Mohamed Khider University of Biskra
Faculty of exact sciences
Matter sciences department

MASTER MEMORY

Science of matter
Chemistry
Pharmaceutical Chemistry

Réf. :

Presented by

Zeghou Faiza

03-06-2025

THE:

Quantitative structure –activity relationship study of isatin derivatives

---

Jury:

Dr.	Kennouche Samir	S.L.A	Mohamed Khider university Biskra	President
Dr.	Hazhazi Halima	S.L.B	Mohamed Khider university Biskra	Supervisor
Dr.	Benbrahim Imane	S.L.B	Mohamed Khider university Biskra	Examiner

AcademicYear: 2024-2025

Dedication

To my parents

to my husband

To my children

To all the people whom love me

Acknowledgments

First and foremost, I thank Almighty God for granting me the strength to endure and the courage to overcome all difficulties.

I would like to express my heartfelt gratitude to my supervisor, Mrs Halima Hazhazi from the University of Biskra, for proposing this topic, accepting to supervise it, and guiding me throughout the process. I am deeply thankful for her constructive criticism, scientific guidance, Support, patience, and encouragement during the course of this work, which have greatly contributed to the completion of this thesis.

I also extend my sincere thanks to the jury members, Mr. Kennouche Samir and Mrs Imane Benbrahim, for the honor of accepting to evaluate my work.

Finally, I wish to thank my entire family, especially my mother.

Contents

Chapter II

Background on theoretical bases

Chapter III

Results and discution

List of tables

List of figures

List of Abbreviations

1D	One-Dimensional
B3LYP	Becke 3-parameter lee-yang-parr
DFT	Density-functional theory
Eg	HOMO-LUMO energy gap
HF	Hartree-Fock
HOMO	Highest Occupied Molecular Orbital
IC50	The Half Maximal Inhibitor Concentration
LOO	Leave one out
LUMO	Lowest Unoccupied Molecular Orbital
MLR	Multiple Linear Regression
MM	Molecular Mechanics
MR	Molecular Refractivity
MW	Molecular Weight
Pol	Polarizability
PRESS	Predicted Residual Sum of Squares
QSAR	Quantitative Structure-Activity Relationship
SAG	Surface Area Grid
SSY	Sum of The Squares of The Response Value

General Introduction

General Introduction

Isatin or (1H-indole-2,3-dione) is a heterocyclic compound with an oxidized indole structure, first isolated from the oxidation products of indigo in 1841. Since its discovery, isatin has attracted considerable attention due to its wide range of biological activities, as anticancer activity, cytotoxic and antineoplastic activites [1,2].

Isatin derivatives form an important class of bioactive molecules in medicinal chemistry. These derivatives, obtained by modifying the core structure of isatin, have shown enhanced pharmacological efficacy against various biological targets, making them promising candidates in the development of new therapeutic agents, especially against cancer diseases [3,4].

Nowadays, molecular modeling has become an essential tool in studying the chemical and biological properties of molecules. This approach allows for the visualization and understanding of molecular interactions and helps predict the behavior of new chemical structures. It relie on quantum chemical calculations and simulations to estimate electronic, geometric, and energetic properties of organic compounds [5,6].

Quantitative Structure–Activity Relationship (QSAR) is one of the most commonly methods used in this field. This approach aims to establish a mathematical correlation between molecular descriptors (such as frontier orbitals, lipophilicity, electrostatic potentials...) and experimentally observed biological activity. QSAR models make it possible to predict the potential effectiveness of new compounds before synthesis [7].

The prediction of biological activity is based on mathematical and statistical calculations; therefore, a statistical method, known as multiple linear regression (MLR), is employed [8].

Multiple Linear Regression (MLR) is a statistical method used to quantify the relationship between independent variables obtained through calculations and a dependent variable determined experimentally. This method relies on a set of statistical parameters that define the

linear combination between the independent variables (molecular descriptors) and the biological activity of the studied molecules [9].

The main objective of this study is to develop QSAR model between molecular descriptors of isatin derivatives and their anticancer activity.

The manuscript of this thesis is divided into three parts:

- ***Chapter I:*** General Information on Isatin (1H-indole-2,3-dione) and Anticancer Activity

This chapter contains general informations about cancer disease, isatin and effect of isatin derivatives for anticancer activity against U937 cells.

- ***Chapter II:*** Background on theorical bases

In this chapter, we present theoretical background of the methods of quantum chemistry we used in this study and information about QSAR (objectives of QSAR, molecular descriptors, statistical parameters .....).

- ***Chapter III:*** Results and Discussion

This part presents results and interpretations obtained from the QSAR study of isatin derivatives for anticancer activity against U937 cells using physicochemical, electronic…. descriptors.

References

[1]. Pandeya, S. N., Sriram, D., Nath, G., & DeClercq, E. (1999). Synthesis and antimicrobial activity of isatin derivatives. *Indian Journal of Pharmaceutical Sciences*, 61(6), 358–361.

[2]. Silva, L. F., & Garden, S. J. (2001). The chemistry of isatins: a review. *Tetrahedron*,57(38), 7841–7852. https://doi.org/10.1016/S0040-4020(01)00715-4

[3]. da Silva, A. F., de Souza, M. V. N., & Ferreira, V. F. (2018). Biological potential of isatin and its derivatives: A review. *European Journal of Medicinal Chemistry*, 150, 920–928. https://doi.org/10.1016/j.ejmech.2018.03.030

[4]. Verma, M., Chandra, R., & Kumar, S. (2014). Isatin and its derivatives: A survey of recent patent literature. *Expert Opinion on Therapeutic Patents*, 24(9), 1027–1041. https://doi.org/10.1517/13543776.2014.943261

[5]. Leach, A. R. (2001). *Molecular Modelling: Principles and Applications* (2nd ed.). Pearson Education.

[6]. Cramer, C. J. (2013). *Essentials of Computational Chemistry: Theories and Models* (2nd ed.). Wiley. https://doi.org/10.1002/9781118712276

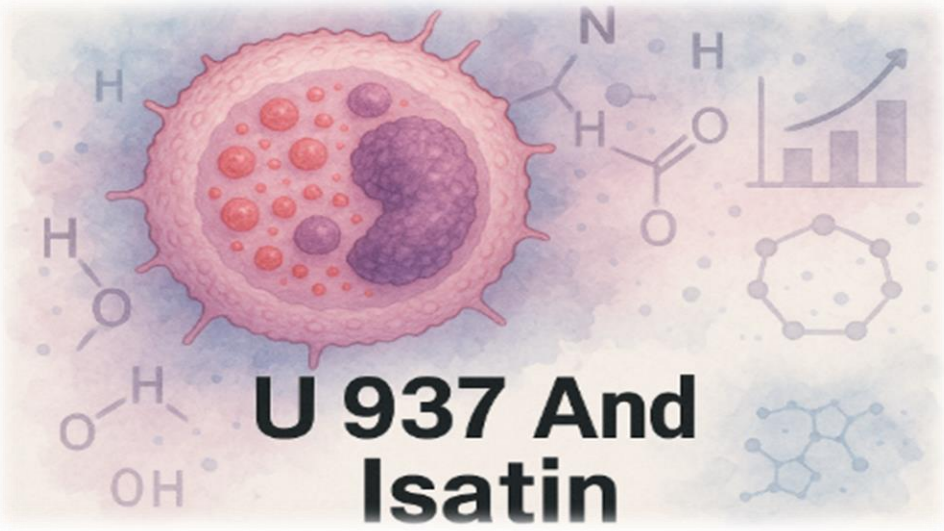
[7]. Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics* (2nd ed.). Wiley-VCH. https://doi.org/10.1002/9783527628766

[8]. Roy, K., Kar, S., & Das, R. N. (2015). *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*. Springer. https://doi.org/10.1007/978-3-319-16706-0

[9]. Goudarzi, N., Goodarzi, M., (2012). QSAR prediction of HIV inhibition activity of styrylquinoline derivatives by genetic algorithm coupled with multiple linear regressions. Medicinal chemistry research vol 21, pp. 437-443.

Chapter I

General Information on Isatin (1H-indole-2,3-dione) and Anticancer Activity



I. 1. Cancer

I.1.1. Introduction

The human body is composed of more than 60 trillion cells, which form the tissues and organs such as the heart, liver, and lungs. These cells constantly renew themselves to replace damaged or aged ones, allowing tissues to maintain their structure and function over time. This renewal process is tightly controlled by the nucleus of the cell, which contains chromosomes made up of DNA and genes. Occasionally, mutations occur in some of these genes, leading the nucleus to issue faulty instructions. As a result, the affected cell begins to divide abnormally and uncontrollably, passing the same mutation to its daughter cells. These abnormal cells proliferate chaotically, forming a tumor**.**

This process may take a long time, between the emergence of the first abnormal cell and the development of a tumor measuring approximately one cubic centimeter. During this time, the tumor stimulates the formation of new blood vessels to sustain itself, a phenomenon known as *angiogenesis*. The real danger arises when cancerous cells invade nearby tissues and spread through blood or lymphatic vessels to other parts of the body, forming metastases [1]. This is demonstrated in the figure I.1 below.

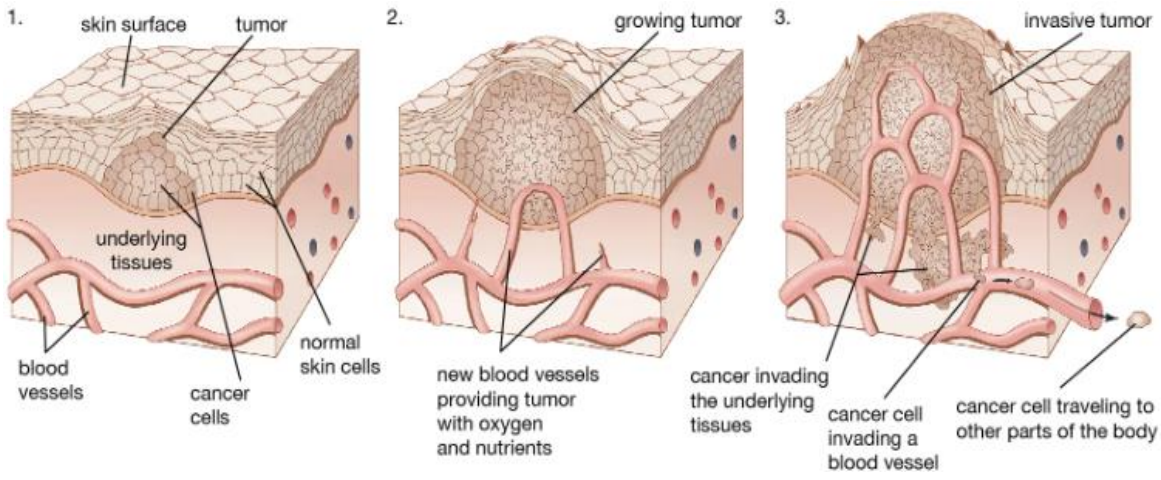


Figure I.1. Phenomenon angiogenesis and disease transmition.

I.1.2. Classification

I.1.2.1. Tumor

A tumor is an excessive cellular proliferation resulting in a new tissue formation that tends to persist and grow indefinitely, indicating a certain degree of biological autonomy.

⇨Types of Tumors

Tumors can be classified into two categories: benign tumors and malignant tumors, also known as cancers.

a) Benign tumors are surrounded by a capsule, which makes them non-invasive. They have limited growth and are rarely fatal.

b) Malignant tumors consist of cells capable of forming secondary tumor sites (metastases) that are located far from the primary tumor. Moreover, malignant tumors often show greater local aggressiveness in terms of invading and destroying the surrounding tissues. Figure I.2 below provides a clear illustration of this [2].

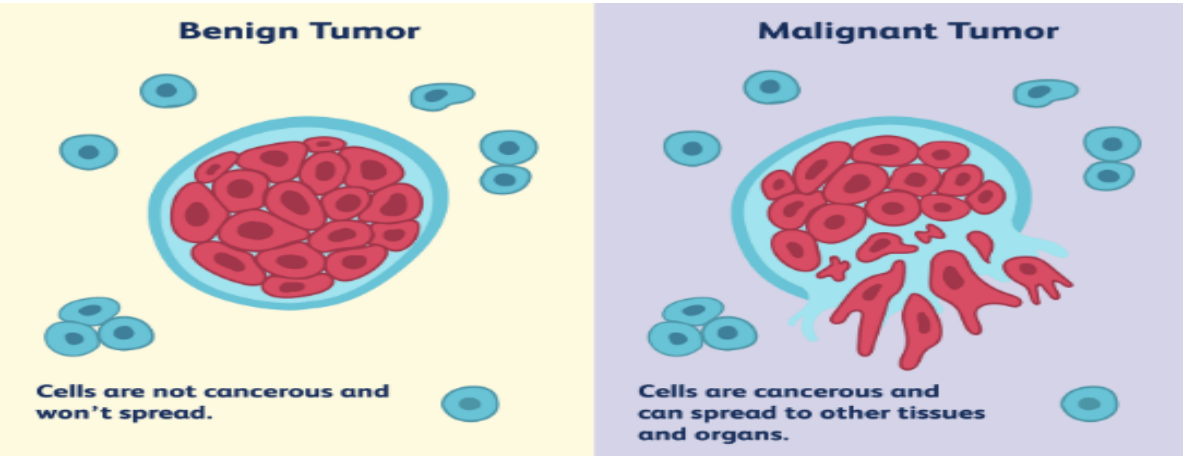


Figure I.2. Comparison between benign and malignant tumors [3].

I.1.2.2. Metastases

The term metastasis, meaning "change of place" in Greek, refers to the growth of secondary tumors at sites distant from a primary neoplasm (WHO). This term distinguishes malignant lesions from benign ones and characterizes the final stage of tumor progression. Metastatic growth is the leading cause of treatment failure and cancer-related deaths. In fact, **90%** of cancers patients die due to their metastases.

The term metastasis is generally reserved for the spread of tumor cells through the bloodstream or lymphatic system. However, dissemination through cerebrospinal fluid and transcoelomic spread is also possible. The majority of cancer patients (**60 to 70%**) present with either overt or occult metastases at the time of diagnosis, and the prognosis for most of them is poor. Figure3.Shows how cancer cells travel from the primary tumor to other organs like the lungs and liver through the blood or lymphatic system [2].

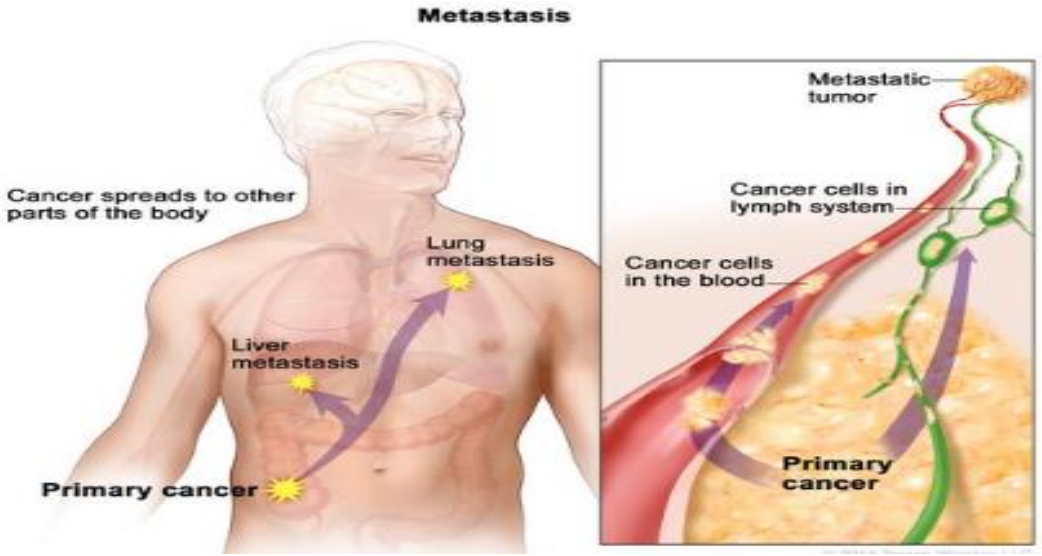


Figure I.3. How cancer cells spread through the blood and lymphatic systems [4].

