MÉMOIRE DE MASTER

Domaine : Sciences de la Matière

Filière : Chimie

Spécialité : chimie pharmaceutique

Réf. : /

Présenté et soutenu par :

Aouisset Darine

Le : 27/05/2022

IN SILICO ANALYSIS OF PYRAZOLE DERIVATIVES APPLIED TO DRUG DESIGN

Jury :

Mme.	Hanane Djouama	M.C.B	Université de Biskra	Présidente
Mr.	Salah Belaidi	Professeur	Université de Biskra	Rapporteur
Mme.	Saida Khamouli	M.C.B	Université de Biskra	Examinatrice

ACKNOWLEDGMENTS

DEDICATION

This thesis is profoundly dedicated to:

To my father, who supported me throughout my life, to my super mother, who made my academic career successful, and my three brothers.

I also make a special dedicated to my grandma (my second mother), aunt and all my little family.

TABLE OF CONTENTS

2ᴿᴰ PART: RESULTS AND DISCUSSION

Chapter III : Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation methods, and the experimental synthesis protocol

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

Chapter V : Quantitative study of the QSAR properties of a series of 1,3,5-triazine derivatives and application of chemometric methods

Chapter VI : Study by molecular docking of the interactions between the EGFR enzyme and pyrazole derivatives

LIST OF FIGURES

LIST OF TABLES

GENERAL

INTRODUCTION

Heterocycles are chemical compounds whose carbon chain, cyclic, contains one or more atoms other than the carbon "heteroatoms". They occupied an increasingly important place in organic chemistry. These molecules in fact combine in the same structure the remarkable characteristics of saturated, partially saturated or aromatic cyclic compounds and especially Nitrogen-containing heterocycles are of considerable importance in the pharmaceutical area. These types of compounds are indeed found in numerous bioactive structures patented in the last few years. [1]

Pyrazoles are reported to possess a wide range of biological activities in literature such as anti-microbial, anti-fungal, anti-tubercular, anti-inflammatory, anti-convulsant, anticancer, anti-viral, angiotensin converting enzyme (ACE) inhibitory, neuroprotective, cholecystokinin-1 receptor antagonist, and estrogen receptor (ER) ligand activity, etc. Which are as shown in figure 1.

Withasomnine

Celecoxib (anti-inflammatory)

Rimonabant (anti-obesity)

CDPPB
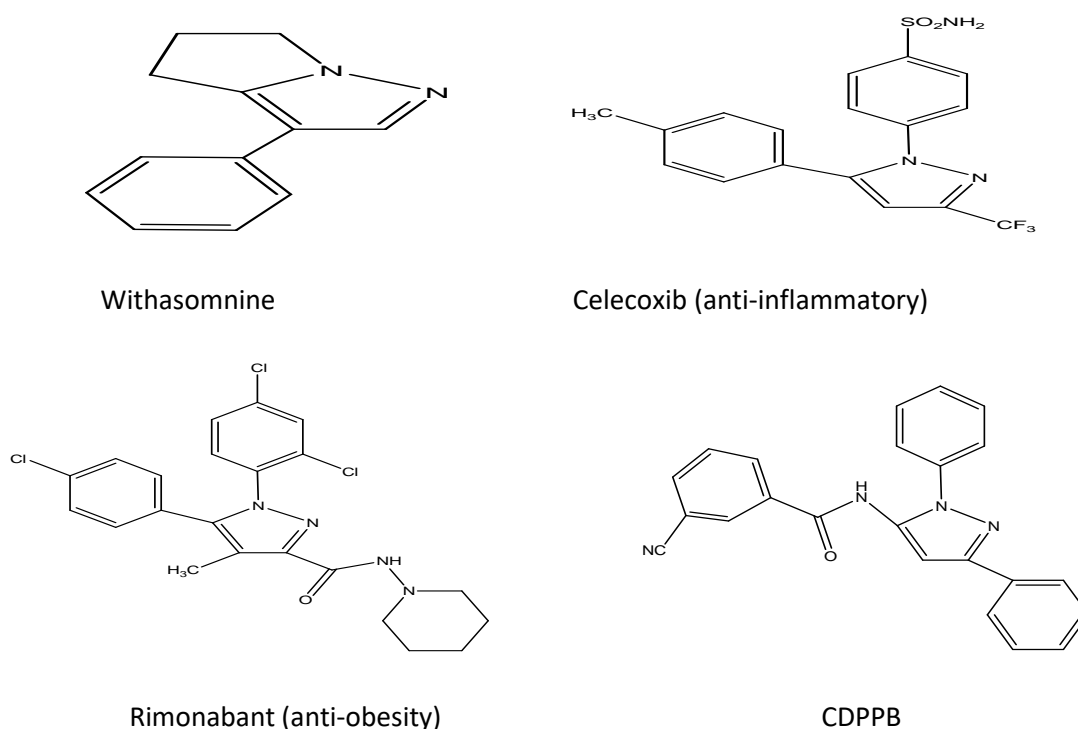
Figure 1

In our research, we report a computational study of 1,2-diazole, and a series of its derivatives and the objective of this work is the study of structural and electronic properties of pyrazole by the methods provided by computational chemistry, Correlation study between physico-chemical parameters and biological properties of bioactive molecules (derivatives of pyrazole) for the contribution to the discovery of new anti-cancer drugs and study the interactions between the bioactive molecules studied and the EGFR enzyme by MOE software.

The pyrazoles variously substituted by aromatic and heteroaromatic groups possess numerous biological activities, which makes them particularly interesting. [2]

Molecular modeling involves the use of theoretical calculation methods (*molecular mechanics, molecular dynamics, Ab initio or semi-empirical quantum mechanics for example PM3 ...*) allowing determining the graphic representation of the geometry or the configuration of atoms of a molecule and evaluating the associated physicochemical properties. [3]

The aim of our work is to study the structure and activity of pyrazole and the biological activity of the selected series and their efficacy in cancer disease (anticancer drugs), and so that our work is more organized we have divided into five chapters advanced by a *General Introduction:*
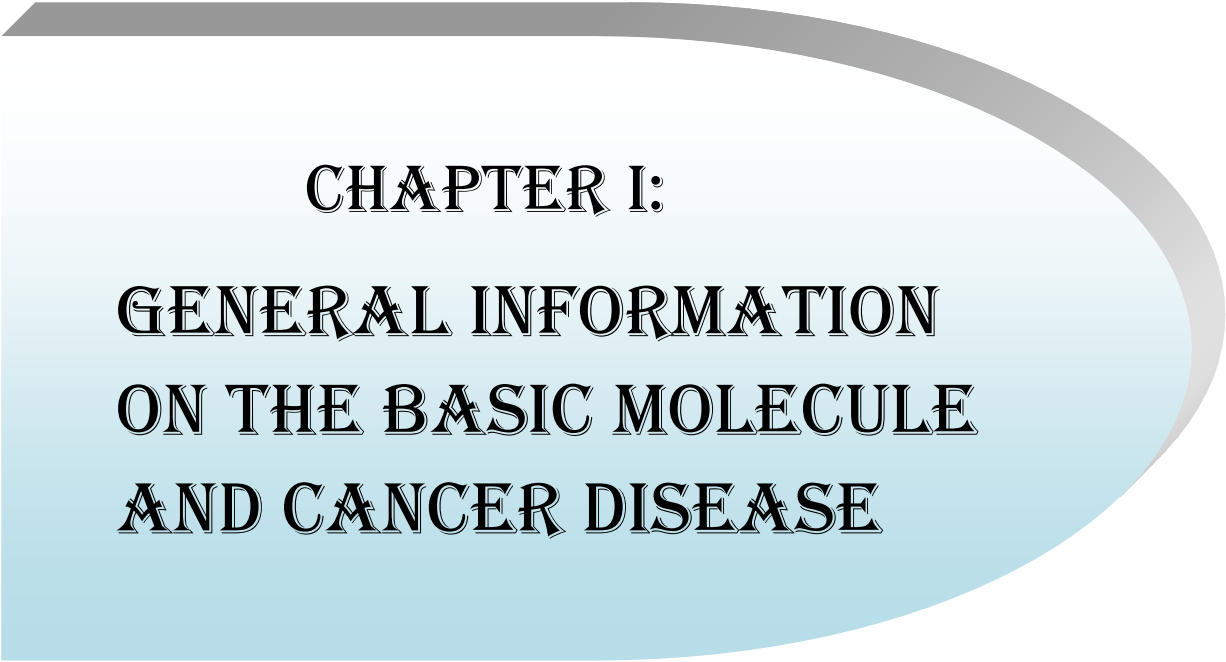
- ✤ **Chapter1:** Bibliographical research on the basic molecule, methods of pyrazole synthesis and research on cancer disease.
- ✤ **Chapter2:** Bibliographical research on molecular modeling methods used in drug selection.

- ✤ **Chapter3:** Structural study, Electronics, MESP on the basic core: 1,2-diazole, using several molecular and quantum mechanical calculation methods and the experimental synthesis protocol.

↳ **Chapter4:** Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

↳ **Chapter5:** Quantitative study of the QSAR properties of a series of 1, 2-diazole derivatives and application of chemometric methods.

↳ **Chapter6:** Study by molecular docking of the interactions between the EGFR enzyme and pyrazole derivatives.

At the end of this manuscript, we end with a ***general conclusion***, which sums up our work.

References:

[1] Sophie Régnier, doctoral thesis, University of Montréal, 2016.

[2] Memory of Thierry Delaunay, Claude Bernard University - Lyon I, 2010.

[3] S.Belaidi, doctoral thesis, University of Batna, 2002.

CHAPTER I:

GENERAL INFORMATION ON THE BASIC MOLECULE AND CANCER DISEASE

I.1. Introduction:

Pyrazole is a parent organic skeleton of the pyrazole class, was donated to this class of compounds by Ludwig Knorr in 1883 [1], And consists of an aromatic heterocycle characterized by a 5-ring structure with three carbon atoms and two nitrogen atoms in adjacent positions. Are classified among alkaloids, but this structure is particularly rare in nature.

Indeed, to our knowledge there are two natural sources of pyrazole. From watermelon seeds, the first natural pyrazole, 1-pyrazolyl-alanine, was isolated in 1959 [2]. In addition, Withasomnine has been isolated from a plant (*withania somnifera Dum*) [3], used in traditional Indian medicine for the treatment of mild disorders (analgesic and antidepressant effects).

Figure .I. 1 . The withania somnifera plant.

I.2. 1,2-Diazole Core Chemical Appearance:

I.2.1 General:

Pyrazole is a compound of the structural isomer of imidazole. The name pyrazole comes from the pyrrole ring to which a nitrogen atom has been added: "azole" (Figure 2). The two nitrogen atoms have different properties: one behaving like pyridine can undergo protonation in acidic conditions; the other has the property of pyrrole nitrogen, the doublet participating in the aromaticity of the cycle. [3]
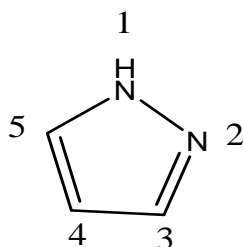
Figure .I. 2. Structure of pyrazole

While the structures of unsubstituted pyrazole were determined in 1887, and can be presented in three tautomeric forms: [1]

1H-Pyrazole   3H-Pyrazole   4H-Pyrazole
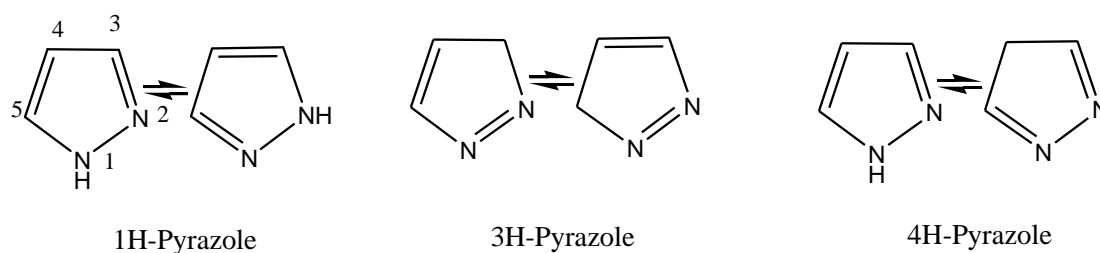
Figure .I. 3. Tautomeric forms of pyrazole

I.2.2. Synthesis of 1,2-diazole:

As we saw in the introduction, pyrazolates, which are substituted, possess several biological activities that make them particularly important.

After extensive research, it was found that Pyrazoles have occupied a unique place in the synthesis of outstanding anti-cancer agents.

The synthesis of pyrazole is also used in the preparation of larger structures such as pyrazolopyrimidines, pyrazoloisoquinolines, and pyrazolopyrazines. Which find wide use in the drug industry? [1]

The synthesis of pyrazoles is generally obtained by the action of β-diketones or their derivatives on hydrazine. **Li-ping Song et col.** Have developed an easy route of access to these compounds by the action of trifluorometyl-1,3-diketones on hydrazine. In order to enhance the value of the F-alkylhydrazinoalcohols that we have synthesized, we made them react with β-diketones in order to synthesize the corresponding pyrazole derivatives. [4]

We will first recall the methods that have been proposed in the literature to prepare many compounds of pyrazole structure have been carried out.

The importance of these molecules and related to their interesting biological activity .We present below some recent methods of preparation of compounds of pyrazole structure:

a) - we cite that of Kira et al. who reported the vilsemier-haack reaction on acetophenones /acetonehydrazones, which led to the corresponding formylpyrazoles: [5]
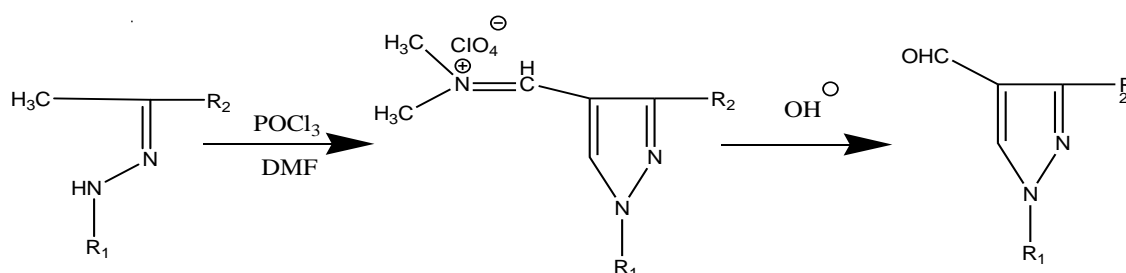
b) - The formation of a pyrazole structure was achieved by condensation of hydrazine on 1,3 diketones.
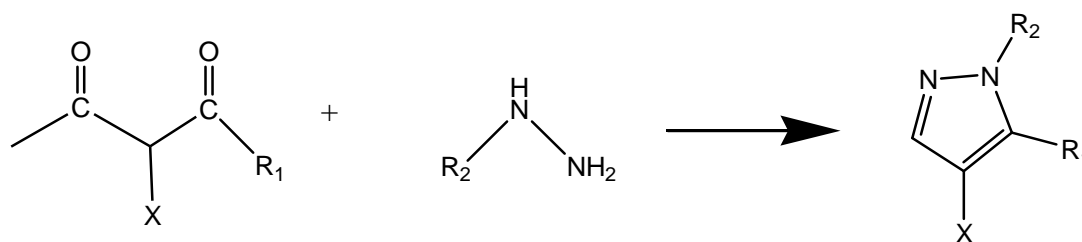
R1=Me, Oct/ R2=Ph, 4-ClPh, COPh, CO-Furyl, CO-thienyl/ X=H, Et, Cl

c)- The cycloaddition of the prop-2-enone substituted by halogens or alkoxy, amino or alkylsulfanyl groups with hydrazine lead to 3,5-diaryl-1H-pyrazoles.
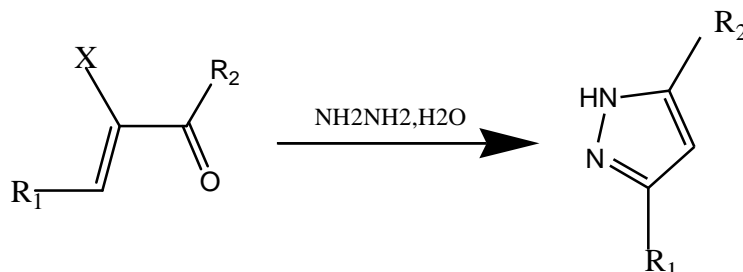
I.3. biological activity of symdiazole derivatives:

Pyrazole is the core of drug structure and the importance of this molecule and related to their biological activity interesting.

The great interest among the most famous: no steroidal anti-inflammatory, and Viagra, a peripheral vasodilator antihypertensive **[3],** marketed by Pfizer laboratories.

They are also applied as insecticides, and fungicides, and in the photography industry **[2]** .Anti-convulsant, anticancer, anti-viral, angiotensin converting enzyme (ACE) inhibitory, neuroprotective, cholecystokinin-1 receptor antagonist, and estrogen receptor (ER) ligand activity. The effect of pyrazoles on the inhibition of corrosion of steel and iron in acid medium concentrate has been studied using different electrochemical methods.

I.4. Symdiazole as medicine:

Pyrazole the most important nucleus in the pharmaceutical field which fined a wide use in the drug industry; Indeed, there are many active ingredients containing symdiazole.

The following table presents drugs containing derivatives of Symdiazole:

Table I. 1. Various biological activities of 1, 2-diazoles

The drugs	Biologic Effect	Structure
Celecoxib	Anti-inflammatory	
Rimonabant	Anti-obesity	
Sulfaphenazole	Antibacterial	
Fluazolate	Herbicide	
Penthiopyrad	Fungicide	
Fipronil	Insecticide	

I.5.Cancer and EGFR inhibitors for the treatment of cancer:

I.5.1.The cancer:

a) Generality:

Despite many advances in cancer research in recent years, this disease still remains one of the leading causes of death in developed countries and the leading cause in Denmark.

Cancer is considered a worldwide health risk nowadays. It is the second driving cause of death in the world, and it is anticipated to be the primary cause of death in the upcoming years [6]. It's a pathology characterized by the presence of one (or several) malignant tumor formed from the transformation by mutations or genetic instability (cytogenetic abnormalities), of an initially normal cell.

The invention of novel small molecules that are both potent and selective antitumor agents continues to be a serious challenge to medicinal chemistry researchers.

Current accessible treatments still have two significant limitations, the primary being the shortage of selectivity for cancer tissues, inflicting unwanted side effects. The second is the acquisition of multiple-drug resistance by cancer cells, rendering them unresponsive to standard therapy. Unwanted side effects of anticancer medication can be overcome with agents capable of discriminating tumor cells from normal cells. Hepatocellular carcinoma is taken into account because it is the third most common cancer in males, the seventh in females, and the third most prevalent cause of mortality associated with cancer round the world, and this is why early diagnosis is of great importance.

b) Symptoms and consequences of cancer:

Cancer can cause the following symptoms (most often there are no symptoms):

- A need to urinate often day and night (frequent urination).
- An urgent need to urinate (urgent urination).
- Difficulty starting or stopping urination.
- A weak or slow stream of urine.

- A stream of urine that stops.

- A feeling of having trouble emptying your bladder.

- Difficulty controlling the bladder (urinary leakage).

- Pain or burning sensation when urinating.

- Difficult or painful ejaculation.

- A presence of blood in the urine or semen (rare). [7]

I.5.2. EGFR receptor:

Epidermal growth factor receptor (EGFR) kinase has been commonly associated with cancers such as lung, ovarian, hormone-refractory prostate, metastatic colorectal, glioblastoma, pancreatic, and breast cancers.
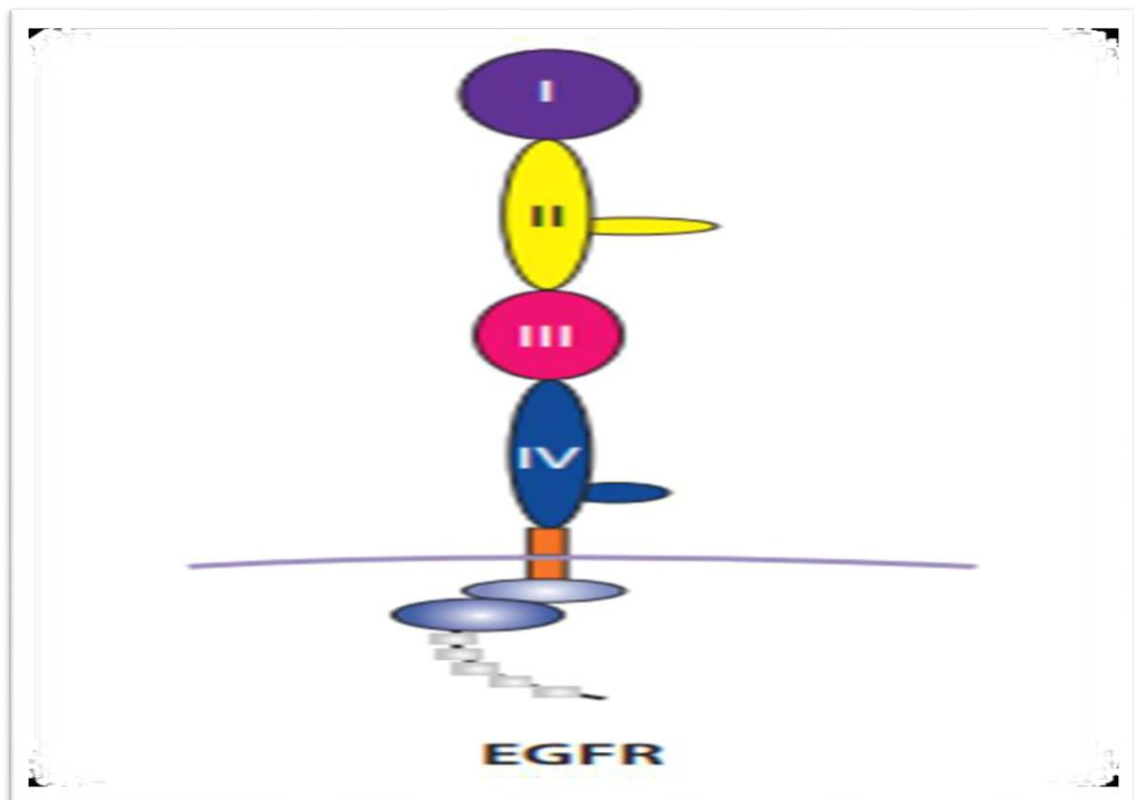
Figure. I. 4. Schematic representation of EGFR

The ERbB or epidermal growth factor (EGF) family of receptor tyrosine kinases (RTKs) is composed of four structurally related members:

EGFR/ErbB1/HER, ErbB2/Neu/HER-2, ErbB3/HER3, and ErbB4/HER4 EGF receptor family members are essential in the etiology of numerous tumors, including those of the breast, ovary, lung, colon, nervous system, head and neck, prostate, and pancreas. The EGFR is more commonly studied because it plays an important role in many vital processes such as cell proliferation, DNA damage and repair, and DNA replication and transcriptional regulation.**[8,9]**
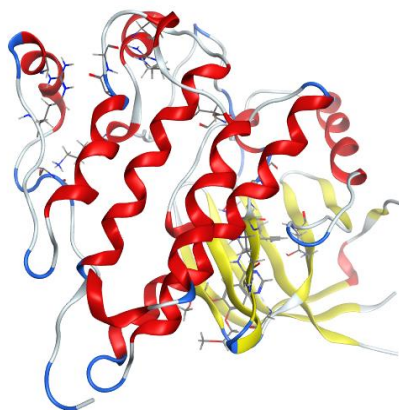
Figure .I. 5. Three-dimensional structure of the protein (PDB ID: 4hjo) with the binding site pocket of EGFR kinase. (By MOE software)

I.5.3. EGFR protein inhibitors as anti-cancer agents:

Cancer chemotherapy has entered a new era of molecularly targeted therapeutics, which is highly selective and not associated with the serious toxicities of conventional cytotoxic drugs. **[10]**

Pyrazoles are one of many significant compounds that have been studied and reported for their antitumor activity in vitro and in vivo against a broad range of cancers. **[11]**

To our knowl-edge, few reports have been dedicated to the synthesis and EGFR inhibitory activity a series of pyrazole derivatives containing thiourea skeleton as anti-cancer agents. **[10]**

On the other hand, many computational analyses for pyrazole derivatives have been reported, including quantitative structure−relationship activity (QSAR) and molecular docking studies. Sunayana et al. studied the two-dimensional (2D) and three-dimensional (3D) group-based quantitative structure−activity relationship (G-QSAR) for evaluating the activity of a set of thiazolyl-pyrazole derivatives as EGFR inhibitors using the multiple regression method and k-nearest neighbor molecular. [11]
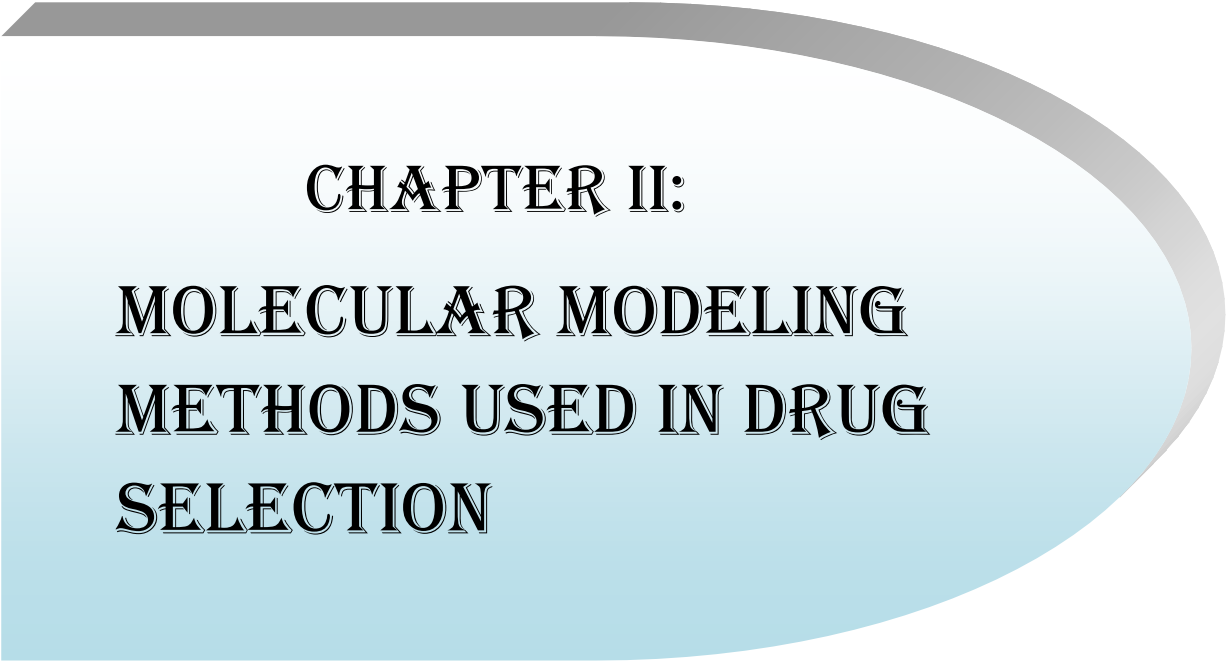
I.6. Conclusion:

In this chapter, we have devoted to a bibliographic review of pyrazole and the important methods of synthesis of pyrazole skeleton derivatives and their biological activities in particular, the anti-cancer activity.

