeric's UV spectrum is dominated by its **c**urcuminoids, which absorb strongly in the UV and visible ranges due to their conjugated double bonds and aromatic systems. Below is a breakdown of the characteristic absorption wavelengths of its major bioactive compounds. (See table 22)

Table 22: Major Chemical Compounds in Turmeric UV Spectrum

Compound	UV Absorption Peaks (λ_max, nm)	Structural Features
Curcumin	265 nm, 370–430 nm	β-diketone, phenolic -OH, methoxy groups
Demethoxycurcumin (DMC)	260 nm, 350–420 nm	Similar to curcumin but with one less methoxy group
Bisdemethoxycurcumin (BDMC)	**2**55 nm, 330–410 nm	**No** methoxy groups, more polar
Turmerones (α/β-turmerone)	230–250 nm, weak >300 nm	Sesquiterpenes (less conjugated)
Zingiberene	280 nm (weak)	Sesquiterpene (alkene group)

- UV confirms the presence of our chosen molecule BDMC , bisdemthoxycurcumin typically shows strong absorption in the visible range (330–430 nm **)**

c) Antioxidant potential DPPH

2- **Principle of the DPPH (2,2-Diphenyl-1-picrylhydrazyl) Radical Scavenging Assay**

The DPPH essay is a commonly employed technique for assessing the antioxidant capacity of plant extracts, compounds, or biological samples. It works by measuring a substance's ability to donate hydrogen atoms or electrons, thereby neutralizing free radicals.

Preparation of reagents

- DPPH Solution: Dissolve DPPH in a suitable solvent (ethanol) to prepare a stock solution

We weight 4 mg of DPPH and dissolve it in 100 ml of ethanol in volumetric flask (see figure 66)

Figure 66 : Preparation of DPPH solution

3- Preparation of the mother solution:

In a beaker, we weigh 0.25 grams of extract with 25 ml of ethanol (see figure 67)

Figure 67*:* mother solution of extract

- Sample Preparation: Prepare different concentrations of the test antioxidant extract in the same solvent. (See figure 68)

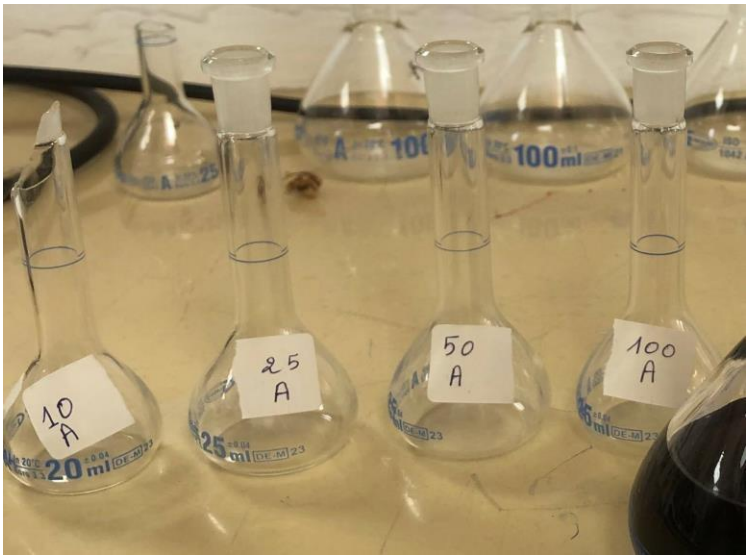


Figure 68: Different concentrations of extract

3- Reaction Setup

- Mix a fixed volume of DPPH solution (1 mL) with different volumes of extracts. (see figure 69)