And figure I.4. Explains how cancer spreads from the original organ to others via blood vessels and the lymphatic system.

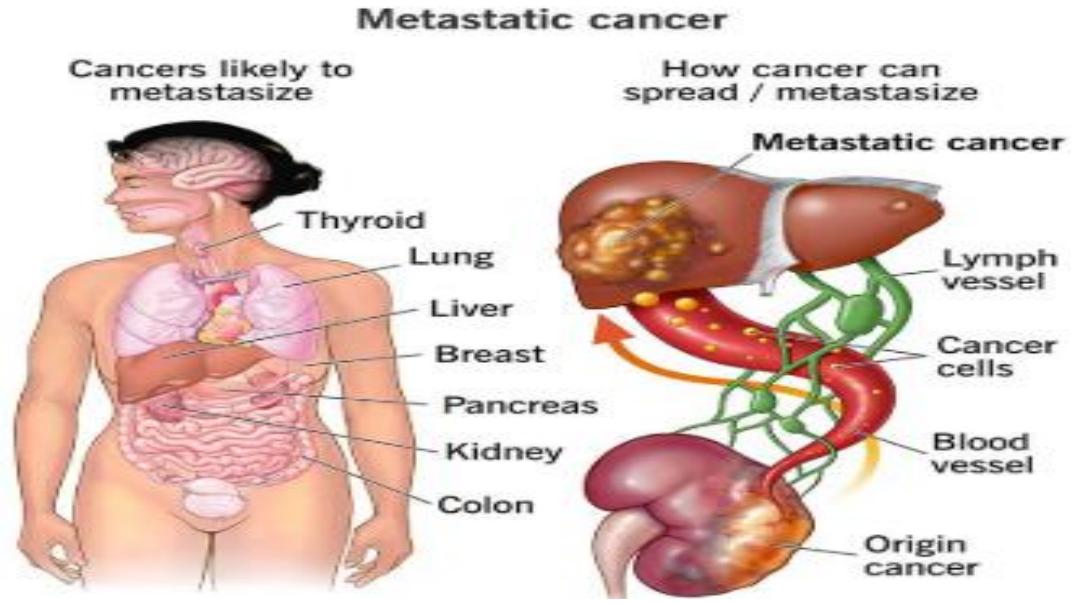


Figure I.4. Common sites of cancer metastasis in the human body [5].

Cancers can also be classified according to the type of tissue affected figure I.5.

- **Carcinomas:** These are the most common types of cancer. They originate from epithelial cells that cover the internal and external surfaces of the body, such as the lungs and colon.
- **Sarcomas:** These cancers affect connective tissues that support the structure of the body, including bone, cartilage, muscle, adipose, and vascular tissues. Their incidence is rare.
- **Lymphomas:** These affect hematopoietic tissues, particularly the lymph nodes and immune system organs.
- **Leukemias:** These involve the bone marrow, which is responsible for the production of white blood cells. This type of cancer is also rare.
- **Myelomasare:** are a type of blood cancer that begins in plasma cells [2].

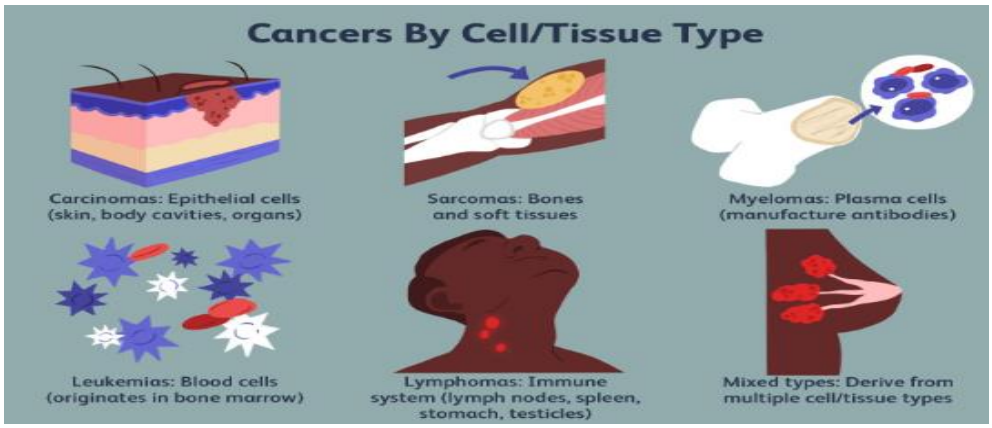


Figure I.5. The different cancer [6].

I.1.3. General Characteristics of Cancer Cells

- **Residual Characteristics**

Cancer cells retain, to a certain degree, the function of their differentiation and the characteristics of their tissue of origin.

- **Acquired Morphological Characteristics**

Morphological abnormalities in cancer cells are numerous, but none are constant or specific to cancer [2].

-**Nuclear abnormalities**: The nucleus is often large, irregular, and sometimes multiple. Nucleoli are visible. Chromosomal abnormalities

are frequent, most often of the hyperploidy type (increased chromosome number).

- **Cytoplasmic abnormalities**: The cytoplasm is often very basophilic with an increased nucleus-to-cytoplasm ratio.

-**Mitosis abnormalities**: Mitoses are generally larger than in normal tissues, with disoriented spindle formation, and may even be multipolar.

- **Cytoplasmic membrane abnormalities**: The only nearly universal definition of a cancer cell is its ability to proliferate without being properly controlled by normal regulatory mechanisms.

This inability to respond to regulatory signals from other cells and tissues naturally leads to studying the membranes of cancer cells, which are key intermediaries in intercellular communication. Major abnormalities include loss of contact inhibition, altered adhesiveness, and changes in surface antigens [2].

- **Dynamic Characteristics**

The proliferation rate of cancer cells is abnormal, high, autonomous, anarchic, fragile (with significant cell death due to hypoxia), and indefinite. Biochemical abnormalities in cancer cells are frequent and diverse [1]. All of this can generally be summed up in the following plan.

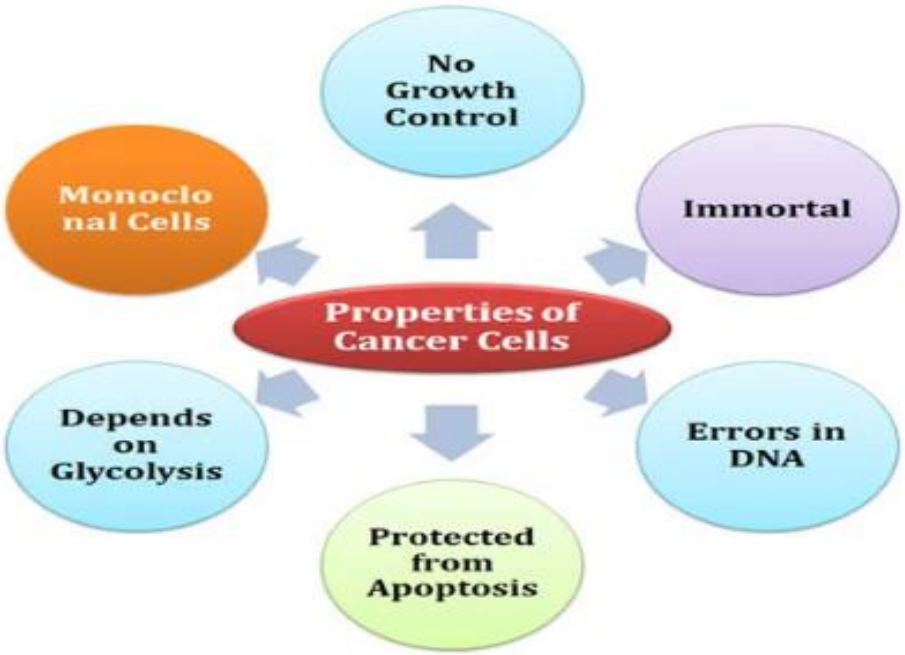


Figure I.6. Properties of cancer cells [2].

I.1.4. Comparison between normal and cancer cells

⇨ Normal cell

Cells are the fundamental building blocks of all living organisms. Under normal conditions, cells grow, divide, and die in a regulated manner.

⇨ Cancer cell

A cancer cell is a mutated cell that has lost its ability to regulate its growth and death. Instead of undergoing natural cell death (apoptosis), it continues to grow and divide abnormally, often forming a mass known as a tumor.The figure I.7 illustrates the general structure of each of them [1].

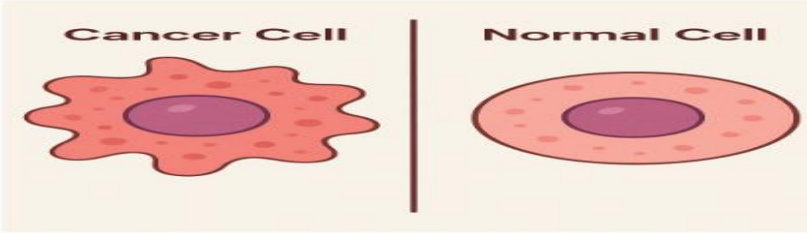


Figure I.7. Diagram comparing a normal cell and a cancer cell.

I.1.5. Major contributing factors to cancer

Recent epidemiological studies highlight several key contributors to cancer development:

⇨ Behavioral and Environmental Factors

- Tobacco Smoking: The single largest présentable cause of cancer. Responsible for approximately 30% of all cancer cases, particularly lung, oral, pharyngeal, bladder, pancreatic, and cervical cancers.
- Alcohol Consumption: Increases the risk of several cancers, including liver, breast, oral, and esophageal cancer.
- Dietary Habits: Low intake of fruits and vegetables, along with high consumption of processed and red meats, is associated with increased cancer risk.

- Obesity and Physical Inactivity: Strongly linked to colorectal, breast, pancreatic, endometrial, and kidney cancers. Sunlight/UV Exposure: A well-established cause of skin cancers, particularly melanoma [7].

⇨ Biological Agents

- Viral Infections: Human papillomavirus (HPV) is associated with cervical cancer; hepatitis B and C viruses with liver cancer; Epstein-Barr virus (EBV) with certain lymphomas.
- Bacterial and Parasitic Infections: Helicobacter pylori increase the risk of gastric cancer.

⇨ Physical and Chemical Carcinogens

- Ionizing Radiation: From medical procedures or environmental exposure can induce DNA mutations leading to cancer.
- Chemical Carcinogens: Exposure to asbestos, benzene, formaldehyde, and some pesticides increases cancer risk.

⇨ Genetic and Random Mutations

- Fewer than 10% of cancers are directly due to inherited genetic mutations (e.g., BRCA1/2 mutations in breast cancer).
- Most cancers result from random mutations accumulated over time due to internal or external factors [8].

I.1.6. Treatment of cancer

There are however means to combat cancer disease. These complementary therapies are sometimes used on their own or in conjunction, depending on the type of cancer and its status. The purpose of these therapies is to make possible to remove the tumor and heal a patient with early stage cancer or like a chronic disease in order to monitor its growth. Common and newer forms of medication (surgery, radiation therapy, chemotherapy, targeted therapy, and immunotherapy) are predominantly associated with adverse outcomes which have a detrimental

impact on quality of life. Thus, the battle for more successful, more tolerable anti-cancer therapy continues [9].

Isatin and its derivatives are considered promising compounds in the field of cancer treatment, as several studies have shown that these compounds possess anti-cancer properties. For example, one study demonstrated that isatin exhibited antioxidant activity and cytotoxicity against human leukemia cells, indicating its potential as an anti-cancer agent.

Additionally, various derivatives of isatin, such as hydrazones and thiosemicarbazones, have been developed, showing anti-cancer activity against a wide range of cancer cell lines. These findings suggest that isatin and its derivatives could serve as a foundation for developing new anti-cancer drugs [10].

Moreover, other studies have shown that isatin derivatives can act as inhibitors of specific enzymes linked to cancer progression, such as kinase enzymes. This indicates that isatin's role is not limited to direct effects on cancer cells, but it may also influence biochemical pathways associated with the development of the disease.

Based on this evidence, isatin and its derivatives can be considered promising compounds for the development of novel cancer therapies, whether as standalone agents or components in multi-target drug formulations [11].

I.2. Isatin

I.2. 1. Introduction

Heterocyclic compounds are a type of organic compounds that exhibit a wide range of biological and pharmacological activities. One such biologically active heterocyclic compound is isatin, or 1*H*-indole-2,3-dione, also known as indole quinone or indenedione. It has a nitrogen atom in position 1 and two carbonyl groups in positions 2 and 3. It is consisting of two cyclic rings, one with six members and the other with five. The two rings are flat. The ring with six members is aromatic and the ring with five members has anti-aromatic character [12].

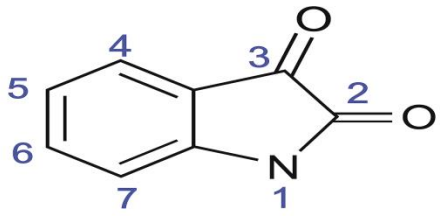


Figure I.8. Structure of isatin (1H-indole-2,3-dione).

I.2. 2. Origin of Isatin

Isatin was first discovered in 1841 by the German chemist Erdmann and the French chemist Laurent, through the oxidation of indigo, a natural blue dye. Therefore, the origin of isatin is closely related to plant-derived indigo, especially from plants like Indigofera tinctoria(fugure1), This Isatin was considered a synthetic moiety for almost 140 years until it was isolated from the plants of Isatis genus [13], Calanthe discolor LINDL [14], the fruit of the cannon ball tree Couroupita guianensis Aubl [15], as a constituent of the secretion from the parotid gland of Bufo frogs , and as metabolic derivatives of adrenalin in humans [16,17].

Moreover, isatin was found to be a component of coal tar while its derivatives fall out in variable dye, Pharmaceuticals, and agriculture chemicals [15] for example, Isatin extracted from the plant Couroupita guianens is exhibits antioxidant activity and cytotoxicity against human leukemia cells, indicating its potential as an anti-cancer compound [18]. Nowadays, isatin has achieved a position, in the design and development of medicinally active analogues because of its efficacy.

Figure I.9. Indigofera tinctoria [19].

I.2. 3. Chemical and Physical Properties of Isatin

Table.I.1. Chemical and Physical Properties of Isatin [20].

Property	Value / Description
IUPAC Name	1H-indole-2,3-dione
Molecular Formula	$C_8H_5NO_2$
Molar Mass	147.13 g/mol
Appearance	Orange to red crystals
Melting Point	197–200 °C
Solubility	Slightly soluble in water; soluble in ethanol, chloroform, and ether
Polarity	Polar compound due to the presence of carbonyl groups (C=O)
Acidic/Basic Properties	Contains a relatively acidic NH proton
Chemical Structure	Contains an oxidized indole ring at positions 2 and 3

I.2. 4. Isatin Synthesis

Isatin can be synthesized through several laboratory methods. The most well-known are based on either oxidation of indigo or derivatization of aryl amines. Below are the main synthetic routes:

⇨ Classical Synthesis from Indigo (Indigo Oxidation)

- This is the earliest and traditional method of isatin synthesis.
- It involves the oxidation of natural indigo dye using strong oxidizing agents such as nitric acid (HNO3).
- The reaction cleaves the double bond in the indigo molecule, forming the two carbonyl groups characteristic of isatin.

⇨Modified Sandmeyer Method (Modern Approach)

- This method starts with o-aminobenzaldehyde and involves reaction with chloral hydrate and hydroxylamine in the presence of hydrochloric acid.
- It is preferred in modern laboratories as it avoids the need for natural indigo.

⇨Synthesis from Aniline Derivatives

- This route involves the use of 2-nitroaniline, which is converted through cyclization and oxidation steps into isatin.
- It is useful for generating substituted isatin derivatives [21].

I.2. 5. Isatin Derivatives and Substitutions

Isatin is a chemically versatile scaffold that can be readily modified to yield a wide range of derivatives with significant biological activities, particularly in anticancer drug discovery. Substitutions on the isatin core are commonly performed at positions C_5, C_6, and C_7 of the benzene ring, or at the C_3 carbonyl group, by introducing functional groups such as halogens, amines, sulfonamides, or heterocyclic moieties. These modifications can enhance pharmacological properties such as solubility, permeability, and binding affinity to cancer cell targets, making isatin derivatives a promising focus in anticancer drug development. The possible substitutions for isatin hybrids are depicted in figure I.10 [22].