On the other hand, to present a detailed description of the target of our interest, the EGFR enzyme, we will also present in this first chapter a state of the art on the sym-diazole based inhibitors of EGFR already known or still in clinical trials.

References:

[1] SOULEH Kheira, Master's thesis in organic chemistry, ZIANE ACHOUR University of DJELFA, 2014 / 2015.

[2] Sunil Kumar Sahu, Ravi Sharma, RNT Journal of Current Discovery in Chemistry .2017 June, 2, 15-25.

[3] Memory of Thierry Delaunay, Claude Bernard University - Lyon I, 2010 .French.

[4] Iklas Mastouri, Ahmed Baklouti Journal of the Chemical Society of Tunisia, 2008**,** *10*, 93-99.

[5] Kira M A, Abdel-Rahman M O, Gadalla K Z, Tetrahedron , 2, 1969, 109.

[6] Nashwa M. Saleh1, Marwa G. El-Gazzar, Hala M. Aly 1 and Rana A. Othman1, Journal of Frontiers in Chemistry 20 January 2020.

[7] American Cancer Society .Coping with Hair Loss. August 2020.

[8] Yarden, Y.; Sliwkowski, M. X. Untangling the ErbB signaling network .2; 127−137; 2001.

[9] Hsu, J. L.; Hung, M. C, Cancer Metastasis. 35; 575−588; 2016.

[10] Peng-Cheng Lv, Huan-Qiu Li, Juan Sun, Yang Zhou, Hai-Liang Zhu Bioorganic & Medicinal Chemistry , 18; 4606–4614; 2010.

[11] Tawassl T. H. Hajalsiddig, Abu Baker M. Osman, and Ahmed E .M. Saeed ACS Omega, 5; 18662−18674; 2020.

CHAPTER II:

MOLECULAR MODELING METHODS USED IN DRUG SELECTION

II.1. Introduction:

Drug discovery is the process through which potential new medicines are identified. It involves a wide range of scientific disciplines, including biology, chemistry and pharmacology.

Drug research, as it's called today, began when chemistry had reached the peak of its career, allowing chemical principles and theories to be applied to problems outside the scope of chemistry, and when pharmacology became an independent scientific discipline on its own. **[1]**

All the processes necessary for the development of a drug, these processes can be subdivided and divided into four phases or stages:

• The research phase.

• The development phase.

• The clinical phase.

• The marketing phase. **[2]**

The process by which new drugs are discovered and developed involves the identification of screening results with the aim of optimizing these results in order to increase affinity, selectivity (to reduce the risk of side effects), efficacy/potency, metabolic stability (to increase by half-life), and oral bioavailability.

Once a compound that meets all of these needs has been identified, the drug development process will begin prior to clinical trials and market entry.

II.2. Representation of calculation methods:

II.2.1. Molecular Mechanics:

MM is commonly applied in large systems to calculate molecular structures and relative potential energies of a molecular conformation or atom arrangement [3, 4]. The exclusion of electrons in MM is justified on the basis of Born–Oppenheimer approximation [5], which states that electronic and nuclear motions can be uncoupled from each other and considered separately. Energy differences between conformations are significant in such calculations, rather than absolute values of potential energies.

MM can simply be viewed as a ball-and-spring model of atoms and molecules with classical forces between them [6]. The MM or rather the total potential energy of a molecule is described as the sum of bond-stretching energy (E_{str}), bond angle-bending energy (E_{bend}), torsion energy (E_{tor}), and energy of interactions among unbound atoms (E_{nb}). Energy contributions of the latter constitute vander Waals (E_{vdw}) and electrostatic interactions (E_{elec}): [1]

$$E_{tot} = E_{str} + E_{bend} + E_{tor} + E_{vdw} + E_{elec}$$

$$E_{tot} = \Sigma_{bonds} K_r (r - r_{eq})^2 + \Sigma_{angles} K_\theta (\theta - \theta_{eq})^2 + \Sigma_{dihedrals} \frac{vn}{2}(1 - \cos(n\phi - r)] +$$

$$\Sigma_{i<j}[\frac{Aij}{rij^{12}} + \frac{Bij}{rij^6} + \frac{qiqj}{\sum rij}]$$

- **Force fields in molecular mechanics:**

We call force field the mathematical model representing the potential energy of a molecule in molecular mechanics. It is important to note that the force fields constitute a purely empirical approach. It designates both the mathematical equation (function of potential energy) and the experimental parameters that compose it. There are different force fields in molecular mechanics:

MM2/MM3/MM4:

MM2 is the first force field developed by Allinger et al. It was designed to beginning for simple molecules (alkanes, alkenes, unconjugated alkynes, amines,

etc.), but its improved versions MM3 (1989) and MM4 (1996) allow it to process increasingly complex organic molecules.

CHARM (Bio+):

Developed by Karplus et al, for the calculation of biomolecules .Its design is similar to that of AMBER. Although initially this force field was designed to amino acids and proteins, it is now about other bimolecular.

AMBER:

AMBER (Assisted Model Building with Energy Refinement), was written by Kollman the field is set for proteins and nucleic acids (UCSF, 1994). It has been used for polymers and for other small molecules.

II.2.2. Quantum mechanics:

Quantum mechanics describes the electronic cloud in the form of orbitals whose shape reflects the probability of the presence of each electron in space. This description in the form of orbitals makes it possible to describe and understand how atoms come together to form molecules or solids.

Thus, the molecular energies are calculated using the Schrödinger equation with the molecular orbital (M.O.) formalism. The Schrödinger equation of a molecular system can be solved by introducing approximations.

Molecular mechanics uses the following approximations:

- Each atom constitutes a particle.
- An atom is considered as a rigid sphere having a radius and a determined load.
- Energies are calculated by formulas derived from mechanics classic.

$$H=T+V$$

Where **H** is the Hamiltonian operator (sum of kinetic energy), **T** the potential energy, and **V** the operator .**H** can also be defined as:

$$H= [-\frac{h^2}{8\pi^2}\sum i\,\frac{1}{mj}(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2})] + \sum i \sum j(\frac{e_i e_j}{r_{ij}})$$

These methods are techniques for solving the Schrödinger equation of systems with many electrons. QM methods include Ab initio [29] density functional theory (DFT) [30-32] and semi empirical calculations.

They use data fitted to experimental results in order to simplify the calculations[7] .The semi-empirical methods contain a default minimum basis (STO-3G) on the other hand the pure quantum methods use different bases and correlations according to the type of calculation. The HyperChem software provides these semi-empirical (CNDO, INDO, MINDO3, PM3, RM1...) and quantum (Ab-initio and DFT) methods.

- **Density functional theory:**

Density functional theory (DFT) is presently the most successful quantum mechanical modeling method used in physics and chemistry to compute the electronic structure (principally the ground state) of many-body systems, in particular atoms, molecules, and the condensed phases. In chemistry, DFT is used to predict a variety of molecular properties, such as molecular structures, vibrational frequencies, atomization and ionization energies, electric and magnetic properties, reaction paths, etc. The modern DFT calculations are based on two Hohenberg and Kohn theorems, which proves that the electronic energy of a molecule in a ground state could be determined completely by electron density $\rho(\mathbf{r})$. The electron density $\rho(\mathbf{r})$ can be defined as in following Equation, where r is spatial variable of electrons and s is the spin variable of electrons. [1]

$$P(r) = N \, \Sigma s_1 \ldots \Sigma \, s_N \int dr_2 \ldots \int r_N |\Psi \, (r_1, \, s_1, \, r_2, \, s_2 \ldots . r_N, \, s_N) \, |^2 \int P(r)dr$$

II.3. Quantitative structure-activity relationships (QSAR):

II.3.1. Introduction:

Quantitative structure – activity relationship (QSAR) modeling pertains to the construction of predictive models of biological activities as a function of structural and molecular information of a compound library.

The concept of QSAR has typically been used for drug discovery and development and has gained wide application for correlating molecular information with not only biological activities but also with other physicochemical properties, which has therefore been termed quantitative structure – property relationship (QSPR). [1]

QSAR is widely accepted predictive and diagnostic process used for finding associations between chemical structures and biological activity.

QSAR appeared and evolved in trying to respond to the need and desire of the specialist chemist to predict the biological response [8]. It found its way into the practice of agro chemistry, pharmaceutical chemistry, and eventually most facets of chemistry [9]. The final outputs of QSAR computations are set of mathematical equations relating chemical structure to biological activity.

QSAR models are not only used for the prediction of properties but are also helpful in selection of alternative mechanism of action, determination of useful structural characteristics, projecting new design methodologies and help in proposing new hypotheses for future research work. [1]

Generates a HyperChem module that allows you to calculate a number of QSAR physico-chemical properties of molecules (the descriptors); these calculations generally add up the contributions of the atoms and are very fast. [10]

II.3.2. Molecular Descriptors:

A. Physico-Chemical Descriptors:

Constitutional descriptors capture properties of the molecules that are related to elements constituting its structure. Examples of constitutional descriptors include molecular weight, Surface molecular, Volume molecular, Polarizability, Hydration Energy, Refractivity, and LogP

B. Hydrophobicity Descriptors:

Hydrophobicity descriptor is an important group of descriptors that are widely used in drug design and discovery as they can be applied to modeling both

pharmacodynamic (receptor binding) and pharmacokinetic properties (e.g. the uptake and distribution of a xenobiotic relying on partitioning through biological membranes). Partition coefficient (log P), distribution coefficient (log D) and aqueous solubility (log S) are important hydrophobicity/hydrophilicity descriptors. [11]

II.3.3. Validation of QSAR Models:

After the model equation is obtained, moreover the stability and the goodness of fit of the model, it is also significant to estimate the power and the validity of the model before using it to predict the biological activity. Validity is to establish the reliability and significance of the method for a particular use. Therefore, validation of a QSAR model must be done. There are two validation methods used for a QSAR model: internal and external validation techniques to establish the confidence and strength of the model. In general, QSAR modeling involves a systematic process with multiple steps. These include dataset preparation, molecular descriptors selection and generation, mathematical or statistical models derivation, model training and validation using a training dataset and model testing on a testing dataset. [1]

II.3.4. Applications of the QSAR study:

Some QSAR studies seem to be little more than academic studies, there are a large number of applications of these models such as [12]:

- Optimization of pharmacological activity.

-The rational design of many other products such as surfactants, perfumes, dyes and fine chemicals.

-Identification of hazardous compounds in early stages of product development or projection of inventories of existing compounds.

-Prediction of toxicity and side effects of new compounds.

-The prediction of toxicity for environmental species.

-The selection of compounds with optimal pharmacokinetic properties, either stability or availability in biological systems.

-The prediction of a variety of physico-chemical properties of molecules.

- Prediction of the fate of molecules released into the environment.

-The prediction of the combined effects of molecules, whether in mixtures or formulations. **[2]**

II.4. Molecular Docking:

II.4.1. Introduction:

Molecular "docking" consists in predicting the structure or structures of the complexes formed between an active molecule and a protein.

Calculation algorithms randomly generate a large number of possible orientations to find "the best way to insert" the molecule into a protein (at the level of the receptor or the active site).

The program takes into account all the degrees of freedom of the molecule (translation and rotation). And For each possibility the energy is calculated in Molecular Mechanics, thus taking into account all the ligand-receptor interactions (Van der Waals, bonds H, hydrophobicity, etc). A "score" is thus obtained to estimate the best ligand/receptor interaction receiver (binding).

II.4.2. Scoring Fonctions:

Scoring function is the most important component in structure based drug design for evaluating the efficacy of ligands binding to their target proteins **[13]**.

The docked poses are ranked and evaluated using scoring functions that approximate the binding free energy of a ligand to a receptor, which is a crucial step to differentiate correct poses from incorrect ones. **[1]**

The scoring functions make various assumptions and simplification in the evaluation of binding free energy for modeled complexes and do not fully consider some of the physical phenomena that are important for molecular recognition, i.e., entropic effects, as the scoring functions must be calculated rapidly during the docking run. Normally the scoring functions are expressed as a sum of separate terms that describe the various contributions to ligand binding. **[1]**

A large number of scoring functions are available, such as force filed-based, empirical and knowledge-based scoring functions, which differ in which terms that are included in the expression of the binding free energy. Terms expressing no bonded interactions, including Van der Waals interactions and electrostatic interactions, and solvation effects are commonly included **[14]**.

II.5. Conclusion:

This memory work was carried out within the computational and pharmaceutical chemistry team.

The study of the electronic and structural properties of pyrazole and its derivatives was carried out by molecular modeling (molecular mechanics, Ab-initio, and QSAR), using HyperChem (8.0.7) and Gaussian (09) software in a PC and in an HP 6000 computing station.

Qualitative study of polar area (PSA) log D and number of rotary bonds using Marvin Sketch software. And For the molecular docking study, Moe 2014 software was used.
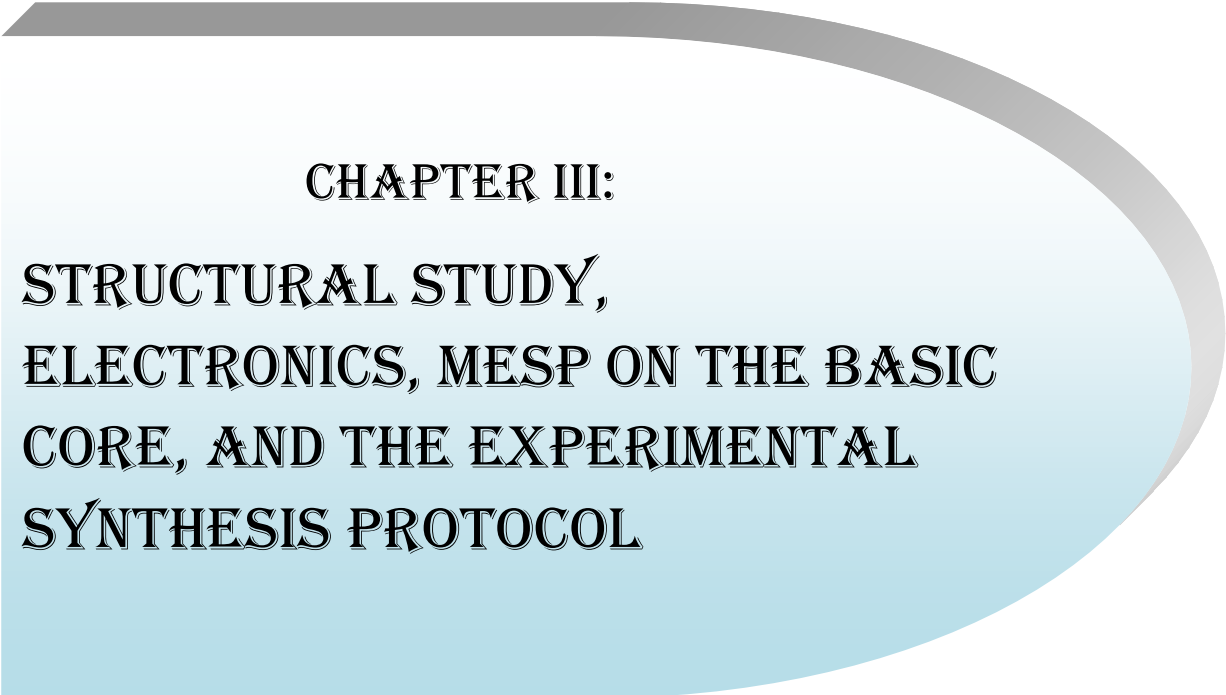
References:

[1] S.khamouli, doctoral thesis, University Mohamed Khider of Biskra, 2019.

[2] Fatouche Maroua, Master's thesis in pharmaceutical chemistry, Mohamed Khider University of Biskra, 2016**.**

[3] D.W. Rogers. 3rd ed. Hoboken (NJ): Wiley; 2003.

[4] J, C. Tully, Theor Chem Acc; 103; 3; 173–176; Feb. 2000.

[5] M. Hotokka, In: Chalmers JM, Griffiths PR, editor's .Handbook of Vibrational Spectroscopy. Hoboken (NJ): Wiley; 2006.

[6] D. J. Tannor, Sausalito, Calif: University Science Books, 2007.

[7] Salah Toufik, Med Khider Biskra University, 2012-2013.

[8] A. F. A. Cros, Université de Strasbourg (1538-1969), and Faculté de médecine, [s.n.], Strasbourg, 1863.

[9] L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, and P. Gramatica, *Environ Health Perspect*, vol. 111; 10; 1361–1375; Aug, 2003.

[10] HyperChem help.

[11] N. K. Mahobia, R. D. Patel, N. W. Sheikh, S. K. Singh, A. Mishra, and R. Dhardubey," *Der Pharma Chemica*, vol. 2; 5; 260–271; 2010.

[12] Jerzy leszczynski, department of chemistry, Jackson state university, U.S.A, 2010.

[13] R, Wang, Y. Lu, and S .Wang." *J.Med. Chem* .46; 12; 2287–2303; Jun. 2003.

[14] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, *Nat Rev Drug Discov*, 3; 11; 935– 949; 2004.

[15] Prasad RY, Kumar PP and Kumar RP, *Eur. J. Chem,* 5: 144-148; 2008.

[16] Jevwon S, Liv CT, Sao LT, Weng JR and Ko HH, Eur. J. Med. Chem, 40: 103-112;2005.

[17] Claisen, L.; Claparède, A. Berichte der Deutschen Chemischen Gesellschaft, *14;* 2460–2468; 1880.

[18] Schmidt, J. G. Berichte der Deutschen Chemischen Gesellschaft .14; 1459–1461; 1881.

CHAPTER III:

STRUCTURAL STUDY, ELECTRONICS, MESP ON THE BASIC CORE, AND THE EXPERIMENTAL SYNTHESIS PROTOCOL

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation methods, and the experimental synthesis protocol

2022

✚ First Part :

III.1. Introduction:

The development of a drug is not only complex but also the returns on investment are not always those desired or anticipated? This translates into a decreasing number of drugs being marketed. There is therefore a need to improve the traditional drug development process in order to facilitate the availability of new products to the patients who need them. In addition, it was intended to use the principles of advanced molecular modeling and clinical trial simulations.

The enormous development of computer science from the end of the twentieth century to the present day and the rise of intensive computing have made molecular modeling a real challenge **[1].**

La modélisation moléculaire implique l'utilisation des méthodes de calcul théorique (mécanique moléculaire, mécanique quantique et semi empirique) permettant de déterminant la géométrie d'une molécule et évaluer les propriétés physico-chimiques associées**[2].**

There are many methods of theoretical chemistry aimed at determining the physical or chemical properties of isolated molecules, quantum chemical methods that allow to accurately determine the electronic properties of molecules **[3]**, and play an important role in obtaining molecular geometries and predicting various properties **[4]** , on the other hand molecular mechanics methods which are based on empirical parameters which make it possible in particular to determine the structural parameters**[3].**

Two molecular orbitals, called frontier orbitals, play a role particular:
HOMO (Highest Occupied Molecular Orbital) translates the character electron donor (nucleophile) of the molecule. The more the energy of this OM is high, the easier the

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation methods, and the experimental synthesis protocol

2022

molecule will give up electrons. LUMO (Lowest Unoccupied Molecular Orbital) translates the character electro-acceptor (electrophile) of the molecule. The more the energy of this OM is weak, the easier the molecule will accept electrons [5].

In this way, it turned out over time that the electronic delocalization between HOMO and LUMO generally became the main factor determining the ease of a chemical reaction and the stereo-selective path, independently of intra- and intermolecular processes [6].

We are interested in this chapter in a structural study of the basic core 1,2-diazole. In addition, we have studied in detail the structural and electronic parameters of the privileged conformation of the basic core of pyrazole.

The optimization process (the overall minimum energy condition) was applied using the molecular mechanics calculation method, force field MM + using the Polak-Ribière algorithm has as criterion a root of the mean square of the gradient equal to 0.1 kcal. mol^{-1}. The obtained structure was re-optimized using the PM3 method by the semi-empirical method.

	MM+	PM3
Pyrazole	E=19.649	E=-898.362
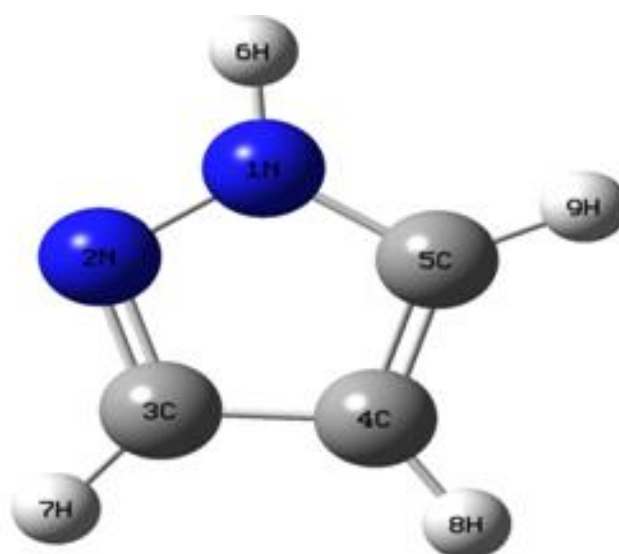
Figure .III. 1. Structure 3D of pyrazole.

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation methods, and the experimental synthesis protocol

2022

Structural and electronic comparison of the basic core:

III.2. Study of the basic core of 1,2-diazole:

To determine these structural parameters and properties, calculations have been carried out which target the following characteristics:

• The charges of each atom by the DFT method and the Ab initio method (Mulliken's Charges). **Table III.1**

• The distances between the linked atoms. **Table III.2**

• The valence angles formed by three atoms linked. **Table III.3**

• Dihedral angles formed by four successive atoms. **Table III.4**

By using different theoretical calculation methods:

- Molecular Mechanics:

• (MM+); (HyperChem8.0.7 software).

-Quantum mechanics:

• Semi-empirical: PM3 (HyperChem8.0.7 software).

• Ab initio/MP2: (Medium (6-31G*), cc-pVDZ) (Logiciel Gaussian09).

• DFT /B3LYP: (Medium (6-31G*), cc-pVDZ) (Logiciel Gaussian09).

	Ab initio/MP2	DFT/B3LYP
Energy stable	-141060.0621	-141857.2753

Table III. 1. The net atomic charges of pyrazole.