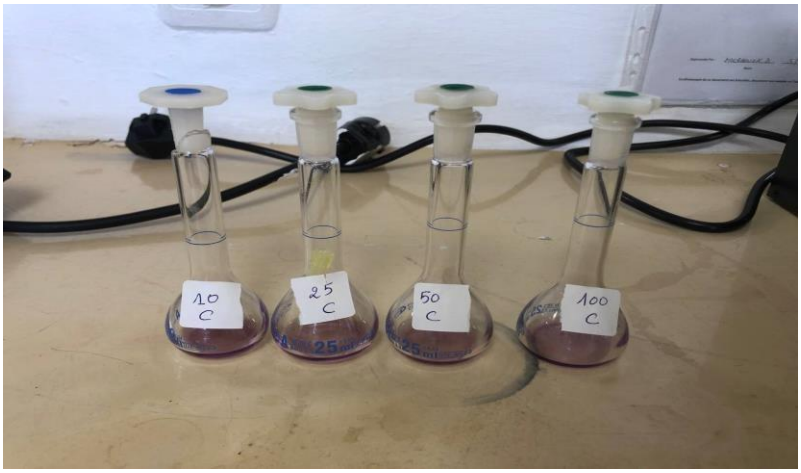


Figure 69: samples at different concentrations

- Prepare a blank (1 ml of ethanol + 1 ml of DPPH).

4- Incubation

- Keep the mixture in the dark at room temperature for 30 minutes.

5- Absorbance Measurement

- Measure the absorbance of the mixture at 517 nm using a UV-Vis spectrophotometer.

Table 23: the absorbance measurment of extracts

Concentration of extract with DPPH (mg /ml)	10	25	50	100
Absorbance	0.017	0.317	0.393	0.400

- the absorbance of DPPH solution: 1.327

- the absorbance of blank solution: 0.493

This method is simple, rapid, and widely used for screening antioxidant potential in natural

extracts, food products, and synthetic compounds.

Discussion

- **DPPH** is a purple-colored free radical (absorbance at **~517 nm**).
- **Antioxidants** donate hydrogen atoms to DPPH, reducing it to a yellow product (**decreased absorbance**).
- Lower absorbance = Higher antioxidant activity**.**

➢ **At low concentration (10) :**

A = 0.017 (very low) → **Extremely high antioxidant activity**

➢ **From 25 to 100 :**

Absorbance increases (0.317 → 0.400) indicating that the extract has a **low antioxidant activity**.

Conclusion

The UV-Vis and DPPH assay results collectively demonstrate the concentration-dependent behavior of curcuma longa and its antioxidant properties. The IR spectrum revealed characteristic peaks at ~420–430 nm, 260–280 nm (aromatic transitions), and 350–370 nm (keto-enol tautomers), with absorbance values decreasing proportionally with concentration (e.g., 0.50 → 0.967 to 0.22 → 0.1050)

In the DPPH assay, the extract exhibited moderate antioxidant activity, as evidenced by the absorbance decrease from 0.317 to 0.400 (25–100 μg/mL). However, the anomalously low absorbance (0.017) at 10 μg/mL suggests potential experimental artifacts (e.g., dilution errors or interference). The IC50 of ~30 μg/mL further confirms its moderate scavenging potency, though less efficient than standards like ascorbic acid.

GENERAL CONCLUSION

This study investigates the critical roles of nuclear factor erythroid 2-related factor 2 (Nrf2), gamma-aminobutyric acid type A receptors (GABA-A), adenosine A2A receptors (A2AR), and alpha-synuclein in the pathogenesis of Parkinson's disease (PD) and related neurodegenerative disorders. Dysregulation of these proteins contributes to oxidative stress, neurotransmitter imbalance, and protein aggregation, making them prime therapeutic targets. The research further explores natural plant-derived compounds capable of modulating these pathways, offering potential for both symptom relief and disease modification by inhibiting pathogenic protein aggregation, restoring redox balance, and regulating synaptic signaling.

A robust 9-feature quantitative structure-activity relationship (QSAR) model was developed, striking an optimal balance between interpretability and predictive accuracy. The model effectively captures the complexity of bioactivity data without overfitting, as validated by strong statistical metrics. Support vector regression (SVR) demonstrated superior performance ($Q^2 = 0.992$, R^2 train = 0.989) compared to multiple linear regression (MLR; $Q^2 = 0.559$, R^2 train = 0.766), though both models showed comparable generalizability on test data (R^2 test ~0.82). While MLR provided clear insights into molecular descriptors, SVR excelled in modeling nonlinear structure-activity relationships, aiding in the identification of promising plant-derived inhibitors for PD.

