PEOPLE'S AND DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

MOHAMED KHIDER UNIVERSITY OF BISKRA

FACULTY OF EXACT SCIENCES

Department of Matter Sciences

THESIS

Presented by

Siham AGGOUN

Submitted for the award of the diploma of

DOCTORATE IN CHEMISTRY

Option :

Computational and pharmaceutical chemistry

Entitled:

Computer-aided design of a few series of heterocyclic molecules for therapeutic purposes

Publicly defended on:28/04/2025

In front of the jury composed of:

Mr. Mahmoud OMMARI	Prof	Mohamed Khider University-Biskra	President
Mr. Salah BELAIDI	Prof	Mohamed Khider University-Biskra	Supervisor
Mr. Lazhar BOUCHLALEG	MC/A	Hadj Lakhder University- Batna 2	Co-Supervesor
Mr. Ammar DIBI	Prof	Hadj Lakhder University- Batna 1	Examiner
Mr. Majdi HOCHLAF	Prof	Gustave Eiffel University-Paris	Examiner
Miss Saida KHAMOULI	MC/A	Mohamed Khider University-Biskra	Examiner

ACKNOWLEDGEMENT

First of All, I have to thank ALLAH, the almighty for everything I have and I could do…

I would like to thank my Supervisor Pr. BELAIDI Salah for accepting to supervise my doctoral research and for his precious advices, I would also thank my Co-Supervisor Mr .BOUCHLALEG Lazhar Lecturer "A" at the University of Batna 2.

I express my deep and respectful gratitude to Mr. OMMARI Mahmoud, Professor at the University of Biskra, who did me the honor of agreeing to chair the jury for this thesis.

I would like to express my sincere thanks and the expression of my to deep respect the members of the jury: to Mr. DIBI Ammar Professor at the University of Batna 1, Pr. HOCHLAF Majdi Professor at Gustave Eiffel University-Paris for their scientific aid. And Miss khamouli Saida Lecturer "A" at the University of Biskra for agreeing to examine and judge this work..

I would like to acknowledge my colleagues at University of Biskra for their support.

I would also thank my Mother,my husband and All my family members for their aid ….

THANK YOU…

This dissertation is

dedicated to the memory of my Father

And for my paradise my Mother,

They are the reason for my success.

To my husband and my children.

To my Brothers and Sisters

To all those who are dear to me.

CONTENTS

CHAPTER III: QSAR and drug-likeness study for biology activity of 1.4 dihydropyridine

LIST OF ABBREVIATIONS

- Absorption, Distribution, Metabolism, and Excretion ADME
- Austin Model 1 AM1
- Becke, three-parameter, Lee-Yang-Parr B3LYP
- Computer-Assisted Drug Design CADD
- Calcium channel blockers CCB
- Cross-Validation CV
- Density Functional Theory DFT
- Half maximal Inhibitory Concentration C50
- Hard and Soft, Acids and Bases theory HSAB
- Hartree-Fock HF
- Highest Occupied Molecular Orbital Homo
- Ligand Efficiency LE
- Ligand lipophilicity efficiency LipE
- Linear Regression LR
- Lowest Unoccupied Molecular Orbital LUMO
- Molar Refractivity MR
- Molecular electrostatic potential MEP
- Molecular electrostatic surface map MESP
- Molecular dynamics MD
- Molecular Weight MW
- Møller-Plesset level 2 MP2
- Molecular Mechanic 2 MM2

LIST OF ABBREVIATIONS

- Multiple Linear Regression MLR
- Number of Hydrogen-Bond Donors and Acceptors NHBD and NHBA
- Number of Rotatable Bonds nrotb
- Partial Least Squares PLS
- partition coefficient octanol/water logP
- Parameterized Model number 3 PM3
- Polar Surface Area PSA
- Predictive Residual Sum of the Squares PRESS
- Protein Data Bank PDB
- Quantitative Structure–Activity Relationship QSAR
- Quantitative Structure–Property Relationship QSPR
- Root-Mean Squared Error RMSE
- Structure–Activity Relationships SARs
- Topological polar surface area TPSA
- Two-Dimensional or Three-Dimensional (2D or 3D) QSAR

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

Scheme

LIST OF TABLES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

GENERAL INTRODUCTION

GENERALINTRODUCTION

Dihydropyridines (DHPs) are a family of bioorganic compounds based on the pyridine nucleus. Among the existing isomeric forms of DHPs is 1, 4-dihydropyridine (DHP14). The synthesis of the first nucleus of 1, 4-dihydropyridine was reported about 130 years ago by Hantzsch [1]. DHPs are structures that can produce selective and robust ligands for a wide variety of biological targets [2]. Its original use was as calcium channel antagonists for the treatment of hypertension disease from the 1970s [3].

Since then, DHP14 derivatives were tested for their biological activity and were introduced into clinical medicine for widely differing biological targets. For instance, they are known as vasodilators, analgesic, bronchodilators, antihero sclerotic, anti-tumor, anti- inflammatory and anti-diabetic agents [4-12].

Its derivatives have become over time one of the most used classes of drugs for the treatment of various cardiovascular diseases [13-15].They act as calcium (Ca2+) channel blockers (CCBs), which, as the Calcium Channel antagonists or calcium antagonists, disrupt the transit of Ca2+ cation through calcium channels [3,16].In particular, the short half-lives of the DHPs 14, like benidipine, felodipine, isradipine, clevidipine [17],nifedipine, nitrendipine, reduce systemic vascular resistance and arterial pressure. The Hantzsch synthesis generally produces racemic mixtures of such compounds. Thus, the development of the controlled stereoselective synthesis of 1,4-dihydropyridines is one of the targets of current research in medicinal chemistry [18].

Nowadays, the use of Quantitative Structure–activity relationship (QSAR) studies for the development and design of drugs is a current use because it allows bettering understanding the chemical structure and the mechanisms of action of drugs on the targets at the molecular levels. QSAR analysis is useful tool in the field of designing rational drugs and discovering the mechanism of pharmacological actions. The basic concept of QSAR methodology is to find a reliable correlation between biological activity and molecular structure. Recent works were devoted into using artificial neural network and multiple regression analysis to provide predictive QSAR models with high reliability [19]. This is the main focus of the present work dealing with 1,4-dihydropyridines as CCBs using similar approaches.

In the literature, there are many QSAR studies of DHP14 derivatives targeting the design of novel CCBs. Back in 2003, Safarpour et al [20] conducted various quantum chemical-QSAR studies of some newly synthesized DHP14 using density functional theory (DFT) calculations. For many of these studies, a nonlinear relationship between the DHPs activity and calculated descriptors was concluded. One year later, Hemmateenejad et al. [15] have used GA-MLR and PC-GA-ANN procedures to study

the CCBs activity of DHP14–based Nifedipine analogous. In 2005, Yao et al. [21] used Least Squares Support Vector Machine (LSSVM) correlation to construct selective models with classification as a major screening mechanism for a novel series of DHP14 antagonists. In 2013, Hadizadeh et al. [22] used molecular descriptors calculated by the DRAGON software to develop MLR model that can be used to design new compounds with CCBs activity. Also, Da Mota et al. [23] proposed two compounds with high predicted activities and pharmacokinetic data comparable with those of known CCBs after MIA-QSAR, molecular docking and computational drug-likeness studies. More recently, Jardínez et al. [24] showed that the reduced density gradient approach can be used for estimating QSAR descriptors. They applied this approach to 1, 4-dihydropyridine derivatives with potential antihypertensive effects. Additionally, El-Moselhy et al. [25] developed a 3D-QSAR model; the latter is based on synthesized compounds, but previously inspired by molecular docking calculations before moving to the experimental stage. The proposed model indicates that the importance of lipophilicity over the electronegativity of substituents deduced by the high values of biological activity. Relying on the afore mentioned statement, thirty six DHP14 derivatives previously synthesized and evaluated as Calcium Channel Blockers (CCBs) were selected to build both robust and reliable linear and nonlinear models. This series was proposed by Navidpour et al.[26] and the corresponding DHP14 derivatives containing lipophilic 4-imidazolyl substituents. Here, multiple regression linear (MLR) and artificial neuron network (ANN) were employed for the construction of QSAR models. DFT method was used for geometry optimization of molecules and calculations of their electronic descriptors. The obtained models were further used as guide to design new compounds with enhanced CCBs activities.

Our main contributions are summed in these essential points, namely:

- Structural and electronic study of the basic nuclei of heterocyclic compounds of 1,4-dihydropyridine.
- Drug-likeness study using several empirical rules such as Lipinski's rules ,Veber Score and Ghose rules.
- Establish at the molecular level 2D-QSAR (MLR/ANN) models for 1,4-dihydropyridine derivatives.
- Analysis by the molecular docking method of the most active chemical of 1.4 dihydropyridine derivatives and the reference ligand.

In order to carry out this work properly and to achieve the main objectives, we have organized our thesis into four chapters:

CHAPTERI: DIHYDROPYRIDINES (DHPS) AND THEIR USE IN THE HYPERTENSIVE

TREATMENT

In the chapter I we will present dihydropyridine molecule and describe the hypertension diseases, their pathogenesis and their treatment. In addition, it contains a description of some their inhibitors resistance.

CHAPTER II: DEVELOPMENT , VALIDATION AND APPLICATION OF QSAR/QSPR METHODS

The second chapter present a bibliographical study on various QSAR/QSPR methodologies, covering their stages of development, validation, and application.

CHAPTERIII: DRUG-LIKENESS AND STRUCTURE-ACTIVITY/QUALITATIVE PROPERTIES RELATIONSHIP OF 1.4 DIHYDROPYRIDINE DERIVATIVES

- The 1st point, a 2 Dimension Quantitative structure-activity relationship (2D-QSAR) models were generated using MLR and ANN methods for series of 31 derivatives of 1,4-dihydropyridine with the use of 29 molecular descriptors.
- In the 2nd point, we will present drug-likeness screening studies of 1,4-dihydropyridine derivatives.
- At last, the obtained QSAR models were employed to define biological activities of potentially novel active compounds.

CHAPTER IV: MOLECULAR DOCKING STUDIES OF 1.4 DIHYDROPYRIDINE DERIVATIVES

A molecular docking analysis recognizes which molecule; the most active compound or the reference ligands can then form novel drugs.

We conclude this thesis with a general conclusion with perspectives.

REFERENCES

[1] Hantzsch A(1881) Condensation products made of aldehydammoniak and keton-like connections. Chem. Ber 14:1637-1638.

[2] Eisner E, Kuthan E,J(1972) Chemistry of dihydropyridines. Chem. Rev 72:1-42.

[3] Triggle D.J (2003) 1,4-Dihydropyridines as calcium channel ligands and privileged structures .Cell. Mol. Neurobiol 23: 293-303.

[4] Edraki N,MehdipourA.R ,Khoshneviszadeh M , Miri R (2009) Dihydropyridines: evaluation of their current and future pharmacological applications. Drug Discov.Today 141058-66.

[5] Bossert F, Vater W (1989) 1,4-Dihydropyridines-a basis for developing new drugs. Med. Res. Rev :9291-324.

[6] Bossert F, Vater W (1971)Dihydropyridine, eineneue Gruppe stark wirksamerCoronartherapeutika [Dihydropyridines, a new group of strongly effective coronary therapeutic agents].Naturwissenschaften 58578.

[7] Epstein M, Black H.R(2001) Arterial Calcification and Calcium Antagonists.Hypertension 37:1414–1415.

[8] Klusa Atypical V 1,4-dihydropyridine derivatives, an approach to neuroprotection and memory enhancement.Pharmacol. Res 113:754-759.

[9] Sharma V.K, Singh S.K(2017) Synthesis, utility and medicinal importance of 1,2- & 1,4-dihydropyridines. RSC Adv 7:2682–2732.

[10] Frank C.A, Forst J.M, Harris R.J, Kau S.T, Li J.H, Ohmmachb C.J, Smith R.W, Trainor D.A, Trivedi S(1993) Dihydropyridine KATP potassium channel openers. Bioorg. Med. Chem. Lett 3:2725–2726.

[11] Bocker R.H, Guengerich F.P(1986) Oxidation of 4-aryl- and 4-alkyl-substituted 2,6-dimethyl-3,5-bis(alkoxycarbonyl)-1,4-dihydropyridines byhuman liver microsomes and immunochemical evidence for the involvement of a form of cytochrome P-450.J. Med. Chem 29:1596-1603.

[12] Ioan P ,Carosati E , Micucci M , Cruciani G , Broccatelli F ,Zhorov B. S, Chiarini A , Budriesi R (2011)1,4-Dihydropyridine Scaffold in Medicinal Chemistry, The Story so Far And Perspectives (Part 1): Action in Ion Channels and GPCRs.Current Medicinal Chemistry 18:4901-4922.

[13] Huang Q , Li Y, Sheng C, Dou Y, Zheng M , Zhu Z , Wang J (2015) Blood pressure lowering efficacy of amlodipine and nifedipine-Gits in ambulatory hypertension .J. Hypertens33:94.

[14] Gordeev M.F, Patel D.V, Gordon E.M (1996) Approaches to the combinatorial synthesis of heterocycles: a solid-phase synthesis of 1,4-dihydropyridines.J.Org.Chem 61:924-928.

[15] Hemmateenejad B, Safarpour M.A, Miri R, Taghavi F(2004) Application of ab initio theory to QSAR study of 1,4-dihydropyridine-based calcium channel blockers using GA-MLR and PC-GA-ANN procedures. J. Comput. Chem 25:1495–1503.

[16] Mannhold R , Jablonka B, Voigdt W, Schoenafinger K, Schravan K (1992) Calcium- and calmodulin-antagonism of elnadipine derivatives: comparative SAR. Eur. J. Med. Chem 27:229-235.

[17] Yet L(2018)1,4-dihydropyridines. In: Privileged Structures in Drug Discovery. Hoboken, NJ, USA: John Wiley & Sons. Inc 5:9–82.

[18] Rucins M, Plotniece A ,Bernotiene E, Tsai W- B ,Sobolev A(2020) Recent Approaches to Chiral 1,4-Dihydropyridines and their Fused Analogues.Catalysts 10:1019.

[19] Hanachi R, Ben Said A , Allal H, Rahali S ,Alkhalifah M.A.M, Alresheedi F ,Tangour B, Hochlaf M (2021**)** Structural, QSAR, machine learning and molecular docking studies of 5-thiophen-2-yl pyrazole derivatives as potent and selective cannabinoid-1 receptor antagonists. New J. Chem 45:17796-17807.

[20] Safarpour M. A, Hemmateenejad B, Miri R, Jamali M (2003) Quantum Chemical-QSAR Study of Some Newly Synthesized 1,4-Dihydropyridine Calcium Channel Blockers. QSAR & Combinatorial Science 22: 997–1005.

[21] Yao X, Liu H, Zhang R, Liu M, Hu Z, Panaye A, Doucet J.P, Fan B (2005) QSAR and classification study of 1,4-dihydropyridine calcium channel antagonists base on least squares support vector machines. Mol. Pharm 2(5): 348-56.

[22] Hadizadeh F, Vahdani S, Jafarpour M (2013) Quantitative Structure-Activity Relationship Studies of 4-Imidazolyl- 1,4-dihydropyridines as Calcium Channel Blockers. Iran. J. Basic Med. Sci 16:910-916.

[23] Da Mota E. G, Silva D.G, Guimarães M. C, Da Cunha E. F. F,Freitas M. P(2014) Computer-assisted design of novel 1,4-dihydropyridine calcium channel blockers. Mol. Simul 40:959-965.

[24] Jardínez C,Vela A,Cruz-Borbolla J, Alvarez-Mendez R.J, Alvarado-Rodríguez J.G (2016) Reduced density gradient as a novel approach for estimating QSAR descriptors, and its application to 1,4-dihydropyridine derivatives with potential antihypertensive effects. J. Mol. Model 22:296.

[25] El-Moselhy T. F, Sidhom P. A , Esmat E. A, El-Mahdy N. A (2017) Synthesis, Docking Simulation, Biological Evaluations and 3D-QSAR Study of 1,4-Dihydropyridines as Calcium Channel Blockers. Chem. Pharm. Bull. (Tokyo) 65:893-903.

[26] Navidpour L,Miri R ,Shafiee A(2004) Synthesis and calcium channel antagonist activity of new 1,4-dihydropyridine derivatives containing lipophilic 4-imidazolyl substituents.Arzneim. Forsch. Drug Res 54:499-504.

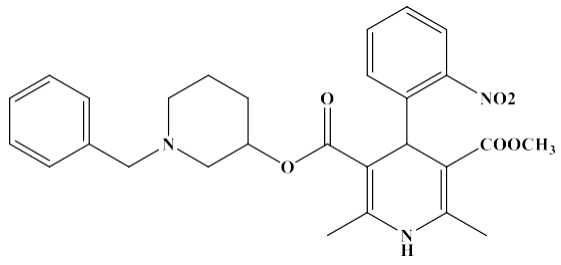
CHAPTER I : DIHYDROPYRIDINES (DHPS) AND THEIR USE IN THE HYPERTENSIVE TREATMENT

I.1. INTRODUCTION

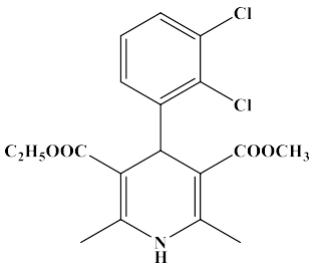
Dihydropyridines (DHPs) are vital nitrogen-containing heterocycles recognized as fundamental frameworks in various biologically active compounds and natural substances. They also play crucial roles as medicinal agents and important starting materials for intricate molecules. Moreover, certain DHPs display blue fluorescence in organic solvents, indicating potential applications in functional materials. With their versatile applications in medicinal chemistry and materials science, there's been considerable focus on establishing efficient and reliable synthetic approaches for their synthesis. The proliferation of synthetic routes for DHPs underscores the escalating interest in their production[1].

Hantzsch reported the first synthesis of the initial nucleus of 1,4-dihydropyridine approximately 130 years ago [2]. In the mid-1970s, its primary applications emerged as calcium channel antagonists for treating hypertension [3]. Since then, derivatives of DHP14 have undergone testing for their biological effects and have been incorporated into clinical practice for various purposes. They are recognized for their roles as vasodilators, analgesics, bronchodilators, anti-atherosclerotic, anti-tumor, anti-inflammatory, hepatoprotective, and anti-diabetic agents [4-12]. Over time, derivatives of these compounds have become one of the most utilized drug classes for treating a range of cardiovascular diseases [13-15]. They function as calcium channel blockers (CCBs), also known as calcium antagonists, inhibiting the passage of Ca2+ ions through calcium channels [16]. The brief half-lives of DHPs 14, including benidipine, felodipine, isradipine, clevidipine [17], nifedipine, and nitrendipine (*Figure. I.1*), specifically lower systemic vascular resistance and arterial pressure. The Hantzsch synthesis typically yields racemic blends of these compounds. Hence, achieving controlled stereoselective synthesis of 1,4-dihydropyridines is a focal point of ongoing research in medicinal chemistry [18].

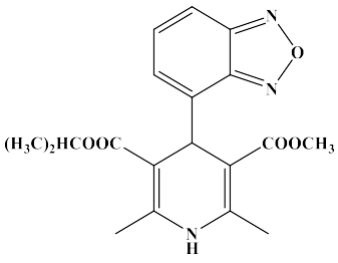
-



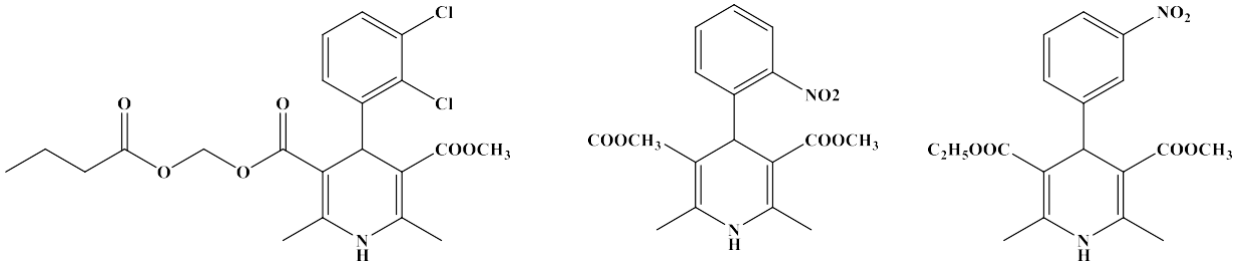
Benidipine



Felodipine



Isradipine



Clevidipine **Nifedipine** **Nitrendipine**

Figure I.2. Structures of some 1,4-dihydropyridine derivatives.

I.2. DIHYDROPYRIDINE DERIVATIVES

Dihydropyridine derivatives are categorized as calcium antagonist medications employed to treat hypertension and angina pectoris.

I.2.1. Classification

I.2.1.1. Compounds with predominant vascular effects

Drugs belonging to the dihydropyridine class bind to slow voltage-gated (or L-type) calcium channels, leading to their inactivation. These channels are present in vascular and cardiac smooth muscles, as well as other tissues. These molecules exhibit a significant peripheral arterial and coronary vasodilator effect while exerting minimal influence on cardiac tissue conduction. Consequently, they cause a reduction in blood pressure without affecting heart rate. For instance, nimodipine, a member of this class, is effective in treating cerebral vascular spasms following subarachnoid hemorrhage.

Main classes of these compounds: amlodipine, nifedipine , isradipine , felodipine, , lacidipine, nicardipine, nitrendipine.

I.2.1.2. Compounds with vascular and cardiac effects

Compounds with vascular and cardiac effects, such as benzothiazepines and phenylalkylamines, act on activated voltage-gated L calcium channels. Like dihydropyridines, these molecules produce similar effects at the vascular level. However, they also influence the cardiac rhythmic center, resulting in a decrease in the frequency of action potentials at the sinus node and a slowing of intracardiac conduction at the atrioventricular node and the bundle of His. These interactions collectively lead to a reduction in blood pressure by altering both heart rate and peripheral resistance of blood vessels.

Some examples of molecules - phenylalkylamines (Verapamil); benzothiazepines (Diltiazem).

I.2.1.3. Calcium Channel Blocker (CCB)

Calcium channel blockers (CCBs) are molecules with diverse chemical compositions, primarily originating from pyridine. They are prescribed for managing various cardiac conditions including

angina, arrhythmias, and hypertension. These compounds are divided into two groups based on their sites of action, acting on voltage-gated calcium channels to impede the typical influx of calcium ions into vascular and cardiac striated smooth muscle cells. Adverse effects have been documented, including instances of severe poisoning.

Calcium antagonists inhibit the entry of calcium into cells, particularly within the cardiovascular system; however, distinct classes produce notably different effects.

Phenylalkylamines such as gallopamil and verapamil exert a reduction in cardiac frequency (negative chronotropic effect), contractility (negative inotropic effect), conductivity (negative dromotropic effect), and contraction of vascular smooth muscles.

Dihydropyridines primarily induce vasodilation with lesser impact on cardiac function. Examples include amlodipine, barnidipine, felodipine, isradipine, lacidipine, lercanidipine, nicardipine, nifedipine, nimodipine, nisoldipine, and nitrendipine.

Benzothiazepine derivatives, such as diltiazem, exhibit effects intermediate between those of verapamil and dihydropyridines.

Etripamil, a distinct calcium channel blocker, can be administered via intranasal spray and is in development for the treatment of supraventricular tachycardia attacks.

- **Amlodipine besylate** is a long-acting L-type calcium channel antagonist used as an antihypertensive and for the treatment of angina.
- **Felodipine** is a calcium antagonist belonging to the dihydropyridine class. Its primary uses include treating angina and hypertension.
- **Lacidipine**, classified as a dihydropyridine calcium antagonist, is primarily prescribed for hypertension treatment.

- **Isradipine** is a dihydropyridine calcium antagonist. Its main indication is hypertension.
- Lercanidipine, belonging to the dihydropyridine family, functions as a calcium antagonist and is primarily utilized as an antihypertensive medication.

- **Nicardipine,** classified as a dihydropyridine calcium antagonist, is primarily indicated for the treatment of angina and hypertension.

- **Nifedipine**, a dihydropyridine calcium antagonist, is primarily prescribed for angina and hypertension. However, it has also been increasingly used for various other indications,

including Raynaud's disease, among others.

- **Nimodipine**, categorized as a dihydropyridine calcium antagonist, is primarily used to prevent cerebral arterial spasms, a complication of subarachnoid hemorrhage. Additionally, it can serve as a first-line treatment for reversible cerebral vasoconstriction syndrome (RCVS).

- **Barnidipine** is a dihydropyridine calcium antagonist. It is indicated against hypertension.

- **Nisoldipine,** belonging to the dihydropyridine family, functions as a calcium antagonist primarily employed as an antihypertensive medication.

- **Nitrendipine**, belonging to the dihydropyridine family, functions as a calcium antagonist predominantly employed as an antihypertensive medication.

I.2.1.3. a . Pharmacological Characteristics:

Calcium channel blockers adhere to slow voltage calcium channels (L-type). These channels regulate the electrochemical gradient, which is significantly higher in extracellular calcium concentration compared to the intracellular environment, by a factor of 10,000. The channels transition from a closed to an open state in response to a depolarizing membrane potential initiated by acetylcholine release at a neuromuscular junction. This intracellular calcium release disengages the actin-myosin binding sites, leading to muscle contraction. Consequently, they regulate cardiac automatism at the atrioventricular sinus node and the bundle of His, as well as vascular tone, as these channels are predominantly found in the cardiovascular system. When inhibitors attach to the channels, they curtail the release of calcium from the channel pores, thereby limiting muscle contraction.