Atom	Ab initio/MP2		DFT/B3LYP	
	6-31G*	**cc-PVDZ**	**6-31G***	**cc-pVDZ**
N1	-0.163	-0.304	-0.176	-0.003

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation   methods, and the experimental synthesis protocol

2022

N2	-0.264	-0.638	-0.209	-0.187
C3	0.028	0.569	-0.198	0.117
C4	-0.350	0.446	0.057	-0.111
C5	0.163	0.657	-0.223	0.129
H6	0.409	-0.003	0.317	0.099
H7	0.224	-0.202	0.139	-0.013
H8	0.221	-0.272	0.140	-0.030
H9	0.246	-0.253	0.153	-0.002

Table III. 2. Calculated bond length values of pyrazole.

Distance	Ab initio/MP2		DFT/B3LYP		EXP [7]
	6-31G*	**cc-pVDZ**	**6-31G***	**cc-pVDZ**	
N1-N2	1.394	1.350	1.350	1.346	1.351
N1-C5	1.383	1.367	1.334	1.334	1.332
N2-C3	1.375	1.357	1.359	1.360	1.360
C3-C4	1.401	1.399	1.383	1.385	1.374
C4-C5	1.425	1.417	1.415	1.417	1.417
N1-H6	1.009	1.013	1.008	1.012	1.002
C3-H7	1.082	1.088	1.081	1.089	1.082
C4-H8	1.083	1.088	1.079	1.086	1.080
C5-H9	1.083	1.088	1.079	1.087	1.083

Table III. 3. Calculated values of the valence angle of pyrazole

Angle	Ab initio/MP2		DFT/B3LYP		EXP [7]
	6-31G*	**cc-pVDZ**	**6-31G***	**cc-pVDZ**	
N1-N2-C3	113.1	114.1	113.2	113.5	113.0
N1-N2-C5	103.3	103.7	104.2	104.0	104.1
C4-C5-N1	111.8	111.8	111.9	112.1	/

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation methods, and the experimental synthesis protocol

2022

C3-C4-C5	105.8	104.8	104.5	104.4	104.5
N2-C3-C4	106.0	105.6	106.2	106.0	106.4
N2-N1-H6	118.5	118.5	118.9	119.1	118.4
N1-H6-C5	128.4	127.5	127.9	127.4	/
C5-H9-N1	119.1	119.1	119.5	119.4	/
C3-H7-C4	129.2	129.1	128.6	128.5	/
C4-H8-C5	127.4	128.2	128.2	128.3	128.7
C3-C4-H8	126.7	127.0	127.3	127.3	127.6
C4-C5-H9	131.8	132.5	131.9	132.1	/
N2-C3-H7	122.2	121.8	121.9	121.9	121.4

Table III. 4. Calculated values of the torsion angle of pyrazole

Angle	Ab initio/MP2		DFT/B3LYP	
	6-31G*	**cc-pVDZ**	**6-31G***	**cc-pVDZ**
C5-N1-N2-C3	0.0239	0.0221	0.0149	0.0139
N2-N1-C5-C4	-0.0115	-0.0098	-0.005	-0.004
N1-N2-C3-C4	-0.0273	-0.0259	-0.0191	-0.0185
N2-C3-C4-C5	0.0216	0.0208	0.0166	0.0166
C3-C4-C5-N1	-0.0058	-0.0065	-0.0066	-0.0072
H6-N1-N2-C3	180.0135	180.013	180.0098	180.0094
N2-N1- C5-H9	-180.0124	-180.0108	-180.007	-180.0064
H6-N1-C5-C4	180.0002	180.0002	180.0007	180.001
H6-N1-C5-H9	-0.0007	-0.0008	-0.0013	-0.0015
H9-C5-C4-C3	-180.0048	-180.0053	-180.0043	-180.0044
H9-C5-C4-H8	0.0019	0.0014	0.0013	0.0014
H8-C4-C3-H7	-0.0043	-0.0046	-0.0041	-0.0039
H8-C4-C5-N1	-179.9991	-179.9998	-180.0011	-180.0015
N2-C3-C4-H8	180.0148	180.014	180.0109	180.0108
H7-C3-C4-C5	180.0024	180.0022	180.0015	180.0019
H7-C3-N2-N1	-180.0103	-180.0094	-180.0056	-180.0054

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation   methods, and the experimental synthesis protocol

2022

1. Interpretation of the results:

The theoretical results obtained by the different methods of semi-empirical and quantum calculation (Ab-initio, DFT) with different bases to note a good correlation between the calculated and experimental values for the geometric parameters.

> **For the atomic distances:**

According to Table III.2 we see that the variation of Ab initio [0.000-0.051] between the results obtained by the two bases and the experimental values. On the other, hand that the variation in the DFT method [0.000-0.011] between the results obtained by the two bases and the experimental values.

However, the variation between the values calculated by two bases of the DFT method values Medium (6-31G*) [0.001-0.009] and cc-pVDZ [0.000-0.011] compared to the experimental.

> **For the valence angles:**

According to Table III.3 the difference between the experimental values and the values calculated by the two bases of the Ab initio method [0.1˚-1.3˚], are very large compared to the values of the two DFT bases [0.0˚-0.7˚].

In addition, the difference between the experimental values and the values of two separate bases of the DFT method are: Medium (6-31G*) [0.0˚-0.5˚] and cc-pVDZ [0.1˚-0.7˚].

After having compared the results obtained by the two methods Ab initio and DFT by different bases, we see that the values are close, but the most precise are the results of the base Medium (6-31G*) of the DFT method.

> **For net charges:**

According to Table III.1 the atoms N1, N2, C3 and C5 have negative charges which favor the electrophilic attack, and the atom C4 have positive charges which favor the nucleophilic attack.

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation methods, and the experimental synthesis protocol

2022

The important aspect of the frontier electron theory is the emphasis on the highest occupied and lowest vacant molecular orbitals (HOMO and LUMO), instead of thinking about the total electron density in a nucleophile, we should be thinking to the location of the HOMO orbital because the electrons in this orbital are freer to participate in the reaction. Similarly, frontier orbital theory predicts that a site where the lowest unoccupied orbital is located is a good electrophilic site. [8]

III.3. Study of the effect of substitution on the pyrazole base core skeleton:
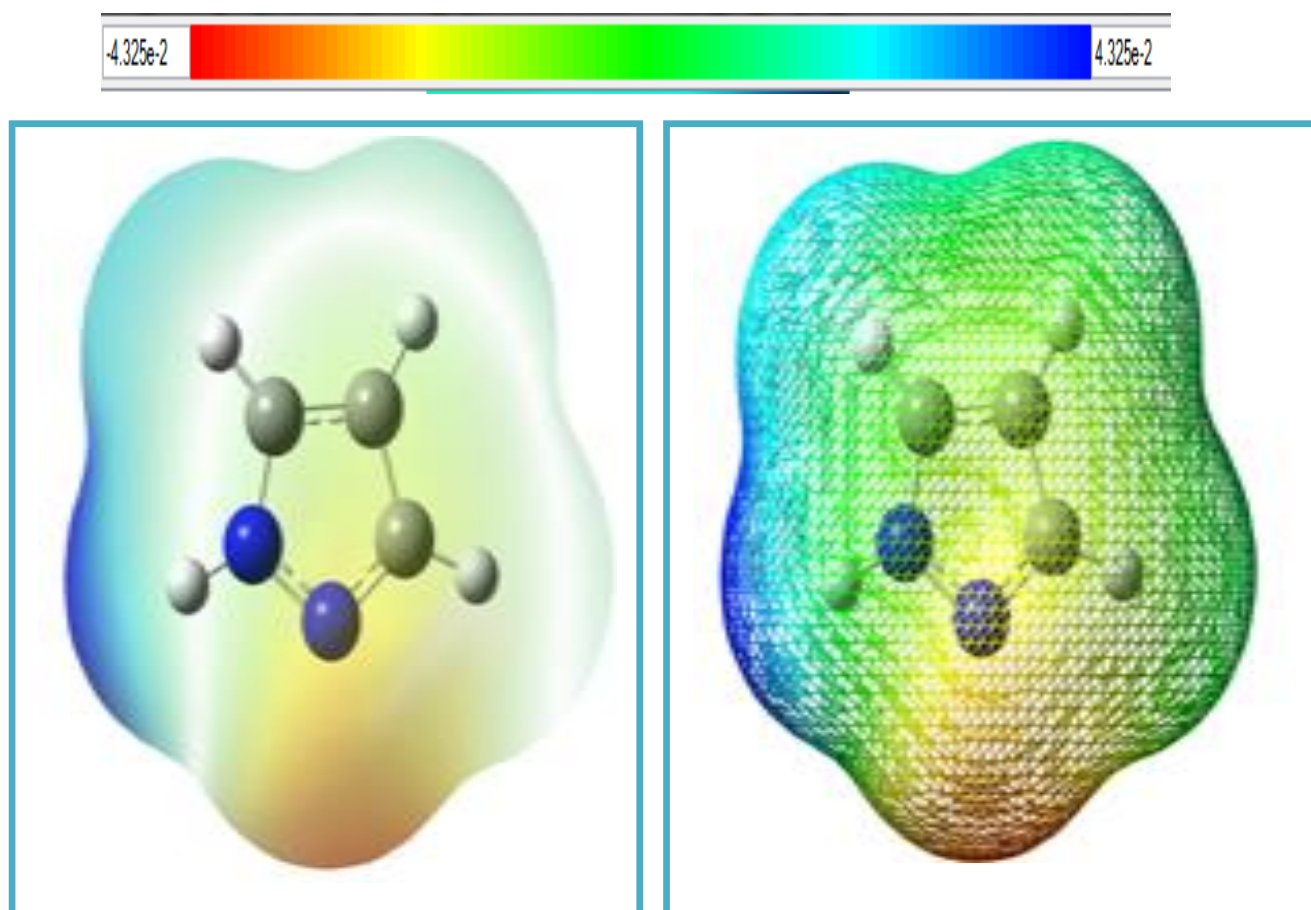
Figure .III. 2. 3D MESP surface map for pyrazole

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation   methods, and the experimental synthesis protocol

2022

The MEPS map of 1,2-diazole (Figure III.2) suggests that there is an electron-rich region (red to yellow) around the N2, C3 nitrogen atom of the heterocyclic ring, which accepts electrophilic attack . On the other hand, the blue color of the four hydrogens characterizes an electrophilic region poor in electrons. The green color that surrounds the C4 atom characterizes the neutral region.

Second Part :

In this part, we will prepare a synthesis of a pyrazole derivative from chalcone. The chalcones are characterized by their possession of a structure in the witch to the aromatic rings A and B are linked by the aliphatic chain of three carbons [15].

 The chalcones are natural biocides and are also intermediates well-known principles in the synthesis of heterocyclic compounds [16] .The synthesis of chalcone which could lead to a reversible reaction (hydrolysis). This reaction is usually carried out under reflux.
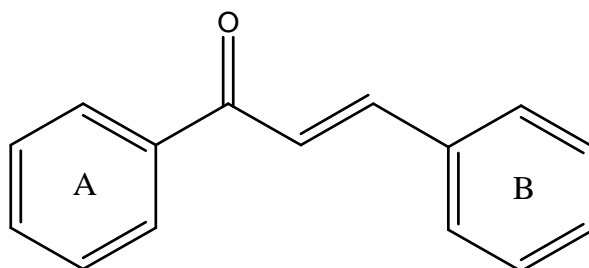
Figure .III. 3. General structure of chalcone

The work consists in preparing, first, the chalcone from acetophenone. And then prepare pyrazole from chalcone and hydrazine.

Chalcones were synthesized by a catalyzed **Claisen-Schmidt** condensation reaction of appropriately substituted acetophenones and aldehydes.

The Claisen-Schmidt reaction is a condensation reaction between a ketone enolizable and a non-enolizable carbonyl compound (example: aromatic aldehyde) in presence of a base or an acid to form α, β-unsaturated ketone with a high chemo

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation   methods, and the experimental synthesis protocol

2022

selectivity. This reaction has been applied to the preparation of chalcones, flavanones and other biologically active compounds. **[17, 18]**

These reactions are named after Rainer Ludwig Claisen and JG Schmidt, who published independently on this subject in 1880 and 1881. **[18]** The method is attractive since it specifically generates (E)-isomer.

III.4. Synthesis of chalcone:
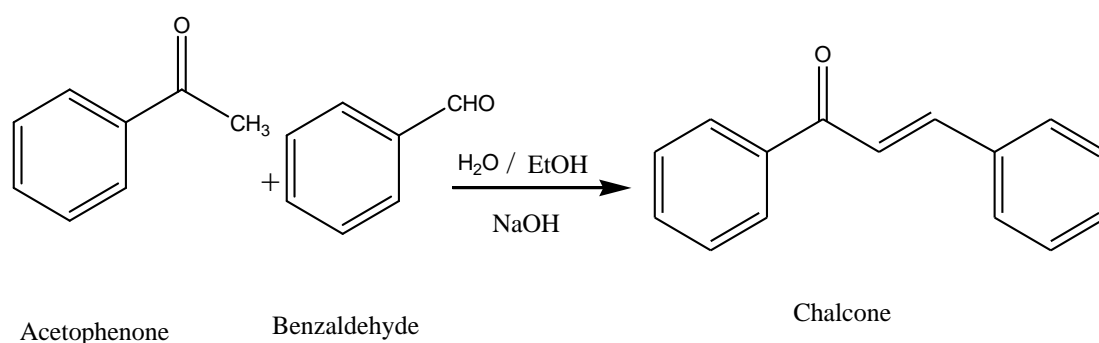
Figure .III. 4. The General Chalcone Synthesis Reaction

Procedure 1:

- In a 250 ml three-necked flask, dissolve 2.5 g of potassium hydroxide in 15 ml of water.
- Then add 10 ml of absolute ethanol. Add 6 g of acetophenone then 5.3 g of pure benzaldehyde.
- Fit a reflux condenser. Heat in a bain-marie and stir for 1 hour.
- Cool the reaction mixture in an ice bath, while stirring.
- The chalcone is rushing. After 30 min, scrape the sides of the balloon with a glass rod if the product is oily.
- Filter the crystals obtained on Buchner.
- Wash with ice cold distilled water until pH neutral, then with a little of ice cold ethanol.
- Dried the product then recrystallize from ethanol.

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation   methods, and the experimental synthesis protocol

2022
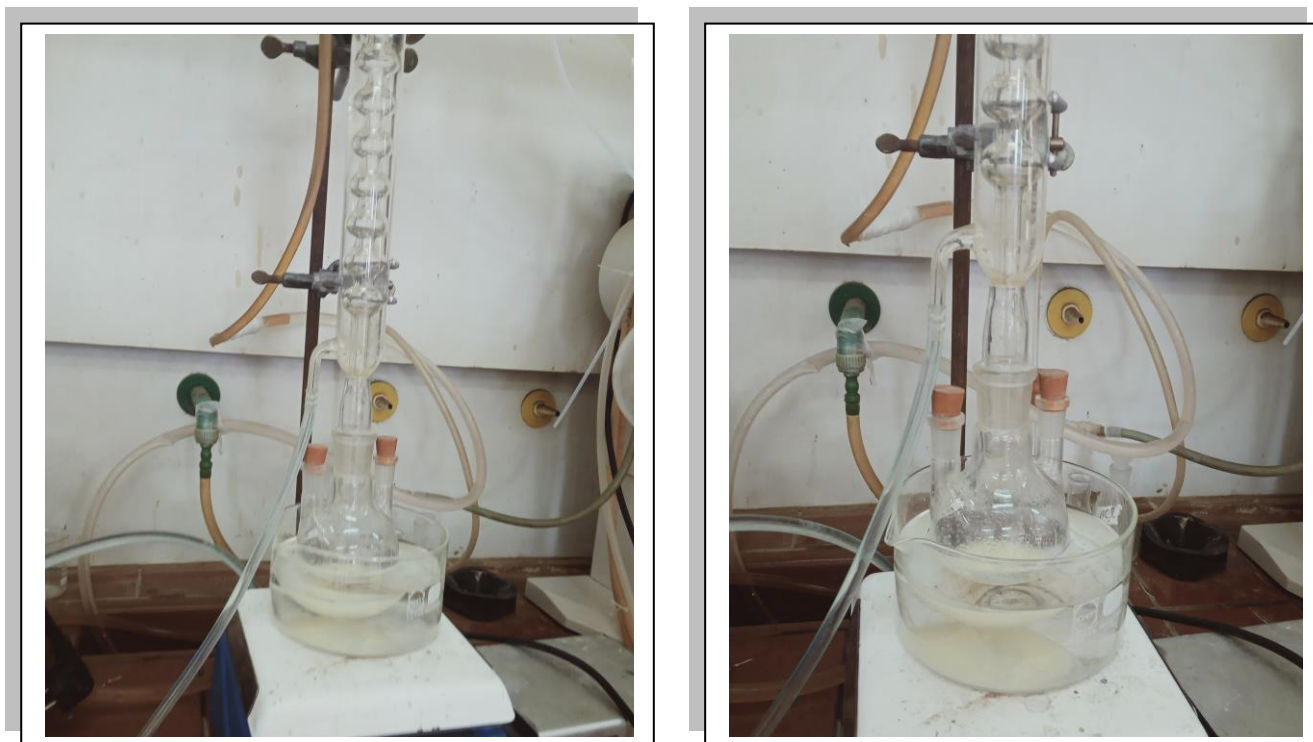
Figure .III. 5. The chalcone reaction assembly

After the synthesis passes the product to IR (Infrared Spectroscopy) analysis:
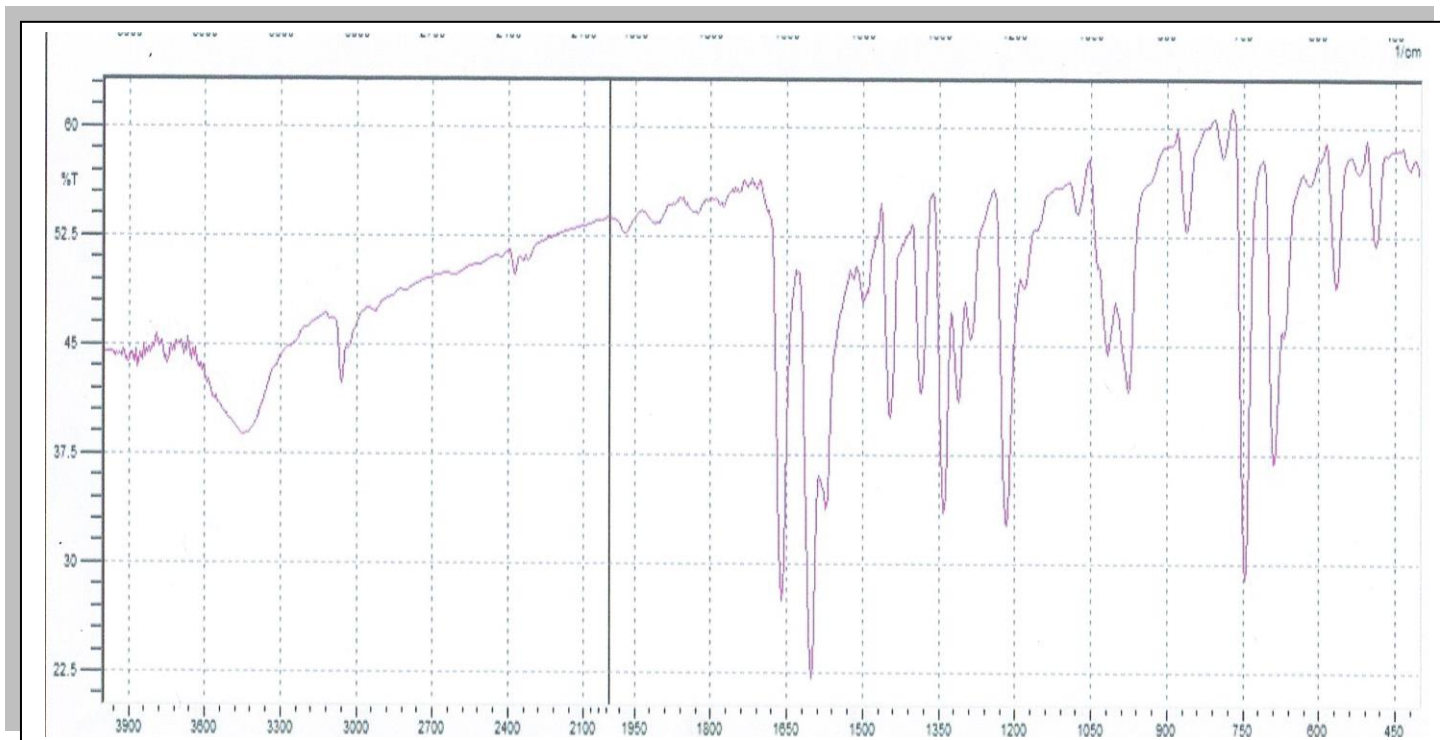
Figure .III. 6. IR spectral results of chalcone

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation methods, and the experimental synthesis protocol

2022

According to the IR spectrum of chalcone (figure 4) we can draw the bands following characteristics:

- ✓ The thin band of high intensity at 750 cm^{-1} corresponding to the vibrations aromatic deformation (C-H) .

- ✓ The fine band of average intensity at 990 cm^{-1} corresponding to the vibrations deformation (C-H) alkene.

- ✓ The fine high intensity bands at 1590-1560 cm^{-1} corresponding to the stretching vibrations (C=C) of aromatic double bonds.

- ✓ The thin band of high intensity at 1660 cm^{-1} corresponding to the carbonyl stretching vibrations (C=O).

- ✓ The thin band of low intensity at 3050 cm^{-1} corresponding to the elongation vibrations of (C=C) alkene.

III.5. Synthesis of pyrazole

Chalcone   +   phenylhyrdazine   → (NaOH / EtOH) →   1,3,5-triphenyl-4,5-dihydro-1H-pyrazole
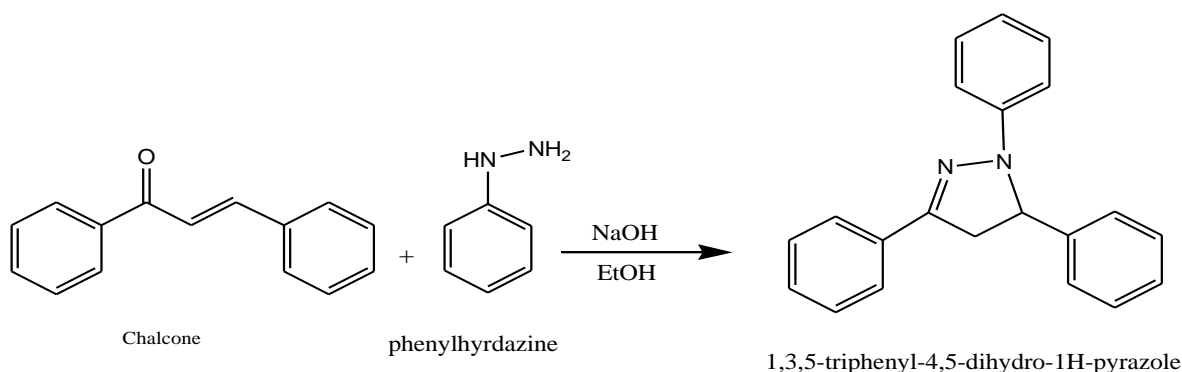
Figure .III. 7. IR spectral results of chalcone

Procedure 2:

- In a 250 ml flask, place 2 g of NaOH in 30 ml of ethanol. Then 4.16 g (0.02mol) of the chalcone and 2 g (0.02mol) of thiosemicarbazide were added.

- The mixture is stirred and refluxed for 8-10 h. At the end of this time, the reaction mixture is cooled in an ice bath.

- The DAP and DDP pyrazoles, which are orange and brick red in color respectively, are recrystallized from a water-ethanol mixture.