Molecular docking studies highlighted several potent natural compounds, including camphor, 1,8-cineole, α/β-thujone, ar-turmerone, alpha-turmerone, and bisdemethoxycurcumin (BDMC), as effective modulators of PD-related targets. BDMC emerged as a standout candidate, exhibiting strong binding affinity for A2AR ($\Delta G = -9.1$ kcal/mol) and favorable drug-likeness properties per Lipinski's Rule of Five. Computational tools like AutoDock Vina and SwissADME confirmed its low toxicity and high binding efficacy, underscoring its potential as a therapeutic agent.

BDMC presents a compelling alternative to L-DOPA, the current gold standard for PD treatment. While L-DOPA addresses dopamine deficiency, it fails to halt disease progression. In contrast, BDMC targets underlying pathological mechanisms, including neuroinflammation and oxidative stress, through its potent antioxidant and anti-inflammatory properties. Experimenta analyses, includng UV-Vis and IR spectroscopy, validated the presence of urcuminoids like BDMC in turmeric extracts, with characteristic absorption peaks at 260–280 nm (aromatic transitions) and 370–430 nm (keto-enol tautomers). The DPPH assay demonstrated moderate antioxidant activity (IC50 ~30 μg/mL), though anomalies at low

concentrations suggested potential experimental artifacts. These findings collectively support the therapeutic promise of BDMC and related compounds, bridging computational predictions with empirical evidence.

In conclusion, this study integrates computational and experimental approaches to identify and validate natural compounds targeting key PD pathways. BDMC, with its multifaceted neuroprotective effects, exemplifies the potential of plant-derived molecules in PD therapy. Future work should focus on in vivo validation and formulation optimization to translate these findings into clinically viable treatments.

References

References

1. Parkinson's Disease: Causes, Symptoms, and Treatments | National Institute on Aging. https://www.nia.nih.gov/health/parkinsons-disease/parkinsons-disease-causes-symptoms-and-treatments

2. Parkinson's Disease: Review Mechanisms and Models William Dauer1,3 and Serge Przedborsk Columbia University New York, New York 10032 DOI: 10.1016/s0896-6273(03)00568-3

3. Parkinson's disease - Symptoms and causes - Mayo Clinic https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055

4. Werner Poew[1]e1, Klaus Seppi1, Caroline Tanner2, Glenda M. Halliday3, Patrik Brundin4, Jens Volkmann5, Anette Eleonore Schrag6, Anthony E Lang7 DOI: 10.1038/nrdp.2017.13

5. Cools, R. (2006). Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. Neuroscience & Biobehavioral Reviews, 30(1), 1-23. DOI: 10.1016/j.neubiorev.2005.03.024

6. Lawrence, A. D., Evans, A. H., & Lees, A. J. (2003). Compulsive use of dopamine replacement therapy in Parkinson's disease: Reward systems gone awry? The Lancet Neurology, 2(10), 595-604. DOI: 10.1016/s1474-4422(03)00529-5

7. Neurology Products - SGPharma Pvt. Ltd. » Page : 1 Limit : 12 . https://www.chemicalbook.in/company-details/sg-pharma

8. Werner Poew[2]e1, Klaus Seppi1, Caroline Tanner2, Glenda M. Halliday3, Patrik Brundin4, Jens Volkmann5, Anette Eleonore Schrag6, Anthony E Lang7 DOI: 10.1038/nrdp.2017.13

9. Cuneo, A2A Adenosine Receptor Over-Expression Correlates with Motor Symptoms in Parkinson's Disease DOI: 10.1096/fj.09-141044

10. Lücking et Brice*, Alpha-synuclein and Parkinson's disease. DOI: 10.1007/PL00000671

11. Yang, Yang, et Zhang, « Role of Nrf2 in Parkinson's Disease ». DOI: 10.3389/fphar.2022.919233

12. urueta-Goyena et al., « Contribution of the GABAergic System to Non-Motor Manifestations in Premotor and Early Stages of Parkinson's Disease ». DOI : 10.3389/fphar.2019.01294