Dihydropyridines exhibit a strong attraction to inactivated channels, primarily found on smooth muscle cells owing to their extended depolarization.

I.2.1.3. b . Therapeutic Uses:

1. Treatment of hypertension: Calcium channel blockers induce vasodilation, diminishing peripheral resistance, and consequently lowering blood pressure. They can be administered alone or in conjunction with β-blockers (e.g., Atenolol) or ACE inhibitors.
2. Management of arrhythmias: Calcium channel blockers are employed to lower heart rate by acting on the heart's rhythmic center.

3. Prevention of angina: Angina is chest discomfort that arises during physical exertion and subsides upon cessation. It is a manifestation of myocardial ischemia. Calcium channel blockers aid in reducing coronary spasms that lead to ischemia or chest pain.
4. Raynaud's syndrome: This condition involves the constriction of small arteries in the extremities without cold-induced triggers. Peripheral vasodilation induced by calcium channel blockers can alleviate this issue.

I. 3. HYPERTENSION

Cardiovascular diseases are the leading cause of death in developed countries and most developing countries . constituting a major public health concern across different regions. They are the primary cause of mortality in Algeria [19]. Epidemiological research on health in Algeria has highlighted that cardiovascular diseases are the most prevalent ailments .

Changes in lifestyle, dietary habits, and behaviors resulting from significant urbanization have been associated with an increase in cardiovascular disease risk factors in urban areas. Epidemiological studies conducted in various European populations have demonstrated the considerable influence of environmental factors and explained the variability of cardiovascular disease across Europe. In Algeria, there is limited research on the influence of residence on the incidence of cardiovascular diseases.

In a survey conducted in Algeria, the prevalence of cardiovascular risk factors , in a representative sample of the adult urban and rural populations. The overall participation rate in the study was 72.0%, with higher participation in rural areas (87.3%) compared to urban areas (61.6%). This discrepancy can be partially explained by the lower rate of professional activity among women in rural areas compared to their urban counterparts.

The prevalence of arterial hypertension is 22.3% in the studied population and is higher in urban areas (28.0-40.0%) compared to rural areas (16.8%), especially among women. Previous studies have already observed that arterial hypertension increases with age. Additionally, numerous epidemiological studies have demonstrated that arterial hypertension is a major risk factor for ischemic heart disease in both developing and developed countries. [20].

I. 3. 1.High blood pressure (hypertension)

High blood pressure, also known as adult essential hypertension or simply hypertension, often has no symptoms but is a major risk factor for cardiovascular disease. Adopting a healthier lifestyle can help prevent hypertension or reduce its complications.

Arteries carry blood from the heart to the organs. Blood pressure measures the force exerted by blood on the walls of the arteries. A certain level of pressure is necessary for blood to circulate throughout the

body.

Blood pressure is assessed using two values. The first value is the pressure during heart contraction, known as systolic or maximum pressure. The second value is the pressure during heart relaxation, known as diastolic or minimum pressure. These pressures are typically measured with a blood pressure monitor placed around the arm. Blood pressure naturally increases with age. On average, every 10 years, systolic pressure rises by 0.5 and diastolic pressure by 0.2. In more than half of people over 60, even those in good health, systolic pressure (the first number) increases above 14. (Figure I. 2)

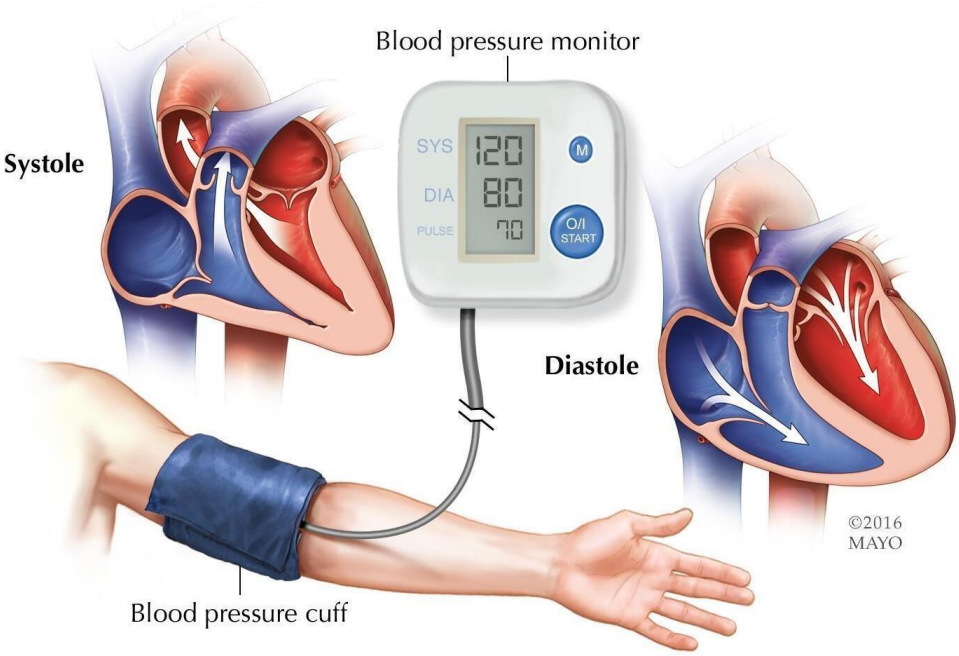


Figure I. 2. Method of measuring blood pressure

High blood pressure occurs when the pressure in the arteries is too high. Normal blood pressure values typically range between 10 and 14 for systolic (maximum) pressure and 6 and 8 for diastolic (minimum) pressure, with 12/8 considered normal. High blood pressure is a major risk factor for heart disease, kidney failure, and stroke. It generally develops with age and is often associated with excess weight.

Various degrees of hypertension		
	Systolic pressure	**Diastolic pressure**
HTA sévère	**> 180 mmHg**	**> 110 mmHg**
HTA stade 2	**> 160 mmHg**	**> 100 mmHg**

HTA stade 1	**> 140 et ≤ 159 mmHg**	**> 90 et ≤ 99 mmHg**
pré HTA[1]	**> 120 et ≤ 139 mmHg**	**> 80 et ≤ 89 mmHg**

Table I. 1. Various degrees of hypertension

I. 3. 2.Symptoms of high blood pressure

Most people with hypertension do not exhibit any symptoms. However, extremely high blood pressure can cause headaches, blurred vision, chest pain, and other symptoms. The best way to determine if you have high blood pressure is by checking it regularly. Untreated hypertension can lead to serious health issues such as kidney disease, heart disease, and stroke.

People with very high blood pressure (typically 180/120 or higher) may experience symptoms including:

- Severe headache
- Dizziness
- Nausea
- Anxiety
- Nosebleeds
- Ringing in the ears
- Confusion
- Irregular heart rhythm
- Chest pain
- Blurred vision or other vision changes
- Difficulty breathing
- Vomiting

I. 3. 3. Causes of arterial hypertension

In over 95% of cases, the cause of hypertension is unknown, and treatment focuses on reducing blood pressure without addressing the underlying cause. In other cases, hypertension is secondary to conditions like kidney, adrenal gland, or thyroid dysfunction.

Several factors are known to worsen high blood pressure:

- Excessive salt consumption
- Physical inactivity
- Smoking
- Stress
- Obesity

High blood pressure tends to develop earlier in men. Women of childbearing age are relatively protected due to the beneficial effects of certain sex hormones, particularly estrogens. However, the incidence of hypertension in women rises to match that of men after menopause.

Excess weight, obesity, and type 2 diabetes are increasingly common in individuals with high blood pressure. Hypertension is twice as prevalent in overweight individuals, and obese elderly people are 1.5 times more likely to have high blood pressure compared to those of normal weight. A study of type 2 diabetes patients found that high blood pressure affected one-third of men and half of women.

Additionally, certain medications or substances can promote or worsen high blood pressure, or even destabilize treated hypertension. These include estrogens, nasal decongestant sprays, non-steroidal anti-inflammatory drugs (such as aspirin, ibuprofen, and ketoprofen), glucocorticoids (like cortisone and dexamethasone), alcohol, licorice, and anise-flavored drinks such as pastis.

I. 3. 4. Complications of high blood pressure

Untreated hypertension can lead to several serious arterial problems, including:

- Cerebrovascular accidents (strokes or transient ischemic attacks)
- Intracranial hemorrhages
- Heart failure
- Retinal damage, sometimes resulting in vision loss.
- Myocardial infarctions (heart attacks)
- Kidney damage, potentially causing kidney failure

I. 3. 5. Medical treatment

When lifestyle changes and dietary adjustments are insufficient to reduce blood pressure, medication is prescribed. Treatment may involve several types of medications. Five classes of antihypertensive drugs are commonly used: diuretics, beta blockers, calcium channel blockers, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin II receptor blockers (ARBs). These medications have demonstrated effectiveness in preventing cardiovascular events in individuals with hypertension.

Groups		DCI	Name of specialities
Dihydropyridines		Amlodipine	AMLOR
		Félodipine	FLODIL

		Isradipine	**ICAZ**
		Lacidipine	**CALDINE**
		Nicardipine	**LOXEN**
		Nitrendipine	**NIDREL**
		Lercanidipine	**LERCAN ® ZANIDIP**
		Manidipine	**IPERTEN**
		Nifédipine	**ADALATE CHRONADALATE LP**

Table I. 2. Some blood pressure medications derived from dihydropyridine

Hypertension poses a widespread medical and social challenge, contributing to cardiovascular diseases on a global scale. Antihypertensive medications are commonly used in clinical practice to reduce both the morbidity and mortality associated with hypertension and its related complications. The 2014 hypertension guideline issued by the Eighth Joint National Committee (JNC8) in the United States introduced several notable alterations in the clinical approach to managing hypertension and the primary medications recommended, marking a departure from previous guidelines [21].

In addition to the guidance specifying the initiation of pharmacological treatment when blood pressure (BP) reaches 150/90 mmHg or higher in adults aged over 60, 140/90 mmHg in adults under 60, or 140/90 mmHg or higher (regardless of age) in individuals with hypertension and diabetes, healthcare providers should contemplate initiating treatment with a thiazide-type diuretic, calcium (Ca2+) channel blocker (CCB), angiotensin-converting enzyme inhibitor (ACEI), or angiotensin receptor blocker (ARB) as the initial antihypertensive medication in non-black populations. For black populations, whether with or without diabetes, initial antihypertensive therapy should involve a thiazide-type diuretic or CCB. [22]

As a result, calcium channel blockers (CCBs) have emerged as crucial first-line agents for monotherapy in treating hypertension. Moreover, their demonstrated safety profile in not elevating the risk of coronary events and stroke further solidifies their significance. [23]

I. 5. CONCLUSION

Dihydropyridines (DHPs) have significantly transformed pharmaceutical research due to their remarkable biological properties. DHPs are highly reactive, allowing for the synthesis of diverse compounds with substantial medicinal value, including natural products. Notable examples include alkaloids such as deplancheine, tangutorine, dihydroakummicine, olivacine, -guatambuine, l-pipecolic acid, -geissoschizine, -akagerine, lyaline, lyadine, harman-dihydropyrimidine, camptothecin, 20-deoxycamptothecin, akuammiline alkaloid precursors, silicine–methuenine alkaloids, vinoxine, -2,7-dihydropleiocarpamine, tubifoline, tubifolidine, and ervitsine .

this remains a dynamic and challenging area for medicinal chemists to explore new synthetic methodologies or integrate different approaches to develop novel DHP-based drugs. Many DHP-based drugs have been blockbusters in their time; hence, we have also briefly highlighted the commercial value of previously approved DHP-based drugs.

Calcium channel blockers that are often used to reduce systemic vascular resistance and arterial pressure. The majority of the dihydropyridines contain either a 2-nitrophenyl, as in nifedipine and nisoldipine, or a 3-nitrophenyl substituent at the 4-position of the dihydropyridine moiety, as in lercanidipine, nicardipine, nimodipine, nitrendipine, and nivaldipine.

REFERECES

[1] Wang1 A. L, Iadecola C, Wang G (2017) New generations of dihydropyridines for treatment of hypertension . Journal of Geriatric Cardiology 14: 67–72 Premed Program, Washington University in St. Louis, St. Louis, MO, USA 2Feil Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY, US

[2] Hantzsch A (1881) Condensation products made of aldehydammoniak and keton-like connections. Chem. Ber 14: 1637-1638.

[3]. Triggle D.J, (2003) 1,4-Dihydropyridines as calcium channel ligands and privileged structures .Cell. Mol. Neurobiol 23: 293-303.

[4] Edraki N,Mehdipour A.R , Khoshneviszadeh M , Miri R (2009) Dihydropyridines: evaluation of their current and future pharmacological applications. Drug Discov. Today 14 :105866.

[5] Bossert F , Vater W (1971) Dihydropyridine, eine neue Gruppe stark wirksamer Coronartherapeutika. Natur wissenschaften 58:578.

[6] Bossert F , Vater W (1989) 1,4-Dihydropyridines-a basis for developing new drugs. Med. Res. Rev 9 :291-324.

[7] Frank C.A , Forst J.M ,Harris R.J ,Kau S.T ,Li J.H , Ohmmachb C.J ,Smith R.W,Trainor D.A,Trivedi S (1993) Dihydropyridine KATP potassium channel openers. Bioorg. Med. Chem. Lett 3:2725–2726.

[8] Epstein M ,Black H.R (2001) Arterial Calcification and Calcium Antagonists. Hypertension 37 :1414–1415.

[9] Klusa V(2016) Atypical 1,4-dihydropyridine derivatives, an approach to neuroprotection and memory enhancement. Pharmacol. Res 113 :754-759.

[10] Bocker R.H , Guengerich F.P (1986) Oxidation of 4-aryl- and 4-alkyl-substituted 2,6-dimethyl-3,5-bis(alkoxycarbonyl)-1,4-dihydropyridines by human liver microsomes and immunochemical evidence for the involvement of a form of cytochrome P-450. J. Med. Chem 29 :1596-1603.

[11] Sharma V.K, Singh S.K (2017) Synthesis, utility and medicinal importance of 1,2- & 1,4-dihydropyridines. RSC Adv 7 :2682–2732.

[12] Ioan P , Carosati E , Micucci M , Cruciani G, Broccatelli F ,Zhorov B. S, Chiarini A, Budriesi R (2011) 1,4-Dihydropyridine Scaffold in Medicinal Chemistry, The Story so Far And Perspectives (Part 1): Action in Ion Channels and GPCRs. Current Medicinal Chemistry 18 :4901-4922.

[13] Gordeev M.F , Patel D.V , Gordon E.M (1996) Approaches to the combinatorial synthesis of heterocycles: a solid-phase synthesis of 1,4-dihydropyridines. J.Org.Chem 61:924-928.

[14] Huang Q , Li Y , Sheng C , Dou Y , Zheng M , Zhu Z , Wang J (2015) Blood pressure

lowering efficacy of amlodipine and nifedipine-Gits in ambulatory hypertension. J. Hypertens 33 :94.

[15] Hemmateenejad B , Safarpour M.A, Miri R ,Taghavi F (2004) Application of ab initio theory to QSAR study of 1,4-dihydropyridine-based calcium channel blockers using GA-MLR and PC-GA-ANN procedures. J. Comput. Chem 25 :1495–1503.

[16] Mannhold R , Jablonka B. Voigdt W , Schoenafinger K. Schravan K (1992) Calcium- and calmodulin-antagonism of elnadipine derivatives: comparative SAR. Eur. J. Med. Chem 27 :229-235.

[17] Yet L (2018) 1,4-dihydropyridines. In: Privileged Structures in Drug Discovery. Hoboken, NJ, USA: John Wiley & Sons, Inc 5 :9–82.

[18] Rucins M , Plotniece A , Bernotiene E,Tsai W- B , Sobolev A (2020) Recent Approaches to Chiral 1,4-Dihydropyridines and their Fused Analogues. Catalysts 10:1019.

[19] INSP. Mortalité générale. Projet TAHINA. Institut national de santé publique Alger. Algérie 2002 :1–20.

[20] Boukli Hacène L, Meguenni K(2007)Facteurs de risque cardiovasculaire dans la communauté urbaine de Tlemcen (Algérie). Cah Sante.17(3) :153–7.

[21] Jarari N, Rao N, Peela JR, et al (2015) A review on prescribing patterns of antihypertensive drugs. Clin Hypertens. ;22:7.

[22] Armstrong C, Joint National C (1014) JNC8 guidelines for the management of hypertension in adults. Am Fam Physician. 90:503–504.

[23] Verdecchia P, Reboldi G, Angeli F, et al (2005) Angiotensin-con verting enzyme inhibitors and calcium channel blockers for coronary heart disease and stroke prevention. Hypertension . 46: 386–392.

CHAPTER II

DEVELOPMENT , VALIDATION AND APPLICATION OF QSAR/QSPR METHODS

II.1. INTRODUCTION

There are many theoretical chemistry methods to determine physical or chemical properties of molecules. We can differentiate two main classes of simulation methods: the first, methods of quantum chemistry that can accurately determine the electronic properties molecules, and the second, molecular mechanics methods that are based on empirical parameters which make it possible in particular to determine the parameters. These methods make it possible to calculate the physic chemical parameters used in the QSAR study; where used to identify important structural features responsible for activity of drugs. Quantitative structure–activity relationships (QSARs) are a significant factor in drug design; consequently, it is quite evident why a many users of QSAR are located in industrial research units [1-4].

By employing these methods, it becomes possible to validate existing experimental data and predict the properties or activities of new compounds or those lacking experimental data. This chapter presents a bibliographical study on various QSAR/QSPR methodologies, covering their stages of development, validation, and application [5].

II.2. QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS / PROPERTY QSAR/QSPR

II.2.1. Definition

QSAR/QSPR methods are founded on the principle that the activity or property of a chemical compound is associated with its structure. Specifically, these methods assert that the activity (or property) and structure of chemical compounds can be linked through a mathematical algorithm, based on the fundamental as sumption that "similar chemical compounds exhibit similar activities." Further more, when molecular parameters are expressed as numerical values, it becomes possible to propose a mathematical relationship, known as a quantitative structure-activity/property relationship, between the two ; there fore QSAR/QSPR is a mathematical model which associates one or more quantitative parameters derived from the chemical structure, with a quantitative measure of a property or activity.

II.2.2. Principle

The principle of an QSAR/QSPR study involves establishing a mathematical relationship that quantitatively connects a biological activity or property, measured for a series of similar compounds under identical experimental conditions, with molecular descriptors using statistical methods. The aim of these studies is to analyze the structural data in order to identify the key factors influencing the studied activity or property. To accomplish this, various types of statistical methods can be utilized [6].

The resulting mathematical expression can be employed to predict the activity or property of new molecules or those lacking experimental data. This relationship can be represented by the following

equation :

Activity/Property = f (molecular descriptors)

II.2.3. Global strategy

The development of a model begins with the search for the maximum possible reliable experimental data. Then, the development of a series of descriptors which characterize the molecular structures of the compounds of the database in order to link them to the activity/experimental property studied. Once developed, the model must be validated in terms of correlation (on the training dataset). The influence of the training set compounds on the model (model robustness) is estimated by internal validation methods. To estimate the predictive power of the model, it is necessary to have additional experimental data (external validation dataset) in order to determine the ability of the model to predict these values. Finally, for any model, it is important to know for which type of molecules it is usable or not, i.e. to know its domain of applicability [7].

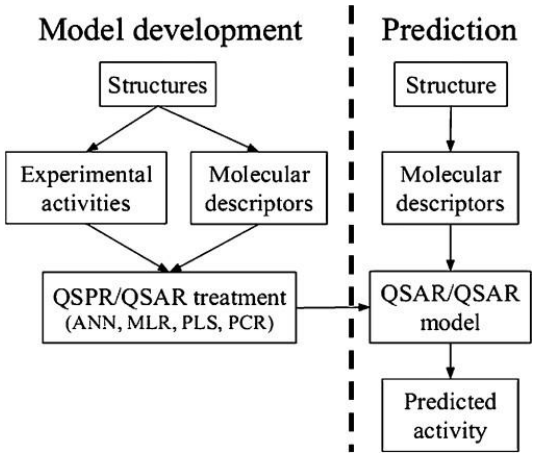


Figure II.1 . Global strategy of QSAR/QSPR model

A QSAR/QSPR model relates, in a qualitative or quantitative way, the structure of molecules to a given activity or property. The strategy for developing such models, respecting the five rules set up by the OECD (Organization for Economic Cooperation and Development) for the validation of QSAR/QSPR models (see below: OECD principles for the validity of QSAR models /QSPR), follows these steps:

- Constitute the structure – activity (or property) database from quantitative, reliable and standardized measurements of the target activity (or property), for each compound, and select molecular descriptors in relation to the activity (or property). target in order to digitally translate the structure of molecules;
- Divide this data set into a training set and a test set.
- Build models from learning games using statistical methods.
- Characterize these models by their statistical indices and by internal validation.

- Validate the models with the test set and calculate their external correlation index.
- Repeat the division operation to obtain other training and test sets, and repeat the same steps (optional).
- Define the area of applicability of the proposed models in order to avoid risky extrapolations.
- Explore and exploit validated models to understand possible mechanisms and make activity/property predictions for new molecules, if possible.

II.3. OECD principles of validity of QSAR/QSPR models

In order to support the development of alternative methods, rules have recently been put in place by the OECD (Organization for Economic Co-operation and Development) for the validation of QSAR/QSPR models [8, 9]. The evaluation of each of the five principles is an important condition in order to propose models applicable in the experimental plan.

According to the OECD, model validation is based on five main principles [10]:

- A defined effect: the database (the targeted activity/property) must be reliable and defined with an identified experimental protocol.
- An unambiguous algorithm: the algorithm on which the model is based must guarantee the transparency and reproducibility of the calculation. Forecasts from a model using an algorithm which cannot verify its operation and whose forecasts cannot be reproduced can hardly be accepted. In particular, caution should be exercised when non-transparent and difficult-to-reproduce methods have been used to develop the RQSA/RQSP model.
- A defined domain of applicability: the domain of applicability and the limitations of the model must be described to allow the evaluation of the chemical space in which predictions can be made with confidence. The most used methods to describe the domain of applicability consist of taking into account the interval of individual descriptors and the presence of structural fragments in the training set. Forecasts from a model containing no information on the applicability domain cannot be accepted.
- Appropriate measures of the degree of fit, robustness and predictability: this principle reflects the need for statistical validation of the model. Statistics relating to internal validation (goodness of fit and robustness) and external validation (predictability) must be available. For example, statistics for the regression model can be reported using the correlation coefficient, cross-validation correlation coefficient, mean squared error of the model, etc. ; the external validation must have been carried out as part of a prediction of compounds from an external set (test set). Statistics relating to external validation make it possible to estimate the uncertainty associated with the forecasts.

A mechanistic interpretation, when possible: a justification of the causal link between the molecular descriptors used in the model and the predicted effect reinforces the reliability of the predictions.

II.4. Molecular descriptors

II. 4.1. Introduction

A molecular descriptor is a parameter (a numerical value) specific to a given chemical structure.

These values can be obtained experimentally or calculated from the structure of the molecule. The calculated descriptors make it possible to make predictions without having to synthesize the molecules, which is one of the objectives of molecular modeling.

Molecular descriptors play a fundamental role in studies of the quantitative structure activity/property relationship. They are used as independent variables to predict a dependent variable (activity or property).

The use of molecular descriptors in the development of QSAR/QSPR models is not an easy task. First of all, a very large number of molecular descriptors, of different complexities and designs, have been introduced in recent years. Then, during this time, no strict rules have been established for the selection of suitable descriptors from the large number of available descriptors. This choice has often been based on the chemical intuition of researchers, or by bowing to tradition [11].

II.4.2. Types of descriptors

The importance of the number of descriptors (more than 6000 descriptors listed [12]) that can describe a molecule makes any classification or presentation of these descriptors non-exhaustive.

We will present the most used molecular descriptors and those that have been used throughout our work, starting with the simplest descriptors, which require little knowledge of molecular structure. We will see then how advances in molecular modeling have made it possible to access the 3D structure of the molecule, and to calculate descriptors from this structure.

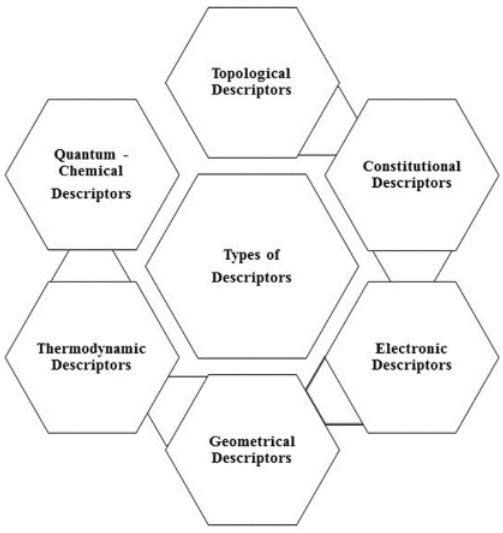


Figure II.2. Types of descriptors

Historically, two main schemes for the classification of molecular descriptors have been established: one based on their origin (constitutional, topological, geometric, quantum, thermodynamic,

etc.), and another based on their dimensionality (1D, 2D, 3D or 4D) [13].

II.4.2.1. 1-D descriptors

These descriptors are calculated from the crude formula of the molecule using the molecular composition, that is to say the atoms which constitute it, and they represent general properties such as: the mass percentages of the atoms, molar mass, molecular weight.