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation   methods, and the experimental synthesis protocol

2022
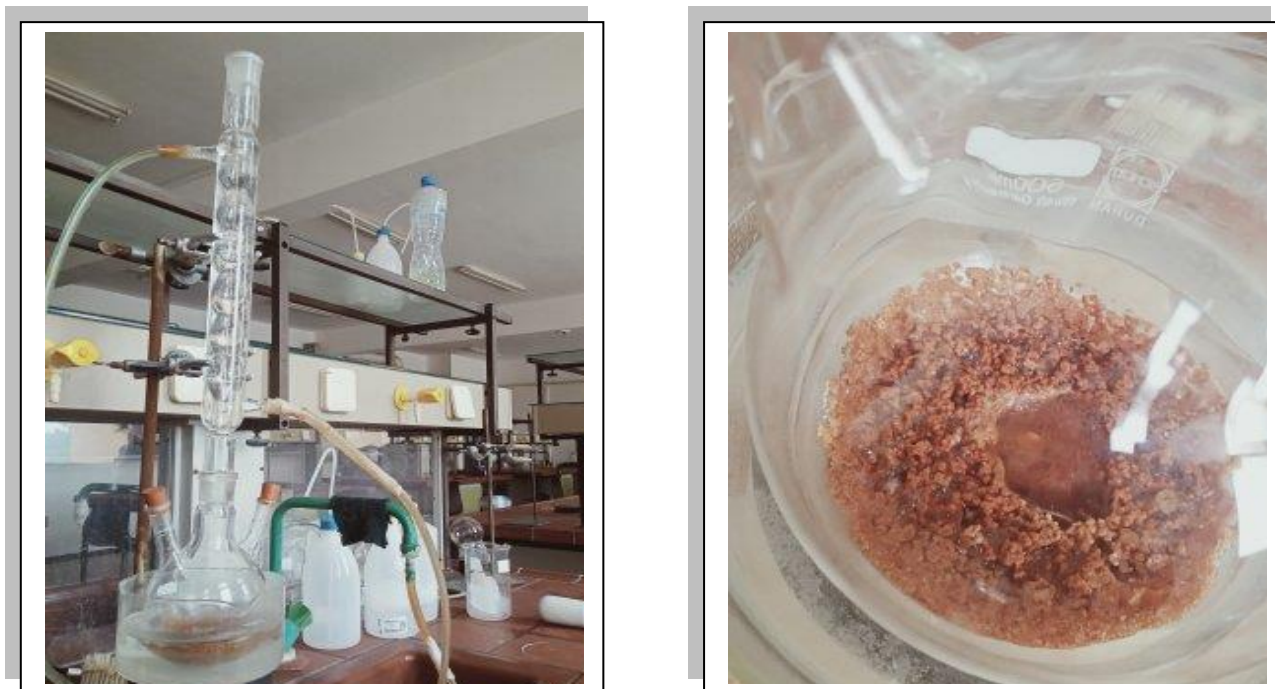
Figure .III. 8. The reaction mixture and the pyrazole derivative product

After the synthesis passes the product to IR (Infrared Spectroscopy) analysis:
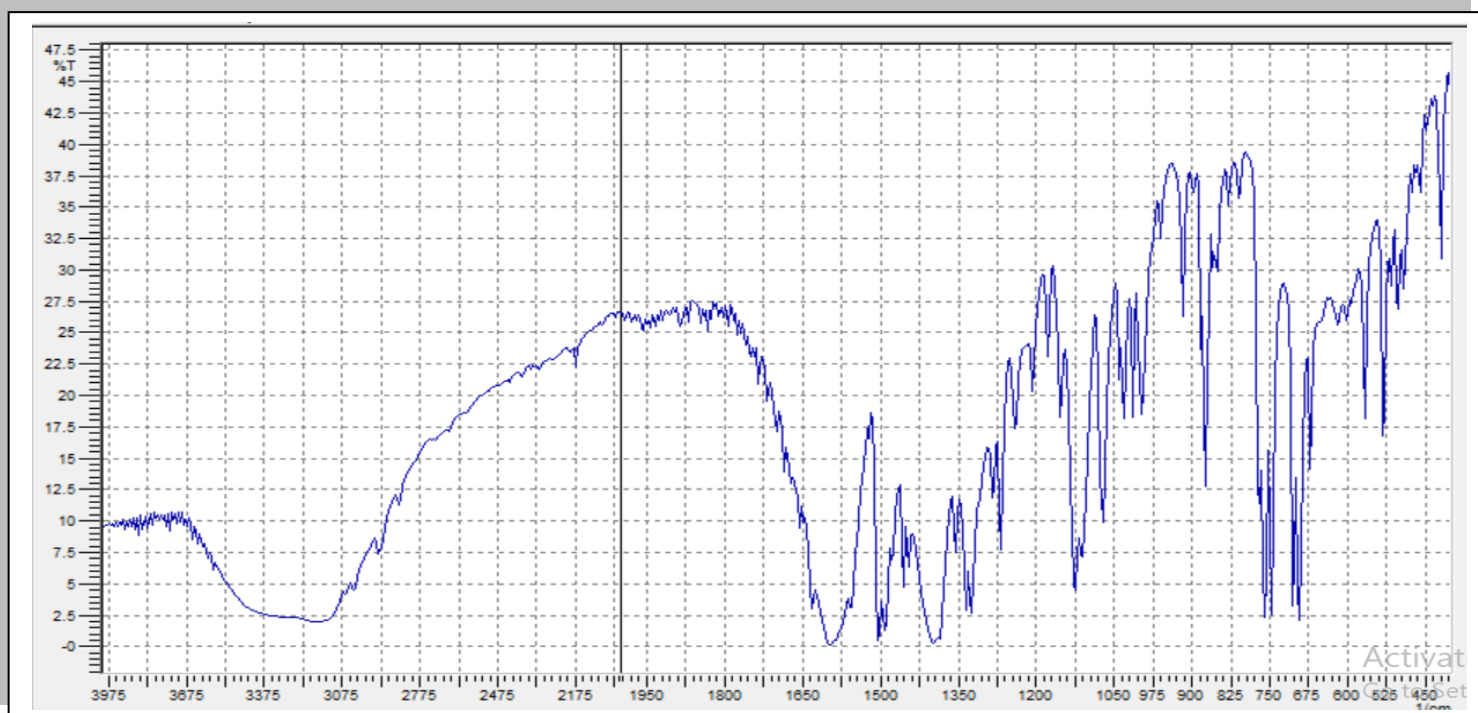
Figure .III. 9. IR spectral results of pyrazole derivative

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation   methods, and the experimental synthesis protocol

2022

According to the IR spectrum of 1, 3, 5-triphenyl-1H-pyrazole (figure 7) we can draw the bands following characteristics:

- ✓ On the infrared spectra of compounds, we note the disappearance of the frequency characteristic vibration of the (C=O) function.
- ✓ A Broad band of strong intensity at [3075-3375] cm^{-1} corresponds to the valence vibrations (O-H) function, which is due to the humidity of KBr.
- ✓ The fine band of strong intensity at [1600-1616] cm^{-1} corresponds to the valence of the vibrations (C=N) function.
- ✓ The fine band of strong intensity at 1133 cm^{-1} corresponds to the valence of the vibrations (C-N) function.
- ✓ That's which attests to the formation of the 1H-pyrazoline cycles.

III.6.Conclusion:

In the first part of the chapter, we carried out a structural and electronic comparison with the different quantum theoretical methods (Ab initio, DFT) by different bases. In addition, it was noticed after the comparison, that there is a similarity between the obtained calculation results and the experimental results. However, The (DFT) with base Medium (6-31G*) method is the most suitable method for doing calculations on the pyrazole nucleus. We found that:
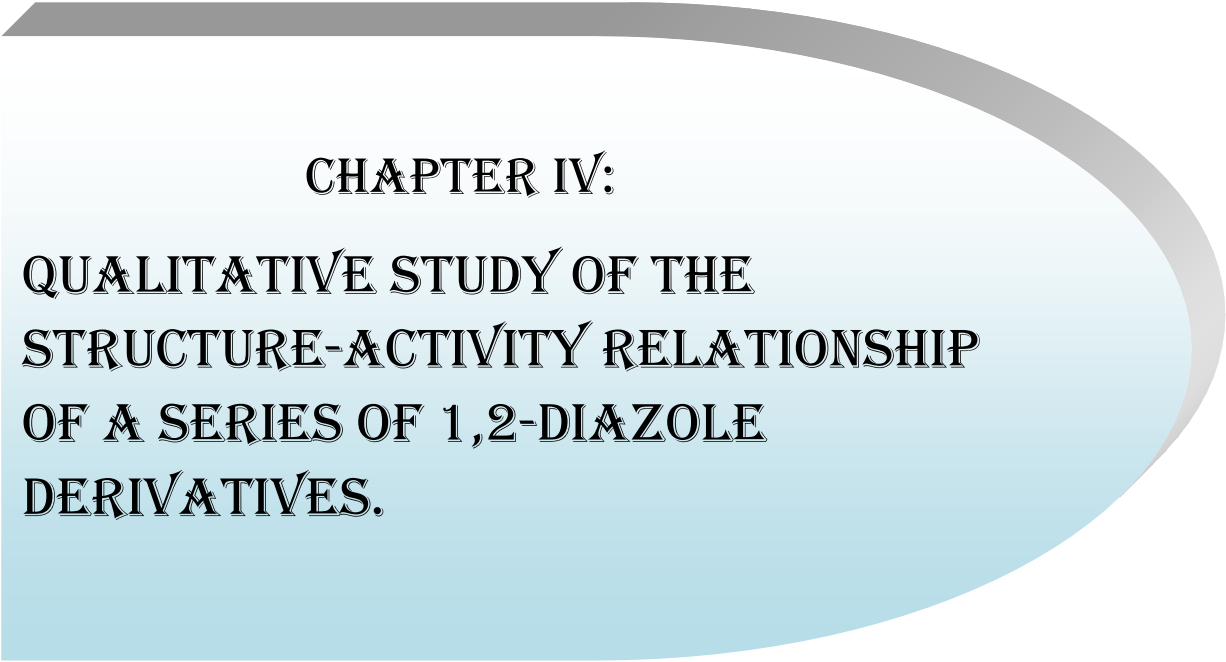
- ✓ Pyrazole has a perfect planar structure.

In the second part of this chapter, we made the synthesis 1,3,5-triphenyl-4,5-dihydro-1H –pyrazole, one of pyrazole derivatives from chalcone the intermediary of this reaction.

After finding the method (DFT) closest to the experimental results, we will begin the next chapter by making in-depth calculations on the drifts of the base nucleus 1, 2-diazole.

Chapter III: Structural study, Electronics, MESP on the basic core: 1, 2-diazole, using several molecular and quantum mechanical calculation   methods, and the experimental synthesis protocol

2022

References:

[1] P. Bultinck, Computational Medicinal Chemistry for Drug Discovery, Dekker, New York, **2004.**

[2] HyperChem 8.0.3(Molecular Modeling System) HyperChem, Inc., 1115 NW 4th Street, Gainesville, FL 32601; USA, **2002.**

[3] N. Melkemi, Thèse de Doctorat, Université de Biskra, **2013**.

[4] E. R. Davidson, Chem. Rev, **1991,** 91- 649.

[5] N. Trong Anh, Frontier orbitals: practical manual, Inter Editions / CNRS Editions, 1995.

[6] K. Kenichi, Department of Hydrocarbon Chemistry, Kyoto University, Sakyo-ku, Kyoto, Japan, 1981.

[7] Piercarlo Fantucci, Venanzio Valenti, J. CHEM. SOC. DALTON .TRANS. Milano, Italy 1992.

[8] Imane Benbrahim, "Study of the QSAR Properties of a Series".

CHAPTER IV:

QUALITATIVE STUDY OF THE STRUCTURE-ACTIVITY RELATIONSHIP OF A SERIES OF 1,2-DIAZOLE DERIVATIVES.

<div align="center">

IV.1. Introduction:

</div>

QSAR, sometimes referred to as a quantitative structure-to-property relationship, is the process by which a chemical structure is correlated with a specific effect such as biological activity or chemical reactivity.

The study of the structure-biological activity relationship of certain molecules makes it possible to establish correlations between the structural parameters and the properties of a molecule. [1]
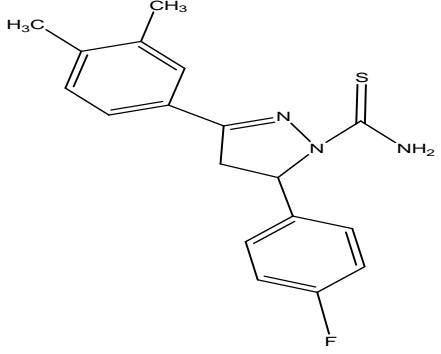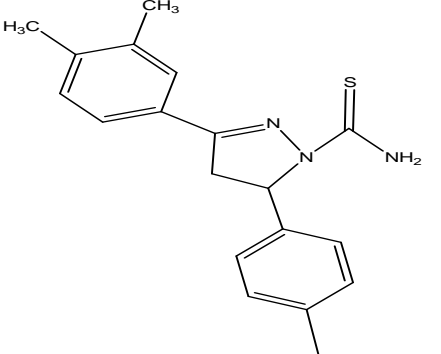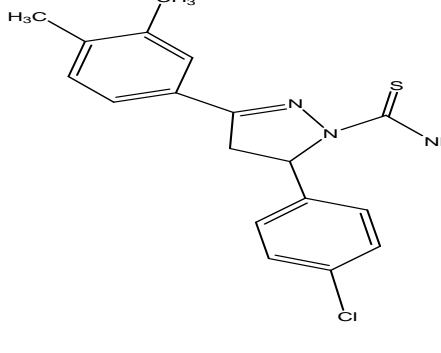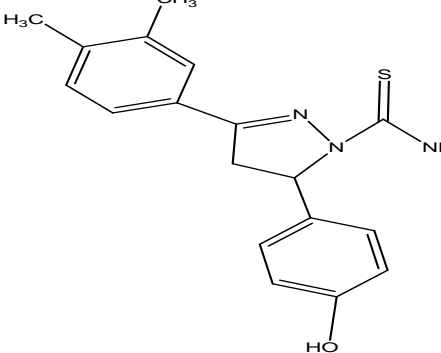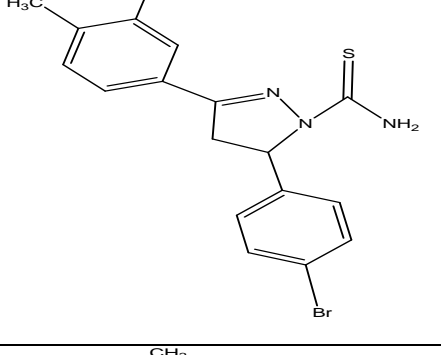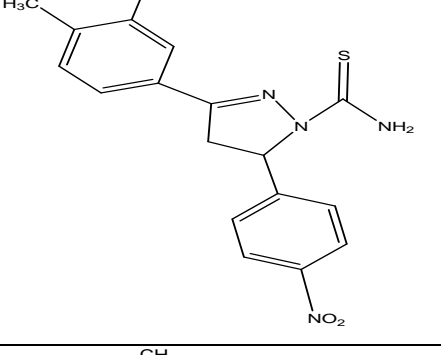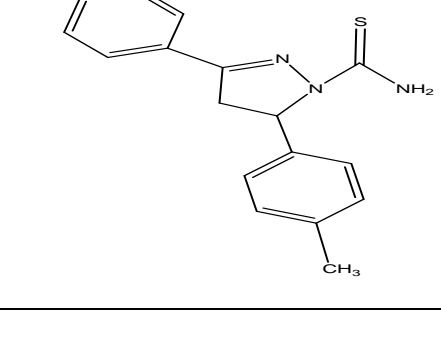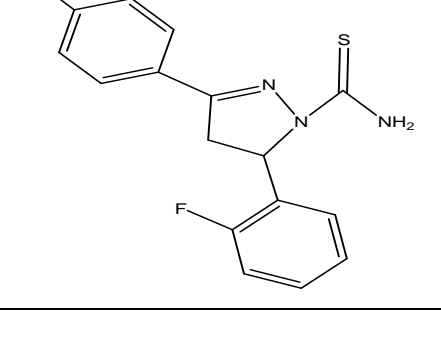
So a quantitative structure-activity or property relationship (QSAR or QSPR) is a methodology consisting in highlighting a quantitative relationship between a macroscopic magnitude (biological activity, physico-chemical property) and the molecular structure of the compounds studied. [2]
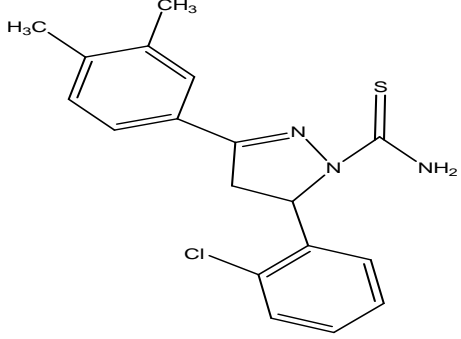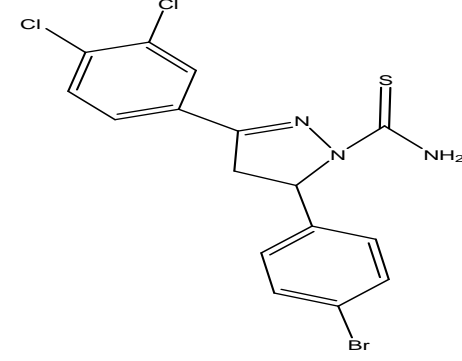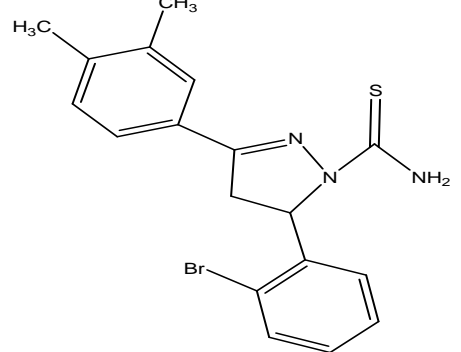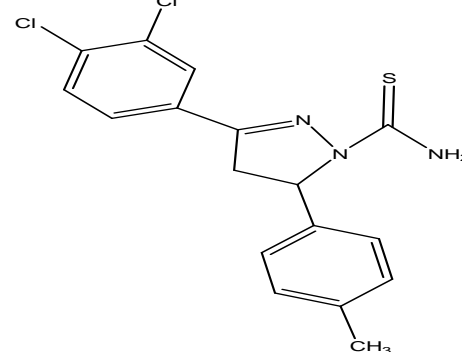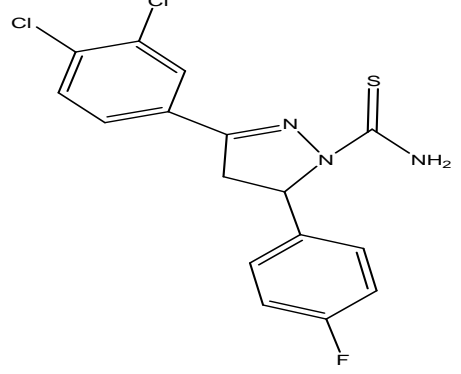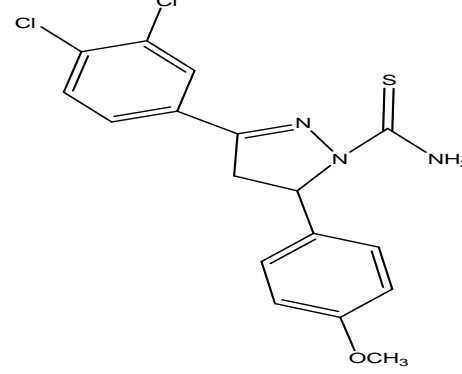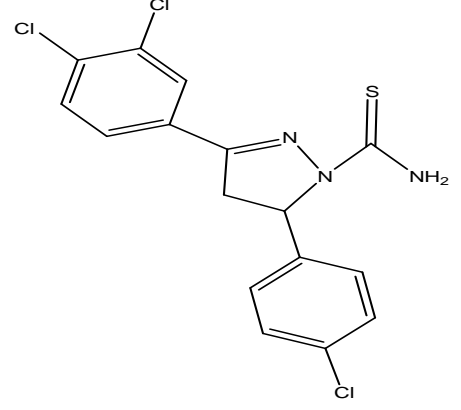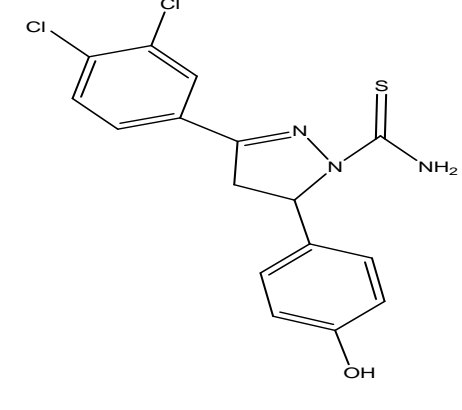
Drug similarity is a qualitative concept used in drug design, which is estimated from the molecular structure even before the substance is synthesized and tested. The calculation of the medicinal property can give us a better hypothesis of the biological activity of certain molecules. The theoretical calculation of certain properties of a molecule can fulfill the essential parameters to demonstrate a certain biological activity. [3]
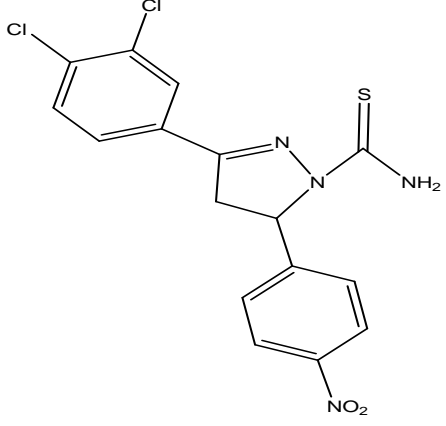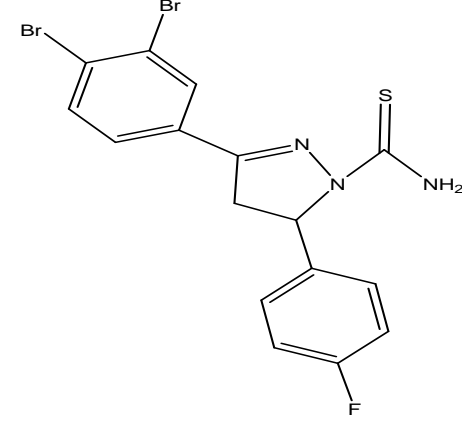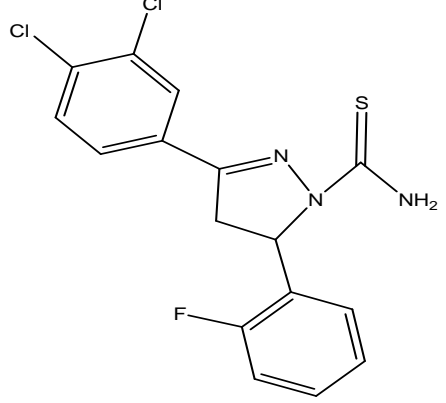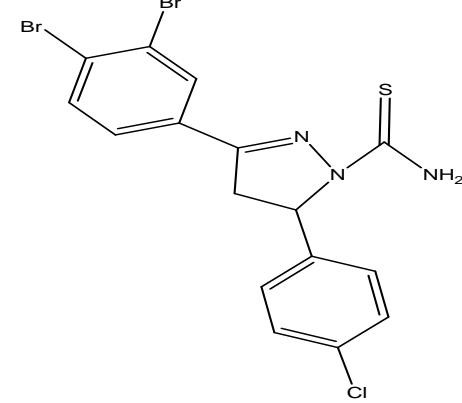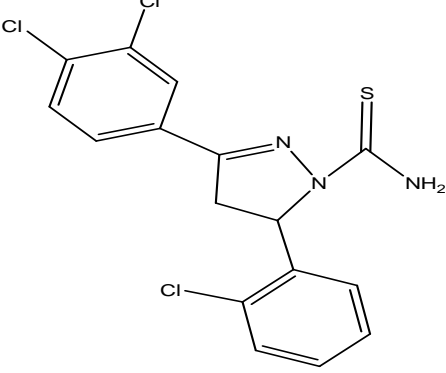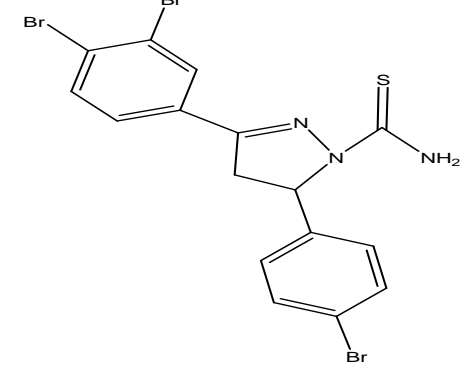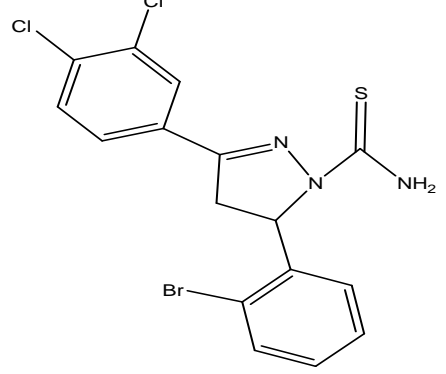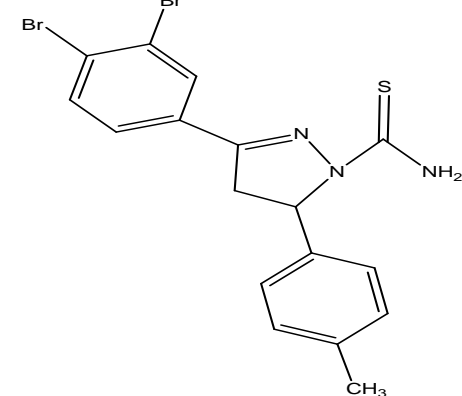
IV.2. Study of the QSAR properties of the series of pyrazole derivatives:

IV.2.1. Chemical structures of 1, 2-diazole derivatives:

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

Table IV. 1. Chemical structures of pyrazole derivatives.