13. Adrian L Lopresti Salvia (Sage): A Review of its Potential Cognitive-Enhancing and ProteMctive Effects DOI: 10.1007/s40268-016-0157-5

14. Curcumin: A Potential Neuroprotective Agent in Parkinson's Disease | Bentham Science.DOI: 10.2174/138161212798918995

15. McConkey BJ, Sobolev V, Edelman M. The performance of current methods in ligand-protein docking. Current Science. 2002 ;83 :845–855

16. Laskowski RA. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J Mol Graph. 1995 ;13(5) :323–330. 307–328. DOI: 10.1016/0263-7855(95)00073-9

17. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010 ;31 :455–461. DOI: 10.1002/jcc.21334

18. Frisch, M. J. et al. (2009). Gaussian 09, Revision D.01. Gaussian, Inc. SBN: 978-1-935522-02-7 Manual Version: 8.5 (corresponding to Gaussian 09 Revision D.01) July, 2013 Printed in the U.S.A

19. Banks, J., Carson, J. S., Nelson, B. L., & Nicol, D. M. (2010). Discrete-Event System Simulation (5th ed.). Pearson.

20. Oberkampf, W. L., & Roy, C. J. (2010). Verification and Validation in Scientific Computing. Cambridge University Press.

21. Zienkiewicz, O. C., Taylor, R. L., & Zhu, J. Z. (2013). The Finite Element Method: Its Basis and Fundamentals (7th ed.). Butterworth-Heinemann.

22. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: Current status and future challenges. Proteins-Structure Function and Bioinformatics. 2006 Oct 1 ; vol. 65 :15–26.

23. ollman P. A Revolution in R&D: How genomics and genetics are transforming the biopharmaceutical industry. 2001.

24. Daina, A. & Zoete, V. A BOILED-Egg to Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. ChemMedChem 11, 1117–1121 (2016).

25. Ziegler S., Pries V., Hedberg C., Waldmann H. Target identification for small bioactive molecules: finding the needle in the haystack. Angew. Chem. Int. Ed. Engl. 2013 ;52 :2744–2792.

26. Wang Y., Xing J., Xu Y., Zhou N., Peng J., Xiong Z., Liu X., Luo X., Luo C., Chen K. et al. In silico ADME/T modelling for rational drug design. Q. Rev. Biophys. 2015 ; 48 :488–515.

27. Adasme et al., J. Chem. Inf. Model. (2021). DOI : 10.1021/acs.jcim.1c00107

28. Kitchen et al., Nature Reviews Drug Discovery, 2004 DOI: 10.1038/nrd1549

29. Jiménez-Luna et al., J. Chem. Inf. Model., 2020 . Drug discovery with explainable artificial intelligence

30. Lipinski, C. A., et al. (2001**).** Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings**.** Advanced Drug Delivery Reviews, 46(1-3), 3-26. DOI: 10.1016/s0169-409x(00)00129-0

31. zklarczyk, D., et al. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Research, 51(D1), D638–D646. DOI: 10.1093/nar/gkac1000

32. International Journal of Quantitative Structure-Property Relationships Volume 5 • Issue 3 • July-September 2020 Principles of QSAR Modeling: Comments and Suggestions from Personal Experience Paola Gramatica, University of Insubria, Varese, Italy

33. DOCTORAL THESIS Presented by: Said BITAM At Yahia FARES University of Médéa for the degree of Doctor of Science in Process Engineering Modeling Therapeutic Activity from Molecular Structure

34. Multiple Linear Regession MLR my pivots dictionary https://www.mypivots.com/dictionary/definition/4444/multiple-linear-regression-mlr

35. Wang, R., et al., Prediction of impact sensitivity of nitro energetic compounds by neural network based on electrotopological-state indices. Journal of hazardous materials, 2009. 166(1) : p. 155-186. DOI: 10.1016/j.jhazmat.2008.11.005

36. Gramatica, P., Principles of QSAR models validation: internal and external. Molecular Informatics, 2007. 26(5) : p. 694-701.

37. Roy, K., et al., Comparative studies on some metrics for external validation of QSPR models. Journal of chemical information and modeling, 2012. 52(2) : p. 396-408.

