University Med Khider of Biskra
Faculty of Exact Sciences and Sciences of Nature and Life
Department of Matter Sciences

MASTER'S GRADUATION THESIS

Domain: Matter science
Spinneret: Chemistry
Field: Pharmaceutical Chemistry

Réf. :

Presented by:
KHALED abderaouf

14-06-2022

In silico study of natural products as potential lead compounds in drug discovery

---

Jury members :

Mme.Harkati Dalal	Professeur	University Med Khider of Biskra	Supervisor
Mme.LARAOUI Habiba	M.C.B	University Med Khider of Biskra	president
Mr.BELAIDI Salah	Professeur	University Med Khider of Biskra	Examiner

2021-2022

Acknowledgments

First of all, I owe a debt of gratitude to Allah, who gave me the strength to accomplish this work and granted me the capability to proceed successfully.

*At the outset, I would like to thank my supervisor Mme **Harkati Dalal** for her efforts, advice, and comments to achieve this work.*

*I wish to extend my sincere thanks to Mme **LARAOUI Habiba** and **Mr.BELAIDI Salah** for accepting reading, and evaluating this research paper.*

I would like also to thank all the teachers who support me with their continual incentives, and their encouraging pieces of advice.

*My deepest thanks go to all my friends who have constantly given me support and strength to continue this research and are extended to my colleagues who never refused to assist me. I am very grateful to **KHELFAOUI Hadjer** who gives me a helping hand during this research.*

I wish also to thank my entire family for providing a loving environment for me.

Abstract

The term drug design describes the systematic search for new compounds with biological activity. Understanding the drug design process is necessary for the development of new drugs and their clinical applications to alleviate modern epidemic diseases. Explain the principles of drug design in the pharmaceutical industry based on targets and selected ligands using molecular docking, pharmacophore modeling, and virtual screening methods. Computational drug design (CADD) is a specialized field that uses computational knowledge-based methods to aid the drug discovery process. The calculation method is estimated to save up to 2-3 years and $300 million. Chemoinformatics enables researchers to perform virtual screening to select lead compounds for synthesis and screening. This enables scientists to make quick decisions on lead compound identification and optimization. Researchers can use in silico ADMET modeling to find a bioavailable drug with suitable drug metabolism properties. The development of selective protein kinase inhibitors that can block or modulate diseases in which these signaling pathways are abnormal is considered a promising approach for drug development. Therefore, the goal of this research is to use computational approaches (virtual screening, Molecular optimization, Molecular docking, and ADMET properties) to explore and identify novel active agents that can act as EGFR inhibitors from five selected molecules: thymoquinone, diosgenin, protodioscin, trigonelline, and Ladanein.

الملخص

يصف مصطلح (drug design) عملية البحث المنهجية عن مركبات جديدة ذات نشاط بيولوجي. إن فهم عملية تصميم الأدوية ضروري لتطوير عقاقير جديدة وتطبيقاتها السريرية للتخفيف من الأمراض الوبائية . شرح مبادئ تصميم الأدوية في صناعة الأدوية بناءً على العلاقة الترابطية بين البروتين المستهدف والدواء المفترض باستخدام الارساء الجزيئي وطرق الفحص الافتراضية. تصميم الأدوية الحسابي (CADD) هو مجال متخصص يستخدم أساليب حسابية قائمة على المعرفة للمساعدة في عملية اكتشاف الدواء. تقدر طريقة الحساب بتوفير ما يصل إلى 2-3 سنوات وما يقارب 300 مليون دولار. تمكن المعلوماتية الكيميائية الباحثين من إجراء فحص افتراضي لاختيار أحسن المركبات للتوليف والفحص. يتيح ذلك للعلماء اتخاذ قرارات سريعة بشأن تحديد المركبات الرئيسية وتحسينها. يمكن للباحثين استخدام نماذج (ADMET) للعثور على دواء متوفر بيولوجيًا بخصائص استقلاب الدواء المناسبة. يعتبر تطوير مثبطات بروتين كينيز الانتقائية التي يمكن أن تمنع أو تعدل الأمراض نهجًا واعدًا لتطوير الأدوية. لذلك ، فإن الهدف من هذا البحث هو استخدام الأساليب الحسابية (الفحص الافتراضي ، التحسين الجزيئي ، الارساء الجزيئي ، وخصائص ADMET لاستكشاف وتحديد المركبات النشطة الجديدة التي يمكن أن تعمل كمثبطات EGFR من خلال خمسة مركبات مختارة: ثيموكينون ، ديوسجينين ، بروتوديوسين ، تريغونيلين ، ولادانين.

Keywords

الكلمات المفتاحية

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CADD: computer-aided drug design.

SBDD: Structure-based drug design.

LBDD: Ligand-based drug design.

QSAR: Quantitative Structure-Activity Relationship.

VS: Virtual screening.

SBVS: structure-based virtual screening.

LBVS: ligand-based virtual screening.

3D: Three-dimensional.

ADMET: Absorption, distribution, metabolism, excretion, and toxicity.

PDB: Protein Data Bank.

EGFR: Epidermal growth factor receptor.

AMBER: Assisted Model Building with Energy Refinement.

DFT: Density Functional Theory.

MOE: Molecular Operating Environment.

FMO: frontier molecular orbitals.

HOMO: highest occupied molecular orbital.

LUMO: lowest unoccupied molecular orbital.

MEP: molecular electrostatic potential.

NBO: Natural Bond Orbitals.

NLS: natural Lewis structure.

General

Introduction

General Introduction

Therapeutic agents are chemical entities that prevent disease, assist in restoring health to the diseased, or alleviate symptoms associated with disease conditions. Medicinal chemistry is the scientific discipline that makes such drugs available either through drug discovery or design processes [1]. The term drug design describes the search for novel compounds with biological activity, on a systematic, rational basis. Basically, it relies on experimental information of the intended molecular target or a similar biomolecule (direct drug design) or known binders of such target (indirect drug design) [2].

Drug development is a complex process that is expensive and time-consuming. the classic method of Drug design takes 10 to 15 years or more. drug development can be divided into several stages; including target identification and verification by finding Disease-related molecular protein structure, lead discovery, drug candidate selection, Preclinical and clinical research. Today, the age of computer methods has become useful and Successful tools for drug discovery and development were submitted. structure-based and ligand-based methods are used to obtain information about ligands and apply homology modeling and Determine binding and interactions between compounds and receptors [2].

Natural anticancer drug development opened up a new concept for drug discovery. Furthermore, various compounds identified in fruits and vegetables have been used in cancer therapy. Identification and development of anticancer agents from the natural product by computer-aided drug design "CADD", virtual screening, and molecular docking have been well documented. *In silico* screening approach is the primary technique for identification of natural products as inhibitors of the target protein and predicting their interactions.

The first chapter of this thesis (Chapter 1 Literature review), provides a review of the literature on computer-aided drug design and moves on to explicit applications of computational methods in drug discovery. The second chapter in this work (Chapter 2 Materials and methods) presents the definition, theoretical details of each method, and software used in this study. And the third chapter (Chapter 3 Results and discussion), consists in discussing all the important results based on different approaches.

Chapter I

Literature review

Chapter I. Literature Review

Introduction

According to the World Health Organization (WHO), Cancer is the second leading cause of death in the world, Cancer is a group of diseases characterized by abnormalities in the cellular growth, proliferation, and survival pathways, which result in an uncontrolled expansion of cancer cells and tumor formation. Collectively, these diseases constitute one of the most pressing public health challenges of the twenty-first century [3].

Protein kinases have become attractive targets in the search for low-molecular-weight therapeutic agents to treat cancer. The complexity and number of protein kinases being used as molecular targets in drug discovery have greatly increased. The Human Genome Project's sequencing effort has revealed that 600 protein kinases and 130 protein phosphatases are most likely present in the human genome. Both of these enzyme classes can be divided into three categories based on their catalytic specificity: those specific for Tyr, those specific for Ser/Thr, and those specific for both Tyr and Ser/Thr. As a result, each cell will have a complement of 50–100 protein kinases [4].

Given the importance of protein kinases in mediating a variety of intracellular signaling pathways, it is not surprising that abnormal kinase activity is a common feature of tumors, and that kinase inhibitors have attracted considerable interest as cancer drug targets. Some of these dysregulated kinases are true oncogenes—tumor growth drivers.

The beginning of a research program aimed at developing a new molecule for therapeutic purposes will be dependent on a variety of parameters and scientific knowledge. Some of them must be considered when developing their research project. They will, for example, be assigned to: Choose a disease; select a target; identify ligands for a receptor; and benefit from the development of a biological test…

This chapter included a review of the literature on the following topics: [topic 1] (Computer-aided drug design) and [topic 2] (Protein kinases as targets for anticancer agents).

I.1 TOPIC 1: COMPUTER-AIDED DRUG DESIGN

I.1.1 What is a drug?

When we talk about drugs, we are referring to the active ingredient that has a therapeutic effect. The molecule that will interact with the target is this active ingredient. To produce a therapeutic effect, i.e., to be effective, an active ingredient must first meet its target.

I.1.2 Computer-aided drug design (CADD)

The term drug design describes the search for novel compounds with biological activity, as we all know, the discovery and development of new drugs is a very complex task, risky and costly process in terms of time, money, and manpower, and requires a lot of resources. Drug design and development takes an average of "10 to 15" years and much more than 1 billion dollars capital in total (**fig I.1**). Consequently, computational approaches to drug design are now widely used to improve the efficiency of the drug discovery process [5].