	Serie		Serie
C1		C5	
C2		C6	
C3		C7	
C4		C8	

C9		**C13**	
C10		**C14**	
C11		**C15**	
C12		**C16**	

C17		**C21**	
C18		**C22**	
C19		**C23**	
C20		**C24**	

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

C25		C28	
C26		C29	
C27		C30	
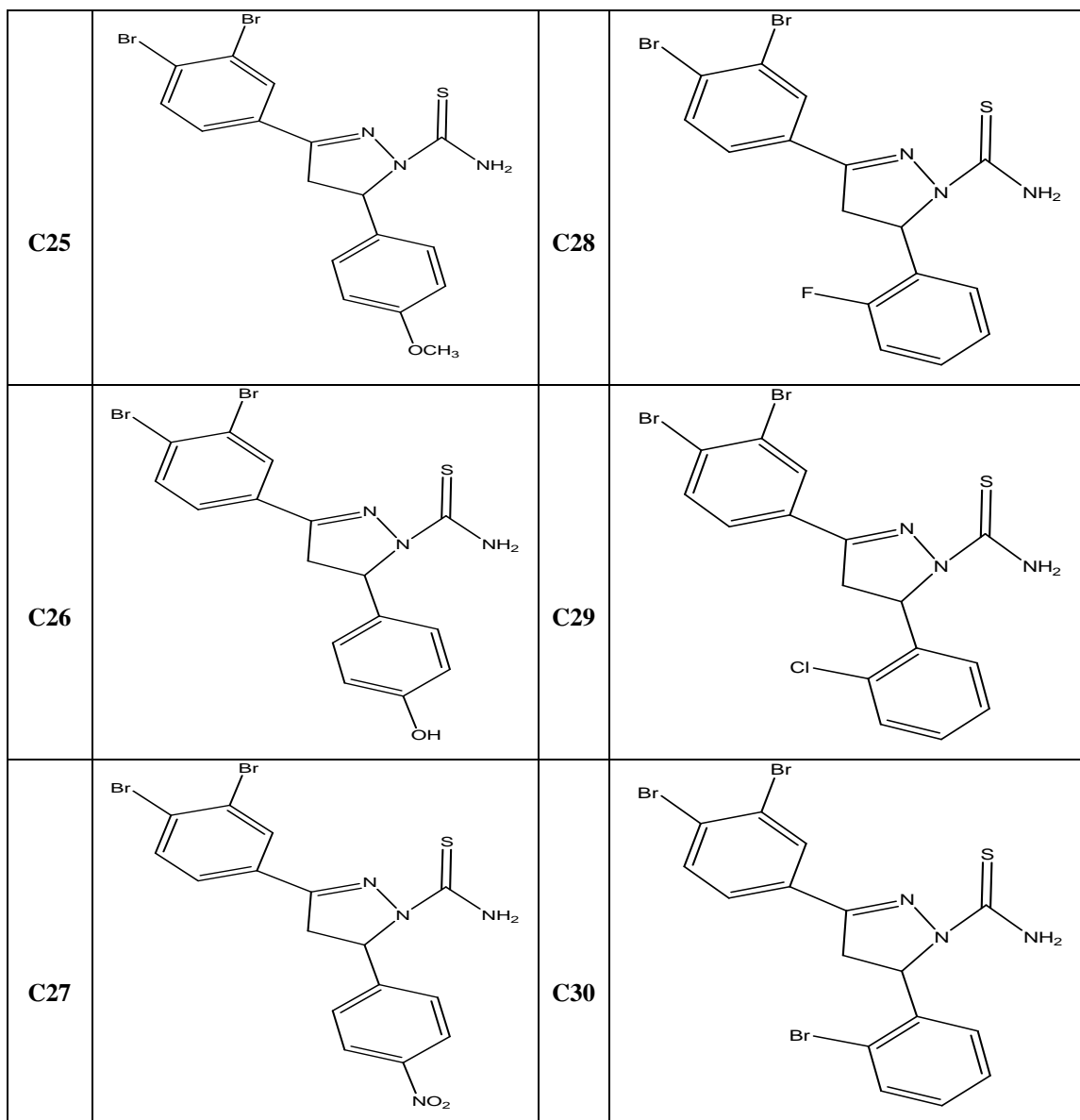
IV.2.2. Study of the physico-chemical properties of pyrazole derivatives:

In this part of work, we studied seven physical and chemical properties of thirty pyrazole derivatives in relation to their activity, was performed by QSAR methods, using HyperChem software (8.0.7).

The properties concerned are surface area (SAG), molar volume (V), energy of hydration (HE), octanol/water partition coefficient (logP), molar refractivity (MR), polarizability (Pol) and molecular weight (MW).

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

The results obtained in the following table:

Table IV. 2. QSAR parameters of 1, 2-diazole derivatives

Compound	Molecular Surface ($Å^2$)	Molecular Volume ($Å^3$)	Molecular Mass (amu)	Polarizability ($Å^3$)	Hydration Energy (Kcal/mol)	Refractivity ($Å^3$)	Log P
1	503.20	829.51	309.28	28.06	-6.23	80.03	1.78
2	523.64	864.80	325.73	30.07	-6.22	84.62	2.16
3	534.32	885.50	370.18	30.77	-6.21	87.44	2.43
4	574.52	978.53	323.46	39.05	-4.96	106.52	3.55
5	590.88	1002.81	339.46	39.68	-7.80	108.62	2.40
6	561.52	947.89	325.43	37.85	-12.95	103.85	2.37
7	565.20	932.72	342.33	34.43	-10.14	96.63	0.93
8	493.85	820.63	309.28	28.06	-6.02	80.03	1.78
9	504.31	846.50	325.73	30.07	-5.98	84.62	2.16
10	509.40	860.57	370.18	30.77	-6.00	87.44	2.43
11	522.09	852.74	356.16	32.69	-6.31	91.91	2.32
12	543.57	888.07	372.62	34.70	-6.29	96.50	2.70
13	555.31	911.70	417.07	35.40	-6.41	99.31	2.98
14	576.88	963.17	364.29	39.23	-6.44	107.39	2.80
15	594.40	988.09	380.29	39.87	-9.26	109.48	1.65
16	560.77	932.30	366.26	38.04	-14.42	104.72	1.62
17	558.80	914.81	383.17	34.62	-11.62	97.50	0.10
18	511.15	844.83	356.16	32.69	-6.09	91.91	2.32
19	521.62	870.61	372.62	34.70	-6.90	96.50	2.70
20	504.95	848.58	417.07	35.40	-7.37	99.31	2.98
21	536.46	882.00	445.06	34.08	-6.25	97.54	2.87
22	556.98	918.06	461.52	36.10	-6.26	102.13	3.25
23	565.41	928.33	505.97	36.80	-3.33	104.95	3.52
24	590.87	999.16	453.19	40.63	-6.44	113.03	3.35
25	605.75	1017.70	469.19	41.27	-9.20	115.12	2.20
26	572.42	961.45	455.17	39.43	-14.38	110.35	2.17
27	568.73	940.90	472.07	36.01	-11.48	103.13	0.73
28	524.48	847.37	445.06	34.08	-6.06	97.54	2.87
29	534.74	899.90	461.52	36.10	-6.11	102.13	3.25
30	524.08	883.87	505.97	36.80	-3.62	104.95	3.52

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

Big value: green   Small value: red

Results interpretation:

According to the results obtained, we notice that the values of the polarizability are generally proportional to the values of the surfaces and the volumes. (Direct Correlation Relationship).

It is also observed that most of the values of the polarizability and of the molar refractivity increase relatively with the size and the molecular weight of the 1,2-diazoles studied.

This result is in agreement with the Lorentz-Lorenz formula, which gives a relationship between the polarizability, the molar refractivity and the molecular size. [4]

Therefore, according to relation shows that the molar refractivity and the polarizability increase with the volume and the molecular mass.

For example, compounds 23, 24 and 25 carry bulky substituents have large values of polarizability (36.80Å^3), (40.63Å^3), (41.27Å^3) with high values of molar refractivity $(104.95\ \text{Å}^3)$,$(113.03\ \text{Å}^3)$,$(115.12\ \text{Å}^3)$ respectively.

On the other hand the compounds carry small substituents have small values of the molar refractivity and the polarizability example: the compounds 8 have (28.06Å^3) of polarizability and $(80.03\text{Å}3)$ of molar refractivity.

The hydration energy in absolute values, the largest are that of compounds 16 (14.42Kcal/mol), 26 (14.38Kcal/mol) and the small values are that of compounds 23 (3.33Kcal/mol) and 30 (3.62Kcal/mol).In fact, in biological environments, Hydration energy is the energy that is released when water molecules attach to ions. These are hydrogen bonds established between them.

The hydrophobic groups in the structures of the derivatives induce a decrease in the energy of hydration; however, the presence of hydrophilic groups as in compound 16 which has three proton donor sites, seven proton acceptor sites, on the

other hand for compound 23 which has two proton donor sites and seven proton acceptor sites. And this explains the high hydration energy in compound 16 and its low in compound 23.

On the contrary, the lipophilicity increases proportionally with the hydrophobic character of the substituents. [5] The fourth compound has a higher Log P value (3.55).

According to the results obtained, the majority of the compounds studied have optimal values of Log P, which vary between 0.10 < Log P < 3.55, there are no negative values. These molecules have good intestinal absorption due to a good balance between solubility and permeability by passive diffusion. In fact, the metabolism is minimized due to the low binding with metabolic enzymes.

IV.3. Drug like and multi-parameter optimization (MPO):

A successful, efficacious and safe drug must have a balance of properties, including potency against its intended target, appropriate absorption, distribution, metabolism, and elimination (ADME) properties and an acceptable safety profile. Achieving this balance of, often conflicting, requirements is a major challenge in drug discovery. Approaches to simultaneously optimizing many factors in a design are broadly described under the term 'multi-parameter optimization' (MPO).

In this review, we will describe how MPO can be applied to efficiently design and select high quality compounds and describe the range of methods that have been employed in drug discovery, including; simple 'rules of thumb' such as Lipinski's rule; and rule of Veber as well as the rule of Ghose.et al. On the other hand, metric methods aim to combine potency with other parameters into a single metric that can be monitored during optimization being the old methods the ligand efficiency (LE) and lipophilic ligand efficiency (LipE) [6], so for modern methods we find the Golden Triangle. [7]

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

IV.3. 1 Representation of "drug-like" calculations based on Lipinski:

A major contributor in the field of the characterization of "drug-like" compounds is Lipinski with the "rule of 5" [8]. This rule is the most used for the identification of "drug-like" compounds [9]. According to this rule, compounds that do not pass at least two of the following criteria are very likely to have absorption or permeability problems:

• Molecular mass $\leq$ 500 Da

• log P $\leq$ 5

• binding acceptors H $\leq$ 10

• H bond donors $\leq$ 5

The "rule of 5" was developed from orally administrable compounds that successfully passed phase II clinical trials. It is therefore not a method for distinguishing compounds that are potentially drugs from those that are not, but rather a method for identifying compounds with low absorption or low permeability.

Table IV. 3. Lipinski parameters of pyrazole derivatives

Compound	HBD	HBA	M (uma)	logP	Number of breaches
1	2	5	309.28	1.78	0
2	2	5	325.73	2.16	0
3	2	5	370.18	2.43	0
4	2	4	323.46	3.55	0
5	2	5	339.46	2.40	0
6	3	5	325.43	2.37	0
7	2	6	342.33	0.93	0
8	2	5	309.28	1.78	0
9	2	5	325.73	2.16	0
10	2	5	370.18	2.43	0
11	2	7	356.16	2.32	0
12	2	7	372.62	2.70	0

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

13	2	7	417.07	2.98	0
14	2	6	364.29	2.80	0
15	2	7	380.29	1.65	0
16	3	7	366.26	1.62	0
17	2	8	383.17	0.10	0
18	2	7	356.16	2.32	0
19	2	7	372.62	2.70	0
20	2	7	417.07	2.98	0
21	2	7	445.06	2.87	0
22	2	7	461.52	3.25	0
23	2	7	505.97	3.52	0
24	2	6	453.19	3.35	0
25	2	7	469.19	2.20	0
26	3	7	455.17	2.17	0
27	2	8	472.07	0.73	0
28	2	7	445.06	2.87	0
29	2	7	461.52	3.25	0
30	2	7	505.97	3.52	0

Results interpretation:

All values for log P are positive indicating that the compounds are too lipophilic. Thus, it has good permeability through the biological membrane, better binding to plasma proteins, elimination by metabolism, but low solubility and poor gastric tolerance. [10]

All the compounds of the series comply with rules 3 and 4, this means that the hydrogen acceptor numbers less than 10 (O, S, F, Br, Cl) and the number of hydrogen donors less than 5 (OH, NH) .results in increased hydration energy and low permeability .Because the HBAs which are of a large number leads to a low permeability through a bilayer membrane. The smaller number leads to better permeability. [11]

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

And for weight All compounds in the series have molar masses below 500 Da except 23 and 30 (505.97), and all compounds are likely soluble and easily cross cell membranes.

Based on these criteria it can be said that the compounds of the series are acceptable to be administered orally. **[12]**

IV.3.2. Veber's rules:

Two other criteria introduced by Veber PSA (Polar Surface area) and number of links that can rotate ≤ 10, are often used in addition to the "rule of 5". These limits were established from measurements of the oral bioavailability of drug candidates. **[13]**

For ideal oral bioavailability, there are two identified descriptors:

- Rotating links are less than 10.
- Polar area is less than 140 Å2.

The PSA was used to calculate the percentage of absorption (% ABS) according to the equation: % ABS = 109 ± 0.345 x PSA. **[14]**

Table IV. 4. Veber's rules for 1, 2-diazole derivatives

Compound	NRB ≤ 10	PSA (Å²) ≤ 140 Å²	%ABS
1	3	73.71	83.57
2	3	73.71	83.57
3	3	73.71	83.57
4	3	73.71	83.57
5	4	82.94	80.39
6	3	93.94	76.59
7	4	119.53	67.76
8	3	73.71	83.57
9	3	73.71	83.57
10	3	73.71	83.57
11	3	73.71	83.57

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

12	3	73.71	83.57
13	3	73.71	83.57
14	3	73.71	83.57
15	4	82.94	80.39
16	3	93.94	76.59
17	4	119.53	67.76
18	3	73.71	83.57
19	3	73.71	83.57
20	3	73.71	83.57
21	3	73.71	83.57
22	3	73.71	83.57
23	3	73.71	83.57
24	3	73.71	83.57
25	4	82.94	80.39
26	3	93.94	76.59
27	4	119.53	67.76
28	3	73.71	83.57
29	3	73.71	83.57
30	3	73.71	83.57

Results interpretation:

According to the results obtained in the table above, it is observed that all the NRB number values of the 1,2-diazole derivatives studied are less than 10. The low number of rotary bonds in the compounds studied indicates that these ligands upon binding to a protein change their conformation only slightly.

For the PSA results, we note that all the compounds of the series studied have values below 140 $Å^2$, which shows the good prediction of oral bioavailability and transport through biological membranes.

For the percent absorption (%ABS) values, all compounds can be assured to have a large %ABS ranging from 67.76% to 83.57%, indicating that these compounds should have good cell membrane permeability.

Therefore, we conclude that all the studied compounds respect Veber's rule.

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

IV.3.3. Efficiency of ligand "LE":

Ligand Efficiency It is a Metric methods originated from the observation that smaller compounds tend to have better physico-chemical properties and good ADMEs than larger compounds.

The smallest compound tends to have the best physicochemical properties and good ADME with respect to ligand efficacy [15], on the other hand larger compounds have weaker physicochemical and ADME properties.

The calculation of this parameter is performed as follows:

$$pIC_{50} = -\log(IC50) \qquad\qquad LE = 1.4\ pIC_{50} / NH$$

pIC_{50}: biological activity.

NH: is the number of heavy atoms.

Table IV. 5. Ligand efficacy of 1, 2-diazole derivatives

Compound	PIC_{50}	N_H	LE
1	6.081	23	0.370
2	5.866	23	0.357
3	5.664	23	0.345
4	6.469	23	0.394
5	7.155	24	0.417
6	6.886	23	0.419
7	5.514	25	0.309
8	5.285	23	0.322
9	5.412	23	0.329
10	5.376	23	0.327
11	5.203	23	0.317
12	5.135	23	0.313
13	5.164	23	0.314
14	5.132	23	0.312
15	5.241	24	0.306
16	5.278	23	0.321
17	5.109	25	0.286

18	5.183	23	0.315
19	5.138	23	0.313
20	5.192	23	0.316
21	5.050	23	0.307
22	5.089	23	0.310
23	4.997	23	0.304
24	5.007	23	0.305
25	5.092	24	0.297
26	4.948	23	0.301
27	4.973	25	0.278
28	4.874	23	0.297
29	4.912	23	0.299
30	4.960	23	0.302

Big value: green   Small value: red

Results interpretation:

From the results obtained in Table IV.5, we observe that LE decreases with the increase in the number of heavy atoms , this means that to obtain a high and significant LE, compounds with weak heavy atoms and this can be explained by the correlation that exists between the size of the compound and their physico-chemical properties.

It is observed that the large value of LE attached to compound 6 of 0.419 because it has less heavy atom number 23, this compound tends to have better physico-chemical and ADME properties than large compounds. On the other hand the small value of LE is equal to 0.278 of compound 27 which carries the greatest number of heavy atom 25.

IV.3.4. Ligand Lipophilicity Efficiency "LLE":

Is a parameter that combines both activity and lipophilicity, It is defined as a measure of the efficiency of a ligand to bind to a given target.

On the other hand, we will study the lipophilic efficiency (LipE) of the ligand, since it is well established that the lipophilicity of a molecule plays a crucial role in determining its ability to be a drug candidate. **[16]** .To maximize potency while keeping lipophilicity as low as possible, due to the association between high lipophilicity and several issues including low solubility, membrane permeability, metabolic stability, etc. **[17]**

For an LLE of 5 to 7 or even more, it can be said that these optimized compounds should be more selective.

Leeson and Springthorpe define Lipophilic Efficiency (LipE) as follows:

$$LLE = LipE = pIC_{50} - \log P$$

Table IV. 6. Ligand Lipophilicity Efficiency of 1, 2-diazole derivatives

compound	PIC_{50}	logP	LLE
1	6.081	1.78	4.301
2	5.866	2.16	3.706
3	5.664	2.43	3.234
4	6.469	3.55	2.919
5	7.155	2.40	4.755
6	6.886	2.37	4.516
7	5.514	0.93	4.584
8	5.285	1.78	3.505
9	5.412	2.16	3.252
10	5.376	2.43	2.946
11	5.203	2.32	2.883
12	5.135	2.70	2.435
13	5.164	2.98	2.184
14	5.132	2.80	2.332
15	5.241	1.65	3.591
16	5.278	1.62	3.658
17	5.109	0.10	5.009
18	5.183	2.32	2.863
19	5.138	2.70	2.438
20	5.192	2.98	2.212

21	5.050	2.87	2.18
22	5.089	3.25	1.839
23	4.997	3.52	1.477
24	5.007	3.35	1.657
25	5.092	2.20	2.892
26	4.948	2.17	2.778
27	4.973	0.73	4.243
28	4.874	2.87	2.004
29	4.912	3.25	1.662
30	4.960	3.52	1.44

Big value: green   Small value: red

Results interpretation:

According to the values obtained in the table above, we see that the LLE changes from 1.44 to 5.009. And we take the only compound 17 reaches an LLE of 5.009 which is belongs to the range 5-7 this indicates that this compound has been successfully optimized.

On the other hand, the LLE values for the other compounds such as (27, 28, 30) are respectively (4.243, 2.004, 1.44), it is observed that none of the compounds reach an ELL greater than 5. In these cases, the affinity gain is accompanied by an increase in lipophilicity. In this respect, the optimization was not as optimal as in the first example.

IV.3.5. Golden Triangle:

The Golden Triangle is a visualization tool developed from in vitro permeability, in vitro clearance and computational data designed to aid medicinal chemists in achieving metabolically stable, permeable and potent drug candidates.

Classifying compounds as permeable, and stable and plotting molecular weight (MW) versus octanol: buffer (pH 7.4) distribution coefficients (logD) or estimated octanol: buffer (pH 7.4) distribution coefficients (logD) reveals useful trends. [7]

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

Analysis of at least two orthogonal trends, such as permeability and clearance, can be extremely effective in balancing and optimizing multiple properties. In addition, molecular weight and logD affect potency-efficiency calculations, allowing potency, clearance and permeability to be optimized simultaneously. [7]

All trends combined lead to the observation that in vitro molecular weight polarity and permeability and low clearance compounds are concentrated with a baseline of **log D=-2.0** to **5.0** for **MW=200** and a top of **log D=1.0** to **2.0** for **MW=450**. These trends lead to a shaped area known as the Golden Triangle and molecules in this area that: are low clearances and permeable should follow the golden triangle rule. [18]

LogD is a measure of logP at a given pH for a compound of a certain pKa. This parameter thus takes into account the concentration of ionized molecules.

Table IV. 7. Distribution coefficients of 1, 2 -diazole

Compound	MW (Da)	LogD (PH = 7.4)
1	309.28	4.63
2	325.73	5.09
3	370.18	5.25
4	323.46	5.00
5	339.46	4.33
6	325.43	4.18
7	342.33	4.42
8	309.28	4.63
9	325.73	5.09
10	370.18	5.25
11	356.16	4.81
12	372.62	5.27
13	417.07	5.43
14	364.29	5.18
15	380.29	4.51

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

16	366.26	4.36
17	383.17	0.53
18	356.16	4.81
19	372.62	5.27
20	417.07	5.43
21	445.06	5.14
22	461.52	5.60
23	505.97	5.76
24	453.19	5.51
25	469.19	4.84
26	455.17	4.69
27	472.07	0.02
28	445.06	5.14
29	461.52	5.60
30	505.97	5.76

Big value: green   Small value: red
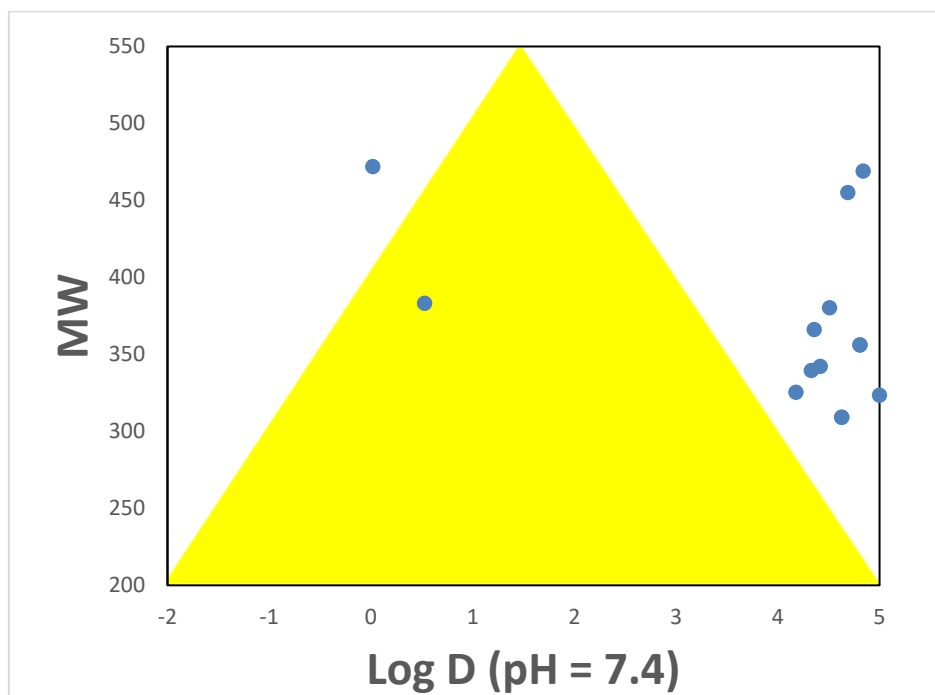
Figure .IV. 1. The Golden Triangle

Chapter IV: Qualitative study of the structure-activity relationship of a series of 1,2-diazole derivatives.

2022

Results interpretation:

Compounds that reside within the golden triangle are more likely to be both metabolically stable and possess good membrane permeability than those that lie outside.

However, in our case, the golden triangle doesn't help in sorting because it shows that all of the studied compounds lie outside the triangle except for compound 17.

In general, lower log D molecules and high molecular weight molecules fail due to low permeability, while higher log D and higher molecular weight compounds fail due to high in vitro clearance **[19].**

IV.4. Conclusion:

In this chapter, a qualitative study of the structure-activity relationship of a series of 1, 2-diazole derivatives has been presented. Also offers a structural comparison between the thirty compounds studied. She also provides an in-depth discussion of drug likeness.

Polarizability values are generally proportional to volume and area values and to refractivity. Compound 25 takes the important values for polarizability 41.27 Å3 and refractivity 115.12 Å3.

The hydration energy in absolute value, the largest being that of compound 16 (14.42 kcal/mol) .It is therefore the best distribution in the tissues.

By comparison, of the calculation results and the criteria of the Lipinski rule, it was noticed that all the compounds of the series confirm these criteria. This shows that the compounds studied theoretically do not present any oral bioavailability problem.