In our work we used:

- The molecular weight, noted MW (also called the formula weight), measured in Daltons (Da). It is the sum of the atomic weights of the different atoms constituting the molecule. It is used in the study of transport including diffusion and mode of operation. Compounds with higher weights are less likely to be absorbed and therefore cannot reach the site of action. Thus, trying to keep molecular weights as low as possible should be the goal for establishing a drug [14]. For drugs delivered orally, the molecular weight must be less than or equal to 500 Daltons (optimum around 300 Daltons) [15].
- The mass percentage, defined by the following formula:

$$\%\ \text{mass} = \frac{\text{the mass of the element in one mole of the compound}}{\text{the mass of one mole of the compound}} * 100$$

The 1D descriptors are easy to calculate, their values are precise, essential and regularly intervene in the RQSA/RQSP models, but they do not make it possible to distinguish the constitutional isomers and do not allow the development of more complex models, it is that is, if we develop models with this type of descriptors only, we will have problems in interpreting the interaction mechanisms involved for the activity or property studied [16]. However, for the vast majority of properties, the position of a substituent modifies its value; 1D descriptors are, in such cases, faulty. It is then necessary to use other classes of descriptors.

II.4.2.2. 2-D descriptors

2D descriptors are obtained from the planar structure of the molecule. In this category we mainly find topological descriptors.

Topological descriptors (or topological indices) describe the atomic connectivities in the molecule. These are more “sophisticated” descriptors that do not necessarily have an obvious chemical meaning but they contain within them information about the overall size of the system, its overall shape and its ramifications [11]. The principle is to find a different value for each molecular skeleton.

These descriptors are easy to calculate, their values are generally precise, they are often used in models. They come from graph theory developed by Euler in 1736 [17]; this theory is applied to the connectivity table, which is a compact representation of the interatomic connectivity within the molecule.

A graph is a set of points, some connected by lines; it makes it possible to represent the topology of the molecule without worrying about the exact spatial geometry of the latter [18].

This theory is invented based on a few simple laws: two adjacent points are connected by lines and two lines with a common point are adjacent. The order or degree of a point is the number of lines connected to that point. A step is a sequence of points and adjacent lines starting with a point and ending with a point. A path is a step in which no point is used more than once. From a molecular point of view, a dot represents an atom and a line represents a covalent bond. The paths are characteristic of the architecture of all the atoms constituting the molecule. Hydrogen atoms are excluded from the graph to simplify the calculations.

The area of the polar surface [19], denoted (PSA), in (Å²), is a very useful parameter for predicting drug transport properties. It is defined as the sum of the surface areas of polar atoms (usually oxygen, nitrogen, sulfur, chlorine and hydrogen attached) in a molecule.

II.4.2.3. 3-D descriptors

This type of descriptor requires a 3D conformation of the molecule; They are evaluated from the relative positions of their atoms in space and describe more complex characteristics; their calculations therefore require knowing, most often by "empirical molecular modeling" or "ab-initio", the 3D geometry of the molecule. Most of these descriptors turn out to be relatively expensive in terms of calculation time, but provide more information and are necessary for modeling properties or activities that depend on the 3D structure. There are several families of 3D descriptors.

II.4.2.3.1 Geometric descriptors

These descriptors can be obtained experimentally or by molecular, empirical or ab-initio modeling. They are based on the spatial arrangement of the atoms constituting the molecule and are defined by the coordinates of the atomic nuclei and the size of the molecule represented. These descriptors include information on the molecular surface obtained by Van Der Waals areas and their superposition [20]. Molecular volumes can be obtained by Van Der Waals volumes [21]. Among the most important ones, which we used in our work, we find:

- The molecular volume, noted MV, in cm3, is defined by the following formula:

$$MV = \mathrm{MW}/d$$

With: MW is the molecular weight and d is the density.

The number of rotating bonds: The rotating bond is defined as a bond of a non-cyclic compound, associated with a non-heavy atom (which is not hydrogen). CN (amide) bonds are not considered due to their high rotational energy barrier. The number of rotatable bonds, denoted NROT, is used to identify the flexibility of the molecule, it has been shown to be a descriptor of very good oral bioavailability of drugs, and for a chemical structure to be able to exhibit good inhibitory and

to be similar to drugs, according to Lipinski's rule, the number of rotatable bonds must be less than or

equal to 5 [15].

The Van Der Waals surface, denoted SVDW, is described as resulting from the set of atomic surfaces defined by the Van Der Waals radius of each atom composing the molecule. The larger this surface area, the greater the possibilities for interactions.

- The Van Der Waals volume, noted VVDW, is the volume occupied by the Van Der Waals envelope, these numerical values depend on the calculation method and the atomic Van Der Waals radii (RVDW). These determine the most favorable position of one atom in relation to another, the appropriate distance where the repulsive and attractive potentials of the atoms are balanced. They are particularly used to model how organic molecules "approach" each other

II.4.2.3.2. Physico-chemical descriptors

Physicochemical descriptors, (or physicochemical indices) some of them reflect the molecular composition of the compound (the number and type of atoms and bonds present in the molecule, the number of rings, the H bond donor/acceptor properties, cation, anion, etc.) [22]. Others represent the hydrophilic or lipophilic nature of the molecule generally evaluated from the Octanol/water partition coefficient represented by log P [23]. Among those we used in our work, we find:

- The Octanol/Water partition coefficient: The transport, passage through membranes and pharmacological activity of a molecule can be conditioned by its sharing between a lipid phase and an aqueous phase, i.e. its hydrophilic character. This can be quantified by the Octanol-Water partition coefficient, denoted (log P), which measures the differential solubility of a solute in these two immiscible solvents [24].

This is an important measure for the identification of drug similarity, according to Lipinski's rule, drugs dispensed orally must have log P values greater than or equal to -2 and less than or equal to 5) [15].

- Molecular refractivity, noted (MR), in m3/mol, is the volume of the substance absorbed per mole of this substance [25].

- The refractive index, noted n [25].

- Polarizability, denoted (αe) in (m3), is the ability to deform the electronic cloud of the molecule under the influence of a uniform electric field. It is one of the parameters which reflect the molecular properties linked to hydrophobicity and consequently to biological activities [26-28].

- The density, noted (d), in (kg/m3), is linked to the mass and size of the molecule. It is the ratio of the molecular weight MW to the molecular volume MV:

$$d = \mathrm{MW}/\mathrm{MV}$$

-The number of hydrogen bond acceptors, denoted (NHA), calculates the number of hydrogen bond acceptors in the molecule. This is the number of atoms having non-bonding doublets (nitrogen, oxygen or fluorine) and capable of forming hydrogen bonds with other molecules.

According to “Lipinski's rule of five” [25], when identifying drug similarity, orally delivered drugs

should have a number of hydrogen bond acceptors (NHA) less than or equal to 10 and a number of hydrogen bond donors (NHD) less than or equal to 5 (optimum of 2).

-The number of hydrogen bond donors, denoted (NHD), calculates the number of hydrogen bond donors in the molecule. This is the number of atoms having an empty quantum cell and containing an acidic hydrogen, that is to say a hydrogen atom linked to a heteroatom

-The tension surface, denoted γ, in dyne/cm [29].

II.4.2.3.3. Quantum/electronic descriptors

These descriptors delineate the charge distribution of molecules (molecular polarity), while also encompassing parameters of quantum chemistry that necessitate more intricate calculations for reliable results. Quantum chemistry methodologies provide access to supplementary information, including structural, energetic, electronic, and spectroscopic data of the systems under scrutiny.m$ù

Electronic properties will be obtained at the end of the calculations, among these properties:

- The dipole moment, symbolized as μ and quantified in debye (D), gauges the over all molecular polarity, elucidating the extent of charge segregation within a molecule where electron density is unevenly distributed among its atoms. The presence of a dipole moment in a molecule arises from variations in electronegativity among its constituent atoms, with greater electron density near the most electronegative atom. Consequently, an asymmetry emerges in the distribution of bonding electrons. Hence, a molecule's dipole moment correlates with the degree of asymmetry within it.

-Total energy , for an isolated molecule in the ground state, the calculated total energy, denoted Et, measured in eV, can be used as a quantum molecular descriptor. This approximate energy was calculated for an optimized conformation of the most stable geometry whose energy structure is minimal.

-The energies of frontier orbitals, play a major role in many chemical reactions and in reaction mechanisms. The energies of these orbitals are very popular parameters in quantum chemistry and in QSAR/RQSP studies.

-The LUMO energy, denoted ELUMO, measured in eV, is the lowest energy level in the molecule which does not contain electrons, it is directly linked to the electron affinity. When a molecule acts as a Lewis acid (an acceptor electron doublet) in forming bonds, incoming electron doublets are received into this orbital. It measures the electrophilicity of a molecule and characterizes the susceptibility of the molecule to attack by nucleophiles [30].

-The HOMO energy, denoted EHOMO, measured in eV, is the highest energy level in the molecule which contains electrons, it is directly linked to the ionization potential. When a molecule acts as a Lewis base (a donor electron pair) in forming a bond, electrons are supplied from that orbital. It measures the nucleophilicity of a molecule and characterizes the susceptibility of the molecule to attack

by electrophiles [30].

-The energy gap, or the HOMO-LUMO gap, noted Egap, measured in eV, reflects the energy between the highest occupied molecular orbital and the lowest vacant. This is an important stability index. This energy difference serves as a measure of the excitability of a molecule. Thus, the lower the energy interval, the more the molecule will be able to interact with the environment.

A large HOMO-LUMO gap implies high stability for the molecule in the sense of its low reactivity in chemical reactions, and similarly, a small gap implies high reactivity of the molecule. The HOMO-LUMO gap has also been used as an approximation of the lowest excitation energy of the molecule [31].

- Hardness and softness, denoted η, and its inverse softness, denoted S, can be obtained.

- Electronegativity, denoted χ, measured in eV, is the opposite of the chemical potential which measures the tendency of the electron cloud to escape from the molecule.

- The electrophilicity index, noted ω, used to characterize the capacity of a molecule to generate an electron transfer.

II.4.2.3.4. Thermodynamic descriptors

These are descriptors little used in RQSA/RQSP studies. They can be expressed by the partition function Q of the molecule used in statistical thermodynamics as well as its derivatives [32-34]. This function describes the way in which the energy of a system of molecules is distributed among the molecular individuals. Its value depends on molecular weight, temperature, molecular volume, internuclear distances, molecular movements and intermolecular forces. The partition function is the most convenient point between the microscopic properties of independent molecules (energy levels, moments of inertia) with the macroscopic properties (The heat of formation (enthalpy), melting point, boiling point, entropy….).

II.4.2.4. 4-D descriptors

They correspond to the measurement of the 3D properties (electrostatic potential, hydrophobicity, hydrogen bonding, etc.) of a molecule at any point in space. They provide information on the structure of the target (protein). We will thus be able to distinguish the 4D descriptors which require an alignment of the molecule guided by the study of the ligand-target complexes (or, at least, by constraints aimed at optimizing the spatial overlap of the electric and steric fields of the ligands, in the absence of information on the true mode of fixation in the target) before being calculated. These descriptors are obtained by calculating the molecular interaction fields (CoMFA, CoMSIA) between a molecule and a probe represented by another molecule (amide, water etc.)

II.5. STATISTICAL METHODS

II.5.1. Definition

Statistics, is the science that deals with the collection, processing, and analysis of data obtained from the observation of phenomena influenced by chance or random factors.

Hence, the primary objective of statistics is to effectively manage uncertainty and extract valuable information from data by analyzing variations within observations. Data analysis serves the purpose of describing, comprehending, and managing the phenomena under study, as well as making predictions and informed decisions.

II.5.2. Application areas

Statistics finds utility in a wide range of disciplines due to its applicability to diverse types of data. It is extensively used in disciplines such as agronomy, medicine, economics, biology, sociology, psychology, geology, chemistry, physics, engineering sciences, information sciences, and communication, among others.

II.5.3. Statistical methods

The practice of statistics involves the study of a collection of similar entities, where we observe specific characteristics referred to as "variables." In our context, the entities or individuals are the molecules under investigation, and the variables correspond to the molecular descriptors discussed earlier in this chapter.

Once the descriptors have been collected, the statistical approach involves processing and interpreting the gathered information regarding these molecules. This approach can be broadly categorized into two main classes: descriptive statistics and inferential statistics, which is further divided into decision-making or predictive statistics.

II.5.3.1. Descriptive statistics

Descriptive statistics (also called data analysis) aims to extract the maximum of the information contained in the data in an efficient, simple and understandable way. It makes it possible to summarize the essential characteristics of the phenomen on studied and to suggest hypotheses for more so phisticated studies. For this, it uses data representations in the form of graphs, tables and statistical indicators. It is also used to divide and classify data into homogeneous classes.

II.5.3.2. Decision-making or predictive statistics

In contrast to descriptive statistics, decision-making or predictive statistics heavily rely on probabilities. The aim of this type of statistics is to make decisions and predictions based on observations. Typically, this involves proposing probabilistic models of the random phenomenon under study and effectively managing the risk of error. In our case, the objective is to establish an approximate relationship between an activity or property and multiple quantitative variables (molecular descriptors).

This relationship can take the form of a linear or non-linear function.

In our work, we have used multiple linear regression (MLR), and artificial neural networks (ANN) to construct QSAR/ QSPR models consistently.

II.5.3.2.1. Multiple Linear Regressio (MLR)

Multiple Linear Regression (MLR) is a widely used modeling method known for its simplicity and interpretability. One of its key advantages is its transparency, as the algorithm is readily available and predictions can be easily made. In much of our work, MLR was also employed for the selection of molecular descriptors used in other statistical methods [35].

There are multiple types of multiple linear regression (MLR), with the most commonly used approches being :

- Progressive ascending MLR : In this method, variables are added to the model step by step. At each stage, the variable with the highest partial correlation with the target variable is selected and included in the model. The process continues until a valid model with the desired correlation is obtained (refer to the validation section).
- Progressive descending MLR : In contrast to progressive ascending MLR, this approach starts with all the variables included in the model and then eliminates them one by one. Variables are removed based on their contribution to the model until the best set of components is obtained, resulting in a valid model with the desired correlation.

Stepwise MLR combines elements from both the progressive ascending and progressive descending methods mentioned earlier. In this approach, variables are added to the model one by one through progressive selection. However, at each step, the significance of the partial correlations of the previously included variables is checked.

II. 5.3.2.2. Multiple nonlinear regression (MNLR)

In the context of multiple nonlinear regression (MNLR), various nonlinear methods such as exponential, logarithmic, polynomial, etc., are utilized to establish a mathematical model that effectively captures the nonlinear relationship between a property or activity and molecular descriptors. In our study, we specifically employed a polynomial model based on the descriptors proposed by the linear model, where the descriptors are raised to the power of 2.

II. 5.3.2.3. The partial least squares regression PLS

PLS regression, which stands for Partial Least Squares regression, is an extension of multiple linear regression. It is employed when there is a high number of descriptors that are strongly correlated [36,37]. This approach combines the principles of PCA (Principal Component Analysis) and multilinear regression. By employing a linear transformation, PLS regression identifies the axes that most effectively represent the data in the given space. This approach enables us to identify the axes that most effectively account for the spread of data points. When data is characterized by n descriptors, Partial

Least Squares (PLS) can identify up to n axes arranged by the variance they capture. This technique involves substituting a predictive data matrix *X*, which has *n*rows and *m*columns, with a new matrix derived from *X*. It also requires that the columns of the resulting matrix *T* are formed as linear combinations of the original variables. In matrix notation, this relationship is expressed as : *T=XW*. Where W (m * k) is a matrix of coefficients defining the linear combinations and *T* is the resulting matrix with columns forming "artificial variables" obtained through linear combinations of the original variables, multiple linear regression is performed on the matrix *T* instead of *X*.

II.5.3.2.4. Artificial Neural Networks (ANN)

Biological Neurons : The human brain comprises an extensive network of nerve cells known as neurons, totaling approximately 100 billion, each with between 1000 to 10,000 synapses or connections [38]. Illustrated in Figure II.3

- The biological neuron is a specialized nerve cell responsible for processing information in the form of electrical signals. It consists of three primary components:
- Dendrites are delicate extensions of the cell body that envelop it in a mesh-like structure, gathering oscillations and information from neighboring nerve cells and conveying them to the cell body.
- The cell body, known as the soma, is responsible for receiving excitations, integrating them, and deciding whether to transmit them further. It houses the nucleus, which sustains the neuron's life functions.
- Axons serve to transmit electrical signals from one neuron's output to another neuron's input. The junction where the axon of one neuron connects with the dendrite of another neuron is termed the synapse [39].

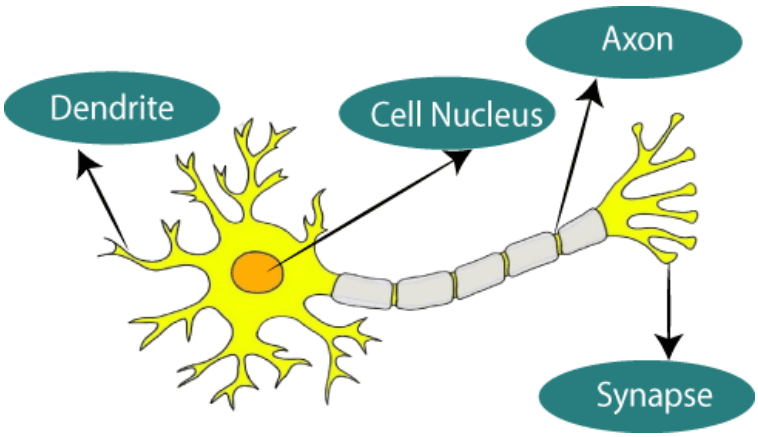


Figure II.3. The biological neuron

At the level of the neuron, an integration (summation) of the signals received occurs and if this sum exceeds a certain threshold the neuron in turn emits an electrical signal to other neurons. This signal can strengthen or decrease the activity of the neurons that receive it depending on whether the synapses are excitatory or inhibitory .

- **Artificial Neural Networks (ANNs)**

History : Neural networks originated as a simplified attempt to mathematically model biological nervous systems. This endeavor began in 1943 with McCulloch and Pitts, who devised the first formal neuron [40]. The first artificial neural network emerged in 1958, credited to Rosenblatt's development of the Perceptron model [41]. This model consisted of a layer of input neurons, known as the perception layer (used for gathering inputs), and a layer of output neurons called the decision layer. Notably, it was the first artificial system capable of learning through experience. In 1960, Widrow and Hoff introduced a model inspired by the perceptron, called the Adaline (Adaptive Linear Element) model [42]. This model subsequently served as the foundation for multilayer neural networks. However, in 1969, Minsky and Papert [43] elucidated the limitations of single-layer neural networks, particularly their inability to address nonlinear problems, as documented in their book "Perceptrons". It wasn't until 1982 that interest in neural networks was reignited, thanks to Hopfield's proposal of associative neurons . Concurrently, Werbos [44] developed the back propagation algorithm, offering a learning mechanism for multilayer Perceptron networks, capable of training neurons within hidden layers. However, it wasn't until after 1986 that this algorithm gained widespread recognition, largely due to Rumelhart [45]. Such networks proved effective in solving nonlinear problems. Additionally, in 1984, the discovery of Kohonen maps [46] introduced an unsupervised algorithm based on self-organization, followed by the unveiling of the Boltzmann machine a year later.

in 1989, Moody and Darken [47] introduced the Radial Basis Function network (RBF), known by its English abbreviation.

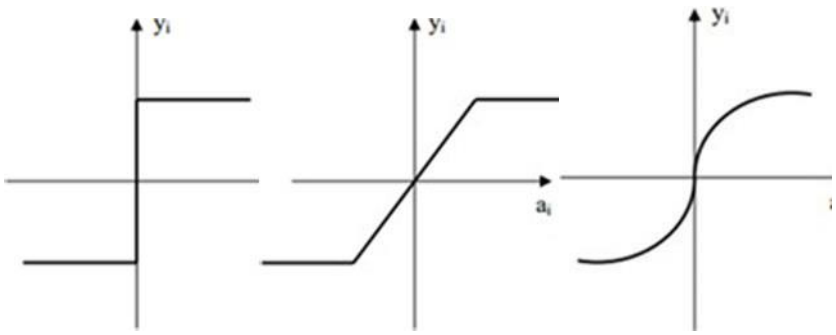
Principle : The ANN methodology mirrors biological neural systems, enabling the processing and transmission of information through the circulation of electrical signals in a network composed of axons. Each artificial neuron serves as a fundamental processor, essentially functioning as a mathematical operator with "inputs" (mathematical function variables) and "outputs" (function values). The significance of neurons lies in the properties that emerge from their integration into networks, specifically from the amalgamation of functions executed by each neuron. A neuron receives a variable number of inputs from up stream neurons or sensors constituting the system it be longs to. Each input is linked to a weight (*wi*), indicating the strength of the connection. Operating as elementary processors, neurons possess a solitary output that subsequently branches out to energize various down stream neurons. An output signal is emitted by the neuron if the weighted sum of the inputs surpasses a certain

threshold.

A neural network consists of several layers: an input layer comprising molecular descriptors, one or more hidden layers, and an output layer representing the properties to be modeled. Neurons within one layer are interconnected with neurons in adjacent layers.

Every neuron in the hidden layer performs weighted summation operations, after which the neuron may or may not be activated. Each neuron in the input layer is linked via synapses to every neuron in the hidden layer, with weights (wi) at these virtual synapses regulating the relative significance of each descriptor. The output layer comprises as many neurons as properties being modeled. In our case, only one property/activity was modeled. During the model's learning phase by a neural network, molecules are sequentially presented to the neurons of the input layer.

The weights (wi) linked to the input neurons undergo iterative adjustments to minimize the discrepancy between the calculated property and the experimental property. Consequently, the output of a neuron is contingent on both its input and its transfer function. Primarily, three types of transfer functions exist: threshold functions, sigmoid functions, and linear functions (refer to *Figure II.4*). The sigmoid function is predominantly favored as it strikes a balance between threshold and linear functions, rendering it widely employed.



The threshold function The linear function The sigmoid function

Figure II.4. Types of transfer functions for the artificial neuron.

Two types of neural networks exist: feed forward networks and recurrent networks. Here, we'll focus solely on feed forward networks. Feed forward neural networks execute one or more algebraic functions of their inputs by composing the functions performed by each neuron. Essentially, they comprise neurons inter connected such that information flows from inputs to outputs without any feedback loops. These networks are commonly referred to as multi layer perceptrons, owing to the inclusion of hidden neurons (*Figure II.5*).

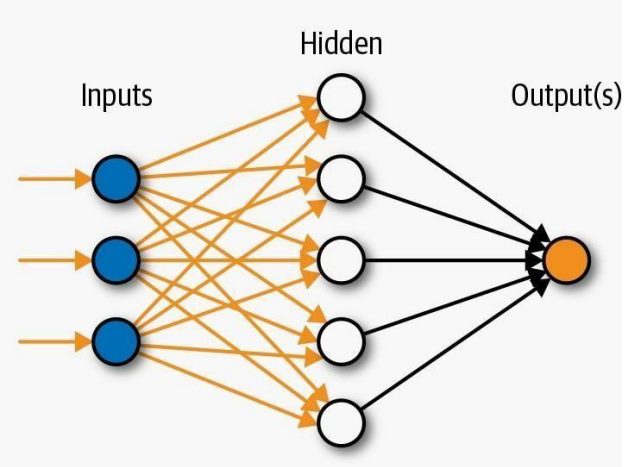


Figure II.5. Topology of a neural network with n inputs and a single output.

- Learning artificial neural networks ANNs

Learning is a critical stage in neural network development, defining the process for establishing the network's structure and parameters. A fundamental attribute of neural networks is their ability to adapt and refine performance by adjusting neuron connections during the learning process [48].

Artificial neural networks are trained using learning algorithms, where the primary objective is to adjust connection weights to align the network's response with the provided experimental examples [48]. Initially, the weights are randomly initialized, and the network is presented with input-output vector pairs from the experiments. Through the application of learning algorithms, the weights are iteratively adjusted to minimize the disparity between the network's calculated outputs (predictions) and the observed experimental outputs.

The database is partitioned into two segments:

- The training set: used for optimizing the weights.
- The test set: employed to assess the network's generalization ability, ensuring that the selected weights yield minimal errors on this dataset

In practice, the initial step involves computing the network weights, which entails estimating the crucial parameters. This necessitates constructing a network that directly links neurons representing the selected molecular descriptors with the output neurons. Subsequently, each descriptor is assigned a weight based on its significance in relation to the property or activity under investigation.

Next, it's essential to determine the architecture of the learning network, which involves selecting the external inputs, determining the number of hidden neurons, and arranging the connections between them. The number of hidden units significantly influences the network's performance. If the count is too low, the network lacks sufficient parameters to capture the dependencies required for modeling and

prediction. Conversely, an excessively high number of hidden neurons may cause the network to overfit the noise present in the training dataset.

Some authors [49, 50] have proposed a parameter ρ, leading to determining the number of hidden neurons, which plays a major role in determining the best architecture of the network (the best number of hidden layers). It is defined as follows

$$p = \frac{\text{Number of data in the training set}}{\text{Sum of number of connections}}$$

Therefore, in order to avoid overfitting or underfitting, it is recommended that the value of ρ be between $1 < p < 3$ [51].

Finally, it is necessary to estimate the quality of the network obtained by presenting it with data that is not part of the learning.

II. 6. VALIDATION METHODS

Assessing the significance of QSAR/QSPR models and their potential in predicting the activities/properties of new compounds necessitates rigorous validation, a critical step in statistical analysis. As models are products of statistical analyses, their interpretation and application should strictly adhere to the domain delineated by the analysis [52]. Applying these models beyond their designated scope requires meticulousness and becomes increasingly precarious the farther one strays from their intended framework. To mitigate errors during validation and application, it's imperative to clearly delineate the model's limitations: verifying its robustness, determining both internal and external predictive power, and constraining the chemical space within which the model is applicable.