Computer-aided drug design (CADD) is widely used to reduce time, cost, and risk factors as a new drug design method. It has been shown that using CADD approaches can reduce drug discovery and development costs by up to **50%** [6]. CADD requires the use of any software-based process for establishing a standard for relating activity to structure. Computational drug design, discovery, and development approaches are rapidly gaining attention, implementation, and admiration, **CADD** uses structural knowledge of targets **structure-based** or known ligands with biological activity **ligand-based**, to facilitate the identification of promising drug candidates (**fig I.2**) [7].

Both methods can be applied to virtual screening, lead identification, and optimization using molecular docking. Recently, computer tools have been widely used in the pharmaceutical industry and research to improve the effectiveness and efficiency of the drug discovery and development process.
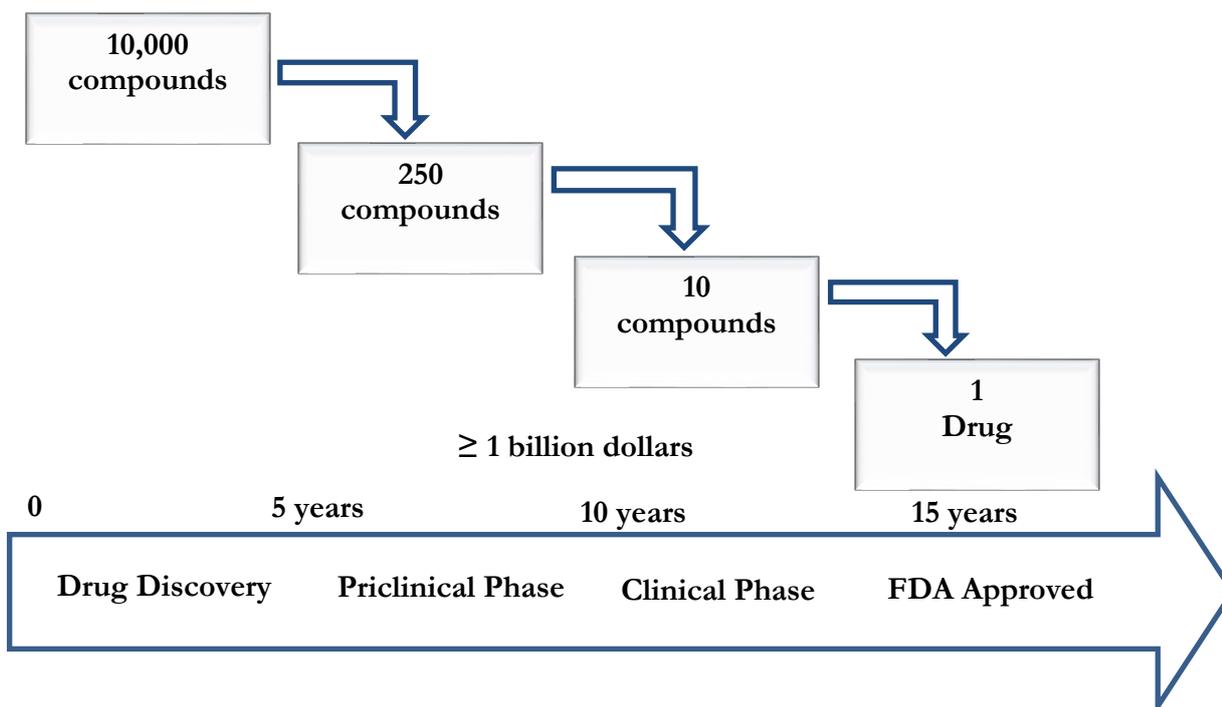
Figure I.1. The traditional process of drug discovery and development.

I.1.3 The most common approaches in CADD

Computer-aided drug design technologies can be classified into two main categories:

- Structure-based drug design / direct approach
- Ligand-based drug design / indirect approach

1) **Structure-based drug design (SBDD):** In **SBDD**, the structure of the target protein is known and the interaction or bio-affinity for all tested compounds calculate after the process of docking; to design a new drug molecule, which shows better interaction with the target protein (**fig I.3**) [8], [9].

2) **Ligand-based drug design (LBDD):** In **LBDD**, the 3D structure of the target protein is not known but the knowledge of ligands that binds to the desired target site is known. These ligands can be used to develop a pharmacophore model or molecule which possesses all necessary structural features to bind to a target active site (**fig I.4**) [10].

Generally, ligand-based techniques are pharmacophore-based approaches and quantitative-structure activity relationships (QSARs). In **LBDD** it is assumed that compounds that have similarities in their structure also have the same biological action and interaction with the target protein [7].
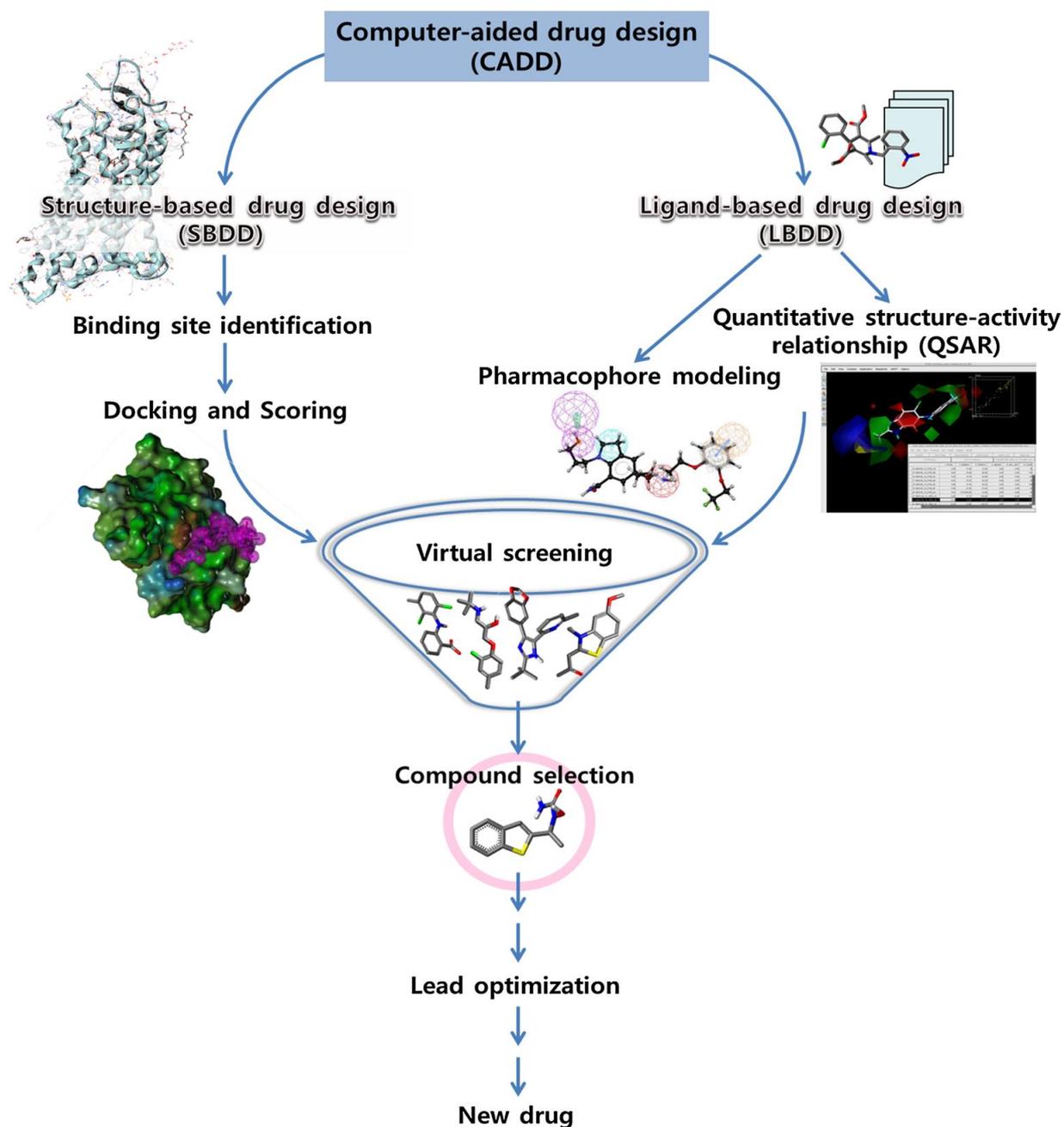
Figure I.2. Representative workflow for computer-aided drug
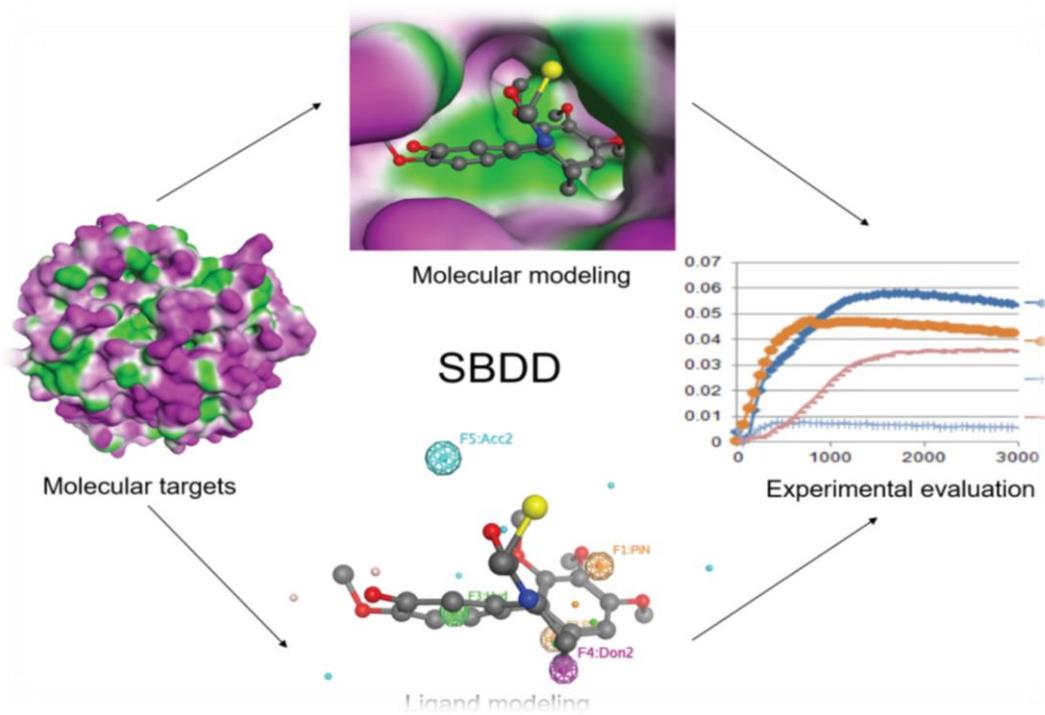
Figure I.3. Structure-based drug design (SBDD) process.
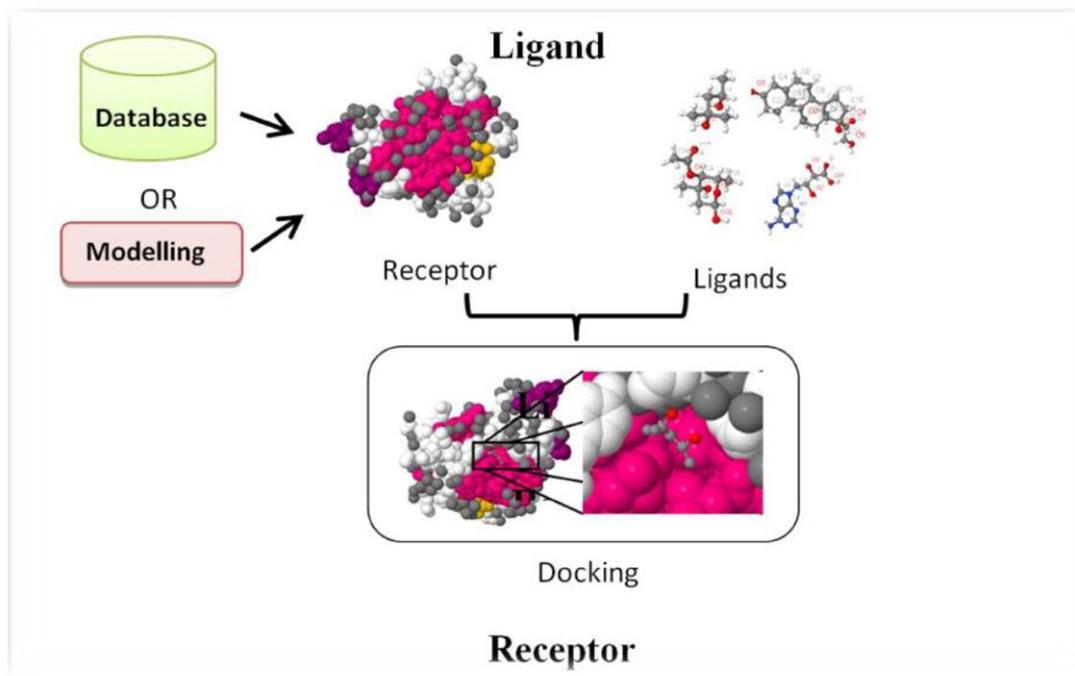
Figure I.4. Ligand-based drug design (LBDD).