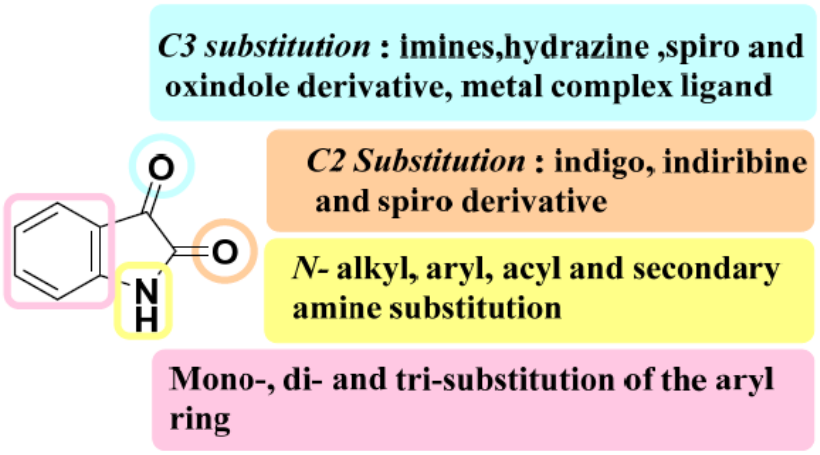


Figure I.10. Probable substitution possible on isatin nucleus [23].

I.2.6. Effect of Substitutions at C_5, C_6, C_7 Positions and N_1 of Isatin

The isatin molecule features an aromatic benzene ring, where substitutions at the C_5, C_4, C_6, and C_7 positions significantly influence its biological activity, particularly its anticancer potential and also the nitrogen atom at the N_1 position is considered a key element in drug design based on the isatin structure, and its modification represents a crucial step in developing more effective and safer compounds.

⇨N_1 position of the isatin ring is a chemically active site that allows for various structural substitutions aimed at enhancing the biological and pharmacological properties of the compound. More complex organic chains, such as peptide derivatives or heterocyclic rings, can also be introduced to increase activity against biological targets such as cancer cells. These modifications directly affect solubility, cellular permeability, and binding affinity to receptors or enzymes, making them central to the drug design strategies based on the isatin scaffold [24]. ⇨C_5 Position Frequently substituted with halogens (Cl, Br, F), nitro (-NO2), hydroxyl (-OH), or amino (-NH3) groups. These modifications enhance the molecule's electronic properties, membrane permeability, and ability to form hydrogen bonds with target enzymes or receptors, thus increasing anticancer activity [25].

⇨ C6 Position often modified with methoxy (-OCH3) or halogen groups. Substitutions here mainly affect the steric configuration of the molecule, influencing its spatial interaction with biological targets and improving solubility and selectivity [25].

⇨C7 Position allows introduction of alkyl groups or fused heterocycles, improving the metabolic stability and potentially enhancing the molecule's binding affinity through hydrophobic or π-π interactions, making it suitable for hybrid drug design [25].

⇨C4 Position although less commonly modified than C_5 to C_7, substitution at C_4 can still influence the isatin molecule's reactivity and bioactivity. Groups such as halogens or electron-with drawing substituents at this position may affect the electronic distribution across the aromatic ring, potentially altering the molecule's interaction with biological targets. Such substitutions can also modulate the acidity of nearby functional groups and may be useful for tuning the molecule's pharmacokinetics.

These structural changes at specific positions are crucial for optimizing the pharmacological profile of isatin derivatives, making them promising scaffolds in anticancer drug development [25].

I.2.7. Pharmacologically active compounds derived from isatin

Isatin derivatives revealed a fascinating array of pharmacological activities, such as anticancer, anti-HIV, antiviral, antitumor, antifungal, antimalarial, Antioxidant, anti-inflammatory, antimicrobial, analgesic, anticonvulsants, and so on. Several Conventional analogues of isatin are currently being used for medicinal purposes, and some représentative examples of marketed drugs containing isatin scaffolds against multiple diseases/disorders are displayed in figure I.11. In the last 15–20 years [23].

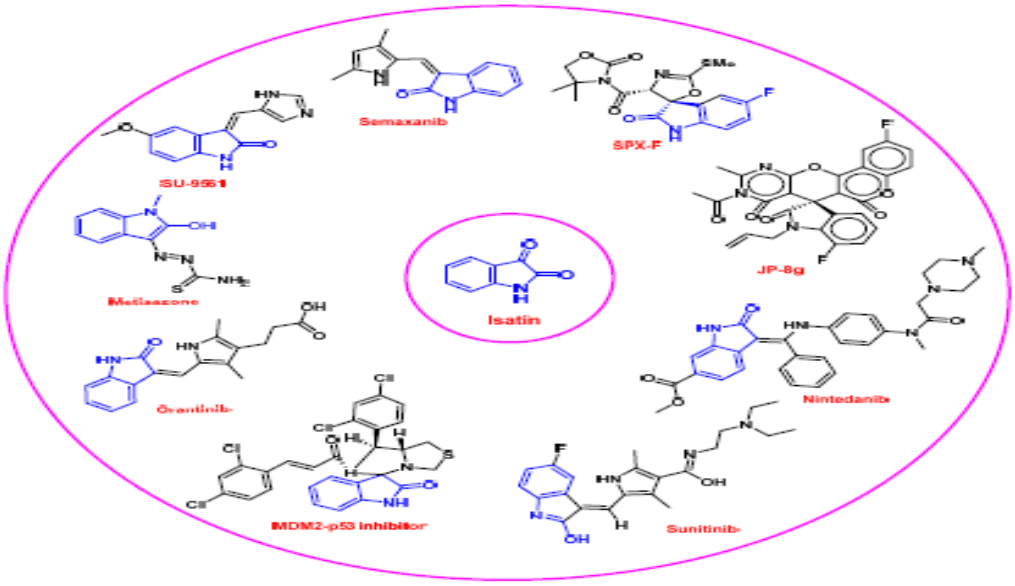


Figure I.11. Examples of marketed drugs containing Isatin.

One of the most well-known anticancer drugs containing isatin is **s**unitinib has received approval for the treatment of several types of cancer, including renal cell carcinoma, gastrointestinal stromal tumors, and pancreatic neuroendocrine tumors. Structurally, Sunitinib features a substituted indolin-2-one moiety, which is a derivative of isatin. This indolinone core is essential for the compound's ability to bind effectively to kinase domains, which leads to the suppression of tumor angiogenesis and cellular proliferation [26].

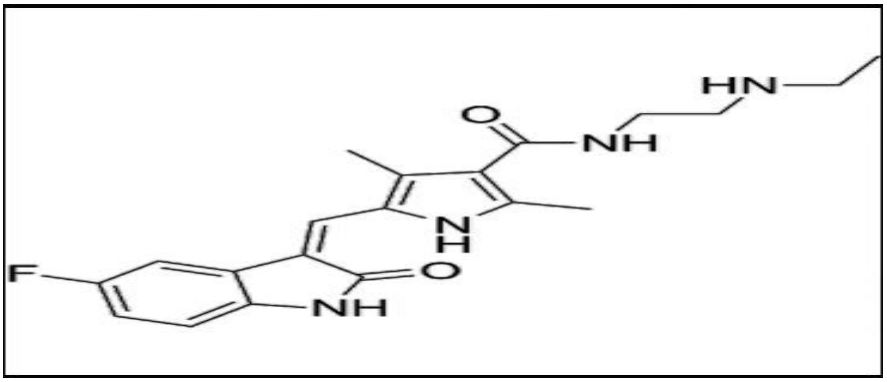


Figure I.12. Chemical Structures of Sunitinib.

I.2.8 Effects of Isatin Derivatives on U937 Cells

Isatin and its derivatives have shown significant potential as anticancer agents, particularly in their effects on U937 cells. They work through several mechanisms:

- **Apoptosis Induction**: Some derivatives trigger cell death by activating enzymes like Caspase-3 and Caspase-9.
- **Cell Cycle Arrest**: These compounds can halt cell division, usually in the G0/G1 or G2/M phases.
- **Oxidative Stress**: By increasing reactive oxygen species, they cause DNA and protein damage that can lead to cell death.
- **Inhibition of Survival Pathways**: Isatin derivatives can suppress key signaling pathways like PI3K/Akt and NF-κB, which are essential for cancer cell survival.

so U937 cells are a powerful tool in cancer and immunology research. Their ability to mimic monocyte and macrophage behavior, respond to various chemical agents, and model rare blood cancers makes them an invaluable resource for studying cellular mechanisms and testing potential cancer therapie **[**30.31].

References

[1]. Almi, I. (2021). *Contribution to drug design through computational studies of several series of bioactive heterocyclic molecules* (Doctoral dissertation). University of Biskra.

[2]. https://cdn.britannica.com/64/91764-050-AD7901A5/tumour-tissues-Cancer-cells- parts-body-travel.jpg

[3]. Cherroun, H. (2019). *Étude quantitative de la relation structure-activité (QSAR) des dérivés pipéraziniques de la phénothiazine* (Master's thesis). Université de Biskra.

[4]. Splane, B. (2023, May 2). *Malignant vs. Benign Tumors: What Are the Differences?* Verywell Health. https://www.verywellhealth.com/malignant-vs-benign-tumors-5208850

[5]. National Cancer Institute. (2020). *Metastasis image.* https://www.cancer.gov/sites/g/files/xnrzdm211/files/styles/cgov_image/public/cgov_image/media_image/2020-09/metastasis.jpg

[6]. Cleveland Clinic. (N.d). *Metastasis – Metastatic Cancer.* https://my.clevelandclinic.org/-/scassets/images/org/health/articles/22213-metastasis-metastatic-cancer.ashx

[7]. Tamang, S. (2024, March 10). *Cancer Cells: Definition, Morphology, Types, Development.* Microbe Notes. https://microbenotes.com/cancer-cells/

[8]. Buffeta, P. (2014). *Overview of human cancer causes.*

[9]. Schwartz, G. (2024). *Epidemiology of cancer.*

[10]. Mun, E. J., Babiker, H. M., Weinberg, U., Kirson, E. D., & Von Hoff, D. D. (2018). Tumor-treating fields: A fourth modality in cancer treatment. *Clinical Cancer Research, 24*(2), 266–275. https://doi.org/10.1158/1078-0432.CCR-17-1117

[11]. Premanathan, M., Radhakrishnan, S., Kulangiappar, K., Singaravelu, G., Thirumalaiarasu, V., Sivakumar, T., & Kathiresan, K. (2012). Antioxidant and anticancer activities of isatin (1H-indole

2,3-dione), isolated from the flowers of *Couroupita guianensis* Aubl. *Indian Journal of Pharmacology, 136*(5), 822–826. https://doi.org/10.4103/0253-7613.103290

[12]. Paiva, R. E. F., Vieira, E. G., Silva, D. R., Wegermann, C. A., & Ferreira, A. M. C. (n.d.). *Anticancer compounds based on isatin-derivatives: Strategies to ameliorate selectivity and efficiency*.

[13]. Borad, M., Bhoi, M., Prajapati, N., Patel, D., & Francis, T. (2013). Review of synthesis of spiro heterocyclic compounds from isatin. *Synthetic Communications, 447*, 897–922.

[14]. Yoshikawa, M., Murakami, T., Kishi, A., Sakurama, T., Matsuda, H., Nomura, M., & Kubo, M. (1998). Novel indole S, O-bisdesmoside, calanthoside, the precursor glycoside of tryptanthrin, indirubin, and isatin from two *Calanthe* species (Orchidaceae). *Chemical and Pharmaceutical Bulletin, 46*, 886–888.

[15]. Bergman, J., Lindström, J.-O., & Tilstam, U. (1985). The structure and properties of some indolic constituents in *Couroupita guianensis* Aubl. *Tetrahedron, 41*, 2879–2881.

[16]. Khan, F., & Maalik, A. (2015). Advances in pharmacology of isatin and its derivatives: A review. *Tropical Journal of Pharmaceutical Research, 14*, 1937.

[17]. Chiyanzu, I., Hansell, E., Gut, J., Rosenthal, P. J., McKerrow, J. H., & Chibale, K. (2003). Synthesis and evaluation of isatins and thiosemicarbazone derivatives against cruzain, falcipain-2 and rhodesain. *Bioorganic & Medicinal Chemistry Letters, 13*, 3527–3530.

[18]. Silva, F. C., et al. (2011). Isatin: A review of its pharmacological properties and synthesis. *Mini-Reviews in Medicinal Chemistry, 11*(6), 451–460.

[19]. Stock. (N.d.). *Isatin-related board.* https://www.istockphoto.com/collaboration/boards/ubIuvPDqr0Sx5wHVHowgyQ

[20]. https://doi.org/10.2174/138955711795445876

[21]. Kumar, D., Patel, G., & Patel, N. (2013). Isatin and its derivatives: A survey of recent advances in medicinal chemistry. *Medicinal Chemistry Research, 22*(3), 1103–1115. https://doi.org/10.1007/s00044-012-0140-1

[22]. Cheke, R. S., Patil, V. M., Firke, S. D., Ambhore, J. P., Ansari, I. A., Patel, H. M., Shinde, S. D., Pasupuleti, V. R., Hassan, M. I., Adnan, M., Kadri, A., & Snoussi, M. (n.d.). *Therapeutic outcomes of isatin and its derivatives against multiple diseases: Recent developments in drug discovery*.

[23]. Verma, M., Gupta, S. J., & Singh, A. K. (2019). Isatin and its derivatives: A survey of recent therapeutic advancements. *European Journal of Medicinal Chemistry, 180*, 562–612. https://doi.org/10.1016/j.ejmech.2019.07.003

[24]. Mendel, D. B., et al. (2003). SU11248, a novel tyrosine kinase inhibitor targeting vascular endothelial growth factor and platelet-derived growth factor receptors: Determination of a pharmacokinetic/pharmacodynamic relationship. *Clinical Cancer Research, 9*(1), 327–337.

[25]. Kumar, D., Rane, R. A., & Shukla, P. K. (2013). Isatin and its derivatives: A survey of recent advances in biological activities. *Medicinal Chemistry Research, 22*(3), 1405–1415. https://doi.org/10.1007/s00044-012-0120-4

[26]. MedChemExpress. (N.d.). *N-Desethyl-Sunitinib*. https://www.medchemexpress.com/N-Desethyl-Sunitinib.html

[27]. Sundström, C., & Nilsson, K. (1976). *Establishment and characterization of a human histiocytic lymphoma cell line (U-937). International Journal of Cancer, 17*(5), 565–577. https://doi.org/10.1002/ijc.2910170507

[28]. ScienceDirect Topics. (N.d.). *U937 Cell – An Overview.*

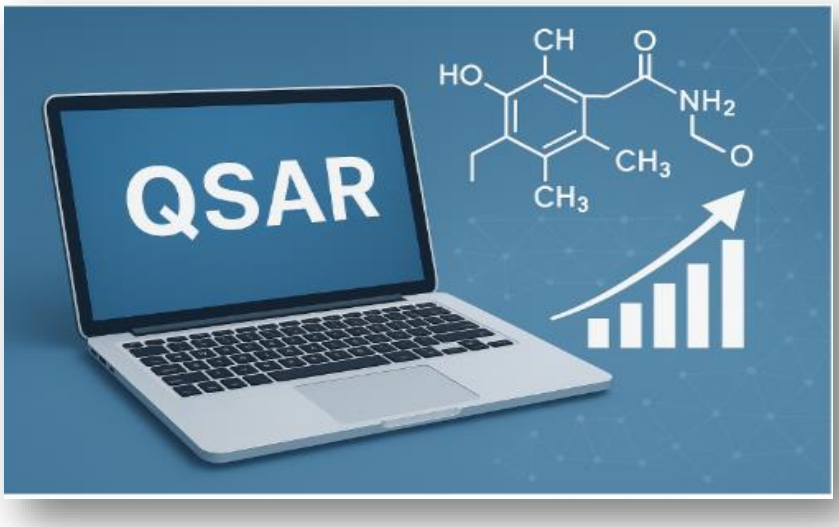
[29]. ATCC. (n.d.). *U-937 (CRL-1593.2) Cell Line*. American Type Culture Collection. Retrieved from https://www.atcc.org/products/crl-1593.2

[30]. ResearchGate. (n.d.). *Images and data-t-on U937 morphology and drug response*. Retrieved from https://www.researchgate.net

[31]. Kumar, D., Sharma, P., Singh, H., Nepali, K., & Gupta, G. K. (2013). Design, synthesis and anticancer evaluation of novel isatin derivatives. *European Journal of Medicinal Chemistry, 67*, 434–443. https://doi.org/10.1016/j.ejmech.2013.06.022

Chapter II

Background on theorical bases



II.1. Molecular modeling

II.1.1. Introduction

Molecular modeling is a scientific discipline that employs computational techniques to represent and simulate the behavior of molecules. The primary goal of molecular modeling is to study the structural, physical, and chemical properties of molecules, as well as their interactions with each other. It relie on the application of classical mechanics or quantum mechanics principles to molecular systems, providing a deeper understanding of the behavior of compounds at atomic and molecular levels.