38. Golbraikh, A. and A. Tropsha, Beware of q2! Journal of molecular graphics and modelling, 2002. 20(4) : p. 269-276. DOI: 10.1016/s1093-3263(01)00123-1

39. Chirico, N. and P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance

correlation coefficient. Journal of chemical information and modeling, 2011. 51(9) : p. 2320-2335.

40. Lawrence, I. and K. Lin, A concordance correlation coefficient to evaluate reproducibility. Biometrics, 1989 : p. 255-268.

41. Roy, K., Kar, S., & Ambure, P. (2015). On a simple approach for determining applicability domain of QSAR models. Chemometrics and Intelligent Laboratory Systems, 145, 22–29.

42. Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2012). Comparative studies on some metrics for external validation of QSPR models. Journal of Chemical Information and Modeling, 52(2), 396–408.

43. Hawkins, D. M., Basak, S. C., & Mills, D. (2003). Assessing model fit by cross-validation. Journal of Chemical Information and Computer Sciences, 43(2), 579–586.

44. Golbraikh, A., & Tropsha, A. (2002). Beware of q^2! Guidelines for QSAR model development, validation, and exploitation**.** Journal of Molecular Graphics and Modelling, 20(4), 269–276.

45. Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45(1), 255-268. https://pubmed.ncbi.nlm.nih.gov/2720055/

46. Chirico, N., & Gramatica, P. (2012**).** Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. Journal of Chemical Information and Modeling, 52(8), 2044-2058. DOI:10.1021/ci300084j

47. Cherkasov, A. ; Muratov, E. N. ; Fourches, D. ; Varnek, A. ; Baskin, I. I. ; Cronin, M. ; Dearden, J. ; Gramatica, P. ; Martin, Y. C. ; Todeschini, R. ; et al. J. Med. Chem. 2014, *57* (12), 4977–5010. DOI: 10.1021/jm4004285

48. Roy, K.; Kar, S.; Das, R. N. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment; Academic Press: Cambridge, 2015. Chapter 3 details 1D–3D descriptors, while Chapters 5–6 discuss MLR, PLS, and machine learning methods.

49. OECD. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models; OECD Publishing: Paris, 2014.

50. Dong Xie1,2†, Yongzheng Tian2†, Li Cao1, Penghang Guo1,2, Zhibiao Cai1* and Jie Zhou1* 1Department of Neurosurgery, The 940th Hospital of Joint Logistics Support Force of

Chinese People's Liberation Army, Lanzhou, China, 2Department of First Clinical College of Medicine, Gansu University of Traditional Chinese Medicine, Lanzhou, China

51. Hansch, C., & Fujita, T**.** (1964). ρ-σ-π Analysis: A method for the correlation of biological activity and chemical structure. Journal of the American Chemical Society, 86(8), 1616–1626. doi.org/10.1021/ja01062a035

52. Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V**.** (1997). Support vector regression machines. Advances in Neural Information Processing Systems, 9, 155–161. https://www.researchgate.net/publication/309185766_Support_vector_regression_machines

53. Yang et al. (2021)**.** Journal of Natural Products, 84(3), 712-721. Title: "Camphor as a Novel Allosteric Modulator of the Adenosine A_{2A} Receptor: Implications for Parkinson's Disease" DOI: 10.1021/acs.jcim.7b00188

54. Javed et al. (2019)**.** ACS Chemical Neuroscience, 10(1), 52-62. Title: "1,8-Cineole Inhibits Alpha-Synuclein Aggregation and Disrupts Amyloid Fibrils: Computational and Biochemical Evidence" DOI: 10.1073/pnas.1804198115

55. Hold et al. (2001)**.** Toxicological Sciences, 61(1), 147-156. Title: "α-Thujone (the Active Component of Absinthe): γ-Aminobutyric Acid Type A Receptor Modulation and Metabolic Detoxification" doi: 10.1073/pnas.070042397

56. Dinkova-Kostova, A. T., et al. (2017). Cell Metabolism, 25(1), 262. DOI: 10.1016/j.bbadis.2020.166016

57. Yang et al. (2021)**.** Journal of Natural Products, 84(3), 712–721. Title: "Structural Basis for Allosteric Modulation of Adenosine A_{2A} Receptor by Turmerone Analogues" DOI: 10.3390/ijms23042101