I.1.4 Virtual Screening (VS)

Virtual screening is now a very useful tool for identifying the most promising bioactive compounds using information about the protein target or known active ligands. In recent years, virtual screening has emerged as a game-changing alternative to high-throughput screening, primarily in terms of cost-effectiveness and the likelihood of discovering the most appropriate novel hit by filtering large libraries of compounds [11].

There are two types of virtual screening **(VS)** approaches structure-based virtual screening **(SBVS)** and ligand-based virtual screening **(LBVS)** (**fig I.5**) [12]**.** The SBVS strategies are based on the structure of the target protein active site and usually rely on the following procedures: (1) molecular target selection and preparation, (2) compound database selection, (3) molecular docking, and (4) analysis of results. whereas the LBVS method is based on the estimation of calculated similarity between the known active and compound from databases.
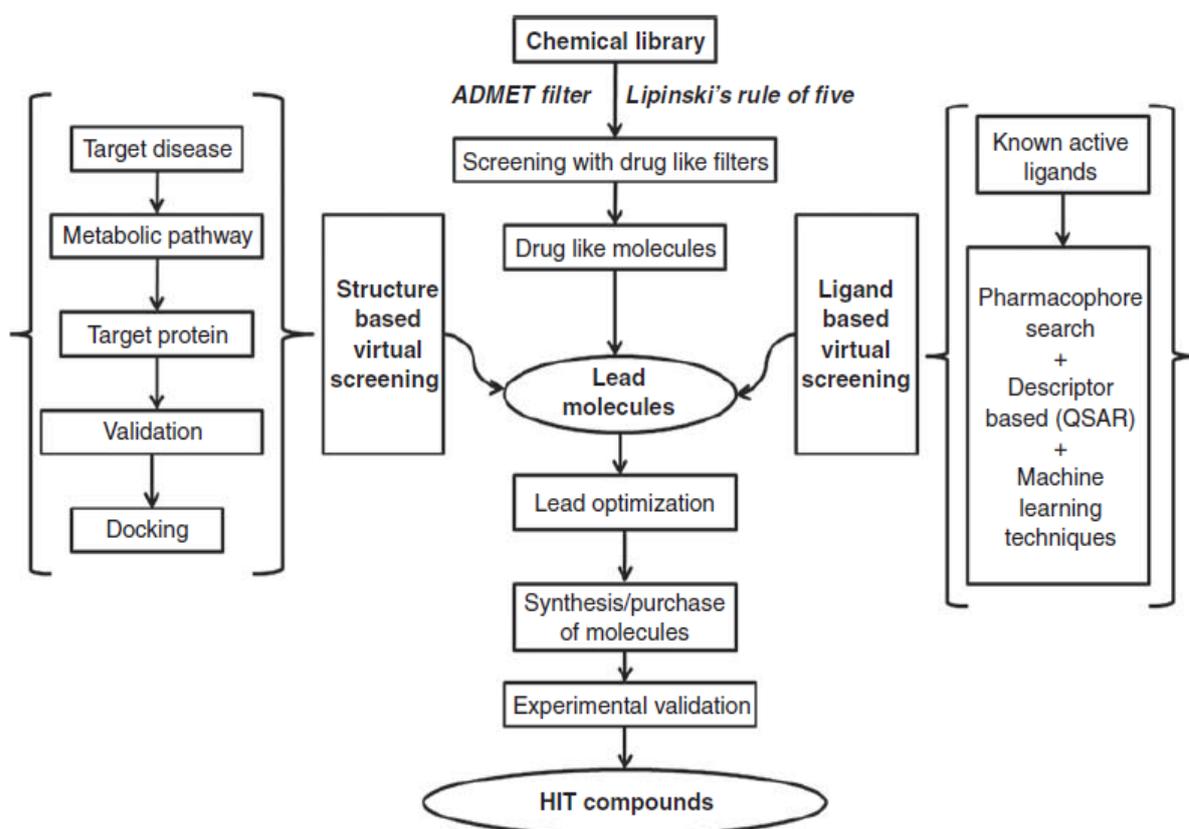
Figure I.5. Schematic diagram of VS process for SBDD & LBDD.

I.1.5 Molecular Docking and Scoring

One of the most well-known **SBDD** methods is **molecular docking** (**fig I.6**), This method is used to predict the most likely 3D conformations of small-molecule ligands within target binding sites, as well as to provide quantitative projections of the energy variations involved in the intermolecular recognition event. Molecular docking is divided into two steps: exploring the ligand conformational space within the binding cavity and calculating the binding energy for each predicted conformation [13].

Molecular docking methodology consists primarily of three interconnected goals: binding pose prediction, bio affinity, and virtual screening. The search algorithm and scoring functions for creating and analyzing ligand conformations form the foundation of the molecular docking method.

Target + Ligand   **Docking**   **Complex**
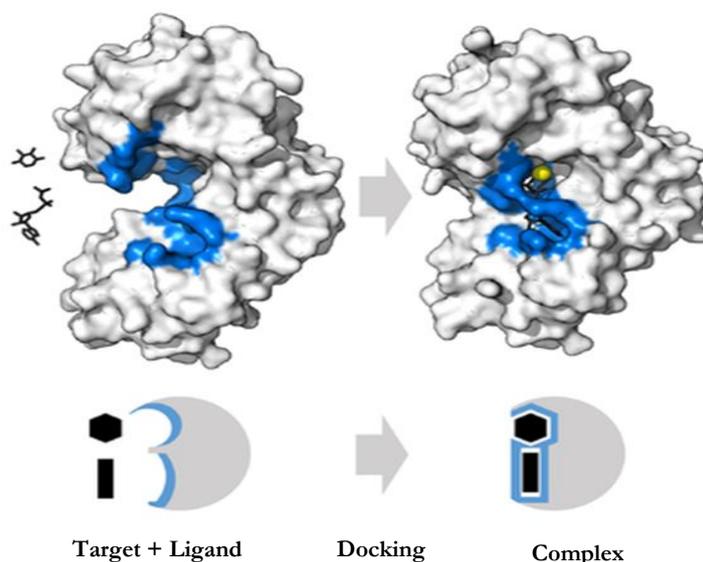
Figure I.6. The Process of Docking.