Another study based on the rules of Veber shows that the compounds of the series studied are all correlated with this rule, that is to say that all the derivatives respect the bars of this rule.
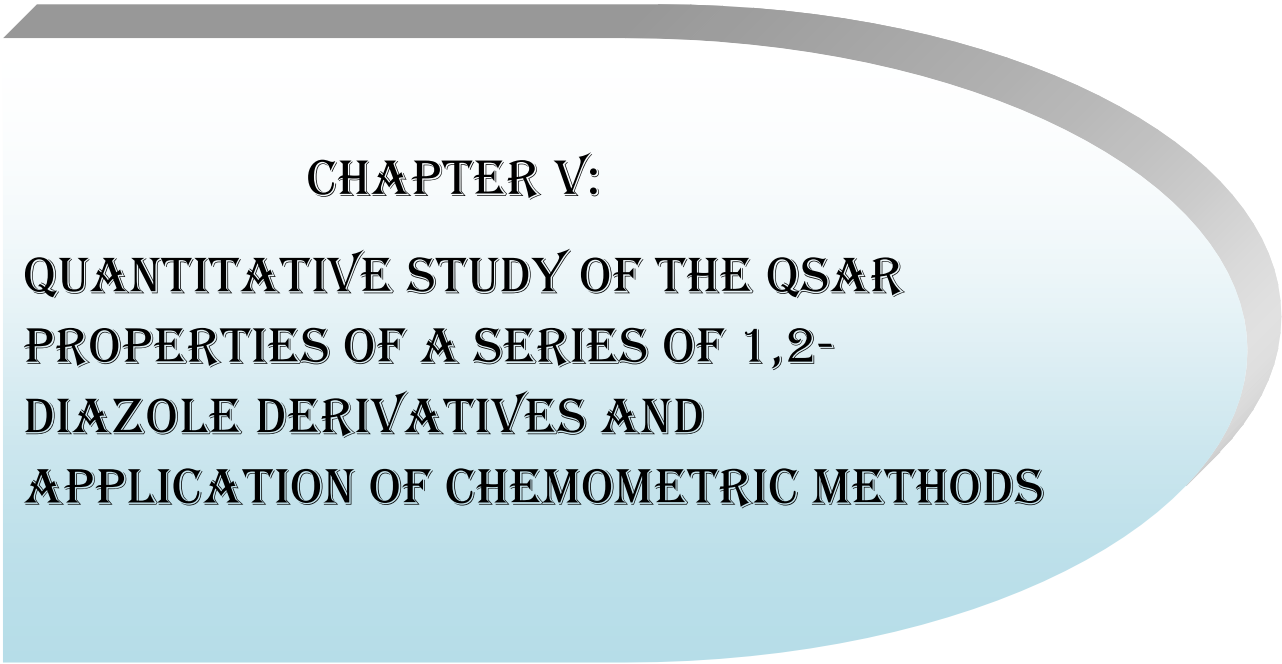
For the study of EL we find that compound 6 has the greatest value which is 0.419, which allows it good physico-chemical and ADME properties.

For the study of ELL we find that the compound 17 reaches an LLE of 5.009 which is belongs to the range 5-7 this indicates that this compound has been successfully optimized.

References:

[1] G. Grant W .Richards, Oxford Chemistry Primers, Oxford, 1995.

[2] Guillaume FAYET, doctoral thesis, University of PIERRE ET MARIE CURIE, 2010.

[3] M. Mellaoui, S. Belaidi, D. Bouzidi, and N.Gherraf, Quantum Matter; 3: 435-441; 2014

[4] Helge Kragh, Substantia journal of the History of Chemistry. 2; 7-18; 2018.

[5] Fatouche Maroua, Master's thesis in pharmaceutical chemistry, Mohamed Khider University of Biskra, 2016**.**

[6] D. M. Segall, J. Curr. Pharm. Des. 18; 1292-1310; 2012.

[7] Ted W Johnson [1], Klaus R Dress, Martin Edwards. Bioorg. Med. Chem. Lett. 19; 5560-5564; 2009.

[8] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. *Adv. Drug.* Deliv. Rev;46 ;3-26; 2001.

[9] Lipinski, C.A .*Drug* Discov .*Today*, 1; 337-341; 2004.

[10] S.Schultes, C. Graaf, E. Haaksma, J. P. Iwan, O. Kramer, J. Drug Discovery Today: Technologies, ; 7: 157;2010.

[11] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, Adv. Drug. Discov. ; 23: 3-25; 1997.

[12] S. Qaneinasab; Z. Bayat; J. Chem. Pharm.**;** 3;561;2011

[13] Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; *J.Med.Chem 45*; 2615-2623; 2002.

[14] M. Remko, M. Swart, F. M. Bickelhaupt, J. Bioorg. Med. Chem.14; 1715-1728;2004

[15] A. L. Hopkins, C. R. Groom, A. Alexander, J. Drug .Discov .T, 9:430; 2004.

[16] P. D.Leeson, B. Springthorpe, Nature Rev. Drug Discov. **;** 6; 881-890; 2007.

[17] M. P. Edwards and D. Price, Annu. Rep. Med. Chem.; 45; 380-391; 2010

[18] Khalfa Nadjla, Master's thesis in pharmaceutical chemistry, Mohamed Khider University of Biskra, 2018

[19] T. W. Johnson, K. R. Dress, and M. Edwards, Bioorg. Med .Chem.Lett .19: 5560-5564; 2009.

CHAPTER V:

QUANTITATIVE STUDY OF THE QSAR PROPERTIES OF A SERIES OF 1,2-DIAZOLE DERIVATIVES AND APPLICATION OF CHEMOMETRIC METHODS

Chapter *V* : Quantitative study of the QSAR properties of a series of 1,2-diazole derivatives and application of chemometric methods

2022

V.1. Introduction:

Quantitative Structure-Activity relationships, Quantitative Structure-property relationships (terms are sometimes used interchangeably), It is the process by which a chemical structure is correlated with a well-defined effect such as biological activity (physico-chemical property).

QSAR/QSPR properties can be used in two aspects: qualitative and quantitative.

To describe properties of a series of molecules using a QSAR/QSPR model, that is to say a set of mathematical expressions, which can then be used as a means of predicting the biological response for similar structures (pharmacophore model, 2D, 3D).

To obtain a valid QSAR/QSPR model it is necessary to Select the descriptors for the different compounds: Quantum descriptors (ex; HOMO, LUMO, Charges…..), Molecular descriptors (ex; Volume; surface...), and Physico-chemical descriptors (Log P; Hydration energy, etc.).

So the principle is to set up a mathematical relationship quantitatively linking the descriptors, with an observable (biological activity, toxicity, etc.), for a series of active chemical molecules

In this work, we are interested in the study of the physicochemical properties of pyrazole and its derivatives which make it possible to predict the physicochemical parameters which are influential on the biological activity, in order to predict the biological activity of new molecules, it is for this reason the choice of a series of thirty derivatives of 1, 2-diazole with different physico-chemical descriptors has been made.

Chapter *V* : Quantitative study of the QSAR properties of a series of 1,2-diazole derivatives and application of chemometric methods

2022

V.2. QSAR tools and techniques:

V.2.1. Biological parameters:

Biological data are usually expressed on a logarithmic scale due to the linear relationship between the response and the logarithm of dose in the central region of the log dose-response curve. Inverse logarithms of activity (log 1/C) are also used to obtain higher mathematical values when the structures are biologically very efficient. [1]

V.2.2. Molecular descriptors:

The crux of the QSPR approach is an appropriate description of molecular structures. The chemical descriptors take into account the different aspects of the chemical. The molecular descriptor expresses the chemical information transformed and encoded from a molecule and effectively solves chemical, pharmaceutical and toxicological problems. The advantage of theoretical molecular descriptors lies in the development of compounds that have never been synthesized or explored experimentally. Molecular descriptors are well known for their ability to establish linear regression relationships with physico-chemical and biological properties. [2]

V.2.3 Multiple Linear Regressions (MLR):

a. Description of the method:

It is based on the assumption that there is a linear relationship between a dependent variable (to explain) Y (here, the activity) and a series of p independent (explanatory) variables Xi (here, the descriptors). [3]

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_P X_P + \varepsilon$$

This equation is linear with respect to the parameters (regression coefficients) $\beta_0, \beta_1 \dots \beta_P$. The determination of equation is then made from a database of n samples (observations) for which both the independent variables and the variable dependent are known.

It is therefore a question of considering a system of equations, which can be given in the form following matrix [4]:

$$
Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ . \\ Y_n \end{bmatrix} \quad
X = \begin{bmatrix} 1 & X_{11} & X_{1P} \\ 1 & X_{21} & X_{2P} \\ 1 & & . \\ . & . & . \\ 1 & X_{n1} & X_{nP} \end{bmatrix} \quad
\beta = \begin{bmatrix} \beta_0 \\ \beta_2 \\ . \\ . \\ \beta_n \end{bmatrix} \quad
\varepsilon = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ . \\ . \\ \varepsilon_n \end{bmatrix}
$$

Multiple linear regression (MLR) techniques based on least squares procedures are widely used to estimate the coefficients involved in the model equation [5, 6].

b. Testing the overall significance of the regression:

It seems reasonable to test the overall significance of the model, i.e. to test whether all the coefficients are assumed to be zero, except for the constant.

The test statistic: Fisher's F statistic.

We use the statistic, denoted F defined by the formula:

$$\mathbf{F} = \frac{ESS}{P} = \frac{n-p-1}{RSS}$$

Thus, the over fitting and chance correlation, due to the excess number of descriptors, can be detected by a Q value, which must be positive. [7].

Q (The quality factor) $\mathbf{Q} = \dfrac{r}{s}$

Where r is the variance and S is the standard deviation: $\mathbf{S} = \sqrt{\dfrac{RSS}{n-p-1}}$

To test the validity of the predictive power of the model, we use some techniques:

LOO (leave-one-out) cross-validation is a model validation technique for evaluating how the results of a statistical analysis will generalize to a set of independent data, in which we will use these statistical parameters:

PRESS (predicted residual sum of squares)

$$PRESS = \sum(Yobs - Ycalc)^2$$

TSS (total sum of squares)

$$SSY = \sum(Yobs - Ymean)^2$$

R^2_{adj} (adjusted R-square)

$$r^2_{adj} = \left([1 - (r^2)]\left(\frac{n-1}{n-p-1}\right)\right)$$

R^2_{CV} (cross-validated correlation coefficient)

$$r^2_{CV} = 1 - \frac{PRESS}{SSY}$$

SPRESS (standard prediction error validation)

$$S_{PRESS} = \sqrt{\frac{PRESS}{n}}$$

PE (prediction error)

$$PE = 0.6745\,(1 - r^2)/\sqrt{n}$$

Where: **n** is the number of observations (molecules); **p** is the number of independent variables (descriptors).

The PRESS statistic (predicted residual sum of squares) seems to be the most important parameter for a good estimation of the actual predictive error of the models. Its small value indicates that the model predicts better than chance and can be considered statistically significant. [8]

V.3. Quantitative studies on structure-activity relationships:

In this step of our study, different isothiazole derivatives were evaluated for their inhibitory activity against EGFR. In order to determine the role of physicochemical properties, Lipinski, Veber and logD parameters on biological activity, we use the QSAR study.

Table V. 1. Values of molecular descriptors used in regression analysis

Compound	Molecular Surface ($Å^2$)	Molecular Volume ($Å^3$)	Molecular Mass (amu)	Polarizability ($Å^3$)	Hydration Energy (Kcal/mol)	Refractivity ($Å^3$)	Log P
1	503.20	829.51	309.28	28.06	-6.23	80.03	1.78
2	523.64	864.80	325.73	30.07	-6.22	84.62	2.16

3	534.32	885.50	370.18	30.77	-6.21	87.44	2.43
4	574.52	978.53	323.46	39.05	-4.96	106.52	3.55
5	590.88	1002.81	339.46	39.68	-7.80	108.62	2.40
6	561.52	947.89	325.43	37.85	-12.95	103.85	2.37
7	565.20	932.72	342.33	34.43	-10.14	96.63	0.93
8	493.85	820.63	309.28	28.06	-6.02	80.03	1.78
9	504.31	846.50	325.73	30.07	-5.98	84.62	2.16
10	509.40	860.57	370.18	30.77	-6.00	87.44	2.43
11	522.09	852.74	356.16	32.69	-6.31	91.91	2.32
12	543.57	888.07	372.62	34.70	-6.29	96.50	2.70
13	555.31	911.70	417.07	35.40	-6.41	99.31	2.98
14	576.88	963.17	364.29	39.23	-6.44	107.39	2.80
15	594.40	988.09	380.29	39.87	-9.26	109.48	1.65
16	560.77	932.30	366.26	38.04	-14.42	104.72	1.62
17	558.80	914.81	383.17	34.62	-11.62	97.50	0.10
18	511.15	844.83	356.16	32.69	-6.09	91.91	2.32
19	521.62	870.61	372.62	34.70	-6.90	96.50	2.70
20	504.95	848.58	417.07	35.40	-7.37	99.31	2.98
21	536.46	882.00	445.06	34.08	-6.25	97.54	2.87
22	556.98	918.06	461.52	36.10	-6.26	102.13	3.25
23	565.41	928.33	505.97	36.80	-3.33	104.95	3.52
24	590.87	999.16	453.19	40.63	-6.44	113.03	3.35
25	605.75	1017.70	469.19	41.27	-9.20	115.12	2.20
26	572.42	961.45	455.17	39.43	-14.38	110.35	2.17
27	568.73	940.90	472.07	36.01	-11.48	103.13	0.73
28	524.48	847.37	445.06	34.08	-6.06	97.54	2.87
29	534.74	899.90	461.52	36.10	-6.11	102.13	3.25
30	524.08	883.87	505.97	36.80	-3.62	104.95	3.52

PIC$_{50}$	LogD (PH = 7.4)	HBD	HBA	PSA (Å²) ≤ 140 Å²	NRB ≤ 10
6.081	4.63	2	5	73.71	3
5.866	5.09	2	5	73.71	3
5.664	5.25	2	5	73.71	3
6.469	5.00	2	4	73.71	3
7.155	4.33	2	5	82.94	4

Chapter*V* : Quantitative study of the QSAR properties of a series of 1,2-diazole derivatives and application of chemometric methods

2022

6.886	4.18	3	5	93.94	3
5.514	4.42	2	6	119.53	4
5.285	4.63	2	5	73.71	3
5.412	5.09	2	5	73.71	3
5.376	5.25	2	5	73.71	3
5.203	4.81	2	7	73.71	3
5.135	5.27	2	7	73.71	3
5.164	5.43	2	7	73.71	3
5.132	5.18	2	6	73.71	3
5.241	4.51	2	7	82.94	4
5.278	4.36	3	7	93.94	3
5.109	0.53	2	8	119.53	4
5.183	4.81	2	7	73.71	3
5.138	5.27	2	7	73.71	3
5.192	5.43	2	7	73.71	3
5.050	5.14	2	7	73.71	3
5.089	5.60	2	7	73.71	3
4.997	5.76	2	7	73.71	3
5.007	5.51	2	6	73.71	3
5.092	4.84	2	7	82.94	4
4.948	4.69	3	7	93.94	3
4.973	0.02	2	8	119.53	4
4.874	5.14	2	7	73.71	3
4.912	5.60	2	7	73.71	3
4.960	5.76	2	7	73.71	3

V.4. QSAR model development:

In this step, we tried to develop the QSAR model, in other words, from the relationships identified in isolation, it was a question of using methods of analysis allowing to group the different parameters into a single relationship to explain correlations between physico-chemical and IC50 biological activities of 1,2 -diazole derivatives. , this analysis was performed on the descriptors (Table V.1) using the statistical software SPSS 21.

Chapter*V* : Quantitative study of the QSAR properties of a series of 1,2-diazole derivatives and application of chemometric methods

2022

After eliminating the descriptors whose value does not vary or varies little over all the molecules, a multivariate analysis followed by a statistical evaluation is carried out to develop the best QSAR model. [9]

Among the various QSAR equations, the best QSAR models were selected based on various statistical parameters such as:

- Correlation coefficient R that measures the degree of line association between two variables. It varies in value from 0 to 1.

- The square of Correlation coefficient (R 2 > 0.6) which is the relative measure of the goodness of fit.

- The standard error of the estimate representing the absolute measure of the goodness of fit.

- The Fischer value (F) (F the Fisher ratio), reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High F-test values indicate that the model is statistically significant. [10]

The correlation between the biological activity and the descriptors, expressed by the following relationships:

Log (1/IC50) = 1.859 - 0.009M - 0.128Polarisability+ 1.483LogP + 2.167NRB + 0.821HBD - 0,302LogD

$n = 30$; $R = 0,953$; $R^2 = 0,909$; $SE = 0,193$; $F = 38.324$; $Q = 4.938$

Where: **n:** is the number of compounds, **R:** is the correlation coefficient, **F:** is Fisher's statistic, **SE:** is the standard error of estimate and **Q:** is the goodness of fit or adaptation.

V.4.1. Results and discussion:

The values of the variance fraction can vary between 0 and 1. The QSAR model must consider an r2 > 0.6 for it to be valid.

Chapter V : Quantitative study of the QSAR properties of a series of 1,2-diazole derivatives and application of chemometric methods

2022

For example, the values r = 0.953 and r2 = 0.909 allowed us to firmly indicate the correlation between the different parameters (independent variables) having a specific activity.

The calculated F value for the generated QSAR model exceeds the tabulated F value by large margin as desired for a meaningful regression. Additionally, the F value was found to be statistically significant at the 95% level for this model.

The positive value of the quality factor (Q) for this QSAR model suggests its high predictive power and lack of fit, and the low standard deviation of the model demonstrates the accuracy of the model.

V.4.2. Quantification of descriptors:

RML statistical analysis determines and quantifies the correlations between the descriptors and the target variable. It also indicates the relative contribution of each descriptor to the overall explanation of the activity.

In the model equation, we notice that Log P, NBR and HBD with positive coefficients suggest that biological activity increases with increasing values of these descriptors. On the one hand, the negative coefficients of Log D, Polarizability, and MW suggest the opposite.

The correlation matrix for the pIC50 biological activity and the descriptors selected to build the QSAR-2D model are presented in the following table:

Table V. 2. Model correlation matrix

	PIC50	MW	Polarizability	LogP	NRB	HBD	logD
PIC50	1,000						
MW	-0.634	1,000					
Polarizability	0.003	0.452	1,000				
LogP	-0,042	0,350	0,224	1,000			
NRB	0,120	0,037	0,333	-0,649	1,000		
HBD	0,193	-0,062	0,295	-0,140	-0,167	1,000	
LogD	-0,034	0,054	-0,018	0,806	-0,637	-0,081	1,000

Chapter*V* : Quantitative study of the QSAR properties of a series of 1,2-diazole derivatives and application of chemometric methods

2022

V.5. Model validation:

For model validation using the cross-validation method, is a method of estimating the reliability of a model based on a sampling technique. In fact, there are at least three cross-validation techniques: (test set validation) or (hold out method), (k-fold cross-validation) and (leave-one-out cross-validation LOO).

In order to verify the predictive capacities of our selected MLR model, we resorted to its validation by the use of the "Leave-One-Out" technique (LOO-technique). The model developed was validated by calculating the following statistical parameters: **PRESS** (sum of squares of predicted residual), **SSY** (sum of squares of response value); R^2_{cv} (global predictive ability), **adjusted R^2**, **PE** (the predictive error of the correlation coefficient) and **SPRESS** (prediction uncertainty).

Table V. 3. Cross-validation parameters

Model	PRESS	SSY	PRESS/SSY	Spress	r^2_{cv}	r^2_{adj}	6PE
1	0.859	9,442	0.091	0.169	0.909	0,885	0.067

PRESS is an important cross-validation parameter because it is a good approximation of the actual predictive error of the model. Its value being lower than SSY indicates that this model predicts better than chance and can be considered statistically significant. According to the results presented in (table V.3) this value is equal to 0.859, the model is statistically significant.

In our case, for the proposed model Press < SSY indicating that it has good predictive power and is better than chance .For a good model, Press/SSY should be less than 0.4 and its value in our model is 0.091.

The indication of model performance is obtained from R2cv (the global prediction capability). The high value of r^2_{cv} and r^2_{adj} are essential criteria for the best qualification of the QSAR model. Our result of these two values for this QSAR model was 0.909 and 0.885 respectively.

SPRESS (prediction uncertainty) is a good parameter to use to decide the uncertainty of the prediction .The lower the value of this parameter, the better the predictive ability of the model. In our case, this parameter has a small value of 0.169, which is why the prediction ability is the best for this model.

The prediction error of the correlation coefficient (PE) is another parameter used to determine the predictive power of the proposed model. We calculated the value of 6PE from the proposed model and it is presented in Table V.3. For this model the condition r > 6PE is satisfied and therefore they can be considered as having a good predictive power.

From the set of results of the cross-validation method (Table V.3), we conclude that our 7 parameter model containing 30 compounds is not only statistically excellent, but also has excellent predictive power.

The experimental activity against EGFR, predicted and residual of isothiazole and its derivatives (Table V.4) were deduced by SPSS software. **[11]**

Table V. 4. Experimental, predicted and residual values of (log (1/IC50)) of bioactive series

Compound	PIC_{50} EXP.	PIC_{50}Pred.	PIC_{50}Resid.
1	6,081	5,597	0,484
2	5,866	5,611	0,255
3	5,664	5,460	0,205
4	6,469	6,574	-0,105
5	7,155	7,008	0,147
6	6,886	6,849	0,037
7	5,514	5,445	0,069
8	5,285	5,597	-0,312
9	5,412	5,611	-0,199
10	5,376	5,460	-0,083
11	5,203	5,315	-0,112
12	5,135	5,329	-0,194
13	5,164	5,192	-0,028
14	5,132	5,004	0,128

15	5,241	5,437	-0,196
16	5,278	5,277	0,001
17	5,109	4,984	0,125
18	5,183	5,315	-0,132
19	5,138	5,329	-0,191
20	5,192	5,192	-0,000
21	5,050	5,024	0,026
22	5,089	5,037	0,052
23	4,997	4,885	0,112
24	5,007	4,712	0,295
25	5,092	5,145	-0,053
26	4,948	4,986	-0,038
27	4,973	5,065	-0,092
28	4,874	5,024	-0,150
29	4,912	5,037	-0,125
30	4,960	4,885	0,075

Figure V.1 below presents the linear regression curves of the predicted values with respect to the experimental values of the biological activity of the pyrazoles. The curves of this model represent for convenience with $r^2 = 0.909$

It indicates that this model can be successfully applied to predict cancer inhibitory activity in this series of molecules.
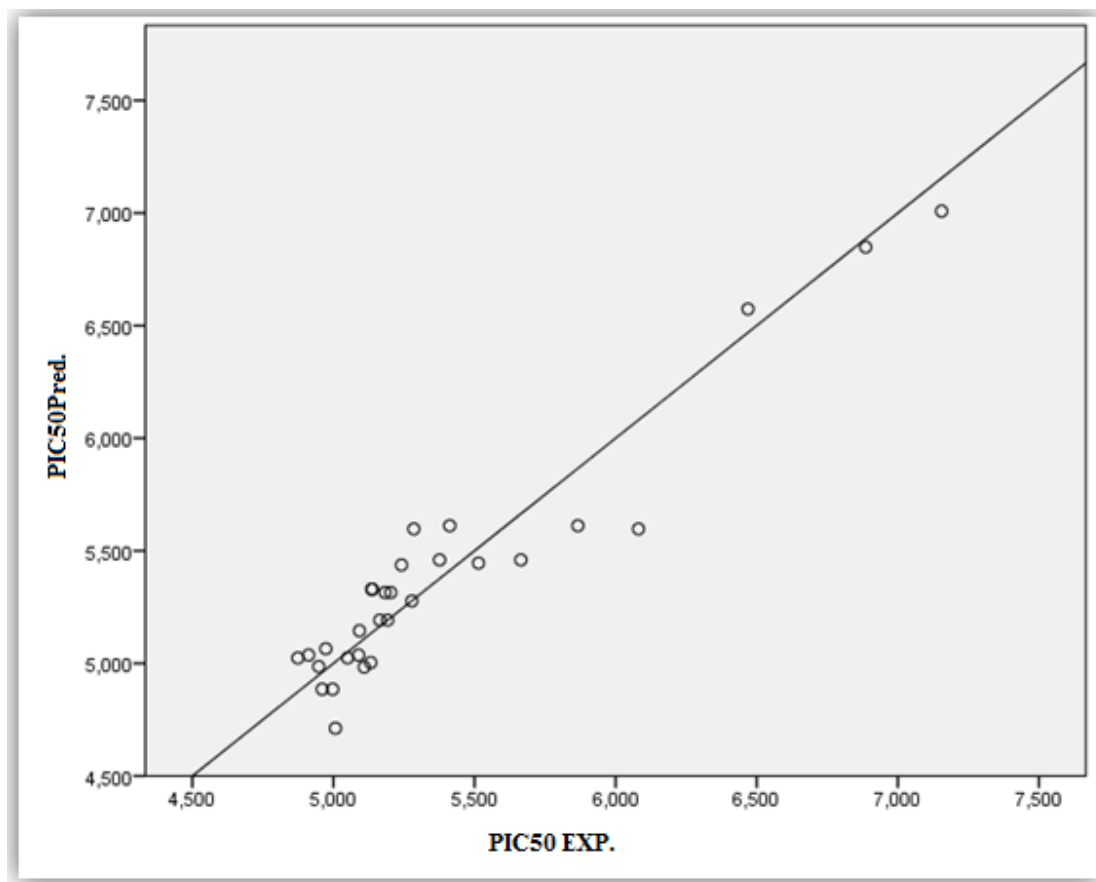
Figure .V. 1. Curve of predicted values as a function of experimental values of log (1/IC50)

To investigate the presence of systematic error in QSAR model building, the residuals of predicted values of log biological activity (1/IC50) were plotted against experimental values, as shown in Figure V.2.