II. 6. 1. Model Coefficients and Standard Statistical Tests

Evaluating the efficacy of a model involves employing various statistical parameters, including Mean Square Errors and correlation coefficients, which are commonly utilized in QSAR/QSPR studies. This section provides a detailed description of these statistical metrics.

II. 6. 1.1. Correlation Coefficient r (and Coefficient of Determination r^2)

The correlation coefficient is a widely used statistical measure that as sesses the proportion of variance in the activity/target property explained by the model.

$$r = \sqrt{1 - \frac{\sum(yi - \hat{y}i)2}{\sum(yi - \bar{y})2}}$$

Where: r represents the correlation coefficient; yi and ŷi denote the observed and predicted values of the dependent variable, respectively; $\bar{y}$ represents the average value of the observed values.

These coefficients are independent of the chosen unit of measurement and indicate a strong correlation between the target activity and the initial activity when r^2 approaches 1 (ideal scenario).

However, the assessment of r or r^2 values is highly subjective. While this coefficient is straight forward

to interpret, it should not be overly relied upon as a sole criterion for evaluating regression quality. Using r^2 to compare models with different numbers of descriptors is not recommended, as it tends to favor models with a larger number of descriptors, even if those variables have no effect on the response (the activity or property under study).

The value of r^2 is influenced by the sample size and the number of predictor variables in the equation. It remains constant or increases when a new predictor variable is added to the regression equation, even if the added variable does not contribute to reducing unexplained variance. Therefore, another statistical parameter called adjusted r^2 (r^2adj) can be employed. Additionally, another indicator is the mean square error (MSE), sometimes preferred over the standard deviation s.

II. 6. 1.2. Adjusted Coefficient of Determination

This coefficient is applied in multiple regression analysis as it incorporates the degree of freedom:

$$r^2\text{adj} = \sqrt{\frac{r^2(n-1)-p}{n-p-1}}$$

Where: *n*, represents the number of observations (molecules); *p* denotes the number of independent variables (descriptors); r^2 signifies the coefficient of determination of the model.

II. 6. 1.3. The mean square error “MSE”

$$MSE = \frac{\sum|(\hat{y}_i - y_i)^2|}{n}$$

With : yi and ŷi are, respectively, the observed and calculated values of the dependent variable ; n is the number of observations.

These parameters measure the variation in target activity not explained by the model

RQSA/RQSP. In particular, the smaller the standard deviation, the better the correlation. Its value always depends on the unit of measurement of the target activity and also takes into account experimental errors, which explains why a value that is too small has no meaning.

II. 6. 1.4. Fisher’s F Test

The Fisher F-test index evaluates the level of statistical significance of the model at a given confidence level (typically 95%), indicating the quality of the parameter selection. It's important to note that the conclusion drawn does not imply that the correlation has a "x%" chance of being true, but rather that the correlation holds true for "x%" of the reference compounds, with an assumption made for the others.

Hypotheses:

- *H*0: Sample variances are homogeneous
- *H*1: Sample variances are not homogeneous

The calculated value is determined as follows: We compute the observed F (F (observed) using the formula:

$$F\ (\text{observed}) = \frac{\sum(y-\bar{y})^2 n-p-1}{\sum(yi-\hat{y})^2\ p}$$

With: *F* represents the Fisher index;*yi* and ŷ *i* denote, respectively, the observed and predicted values of the dependent variable; ȳ is the average value of the predicted values; *n* is the number of observations (molecules); *p* is the number of independent variables (descriptors).

Following the computation of the observed *F* (F(observed)), we compare it with the theoretical *F* obtained from standard statistical tables (the Fisher table).

If the observed *F* exceeds the theoretical *F*: rejection of the null hypothesis

*H*0, indicating that the sample variances differ significantly and can not be considered homogeneous.

If the observed *F* is lower than the theoretical *F*: acceptance of the null hypothesis *H*1, suggesting that the two variances are sufficiently similar to be deemed homogeneous.

II. 6. 2. Forecasting power

To assess the predictive reliability of a model and investigate the impact of individual samples (compounds) on the final model, cross-validation procedures are frequently employed. These validation techniques typically enable the assessment of model robustness, indicating the consistency of QSAR/QSPR model parameters concerning the molecules within the training set. However, it's important to note that these techniques do not directly demonstrate the predictive capability of the models [51,52].

The essence of these methods involves selecting a specific number of molecules from the training set and constructing a new model with the remaining molecules, utilizing the designated descriptors (only the regression constants vary). This newly formed model is the nutilized for predicting the values of the removed molecules. This iterative process is repeated until all molecules in the training set have been both removed and predicted. The correlation coefficient .2*q* (or r^2_{cv}) between the calculated and observed activities reflects the internal predictive capability of the model ; fgtthe closer the coefficient's value is to 1, the stronger the predictiveability. For the model to be considered acceptable, the internal predictive capability should exceed 0.5 .

II. 6. 2.1. Leave-Many-Out Cross-Validation

The "k-fold cross-validation" procedure involves partitioning the training dataset into k subsets. One of these subsets is designated as the validation set, while the remaining (k-1) subsets constitute the training set. A QSAR/QSPR model is then constructed using the training set, and the activities/properties of the validation set are predicted. This process is iterated k times, ensuring that each sub set is used exactly once as a validation set. The correlation coefficient 2*q* (or r^2_{cv}) between the predicted and observed activities is then calculated to assess the internal predictive capability of the model.

II. 6. 2.2. Leave-One-Out Cross-Validation

This approach is a specific instance of "k-fold" cross-validation where k equals n, meaning that we train the QSAR/QSPR model using (n-1) observations and then validate it on the nth observation. This process is repeated n times, ensuring that each observation is utilized exactly once as a validation set.

II. 6. 2.3. Randomization Test (Y-Randomization Test) for Validation

The internal predictive power assessed through cross-validation procedures tends to be susceptible to over estimation. A high correlation coefficient value $2q$ (or r^2cv) can arise either from chance correlations or from structural redundancies, particularly when the differences between compounds in the training set are minimal (autocorrelation).

To verify the robustness of a model, the randomization test is often employed [53]. In this validation approach, the values of the target variable are randomly shuffled across the entire training set, and a new model is generated. This process is repeated multiple times, and if the average correlation coefficients obtained remain high, it suggests that no acceptable model can be derived using this statistical method with the given dataset.

II. 6. 3. External Predictive Power

A model exhibiting high values of internal indices such as $2q$ (or r^2cv) is not automatically deemed valid; thus, internal validation alone is necessary but insufficient.

The true predictive capability of a QSAR/QSPR model lies in its ability to accurately fore cast the activity/property of compounds from an external test set (i.e., compounds not utilized for model development). A robust QSAR/QSPR model should not only predict the activities of compounds within the training set but also those of test molecules. The QSAR/QSPR model is constructed using the training set and validated using the test set. The predictive performance of the model is assessed based on the correlation coefficient test $2r$ test 2 between the observed and predicted activities for the test set, with a higher test $2r$ test 2 value (>0.5) indicating superior model performance.

II. 6. 4. Applicability Domain

A QSAR/QSPR model cannot be regarded as universally applicable since it is developed based on a finite number of compounds that fail to encompass the entirety of chemical space. Consequently, predicting the activity/property of a chemically dissimilar compound, not represented in the training set, cannot be deemed reliable [54].

An ideal model would possess the capability to predict the activity or property of any conceivable molecule. However, achieving this is often unattainable. The constrained size of the training set restricts the chemical space covered by the models developed. Therefore, when a molecule falls out side this chemical space, the prediction becomes less reliable.

REFERENCES

[1] Young D.C (2001) Computational chemistry: A practical guide for applying technique store al-world problems. John Wiley&Sons , Inc. New York.

[2] Yee L.C, Wei Y.C(2012) Current modeling methods used in QSAR/QSPR, Statistical modelling of molecular descriptors in QSAR/QSPR. Dehmer In, Varmuza K and Bonchev D (eds).

[3] Ghamali M, Chtita S , Bouachrine M, Lakhlifi T(2016) Méthodologie générale d'une étude RQSA/RQSP. Revue Interdisciplinaire. 1(1).

[4] Fang J, Pang X, Wu P, Yan R, Gao L, Li C, Lian W, Wang Q, Liu A.l, Du G.h (2016) Molecular Modeling on Berberine Derivatives to ward BuChE: An Integrated Study with Quantitative Structure–Activity Relationships Models, Molecular Docking ,and Molecular Dynamics Simulations. Chemical Biology& Drug Design 87 (5) :649-663.

[5] Muhammad U, Uzairu A, Ebuka Arthur D (2018) Review on: quantitative structure activity relationship (QSAR) modeling.J Anal Pharm Res. 7(2):240–242.

[6] Li F, Wu H, Li L, Li X ,Zhao J, Peijnenburg W. J (2012) Docking and QSAR study on the binding interactions between polycyclic aromatic hydrocarbons and estrogen receptor. Ecotoxicology and Environmental Safety 80: 273-279.

[7] Tropsha A, Gramatica P, Gombar V. K (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR&Combinatorial Science *22* (1) :69-77.

[8] (Organization for Economic Co-operation and DevelopmentOECD) october 20 th ,2009.

[9] Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods‖. Organization for Economic Co-operation and Development. Paris.2009.

[10] Principles for the Validation, for Regulatory Purposes, of Structure-Activity Relationship Models‖.Organization for Economic Co-operation and Development,Paris,2009.

[11] Karelson M(2000) Molecular descriptors in QSAR/QSPR .Wiley, New York.

[12] Todeschini R,Consonni V ,and Mannhold R (2009) Molecular Descriptors for .Drug Discovery & Development 41(2).

[13] Dudek A.Z, Arodz T and Gàlvez J (2006) Computational methods in developing quantitative structure-activity relationships (QSARs):A review, Combinatorial Chemistry & High Throughput Screening 9 : 213–228.

[14] Rekkab,S (2014) Drug Design et synthèse de nouveaux calix[8]arènessulfoniques flexibles à activités anticorrosive et anti coagulante‖. Thèse de Doctorat,Univ.Constantine,Alger.

[15] Lipinski C.A ,Lombardo F ,Dominy B.W ,and Feeney P.J (1997) Experimental and computational

approaches to estimate solubility and permeability in drug discovery and development settings. Advance Drug Delivery Reviews 6(1–3) :3–25.

[16] Goulon-Sigwalt-Abram A Une nouvelle méthode d'apprentissage de données structurées : applications à l'aide à la découverte de médicaments. Thèse de Doctorat, Université Pierre et Marie Curie(Paris 6), France, 2008.

[17] Euler L (1741) The solution of a problem relating to the geometry of position . Commentarii academiae scientiarum Petropolitanae 8 : 128–140.

[18] Schultz H.P (1989)Topoligical organic chemistry : Graph theory and topological indices of alkanes. Journal of Chemical Information and Modeling 29 :227–228.

[19] Saunders R.A, Platts J.A (2004) Scaled polar suface area descriptors :development and application to three sets of partition coefficients .New Journal of Chemistry 28 :166–172.

[20] Labute P(2000) A widely applicable sef of descriptors3Journal of Molecular Graphics and Modelling 18 : 464–477.

[21] Higo J and.Go N (1989) Algorithme for rapid calculation of excluded volume of large molecules. Journal of Computational Chemistry 10 :376–379.

[22] Bosque R, Sales J, Bosch E, Rosès M, Garcia-Alvarez-Coque M.C, and Torres-Lapasio J.R(2003)A QSPR study of the p-solute polarity parameter to estimate retention in HPLC. Journal of Chemical Information and Modeling 43 : 1240–1247.

[23] Viswanadhan V.N,Ghose A.K ,Revankar G.R and Robins R.K (1989) Atomic physicochemical parameters for 3D structure directed quantitative structure-activity relationships: Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occuring nucleoside antibiotics. Journal of Chemical Information and Modeling 29 : 163–172.

[24] Taylor P.J(1990) Hydrophobic Properties of Drugs, In Quantitative Drug Design. Pergamon Press,Oxford(UK) 4 : 241–294.

[25] H.A. Lorentz (1880) Über dieBeziehungzwischen der Fortpflanzungsgeschwindigkeit des Lichtes der K€orperdichte . Wiedemann's Annalen der Physik 9 :641–665.

[26] Cammarata.A (1967) An Apparent Correlation between the in Vitro Activity of Chloramphenicol Analogs and Electronic Polarizability.Journal of Medicinal Chemistry10 :525–527.

[27] Leo A,Hansch C,and Church C (1969) Comparison of parameters currently used in the study of structure-activity relationships. Journal of Medicinal Chemistry12(5) : 766–771.

[28] Hansch C and Coats E (1970) α-Chymotrypsin:A Case Study of Substituent Constants and Regression Analysis in Enzymic Structure—Activity Relationships. Journal of Pharmaceutical Sciences 59(6) : 731–743.

[29] Sudgen S (1924) The variation of surface tension with temperatureand some related functions.

Journal of the Chemical Society 125 : 32–41.

[30] Atkins P.W and dePaula J (2002) Atkins' Physical Chemistry, 7thed,Oxford University Press, Oxford

[31] Kikuchi O (1987) Systematic QSAR Procedures with Quantum Chemical Descriptors. Molecular Informatics 6(4) :179–184.

[32] McQuarrie D.A(1973) Statistical Thermodynamics .harper row publishers, NewYork.

[33] Akhiezerand A.I, Peltminskii S.V(1981) Methods of Statistical Physics, Pergamon Press, Oxford.

[34] Atkins P.W(1982) Physical Chemistry.2ndedition,W.H.Freeman and Company, San Francisco.

[35] Roy K ,Kar S and Narayan Das R (2015) Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, Chapter 6 – Selected Statistical Methods in QSAR. Academic Press, Boston 191–229.

[36] Puzyn T ,Leszczynski J and Cronin M.T(2010) Recent Advances in QSAR Studies :Methods and Applications: Part I Theory of QSAR. Challenges and Advances in Computational Chemistry and Physics 8.

[37] Tobias R.D (2002) An Introduction to Partial Least Squares Regression .Statistical Analysis System Institute Inc, Cary, USA.

[38] Rolland J, and Blouch P(2002) Les bouées météorologiques : L'exemple de Météo-France. La Météorologie 39 :83–88.

[39] Torres-Moreno J.M (1992) Apprentissage et généralisation par des réseaux de neurones : étude de nouveaux algorithmes constructifs, Thèse de Doctorat, Institut Nationale Polytechnique de Grenoble.

[40] McCulloch W.S and Pitts W(1943)A logical calculus of ideas immanent in nervous activity.The Bulletin of Mathematical Biophysics 5 : 115–133.

[41] Rosenblatt F (1958) The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological *Review* 65 : 386–408.

[42] Widrow B and Hoff M.E (1960) Adaptive Switching Circuits. IRE WESCON Convention Record 4 :96–104.

[43] Minsky M and Papert S (1972) Perceptrons: An Introduction to Computational Geometry. 2ndedition with corrections, first edition 1969.The MIT Press;CambridgeMA.

[44] Werbos P(1974) Beyond Regression: New Tools for Predictionand Analysis in the Behavioral Sciences, *PhD Thesis, Harvard University, Cambridge*.

[45].D.E. Rumelhart D.E, Hinton J.E and Williams R.J (1986) Learning representations by back-propagating errors .Nature 323 :533–536

[46] Kohonen T(1984) Self-Organization and Associative Memory.Springer, Berlin.

[47] Moody J and Darken C (1989) Fast learning in network soflocally-tuned processing units.Neural Computation 1 : 281–294.

[48] Ammar M.Y(2007) Mise en œuvre de réseaux de neurones pour la modélisation de cinétiques réactionnelles en vue de la transposition Batch/Continu. Thèse de Doctorat, Institut Nationale Polytechnique de Toulouse.

[49] So S.S and Richards W.G(1992) Application of neural networks: quantitative structure-activity relationships of the derivatives of 2,4 diamino (substituted-benzyl) pyrimidines as DHFR inhibitors. Journal of Medicinal Chemistry 35(17) : 3201–3207.

[50] Andrea T.A and Kalayeh H(1991) Applications of neural networks in quantitative structure-activity relationships of dihydrofolatereductase inhibitors. Journal of Medicinal Chemistry 34(9) :2824–2836.

[51] Golbraikh A and Tropsha A (2002) Bewareofq^2!. Journal of Molecular Graphics and Modelling 20(4) :269–276.

[52].A.Tropsha A , Gramatica P ,and Gombar V.K (2003) the importance of Being Earnest: Validation is the Absolute Essential for Successful Application and interpretation of QSPR Models. QSAR and Combinatorial Sciences 22(1) :69–77.

[53] Wold S and Eriksson L (1995) Statistical validation of QSAR results. Validation tools. Methods and Principles in Medicinal Chemistry 2 : 309–18.

[54] Tetko I.V, Sushko I, Pandey A.K, Zhu H, Tropsha A, Papa E, Oberg T, Todeschini R, Fourches D, and Varnek A (2008) Critical assessment of QSAR models of environment altoxicity against Tetrahymena-pyriformis: focusing on applicability domain and overfitting by variables election. Journal of Chemical Information and Modelin 48 :1733–1746

CHAPTER III

QSAR and drug-likeness study for biology activity of 1.4 dihydropyridine

III .1. INTRODUCTION

Identifying and discovering new molecules early and reliably for potential drug development remain pivotal main in pharmaceutical research, posing a significant challenge for the foreseeable future. The process is notably complex, influenced by factors including cost, time constraints, and the availability of laboratories equipped for synthesis and testing. [1]

Typically, only one innovative drug emerges on the market for every thousand molecules synthesized and tested. Moreover, drug development typically requires over a decade of research. The primary challenge is to identify a molecule with targeted therapeutic properties while mitigating undesirable side effects. The substantial cost of drug development largely arises from protracted, costly, and frequently unproductive synthesis processes. Consequently, the pharmaceutical industry is transitioning towards innovative research methods that entail predicting the properties and activities of molecules before synthesis. [2] In recent years, advancements in technologies enabling the simultaneous synthesis of a vast array of molecules and their evaluation on therapeutic targets have yielded promising outcomes. This forms the core focus of studies on Quantitative Structure-Activity Relationships (QSAR) [3] and Quantitative Structure-Property Relationships (QSPR). These investigations primarily revolve around identifying resemblances among molecules within extensive databases containing known activities or properties. Uncovering such relationships facilitates the prediction of activities and properties for novel compounds [4], there by steering the synthesis of new molecules without necessitating their physical creation. Relationships between molecular structures and their activities or properties are typically forged through a combination of molecular modeling and statistical methodologies. Common approaches involve characterizing molecules using a series of descriptors, which are real numbers derived from measured or calculated features of molecular structures. Subsequently, these descriptors can be correlated with the modeled quantity to establish a relationship [5]. The continuous progress in heterocyclic chemistry plays a crucial role in organic synthesis, [6], as heterocycles form the essential scaffold in a broad spectrum of compounds of considerable chemical, biological, pharmacological, and industrial importance. Nitrogen-

containing heterocycles such as pyrimidines, pyridines , phenothiazines, indoles, acridines and others are widely distributed in various natural products and hold significant relevance in medicinal chemistry. The exploration for novel nitrogen heterocycles with potential biological activity is essential in the endeavor to create fresh compounds that can meet the rising demand for innovative molecules. [7]

III.2. MATERIAL AND METHODS

Accurate predictions of molecular geometries are sensitive to the choice of the electronic structure method and the atomic basis set used for the description of the atoms. We started our investigations by selecting a suitable methodology to be used for the determination of the equilibrium structures of 1,4 dihydropyridine derivatives under study. Our strategy consists in performing benchmark computations on the subunit of the series

Chemical structure of each molecule was built and optimized by HyperChem (version 8.08) packages [8], using the AM1 semi-empirical method. GAUSSIAN 16 program [9] was operated to reoptimize the molecular structures. The structures were reoptimized by density functional theory (DFT) at the level of B3LYP /6-31G++ (d, p).

The geometries of 1.4 dihydropyridine, were fully optimized with ab initio/HF (6-31G, 6-31G^{+} (d,p), 6-311G^{++}(d,p)) and DFT/B3LYP(6-31G, 6-31G^{+} (d,p), 6-311G^{++}(d,p)).

integrated in Gaussian 16 program package. The calculation of QSAR properties is performed through the module QSAR Properties (HyperChem version 8.08), and allows the calculation of several properties commonly used in QSAR studies.

Molinspiration [10], was used to obtain parameter such as TPSA (topological polar surface area), nrotb (number of rotatable bonds) and drug likeness. The calculated results are reported in the present work.

- GEOMETRIC AND ELECTRONIC STRUCTURE OF 1,4 DIHYDROPYRIDINE

The optimized geometrical parameters of 4-Imidazolyl- 1,4-dihydropyridines (Figure.III.1) are obtained using ab-initio/*HF* and *DFT* methods, listed in (Table.III.1) and (Table III.2)

We found good agreement between predicted geometries (bond lengths, angles) and corresponding experimental data, especially the DFT/B3LYP results. From that, we can say the DFT method is more appropriate for our next study part. Charge densities calculated by DFT/B3LYP are almost similar to Ab initio/HF methods.

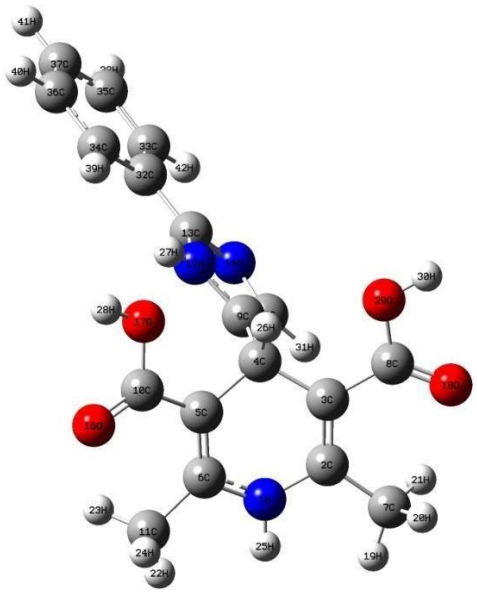


Figure III.1. 3D conformation of 4-Imidazolyl- 1,4-dihydropyridines(Gauss View 3.0.9)

Table III. 1. Calculated bond lengths and angles of 4-Imidazolyl- 1,4-dihydropyridines .

Bond lengths are in Å and bond angles are in degrees.