I.1.6 In silico ADMET

Due to high attrition rates due to poor pharmacokinetic profiles, it was necessary to determine the ADMET properties of leads in the early stages of drug screening. However, in terms of money and time, experimental evaluation of pharmacokinetic properties of millions of compounds is not a viable option, Thus, virtual screening can be used to filter hits and eliminate compounds with undesirable properties to quickly assess the drug-likeness of a lead compound before extensive experimental testing. Chemical or molecular descriptors are used to generate in silico ADMET filters, which are used to predict the drug-like properties of compounds [14].

It is important to remember that the in silico ADMET model tools are more useful in the qualitative analysis of **hits** or compound sets than in accurately predicting quantitative values. These methods help recognize a specific class of compounds for **in vitro** or **in vivo** assessment or evaluation of a specific descriptor and SAR [15].

I.1.7 The roles of CADD in drug discovery

- We can reduce our reliance on synthetic and biological testing as a result of it. It provides the most promising drug candidate by removing compounds with undesirable properties (low efficacy, poor ADMET, etc.) using in silico filters [16].
- It is a low-cost, time-saving, rapid, and automated process.
- We can learn about the drug-receptor interaction pattern through it.
- In comparison to traditional high throughput screening, it yields compounds with high hit rates by searching vast libraries of compounds in silico [17].
- These approaches reduce the likelihood of failures in the final phase.

I.2 TOPIC 2: PROTEIN KINASES AS TARGETS FOR ANTICANCER AGENTS

Protein kinases have emerged as a class of molecular targets with the potential to be **"cancer-specific,"** allowing for the selective targeting of cancer cells versus normal cells over the last decade. The cytotoxic side effects associated with conventional cancer chemotherapy would be eliminated by these selective anticancer drugs. Protein kinases are a large family of proteins that catalyze the transfer of the terminal phosphate group from ATP to the hydroxy group of a substrate protein's serine, threonine, or tyrosine [18].

Almost, any protein kinase has a unique homologous kinase domain, which is made up of (250-300) amino acid residues [19]. Based on kinase domain phylogeny, this superfamily of proteins has been divided into two major subdivisions (serine/threonine kinases and tyrosine kinases) and numerous subgroups with similar substrate specificities and regulatory modes. Protein kinase and phosphatase-dependent phosphorylation and dephosphorylation act as molecular switches to control the extent and duration of cellular signaling cascades.

The complex interplay of these molecular switches in specific signaling pathways controls cellular functions such as cell-cycle progression and differentiation. A malfunctioning signal transduction mechanism causes a variety of diseases, including cancer. Many protein kinase genes are mutated or amplified in tumor cells, and the proteins are over-expressed. Because of this difference between tumor and normal cells, protein kinases are appealing therapeutic targets for cancer treatment [18].

I.2.1 Cell structure

Because life is made up of cells, drugs must surely act on cells. (**Fig I.7**) [20]. shows the structure of a typical mammalian cell. All cells in the human body have a boundary wall called the cell membrane that encloses the cell's contents, known as the cytoplasm.
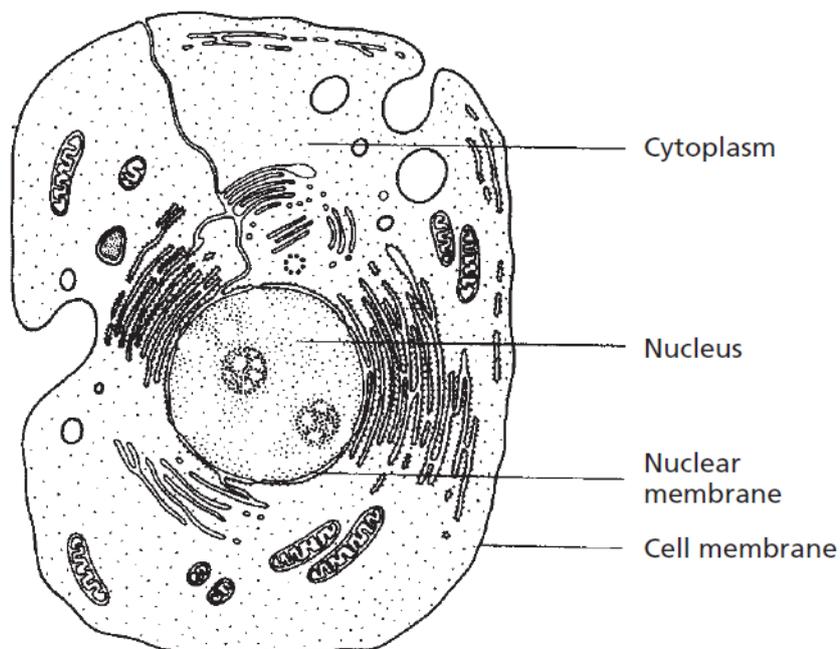
Figure I.7. A typical mammalian cell.

The membrane is not just made up of phospholipids, however. There are a large variety of proteins situated in the cell membrane (**Fig I.8**) [20]. Some proteins lie attached to the inner or the outer surface of the membrane. Others are embedded in the membrane with part of their structure exposed to one surface or both. The extent to which these proteins are embedded within the cell membrane structure depends on the types of amino acids present [21].
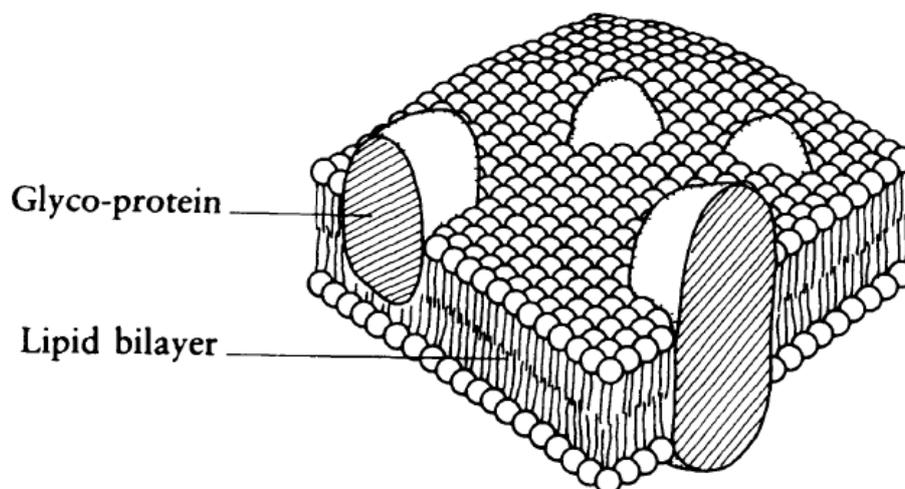
Figure I.8. Cell membrane.

I.2.2 Types of Drug targets

Drug action is frequently linked to a direct effect on cellular function. However, modifying the activity of a transport system can control a cellular function, or the substances can have direct intracellular action by inhibiting or activating an enzyme. Events that occur in the core can also be altered. The two most important classes of targets, from a biological standpoint, are membrane mediator receptors and enzymes. Ion channels, nuclear receptors, and other drug targets. Mediators are substances that allow different cell types to interact with one another. Drugs bind to specific sites on the body known as **receptors**. They interact with these mediator receptors and cause a cellular response via the signaling pathway, which is a collection of intracellular biochemical processes. As a result, they will be able to modify, block, or reverse the message that the mediator normally transmits to his target.

The main molecular targets for drugs are [21]:

- proteins: mainly enzymes, receptors, and transport proteins (nuclear receptors, G-protein-coupled receptors (GPCRs), **and tyrosine kinase receptors (RTKs)**.
- nucleic acids (DNA and RNA).
- Lipids: prostaglandins, thromboxanes, ceramides, sphingolipids …

The interaction of a drug with a macromolecular target involves a process known as binding. There is usually a specific area of the macromolecule where this takes place, known as the binding site (**fig I.9**) [20].
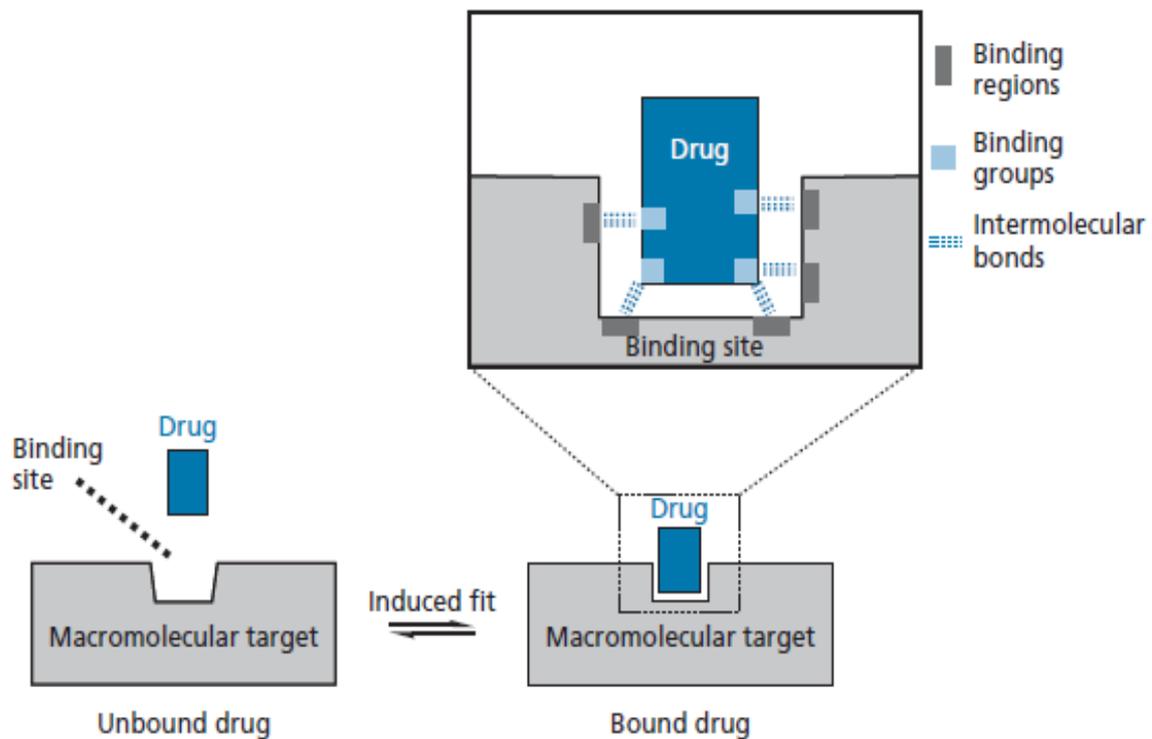
Figure I.9. The equilibrium of a drug being bound and unbound to its target.