Molecular modeling encompasses a range of methods, including molecular mechanics, molecular dynamics, and quantum chemistry, which are used to study and analyze molecular systems with varying levels of complexity and precision. These methods complement experimental studies by offering detailed insights into molecular geometry and dynamics that are often difficult to obtain experimentally [1,2].

Molecular modeling has proven to be an effective tool in predicting the biological activity of compounds prior to synthesis, which significantly reduces the time and cost of drug development. These predictions are often based on quantum chemical calculations or statistical models such as quantitative structure-activity Relationship (QSAR), which link molecular descriptors to biological responses. Studies have demonstrated that molecular modeling can accurately forecast ligand-receptor binding and pharmacological activity, making it a cornerstone in modern drug design [3].

II.2. The basics of quantum chemistry

The foundations of quantum chemistry appeared and developed significantly in the 1920s thanks to scientists such as Bohr, Schrödinger, Born, Oppenheimer, Hartree, and also Slater. In 1930, Hartree and Fock developed the self-consistent field method, which allowed for the first calculations for diatomic systems. However, it was not until the early 1950s that calculations began for systems with more atoms.

In 1964, Hohenberg and Kohn defined a theorem, the foundation of Density Functional Theory (DFT). In 1970, Pople created Gaussian, which is still the most widely used modeling software. Semi-empirical methods and the use of DFT methods became increasingly developed between the 1970s and 1980s, with a significant rise starting in the 1990s. With the advancement of computing, modeling entered our computers. In 1993, the B3LYP method appeared a hybrid method that allows DFT calculations. In 1998, the Nobel Prize in Chemistry was awarded to John A. Pople and Walter Kohn for their work in the field of computational chemistry and molecular modeling (quantum chemistry) [4,5]. Quantum chemistry involves using methods based on solving the time-independent Schrödinger equation (stationary state). By solving the equation for eigenvalues and eigenvectors, $H\Psi = E\Psi$, where H is the non-relativistic Hamiltonian, E is the total energy, and Ψ is the wave function of the system, it becomes possible to determine all the system's quantum information. It is thus possible to rigorously solve such an equation using quantum chemistry theory proposed since the 1920s in order to reproduce this equation approximately [6,7] although the study of structure-activity relationship began at the end of the 19th century, it was not until the 1960s that Corwin Hansch proposed the first mathematical model to correlate biological activity with chemical structure. Over the following decades, this domain became well established, and the bibliographic resources available for this approach are now extensive [8].

The QSAR method includes all statistical methods by which biological activities (most often expressed through the logarithms of molar or equipotent activities) are linked with structural elements (Free Wilson analysis), physicochemical properties (Hansch analysis), or different parameters related to the notion of a field aiding in structural description (3D QSAR) [9].

II.3. Calculation Methods

II.3.1. Semi-Empirical Methods

Semi-empirical methods are entirely based on experience. They are derived from ab initio methods and follow the same principles; except they do not consider all electrons in the valence shell of each atom and neglect some integral calculations by using empirical

parameters. These approximations are often compensated by fitting to experimentally observed properties. Semi-empirical methods are generally used to study large molecular groups and obtain useful molecular properties (structure, reactivity, etc.). Several variants exist: **CNDO, INDO, MINDO/3, MNDO, AM1, PM3, SAM1 [**10].

II.3.2. Quantum Methods

Pure quantum methods are based solely on quantum mechanics. Quantum mechanics is a rigorous mathematical technique based on the Schrödinger equation. Solving this equation provides precise information about the geometric and electronic properties of the molecule [11] these methods are techniques for solving the Schrodinger equation of multi-electron system. They use data adjusted to experimental results to simplify calculation. semi-empirical methods contain a minimal basis by default (STO-3G) whereas pure quantum methods use different basis sets and correlation, depending on the type of calculation [12].

II.3.2.1. Density Functional Theory (DFT)

The foundations of Density Functional Theory (DFT) were established in 1927 by Thomas and Fermi, who calculated the energy of an atom by expressing its kinetic energy as a function of the electron density [13]. In 1928, Dirac introduced the exchange term predicted by Hartree, but there was still no consideration of electron correlation, which was eventually added by Wigner.

In this model, the *n* **e**lectrons, which normally depend on 3*n* spatial coordinates, are replaced by their density ρ(r), which depends on only 3 variables. The ground state of this system is described by the wave function Ψ_0 (r_1, r_2, ..., r_n), which corresponds to a unique electron density ρ(r)**.** This wave function, and the associated energy E_0, are determined by minimizing the total energy of the system. The external potential $v_{ext}(r)$ created by the *N* nuclei of the system is then completely determined and thus defines the Hamiltonian. As such, the number of electrons *n* and the potential $v_{ext}(r)$ defined all the properties of the ground state [14].

II.3.2.2. Ab-initio

Ab initio methods are characterized by the introduction of an arbitrary basis set to expand the molecular orbitals, and then the explicit calculation of all the required integrals involving this basis. Ab initio methods are divided into two subfamilies: the Hartree-Fock methods (HF, RHF, UHF, ROHF) (Hartree,1928, Fock, 1930), and the post-Hartree-Fock methods (MN, CAS...) Møller, 1934.The main difference between these two approaches is that electron-electron interactions are neglected in HF methods and reintroduced in post-HF methods. These methods can only be applied to systems with a few dozen atoms in the case of HF methods, and to systems with only about ten atoms in the case of post-HF methods [15].

II.3.3 Molecular mechanics

The term '**Molecular Mechanics**' refers to a computational method that makes it possible to obtain molecular geometry and energy results based on classical mechanics. This method first appeared in 1930, but developed further from the 1960s onwards, when computers became more accessible and more powerful [16].

MM is often applied to large systems to calculate molecular structures and relative potential energies of molecular conformations or atomic arrangements .Electrons in the studied system were not explicitly considered, but each atom - specifically the nucleus and associated electrons- was considered an asingle particle. The exclusion of electrons in MM is justified by the Born-Oppenheimer approximation, which states that electron and nuclear motions can be separated and considered separately. The energy difference between conformations is important in such calculations, not the absolute value of the potential energy.

MM can be simply viewed as a ball and spring model with classical forces between atoms and molecules. These forces are explained by potential energy functions incorporating structural features such as length, bond angle, and torsion angle.

The potential energy function is equipped with parameters designed to reproduce the experimental properties. The MM or rather the total potential energy of a molecule is

described as the sum of the bond stretch energy (Estr), the bond angle bending energy (Ebend), the torsion energy (Etor) and energy of interactions between non-bonded atoms (Enb). The energetic contributions of the latter constitute the van der Waals (Evdw) and electrostatic (Eelec) interactions

$$\mathbf{E}_{\mathbf{tot}} = \mathbf{E}_{\mathbf{str}} + \mathbf{E}_{\mathbf{bend}} + \mathbf{E}_{\mathbf{tor}} + \mathbf{E}_{\mathbf{vdw}} + \mathbf{E}_{\mathbf{elec}}$$

$$\mathbf{E}_{\mathbf{tot}} = \sum_{bonds} \mathbf{K}_{\mathbf{r}} (\mathbf{r} - \mathbf{r}_{eq})^2 + \sum_{angles} \mathbf{K}_{\mathbf{r}} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\mathbf{eq}})^2$$

$$+ \sum_{dihedrals} \frac{\mathbf{vn}}{\mathbf{2}} [\mathbf{1} - \cos(\mathbf{n}\emptyset - \mathbf{r})] + \sum_{\mathbf{i}<j} \left[\frac{\mathbf{Aij}}{\mathbf{rij}^{12}} + \frac{\mathbf{Bij}}{\mathbf{rij}^{6}} + \frac{\mathbf{qiqj}}{\sum \mathbf{rij}} \right]$$

- **Force fields in molecular mechanics**

We call the mathematical model that represents the potential energy of a molecule in molecular mechanics a force field. It is important to note that force fields are a purely empirical approach. Designates both the mathematical equation (potential energy function) 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 initially designed for simple molecules (alkanes, alkenes, non-conjugated alkynes, amines, etc.), but its improved versions MM3 (1989) and MM4 (1996) allow it to process increasingly complex organic

⇨ CHARM (Bio+)

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

⇨AMBER

AMBER (Assisted Model Building with Energy Refinement), was written by Kollman the field is established for proteins and nucleic acids (UCSF, 1994). It has been used for polymers and for other small molecules [17].

⇨MM+

Is an extension of the MM2 force field, with the addition of some extra parameters. MM+ is a robust force field; it has the ability to takes into account parameters that are neglected in other force fields and can therefore be applied to more complex molecules such as inorganic compounds.We will use this field in our work [16].

II.4. Quantitative structure activity relationship (QSAR)

II.4.1. QSAR History

QSAR dates back to the 19th century. In 1863, A.F.A. Cros at the University of Strasbourg observed that toxicity of alcohols to mammals increased as the water solubility of the alcohols decreased [18]. In the 1890's, Hans Horst Meyer of the University of Marburg and Charles Ernest Overton of the University of Zurich, working independently, noted that the toxicity of organic compounds was dependent on the lipophilicity [18, 19].

Little additional development of QSAR occurred until the work of Louis Hammett [20], who correlated electronic properties of organic acids and bases with their equilibrium constants and reactivity. Hammett observed that adding substituents to the aromatic ring of benzoic acid had an orderly and quantitative effect on the dissociation constant. Hammett also observed that substituents have a similar effect on the dissociation of other organic acids and bases. QSARs based on Hammett's relationship utilize electronic properties as descriptors. Difficulties were encountered when investigators attempted to apply Hammett type relationships to biological systems, indicating that other structural descriptors were necessary.

Robert Muir, a botanist at Pomona College, was studying the biological activity of compounds that resembled indole acetic acid and phenoxyacetic acid [21], which function as plant growth regulators. In an attempt to correlate the structures of the compounds with their activities, he consulted Corwin Hansch. Using Hammett sigma parameters to account for the electronic effect of substituents did not lead to a meaningful QSAR. However, Hansch recognized the importance of lipophilicity, expressed as the octanol water partition coefficient, on biological activity [22].

This parameter is recognized to provide a measure of membrane permeability, since a compound needs to have lipophilic properties to enter a membrane and hydrophilic properties to pass through. The octanol-water partition coefficient is also a driving force when drugs bind into targets.

QSAR models are now developed using a variety of parameters such as descriptors of the structural properties of molecules, descriptors to account for the shape, size, lipophilicity, polarizability, and other properties [23].

II.4.2. Object of QSAR

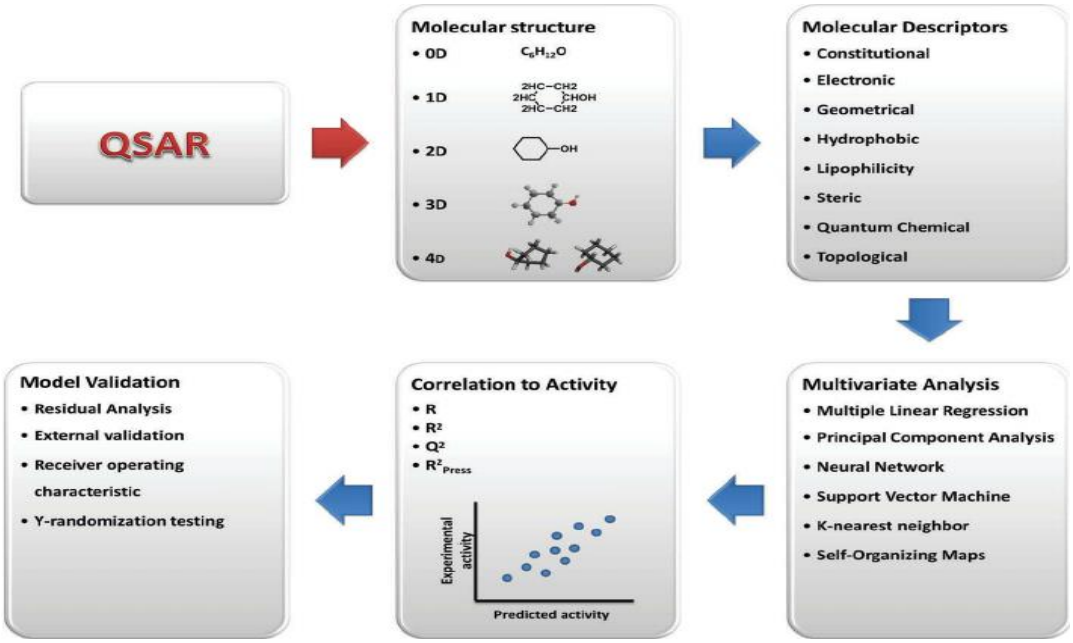
The main objective of the QSAR study is the rational creation of a mathematical model, followed by an examination of the involved chemical information, in order to gain insight into the mechanism and behavior of the system to be studied.

It is also useful in identifying alternative modes of action, in selecting useful structural features, in preparing new design methodologies, in developing new drugs and in helping to formulate new hypotheses for future research studies. As a result, QSAR reduces costs, time and human capital to make the pharmaceutical product available to patients. QSAR models are also used in anticipation of pharmacokinetic and pharmacodynamics properties. QSAR also predicts properties, such as: permeability, and solubility [24].

II.4.3. Main Steps in QSAR Analysis

To develop QSARs, a series of compounds, called a training set, is used. The compounds in the training set ideally; the same or similar mechanism of biological action to ensure that the

same factors influence the activity of all compounds under investigation. For all compounds in the series, biological activities are evaluated and compound structural descriptors are calculated. Statistical tools are then used to derive QSARs [25]. Figure II.1 below shows the the main steps in QSAR analysis.



FigureII.1. The main steps in QSAR analysis.

II.4.4. Tools and Techniques of QSAR

II.4.4.1. Biological Descriptor

QSAR models depend on the experimental data used for their construction. The modeler must take into account the data to be modeled. Therefore, the selection of the database is a very important step in the development of QSAR models. Ideally, this data should be of high quality, meaning it should be reliable and consistent. It is thus important to choose data with low uncertainty in order to limit experimental error bars. Furthermore, the modeler must

ensure that the experimental data used have been obtained under the same protocol. Indeed,

experimental conditions generally have a strong influence on the values obtained. It is also necessary for the distribution of the data to be as homogeneous and normal as possible, since most statistical methods are based on this type of distribution.

The effectiveness of a QSAR model also depends on the type of molecules included in it. The more the model includes compounds with closely related and similar structures, the more likely it is to be effective.

Biological data are generally expressed in inverse logarithmic values (log1/C) in order to obtain higher numerical values when the structures are biologically very effective [22]. A few common endpoints are outlined in Table II.1.

Table II.1. Types of biological data utilized in QSAR analysis [26].