	DFT/B3LYP			***Ab initio/HF***		
	6-31G	**31G+ (d,p)**	**31G++ (d,p)**	**6-31G**	**31G+ (d,p)**	**31G++ (d,p)**
N1-C2	1.39	1.38	1.38	1.38	1.38	1.38
C2-C3	1.36	1.36	1.36	1.34	1.34	1.34
C3-C4	1.52	1.52	1.52	1.52	1.52	1.52
C4-C5	1.53	1.53	1.53	1.52	1.52	1.52
C5-C6	1.37	1.36	1.36	1.35	1.34	1.34

C6-N1	1.38	1.38	1.38	1.37	1.37	1.37
C2-C7	1.50	1.50	1.50	1.50	1.50	1.50
C3-C8	1.45	1.46	1.46	1.45	1.47	1.47
C4-C9	1.51	1.51	1.51	1.51	1.51	1.51
C5-C10	1.45	1.46	1.46	1.45	1.46	1.46
C6-C11	1.50	1.50	1.50	1.50	1.50	1.50
C9-N12	1.38	1.38	1.38	1.37	1.37	1.37
N12-C13	1.38	1.37	1.37	1.36	1.35	1.35
C13-N14	1.34	1.32	1.32	1.30	1.29	1.29
N14-C15	1.38	1.37	1.37	1.38	1.37	1.37
C10-O16	1.24	1.22	1.22	1.22	1.19	1.19
C10-O17	1.40	1.38	1.38	1.36	1.34	1.34
C8-O18	1.24	1.22	1.22	1.22	1.19	1.19
C8-O29	1.38	1.36	1.36	1.35	1.33	1.33
O29-H30	0.98	0.97	0.97	0.95	0.94	0.94
O17-H28	0.98	0.97	0.97	0.95	0.94	0.94
N1-H25	1.00	1.01	1.01	0.99	0.99	0.99
C4-H26	1.09	1.09	1.09	1.08	1.08	1.08
N12-H27	1.01	1.01	1.01	0.99	0.99	0.99
C15-H31	1.07	1.08	1.08	1.06	1.07	1.07
C7-H19	1.09	1.09	1.09	1.08	1.08	1.08
C7-H20	1.09	1.09	1.09	1.08	1.08	1.08
C7-H21	1.09	1.09	1.09	1.07	1.07	1.07
C11-H22	1.09	1.09	1.09	1.08	1.08	1.08
C11-H23	1.09	1.09	1.09	1.07	1.07	1.07
C11-H24	1.09	1.09	1.09	1.08	1.08	1.08
C13-C32	1.46	1.46	1.46	1.46	1.47	1.47
C32-C33	1.41	1.40	1.40	1.39	1.39	1.39
C32-C34	1.40	1.40	1.40	1.39	1.39	1.39

C33-C35	1.39	1.39	1.39	1.38	1.38	1.38
C34-C36	1.39	1.39	1.39	1.38	1.38	1.38
C35-C37	1.40	1.40	1.40	1.39	1.39	1.39
C36-C37	1.39	1.39	1.39	1.38	1.38	1.38
C33-H42	1.08	1.08	1.08	1.07	1.07	1.07
C34-H39	1.08	1.08	1.08	1.07	1.07	1.07
C35-H38	1.08	1.08	1.08	1.07	1.07	1.07
C36-H40	1.08	1.08	1.08	1.07	1.07	1.07
C37-H41	1.08	1.08	1.08	1.07	1.07	1.07
N1-C2-C3	118.53	118.34	118.35	118.70	118.67	118.67
C2-C3-C4	120.32	119.87	119.87	120.46	120.03	120.03
C3-C4-C5	110.48	110.25	110.96	110.14	109.81	109.81
C4-C5-C6	120.05	119.70	119.70	120.17	119.81	119.81
C5-C6-N1	118.51	118.26	118.27	118.73	118.64	118.64
C6-N1-C2	124.22	124.12	124.11	123.89	123.65	123.25
N1-C2-C7	114.75	114.75	114.58	114.44	114.09	114.09
C7-C2-C3	126.65	127.03	127.02	126.81	127.22	127.22
N1-C6-C11	114.79	114.66	114.66	114.29	113.97	113.95
C2-C3-C8	120.53	120.43	120.44	120.60	120.43	120.43
C6-C5-C10	120.02	120.03	120.02	120.07	120.02	120.03
C3-C4-C9	110.40	110.70	110.69	110.39	110.69	110.69
C5-C4-C9	110.58	111.76	111.77	111.61	112.01	112.04
C3-C4-H26	108.38	108.14	108.14	108.38	108.30	108.28
C5-C4-H26	108.13	108.26	108.26	108.58	108.40	108.39
C5-C10-O16	128.24	127.61	127.61	127.55	127.05	127.06
C5-C10-O17	113.02	112.76	112.77	113.69	113.14	113.14
O16-C10-O17	118.73	119.61	119.61	118.74	119.80	119.79
C10-O17-H28	108.01	105.36	105.35	111.78	107.39	107.37

C3-C8-O18	127.33	126.73	126.74	126.85	126.25	126.26
C3-C8-O29	112.53	112.56	112.56	113.21	112.96	112.96
C8-O29-H30	109.03	106.09	106.09	112.83	108.03	108.03
C4-C9-C15	133.23	133.34	133.34	133.06	133.26	133.26
C4-C9-N12	121.71	122.13	122.14	121.81	122.10	122.11
C9-N12-C13	108.34	108.24	108.24	108.22	107.75	107.74
N12-C13-N14	109.86	110.30	110.30	109.93	111.05	111.06
C13-N14-C15	106.25	106.03	106.03	106.83	105.97	105.96
N14-C15-C9	110.48	110.89	110.89	109.89	110.59	110.59
C9-N12-H27	122.71	123. 66	123.66	123.62	124.79	124.80
C9-C15-H31	128.25	127.73	127.73	128.85	128.17	128.16
C13-N12-H27	128.85	127.95	127.95	128.15	127.35	127.34
N14-C15-H31	121.25	121.36	121.36	121.25	121.24	121.24
N12-C13-C32	124.95	124.31	124.30	124.79	132.72	123.69
N14-C13-C32	125.17	125.38	125.39	125.27	125.22	125.24
C13-C32-C33	118.87	119.07	119.08	118.92	119.10	119.12
C13-C32-C34	122.40	122.33	122.32	122.19	121.93	121.91
C32-C33-C35	120.47	120.53	120.52	120.41	120.36	120.36
C32-C34-C36	120.61	120.71	120.71	120.63	120.59	120.59
C33-C35-C37	120.44	120.49	120.49	120.37	120.38	120.38
C34-C36-C37	120.26	120.27	120.26	120.17	120.13	120.12
C35-C37-C36	119.48	119.40	120.26	119.56	119.57	119.57
C32-C33-H42	118.21	118.48	118.48	118.60	118.89	118.89
C32-C34-H39	120.43	120.60	120.60	120.58	120.55	120.53

Table III.2. NBO charges distribution of 4-Imidazolyl- 1,4-dihydropyridines

	Ab initio/HF			**DFT/B3LYP**		
	6-31G	**6-31G+(d,p)**	**6-31G++(d,p)**	**6-31G**	**6-31G+(d,p)**	**6-31G++(d,p)**
N 1	-0.667	-0.685	-0.680	-0.569	-0.585	-0.581
C 2	0.354	0.351	0.354	0.261	0.262	0.265
C 3	-0.269	-0.282	-0.279	-0.188	-0.201	-0.199
C 4	-0.279	-0.267	-0.255	-0.317	-0.313	-0.300
C 5	-0.280	-0.295	-0.294	-0.198	-0.214	-0.213
C 6	0.370	0.369	0.371	0.275	0.278	0.280
C7	-0.709	-0.679	-0.635	-0.732	-0.726	-0.689
C8	0.933	0.973	0.971	0.755	0.788	0.786
C9	0.148	0.136	0.137	0.126	0.112	0.116
C10	0.928	0.969	0.967	0.755	0.783	0.782
C11	-0.710	-0.679	-0.635	-0.732	-0.727	-0.689
N12	-0.629	-0.632	-0.630	-0.552	-0.563	-0.556
C13	0.419	0.436	0.451	0.347	0.366	0.347
N14	-0.539	-0.557	-0.555	-0.479	-0.492	-0.489
C15	-0.108	-0.083	-0.077	-0.109	-0.093	-0.089
O16	-0.708	-0.730	-0.732	-0.600	-0.627	-0.629
O17	-0.811	-0.807	-0.801	-0.734	-0.758	-0.752
O18	-0.714	-0.736	-0.738	-0.607	-0.634	-0.637
H19	0.224	0.215	0.203	0.232	0.233	0.223
H20	0.279	0.268	0.250	0.279	0.282	0.268
H21	0.283	0.274	0.259	0.288	0.290	0.276
H22	0.227	0.218	0.206	0.234	0.236	0.225
H23	0.283	0.274	0.259	0.286	0.289	0.274
H24	0.276	0.264	0.247	0.278	0.281	0.267
H25	0.440	0.437	0.531	0.432	0.438	0.433

H26	0.303	0.285	0.270	0.308	0.304	0.293
H27	0.464	0.456	0.452	0.452	0.459	0.450
H28	0.524	0.532	0.527	0.229	0.229	0.522
O29	-0.788	- 0.784	- 0.778	-0.704	-0.729	-0.723
H30	0.523	0.529	-0.525	0.499	0.524	0.519
H31	0.238	0.229	0.220	0.239	0.238	0.231
C32	-0.091	-0.077	-0.097	-0.088	-0.097	-0.080
C33	-0.194	-0.197	-0.185	-0.205	-0.205	-0.189
C34	-0.224	-0.225	-0.215	-0.233	-0.231	-0.230
C35	-0.236	-0.231	-0.222	-0.238	-0.240	-0.238
C36	-0.237	-0.234	-0.224	-0.238	-0.241	-0.233
C37	-0.236	-0.233	-0.224	-0.244	-0.247	-0.238
H38	0.242	0.239	-0.230	0.243	0.247	0.239
H39	0.225	0.221	-0.217	0.266	0.229	0.222
H40	0.239	0.237	0.228	0.244	0.245	0.237
H41	0.240	0.237	0.228	0.242	0.245	0.237
H42	0.269	0.263	0.253	0.267	0.268	0.261

Molecular electrostatic potential

The molecular electrostatic potential (MEP) serves as a well-established method for elucidating the reactive characteristics of diverse chemical systems, encompassing both electrophilic and nucleophilic reactions. In this study, the focus was on investigating biological recognition processes and hydrogen bonding interactions to anticipate the reactive sites susceptible to electrophilic and nucleophilic attack within the analyzed molecule [11]. The MEP calculations were conducted based on the DFT optimized geometry (see Figure III.2).

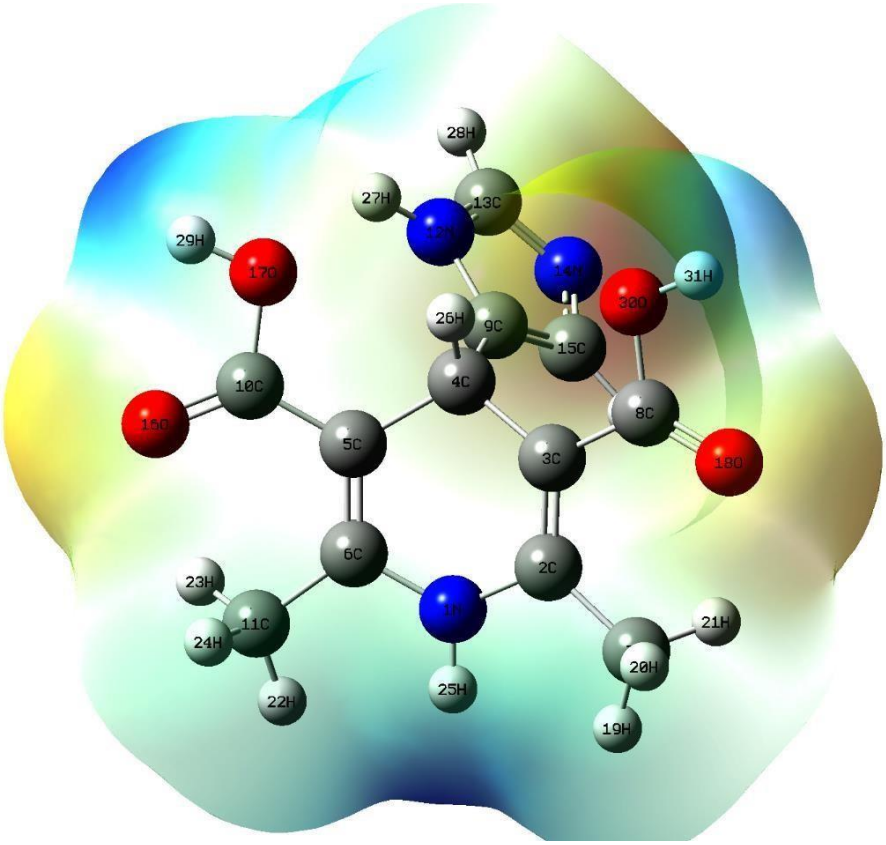


Figure III .2. 3D molecular electrostatic potential surface map (3D MESP) for 1,4-dihydropyridines .

The electrostatic potential values are depicted using various colors, illustrating the positive, negative, and neutral regions of molecules through a gradient of colors. Typically, the potential ascends in the sequence: red < orange < yellow < green < blue. Red signifies the most negative region, while blue indicates the most positive region [12].

For the 4-Imidazolyl- 1,4-dihydropyridines molecule, we carried out NBO population analysis and we mapped its 3D molecular electrostatic potential surface (MESP). The corresponding data are listed in Table III .2 and displayed in Figure III .2. This figure shows that this MESP exhibits one region characterized by red color (negative electrostatic

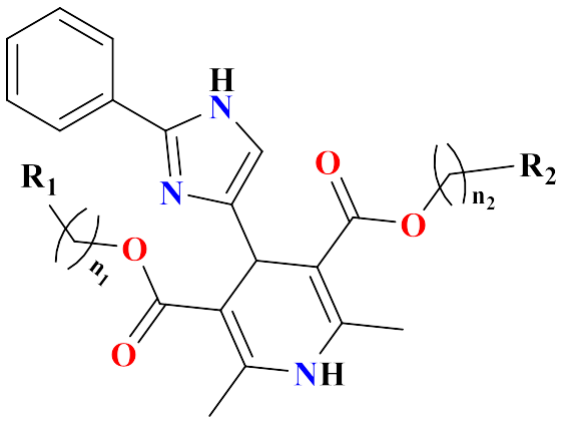
potential) around N1 nitrogen atom. This indicates area of excess negative charge enabling electrophilic attacks on this position. Whereas a blue color (positive electrostatic potential)

can be seen around all the other atoms of 1,4-dihydropyridines indicating an electron deficiency. Thus, these correspond to regions susceptible of a nucleophilic attack. The NBO analysis [13] confirms these findings as can be seen in Table III. 2. This table shows however that DFT charges exhibit some differences with computed using HF.

III.3. QSAR STUDY OF 1,4 DIHYDROPYRIDINES DERIVATIVES

III.3.1. Dataset

A series of thirty-one 1,4-dihydropyridine derivatives acting as Calcium Channel blockers was collected from the literature [14]. The chemical structures of these compounds are given in Scheme 1 and Table III .3. This table lists also their biological activity values (IC_{50}) after conversion to logarithmic scale (pIC_{50}= -log IC_{50} and used as dependent variables to develop QSAR models. The structures of the collected molecules and their biological activity values are given in Table III .3. These compounds consist on a substituted 1,4-dihydropyridine by a phenyl imidazolyl and ester groups in positions 3 and 5. The various substituents are detailed in Table III .3.



Scheme 1: *Structures of the series of 1,4-dihydropyridines treated presently. (See Table. III. 3 for the definition of R_1, R_2, n_1 and n_2).*

Table III.3. *Chemical structures of 1,4-dihydropyridinederivatives.*

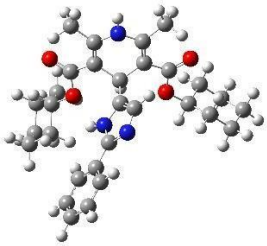
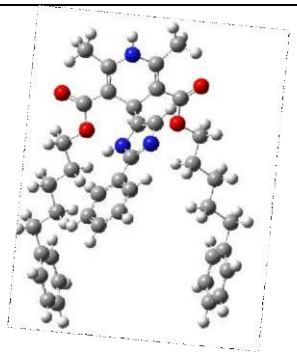
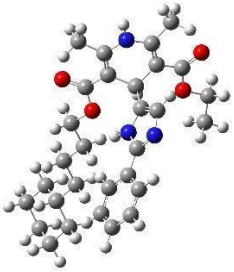
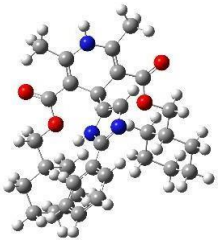
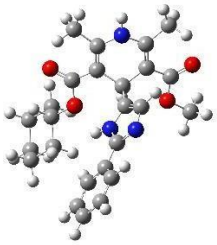
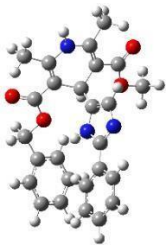
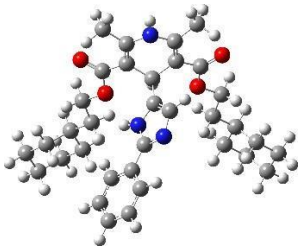
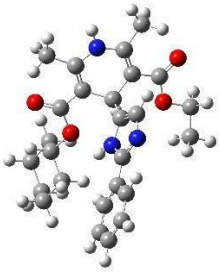
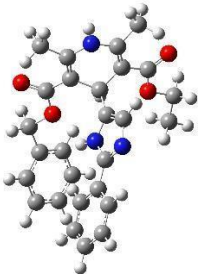
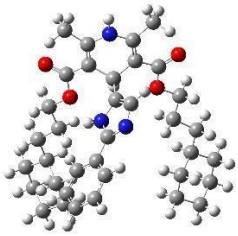
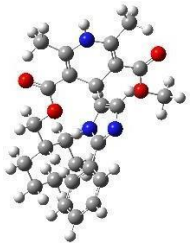
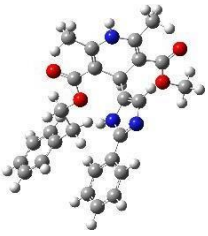
** denotes compounds used for external statistical validation (test set). We quote also their experimental biological activity values (IC_{50} and pIC_{50}= -log IC_{50}) as taken from Ref. [14].*

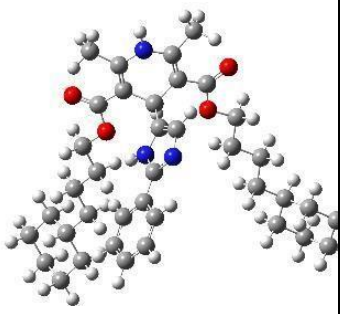
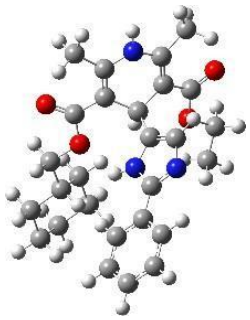
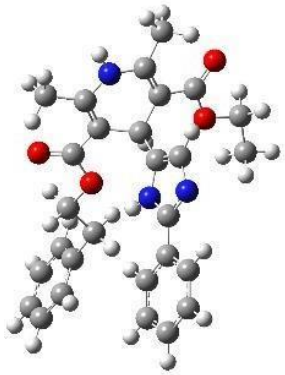
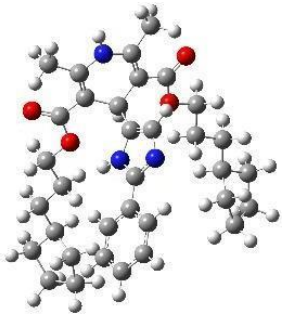
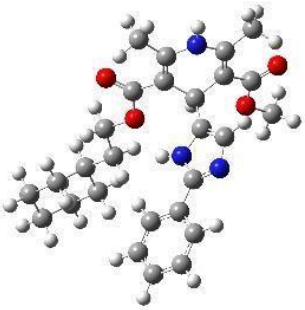
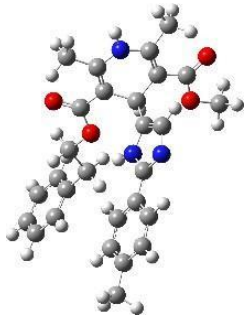
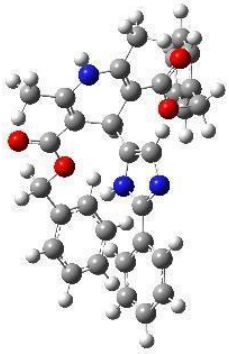
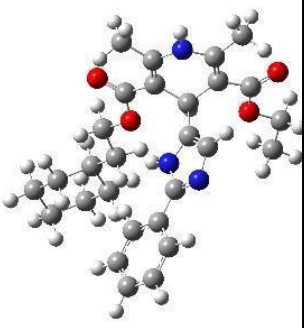
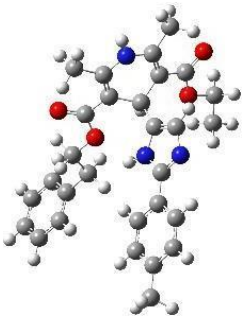
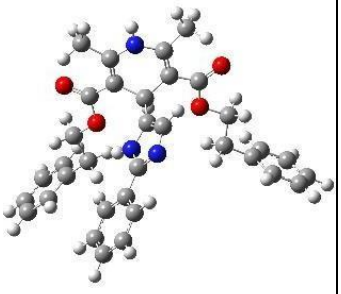
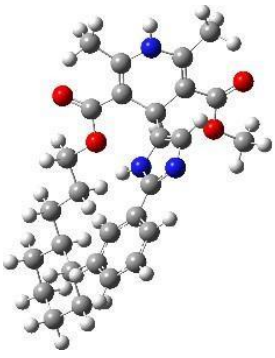
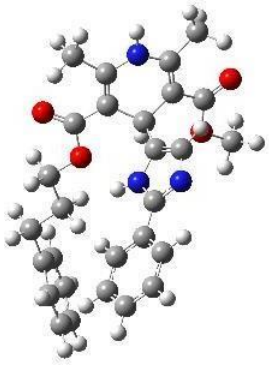
N	R_1	n_1	R_2	n_2	Molecular Formula	IC_{50} mmol/l	pIC_{50}
1*	C_6H_{11} (cyclohexyl)	0	C_6H_{11} (cyclohexyl)	0	$C_{30}H_{37}N_3O_4$	3.02×10^{-10}	9.52
2	C_6H_{11} (cyclohexyl)	1	C_6H_{11} (cyclohexyl)	1	$C_{32}H_{41}N_3O_4$	3.60×10^{-11}	10.44
3	C_6H_{11} (cyclohexyl)	2	C_6H_{11} (cyclohexyl)	2	$C_{34}H_{45}N_3O_4$	1.79×10^{-11}	10.74
4	C_6H_{11} (cyclohexyl)	3	C_6H_{11} (cyclohexyl)	3	$C_{36}H_{49}N_3O_4$	1.20×10^{-11}	10.92
5	C_6H_{11} (cyclohexyl)	4	C_6H_{11} (cyclohexyl)	4	$C_{36}H_{49}N_3O_4$	6.43×10^{-10}	9.19
6	C_5H_9 (cyclopentyl)	3	C_5H_9 (cyclopentyl)	3	$C_{34}H_{45}N_3O_4$	2.79×10^{-9}	8.55
7*	C_6H_5	1	C_6H_5	1	$C_{32}H_{29}N_3O_4$	5.61×10^{-10}	9.25
8*	C_6H_5	2	C_6H_5	2	$C_{34}H_{33}N_3O_4$	4.52×10^{-10}	9.34
9	C_6H_5	3	C_6H_5	3	$C_{36}H_{37}N_3O_4$	9.72×10^{-11}	10.01
10	C_6H_5	4	C_6H_5	4	$C_{40}H_{45}N_3O_4$	6.42×10^{-10}	9.19
11	C_6H_5	5	C_6H_5	5	$C_{40}H_{45}N_3O_4$	8.91×10^{-8}	7.05
12	C_6H_{11} (cyclohexyl)	0	CH_3	0	$C_{25}H_{29}N_3O_4$	1.75×10^{-10}	9.75
13	C_6H_{11} (cyclohexyl)	0	CH_2CH_3	0	$C_{26}H_{31}N_3O_4$	8.45×10^{-10}	9.07
14	C_6H_{11} (cyclohexyl)	1	CH_3	0	$C_{26}H_{31}N_3O_4$	2.80×10^{-9}	8.55
15*	C_6H_{11} (cyclohexyl)	1	CH_2CH_3	0	$C_{27}H_{33}N_3O_4$	1.32×10^{-9}	8.88
16	C_6H_{11}	2	CH_3	0	$C_{27}H_{33}N_3O_4$	3.02×10^{-9}	8.52

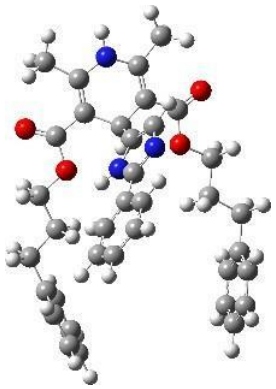
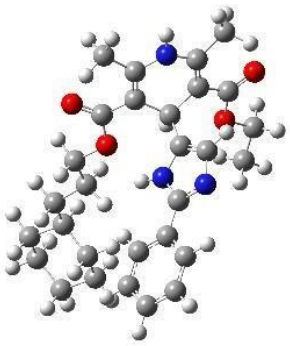
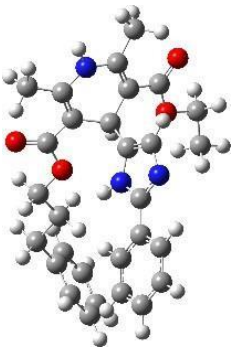
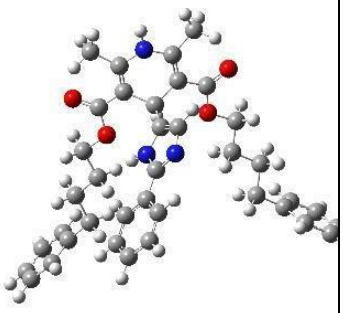
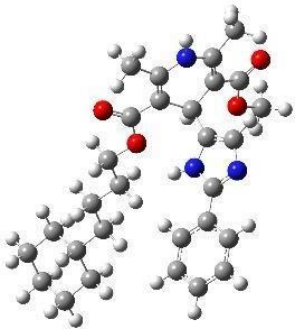
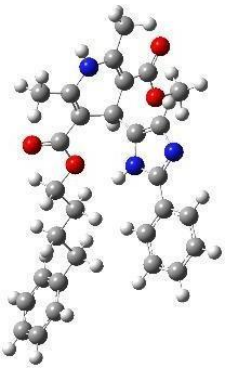
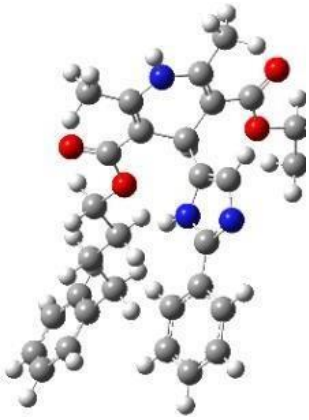
	(cyclohexyl)						
17	C_6H_{11} (cyclohexyl)	2	CH_2CH_3	0	$C_{28}H_{35}N_3O_4$	1.09×10^{-9}	8.96
18	C_6H_{11} (cyclohexyl)	3	CH_3	0	$C_{28}H_{35}N_3O_4$	3.52×10^{-9}	8.45
19	C_6H_{11} (cyclohexyl)	3	CH_2CH_3	0	$C_{29}H_{37}N_3O_4$	2.23×10^{-9}	8.65
20	C_6H_{11} (cyclohexyl)	4	CH_3	0	$C_{29}H_{37}N_3O_4$	4.94×10^{-9}	8.30
21	C_6H_{11} (cyclohexyl)	4	CH_2CH_3	0	$C_{30}H_{39}N_3O_4$	6.31×10^{-9}	8.20
22	C_6H_5	1	CH_3	0	$C_{26}H_{25}N_3O_4$	2.71×10^{-10}	9.56
23	C_6H_5	1	CH_2CH_3	0	$C_{27}H_{27}N_3O_4$	2.11×10^{-10}	9.67
24*	C_6H_5	2	CH_3	0	$C_{27}H_{27}N_3O_4$	5.23×10^{-10}	9.28
25	C_6H_5	2	CH_2CH_3	0	$C_{28}H_{29}N_3O_4$	1.90×10^{-10}	9.72
26	4-CH_3-C_6H_4	2	CH_3	0	$C_{28}H_{29}N_3O_4$	3.24×10^{-10}	9.49
27	4-CH_3-C_6H_4	2	CH_2CH_3	0	$C_{29}H_{31}N_3O_4$	5.75×10^{-10}	9.24
28	C_6H_5	3	CH_3	0	$C_{28}H_{29}N_3O_4$	5.84×10^{-10}	9.23
29*	C_6H_5	3	CH_2CH_3	0	$C_{29}H_{31}N_3O_4$	8.75×10^{-9}	8.05
30	C_6H_5	4	CH_3	0	$C_{29}H_{31}N_3O_4$	6.14×10^{-9}	8.21
31	C_6H_5	4	CH_2CH_3	0	$C_{30}H_{33}N_3O_4$	5.02×10^{-9}	8.30

* denotes the compounds selected for external validation (test set).