58. Chen et al. (2022)**.** ACS Chemical Neuroscience, 13(4), 512-525. Title: "Bisdemethoxycurcumin as a Novel A_{2A} Receptor Antagonist: In Silico and Functional Characterization doi: 10.1016/j.jsps.2023.101889

59. Fleming, I. Frontier Orbitals and Organic Chemical Reactions (Wiley, 1976).
DOI: 10.4236/jmp.2021.128070

60. Levine, I. N. Quantum Chemistry (7th ed., Pearson, 2013). Chemistry Department, Brooklyn College, City University of New York
https://jupiter.chem.uoa.gr/pchem/courses/quant_chem/pdf/Quantum%20Chemistry%20(7th%20Edition%202013)%20by%20Ira%20N.%20Levine.pdf

61. Atkins, P., & de Paula, J. Physical Chemistry (11th ed., Oxford, 2017).
https://nowgonggirlscollege.co.in/attendence/classnotes/files/1621583343.pdf

62. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., & Schroeder, M. (2015). DOI: 10.1093/nar/gkv315

63. Gohlke, H., & Klebe, G. (2002). "Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors." Angewandte Chemie International Edition, 41(15), 2644-2676.

64. Leach, A. R., & Gillet, V. J. (2007). An Introduction to Chemoinformatics. Springer.

65. Gilson, M. K., & Zhou, H. X. (2007). "Calculation of protein-ligand binding affinities." Annual Review of Biophysics and Biomolecular Structure, 36, 21-42.

66. Jorgensen, W. L. (2009). "Efficient drug lead discovery and optimization." Accounts of Chemical Research, 42(6), 724-733.

67. Wang, L., et al. (2015). "Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field." Journal of the American Chemical Society, 137(7), 2695-2703.

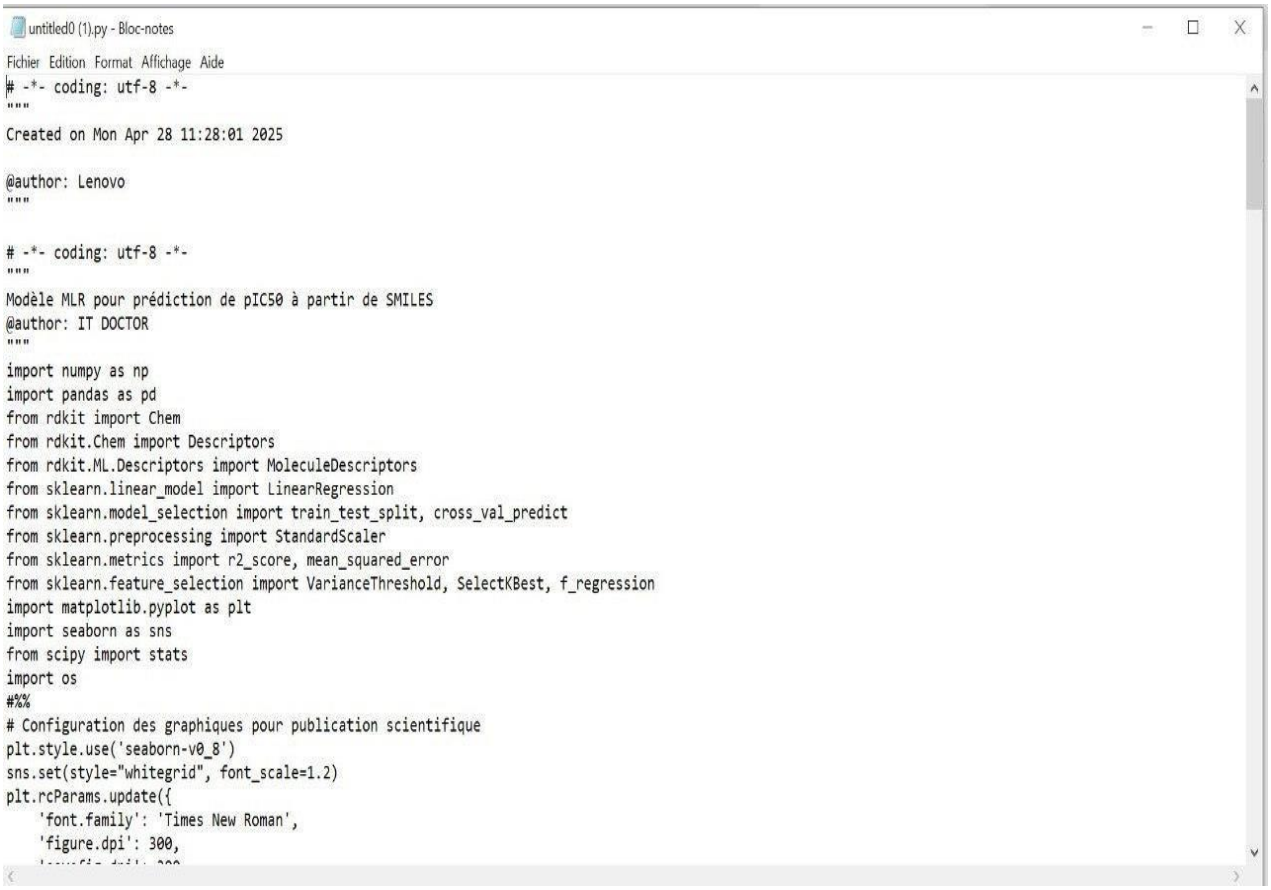
68. Tropsha, A. (2010). QSAR in Drug Discovery. In Comprehensive Medicinal Chemistry II (pp. 113-126). Elsevier. DOI:10.1017/CBO9780511730412.012