I.2.3 Role of the Receptor

Receptors are proteins that are, by far, the most important drug targets in medicine. They are implicated in ailments such as pain, depression, Parkinson's disease, psychosis, heart failure, asthma, and many other problems [21].

A receptor is a protein molecule embedded within the cell membrane with part of its structure facing the outside of the cell, The protein surface will be a complicated shape containing hollows, ravines, and ridges, and somewhere amidst this complicated geography, there will be an area which has the correct shape to accept the incoming messenger. This area is known as the binding site and is analogous to the active site of an enzyme. When the chemical messenger fits into this site, it 'switches on' the receptor molecule, and a message is received (**fig I.10**) [20].
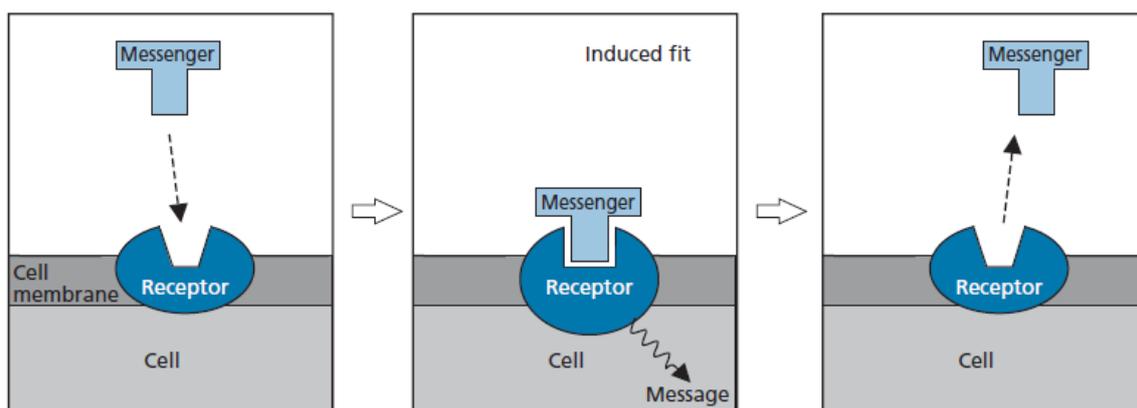
Figure I.10. Binding of a chemical messenger to a protein receptor.

I.2.4 The Protein-tyrosine Kinases Receptor (RTKs)

Communication between cells is mediated by compounds such as neurotransmitters and hormones, which, when released, activate receptors in target cells. This communication is critical for many physiological functions, and dysfunction in cell communication pathways can have serious consequences. Many diseases are caused by pathway dysfunction, and in these cases, drugs that act on receptors can be beneficial. As a result, receptors are critical drug targets [1].

Receptor tyrosine kinases (RTKs) are a widespread family of single-pass membrane proteins. They all share a similar structure, comprising an extracellular ligand-binding domain, a single-pass transmembrane helix, and an intracellular kinase domain that differentiates them from all other receptors. Given their central role in several cellular functions and a variety of human pathologies, RTKs are one of the most widely studied of the protein classes.

The reversible phosphorylation of receptor tyrosine kinases (RTKs) regulates RTK activity. In general, ligand binding to the receptor's extracellular domain causes autophosphorylation of tyrosine residues in the kinase domain, which results in kinase activation. Additional phosphorylation sites on the intracellular portion of the receptor can negatively and/or positively regulate RTK activity and create docking sites for downstream cytoplasmic targets [22].

RTKs are classified into twenty families based on the structure of their extracellular and intracellular domains, as shown in (**Fig I.11**) [23]. These various families of receptors for epidermal growth factors are distinguished (**EGFR**).

EGFR	PDGFR	FGFR	NGFR	EPHR	TIE	DDR	ROS	ROR	LMR

InsulinR	VEGFR	KLG/CCK	HGFR	AXL	RYK	RET	LTK	MUSK	?

EGFR	INSR	PDGFR-α	VEGFR1	FGFR-1	CCK4*	TRKA	MET	EPHA1	AXL	TIE	RYK*	DDR1	RET	ROS	LTK	ROR1	MUSK	RTK106
ERBB2	IGF-1R	PDGFR-β	VEGFR2	FGFR-2		TRKB	RON	EPHA2	MER	TEK		DDR2			ALK	ROR2
ERBB3*	IRR	CSF-1R	VEGFR3	FGFR-3		TRKC		EPHA3	TYRO3
ERBB4		KIT/SCFR		FGFR-4				EPHA4
			FLK2/FLT3							EPHA5
										EPHA6
										EPHA7
										EPHA8
										EPHB1
										EPHB2
										EPHB3
										EPHB4
										EPHB5
										EPHB6

AATYK
AATYK2
AATYK3

Figure I.11. Human receptor protein-tyrosine

I.2.5 Structure of tyrosine kinase receptors

The basic structure of a tyrosine kinase receptor consists of a single extracellular region (the N -terminal chain) that includes the binding site for the chemical messenger, a single hydrophobic region that traverses the membrane as an α-helix of seven turns, and a C -terminal chain on the inside of the cell membrane (**Fig I.12**) [21]. The C -terminal region contains the catalytic binding site.
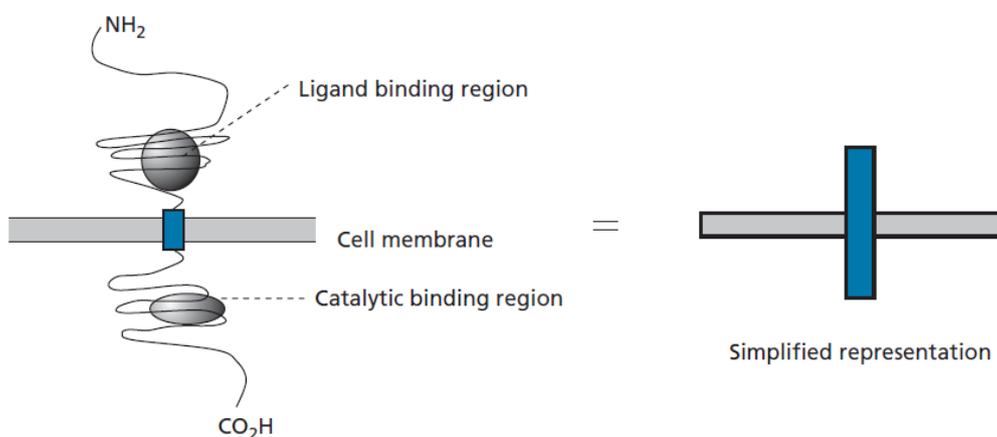
Figure I.12. Structure of tyrosine kinase receptors.

I.2.6 Activation mechanism for tyrosine kinase receptors

A specific example of a tyrosine kinase receptor is the receptor for a hormone called **epidermal growth factor** (EGF). EGF is a bivalent ligand that can bind to two receptors at the same time. This results in receptor dimerization, as well as activation of enzymatic activity. The dimerization process is important because the active site on each half of the receptor dimer catalyzes the phosphorylation of accessible tyrosine residues on the other half (**Fig I.13**) [21]. The important point to grasp at this stage is that an external chemical messenger has managed to convey its message to the interior of the cell without itself being altered or having to enter the cell.
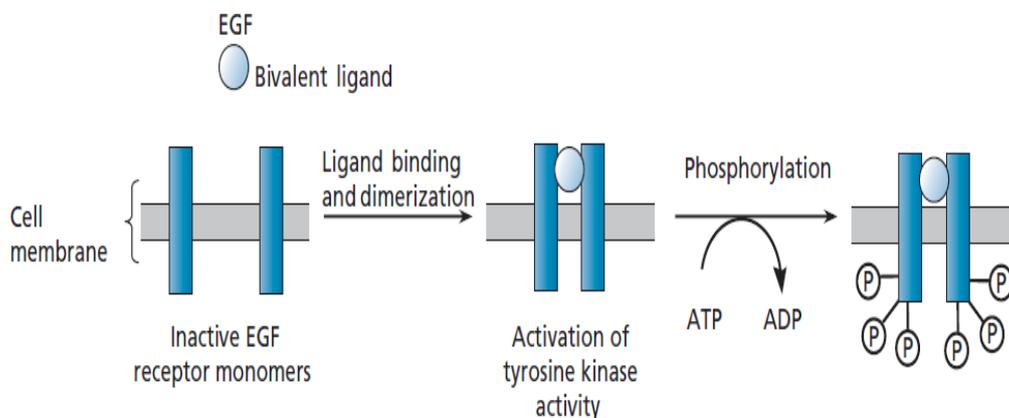
Figure I.13. Activation mechanism for the epidermal growth factor (EGF)