Table III .4 : *3D Chemical structures of 1, 4-dihydropyridines derivatives under study.*

NO	Compound structure	NO	Compound structure	NO	Compound structure
1		**11**		**21**	
2		**12**		**22**	
3		**13**		**23**	
4		**14**		**24**	

5		**15**		**25**	
6		**16**		**26**	
7		**17**		**27**	
8		**18**		**28**	

9		**19**		**29**	
10		**20**		**30**	
		31			

III.3.2. Drug likeness scoring.

Drug-likeness is a qualitative method applied in the research of the structure/activity relationship in biomolecules [15]. These methods study the balance between molecular prop erties affecting the pharmacodynamics and pharmacokinetics of molecules that have a significant impact on their absorption, distribution, metabolism and excretion (ADME)

in the human body. Molecular properties such as bioavailability, solubility and membrane permeability and are related to several basic molecular descriptors such as molecular weight (MW), topological polar surface (TPSA), partition coefficient (LogP), molar refractivity (MR),number of rotatable bonds (Nrotb), hydrogen bond donors (HBD) and hydrogen bond acceptors (HBA).These descriptors were calculated using HyperChem 8 and the tool Molinspiration server (http://www.molinspiration.com).

According to Lipinski's rule [16], good absorption of the drug is obtained if: the molecular weight is less than 500 Da, log P is less than 5, the H donor bonds are less than 5, the H acceptor bonds are less than 10 .

According to Veber's rule, a good bioavailability is more present if the compounds have rotatable bonds (nrotb) less than 10 and a total polar surface area (TPSA) less than 140 Å2 [17].

Ghose's rule defines a ligand as a good drug if: the molecular weight is between 160 and 480, the LogP is between -0.4 and 5.6, the molar refraction (MR) between 40 and 130, and the total number of atoms between 20 and 70 [18].

III.3.3. Molecular structure and molecular descriptors calculations

GAUSSIAN16 program was used to optimize the investigated structures and further generate their electronic descriptors by using DFT method at the B3LYP/6-31G+ (d,p) level as implemented in GAUSSIAN 16. To better describe their structural features, others various molecular descriptors were generated using Chem3D software [19].

Molecular frontier orbitals, HOMO and LUMO are major parameters in quantum chemistry, because they determine how the molecule interacts with other chemical entities. The value of the frontier orbital gap allows us to estimate and characterize the kinetic stability and the chemical reactivity of the studied molecule ($E_{gap} = E_{LUMO} - E_{HOMO}$). A molecule which has a small gap is more polarizable, ie associated with a high chemical reactivity and a low kinetic stability and is qualified as soft molecule, according to the concept of HSAB (hard and soft, acids and bases)

The conceptual descriptors based on the DFT quantum method have facilitated a good understanding of the 3D structure of molecules and their reactivity by calculating the global hardness (η) and the chemical potential (μ). Using the frontier orbital energies (E_{HOMO} and E_{LUMO}), the chemical potential μ is equal to ($E_{HOMO}+E_{LUMO}$)/2 and the hardness is given by η =

$(-E_{HOMO} + E_{LUMO})/2$ [20]. This index measures the stabilization of the molecule when the system has obtained an additional charge from the environment, therefore, the electrophilicity presents a double capacity, the first to acquire an additional electronic charge and the second to resist an exchange of an electronic charge with the environment. The global electrophilicity (ω) power of a ligand is given by $\omega = \mu^2/2\eta$

III.3.4. QSAR study

In the current QSAR study, two different methods of cheminformatics techniques, namely multiple linear regression (MLR) and artificial neural networks (ANNs), were used to model the relationship between the observed activity of the investigated compounds and their molecular descriptors. Before starting modeling, the collected series was randomly divided into a training set and a test set. The quality of developed models was tested by calculating their internal validation indices and external validation statistics parameters including the determination coefficient of external validation estimating their predictive abilities and y-Randomization test parameters giving insight into their robustness. The model's domain of applicability was also identified to be used as guide to design new compounds with enhanced CCBs activities.

III .4. RESULTS AND DISCUSSION

III .4.1. Molecular structures

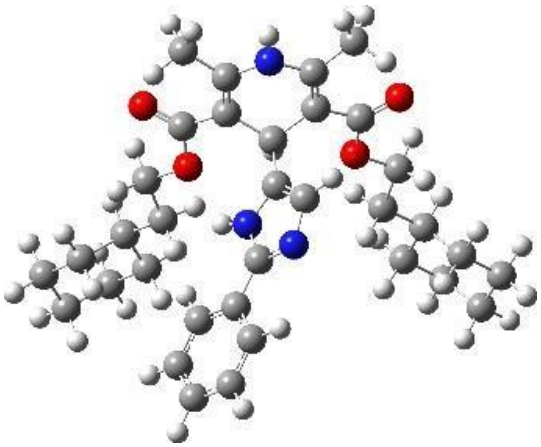


Figure. III 3. *3D conformation of compound 3*

As discussed in Refs. [21-25], the study of the subunit common to all members of a series may help in understanding of their 3D structure and their biological activities. Nevertheless, we found that the 3D structure of the 4-imidazolyl- 1,4-dihydropyridine molecule, which is the subunit of the series under investigation, has different 3D shape with those of the series. As can be seen in Figure III.3, the 4-imidazolyl- 1,4-dihydropyridine structure consists of two planar phenyl and imidazolyl group, whereas the 1,4-dihydropyridine moiety is close to orthogonal to them. Within the series, three cycles are however close to planar favoring their stabilization by electron delocalization over the three cycles. Such structural differences between the isolated subunit and the series is due to the additional long range interactions between the phenyl and the alkoxy substituents attached to ester functions of the present 1,4-dihydropyridine derivatives (e.g. CH- interactions).

III.4.2. Drug likeness screening of 1,4-dihydropyridines derivatives

All compounds under study have a LogP between 0 and 5 except compounds 22, 23, 24 and 26 LogP values that are in the range of 1 to 3, resulting in good oral bioavailability; therefore the ligand has a good aqueous solubility to dissolve in the digestive tract. However for a value of LogP > 3, the ligand has a low solubility in the stomach and the intestines (compounds 4,5 and 11) and for LogP < 1 the drug has a good solubility, but it has a bad penetration through lipid membranes. [26]

Molecular weight (MW) is an important factor determines drug permeability, for molecules with a molar mass < 450, the result is better cerebral permeability and good oral absorption (compounds 12, 13, 14 and 22) [27].

All compounds have H-bond acceptors less than 10 and H-bond donors less than 5. HBAs those are large in number leads to low permeability through a lipid bilayer membrane, while smaller number leads to better permeability. [28]

The topological polar surface (TPSA) is a necessary parameter to predict the intrinsic properties of molecular transport, especially in permeability, that is to say the speed of passage of molecules through the blood-brain barrier and also in intestinal absorption [29]. For our results, all 1,4 dihydropyridines derivatives have values below 140 $Å^2$ (93.32$Å^2$), which shows the good prediction of oral bioavailability and transport through biological membranes. The total number

of atoms for all compounds is between 20 and 70.

In conclusion the majority of ligands agree with Lipinski rules, Veber rules and Ghose rules, in particular compound 12 Lipinski who has the top score in the three rules. These rules are used as filters to select the best promising molecules.

Table III 5. *Pharmacological activities and properties for1, 4 dihydropyridines derivatives under study; with application of the following rules: Lipinski, Veber and Ghose.*

comp	*Log P*	*MW(amu)*	*HBA*	*HBD*	*Lipinski Score*	*TPSA*	*nrotb*	*Veber Score*	*MR*	*N atoms*	*Ghose and al Score*
1	1.64	503.28	7	2	3	93.32	8	2	150.09	37	2
2	2.27	531.31	7	2	3	93.32	10	2	159.55	39	2
3	2.91	559.34	7	2	3	93.32	12	1	168.91	41	2
4	3.71	587.37	7	2	3	93.32	14	1	178.11	43	2
5	4.50	615.40	7	2	3	93.32	16	1	187.31	45	2
6	2.91	559.34	7	2	3	93.32	14	1	168.91	41	2
7	0.50	519.22	7	2	3	93.32	10	2	165.57	39	2
8	1.00	547.25	7	2	3	93.32	12	1	175.08	41	2
9	1.79	575.28	7	2	3	93.32	14	1	184.28	43	2
10	2.58	603.31	7	2	3	93.32	16	1	193.48	45	2

11	3.38	631.34	7	2	3	93.32	18	1	202.68	47	2
12	0.05	435.22	7	2	4	93.32	7	2	129.08	32	4
13	0.40	449.23	7	2	4	93.32	8	2	133.82	33	3
14	0.37	449.23	7	2	4	93.32	8	2	133.81	33	3
15	0.71	463.25	7	2	4	93.32	9	2	138.56	34	3
16	0.69	463.25	7	2	4	93.32	9	2	138.49	34	3
17	1.03	477.26	7	2	4	93.32	10	2	143.23	35	3
18	1.09	477.26	7	2	4	93.32	10	2	143.09	35	3
19	1.43	491.28	7	2	4	93.32	11	1	147.84	36	2
20	1.48	491.28	7	2	4	93.32	11	1	147.69	36	2
21	1.83	505.29	7	2	3	93.32	12	1	152.44	37	2
22	-0.52	443.18	7	2	4	93.32	8	2	136.81	33	3
23	-0.17	457.20	7	2	4	93.32	9	2	141.56	34	3
24	-0.27	457.20	7	2	4	93.32	9	2	141.57	34	3
25	0.08	471.22	7	2	4	93.32	10	2	146.32	35	3

26	-0.11	471.22	7	2	4	93.32	9	2	145.85	35	3
27	0.23	485.23	7	2	4	93.32	10	2	150.60	36	2
28	0.13	471.22	7	2	4	93.32	10	2	146.17	35	3
29	0.47	485.23	7	2	4	93.32	11	1	150.92	36	2
30	0.53	485.23	7	2	4	93.32	11	1	150.77	36	2
31	0.87	499.25	7	2	4	93.32	12	1	155.52	37	2

III.4.3. Molecular descriptors

The molecular descriptor is the consequence of a mathematical procedure which transforms the chemical information in a symbolic character representation of a molecule into a useful number with standard numerical character. To assess chemical properties and to reduce the number of laboratory tests which are very expensive, so it is essential to build reliable models using computational chemistry to establish a quantitative mathematical relationship between the structures of molecules and the properties desires (QSAR). Descriptors can be classified in different forms. In particular, two main categories are distinguished, calculated descriptors and purely experimental descriptors. [30, 31]. The process of encoding the structure of chemical compounds by descriptors is an essential step in QSAR studies. [32]

We calculated 29 descriptors of 1,4-dihydropyridine derivatives, including Exact Mass (EM), Intrinsic Solubility (Log S), Number Rotatable Bonds (NRB), Shape Coefficient (SC), Sum of Valence Degrees (SVD), Topological Diameter (TD), Molar Refractivity (MR), Hydration Energy (HE), Molar Volume (V), Surface Area Grid (SAG) and Partition Coefficient Octanol/Water (Log P). The values of the calculated descriptors are given in Table III. 6. These descriptors will be used later to establish the QSAR models.

Table III 6. *Calculated molecular descriptors.*

N	EM	LogS	NRB	SC	SVD	TD	MR	HE	V	SAG	logP	E_{HOMO}	E_{LUMO}	E_{HOMO-1}	E_{LUMO+1}	D	E	η	χ	E_{Gap}
1	503.28	-7.50	8	1	122	14	150.09	-4.21	1451.63	812.16	1.64	-5.24	-1.63	-5.83	-0.47	3.92	-44361.11	1.81	3.44	3.61
2	531.31	-8.40	10	1	126	16	159.55	-3.62	1569.04	870.21	2.27	-5.28	-1.68	-5.88	-0.51	3.75	-46499.93	1.80	3.48	3.60
3	559.34	-8.78	12	1	130	18	168.91	-2.24	1756.62	1003.51	2.91	-5.26	-1.65	-5.85	-0.49	3.81	-48639.02	1.80	3.46	3.61
4	587.37	-9.46	14	1	134	20	178.11	-3.07	1846.21	1052.63	3.71	-5.26	-1.66	-5.86	-0.50	3.73	-50778.11	1.80	3.46	3.60
5	615.40	-10.14	16	1	138	22	187.31	-1.01	1988.95	1150.05	4.50	-5.26	-1.66	-5.86	-0.50	3.77	-52916.92	1.80	3.46	3.61
6	559.34	-8.67	14	1	130	18	168.91	-2.96	1713.28	979.85	2.91	-5.26	-1.66	-5.87	-0.50	3.73	-48638.47	1.80	3.46	3.60
7	519.22	-7.80	10	1	138	16	165.57	-10.45	1448.30	814.67	0.50	-5.21	-1.71	-5.89	-0.57	4.08	-46302.92	1.75	3.46	3.50
8	547.25	-8.09	12	1	142	18	175.08	-9.61	1599.33	915.66	1.00	-5.31	-1.72	-5.92	-0.55	3.79	-48442.01	1.80	3.52	3.59
9	575.28	-8.63	14	1	146	20	184.28	-8.88	1684.03	954.11	1.79	-5.28	-1.70	-5.89	-0.51	3.70	-50580.82	1.79	3.49	3.59
10	603.31	-9.47	16	1	150	22	193.48	-8.21	1793.49	1008.54	2.58	-5.30	-1.70	-5.90	-0.53	3.74	-52719.91	1.80	3.50	3.60
11	631.34	-10.31	18	1	154	24	202.68	-7.66	1880.58	1063.00	3.38	-5.28	-1.69	-5.89	-0.51	3.74	-54859.00	1.79	3.48	3.59
12	435.22	-5.79	7	1	110	14	129.08	-5.39	1252.30	713.83	0.05	-5.28	-1.68	-5.87	-0.49	4.12	-39046.18	1.80	3.48	3.60
13	449.23	-6.13	8	1	112	14	133.82	-4.81	1308.63	745.27	0.40	-5.27	-1.66	-5.85	-0.49	4.08	-40115.86	1.81	3.46	3.62
14	449.23	-6.24	8	0	112	15	133.81	-5.02	1324.04	761.86	0.37	-5.28	-1.70	-5.90	-0.52	3.77	-40115.59	1.79	3.49	3.58
15	463.25	-6.58	9	0	114	15	138.56	-4.41	1376.32	793.15	0.71	-5.28	-1.68	-5.88	-0.51	3.77	-41185.27	1.80	3.48	3.60
16	463.25	-6.81	9	1	114	16	138.49	-4.43	1447.98	825.77	0.69	-5.27	-1.69	-5.89	-0.50	3.97	-41185.00	1.79	3.48	3.58
17	477.26	-7.15	10	1	116	16	143.23	-3.85	1483.56	854.08	1.03	-5.26	-1.67	-5.87	-0.50	3.91	-42254.68	1.80	3.47	3.60
18	477.26	-7.22	10	0	116	17	143.09	-4.40	1462.37	844.40	1.09	-5.27	-1.69	-5.89	-0.51	3.79	-42254.41	1.79	3.48	3.58

19	491.28	-7.56	11	0	118	17	147.84	-3.88	1524.98	886.22	1.43	-5.27	-1.67	-5.87	-0.50	3.86	-43324.09	1.80	3.47	3.60
20	491.28	-7.64	11	1	118	18	147.69	-3.78	1511.14	880.51	1.48	-5.27	-1.69	-5.90	-0.51	3.89	-43324.09	1.79	3.48	3.58
21	505.29	-7.98	12	1	120	18	152.44	-3.29	1562.52	899.07	1.83	-5.27	-1.67	-5.87	-0.50	3.83	-44393.77	1.80	3.47	3.60
22	443.18	-5.93	8	0	118	15	136.81	-8.27	1262.50	723.19	-0.52	-5.22	-1.70	-5.89	-0.57	3.76	-40016.81	1.76	3.46	3.51
23	457.20	-6.27	9	0	120	15	141.56	-7.69	1318.10	552.38	-0.17	-5.22	-1.68	-5.87	-0.57	3.71	-41086.49	1.77	3.45	3.54
24	457.20	-6.08	9	1	120	16	141.57	-8.15	1336.03	772.92	-0.27	-5.30	-1.72	-5.92	-0.54	3.92	-41086.49	1.79	3.51	3.58
25	471.22	-6.42	10	1	122	16	146.32	-7.57	1389.85	801.39	0.08	-5.30	-1.70	-5.90	-0.53	3.84	-42156.17	1.80	3.50	3.59
26	471.22	-6.44	9	0	122	17	145.85	-6.95	1381.48	801.95	-0.11	-5.19	-1.70	-5.89	-0.48	3.61	-42156.17	1.75	3.44	3.49
27	485.23	-6.78	10	0	124	17	150.60	-6.38	1436.92	829.38	0.23	-5.19	-1.68	-5.87	-0.47	3.56	-43225.85	1.76	3.43	3.51
28	471.22	-6.35	10	0	122	17	146.17	-7.72	1381.76	799.42	0.13	-5.28	-1.71	-5.91	-0.52	3.78	-42155.90	1.79	3.50	3.57
29	485.23	-6.69	11	0	124	17	150.92	-7.14	1436.81	831.57	0.47	-5.28	-1.69	-5.89	-0.51	3.73	-43225.58	1.80	3.49	3.59
30	485.23	-6.77	11	1	124	18	150.77	-7.42	1491.96	873.56	0.53	-5.29	-1.71	-5.92	-0.53	3.78	-43225.58	1.79	3.50	3.58
31	499.25	-7.11	12	1	126	18	155.52	-6.84	1547.91	903.90	0.87	-5.29	-1.69	-5.90	-0.53	3.73	-44295.26	1.80	3.49	3.60

EM: Exact Mass; Log S: Intrinsic Solubility; NRB: Number of Rotatable Bonds; SC: Shape Coefficient; SVD: Sum of Valence Degrees; TD: Topological Diameter; MR: Molar Refractivity; HE: Hydration Energy; V: Molar Volume; SAG: Surface Area Grid; Log P: Partition Coefficient Octanol/Water;E_{HOMO}(eV), E_{LUMO}(eV), E_{HOMO-1}(eV)and E_{LUMO+1} (eV): Energies of the HOMO, LUMO, HOMO-1 and LUMO+1 orbitals, respectively; D (Debye): Dipole moment; E (eV): Total energy; η(eV): Hardness; χ(eV): Electronegativity; E_{Gap}(eV): Energy Gap.

III.4.4. Data analysis

For building QSAR models, we need a data analysis technique. The latter quantifies the relationship between the structure and the activity (descriptor). There are different methods for constructing a QSAR model and analyzing the statistical data of the latter. We used in our study the XLSTAT software [33] for multiple linear regression and Matlab software [34] for artificial neural network investigations.

III.4.4.1. Multiple linear regression (MLR)

The MLR statistical method is one of the most widely used approaches for developing QSAR models, either alone or in combination with other statistical methods, this method is simple to use and very practical, the interpretation is easy with clear transparency. [35]

We have established a remarkable correlation between CCBs activity and molecular descriptors which is represented by the following equation:

$pIC_{50} = 4.089 + 1.871\ Log\ S - 1.242\ NRB - 0.494\ SVD + 0.568\ MR - 12.480\ E_{LUMO+1}$

The previous equation shows that the biological activity depends solely on the intrinsic solubility (**Log S**), the number of rotatable bonds (**NRB**), sum of valence degrees (**SVD**), molar refractivity (**MR**) and the energy of the LUMO+1 ($\mathbf{E_{LUMO+1}}$) physicochemical descriptors. Indeed, these descriptors are strongly correlated with the pIC50 value quantifying the target activity. Moreover, we also note that Log S, and MR are preceded by a positive sign, while the remaining descriptors are preceded by a negative sign. Therefore, increasing the values of log S and MR or decreasing the values of NRB, SVD and $\mathbf{E_{LUMO+1}}$ will contribute to the improvement of the target activity.

Table III .7. *Predicted and Experimental pIC_{50} values using ANN and MLR methods.*

N°	Exp. pIC_{50}	Pred. pIC_{50}		N°	Exp. pIC_{50}	Pred. pIC_{50}	
		MLR	ANN			MLR	ANN
1*	9.52	10.93	10.93	16	8.52	8.81	8.97
2	10.44	10.75	10.35	17	8.96	8.59	8.68
3	10.74	10.64	10.69	18	8.45	8.48	8.74
4	10.92	10.21	10.43	19	8.65	8.24	8.42

5	9.19	9.68	9.45	**20**	8.30	8.08	8.48
6	8.55	8.41	8.90	**21**	8.20	7.86	8.15
7*	9.25	10.01	10.01	**22**	9.56	9.56	9.37
8*	9.34	10.17	10.17	**23**	9.67	9.46	9.19
9	10.01	9.50	10.09	**24***	9.28	9.38	9.38
10	9.19	8.98	8.92	**25**	9.72	9.16	8.72
11	7.05	7.90	8.11	**26**	9.49	9.44	9.41
12	9.75	9.65	9.25	**27**	9.24	9.18	9.35
13	9.07	9.45	9.02	**28**	9.23	9.02	8.62
14	8.55	9.59	9.13	**29***	8.05	8.79	8.79
15*	8.88	9.35	9.35	**30**	8.21	8.77	8.34
				31	8.30	8.55	7.92

(* denotes the compounds selected for test set).

The values of calculated activities using MLR model are given in Table III 7, Figure III.4 shows the plot of the experimental activities against the predicted values. We notice that the predicted pIC50 values are in good agreement with the experimental values with correlation coefficient R^2=0.762 for the training set and R^2_{test}=0.684 for test set shows the good correlation between different independent variables with the activity of BCCs.

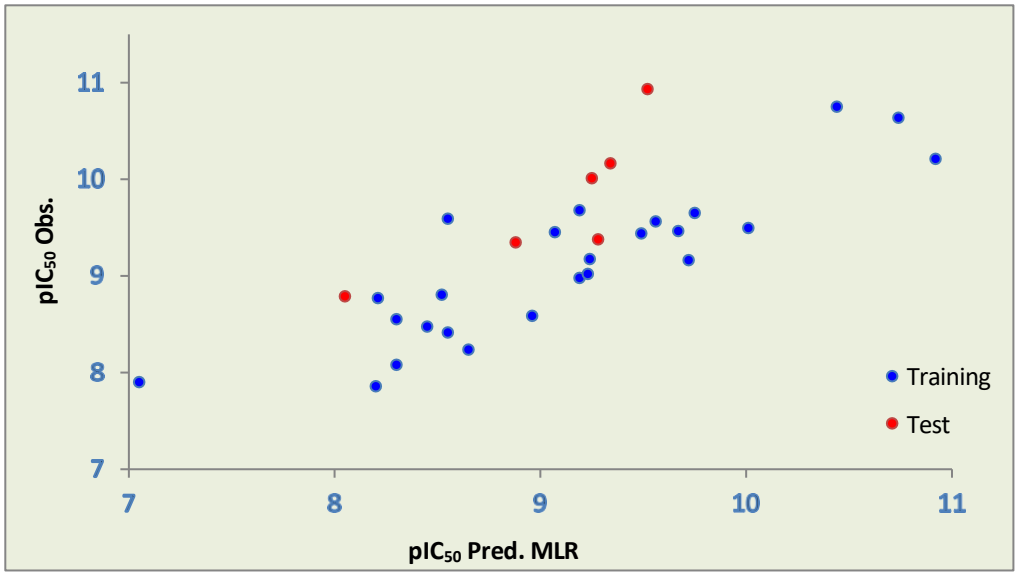


Figure. III .4. *Correlation between predicted and experimental pIC50 values using MLR.*

The domain of applicability (DA) of each QSAR model is considered a necessary and essential step to assess the accuracy of the prediction. The AD of the built MLR model was determined by using the leverage approach to represent the William's plot accessible in Matlab software [36-37]. The leverage of a given chemical entity hi is defined: $hi = x_i^T(X^TX)^{-1}x_i$ (i= 1, 2…n), where X is the descriptor matrix of the chemical compounds in the training set used to build the model and xi is the query compound descriptor row. As a prediction tool, the warning leverage h* is defined as: h*=3(P+1)/n, where P is the number of descriptors in the model and n is the number of training compounds. The domain of applicability is realized in a known area squared inside a standard deviation (x= ±2.5) used as a cut of value for accepting predictions [38–39].