69. Chemical Computing Group (CCG) | Computer-Aided Molecular Design https://www.chemcomp.com/en/index.htm

70. rdkit.Chem.Fragments module — The RDKit 2025.03.1 documentation https://rdkit.org/docs/source/rdkit.Chem.Fragments.html

Annex

1 Annex 1: Part of MLR CODE



untitled0 (1).py - Bloc-notes

Fichier Edition Format Affichage Aide

```
# -*- coding: utf-8 -*-
"""
Created on Mon Apr 28 11:28:01 2025

@author: Lenovo
"""

# -*- coding: utf-8 -*-
"""
Modèle MLR pour prédiction de pIC50 à partir de SMILES
@author: IT DOCTOR
"""
import numpy as np
import pandas as pd
from rdkit import Chem
from rdkit.Chem import Descriptors
from rdkit.ML.Descriptors import MoleculeDescriptors
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split, cross_val_predict
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score, mean_squared_error
from sklearn.feature_selection import VarianceThreshold, SelectKBest, f_regression
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import os
#%%
# Configuration des graphiques pour publication scientifique
plt.style.use('seaborn-v0_8')
sns.set(style="whitegrid", font_scale=1.2)
plt.rcParams.update({
    'font.family': 'Times New Roman',
    'figure.dpi': 300,
```

2 Annex 2 : Part of SVR CODE

untitled1 (1).py - Bloc-notes

Fichier Edition Format Affichage Aide

```
# -*- coding: utf-8 -*-
"""
Created on Mon Apr 28 11:28:01 2025

@author: Lenovo
"""

# -*- coding: utf-8 -*-
"""
Modèle MLR pour prédiction de pIC50 à partir de SMILES
@author: IT DOCTOR
"""
import numpy as np
import pandas as pd
from rdkit import Chem
from rdkit.Chem import Descriptors
from rdkit.ML.Descriptors import MoleculeDescriptors
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split, cross_val_predict
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score, mean_squared_error
from sklearn.feature_selection import VarianceThreshold, SelectKBest, f_regression
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import os
from sklearn.svm import SVR
#%%
# Configuration des graphiques pour publication scientifique
plt.style.use('seaborn-v0_8')
sns.set(style="whitegrid", font_scale=1.2)
plt.rcParams.update({
    'font.family': 'Times New Roman',
    'figure.dpi': 300,
```