I.2.7 The EGFRs Subtypes and ligands

The HER family of tyrosine kinase receptors are encoded by the ERBB oncogenes and include EGFR (or HER1), HER2, HER3, and HER4. All HER-family receptors comprise an extracellular ligand-binding domain, a hydrophobic transmembrane domain, and an intracellular tyrosine kinase domain. The principal ligands for EGFR are the epidermal growth factor (EGF), transforming growth factor α (TGF-α), and amphiregulin, which binds specifically to EGFR. Stimulation of the receptor, caused by a ligand binding to the extracellular receptor domain, induces interaction with either a second EGFR (homodimerization) or with a different HER-family receptor (heterodimerization). Dimerization stimulates autophosphorylation of the intracellular tyrosine kinase domain and activation of a downstream signaling cascade. Ligand and composition of the receptor dimer determine which intracellular signaling pathways are activated; processes influenced by HER activation include cell-cycle progression, cell division, motility, survival, invasion, and adhesion (**fig I.14**) [24].
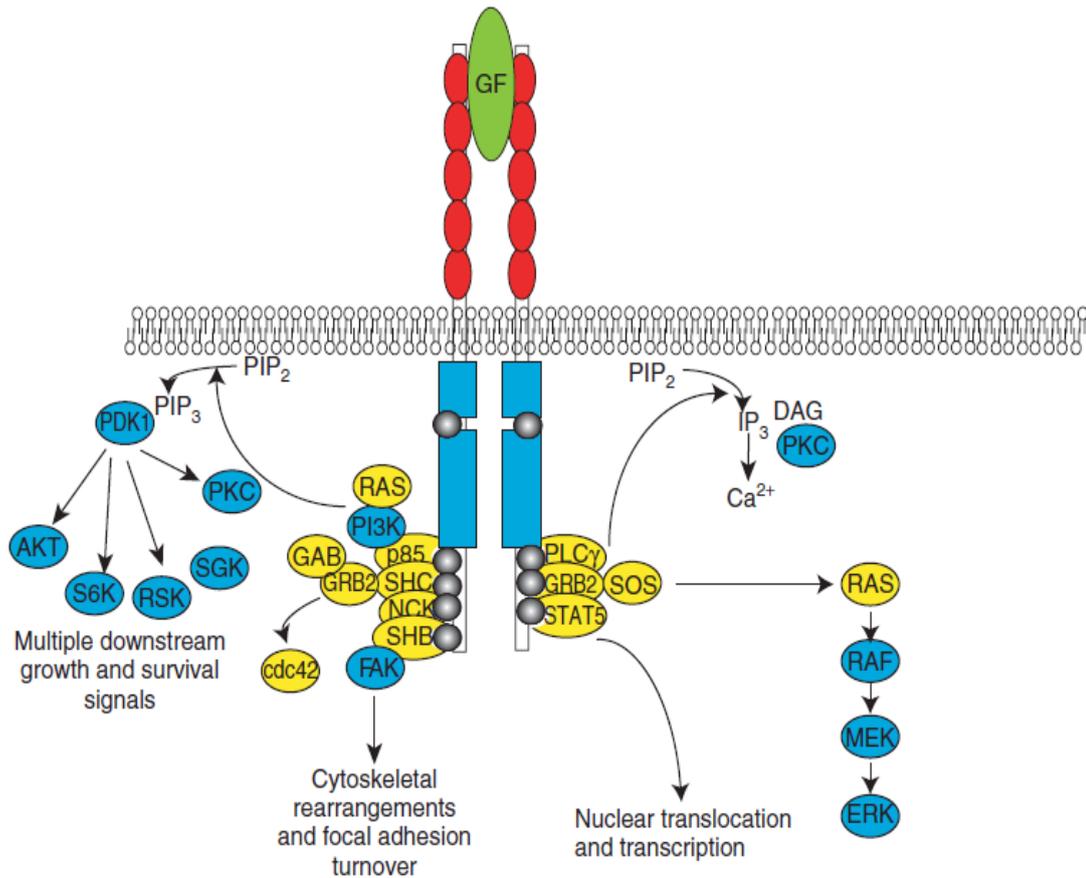
Figure I.14. The EGFR pathway.

I.2.8 EGFR (ERBB1)

a) The EGFR Protein

➢ Processing

EGFR is initially produced as a precursor, which undergoes cleavage of its aminoterminal signal and several glycosylation steps. The maturation process is initiated upon EGFR entry into the lumen of the endoplasmic reticulum (ER) and involves the transfer of carbohydrate moieties to specific asparagine residues. The carbohydrate side chains are then processed in the ER and in the Golgi network to produce a mature glycoprotein, which is exported through the secretory machinery to the plasma membrane. In addition to asparagine-directed glycosylation, EGFR was reported to undergo fucosylation and sialylation, side-chain modifications critical for its activity [25], [26].

➢ **Domain Structure**

EGFR was the first RTK to be characterized, and it shares with the later discovered ERBB members a characteristic domain structure (**Fig I.15**), including an extracellular ligand-binding region, that comprises two leucine-rich repeat domains (denoted I and III) and two cysteine-rich domains (II and IV), a single membrane-spanning region, and intracellular bilobular tyrosine kinase domain, and a carboxyl-terminal tail harboring multiple phosphorylation sites Note that domain II serves as a receptor dimerization site, which stabilizes homodimers of EGFR, as well as heterodimers with the other three ERBB proteins. Despite the common domain structure of all ERBBs, it is notable that two members of the family are peculiar; ERBB2 lacks the ligand-binding domain but carries an evolutionary conserved kinase domain, whereas ERBB3 retains the ligand-binding ability, but its kinase function is impaired [27].

b) EGFR Activation in Cancer

Recent genome sequencing efforts of a wide range of human tumors have repeatedly confirmed that the EGFR gene is one of the most frequently mutated genes in non-hematopoietic cancer. Importantly, all EGFR mutants have a fully functional or hyperactivated kinase domain. EGFR levels were initially found to be elevated in lung and head/neck carcinoma specimens [28].
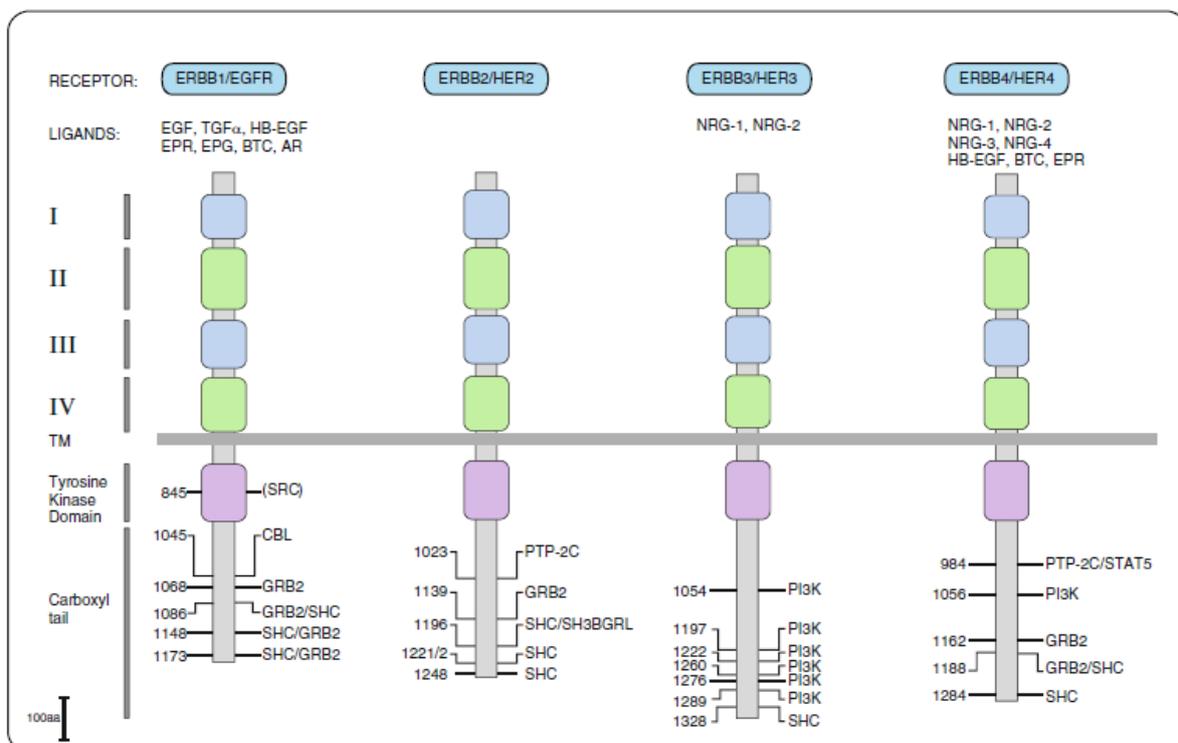
Figure I.15. Linear representation of the ERBB family of receptors.