The spread of residuals on both sides of zero indicates that no systemic error exists, as suggested by Jalali-Heravi and Kyani. [12] It indicates that this model can be successfully applied to predict the inhibitory activity against cancer of this class of molecules.
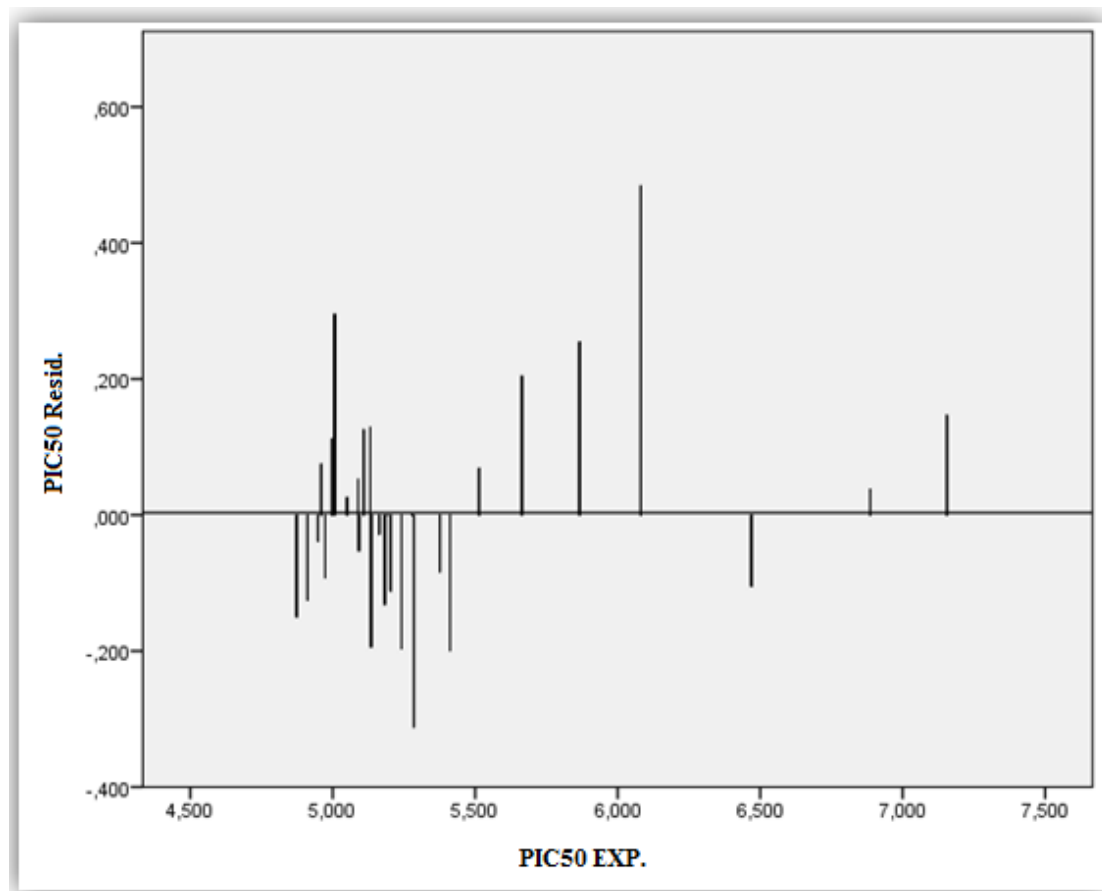
Figure .V. 2. Curve (bar chart) of the residual values compared to the experimentally observed

V.6. Conclusion:

The present study of QSAR has made it possible to determine the quantitative relationship between the physico-chemical properties of the compounds with their activities has been carried out.
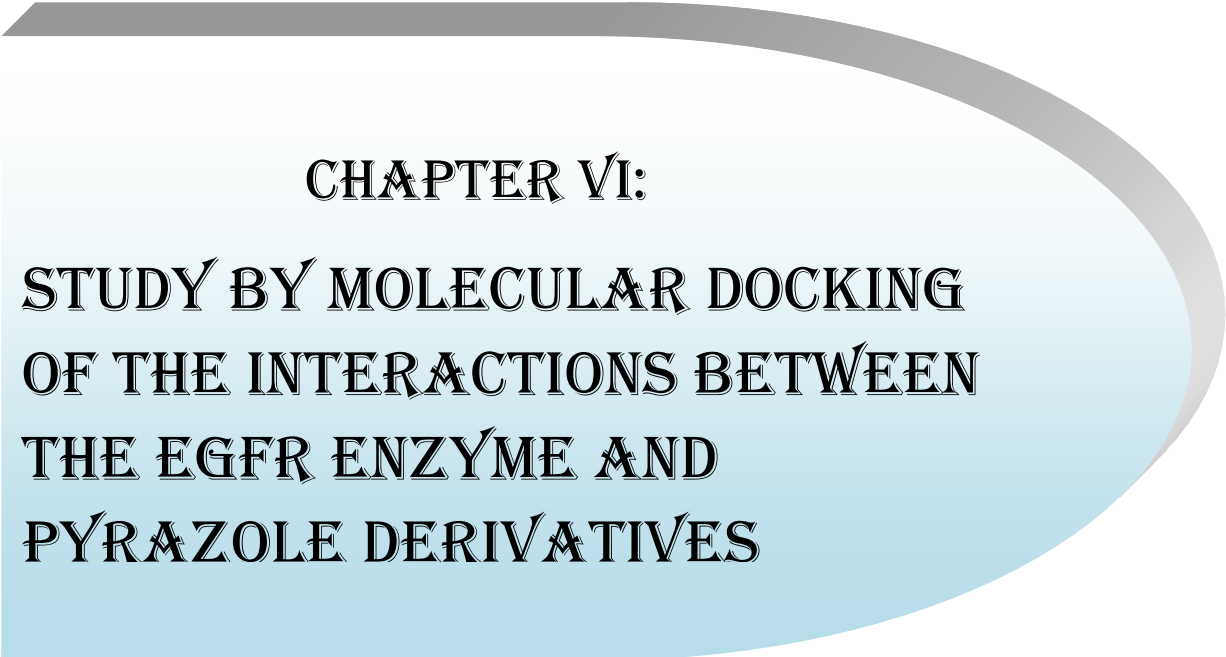
MLR regression analysis was used to develop the model and predict biological activity from molecular descriptors belonging to the 1, 2 -diazole derivative series. Our developed QSAR model is based on the following descriptors: HBD, logD, MW, logD, NRB, Polarizability.

The QSAR model indicates that these descriptors have significant relationships with the observed bioactivity. We observed a high similarity between the

Chapter*V* : Quantitative study of the QSAR properties of a series of 1,2-diazole derivatives and application of chemometric methods

2022

values experimental values and the predicted values of the activity, which indicates the excellent quality of the QSAR model.

References:

[1] K.Tuppurainen, M. Viisas, R. Laatikainen, M. peraekylae .J .Chem. Inform .Comput. Sci., 42: 607-613; 2002.

[2] M .Karelson .Molecular descriptors in QSAR/QSPR .Wiley- Interscience, 2000.

[3] Yagoub Youcef ,Lakrouche Mohammed Seghir, Neguia Seif Eddine, master's thesis, in Chemical Engineering University of El Oued., 2021.

[4] O. FADIA, master's thesis, University ABU-BEKR BELKAID, of Tlemcen, 2019.

[5] J.N .Miller, J .C .Miller, Prentice Hall, London, 2000.

[6] N.R .Draper, H .Smith, Applied Regression Analysis, 2nd .Ed .Wiley, New York, 1981.

[7] S .Riahi, M.F .Mousavi, M .Shamsipur, Talanta; 69; 736-740; 2006.

[8] J. Ghasemi, S. Saaidpour, S. D. Brown. J. Mol. Struc.: Theochem; 805; 27-32; 2007.

[9] Damale, M.G.; Harke, S.N.; Kalam Khan, F.A.; et al.Med .Chem; 14; 35-55; 2014.

[10] K. J. Sanmati, J. Rahul, S. Lokesh, K. Y. Arvind, J. Chem. Pharm. **;** 4: 3215-3223**;**2012.

[11] SPSS 19 for Windows.

[12] J.P. Hughes, S. Rees, S.B. Kalindjian et al .J .Pharmacol; 162; 1239-49; 2011.

CHAPTER VI:

STUDY BY MOLECULAR DOCKING OF THE INTERACTIONS BETWEEN THE EGFR ENZYME AND PYRAZOLE DERIVATIVES

VI.1. Introduction:

Protein-ligand interactions play a key role in the organization of biological systems. They allow the regulation of certain biological processes. However, molecular modeling tools are particularly effective for studies at the atomic level of interactions between two molecular entities and this type of study most often use of **MOLECULAR DOCKING** techniques.

The study of docking is a crucial step in understanding biological reactions and in the design of drugs. Attempts at docking have long been based on the key-lock concept. [1]

Docking aims to predict the structure of a complex formed by two molecules, the formation of these complexes is based on the recognition of the three-dimensional structure of a ligand by a receptor site.
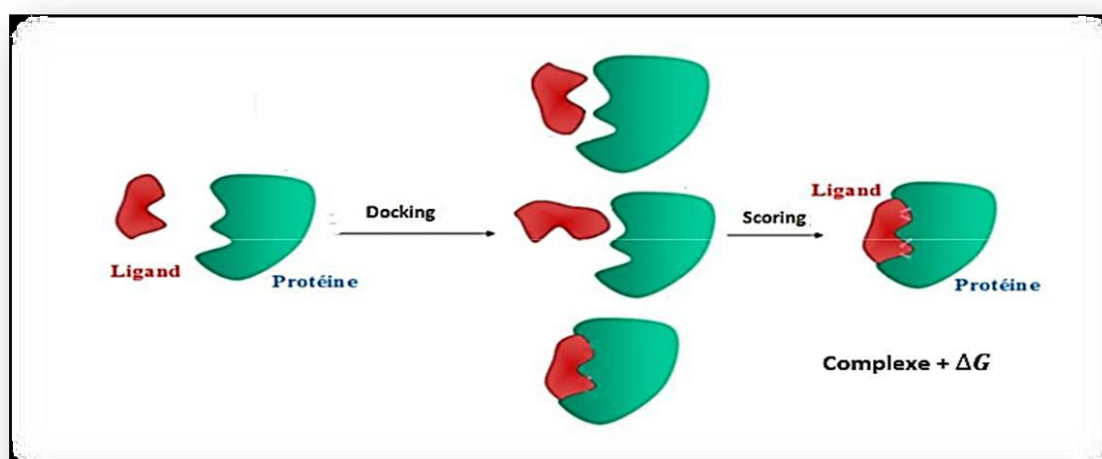

Figure .VI. 1. General principle of a Docking program

VI.2. Drug discovery parameters:

VI.2.1. the RMSD (Root Meany Square Deviation):

In bioinformatics, the RMSD is used to compare superposition of protein structures. Indeed, for this, a simple measure is to calculate the RMSD by taking as values the positions of the atoms in space. We can thus calculate the spatial distance between two atoms that are supposed to be superimposed. We will square, calculate the average of the distances, apply the square root and obtain a measurement of the quality of the superimposition.

Corresponds to the average of the deviation of each of the atoms compared to those of the original molecule .The best possible means that the value of the RMSD between the installation of the ligand calculated by the software and the conformation in the experimental complex is the most small as possible .The positioning, i.e .the correct identification of the binding site on the protein, the orientation and the conformation of the ligand influence the value of the RMSD [3].

The RMSD between two poses is a geometric measure of the distance between the atomic positions of two structures. The more accurate the positioning prediction, the more the differences between the two structures are small, the lower the value of the RMSD [4]. The RMSD value indicates good consistency in the following table:

Table VI.1. The RMSD values given by the MOE software

RMSD	RMSD< 1.5	1.5< RMSD< 3.5	3.5< RMSD< 6	6 < RMSD
Structure	Perfect	Acceptable	Inadequate	Unacceptable

VI.2.2. Receptor-Ligand Interactions:

At the active site, the interaction process takes place by complementary binding of the ligand with the protein structure. The forces involved in the formation of the complex can be electronic or Sterics in nature or both with different contributions [5].

There are four types of ligand-protein interactions. Described in the following:

A. The hydrogen bond:

Hydrogen bond is a dipole-dipole non-covalent chemical bond between an atom of hydrogen (H), bonded by an electronegative atom donor (Y, N, F) or a second electronegative atom with an unshared pair of electrons (acceptor).

B. Van Der Waals interactions:

Van der Waals bond is an interatomic potential due to a weak electrical interaction between two atoms or molecules, or between a molecule and a crystal.

These Van Der Waals interactions apply at very short distances and therefore only concern surface atoms. They are numerous and contribute essentially in research the steric match between the ligand and the receptor protein. [6]

Figure .VI. 2. Interactions de Van Der Waals

C. Hydrophobic interactions:

Amino acids whose radical is hydrophobic and a polar have the property of preventing the formation of hydrogen bonds between water molecules. They form hydrophobic areas in the protein structure where water molecules cannot exchange any bonds with radical's amino acids [7].
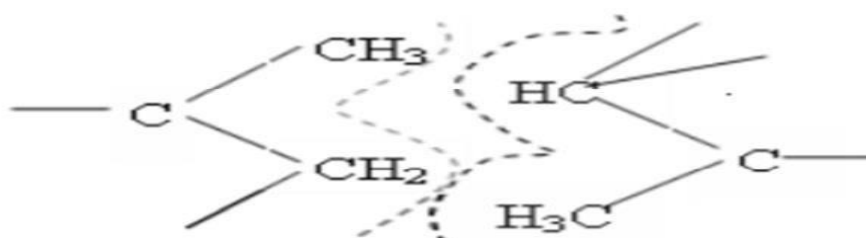

Figure .VI. 3. Hydrophobic interaction

D. Ionic bonds:

Ionic bonds represent fairly large bond strength between functional groups that carry opposite charges. The energies of these bonds about 20 kJ mol^{-1}.

VI.3. Preparation of receptor and ligands:

VI.3.1. Protein Preparation Structure:

The enzyme used in this part is EGFR kinase, the latter was downloaded from a Bookhaven Protein Data Bank (pdb) with the access code is 4hjo.

The 3D structure of EGFR was obtained by X-ray diffraction with a resolution of 2.73 Å. Note that EGFR co-crystallizes with the reference ligand pyrazole in monomer form (A chain). Water molecules were removed and the cofactors to obtain a simple model of the enzyme, and then we optimized the receiver using MOE with a MMFF94X force field and we minimized its energy so as to have the better conformation. This allowed us to obtain the model shown in (Figure VI .4).

Figure .VI.4. The simplified three-dimensional crystal structure model of the target protein

VI.3.2. Preparation of the ligands:

The molecules used in this work are derivatives of pyrazole, their structures are shown in (Table IV.1), using ChemDraw ultra software (8.0).

All structures have been optimized using the HyperChem 8.0 program. The molecules thus obtained are recorded in mol format, using the mechanics molecular (MM+) or the force field. The MM+ model is based on an approach semi-empirical, the objective of which is to optimize the geometry of the molecule and to determine energy and electronic distribution. [2]

Then we optimized again the geometry of the ligands was carried out using the force field (MMFF94X) implanted in the MOE software, professional version to determine the most conformation steady.

In our work, we studied the existing interactions between inhibitors (pyrazole derivatives) and the EGFR enzyme using molecular docking as a calculation method.

Table VI. 2. Optimization energies of all inhibitors by HyperChem 8.0

Ligand	MM (Kcal/mol)	AM1 (Kcal/mol)
1	14.0343	-2533.9327
2	13.8691	-2506.1755
3	13.8289	-2492.5059
4	8.8722	-4639.6990
5	13.2929	-4629.7116
6	9.1325	-4460.7397
7	13.8578	-3242.2649
8	16.1008	-2527.8966
9	18.7745	-2500.7746
10	19.5063	-2486.5378
11	15.6083	-2471.5909
12	15.4498	-2443.8865
13	15.4134	-2432.6228
14	11.2989	-4040.4816
15	15.7375	-4130.4902
16	11.6656	-3861.8796

17	18.2461	-2631.9497
18	17.5682	-2462.6056
19	20.2533	-2435.2673
20	20.9919	2424.6220
21	16.4884	-2443.3070
22	16.3280	-2415.6037
23	14.2355	-2405.0882
24	12.7200	-4012.8158
25	17.1587	-4102.6526
26	13.1178	-3833.6779
27	19.1258	-2603.6286
28	18.4506	-2384.4062
29	21.3197	-2410.5441
30	16.2074	-2400.2549

VI.3.3.Cavity detection:

The identification of active sites is of great importance for understanding the function of a protein and the mechanism of interactions. Moreover, knowledge of these functional sites can be used to guide mutagenesis experiments. There is a number of cavities or pockets on the surface of proteins where small molecules. Therefore, the identification of such cavities is often the starting point. Protein-ligand binding site prediction for protein function annotation and structure-based drug design [8].

In our work, we chose cavity 1. The MOE 2014 software allowed identify and present the residues that form the active sites (the cavities) using the "Site Finder" module. We give the properties of the first cavity detected by MOE in the table (Table VI.3).

Cavity 1 was chosen for our docking calculation because it has:

1. The ligand that co-crystallized from EGFR.

2. A large volume compared to other cavities.

3. The same active site residues from the literature.

Table VI. 3. Different cavity properties detected by MOE of EGFR

| Site	Size	PLB	Hyd	Side	Residues
1	309	3.79	91	151	LEU694 GLY695 SER696 GLY697 ALA698 PHE699 GLY700 THR701 VAL702 ALA719 ILE720 LYS721 MET742 VAL745 VAL750 CYS751 ARG752 LEU753 LEU764 ILE765 THR766 PRO770 PHE771 GLY772 CYS773 LEU775 ASP776 ARG779 ARG812 ASP813 LEU814 ARG817 ASN818 LEU820 ILE829 TYR830 ASP831 PHE832 LEU834 ALA835 LEU838 GLY839 ALA840 TYR845 HIS846 ALA847 GLY850 LYS851 VAL852 PRO853 ILE854 LYS855 TRP856 MET857 ALA 858 SER861 ILE862 ARG865 TYR867 SER871 ASP872 TRP874 SER875 GLU882 GLY887 SER888
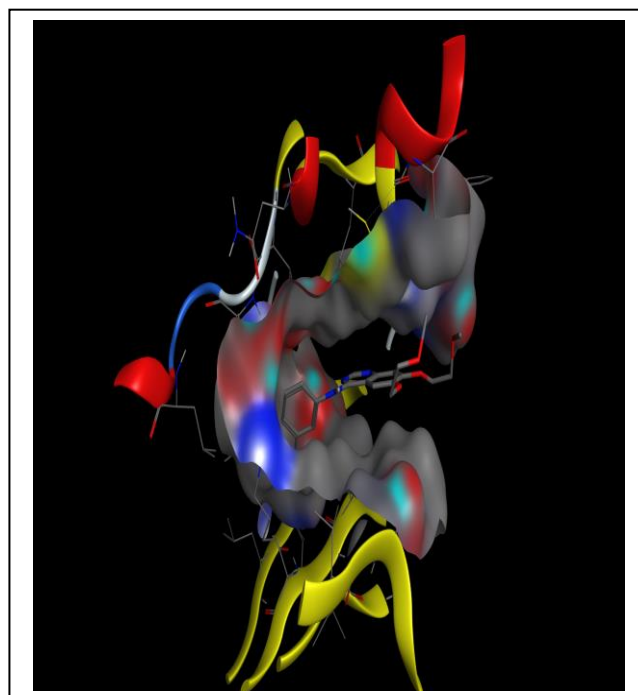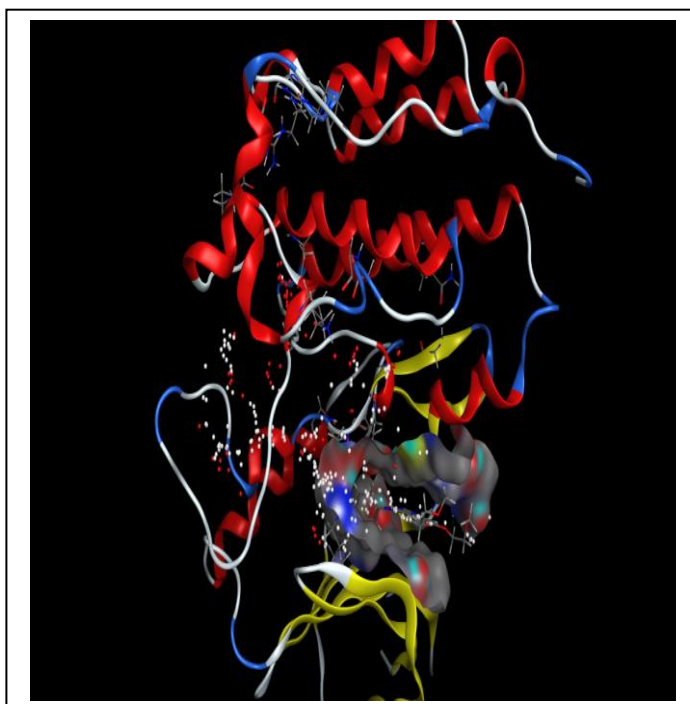
Figure .VI. 5. Cavity 1 of EGFR enzyme

VI.4. Results and interpretations:

VI.4.1. Protein-ligand interactions:

We used the Docking method using MOE 2014 software. It has based on a type of semi-flexible docking and generally used for protein ligand docking, the ligand being considered flexible and the main chain of the enzyme has been kept rigid, while the side chains remain flexible. Once the ligand-receptor complex is formed, it will adapt the most stable conformation, i.e. with the lowest energy level. The results in the table follow:

Table VI. 4. SCORE results of molecular docking of pyrazole derivatives

Ligand	S	RMSD_refine	Ligand	S	RMSD_refine
C1	-6.3143	1.7209	C16	-6.6341	1.1597
C2	-6.6931	1.9648	C17	-7.0383	3.0050
C3	-6.6379	3.0623	C18	-6.7453	2.6540
C4	-6.6322	1.2784	C19	-6.9612	2.3584
C5	-6.7723	1.7595	C20	-6.7033	2.3683
C6	-6.5837	0.5853	C21	-6.5923	4.7629
C7	-6.8960	3.1182	C22	-6.9883	1.5917
C8	-6.5186	4.1428	C23	-6.9689	1.0826
C9	-6.9182	2.3193	C24	-6.7433	2.0012
C10	-6.8799	1.5760	C25	-6.5943	1.3674
C11	-6.7215	2.0635	C26	-6.5401	2.1310
C12	-6.8927	2.8498	C27	-7.2198	1.4879
C13	-6.8237	1.1802	C28	-6.7459	4.3383
C14	-6.9248	2.6934	C29	-6.8776	1.6707
C15	-6.8556	1.2500	C30	-6.9767	1.9104

Table VI. 5. SCORE results of molecular docking of reference ligand

Ligand	S	RMSD_refine
aq4 co-crist	-7.6483	1.4010

We selected five molecules C17, C22, C23, C27 and C30 with which have the closest score energy to the reference ligand score energy this shows that these complexes are more stable. We can classify in the following order:

Aq4 < C27< C17< C22< C30< C23.

According to the RMSD results of the selected molecules we have: C23 and C27 have the perfect docking accuracy. Because the RMSD of molecules less than 1.5.

On the one hand, the other molecules C22, C17 and C30 have acceptable docking accuracy. Because are RMSD between [1.5-3.5].

• **Reference ligand:**

The study of the interaction of (enzyme-ligand of reference) is important to make by comparing it with other ligands the following table represents the different interactions between active site residues and reference ligand aq4.