Source of Activity	Biological Parameters	Logarithmic Form
Isolated Receptors	Reaction rate constant	Log K
	Michaelis- Menten constant	$\text{Log}\frac{1}{Km}$
	Inhibition constant	$\text{Log}\frac{1}{Ki}$
Cellular Systems	Inhibition constant	$\text{Log}\frac{1}{IC50}$
	Cross-resistance	Log CR
	In vitro biological data	$\text{Log}\frac{1}{c}$
	Gene mutation	Log TA98
In Vivo Systems	Bioconcentration facteur	Log BCF
	In vivo reaction rates	Log I (induction)
	Pharmacodynamics rates	Log T (total clearance)

II.4.4.2. Molecular Descriptors

A molecular descriptor is a parameter (a numerical value) specific to a given chemical structure. These values can be obtained experimentally or calculated from the molecular structure. Calculated descriptors allow for predictions to be made without having to synthesize the molecules, which is one of the goals of molecular modeling. Molecular descriptors play a fundamental role in quantitative structure–activity/property relationship (QSAR) studies. They are used as independent variables to predict a dependent variable (such as biological activity or a physicochemical property).

The use of molecular descriptors in the development of QSAR models is not an easy task. First, a very large number of molecular descriptors, with varying complexities and theoretical foundations have been introduced over recent years. Moreover, during this time, no strictrules have been established for selecting appropriate descriptors from the vast number available. This choice has often relied on the chemist's intuition or on established traditions [27]. Many software tools have been developed to calculate different molecular descriptors, such as: Gaussian**, ChemOffice, ChemSketch, Marvin Sketch, Dragon.**
In the following, we will present only the molecular descriptors that have been used in the entirety of our work.

1-D Descriptors

These descriptors are calculated from the molecular formula using the molecular composition, that's to say the atoms that make it up, and represent general properties such as: atomic mass percentages, molar mass, molecular weight... In our work we used:

⇨Molecular Weight (MW)

Molecular weight descriptor has been used as a descriptor in systems such as transport studies where diffusion is the mode of operation. It is an important variable in QSAR studies pertaining to cross resistance of various drugs in multi-drug resistant cell lines. Molecular weight is correlated with the size of the molecule. High molecular weight compounds are

likely to show high toxicity as promiscuity of compounds is also likely to increase Additionally, the systemic clearance of a compound is inversely proportional to the molecular weight [22].

Geometrical descriptors

⇨**Molecular Volume (MV)**

This is the volume occupied by a substance, specific at standard temperature and pressure. Its calculation is very similar to that of the surface.

The volume is defined by the following equation [28].

$$\mathbf{MV} = \frac{W}{d}$$

W: The molecular mass

d: The density.

Physicochemical descriptors

⇨**Surface Area Grid (SAG)**

Refers to a computational method used to estimate the surface area grid by mapping the three-dimensional space around a molecule onto a grid. Each point on this grid is evaluated for interaction or exposure, helping in the analysis of molecular shape, volume, and reactivity.

This approach is especially useful in drug design and molecular docking, where understanding the exposed surface of a molecule can aid in predicting solubility, permeability, and binding efficiency [29,30].

⇨Molar Refractivity (MR)

Is a measure of a molecules ability to scatter light. It reflects both the size of the molecule and the flexibility of its electron cloud. A higher MR value indicates a larger or more polarizable molecule. It is commonly used in QSAR studies to help explain how molecular properties influence biological activity [31].

$$\mathbf{MR} = \frac{Mw}{\mathbf{P}} \times \frac{\mathbf{n^2-1}}{\mathbf{n^2+2}}$$

MR: Molar Refractivity

n: Refractive index

Mw: Molecular weight

P : Density

This is known as the Lorentz–Lorenz equation, and it helps estimate molecular polarizability using measurable physical properties [32].

⇨Polarizability (Pol)

Polarizability, denoted as αe\alpha_eαe and expressed in cubic meters (m^3), represents a molecule's ability to have its electron cloud distorted under the influence of a uniform electric field. It is one of the parameters that reflects molecular properties related to hydrophobicity and, consequently, to biological activities [33]. This quantity can be calculated from molar refractivity or molar volume using the following equations:

$$\mathbf{Pol} = \mathbf{0.3964308} \times \mathbf{MR} = \mathbf{0.3964308} \times \frac{\boldsymbol{n^2-1}}{\boldsymbol{n^2+2}} \boldsymbol{MV}$$

These descriptors characterize the charge distribution of molecules (molecular polarity), as well as quantum chemical parameters which, in order to be reliably calculated, require more sophisticated computational methods.

⇨Partition Coefficient (Log P)

Lipophilicity is a property that significantly influences the solubility, absorption, distribution, metabolism, and excretion of drugs. Hansch and Leo estimated that molecules with high lipophilicity are likely to be retained within membrane lipids .The best method to estimate a compound's ability to dissolve in both the aqueous environment of the cytoplasm and the non-polar environment of the cell membrane is by measuring lipophilicity (Figure II.2) The partition coefficient P is calculated as follows [34].

$$\text{Log P} = [\text{drug concentration}]_{\text{octanol}} / [\text{drug concentration}]_{\text{water}}$$

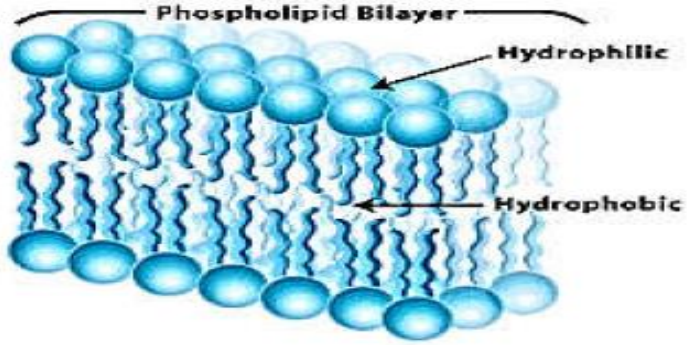


Figure II.2. Polarity of Different Cellular Environments.

Quantum/electronic descriptors

⇨Total Energy (Et)

For an isolated molecule in its ground state, the calculated total energy (Et), measured in electronvolts (eV), can be used as a quantum molecular descriptor. This approximate energy was calculated for an optimized conformation of the most stable geometry, the structure with minimal energy [34].

⇨Dipole Moment (μ)

Measured in Debye (D), it indicates the net molecular polarity and describes the separation of charges in a molecule where the electron density is unequally shared between atoms. The existence of a dipole moment in a molecule originates from the difference in electronegativity between atoms. The electron density is higher near the more electronegative atom, leading to an asymmetry in the distribution of bonding electrons. Thus, the higher the dipole moment of a molecule, the greater the asymmetry.

⇨Frontier Orbital Energies

These play a major role in many chemical reactions and reaction mechanisms. The energies of these orbitals are very popular parameters in quantum chemistry and QSAR studies:

- **HOMO Energy (E_{HOMO})**

Measured in eV, it is the highest occupied molecular orbital. It is directly related to the ionization potential. When a molecule acts as a Lewis base (electron pair donor) during bond formation, electrons are donated from this orbital. It measures the nucleophilicity of a molecule and its susceptibility to electrophilic attack.

- **LUMO Energy (E_{LUMO})**

Measured in eV, it is the lowest unoccupied molecular orbital and is directly related to electron affinity. When a molecule acts as a Lewis acid (electron pair acceptor) during bond formation, incoming electron pairs are accepted into this orbital. It measures the electrophilicity of a molecule and its susceptibility to nucleophilic attack [26].

⇨Energy Gap (Eg)

The HOMO-LUMO gap, measured in (eV), represents the energy difference between the highest occupied molecular orbital and the lowest unoccupied one. It is an important indicator

of molecular stability. This energy difference serves as a measure of a molecule's excitability. The smaller the gap, the more reactive the molecule is with its environment.

A large HOMO-LUMO gap implies high molecular stability in the sense of low reactivity in chemical reactions. Conversely, a small gap indicates high molecular reactivity. The HOMO-LUMO gap has also been used as an approximation of the molecule's lowest excitation energy [34].

⇨Electronegativity (χ)

Denoted as **χ** and measured in electronvolts (eV), is defined as the negative of the chemical potential. It reflects the tendency of the electron cloud to escape from the molecule. It is a global parameter of the molecular system and corresponds to the slope of the total energy E with respect to the number of electrons N, at constant external potential v(r), as defined by Parr and Mulliken [35,36].

$$\chi = -\mu = \left(\frac{\partial E}{\partial N}\right)_{v(r)} = -\frac{(E_{LUMO} + E_{HOMO})}{2}$$

⇨Chemical hardness (η)

denoted as **η**, and its inverse, softness, denoted as **S**, can be derived from the first derivative of the chemical potential [37,38]:

$$\eta = \left(\frac{\partial \mu}{\partial N}\right)_{v(r)} = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(r)} = \frac{1}{S} = \frac{E_{LUMO} - E_{HOMO}}{2}$$

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{v(r)} = -\frac{PI+AE}{2} = -\chi$$

I: is the ionization potential

A: is the electron affinity.

The qualitative definition of hardness is closely related to polarizability, as a decrease in the energy gap generally facilitates the polarization of the molecule. This descriptor allows for differentiation between reaction rates at various molecular sites [39, 40].

⇨Electrophilicity index (ω)

denoted as **ω**, is used to characterize a molecules ability to undergo electron transfer. It is calculated using the following formula [41]:

$$\omega = \frac{\chi^2}{2\eta}$$

II.4.5. Statistical Parameters

Statistical methods are an essential component of QSAR work. They help to build models, estimate a model's predictive abilities, and find relationships and correlations among variables and activities. A suitable statistical method coupled with a variable selection method allows analysis of this data in order to establish a QSAR model with the subset of descriptors that are most statistically significant in determining the biological activity. The statistical method can be broadly divided in to two: linear and non-linear method. In statistics a correlation is established between dependent variables (biological activity) and independent variables (physiochemical properties or molecular descriptor). The liner method fits a line between the selected descriptor and activity as compared to non-linear method. Which fit a curved between the selected descriptor and activity. The statistical method to build QSAR model is decided based on the type of biological activity data.

Following is commonly used statistical methods: Principal component analysis (PCA), Cluster analysis (CA), Simple liner regression (SLR), multiple liner regression (MLR). Stepwise multiple liner regression (MLR), Principle component regression (PCR) Continuum Regression (CR), Partial least squares (PLS), Genetic function approximation (GFA), Genetic partial least squares (GPLS), Logistic regression (LR), K-Nearest Neighbors classification (KNN), Neural Network (NN), Discriminant analysis (DA), Decision Trees (DT), Canonical Correlation (CC) [42].

- **Multiple Linear Regression (MLR)**

Can be considered as an easy interpretable regression-based method, regression analysis correlates independent X variables or descriptors with dependent Y variables (biological data). The regression model assumes a linear relationship between *m* molecular descriptors and the response (biological activity) variable. This relationship can be expressed with the single multiple-term linear equation:

$$\mathbf{Y}=b_0+b_1X_1+b_2X_2+\ldots\ldots\ldots+b_mX_m+\mathbf{e}$$

MLR analysis computes the regression coefficients bi by minimizing the residuals, which quantify the deviation between the data (Y) and the model (Y'), as in simple linear regression [43].

⇨ Description of the Method

Multiple linear regression is the simplest and most widely used method for developing predictive models. It is based on the assumption that there is a linear relationship between a dependent variable Y (in this case, the property) and a series of p independent variables Xi (in this case, the descriptors). The goal is to obtain an equation of the following form**:**

$$\mathbf{Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \ldots + \beta_pX_p + \varepsilon}$$

$\boldsymbol{\beta_i}$ are the regression coefficients.

The determination of equation is done based on a dataset of **n** samples for which both the dependent and independent variables are known. This means considering a system of **n** equations:

$$\mathbf{Y_1 = \beta_0 + \beta_1 X_{1,\,1} + \beta_2 X_{2,\,1} + \cdots + \beta_P X_{n,\,1} + \varepsilon_1}$$

$$\mathbf{Y_2 = \beta_0 + \beta_1 X_{1,\,2} + \beta_2 X_{2,\,2} + \cdots + \beta_P X_{n,\,2} + \varepsilon_2}$$

$$\mathbf{\ldots\ \ldots\ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots}$$

$$\mathbf{Y_n = \beta_0 + \beta_1 X_{1,P} + \beta_2 X_{2,P} + \cdots + \beta_P X_{n,p} + \varepsilon_n}$$

This system of equations can be written in the following matrix form:

$$\begin{Bmatrix} Y_1 \\ \vdots \\ Y_n \end{Bmatrix} = \begin{Bmatrix} 1 & X_{1.1} \cdots & X_{2,2} & \ldots. & X_{1,p} \\ . & . & & & . \\ . & . & \vdots & \ddots & . \\ . & . & & & . \\ 1 & X_{n,1} & X_{n,2} & \ldots. & X_{n,p} \end{Bmatrix} \begin{Bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{Bmatrix} + \begin{Bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{Bmatrix}$$

The matrix **X (n, p)** has, in its first column, a vector consisting entirely of 1s. This vector corresponds to the constant term $\mathbf{X_0}$. Therefore, the matrix **X** is of dimension **(n, p+1)**. The model is then written in the form:

$$\mathbf{Y = X_b + \varepsilon}$$

The method consists of choosing the coefficients of the vector **β** in such a way as to minimize the sum of the squares of the differences between the predicted values and the actual values across the entire dataset, under certain initial assumptions. First, the independent variables $\mathbf{X_i}$, as the name suggests, are assumed to be independent from each other, and their uncertainty is considered negligible. Second, the different samples $\mathbf{Y_i}$ are assumed to be independent from one another.

Finally, by nature, the dependence of **Y** on the $\mathbf{X_i}$ is assumed to be linear [26]. The predicted value of the dependent variable is then:

$$\widehat{Y} = 0 + \widehat{\beta}_1 X_{1,i} + \cdots + \beta_p X_{n,i}$$

The residuals can then be defined as the difference between the observed and predicted values of **Y**:

$$\varepsilon = Y_i - \widehat{Y}_i$$

The goal is to find the coefficients $\hat{\beta}_i$ in order to minimize the sum of the squared residuals (RSS) for the entire dataset [28].

II .4.6. Chemometrics Tools

- **Correlation coefficient (R)**

It is the most widespread statistical indicator is the correlation coefficient, which evaluates the part of the variance of the activity / the target property explained by the model.

$$\mathbf{R} = \sqrt{1 - \frac{\sum(yi - \hat{y}_i)^2}{\sum(yi - \bar{y}_i)^2}}$$

R: is the correlation coefficient.

y_i, $\hat{y}_i$: are, respectively, the observed and calculated values of the dependent variable.

$\bar{y}_i$: is the mean value of observed values.

These coefficients are not affected by the unit of measurement chosen and reflect a good correlation between the target activity and the initial activity if R^2 is close to 1 (ideal case) [44].

- **Coefficient of Determination (R^2)**

The coefficient of determination is found by squaring the correlation coefficient and is used as a more precise way to interpret the correlation coefficient. It is useful because it gives the proportion of the variance in one variable that is "explained" by the other variable. It represents the percent of the data that is the closest to the line of best fit [45].

The correlation coefficient can be determined by the mathematical formula:

$$\mathbf{R^2 = \frac{\sum_{i=1}^{n}(y_{i,obs} - y_{i,cal})^2}{\sum_{i=1}^{n}(y_{i,obs} - \bar{y})^2}}$$

The coefficient of determination is such that $\mathbf{0 < R^2 < 1}$, and the stronger the correlation (R is closer to 1).

- **Adjusted coefficient of determination (R^2_{adj})**

$$\mathbf{R^2_{adj} = \frac{R^2(n-1)-p}{n-p-1}}$$

Where: **n** is the number of dependent variables (the molecules); **p** is the number of independent variables (descriptors); $\mathbf{R^2}$ is the coefficient of determination [44].

- **Standard Deviation (S)**

Standard deviation (S) is a statistical measure of the spread or uncertainty around the mean. It is defined by the equation:

$$\mathbf{S = \sqrt{\frac{\sum(yi - \hat{y}_i)^2}{n-p-1}}}$$

In particular, the smaller the standard deviation, the better the correlation [46].

- **Fischer Statistic (F)**

The Fisher F index is also used to measure the level of statistical significance of the model, that is to say the quality of the choice of descriptors constituting the model

$$\mathbf{F}=\frac{\sum(\hat{\mathbf{y}}_\mathbf{i}-\bar{\mathbf{y}}_\mathbf{i})^\mathbf{2}}{\sum(\mathbf{y}_\mathbf{i}-\hat{\mathbf{y}}_\mathbf{i})^\mathbf{2}}\ \frac{(\mathbf{n}-\mathbf{p}-\mathbf{1})}{\mathbf{p}}$$

$\boldsymbol{\gamma_i}$, $\boldsymbol{\hat{\gamma}_i}$ are, respectively, the observed and calculated values of the dependent variable.