Later studies confirmed EGFR overexpression, with or without gene amplification, in a broader range of carcinomas, and in some cancers, overexpression can predict recurrence or shorten patient survival. Brain tumors of glial origin provide an important example: EGFR gene amplification occurs in approximately 50% of high-grade gliomas. Furthermore, in glioblastoma, large and small deletions are frequently associated with gene amplification. The discovery of point mutations in the kinase domain of EGFR in non-small cell lung cancer was motivated by the findings of international clinical trials testing EGFR's specific kinase inhibitors, which revealed that patients enrolled in Japan had significantly higher response rates [29][30]. Following this and other similar findings, several groups reported activating mutations within the kinase domain of EGFR; these mutations frequently correlate with patient response to kinase inhibitors. Unfortunately, EGFR kinase inhibitors produce initial responses in patients, but their tumors eventually progress. Analyses of patients with acquired drug resistance revealed that, in addition to a primary drug-sensitive mutation, progressing tumors contained a secondary, resistance-conferring mutation resulting in a substitution of threonine for methionine at position 790 [31].

Despite the high frequency of EGFR genetic mutations, it appears that autocrine loops are the more common mode of EGFR activation in tumors. Tumor cells co-express EGFR and one or more EGF-like ligands in these cases, resulting in deregulated receptor activation, which may reduce patient response to conventional and targeted cancer therapies [32].

c) Cancer Therapeutic Strategies Targeting EGFR

Cetuximab is a human-mouse anti-EGFR chimeric antibody that has shown efficacy in colorectal cancer when combined with chemotherapy, as well as head and neck cancer when combined with radiotherapy This antibody binds to EGF and other EGFR ligands, preventing receptor dimerization and downstream signaling. Panitumumab is a fully human antibody that targets EGFR and is both effective and well-tolerated in colorectal cancer.

Chapter I References

[1] K. Stromgaard, P. Krogsgaard-Larsen, and U. Madsen, *Textbook of drug design and discovery*. CRC press, 2009.

[2] M. Gore and U. B. Jagtap, *Computational drug discovery and design*. Springer, 2018.

[3] D. J. Matthews and M. E. Gerritsen, *Targeting protein kinases for cancer therapy*. John Wiley & Sons, 2011.

[4] D. Fabbro *et al.*, "Protein kinases as targets for anticancer agents: from inhibitors to useful drugs," *Pharmacol. Ther.*, vol. 93, no. 2–3, pp. 79–98, 2002.

[5] A. Daina *et al.*, "Drug design workshop: A web-based educational tool to introduce computer-aided drug design to the general public," *J. Chem. Educ.*, vol. 94, no. 3, pp. 335–344, 2017.

[6] M. Xiang, Y. Cao, W. Fan, L. Chen, and Y. Mo, "Computer-aided drug design: lead discovery and optimization," *Comb. Chem. High Throughput Screen.*, vol. 15, no. 4, pp. 328–337, 2012.

[7] S. J. Y. Macalino, V. Gosu, S. Hong, and S. Choi, "Role of computer-aided drug design in modern drug discovery," *Arch. Pharm. Res.*, vol. 38, no. 9, pp. 1686–1701, 2015.

[8] S. S. Imam and S. J. Gilani, "Computer aided drug design: A novel loom to drug discovery," *Org. Med. Chem. Int. J.*, vol. 1, no. 3, pp. 113–118, 2017.

[9] Y.-C. Lo, R. Gui, H. Honda, and J. Z. Torres, "Quantitative Methods in System-Based Drug Discovery," in *Complex Systems, Sustainability and Innovation*, IntechOpen, 2016.

[10] S. Giuliatti, "Computer-Based Methods of Inhibitor Prediction," in *An Integrated View of the Molecular Recognition and Toxinology-From Analytical Procedures to Biomedical Applications*, IntechOpen, 2013.

[11] M. Lill, "Virtual screening in drug design," *Silico Model. Drug Discov.*, pp. 1–12, 2013.

[12] S. Kar and K. Roy, "How far can virtual screening take us in drug discovery?," *Expert Opin. Drug Discov.*, vol. 8, no. 3, pp. 245–261, 2013.

[13] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular docking: a powerful approach for structure-based drug discovery," *Curr. Comput. Aided. Drug Des.*, vol. 7, no. 2, pp. 146–157, 2011.

[14] J. Bajorath, "Integration of virtual and high-throughput screening," *Nat. Rev. Drug Discov.*, vol. 1, no. 11, pp. 882–894, 2002.

[15] M. P. Gleeson and D. Montanari, "Strategies for the generation, validation and application of in silico ADMET models in lead generation and optimization," *Expert Opin. Drug Metab.*

Toxicol., vol. 8, no. 11, pp. 1435–1446, 2012.

[16]   I. Kapetanovic, "Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach," *Chem. Biol. Interact.*, vol. 171, no. 2, pp. 165–176, 2008.

[17]   C. Shekhar, "In silico pharmacology: computer-aided methods could transform drug development," *Chem. Biol.*, vol. 15, no. 5, pp. 413–414, 2008.

[18]   Q. Li and G.-D. Zhu, "Targeting serine/threonine protein kinase B/Akt and cell-cycle checkpoint kinases for treating cancer," *Curr. Top. Med. Chem.*, vol. 2, no. 9, pp. 939–971, 2002.

[19]   S. K. Hanks and T. Hunter, "The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification 1," *FASEB J.*, vol. 9, no. 8, pp. 576–596, 1995.

[20]   J. Mann, *Murder, magic, and medicine*. Oxford University Press, USA, 2000.

[21]   G. L. Patrick, *An introduction to medicinal chemistry*. Oxford university press, 2013.

[22]   S. Germano, *Receptor Tyrosine Kinases: Methods and Protocols*. Springer, 2015.

[23]   P. Blume-Jensen and T. Hunter, "Oncogenic kinase signalling," *Nature*, vol. 411, no. 6835, pp. 355–365, 2001.

[24]   A. Okines, D. Cunningham, and I. Chau, "Targeting the human EGFR family in esophagogastric cancer," *Nat. Rev. Clin. Oncol.*, vol. 8, no. 8, pp. 492–503, 2011.

[25]   J. N. Contessa, M. S. Bhojani, H. H. Freeze, A. Rehemtulla, and T. S. Lawrence, "Inhibition of N-linked glycosylation disrupts receptor tyrosine kinase signaling in tumor cells," *Cancer Res.*, vol. 68, no. 10, pp. 3803–3809, 2008.

[26]   X. Wang, J. Gu, H. Ihara, E. Miyoshi, K. Honke, and N. Taniguchi, "Core fucosylation regulates epidermal growth factor receptor-mediated intracellular signaling," *J. Biol. Chem.*, vol. 281, no. 5, pp. 2572–2577, 2006.

[27]   D. L. Wheeler and Y. Yarden, *Receptor tyrosine kinases: family and subfamilies*. Springer, 2015.

[28]   F. J. Hendler and B. W. Ozanne, "Human squamous cell lung cancers express increased epidermal growth factor receptors.," *J. Clin. Invest.*, vol. 74, no. 2, pp. 647–651, 1984.

[29]   N. Sugawa, A. J. Ekstrand, C. D. James, and V. P. Collins, "Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas.," *Proc. Natl. Acad. Sci.*, vol. 87, no. 21, pp. 8602–8606, 1990.

[30]   V. A. Miller *et al.*, "Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non–small-cell lung cancer," *J. Clin. Oncol.*, vol. 22, no. 6, pp. 1103–1109, 2004.

[31] W. Pao *et al.*, "Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain," *PLoS Med.*, vol. 2, no. 3, p. e73, 2005.

[32] D. L. Wheeler *et al.*, "Mechanisms of acquired resistance to cetuximab: role of HER (ErbB) family members," *Oncogene*, vol. 27, no. 28, pp. 3944–3956, 2008.

Chapter II

Materials and methods

Chapter II. Materials and methods

Introduction

Cancer is a major cause of mortality worldwide with increasing numbers over the years. The toxicity of current chemotherapeutic drugs, coupled with drug resistance and poor prognosis, provides a much-needed impetus for the development of new anticancer drugs. Even so, medicinal plants and natural products are widely used as a primary source of treatment for various pathologies, including cancer, Natural anticancer drugs including plants (Nigella sativa "Ranunculaceae," Trigonella and foenum-graecum "Fabaceae," and Marrubium vulgare L. "Lamiaceae"), are the most preferred by herbalists for cancer treatment in North Africa [1].

Natural anticancer drug development opened up a new concept for drug discovery. Furthermore, various compounds identified in fruits and vegetables have been used in cancer therapy. Traditional medicines are known for their ability to exhibit a wide range of biological activities, including anticancer potential. Furthermore, synthetic analogs of natural compounds with increased potency and safety may be developed, positioning them as a beacon for cancer drug discovery. In fact, the majority of US Food and Drug Administration (FDA)-approved drugs are inspired by natural products. Another notable feature is that natural products can also be synthesized and have played a pivotal role in drug development by providing challenging synthetic targets [2].

Identification and development of anticancer agents from the natural product by computer-aided drug design "CADD", virtual screening, and molecular docking have been well documented. In silico screening approach is the primary technique for identification of natural products as inhibitors of the target protein and predicting their interactions.