Table VI. 6. The ratio of interactions between active site residues with aq4

Ligand	Receptor	Interaction	Distance	E (Kcal/mol)
C14	29 O----MET769	H-donor	3.26	-0.8
C19	44 O----GLN767	H-donor	3.18	-0.9
N2	43 N----MET769	H-acceptor	2.92	-4.8

According to Anne Imbert and col, Interactions having distances between 2.5 Å and 3.1 Å are considered strong and those between 3.1 Å and 3.55 Å are assumed to be average and when their distances are greater than 3.55Å they are considered weak. **[9]**

Visual analysis shows that the reference ligand forms three interactions with the active site of EGFR **(Figure VI.6, Table 6):** the first H-donor type interaction average (between the C14 atom of the reference ligand and the oxygen of the residue MET769) with a distance of 3.26 Å and the average second donor type H-donor (between the C19 atom of reference ligand and the oxygen of residue GLN767) with a

distance 3.18 Å, the third type of strong H-acceptor (between the N2 atom of reference ligand and the NH2 amine function of residue MET 769) with a distance of 2.92 Å.
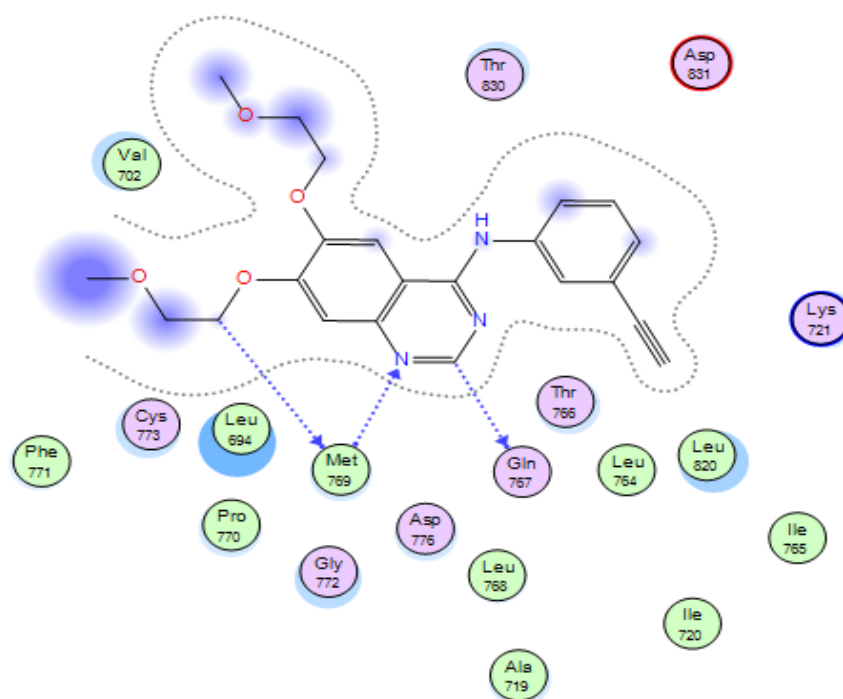

Figure.VI.6. The 2D interaction of reference ligand with active site residues

EGFR

•**Ligand C27:**

Table VI. 7. The ratio of interactions between active site residents with C27

Ligand	Receptor	Interaction	Distance	E (Kcal/mol)
C	21 SD-----MET742	H-donor	4.26	-0.7
N	35 O-----PHE832	H-donor	3.36	-1.5
O	38 O-----CYS751	H-donor	2.71	-3.0
S	34 N-----THR766	H-acceptor	3.75	-1.1
S	34 OG1----THR766	H-acceptor	3.21	-1.3
6-ring	CG2-----VAL702	Pi-H	4.11	-0.6

According to **Table VI.7** it is observed that the ligand C27 carries six types of interactions with the site active EGFR: the first H-donor type interaction weak (between the C atom of ligand C27 and the SD of the residue MET742) with a distance of 4.26 Å and the average second H-donor type (between the N atom of ligand C27 and the oxygen of residue PHE832) with a distance 3.36 Å, the third type of strong H-donor (between the O atom of ligand C27 and the oxygen of residue CYS 751) with a distance of 2.71 Å.

The fourth weak H-acceptor type (between the S atom of ligand C27 and the N of residue THR766) with a distance of 3.75 Å, the average fifth H-acceptor type (between the S atom of ligand C27 and the OG1 of residue THR766) with a distance of 3.21 Å, and the last type weak pi-H (between the 6-ring atom of ligand C27 and the CG2 of residue VAL702) with a distance 4.11Å.
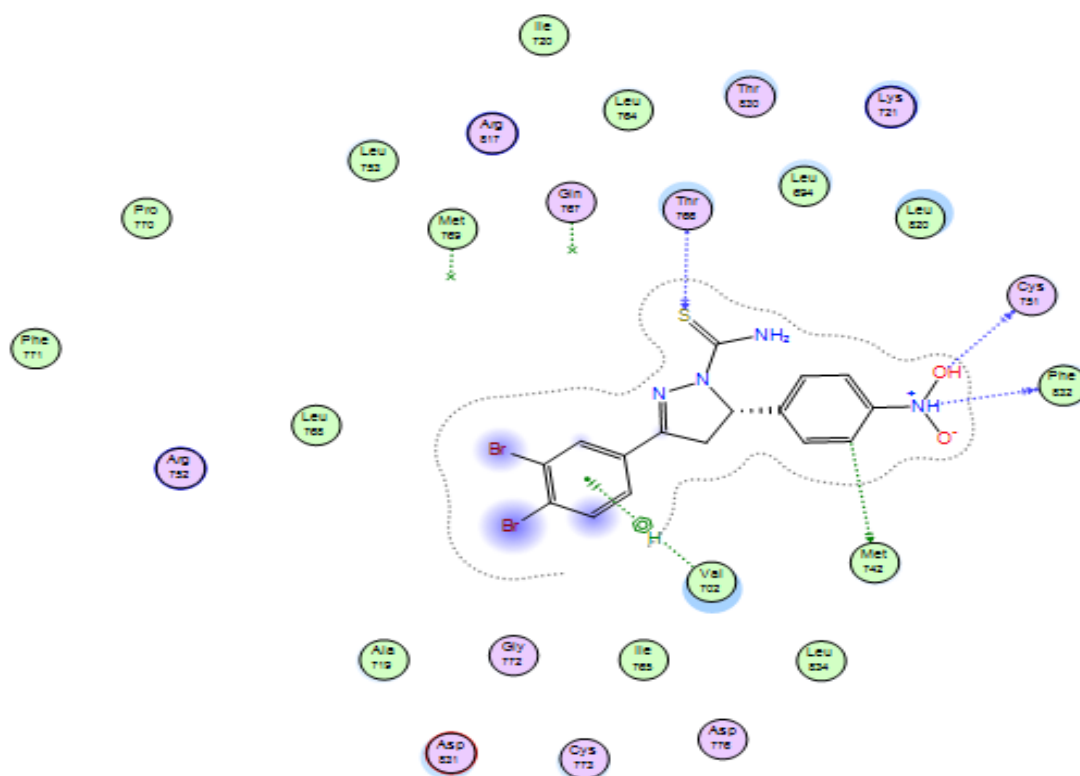

Figure .VI.7. 2D interactions of C27 ligand with EGFR active site residues

•**Ligand C17:**

Table VI.8. The ratio of interactions between active site residents with C17

Ligand	Receptor	Interaction	Distance	E (Kcal/mol)
N	35 O----MET769	H-donor	3.06	-6.4

Visual analysis of ligand C17 and EGFR active site residues indicates the presence one type interaction average H-donor (between the N atom of ligand C17 and the O of residue MET769) with a distance of 3.06 Å.
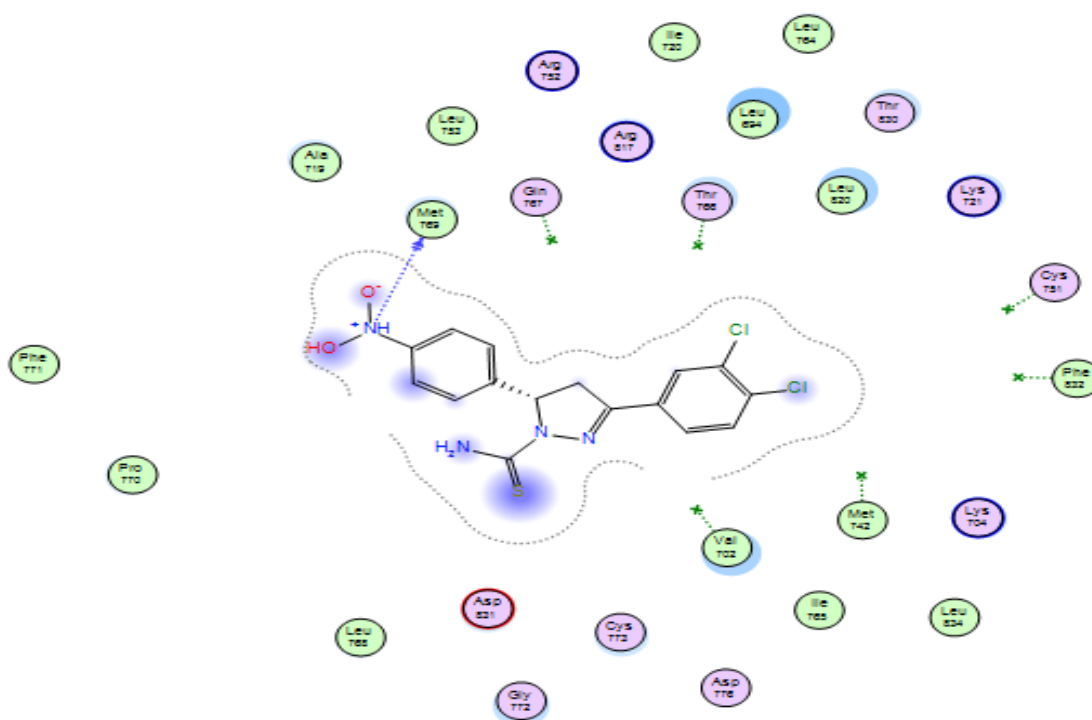

Figure .VI. 8. 2D interactions of C17 ligand with EGFR active site residues

•**Ligand C22:**

Table VI.9. The ratio of interactions between active site residents with C22

Ligand	Receptor	Interaction	Distance	E (Kcal/mol)
Br	28 O----ALA719	H-donor	3.38	-0.5
Br	28 O----LEU764	H-donor	3.40	-0.5
6-ring	CB----LEU694	Pi-H	4.26	-0.9

The ligand C22 has three interactions which represent in (**the table VI.9, the figure VI.9**): the first type weak pi-H (between the 6-ring atom of ligand C22 and the CB of residue LEU694) with a distance 4.26Å.and two type interactions average H-donor (between the Br atom of ligand C22 and the oxygen of residue ALA719 and the oxygen of residue LEU764) with a distance of 3.38 Å and 3.40 Å,
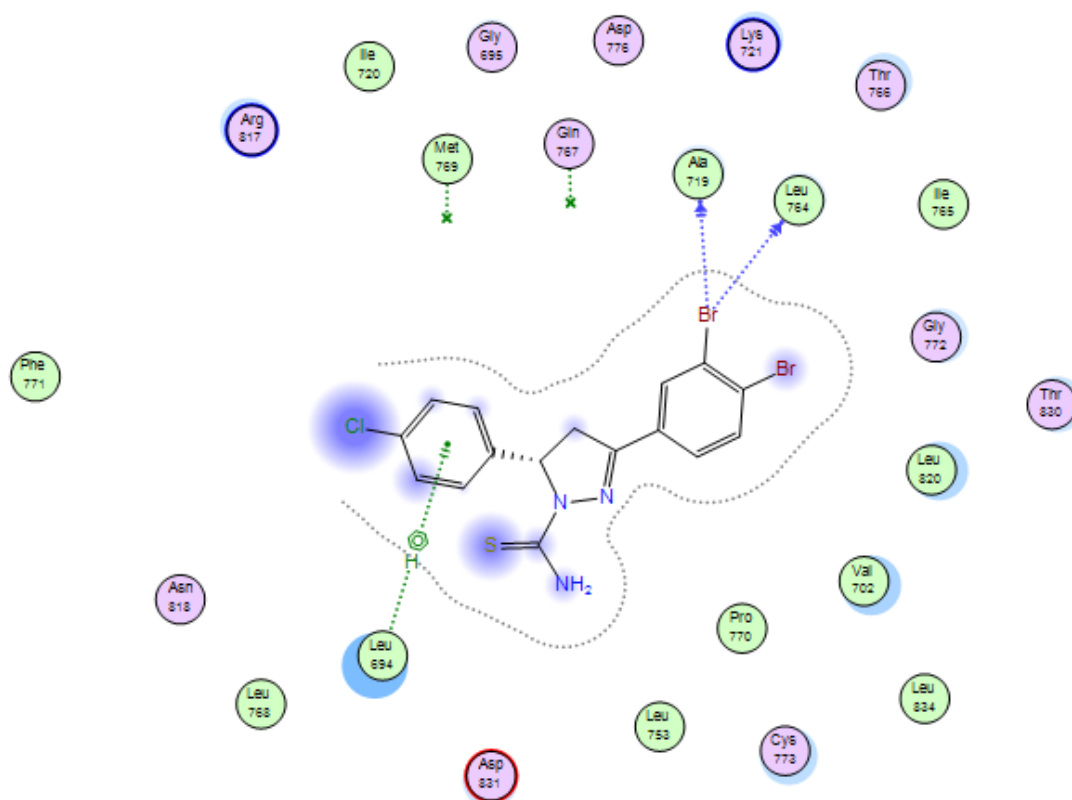

Figure .VI.9. 2D interactions of C22 ligand with EGFR active site residues

•Ligand C30:

Table VI.10. The ratio of interactions between active site residents with C30

Ligand	Receptor	Interaction	Distance	E (Kcal/mol)
Br	28 O----LEU764	H-donor	3.59	-0.5

Visual analysis of ligand C17 and EGFR active site residues indicates the presence one type interaction Weak H-donor (between the Br atom of ligand C30 and the O of residue LEU764) with a distance of 3.06 Å.
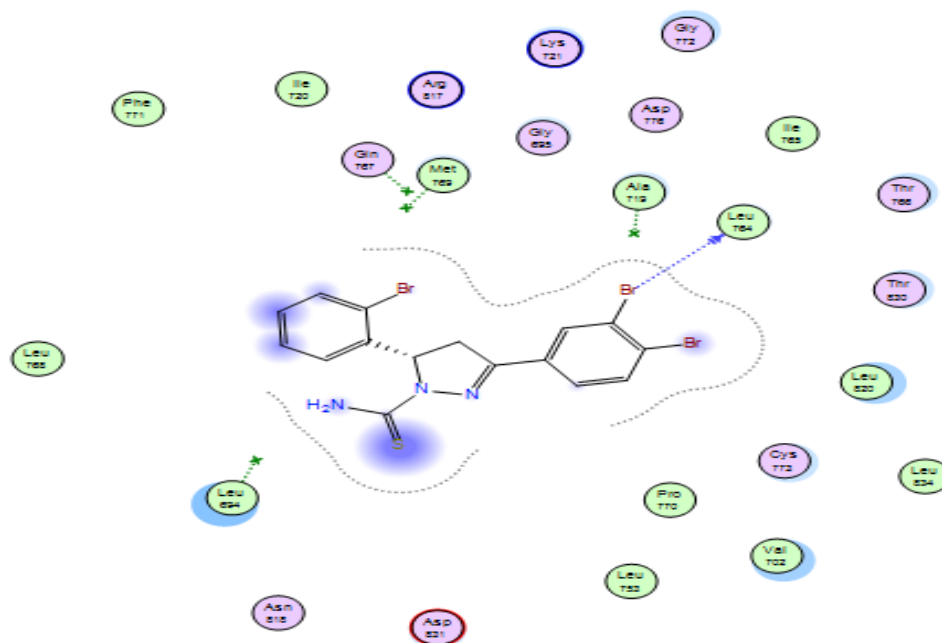
Figure .VI.10. 2D interactions of C30 ligand with EGFR active site residues
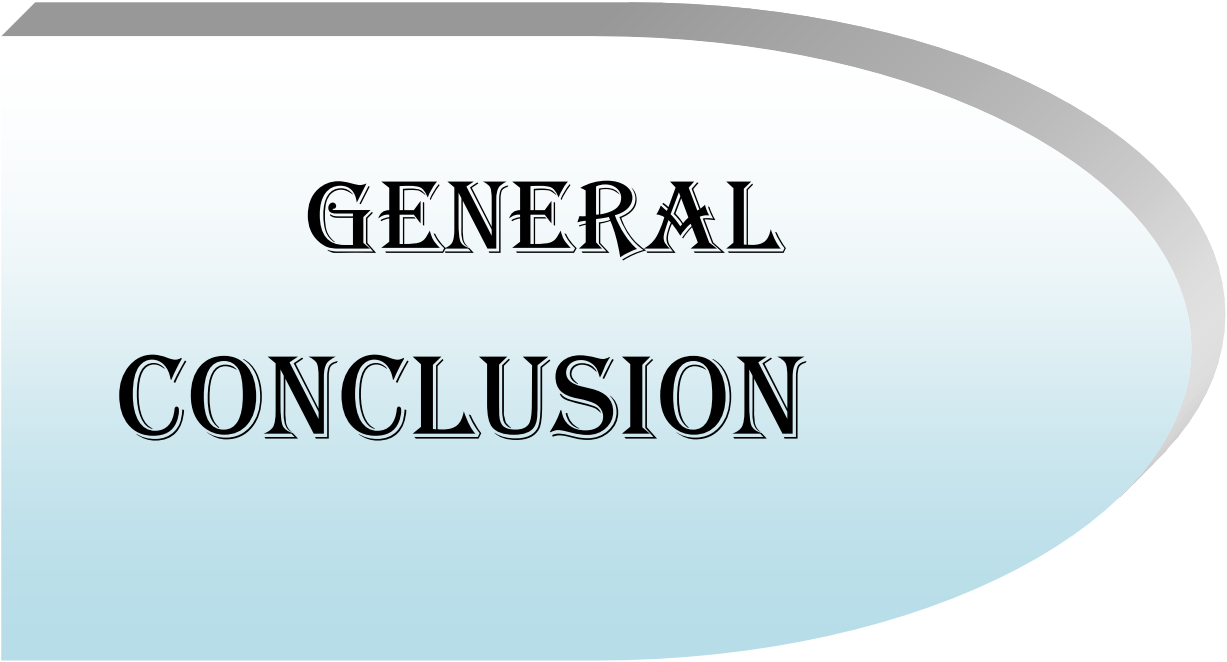
•**Ligand C23:**

Table VI.11. The ratio of interactions between active site residents with C23

Ligand	Receptor	Interaction	Distance	E (Kcal/mol)
S	34 N----THR766	H-acceptor	3.74	-1.0
S	34 OG1----THR766	H-acceptor	3.21	-1.3
6-ring	CG2----VAL702	Pi-H	4.10	-0.7

The ligand C23 has three interactions, which represent in (**the table VI.11, the figure VI.11**): the first type weak pi-H (between the 6-ring atom of ligand C23 and the CG1 of residue VAL702) with a distance 4.10Å.and two type interactions H-donor, the first one weak (between the S atom of ligand C23 and the amine function of residue THR766).in addition, and the second one average (between the S atom of ligand C23 and OG1 of residue THR766) with a distance of 3.74 Å and 3.21 Å.

Figure .VI.11. 2D interactions of C23 ligand with EGFR active site residues

VI.5. Conclusion:

Virtual screening methods are widely used for reducing cost and time of drug discovery process [10].

From this study of in silico drug designing and molecular docking of the 30-pyrazole derivatives, we conclude that those derivatives have the ability to inhibit the EGFR protein kinase. The scoring results reveal from the docking score, and the RMSD obtained by MOE software that C17, C22, C23, C27 and C30 showed better results from 30 docked ligands.

References:

[1] Fatouche Maroua, Master's thesis in pharmaceutical chemistry, Mohamed Khider University of Biskra, 2016**.**

[2]A. Imberty, K. D. Hardman .J.P .Carver, S .Perez, 1; 631- 642; 1991.

[3] Kramer, Bernd, Matthias Rarefied, and Thomas Lindauer, 37:2: 228-24, 1999.

[4] Sehil.M Ben Abdellah Master's thesis in pharmaceutical chemistry, Mohamed Khider University of Biskra, 2021.

[5] BEN MOHAMED.I., M.H, Master's thesis in pharmaceutical chemistry, Mohamed Khider University of Biskra, 2019.

[6] Teno.S, University surrounded me Constantine, 2012.

[7] SALAHI.D, Master memory .University of Mouloud Mammeri Tizi Ouzo, 2014.

[8] Zhang. Z, et al, 27; 15; 2083-2088, 2011.

[9] Imberty .A, Hardman .K.D, Carver .J.P, Pérez. S, 1; 631-642, 1991.

[10] S.khamouli, doctoral thesis, University Mohamed Khider of Biskra, 2019.

GENERAL

CONCLUSION

In this work, we used computational chemistry to study the biological anticancer activity on the series of pyrazole derivatives which present good drug candidates for anticancer treatment.

In the first part, we have two special chapters on bibliographic research; the first chapter is the research on the basic molecule and the methods of synthesis of pyrazole, and the second chapter talks about the calculation method used in this work.

On the one hand, in the second part, we made a theoretical study of the structural and electronic parameters of the privileged conformation of the basic nucleus, in addition we closed chapter 3 with a synthesis of a pyrazole derivative from a synthesized intermediary "chalcone".

From the theoretical study of chapter 3, it was found that the method (DFT) with base Medium (6-31G*) is the most appropriate method to perform calculations on the pyrazole nucleus.

After the basic molecule study, we went on to study QSAR and drug-like properties of a bioactive series of pyrazole derivatives. In our case, all the structures studied respect the rules of Lipinski and Veber .Concluded that, the compounds of the series are acceptable for oral administration.

To simplify the Predicting the biological activity of untested compounds from "descriptors" ... chemical properties of substances we created the QSAR model from physical-chemical descriptors by the SPSS software, and we found a good model by six descriptors (HBD, logD, MW, logD, NRB and Polarizability).

In the last part, we did a docking study using MOE software, was performed on these compounds to find their interactions with the EGFR receptor. The scoring results reveal from the docking score, and the RMSD, We have selected: C17, C22, C23, C27 and C30 showed better results from 30 docked ligands.

Abstract:

The widespread use of the microscope in the 18th century led to the discovery of the cancer disease which remains to this day the deadliest of diseases. What has prompted most researchers and chemists to intensify the study on the discovery of new molecules to inhibit them. In this work, we studied a series of pyrazole derivatives. This occupies an important place in the pharmaceutical field because of the multiplicity of its biological activities.

In this study, Lipophilic Efficacy (LipE), Lipinski and Veber of Multifactor Optimization (MPO) process was used to predict the best balance of properties of these compounds. The Multiple Linear Regression (MLR) method was also applied to derive the QSAR model .In silico molecular docking to present the interactions between the studied molecules and the enzyme (EGFR).

Keywords: cancer disease, biological activities, pyrazole, MPO, MLR, QSAR, docking, EGFR.

Résumé :

La généralisation de l'utilisation du microscope au 18ème siècle a conduit à la découverte de la maladie cancéreuse qui reste à ce jour la plus mortelle des maladies. Ce qui a poussé la plupart des chercheurs et chimistes à intensifier l'étude sur la découverte de nouvelles molécules pour les inhiber. Dans ce travail, nous avons étudié une série de dérivés du pyrazole. Celle-ci occupe une place importante dans le domaine pharmaceutique du fait de la multiplicité de ses activités biologiques.

Dans cette étude, le processus Lipophilic Efficacy (LipE), Lipinski and Veber de Multi factor Optimization (MPO) a été utilisé pour prédire le meilleur équilibre des propriétés de ces composés. La méthode de régression linéaire multiple (MLR) a également été appliquée pour dériver le modèle QSAR. Docking moléculaire in silico pour présenter les interactions entre les molécules étudiées et l'enzyme (EGFR).

Mots clés : maladie cancéreuse, activités biologiques, pyrazole, MPO, MLR, QSAR, docking, EGFR.

ملخص :

أدى الاستخدام الواسع النطاق للميكروسكوب في القرن الثامن عشر إلى اكتشاف مرض السرطان الذي لا يزال حتى يومنا هذا أكثر الأمراض فتكًا. ما دفع معظم الباحثين والكيميائيون إلى تكثيف الدراسة عن اكتشاف جزيئات جديدة تثبطها. في هذا العمل ، درسنا سلسلة من مشتقات البيرازول. يحتل هذا مكانة مهمة في المجال الصيدلاني بسبب تعدد أنشطته البيولوجية.

في هذه الدراسة، تم استخدام عملية محبة الدهون (LipE) و Lipinski وVeber, في(MPO) للتنبؤ بأفضل توازن لخصائص هذه المركبات. تم تطبيق طريقة الانحدار الخطي المتعدد (MLR) أيضًا لاشتقاق نموذج QSAR. في الالتحام الجزيئي السيليكي لعرض التفاعلات بين الجزيئات المدروسة والإنزيم (EGFR).

الكلمات الدالة : مرض السرطان، نشاط البيولوجي، البيرازول،MLR، MPO،QSAR،الالتحام الجزئي، EGFR.