$\boldsymbol{\bar{\gamma}_i}$ is the mean value of predicted values.

n is the number of dependent variables (the molecules).

p is the number of independent variables (descriptors).

- **Quality Factor (Q)**

Quality factor is calculated by equation:

$$\mathbf{Q}=\frac{\boldsymbol{R}}{\boldsymbol{S}}$$

Where R is variance and S is standard deviation. Over fitting and chance correlation, due to excess number of descriptors, can be detected by Q value. Positive value for this QSAR model suggests its high predictive power and lack of over fitting [46].

II.7. Validation of QSAR Models

The predictive powers of the equations were validated by leave–one–out (LOO) cross–validation method [48–50], cross-validation is a practical and reliable method for testing the significance of a model. Hence, to validate the final models generated individually for different activities/properties, leave one-out method is used to do crossvalidation. The leave-one-out method consists of developing a number of models with one compound omitted at the time after developing each model. The omitted sample data are predicted and the difference between observed and predicted values (activities) is calculated.

The predictive ability of the model is quantified in terms of the corresponding leave-one-out cross-validated parameters. The cross-validated parameters often used being: PRESS, SSY, Spress, R^2_{cv} ,R^2_{adj}, PE. These parameters are defined as below

Table II.2. Statistical parameters for cross-validation [51].

Statistic	Definition	Formula
PRESS	Predicted residual sun of squares	**PRESS=**$\sum(Y_{obs} - Y_{cal})^2$
TSS	Total Sun of squares	**Tss=**$\sum(Y_{obc-} \bar{Y})^2$
R^2_{adj}	The square of the correlation adjusted	R^2_{adj}**=1-(1-**R^2**)** $(\frac{n-1}{n-p-1})$
R^2_{cv}	The square of the correlation coefficient	R^2_{cv} **= 1-** $\frac{PRESS}{TSS}$
S_{PRESS}	Standard prediction error validation	**Spress=** $\sqrt{\frac{PRESS}{n}}$
PE	Prediction error	**PE=** $\frac{0.6745 \ (1-R^2)}{\sqrt{n}}$

- **PRESS:** Predicted residual sum of squares is the difference between an observed value and the value predicted by the model.
- **SSY:** is the sum of the squares of the distances of the observed values for a variable compared to the average of this variable, the sum of squares allows to measure the total change in variable.
- **SPRESS:** The predictive ability of the models is evaluated by the root mean square error.
- $\mathbf{R^2_{cv}}$: The change in the R^2 statistic that is produced by adding or deleting an independent variable, if the R^2 change associated with a variable is large, that means that the variable is a good predictor of the dependent variable.
- $\mathbf{R^2_{adj}}$ **:** The sample R^2 tends to optimistically estimate how well the models fit the population. The model usually does not fit the population as well as it fits the sample

from which it is derived. Adjusted R^2 attempts to correct R^2 to more closely reflect the goodness of fit of the model in the population

- **PE:** The prediction error of the correlation coefficient is used to determine the predictive power of the models proposed [52].

II.8. Evaluation of the model

A developed QSAR model can be accepted generally in QSAR studies when it can satisfy the following criterion:

⇨ If correlation coefficient $R \geq 0.8$ (for *in vivo* data).

⇨If coefficient of determination $R^2 \geq 0.6$

⇨ If the standard deviations are not much larger than standard deviation of the biological data.

⇨If its F value indicate that overall significance level is better than 95%.

⇨If its confidence interval of all individual regression coefficients proves that they are justified at the 95% significance level [53].

II.9. QSAR Applications

Drug development: Helps design effective and safe drugs.

Toxicity prediction: Predicts toxicity of chemicals before testing.

Environmental safety: Assesses impact of chemicals on living organisms and nature.

Regulatory use: Used to register chemicals without animal testing.

Pesticide improvement: Designs safer and more effective agricultural chemicals [54].

References

[1]. Cramer, C. J. (2004). *Essentials of Computational Chemistry: Theories and Models*. John Wiley & Sons.

[2]. Mura, C., & McAnany, C. E. (2014). An introduction to biomolecular simulations and docking. *ArXiv preprint*, arXiv :1407.3752.

[3]. Leach, A. R. (2001). *Molecular Modelling: Principles and Applications*. Pearson Education.

[4]. Ohanessian, G. (2005). *Applications de la Chimie Quantique en Chimie - Cours*. École Polytechnique, Orsay.

[5]. Pollet, R. (2006). *Méthodes de la Chimie Quantique*. Université Paris-Sud 11 et ENS Cachan.

[6]. Doucet, J. P., & Weber, J. (1996). *Computer-Aided Molecular Design: Theory and Applications*. Academic Press.

[7]. Levine, I. N. (1999). *Quantum Chemistry* (5th ed.).

[8]. Debnath, A. K. (2001). Quantitative structure-activity relationship (QSAR) paradigm—Hansch era to new millennium. *Mini Reviews in Medicinal Chemistry*, 1(2), 187–195.

[9]. Phuong, H. N. G. (2007). *Synthèse et étude des relations structure-activité quantitatives (QSAR/2D) d'analogues benzo[c]phénathridiniques* [Thèse de doctorat, Université d'Angers].

[10]. Bounehas, R. (2014). *Approche qualitative de la relation structure-activité dans des dérivés bioactifs de 1,2,3-triazine* [Mémoire de master, Université Biskra].

[11]. Aouadi, Y., & Daha, F. N. (2014). *Étude de la relation structure–activité dans des oxazoles antibiotiques* [Mémoire de master, Université d'El-Oued].

[12]. Salah, T. (2013). *Développement des modèles QSAR pour la prédiction des activités inhibitrices antitrypanosomiennes des dérivés cryptolepine* [Mémoire de master, Université Biskra].

[13]. Pople, J. A., Santry, D. P., & Segal, G. A. (1965). *The Journal of Chemical Physics*, 43(S129).

[14]. Mellaoui, M. (2014). *Analyse des relations structures-propriétés pour des dérivés de la céphalosporine et de l'oxazole* [Thèse de doctorat, Université Biskra].

[15]. Piron, C. (1998). *Mécanique quantique : Bases et applications*. PPUR Presses Polytechniques.

[16]. Beleahi, Z. (2023). *Étude par modélisation moléculaire des dérivés de métronidazole* [Mémoire de master].

[17]. Serigi, A. (2023). *Design and selection of new heterocyclic molecules of medical interest by in silico method* [Mémoire de master, Université Biskra].

[18]. Bormann, S. (1990). *Chemical & Engineering News*, 68, 20–23.

[19]. Lipnick, R. L. (1986). *Trends in Pharmacological Sciences*, 7, 161–164.

[20]. Hansch, C., Leo, A., & Taft, R. W. (1991). *Chemical Reviews*, 91, 165–195.

[21]. Lantican, B. P., & Muir, R. M. (1967). *Plant Physiology*, 42, 1158–1160.

[22]. Hansch, C. (1969). *Accounts of Chemical Research*, 2, 232–239.

[23]. Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics*. Wiley-VCH.

[24]. Almi, I. (2021). *Contribution to drug design through computational studies of several series of bioactive heterocyclic molecules* [Doctoral dissertation, University of Biskra].

[25]. Lessigiarska, I. (2006). *Development of structure-activity relationships for pharmacotoxicological endpoints relevant to European Union legislation* [PhD thesis, Liverpool John Moores University].

[26]. Ida, K., & Hirane, F. (2019). *Étude QSAR des dérivés de la coumarine comme inhibiteur de la monoamine oxydase B* [Mémoire de master].

[27]. Karelson, M. (2000). *Molecular Descriptors in QSAR/QSPR*. Wiley-Interscience.

[28]. Cherroun, H. (2019). *Étude quantitative de la relation structure-activité (QSAR) des dérivés pipéraziniques de la phénothiazine* [Mémoire de master].

[29]. Authors Unknown. (2024). Toward grid-based models for molecular association. *Journal of Chemical Theory and Computation*. https://pubs.acs.org/doi/full/10.1021/acs.jctc.4c01293

[30]. Authors Unknown. (2006). *An adaptive grid-based method for computing molecular surfaces*. Department of Computer Science, University of Texas. https://www.cs.utexas.edu/ftp/techreports/tr06-56.pdf

[31]. Karelson, M. (2000). *Molecular Descriptors in QSAR/QSPR*. Wiley-Interscience.

[32]. Atkins, P. W., & de Paula, J. (2010). *Atkins' Physical Chemistry* (9th ed.). Oxford University Press.

[33]. Cammarata, A. (1967). An apparent correlation between the in vitro activity of chloramphenicol analogs and electronic polarizability. *Journal of Medicinal Chemistry*, 10, 525–527.

[34]. [Duplicate of (28)] Cherroun, H. (2019). *Étude quantitative de la relation structure-activité (QSAR) des dérivés pipéraziniques de la phénothiazine* [Mémoire de master].

[35]. Parr, R. G., Donnelly, R. A., Levy, M., & Palke, W. E. (1978). Electronegativity: The density functional viewpoint. *The Journal of Chemical Physics*, 68(8), 3801–3807.

[36]. Mulliken, R. S. (1934). A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. *The Journal of Chemical Physics*, 2, 782–793.

[37]. Parr, R. G., & Pearson, R. G. (1983). Absolute hardness: Companion parameter to absolute electronegativity. *Journal of the American Chemical Society*, 105(26), 7512–7517.

[38]. Yang, W., & Parr, R. G. (1985). Hardness, softness, and the Fukui function in the electronic theory of metals and catalysis. *Proceedings of the National Academy of Sciences*, 82(20), 6723–6726.

[39]. Zhou, Z., & Parr, R. G. (1990). Activation hardness: New index for describing the orientation of electrophilic aromatic substitution. *Journal of the American Chemical Society*, 112(15), 5720–5724.

[40]. Pearson, R. G. (1989). Absolute electronegativity and hardness: Applications to organic chemistry. *The Journal of Organic Chemistry*, 54(6), 1423–1430.

[41]. Parr, R. G., Szentpály, L. V., & Liu, S. (1999). Electrophilicity index. *Journal of the American Chemical Society*, 121(9), 1922–1924. https://doi.org/10.1021/ja983494x

[42]. Echota, O. (2010). *Quantitative structure-property relationship modeling algorithms, challenges and IT solutions* [Master's thesis, Masaryk University Faculty of Informatics].

[43]. Wu, W., Zhang, C., Lin, W., Chen, Q., Guo, X., Qian, Y., et al. (2015). Quantitative structure-property relationship (QSPR) modeling of drug-loader polymeric micelles via genetic function approximation. *PLOS ONE*, 10(3), e0119575.

[44]. Belouerghi, S., & Azri, S. (2020). *Étude qualitative et quantitative des relations structure-activité d'une série de composés phénoliques* [Mémoire de master, Université Mohamed Khider Biskra].

[45]. Solomonov, B. N., Konovalov, A. I., Norikov, V. B., Borbachuk, V. V., & Neklyudov, S. A. (1985). *Journal of General Chemistry*, 55, 1681.

[46]. Verma, R. P., & Hansch, C. (2010). QSAR modeling of taxane analogues against colon cancer. *European Journal of Medicinal Chemistry*, 45(4), 1470–1477.

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

[48]. Purkayastha, S. K., Jha, T., Pal, D. K., & De, A. U. (1993). *Anticancer Drug Design*, 8, 95–100.

[49]. Samanta, S., Srikanth, K., Banerjee, S., Debnath, B., Gayen, S., & Jha, T. (2004). *Bioorganic & Medicinal Chemistry*, 12, 1413–1423.

[50]. Tetko, I. V., Tanchuk, V. Y., & Villa, A. E. (2001). *Journal of Chemical Information and Computer Sciences*, 41, 1407–1421.

[51]. Aouisset, D. (2022). *In silico analysis of pyrazole derivatives applied to drug design* [Mémoire de master, Université Mohamed Khider Biskra].

[52]. Toufik, S. (2017). *Utude par la chimie computunalle des correlations 2D-Qsar et 3D-QSAR de qulques composes bioactifs* [Thèse doctorat, Université Mohamed Khider Biskra].

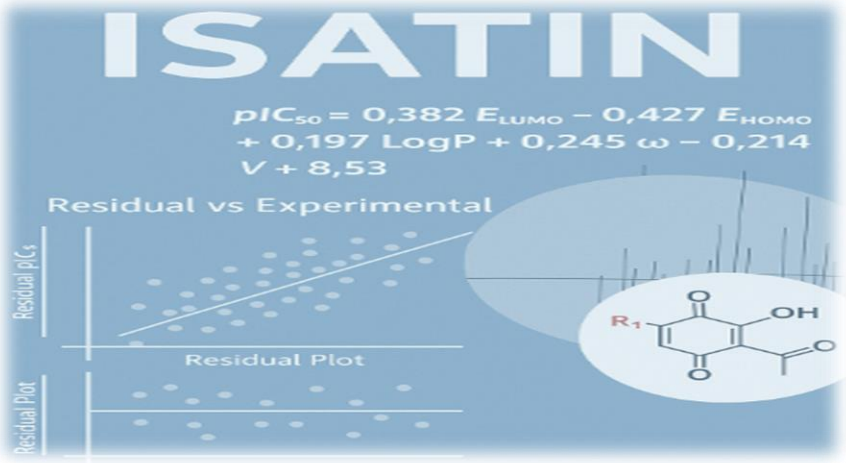
[53]. Ouassaf, M., Belaidi, S., Shtaiwi, A., & Chtita, S. (2022). Quantitative structure-activity relationship (QSAR) investigations and molecular docking analysis of Plasmodium protein farnesyltransferase inhibitors as potent antimalarial agents. *Jordan Journal of Pharmaceutical Sciences*, 15(3), 315–340. https://doi.org/10.35516/jjps.v15i3.407

[54]. Cherkasov, A., et al. (2014). *Journal of Medicinal Chemistry*.

https://doi.org/10.1021/jm4004285

Chapter III

Results and discussion



III.1. Introduction

The quantitative structure-activity relationship (QSAR) study is a method used in chemistry and pharmacology to predict or analyze the quantitative relationship between the chemical structure of a molecule and its biological activity or physicochemical properties.

This based on the hypothesis that specific molecular properties can be quantitatively linked to biological activity or properties of interest.

This study is feasible by establishing a mathematical relationship that quantitatively links molecular properties, called descriptor, with a macroscopic observable (biological activity, toxicity, physicochemical properties, act.) for a series of biologically active molecules using data analysis methods [1].

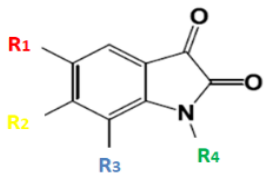
In this work we focus to develop QSAR models able to correlate the structural features of the derivatives of Isatin with their anticancer activity. To achieve this, we have selected a serie of 27derivatives with different physicochemical and electronic properties.

III. 2. Methodology

III.2.1. Experimental Data set

In this study, a series of 27 selected isatin derivatives were analyzed. These compounds had been synthesized and evaluated for their anticancer activity against U937 cells (human monocyte-like histiocytic lymphoma). The experimental biological activity IC_{50} values (μM) has been taken from literature [2- 4] were converted to the negative logarithm of IC_{50}, **$pIC_{50} = -\log_{10}(IC_{50})$.**
The pIC_{50} values were used as indicators of biological activity in the QSAR analysis. The table III.1 below presents the chemical structures of the compounds along with their corresponding pIC_{50}values

Table III.1: Chemical structures of Isatin derivatives used in this study and their experimental activity for anticancer activity against U937 cells.