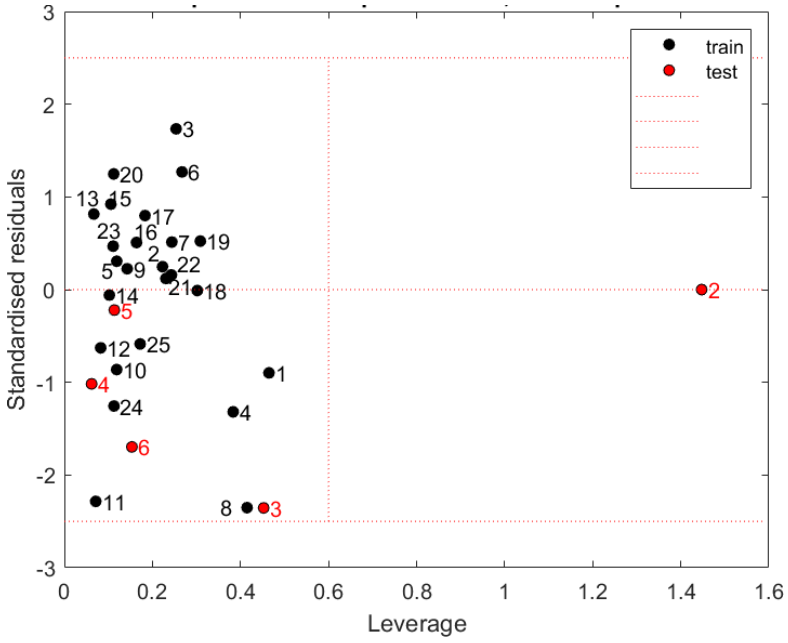


Figure. III .5. *William's plot of standardized residual versus leverage via MLR model, with h* = 0.250 and residual limits ± 2.5. Training samples are in black and test samples are in red. The dashed red lines correspond to the interval delimiting the applicability domain.*

From the William's plot (Figure. III. 5) all compounds in the series are within the applicability domain of the model except test compound **2**. This compound has a leverage value greater than the warning h* value and may be a high leverage compound influencing the model's performance. So their standard residuals are very low and within the established limit. This compound may be

considered significant in fitting the model performance, but there is no need to delete it from the data set. Besides, the standardized residuals of all compounds belongs to the interval [-2.5, 2.5] used for accepting predictions

III.4.4.2. Artificial neuron network

Artificial neural networks (ANN) are very useful for the prediction of biological activity for a dataset of chemical compounds [40]. The latter make it possible to study complex and non-linear relationships, unlike conventional statistical methods (MLR, PLS,..) [41]. They are accustomed to modeling various complex nonlinear systems in several fields such as the pharmaceutical field, medicinal chemistry, engineering, psychology [42].

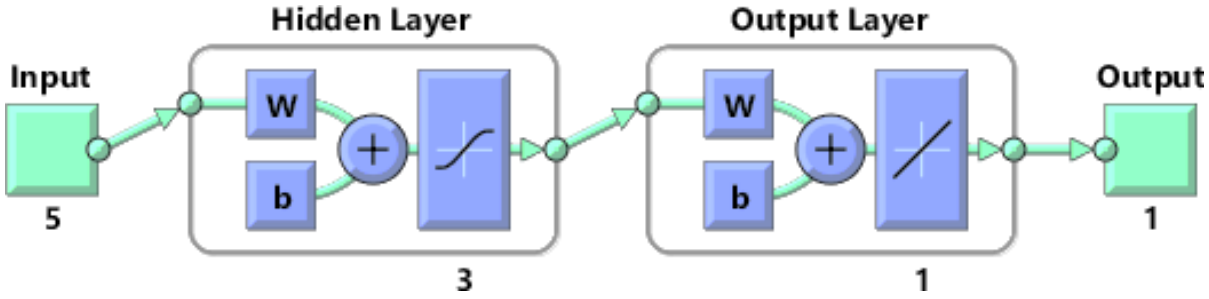


Figure. III. 6. Structure of ANN used in this work

The structure of the ANN is designed as follows, the neural network is a more or less complex system of neurons fully interconnected and structured in three layers. As can be seen in *Figure. III. 6,* the input layer is composed of five neurons, each neuron receiving one of the five descriptors selected in the MLR model (Log S, NRB, SVD, MR and E_{LUMO+1}**)**. The hidden layer also called middle is made up of three neurons that form the deep internal entity that reveals the most important correlations between experimental and predicted data. One neuron is the output layer, which returns the value of pIC50 [43]. The values of calculated activities using ANN model are given in Table III .7.

Figure. III 7. gives the experimental activities versus the predicted values as determined by the ANN model. We notice that the predicted pIC50 values are in very good agreement with the values taken from the experiment., which demonstrates a reliable correlation between the five selected descriptors with CCBs activity. In fact, the statistical parameters of this model present a correlation coefficient of R^2= 0.767. In addition, the reliability and robustness of the model has been confirmed by the

important R^2_{test} value of the test data set (R^2_{test} = 0.876).

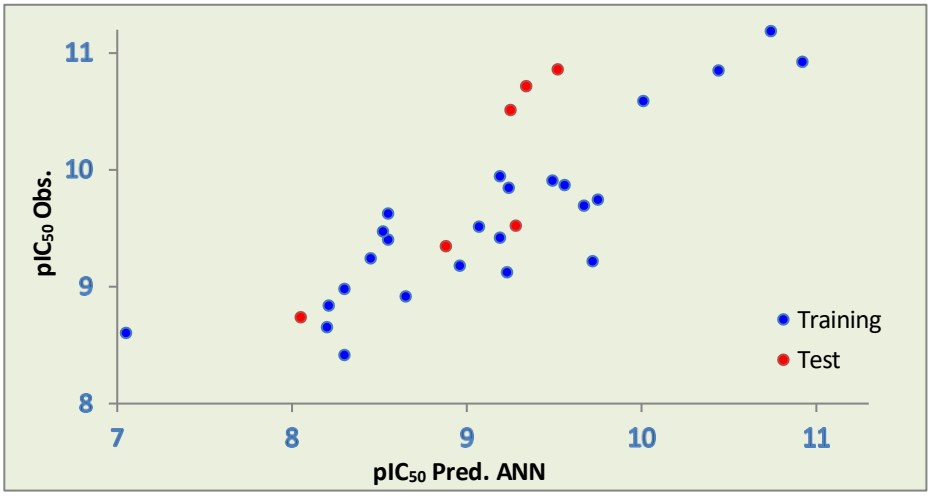


Figure III .7. *Correlation of predicted and experimental pIC50 values validated by the ANN model.*

III .5. CONCLUSION:

The current study presents a structural analysis of thirty-one 1,4-dihydropyridine derivative compounds. Additionally, it includes a discussion of various qualitative approximations of the relationship between their activity and structural properties. All compounds under study have a LogP between 0 and 5 except compounds 22, 23, 24 and 26 LogP values that are in the range of 1 to 3, resulting in good oral bioavailability; therefore the ligand has a good aqueous solubility to dissolve in the digestive tract. However for a value of LogP > 3, the ligand has a low solubility in the stomach and the intestines (compounds 4,5 and 11) and for LogP < 1 the drug has a good solubility, but it has a bad penetration through lipid membranes.

Molecular weight (MW) is an important factor determines drug permeability, for molecules with a molar mass < 450, the result is better cerebral permeability and good oral absorption (compounds 12, 13, 14 and 22) .

All compounds have H-bond acceptors less than 10 and H-bond donors less than 5. HBAs those are large in number leads to low permeability through a lipid bilayer membrane, while smaller number leads to better permeability.

The topological polar surface (TPSA) is a necessary parameter to predict the intrinsic properties of

molecular transport, especially in permeability, that is to say the speed of passage of molecules through the blood-brain barrier and also in intestinal absorption. For our results, all 1,4 dihydropyridines derivatives have values below 140 $Å^2$ (93.32$Å^2$), which shows the good prediction of oral bioavailability and transport through biological membranes. The total number of atoms for all compounds is between 20 and 70.

In conclusion the majority of ligands agree with Lipinski rules, Veber rules and Ghose rules, in particular **compound 12** Lipinski who has the top score in the three rules. These rules are used as filters to select the best promising molecules.

In this work, the CCBs activity of a series of 1,4-dihydropyridines was predicted using MLR and ANN approaches. The external and internal consistency of the built models was confirmed using internal and external validation methods to assess their statistical quality. External validation of these models confirmed their ability to accurately predict the CCBs activity of the studied compounds. Furthermore, the MLR equation can show that physical features, organic functional groups, and chemical molecular fragments are significantly linked to the desired activity of these studied compounds.

REFERECES

[1] Montañez-Godínez N and al (2015) QSAR/QSPR as an application of artificial neural networks. In: Artificial neural networks. Springer : 319-33.

[2] Puri M, Solanki A, Padawer T, Tipparaju S. M, Moreno W. A, & Pathak Y (2016) Introduction to Artificial Neural Network (ANN) as a Predictive Tool for Drug Design, Discovery, Delivery, and Disposition. Artificial Neural Network for Drug Design, Delivery and Disposition : 3–13.

[3] Judson R, Kavlock R, Martin M, Reif D, Houck K, Knudsen T, Richard A, Tice R.R, Whelan M, Xia M, Huang R, Austin C, Daston G, Hartung T, Fowle III J.R, Wooge W, Tong W, Dix D(2013) Perspectives on validation of high-throughput assays supporting 21st century toxicity testing, ALTEX. 30: 51-56.

[4] Berhanu W. M, Pillai G. G, Oliferenko A. A, & Katritzky A. R (2012) Quantitative Structure-Activity/Property Relationships: The Ubiquitous Links between Cause and Effect. Chem Plus Chem. 77(7):507–517.

[5] Ramírez-Galicia G, Garduño-Juárez R, Hemmateenejad B et al (2007) QSAR Study on the antinociceptive activity of some morphinans. Chem Biol Drug Des 70 :53–64.

[6] Arora P, Arora V, Lamba H.S and Wadhwa D (2012) Importance of heterocyclic chemistry: a review. IJPSR. 3(9) : 2947-2954.

[7] Gaba M and Mohan C (2016) Development of drugs based on imidazole and benzimidazole bioactive heterocycles: recent advances and future directions. Medicinal Chemistry Research 25 : 173–210.

[8] HyperChem (Molecular Modeling System) Hypercube, Inc., 1115NW, Gainesville,FL32601,USA,2008.

[9] Frisch M.J, Trucks G.W, Schlegel H.B, Scuseria G.E, Robb M.A, Cheeseman J.R, Scalmani G, Barone V, Mennucci B, Petersson G.A, Nakatsuji H, Caricato M, Li X, Hratchian H.P, Izmaylov A.F, Bloino J, Zheng G, Sonnenberg J.L, Hada M, Ehara M, Toyota K, Fukuda, R., Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery Jr J.A, Peralta J.E, Ogliaro F, Bearpark M, Heyd J.J, Brothers E, Kudin K.N, Staroverov V.N, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant J.C, Iyengar S.S, Tomasi J, Cossi, M., Rega N, Millam J.M,

Klene M, Knox J.E, Cross J.B, Bakken V, Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi R, Pomelli, C, Ochterski J.W, Martin R.L, Morokuma K, Zakrzewski V.G, Voth G.A, Salvador P, Dannenberg J.J, Dapprich, S, Daniels A.D, Farkas O, Foresman J.B, Ortiz J.V, Cioslowski J and Fox D.J(2016) Gaussian 16, Revision C.01, Gaussian, Inc., Wallingford CT,

[10] cf.http://www.molinspiration.com

[11] Li H.-B, Lu N, Zhang Q , Wang Y, Feng D, Chen T, … Yu P. (2018) Electric-field control of ferromagnetism through oxygen ion gating. Nature Communications 9(1) :580.

[12] Salah T, Belaidi S, Nadjib M, Ismail D, & BOUGHDIRI S (2016) In silico investigation by conceptual DFT and molecular docking of antitrypanosoma compounds for understanding cruzain inhibition. Journal of Theoretical and Computational Chemistry, 15(3).

[13] NBO 6.0. Glendening E.D, Badenhoop J.K, Reed A.E, Carpenter J.E, Bohmann J.A, Morales C.M, Landis C.R,Weinhold F, Theoretical Chemistry Institute, University of Wisconsin,Madison,2013.

[14] Bossert F, Vater W (1971) Dihydropyridine, eine neue Gruppe stark wirksamer Coronartherapeutika [Dihydropyridines, a new group of strongly effective coronary therapeutic agents], Natur wissenschaften 58: 578.

[15] Schneider G, Baringhaus K-H,Kubinyi H (2008)Molecular Design: Concepts and Applications .JohnWiley&Sons.

[16] Lipinski C.A, Lombardo F, Dominy B. W, Feeney P (1997) J. Adv. Drug. Deliv. Rev23: 3-25.

[17] Veber D.F, Johnson S., Cheng H.Y , Smith B.R , Ward K.W , Kopple K.D (2002) Molecular Properties That Influence the Oral Bioavailability of Drug Candidates .J.Med.Chem 45:2615-2623.

[18] Ghose A.K , Viswanadhan V.N,. Wendoloski J. J, Comb (1999) J. Chem 1 :55.

[19] ChemOffice software, PerkinElmer Informatics Company, Wellesley, Massachusetts, USA. http://www.cambridgesoft.com

[20] Chtita S , Ghamali M , HmamouchiR , Elidrissi B , Bourass B , Larif M , Bouachrine M , Lakhlifi T(2016) Investigation of Antileishmanial Activities of Acridines Derivatives against Promastigotes and Amastigotes Form of Parasites Using Quantitative Structure Activity Relationship Analysis.Advances in Physical Chemistry 5137289 :16.

[21] Alloui M, Belaidi S , Othmani H , Jaidane N.E , Hochlaf M (2018) Imidazole derivatives as angiotensin II AT1 receptor blockers : Benchmarks, drug-like calculations and quantitative structure-activity relationships modeling. Chem. Phys. Lett 696 :70–78.

[22] Boudergua S , Alloui M , Belaidi S , Mogren Al Mogren M, Abd Ellatif Ibrahim U.A, Hochlaf M (2019) QSAR Modeling and Drug-Likeness Screening for Antioxidant Activity of Benzofuran Derivatives. J. Mol. Struct 1189 :307-314.

[23] Ghamri M , Harkati D, Belaidi S, Boudergua S , Linguerri R, Chambaud G , Hochlaf M (2020) Carbazole derivatives containing chalcone analogues targeting topoisomerase II inhibition: first principles characterization and QSAR modeling. Spectrochim. Acta A 242 :118724.

[24] Almi I , Belaidi S , Zerroug A , Alloui A, Ben Said R , Linguerri R , Hochlaf M(2020) QSAR investigations and structure-based virtual screening on a series of nitrobenzoxadiazole derivatives targeting human glutathione-S-transferases. J. Mol. Struct 1211 :128015.

[25] Manachou M, Gouid Z , Almi Z, Belaidi S, Boughdiri S , Hochlaf M (2020) Pyrazolo[1,5-a][1,3,5] triazin-2thioxo-4-ones derivatives as thymidine phosphorylase inhibitors: Structure, drug-like calculations and quantitative structure-activity relationships (QSAR) modeling. J. Mol. Struct 1199 :127027.

[26] Eros D , Kövesdi I ,Orfi L , Takács-Novák K ,Acsády ,Gy,Kéri V (2002) Reliability of logP predictions based on calculated molecular descriptors: a critical review. Curr. Med. Chem 9 :1819-29.

[27] Ambati A, Canakis C.S, Miller J.W ,Gragoudas E.S , Edwards E, Weissgold D.J , Kim I, Delori F.C, Adamis A.P (2000) Diffusion of High Molecular Weight Compounds through Sclera. Investig. Ophthalmol. Vis. Sci 41 :1181-1185.

[28] Lipinski C.A, Lombardo F, Dominy W , Feeney P (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. J. Adv, Drug Deliv. Rev 46 :3-26.

[29] Schaftenaar G, de Vlieg J (2012) Quantum mechanical polar surface area. J. Comput. Aided Mol.Des 26 (3):311–318.

[30] Dermeche K, Tchouar N , Belaidi S, Salah T (2015) Qualitative Structure-Activity Relationships and 2D-QSAR Modeling of TNF-α Inhibition by Thalidomide Derivatives. J. Bionanosci. 9 :395-400.

[31] Kerassa A, Belaidi S, Harkati D, Lanez T, Prasad O, Sinha L (2016) Investigations on Molecular Structure, Electronic Properties, NLO Properties and Comparison of Drug-Likeness of Triazolothiadiazole Derivatives by Quantum Methods and QSAR Analysis .Rev. Theor. Sci. 4 :85-96.

[32] Mauri A , Consonni V , Todeschini R (2017) Handbook of Computational Chemistry. Springer, Cham :2065-2093.

[33] XLSTAT software, XLSTAT Company. Retrieved from https://www.xlstat.com.

[34] MATLAB 7.9.0 (R2009b) and Statistics Toolbox Release, the Math Works‖, Inc., Natick, Massachusetts, United States, 2011.

[35] Chtita S , Ghamali M , Hmamouchi R , Larif M, Bouachrine M , Lakhlifi T (2017) Quantitative structure–activity relationship studies of anticancer activity for Isatin (1H-indole-2,3-dione) derivatives based on density functional. IJQSPR 2 :90-115.

[36] Eriksson L, Jaworska J , Worth A.P , Cronin M.T , McDowell R.M ,Gramatica P (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classifcation- and regression-based QSARs. Environ. Health Perspect 111 :1361–1375.

[37] Netzeva T.I , Worth A.P , Aldenberg T, Benigni R , Cronin M.T , Gramatica P, Jaworska J.S ,Kahn S, Klopman G, Marchant C.A (2005) Current status of methods for defning the applicability domain of quantitative structure-activity relationships -The Report and Recommendations of ECVAM Workshop 52

33(2) :155–173.

[38] Sabando M.V , Ponzoni I ,Soto A.J (2019) Neural-based approaches to overcome feature selection and applicability domain in drugrelated property prediction. Appl. Soft Comput 85 :105777.

[39] Jaworska J , Nikolova-Jeliazkova N , Aldenberg T (2005) QSAR applicability domain estimation by projection of the training set in descriptor space 33 :445–459.

[40] Luo J , Hu J , Fu L , Liu C, Jin X (2011) Use of Artificial Neural Network for a QSAR Study on Neurotrophic Activities of N-p-Tolyl/phenylsulfonyl L-Amino Acid Thiolester Derivatives. Procedia Eng 15 :5158-5163.

[41] Terfloth L , Gasteiger J (2001) Neural networks and genetic algorithms in drug design, DDT 6 :102-108.

[42] Cheng F ,Sutariya V (2012) Applications of Artificial Neural Network Modeling in Drug Discovery. Clin. Exp. Pharmacol 2 :113.

[43] Ripley B.D (1996) Pattern Recognition and Neural Networks, Cambridge University Press, NY United States, USA,.

CHAPTER IV

MOLECULAR DOCKING STUDIES OF 1.4 DIHYDROPYRIDINE DERIVATIVES

IV .1. INTRODUCTION

Drug design often relies on computer modeling techniques, commonly referred to as "in silico" methods. This approach, known as computer-aided drug design, aims to predict the affinity between a ligand and its target. Thus far, the outcomes of these modeling techniques have generally been satisfactory. however, there exist additional properties that must be fine-tuned for the ligand to achieve optimal safety and efficacy as a drug. These molecular attributes, including the absence of side effects, bioavailability, metabolic half-life, toxicity….pose considerable challenges for rational drug design techniques [1].The mode of action for numerous drugs involves serving as inhibitors or activators of specific receptors relevant to drug development. The elucidation of the three-dimensional structure of proteins implicated in various pathologies has facilitated, through computer simulations, the identification of potent inhibitors targeting these proteins. This has significantly diminished the need for extensive screening tests to develop new drugs [2].

Interactions between molecules are the basis of most biological mechanisms understanding the intricate workings of these interactions on a molecular scale is highly compelling, with methods such as X-ray crystallography or nuclear magnetic resonance (NMR) providing avenues for exploration. However, due to the vastarray of diverse molecules present within a single cell, these techniques of tenfall short of fully elucidating every interaction.

The objective of in silico molecular docking is to forecast the configuration of a molecular complex starting from individual molecules. This approach is notably more convenient, cost-effective, and rapid compared to the experimental techniques [3].

IV .2 . OVERVIEW OF MOLECULAR DOCKING

Docking refers to molecular simulations that integrate various methods to investigate how two molecules interact. Docking software serves as a valuable tool in biology, pharmacy, and medicine, particularly because many active compounds are small molecules that interact with biologically relevant targets. Typically, the macromolecular receptor is a protein, and the term "docking" commonly denotes the interaction between a protein and a ligand [4].

Docking methods integrate a search algorithm to produce potential binding modes of the ligand within the receptor, known as "poses," along with a scoring function to rank these poses based on predicted affinity scores. Consequently [5], docking methods strive to discern the true ligands of the receptor among the studied molecules and ascertain the accurate poses, representing the

conformations adopted by the ligands during receptor binding [6].

Docking essentially comprises two phases: docking and scoring:

- The initial phase (docking) involves the selection process. It entails positioning the ligand within the protein's active site, sampling various conformations, positions, and orientations (poses), and retaining those that depict the most favorable interaction modes. While manual execution is feasible, this step is typically automated using docking algorithms to enhance the efficiency and precision of simulations [4].
- The second step (scoring): the scoring functions are used to mathematically estimate the binding affinity between a receptor and each of the poses generated during docking [7]. The efficiency of these scoring functions is at least as important as that of conformation al search algorithms. Indeed, even if the bioactive conformation of the ligand has been obtained during docking, if the scoring functions do not make it possible to differentiate the correct poses from the incorrect ones, the most promising compounds for the target cannot be identified [8].

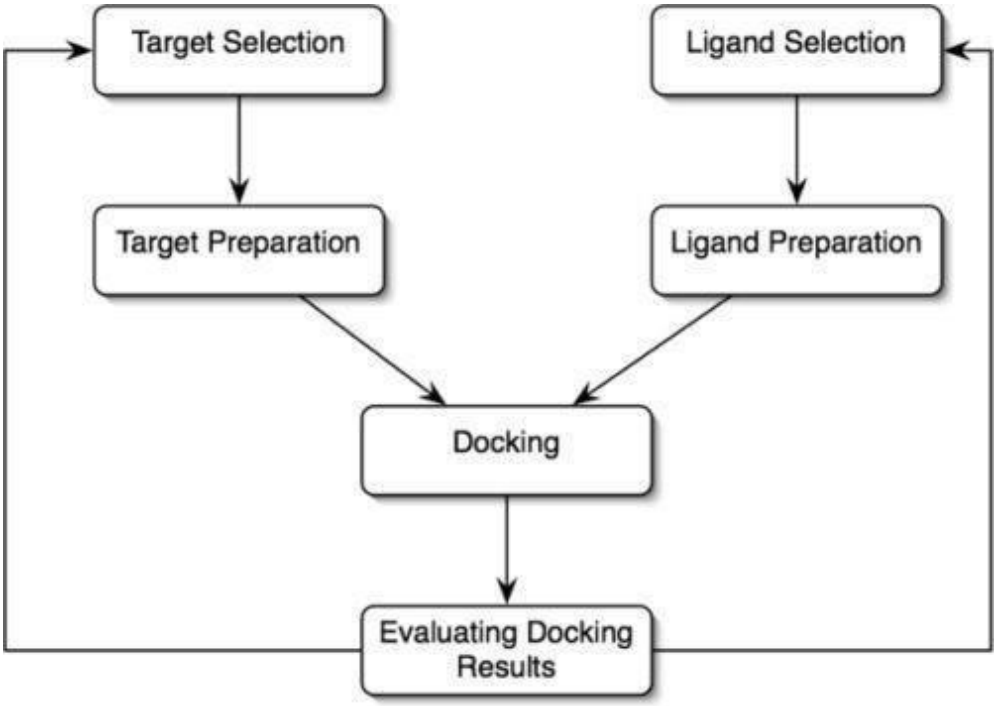


Figure IV.1. *The fundamental concept of molecular docking*

IV .2-1- Various molecular docking techniques

Molecular docking techniques can be categorized into three types, depending on factors such as ligand flexibility, the algorithms employed for pose exploration, and the scoring functions for evaluating binding affinity. Hence, we can delineate three types of docking

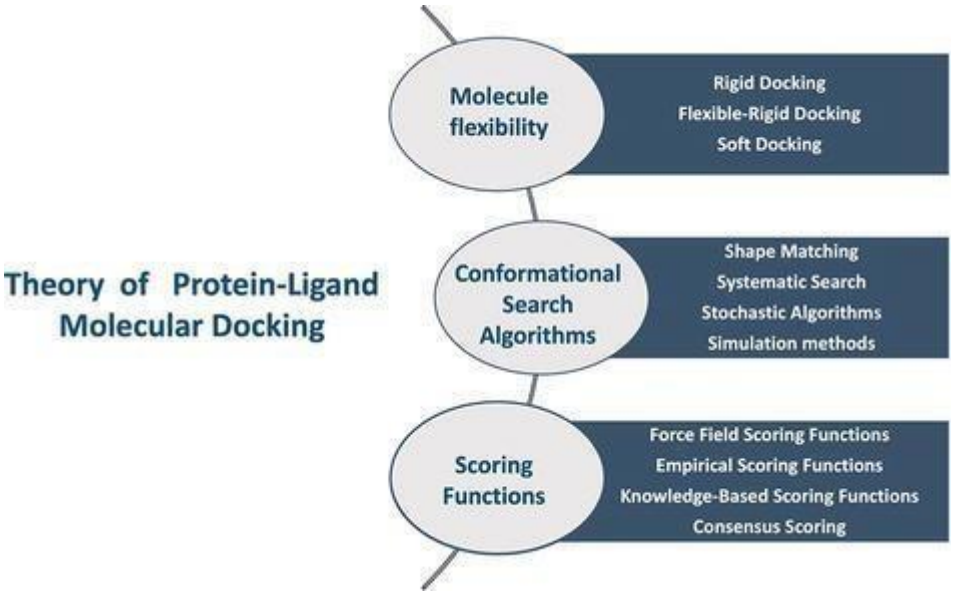


Figure IV.2. *Methods for protein-ligand molecular docking.*

IV .2.1.1.Rigid docking

Rigid docking involves considering both the protein and ligand as entirely rigid entities. This approach aligns with the concept of the "lock and key" mechanism proposed by Emil Fischer in 1890, wherein only a ligand (resembling the key) with the appropriate size and shape can interact with the protein (acting as the lock) [9].