Comp. Number	R1	R2	R3	R4	PIC_{50}
A1	Br	H	Br	$CH_2CH=CH_2$	5.18
A2	Br	H	Br	$CH_2CH_2OCH_3$	5.46
A3	Br	H	Br	$CH_2CH_2CH(CH3)_2$	5.62
A4	Br	H	Br	$CH_2C_6H_5$	5.94
A5	Br	H	Br	$CH_2C_6H_4CH_3$[(b)]	6.31
A6	Br	H	Br	$CH_2C_6H_4OCH_3$[(b)]	5.74
A7	Br	H	Br	$CH_2C_6H_4OCH_3$[(c)]	5.75
A8	Br	H	Br	$CH_2C_6H_4NO_2$[(b)]	6.05
A9	Br	H	Br	$CH_2C_6H_4NO_2$[(d)]	5.64
A10	Br	H	Br	$CH_2C_6H_4CL$[(b)]	6.01
A11	Br	H	Br	$CH_2C_6H_4Br$[(b)]	6.20
A12	Br	H	Br	$CH_2C_6H_4CF_3$[(b)]	6.10
A13	H	Br	H	$CH_2C_6H_4CF_3$[(d)]	5.28
A14	Br	H	Br	$CH_2C_6H_4COOCH_3$[(b)]	5.92
A15	Br	H	Br	$CH_2C_6H_4(CH_3)_3$[(b)]	5.98
A16	Br	H	Br	$CH_2C_6H4C_6H_5$[(b)]	6.12
A17	H	Br	H	H	3.25
A18	Br	H	H	H	4.19
A19	H	H	H	H	4.13
A20	H	Br	Br	H	4.08
A21	F	H	H	H	4.01
A22	NO_2	H	H	H	3.88
A23	OCH_3	H	H	H	3.38
A24	Br	H	Br	H	4.98
A25	Br	Br	H	H	4.94
A26	Br	H	NO_2	H	3.59
A27	Br	Br	Br	H	5.17

(b) Substitutions at Para position., (c) Substitutions at Meta position., (d)Substitutions at Ortho position.

The detailed chemical structures of the derivatives are presented in figure III.1

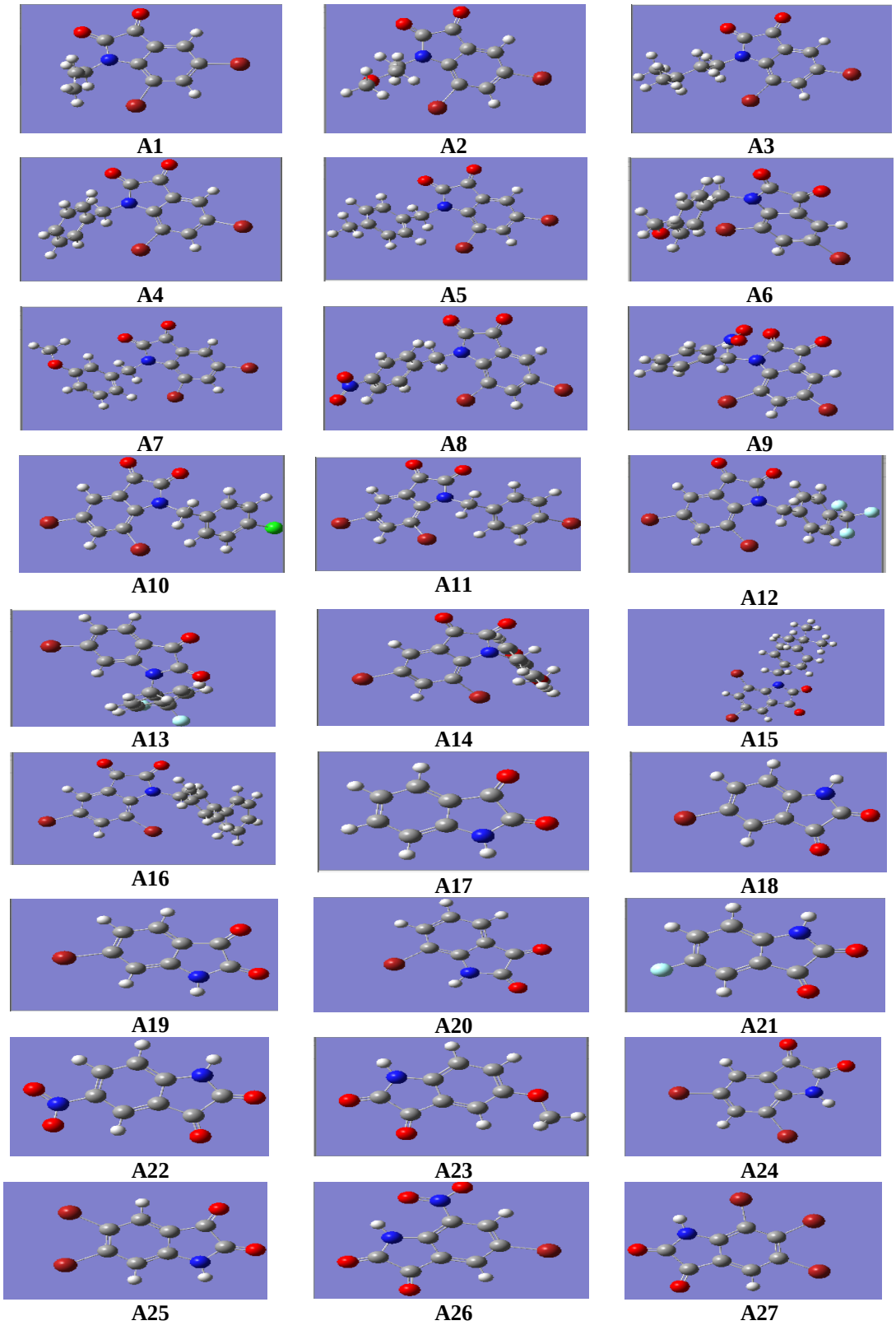


Figure III. 1. 3D Prestation of the structures of Isatin derivatives.

III.2.2. Descriptors

In this work, the descriptors chosen to describe the structure of the molecules constituting the series to be studied are given below:

Quantum descriptors
- Total Energy (E_T)
- Dipole moment(μ)
- Energy Gap
- Lowest unoccupied molecular orbital energy (E_{LUMO})
- Highest occupied molecular orbital energy (E_{HOMO})
- Absolute hardness (η)
- Electrophilicity index (ω)
- Absolute electronegativity (χ)

Geometric descriptors
- Molecular volume (MV)

Constitutional descriptor
- Molecular weight (MW)

Physicochemical descriptors
- Molar Refractivity (MR)
- Polarizability (Pol)
- Molecular surface Area (SAG)
- Partition coefficient (Log P)

Table III.2. Values of molecular descriptors used in this study.

Nbr	E_{HOMO}	E_{LUMO}	E g	η	χ	ω	μ	E_T	*SAG*	MV	LogP	MR	Pol
A1	-6.542	-3.087	3.455	1.727	4.815	6.710	4.100	182.61	385.97	686.69	0.82	71.78	25.42
A2	-6.564	-3.109	3.454	1.727	4.837	6.773	4.168	184.31	438.08	734.51	-0.08	73.66	26.25
A3	-6.477	-3.047	3.429	1.714	4.762	6.612	4.214	179.93	470.00	801.22	1.62	81.04	29.28
A4	-6.519	-3.092	3.427	1.713	4.805	6.739	4.128	183.38	413.73	806.02	1.09	91.37	31.6
A5	-6.448	-3.048	3.399	1.699	4.748	6.632	4.042	180.48	463.87	859.06	1.25	95.65	33.43
A6	-6.049	-3.028	3.021	1.510	4.538	6.818	3.263	185.52	479.70	882.60	0.11	97.74	34.07
A7	-6.038	-3.086	2.952	1.476	4.560	7.050	2.902	191.86	468.32	880.19	0.10	97.74	34.07
A8	-6.822	-3.359	3.462	1.731	5.090	7.483	5.187	203.6	468.84	865.34	-1.65	97.08	33.44
A9	-6.503	-3.009	3.493	1.746	4.756	6.475	7.449	176.20	411.53	848.94	-1.65	97.08	33.44
A10	-6.621	-3.185	3.435	1.717	4.903	6.999	4.086	190.45	456.18	849.75	0.87	96.08	33.53
A11	-6.596	-3.183	3.413	1.706	4.890	7.005	4.045	190.63	465.12	868.41	1.15	98.9	34.23
A12	-6.698	-3.247	3.451	1.725	4.972	7.165	4.061	194.97	473.61	881.57	1.66	96.58	33.16
A13	-6.717	-2.946	3.770	1.885	4.832	6.193	3.950	168.52	447.97	837.05	1.61	89.05	30.54
A14	-6.615	-3.165	3.450	1.725	4.890	6.931	5.289	188.60	510.11	941.44	0.51	102.1	35.99
A15	-6.444	-3.060	3.383	1.691	4.752	6.675	4.334	181.64	553.88	992.16	2.41	109.27	38.94
A16	*-6.187*	*-3.094*	3.093	1.54	4.641	6.962	4.169	189.45	505.39	1017.3	1.70	119.8	41.26
A17	*-6.550*	*-2.650*	3.900	1.950	4.600	5.425	5.934	147.64	240.39	443.29	-0.27	42.65	14.85
A18	-6.565	-2.922	3.642	1.821	4.743	6.177	5.202	168.10	287.19	506.43	-0.22	50.19	17.48
A19	-6.775	-2.872	3.902	1.951	4.824	5.963	4.390	162.27	287.23	506.04	-0.22	50.19	17.48
A20	-6.721	-2.890	3.830	1.915	4.805	6.028	5.033	164.05	278.16	501.26	-0.22	50.19	17.48
A21	-6.539	-2.852	3.687	1.843	4.696	5.981	5.358	162.75	253.46	451.19	-0.87	42.78	14.76
A22	-7.318	-3.307	4.011	2.005	5.312	7.036	5.478	191.48	295.13	502.35	-3.02	48.37	16.69
A23	-5.967	-2.591	3.376	1.688	4.279	5.423	5.388	147.52	299.07	520.26	-1.26	49.03	17.32
A24	-6.726	-3.130	3.596	1.798	4.928	6.755	4.095	183.81	325.91	564.80	-0.17	57.72	20.1
A25	-6.718	-3.060	3.657	1.828	4.889	6.536	4.216	177.86	330.25	563.29	-0.17	57.72	20.11
A26	-7.165	-3.519	3.646	1.823	5.342	7.828	2.765	213.01	322.90	558.40	-2.96	55.90	19.32
A27	-6.804	-3.203	3.601	1.800	5.003	6.952	3.672	189.18	364.86	661.96	-0.11	65.25	22.73

III.2.3. Material and Methods

The twenty-seven investigated molecules were pre-optimized by means of the Molecular Mechanics, with Force Field (MM+) included in HyperChem version 8.03 software [5]. Afterward, these derivatives were re-optimized with Gaussian 09 software [6] at the Density Functional Theory (DFT) level. The Becke, three-parameter, Lee–Yang–Parr (B3LYP) hybrid functional was employed in combination with the 6-31G(d) basis set to obtain reliable descriptors.

- The QSAR properties from HyperChem8.0 [5] were used to calculate the following descriptors: Molar refractivity (MR), molar weight (MW), molar Polarizability (Pol), Surface area grid (SAG), Partition coefficient (LogP) and Molecular volume (V).
- The Quantum Chemical descriptors: dipole moment (DM), the total energy (E_t), Highest Occupied Molecular Orbital energy E_{HOMO}, Lowest Unoccupied Molecular Orbital energy E_{LUMO} and their difference in absolute value Gap, the absolute hardness (η), Electrophilicity index (ω), the absolute electronegativity (χ) and were calculated by the DFT method (B3LYP \6-31G(d). using Gaussian 09[6] and GaussView5.0 [7].
- A relationship between independent (physicochemical and electronic descriptors) with dependent (biological activities) variables were determined statistically using regression analysis. In the present work, Multiple Linear Regression MLR analysis of molecular descriptors was carried out using the stepwise strategy in SPSS version 19 [8].

III .2.4. Quantitative Structure-Activity Relationships Studies

III .2.4.1. Development of QSAR models

This study aimed to develop the best QSAR model to explain the correlations between the physicochemical, electronic parameters and biological activity (IC50 values) of isatin derivatives.

The QSAR analysis was performed using the experimental anticancer activity values of the 27 molecules and the values of the 13 descriptors as shown in table III.3

The multiple linear regression analysis was performed [9] on all descriptors using SPSS 19 software**.** The multiple linear regression based on the elimination of descriptors until a valid model was obtained (including the critical probability: sig<0.05) for all descriptors and the model complete [10].

The QSAR models were obtained and presented by the following mathematical equation:

$$PIC50 = 1.709 + 0.157LogP + 0.003MV - 4.826\omega - 16.640E_{LUMO} + 2.729E_{HOMO}$$

N = 27 R = 0.948 R^2 = 0.899 SEE = 0.34036 F = 37.564.

- **N :** Number of compound
- **R^2 :** Coefficient of determination
- **SEE :** Standard error of estimate
- **F :** Fisher statistic
- **R :** Correlation coefficient

⇨ The best QSAR models were selected on the basis of various statistical parameters such as the correlation coefficient R between 0 to 1. QSAR model having squared correlation coefficient **$R^2 > 0.6$** will only be considered for validation [11]. Fischer's value F is the Fisher ratio, reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F–test indicate that the model is statistically significant [12, 13]. Our QSAR model R= 0.948 and R^2 = 0.899 explains 89.9% variance in biological activity allowed us to indicate firmly the correlation between different molecular descriptors (LogP, MV, ω, E_{LUMO}, and E_{HOMO}) with anticancer activity against U937 cells.

⇨The positive coefficient of Log P, E_{HOMO} and molecular volume shows that any increase in the values of this parameter leads to an increase in the activity of the isatin derivative. Specifically, an elevated Log P suggests increased lipophilicity, which may facilitate improved cellular membrane permeability. A higher molecular volume is indicative of potentially better molecular

accommodation within the active site of the target, enhancing hydrophobic. Moreover, a higher E_{HOMO} value reflects greater electron-donating ability, which may promote stronger interactions with biomolecular targets, thereby improving binding affinity.

⇨The negative coefficient of the lowest unoccupied molecular orbital (E_{LUMO}) and the electrophilicity index (ω) denote an inverse correlation with anticancer activity. A E_{LUMO} value reflects reduced electron-accepting capacity, potentially diminishing the molecules reactivity towards electrophilic biological targets. Similarly, a lower electrophilicity index indicates decreased ability to attract electron density from nucleophilic environments, possibly weakening interactions with key biological moieties. These findings emphasize the importance of maintaining a balanced electronic profile to maximize the anticancer potential of isatin derivatives.

III.2.4.2. Correlation Matrix

The correlation matrix between the descriptors obtained by MLR method analysis and the biological activity pIC50 is presented in Table III.3.

Table III.3. Correlation Matrix for Mode.

	pIC_{50}	**E_{HOMO}**	**E_{LUMO}**	**η**	**ω**	**MV**	**LogP**
PIC50	1.000						
E_{HOMO}	0.270	1.000					
E_{LUMO}	-0.425	0.545	1.000				
η	-0.650	-0.757	0.136	1.000			
ω	0.535	-0.338	-0.972	-0.359	1.000		
MV	**0.915**	0.370	-0.368	-0.724	0.507	1.000	
LogP	0.630	0.438	0.151	-0.399	-0.063	0.596	1.000

This matrix shows that the obtained descriptors are correlated with each other and with the biological activity. It also shows that the descriptor is a volume (MV) the most important parameter in the correlation between the selected descriptors and the biological activity of the Isatin derivatives.

III.2.4.3. Validation of QSAR Models

In order to test the validity of the predictive power of selected MLR model, **Leave-One-Out Cross-Validation (LOO-CV)** method [13,14] was used to estimates the trustworthiness of a model by predicting data. The developed models were validated by calculation of the following statistical parameters:

predicted residual sum of squares (PRESS), total sum of squares deviation (SSY) and cross validated correlation coefficient (R^2adj) (Table III.4).

Table III.4. Cross-validation Parameters.

Model	PRESS	SSY	PRESS\SSY	S_{PRESS}	R^2CV	R^2adj	PE
	2.433	24.081	0.1005	0.300	0.899	0.875	0.013

PRESS is an important cross-validation parameter as it is a good approximation of the real predictive error of the models. Its value being less than SSY points out that model predicts better than chance and can be considered statically significant.

The smaller PRESS value means the better of the model predictability [15]. From the results depicted in Table III.4, model is statistically significant.

To have a dependable QSAR model, **PRESS/SSY** ratio should be smaller than 0.4. The data presented in Table III.4indicate that for the developed models this ratio is 0.100.

The high value of **R^2cv** and **R^2adj** are essential criteria for the best qualification of the QSAR models [16]. Our findings for these QSAR models the value of R^2_{CV} = 0.899 and R^2_{adj} = 0.875.

Spress = 0.300 indicates that the model has acceptable predictive accuracy, with a relatively small deviation between the observed and predicted values. Although an ideal Spress is generally expected to be below 0.200 to ensure high predictive precision [17], Values below 0.500 are indicative of a reliable QSAR model [15]. Therefore, this result supports the strength and reliability of the developed QSAR model in predicting the anticancer activity of isatin derivatives.

The predictive error of the coefficient of correlation (**PE**) is yet another parameter used to evaluate the predictive power of the proposed models. We have calculated the PE value of the proposed models and they are reported in Table III.4.

For the model developed the condition **R> 6 PE** is satisfied and hence they can be said to have a good predictive power [18].

III.2.4.4. Activity prediction

The experimental, predicted, and residual biological activities of isatin derivatives are represented in Table III.5. The residual is due, on the one hand, to the uncertainty of the experimental measurement and, on the other hand, to the imperfection of the model therefore, if the predicted activity is closer to the experimental activity, the model may be applicable [19].

Table III.5. Experimental, predicted and residual activity of Isatin derivatives .

Comp	PIC_{50}exp	Prediction	Residual	Comp	PIC_{50}exp	Prediction	Residual
1	5.180	5.287	-0,107	**15**	5.950	6.556	-0 ,606
2	5.460	5.317	0,142	**16**	6.120	6.401	-0,281
3	5.620	5.781	-0,163	**17**	3.250	3.195	0,054
4	5.940	5.734	0,205	**18**	4.190	4.279	-0.089
5	6.310	5.921	0,388	**19**	4.130	3.913	0.216
6	5.740	5.670	0,069	**20**	4.080	4.020	0.059
7	5.700	5.538	0,161	**21**	4.010	3.846	0.163
8	6.050	5.528	0,521	**22**	3.880	4.028	-0 ,148
9	5.640	5.390	0,249	**23**	3.380	3.919	-0 ,539
10	6.010	5.872	0,137	**24**	4.980	4.720	0,259
11	6.200	5.973	0,226	**25**	4.940	4.631	0,308
12	6.100	6.111	-0 ,011	**26**	3.590	4.356	-0,786
13	5.280	5.598	-0,318	**27**	5.170	5.099	0,076
14	5.920	6.123	- 0,203				

The following figure III.2 shows the plots of linear regression of predicted versus experimental values of anticancer biological activity. The plots for model show a good deal of correspondence with experimentally reported data having R^2= 0.898. Thus, our QSAR model can be successfully applied to predict the anticancer activity in this series of isatin.

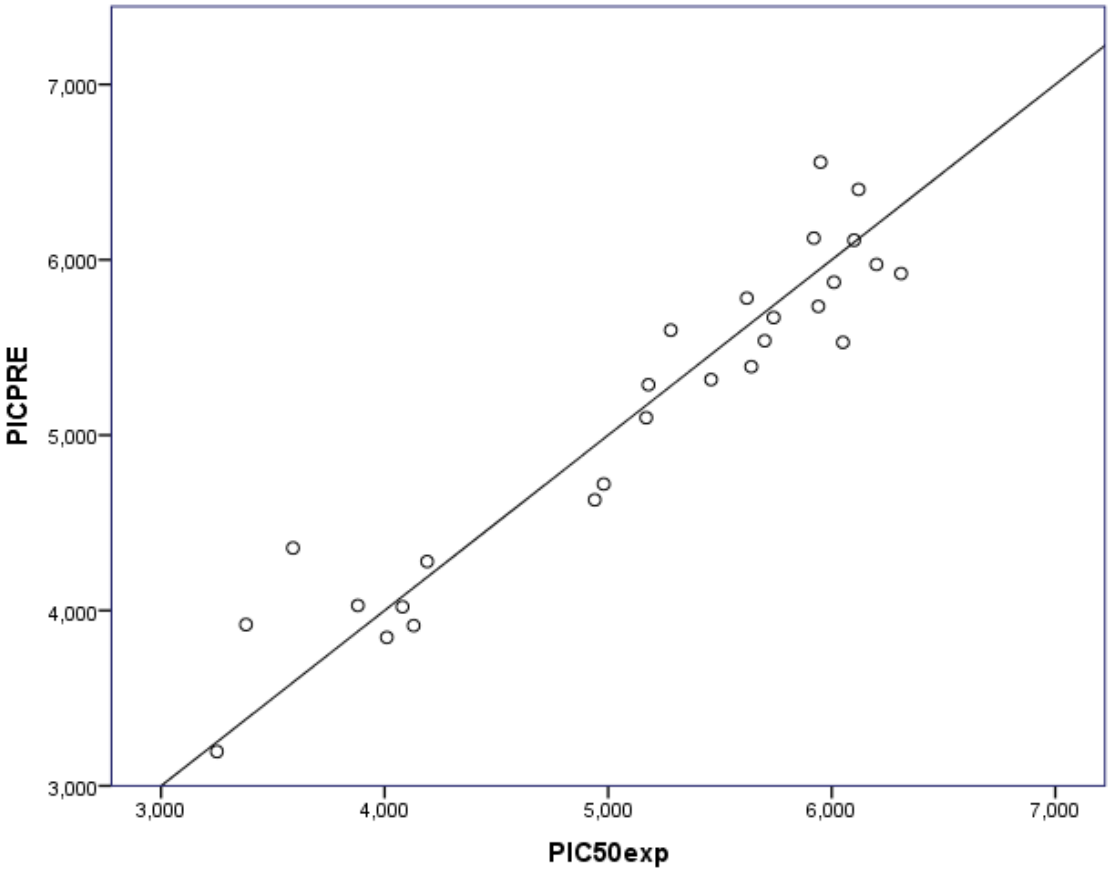


Figure III.2. Scatter plot between the observed and predicted activity of model.

To investigate the presence of a systematic error in developing the QSAR model, the residuals of predicted values of the biological activity (log (1/IC50) was plotted against the experimental values, as shown in Figure III.3.

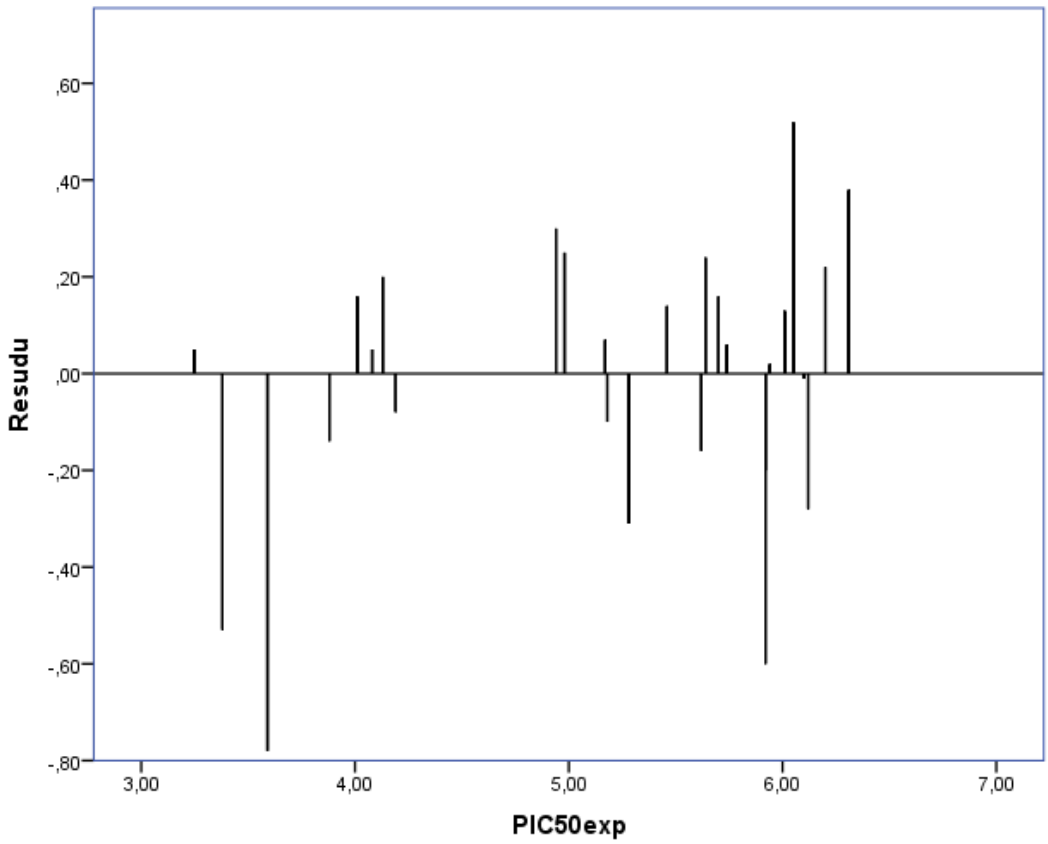


Figure III.3. Plots of the residual values against the experimentally observed.

The propagation of the residuals on both sides of zero indicates that no systemic error exists[21.22]. It indicates that these models can be successfully applied to predict the anticancer activity against U937 cells of isatin derivatives.

References

[1]. Fayet, G. (2010). *Développement de modèles QSAR pour la prédiction des propriétés d'explosibilité des composés nitroaromatiques* (Doctoral dissertation, Université Pierre et Marie Curie, Paris VI).

[2]. Adamo, C., & Barone, V. (2000). "A TDDFT study of the electronic spectrum of stetrazine in the gas-phase and in aqueous solution." *Chemical Physics Letters, 330*, 152–160.

[3]. Parac, M., & Grimme, S. (2003). "Comparison of Multireference Møller−Plesset Theory and Time-Dependent Methods for the Calculation of Vertical Excitation Energies of Molecules." *The Journal of Physical Chemistry A, 106*(29), 6844–6850.

[4]. Becker, L., Hinrichs, K., & Finke, U. (1993). "A New Algorithm for Computing Joins with Grid Files." In *Proceedings of the 9th International Conference on Data Engineering*, Vienna, Austria, 190–197.

[5]. HyperChem (Molecular Modeling System). (2008). Hypercube, Inc., 1115 NW 4th Street, Gainesville, FL 32601, USA.

[6]. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., ... & Fox, D. J. (2009). *Gaussian 09*. Gaussian, Inc., Wallingford CT.

[7]. Dennington, R., Keith, T., & Millam, J. (2009). GaussView, Version 5.0 [Computer software]. Semichem Inc., Shawnee Mission, KS, USA.

[8]. SPSS Inc. (2010). *SPSS 19 for Windows* [Computer software]. SPSS Inc., 444 North Michigan Avenue, Suite 3000, Chicago, IL 60611, USA.

[9]. Carlson, H. A. (2002). "Protein flexibility is an important component of structure-based drug discovery." *Current Opinion in Chemical Biology, 6*, 447–452.

[10]. Damale, M. G., Harke, S. N., Khan, F. A. K., *et al.* (2014). "Recent advances in multidimensional QSAR (4D–6D): A critical review." *Mini Reviews in Medicinal Chemistry, 14*(1), 35–55.

[11]. Ajeet, K., Bijander, (2012). Int. J. Pharm. & Pharm. Sci., 4, 445.

[12]. Derecho, K., et al. (2015). "Qualitative Structure-Activity Relationships and 2D-QSAR Modeling of TNF-α Inhibition by Thalidomide Derivatives." Journal of Bionanoscience, vol.9.no.5, pp.395-400.

[13]. Clark, M., & Cramer, R. D. (1993). "The probability of chance correlation using partial least squares (PLS)." *Quantitative Structure–Activity Relationships, 12,* 45–137.

[14]. Van der Voet, H. (1994). "Comparing the predictive accuracy of models using a simple randomization test." *Chemometrics and Intelligent Laboratory Systems, 25*, 23–313.

[15]. Kuzmanović, P., Sanja, O., Dragoljub, D., Cvetković, D., & Barna, D. (2009). "QSAR analysis of 2-amino or 2-methyl-1-substituted benzimidazoles against *Pseudomonas aeruginosa*." *International Journal of Molecular Sciences, 10*, 1670–1682.

[16]. Ajeet, K., & Bijander. (2012). *International Journal of Pharmacy and Pharmaceutical Sciences, 4*, 445.

[17]. Ida, K., & Hirane, F. (2019). *Étude QSAR des dérivés de la coumarine comme inhibiteur de la monoamine oxydase B* [Mémoire de master].

[18]. Srivastava, A. K., & Shukla, N. (2013). "Quantitative structure–activity relationship (QSAR) studies on a series of imidazole derivatives as novel ORL1 receptor antagonists." *Journal of Saudi Chemical Society, 17*, 321–328.

[19]. Matthew, C., & Cramer, R. D. (1993). "The probability of chance correlation using partial least squares (PLS)." *Molecular Informatics, 12,* 137–145.

[20]. Mellaoui, M., Belaidi, S., Bouzidi, D., & Gherrf, N. (2014). "Electronic structure and physical-chemistry property relationship for cephalosporin derivatives." *Quantum Matter, 3*, 435.

[21]. Heria, M. J., & Kyani, A. (2024). "Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds: A PCA–MLR–ANN approach." *Journal of Chemical Information and Computer Sciences, 44*, 1328.

General Conclusion

General Conclusion

The search for biological active compounds is fundamentally relies on understanding the relationship between chemical structure and biological activity. In this present study we applied quantitative structure–activity relationship (QSAR) of anticancer activity against U937 cells for isatin (1H-indole-2,3-dione) derivatives. Several descriptors, such as: physicochemical descriptors (MR, SAG, log P, Pol), quantum descriptors (E_T, μ, Gap, E_{LUMO}, E_{HOMO}, η, ω, χ), geometric descriptors (MV) and constitutional descriptors (MW), were used in the development of the QSAR model.

A multiple linear regression (MLR) analysis was performed to derive quantitative structure activity relationship model which were further evaluated internally for the prediction of anticancer activity for isatin derivatives.

The predictive power of QSAR model was validated by the cross validation 'Leave-one-out'. Where the best QSAR model has acceptable statistical quality and predictive potential as indicated by the value of cross validation.

These descriptors (E_{LUMO}**,** E_{HOMO}, LogP,ω , MV) are reliable for the prediction of activity. The high correlation observed between experimental and predicted values of anticancer activity which confirmed the predictive power and the good quality of the QSAR models.

Abstract

A series of twenty-seven molecules derived from isatin (1H-indole-2,3-dione) derivatives is based on the quantitative structure-activity relationship (QSAR). The analysis was based on a set of molecular descriptors, including: E_{HOMO}, E_{LUMO}, log P, ω and MV. Multiple linear regression (MLR) was used to establish the mathematical relationship between molecular descriptors and the biological activity of the isatin derivatives.

The prediction of QSAR model obtained was confirmed by the method of LOO cross-validation. A high correlation between experimental and predicted values of anticancer activity was observed in the results, indicating the validation and the good quality of the QSAR models.

Keywords: Isatin, Anticancer activity, QSAR, MLR, Cross Validation.

Résumé

Une série de vingt-sept molécules dérivées de l'isatine (1H-indole-2,3-dione) a été étudiée sur la base de la relation quantitative structure-activité (QSAR). L'analyse repose sur un ensemble de descripteurs moléculaires, incluant : E_{HOMO}, E_{LUMO}, log P, ω et MV. Une régression linéaire multiple (RLM) a été utilisée pour établir la relation mathématique entre les descripteurs moléculaires et l'activité biologique des dérivés de l'isatine.

La prédiction du modèle QSAR obtenus a été confirmée par la méthode de validation croisée « leave-one-out » (LOO). Une forte corrélation entre les valeurs expérimentales et prédites de l'activité anticancéreuse a été observée dans les résultats, ce qui indique la validité et la qualité de modèle QSAR obtenus.

Mots Clés : Isatin, Activité anticancéreuse, QSAR, MLR, Validation croisée.