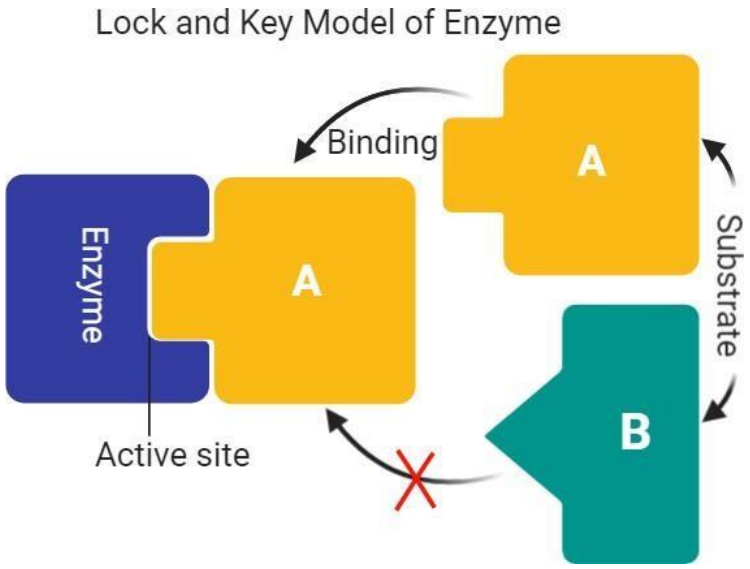


Figure IV. 3. *Diagram illustrating the principle of "Lock and key"*

IV .2.1.2. Semi-flexible docking

Semi-flexible docking involves allowing flexibility in the ligand to explore its various conformations, while keeping the target rigid through out the process. While this approach offers the advantage of shorter calculation times, it over looks any conformational changes necessary in the target [10].

IV .2.1.3. Flexible docking

This technique addresses the flexibility of both the ligand and the receptor. One significant challenge of flexible docking methods lies in navigating the ligands' conformational space to discern both the accurate poses within the protein's active site and the binding modes associated with low binding free energy [11].

IV .2.2. Molecular docking algorithms

Offer various approaches to the process. Initially, docking could be performed manually, where the modeler directly positions the ligand within the protein's active site using a graphical interface. Subsequently, the assembly's geometry is optimized to rectify steric issues and achieve an energetically stable complex. This method is suitable when the actual mode of ligand interaction is well-understood. However, in many cases, the precise mode of interaction remains unknown. In such

instances, manually exploring all ligand conformations and orientations becomes impractical, even when considering the protein as a rigid entity. To address this challenge, docking algorithms have been developed to systematically, swiftly, and efficiently search for the most favorable protein-ligand binding modes [12].

An algorithm's effectiveness in locating the accurate position of the ligand concerning its receptor is typically evaluated using the Root-Mean-Square Deviation (RMSD) of the model generated by the software. The RMSD between two poses serves as a geometric metric, quantifying the distance between the atomic positions of the experimental structure and those of the predicted structure of the ligand/binding site complex [13].

IV .2.3. Scoring function

The scoring function is a numerical value used to measure the extent to which a ligand binds to a receptor. Typically, it approximates the free energy change associated with the transition from the unbound state of the protein and ligand to the formation of a complex. The underlying thermodynamic principle is as follows: [14].

$$\Delta G = \Delta G_{complexe} - \Delta G_{\text{ligand}} - \Delta G_{\text{proteine}}$$

IV .3- MOLECULAR DOCKING STUDIES

The quality of any QSAR model is mainly related to the accuracy of the experimental results based on which it will be developed. Therefore, after QSAR analysis, it is of great importance to evaluate the collected data quality [15,16] .In this context, molecular docking studies were performed using Auto Dock Vina software [17], to re-estimate the activity of the dataset molecules as Calcium Channel antagonists by analyzing their binding energies as well as their mutual interaction types. The crystal structure of Calcium Channel in complex with amlodipine, which is a Calcium Channel blocker, was retrieved from the RCSB Protein Data Bank (PDB ID: 5KMD). Before performing docking, all ligands linked to the Calcium Channel were removed and then Kollman charges as well as polar hydrogen were added using Auto Dock Tools [18]. The docking gird box was set as follow: x = 39.626, y = 36.709, z = 14.846 at 20 Å size and 0.375 Å spacing. Furthermore, the ligands structures were optimized with the steepest Descent method using Avogadro software [19]. and then Gasteiger charges were added to the optimized structures, followed by merging polar hydrogens. Finally, the investigated ligands were docked to the Calcium Channel and the involved interactions were analyzed employing Discovery Studio 2021 software [20].

RESULTS AND DISCUSSION

After developing QSAR models, docking studies were performed to check upon the validating of the collected experimental results. The docking results were analyzed based on two parameters, namely binding energies and ligand-protein interactions. This table shows that all compounds of our series have negatives binding energies while interacting with the Calcium Channel. This is a signature of their strong affinities towards the mentioned biological target. Besides, they exhibit larger binding energies in absolute values compared to Amlodipine, which is a well-known calcium channel blockers drug. Thus the series under investigation have a favorable complexation with the Calcium Channel.

Table IV.1 :*Binding energies (BE, kcal/mol) for docked compounds of the series as ligands*

Compound	BE	Compound	BE
1	-8.3	**17**	-8.3
2	-7.9	**18**	-7.8
3	-7.8	**19**	-7.9
4	-8.7	**20**	-8.1
5	-6.9	**21**	-7.6
6	-8.4	**22**	-7.9
7	-7.7	**23**	-7.7
8	-8.3	**24**	-8.5
9	-9.0	**25**	-8.2
10	-7.2	**26**	-8.8
11	-7.4	**27**	-8.4
12	-8.0	**28**	-8.7
13	-7.8	**29**	-8.5
14	-8.1	**30**	-8.4
15	-7.8	**31**	-8.1
16	-7.9	**Amlodipine**	-5.5

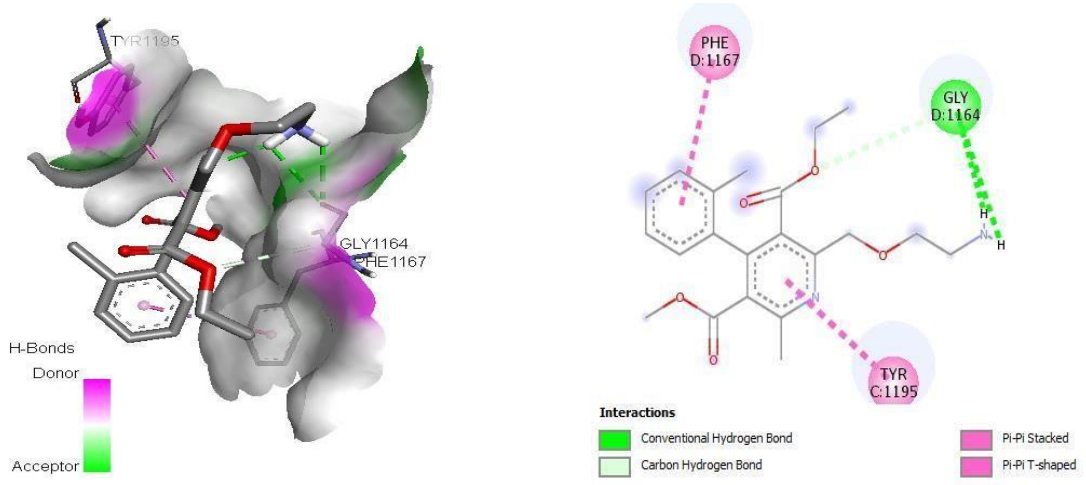


Figure IV .4:*2D (left) and 3D (right) diagrams revealing the interactions between Amlodipine and the Calcium Channel.*

At first glance, we screened the interactions between Calcium Channel and Amlodipine drug. Indeed, the key residues characterizing the active site of the Calcium Channel were identified by analyzing the interacting sites responsible for the activity of Amlodipine. According to *Figure IV .4,* Amlodipine bind to the biological target by implicating three hydrogen bonds with GLY D: 1164, and two π-π interactions with PHE D: 1167 and TYR C: 1195. Therefore, GLY D: 1164, PHE D: 1167 and TYR C: 1195 can be considered as the most potent key residues influencing the Calcium Channel activity. Afterwards, the interactions involved between the most active ligands (i.e. compounds **2**, **3** and **4**) and their biological target were analyzed, to justify their high activity values (high pIC_{50}). As depicted in *Figure IV .5*, compound **2** interacts mainly with two key residues, namely TYR C: 1195, PHE D: 1167 via a π-π interaction with TYR C: 1195 as well as a π-alkyl interaction with PHE D: 1167 in addition to other interactions with PHE D: 1171 and ILE C: 1199, justifying its activity as Calcium channel blocker. Besides, the compounds **3** was found to be complexed with the Calcium Channel by involving five alkyl interactions with ILE C: 1199, TYR C: 1195 and VAL C: 1196, a π-π interaction with TYR C: 1195 and a hydrogen bond with GLY D: 1164. Moreover, TYR C: 1195 and GLY D: 1164 with which interact compound **3**, are among the active sites influencing the Calcium Channel activity, justifying its observed activity toward the Calcium channel. Concerning Compound **4**, which is the most active compound among the collected dataset, it was able to interact with all key residues influencing the Calcium Channel activity. Indeed, it is involved in a hydrogen bond with GLY D: 1164, five alkyl interactions with TYR C: 1195, PHE D: 1167, ILE C: 1199 and VAL C: 1196, two π-πinteractions with PHE D: 1167 and TYR C: 1195, in addition to two π-σ interactions with TYR C: 1195 and PHE D: 1171.

Table IV.2: *The number of interactions involved between the most active compounds and the key residues of the Calcium Channel.*

		Interacting sites		
Compound	**pIC_{50}**	**GLY D: 1164**	**PHE D: 1167**	**TYR C: 1195**
2	10.44	0	1	1
3	10.74	1	0	2
4	10.92	1	2	3

On the other hand, by comparing the number of interactions involved between the most active compounds and the key residues (*Table IV.2*), we can provide an explanation for their enhanced biological activity against the Calcium Channel as found experimentally. Indeed, the higher the number of interactions with the key residues, the higher the activity. Therefore, the docking results corroborate and validate the biological activities exhibited *in vitro* by the collected 1, 4-dihydropyridine derivatives [21].

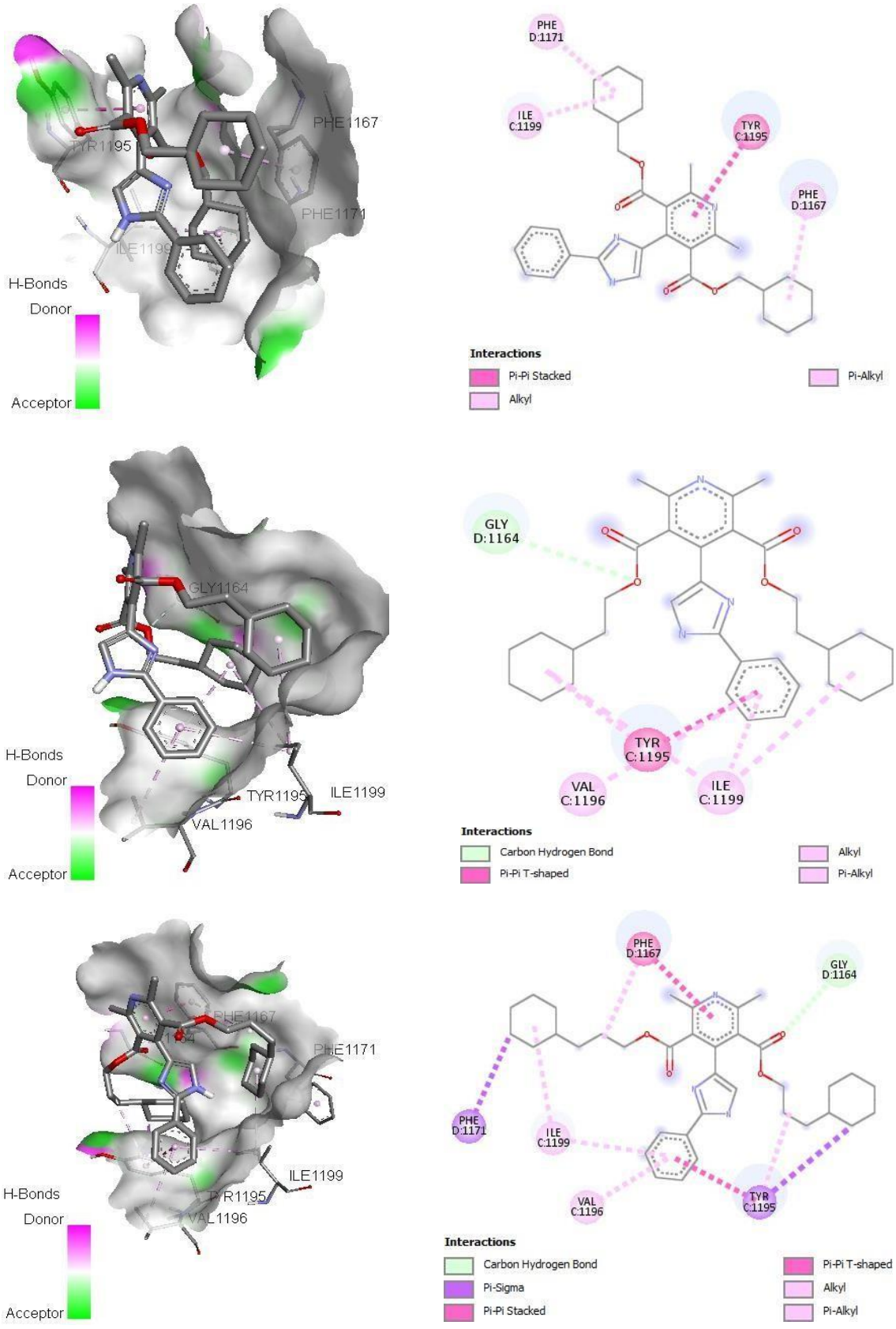


Figure . IV.5. *2D (left) and 3D (right) diagrams revealing the interactions between Calcium channel and ligand (2) (top), ligand (3) (middle) and ligand (4) (bottom).*

IV .4. CONCLUSION

Molecular docking studies were conducted using Auto dock Vina software to reassess the activity of the dataset molecules as Calcium Channel antagonists. This involved analyzing their binding energies and interaction types. The crystal structure of the Calcium Channel complexed with amlodipine (a Calcium Channel blocker) was obtained from the RCSB Protein Data Bank (PDB ID: 5KMD). Prior to docking, ligands bound to the Calcium Channel were removed, and Kollman charges along with polar hydrogen atoms were in corporated using AutoDock Tools. The docking grid box dimensions were set to x = 39.626, y = 36.709, z = 14.846 with a size of 20 Å and spacing of 0.375 Å. Additionally, ligand structures under went optimization using the steepest Descent method in Avogadro software, followed by the addition of Gasteiger charges and merging of polar hydrogens. Subsequently, the ligands were docked to the Calcium Channel, and their interactions were analyzed using Discovery Studio 2021 software.

REFERECES

[1] Asses Y (2011) Conception par modélisation et criblage in silico d'inhibiteurs du récepteur c-Met. Thèse de doctorat, Université Henri Poincaré – Nancy I.

[2] Belaidi S, Salah T, Melkemi M, Sinha L, Prasad O (2015) Structure activity relationship and quantitative structure-activity relationships modeling of antitrypanosomal activities of alkyldiamine cryptolepine derivatives, J. Comput. Theor. Nanos., 12(9) : 2421-2427

[3] Grosdidier A (2007) Conception d'un logiciel de docking et applications dans la recherche de nouvelles molécules actives. Thèse de doctorat, Université Joseph Fourier.

[4] Warren G. L, Andrews C. W, Capelli A.-M, Clarke B, LaLonde J , Lambert M. H , Lindvall , Nevins N, Semus S. F, Senger S ,Tedesco C, Wall I. D ,Woolven J. M , Peishoff C. E , Head M. S (2006) A critical assessment of docking programsand scoring functions. Journal of Medicinal Chemistry **.** 49 (20) :5912-5931.

[5] Barril X , Soliva R (2006) Molecular modelling. Molecular BioSystem s **.**2 (12) : 660-681.

[6] Lagarde N (2014) Méthodes de criblage virtuel in silico: importance de l'évaluation et application à la recherche de nouveaux inhibiteurs de l'interleukine 6. Thèse de doctorat,Paris, CNAM.

[7] Dias R , de Azevedo Jr W. F (2008) Molecular docking algorithms. Current Drug Targets **.** 9 (12) :1040-1047.

[8] Kitchen D. B , Decornez H , Furr J. R, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery **.**3 (11) :935-949.

[9] Lichtenthaler F. W(1995) 100 Years "Schlüssel-Schloss-Prinzip": What Made Emil Fischer Use this Analogy? Angewandte Chemie International Edition in English **.**33 : 2364-2374.

[10] Alvarez J. C (2004) High-through put docking as a source of novel drug leads. Current opinion in chemical biology **.** 8 (4) :365-370.

[11] Sotriffer C , Klebe G , Stahl M , Böhm H. J (2005) Docking and scoring functions/virtual screening. Burger's Medicinal Chemistry and Drug Discovery :281-331.

[12] Leach A. R , Shoichet B. K , Peishoff C. E (2006) Prediction of protein− ligand interactions. Docking and scoring: successes and gaps. Journal of Medicinal

Chemistry .49 (20) :5851-5855.

[13] Hawkins P. C , Warren G. L , Skillman A. G , Nicholls A (2008) How to do an evaluation: pitfalls and traps. Journal of Computer-Aided Molecular Design .22 (3-4) : 179-190.

[14]. Kollman P. A , Massova I , Reyes, C , Kuhn B , Huo S , Chong L , Lee, M , Lee T , Duan Y , Wang W , Donini O , Cieplak P ,Srinivasan J , Case D. A , Cheatham T. E (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Accounts of Chemical Research.33 (12) : 889-897.

[15] Ouassaf M, Belaidi S, Chtita S, Lanez T, Abul Qais F, Md Amiruddin H (2022) Combined molecular docking and dynamics simulations studies of natural compounds as potent inhibitors against SARS-CoV-2 main protease. J. Biomol. Struct. Dyn. 40: 11264-11273.

[16] Ouassaf M , Belaidi S , Lotfy K, Daoud I , Belaidi H (2018) Molecular docking studies and ADMET properties of new 1.2. 3 triazole derivatives for anti-breast cancer activity. J. Bionanosci. 12: 1-11.

[17] Trott O, Olson A.J (2010)AutoDock Vina:improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J. Comput.Chem .31 :455-461.

[18] ADT / AutoDockTools — AutoDock. http://autodock.scripps.edu/resources/adt (Accessed : 28.05. 2022).

[19] Hanwell M.D , Curtis D.E , Lonie D.C , Vandermeerschd T , Zurek E ,.Hutchison G.R (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform," J. Cheminform 4 : 17.

[20] BIOVIA Discovery Studio - BIOVIA - Dassault Systèmes®." https://www.3ds.com/products-services/biovia/products/molecular-modeling-si .

[21] Bossert F , Vater W (1971) Dihydropyridine, eine neue Gruppe stark wirksamer Coronartherapeutika [Dihydropyridines, a new group of strongly effective coronary therapeutic agents]. Naturwissenschaften. 58:578.

GENERAL CONCLUSION

GENERAL CONCLUSION

The aim of this work was to construct dependable QSAR models for predicting specific biological properties and activities of heterocyclic organic molecules, particularly those with varied structures of 1,4-dihydropyridines. A comprehensive range of molecular descriptors was computed, encompassing constitutional, electronic, topological, geometric, physicochemical, and thermodynamic descriptors..,we applied many methods of computational chemistry in this study.Quantum mechanics methods wereused in the study of chemical reactivity of 1,4 dihydropyridine and theirderivatives, with methods: PM3, Ab initio/ (HF / 6-31 + G (d, p)) and DFT (B3LYP / 6-31 + G (d, p)).The qualitative study of the structure-activity relationship was conducted on 31 compounds, all of which possess pharmacological properties. The nature of the groups attached to the basic nucleus of these molecules influences their physicochemical properties, and in turn, their pharmacological effects.
Molecular properties like membrane permeability and oral bioavailability are typically associated with basic molecular descriptors, such as molecular weight (MW), log P (partition coefficient) and the number of hydrogen bond acceptors and donors in a molecule.

In this work, the CCBs activity of a series of 1,4-dihydropyridines was predicted using MLR and ANN approaches. The external and internal consistency of the built models was confirmed using internal and external validation methods to assess their statistical quality. External validation of these models confirmed their ability to accurately predict the CCBs activity of the studied compounds. Furthermore, the MLR equation can show that physical features, organic functional groups, and chemical molecular fragments are significantly linked to the desired activity of these studied compounds. Furthermore, the molecular docking study allowed us to understand the mechanism of the interactions between the required activity and this kind of chemical compounds.

The methodology based on MLR has primarily been employed for predictionpurposes. This approach effectively generates transparent QSAR models, which are characterized by their reliability, explanatory power, predictiveness, and interpretability. By selecting pertinent descriptors, these models facilitate the explanation and interpretation of the structure-activity of the compounds under study, from both statistical and chemical perspectives.
The models presented in these applications offer insights into pharmaceutical research, aiding in the design and synthesis of novel molecules with potential to evolve into drugs.

While the primary objectives of this work have been achieved, to continue our research journey in this field, we anticipate various perspectives moving forward:

- we are in the process of employing alternative molecular modeling techniques. This includes the development of QSAR-3D models using other methods .
- We intend to use the same databases and develop models using other methods such as: genetic algorithms (GA)…
- In experiment alterms, the next steps for advancing the proposed models would involve on going efforts aimed at validating them through the synthesis of the newly proposed molecules. This process should be carried out in collaboration with other research laboratories.

Abstract

Artificial neural networks (ANNs) are useful for predicting biological activities from large datasets of molecules. Unlike traditional statistical methods such as regression analysis, ANNs allow the study of complex and nonlinear relationships such as QSAR studies. Here, we use artificial neural network and multiple linear regression (MLR) methods to generate QSAR models for Calcium Channel Blockers activity of a series of 1,4-dihydropyridine derivatives molecules. The molecular descriptors were calculated by using Density Functional Theory (DFT) method at the B3LYP/6-31G+ (d, p) level. The statistical analyses indicate that the predicted values are in good agreement with the experimental results for both the training and test sets using either MLR or ANN. In addition, we used molecular docking to determine the binding energies, and ligand-protein interactions between these compounds and their biological target.

Keywords:1,4-dihydropyridine, Calcium Channel Blockers, QSAR, DFT, ANN, MLR.

Resume

Les réseaux de neurones artificiels (ANN) sont utiles pour prédire les activités biologiques à partir de grands ensembles de données de molécules. Contrairement aux méthodes statistiques traditionnelles telles que l'analyse de régression, les ANN permettent l'étude de relations complexes et non linéaires telles que les études QSAR. Ici, nous utilisons un réseau de neurones artificiels et des méthodes de régression linéaire multiple (MLR) pour générer des modèles QSAR pour l'activité des bloqueurs de canaux calciques d'une série de molécules de dérivés de 1,4-dihydropyridine. Les descripteurs moléculaires ont été calculés en utilisant la méthode de la théorie fonctionnelle de la densité (DFT) au niveau B3LYP/6-31G+ (d, p). Les analyses statistiques indiquent que les valeurs prédites sont en bon accord avec les résultats expérimentaux pour les ensembles d'apprentissage et de test utilisant soit MLR, soit ANN. De plus, nous avons utilisé l'amarrage moléculaire pour déterminer les énergies de liaison et les interactions ligand-protéine entre ces composés et leur cible biologique.

Mots-clés : 1,4-dihydropyridine, inhibiteurs calciques, QSAR, DFT, ANN, MLR.

ملخص

الشبكات العصبية الاصطناعية ANNs مفيدة للتنبؤ بالأنشطة البيولوجية من مجموعات بيانات كبيرة من الجزيئات. على عكس الاساليب الاحصائية التقليدية مثل تحليل الانحدار، تسمح الشبكات العصبية الاصطناعية بدراسة العلاقات المعقدة وغير الخطية مثل دراسات QSAR. هنا ، نستخدم الشبكة العصبية الاصطناعية وطرق الانحدار الخطي المتعددة MLR إنشاء نماذج QSAR لنشاط حاصرات قنوات الكالسيوم لسلسلة من جزيئات مشتقات 1،4-ديهيدروبيريدين. تم حساب الواصفات الجزيئية باستخدام طريقة نظرية الكثافة الوظيفية)DFT على مستوىp(d p) + 6-31G / B3LYP،
التحليلات الاحصائية إلى أن القيم المتوقعة تتفق جيدًا مع النتائج التجريبية لكل من مجموعات التدريب والاختبار باستخدام إما MLR أو ANN. بالإضافة إلى ذلك ، استخدمنا الالتحام الجزيئي لتحديد طاقات الربط ، وتفاعلات بروتين يجند بين هذه المركبات وهدفها البيولوجي.

الكلمات المفتاحية : 1،4-ديهيدروبيريدين ، حاصرات قنوات الكالسيوم ، QSAR، DFT، ANN، MLR.