This study aims to look into the interactions of five molecules that act as EGFR inhibitors and the EGFR protein receptor by the calculation of the Global and Local Reactivity, Electronic Properties Analysis, ADMET Properties, and Molecular Docking studies. The methodology used in this study is illustrated in **Fig II.1**.

```
┌─────────────────────────────┐                    ┌─────────────────────────────┐
│                             │                    │   Compounds collection      │
│                             │                    │      (screening)            │
│                             │                    └─────────────────────────────┘
│                             │                                  │
│                             │                                  ▼
│ Cristal strucure recived    │                    ┌─────────────────────────────┐
│ from PDB databse (ID:       │                    │ Ligand preparation by using │
│ 1xkk), (ID: 2itv), (ID:     │                    │ ChemDraw Ultra 12.0 and     │
│ 5hg5)                       │                    │ Chem3D Pro                  │
└─────────────────────────────┘                    └─────────────────────────────┘
              │                                                  │
              ▼                                                  ▼
┌─────────────────────────────┐                    ┌─────────────────────────────┐
│  Protien preparation using  │                    │ Study of the global         │
│  MOE software               │                    │ reactivity and Optimisation │
│                             │                    │ using Gausian               │
└─────────────────────────────┘                    └─────────────────────────────┘
                                                                 │
                                                                 ▼
                                                   ┌─────────────────────────────┐
                                                   │ Study of the ADMET          │
                                                   │ properties using swissADME  │
                                                   │ and ADMETlab tools          │
                                                   └─────────────────────────────┘
```

Compounds collection (screening)

Cristal strucure recived from PDB databse (ID: 1xkk), (ID: 2itv), (ID: 5hg5)

Ligand preparation by using ChemDraw Ultra 12.0 and Chem3D Pro

Protien preparation using MOE software

Study of the global reactivity and Optimisation using Gausian

Study of the ADMET properties using swissADME and ADMETlab tools

Induced Fit Docking using MOE software
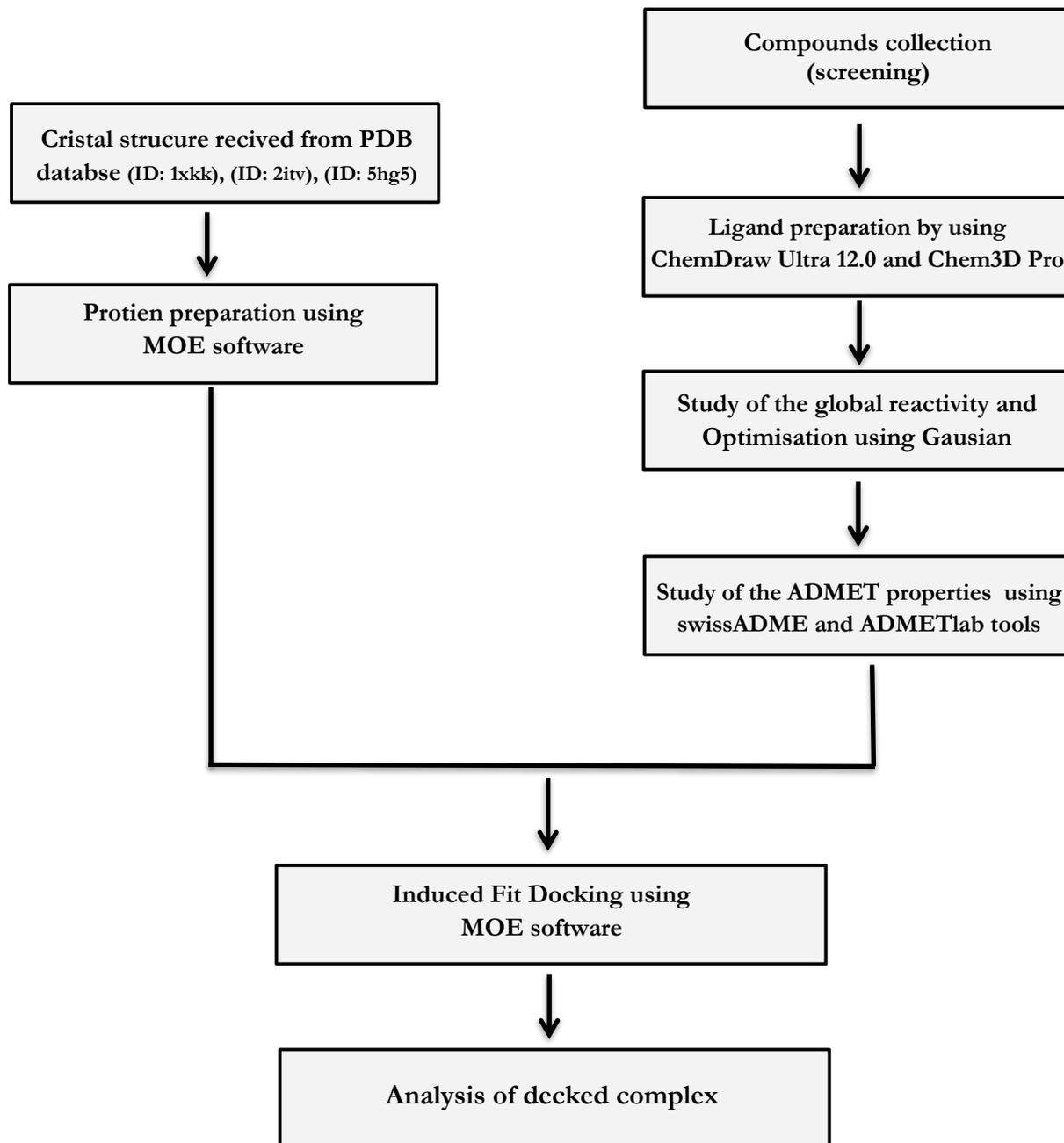
Analysis of decked complex

Figure II.1 Schematic representation of the methodology used in the study.

II.1 COMPOUND COLLECTIONS

An online search of published articles related to **Thymoquinone** from N. Sativa (fig II.2,1). **Diosgenin** (fig II.2,2), **protodioscin** (fig II.2,3), and **trigonelline** (fig II.2,4) from T. foenum-graecum, **ladanein** (fig II.2,5) from M. Vulgare, are the major compounds from selected North African medicinal plants used for cancer therapy [3].
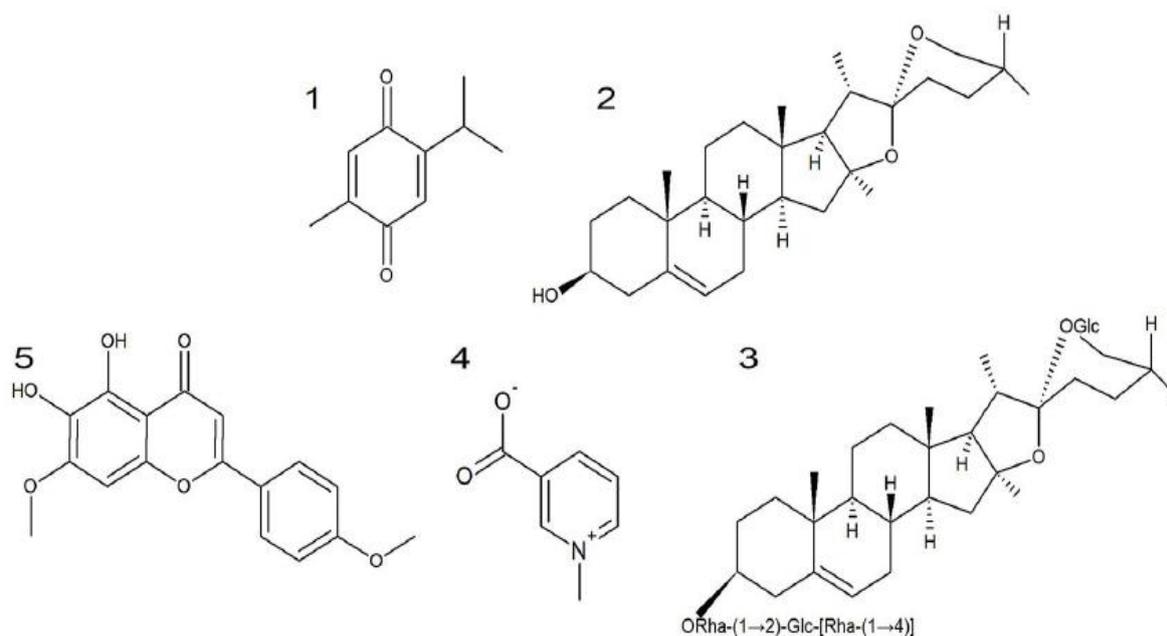
Figure II.2 Chemical 2D structures of the major compounds from selected North African plants.

✓ **Thymoquinone from Nigella sativa (N. sativa)**

N. Sativa (Ranunculaceae) is also known as **"Habbat Al-barakah"** in Arabic and black cumin or black seed in English. This plant is widely used in Arabic medicine to treat a variety of diseases, including cancer. Many active components of N. sativa have been isolated, including **thymoquinone**, thymohydroquinone, dithymoquinone, thymol, and alpha-hederin. Furthermore, quite a few pharmacological effects of N. sativa active components have been identified to include immune stimulation, anti-inflammation, hypoglycemic, antihypertensive, antiasthmatic, antimicrobial, antiparasitic, antioxidant, and anticancer effects [4].

- ✓ **Diosgenin, trigonelline and Protodioscin from Trigonella fenugreek**

   T. foenum-graecum (Fabaceae) is also known as **"helba"** in Arabic and fenugreek in English. This plant (**fig II.3**) has been described as having a variety of pharmacological properties, including anticancer properties. There are two major compounds isolated from fenugreek "Diosgenin and protodioscin" [5]. The most powerful anticancer components of fenugreek identified are trigonelline, diosgenin, and protodioscin, these compounds were found to be effective against a large variety of cancer lines and animal models as well as to target many signaling pathways involved in cancer hallmarks [6].

Diosgenin is the main compound of fenugreek, it reduces cell proliferation in a time- and dose-dependent manner in HT-29 colon cancer cells. In addition, it induced cell death in several breast, pancreatic, and prostate cancer cell lines. While Protodioscin Inhibits the growth of   HL-60 leukemia cells while showing weak inhibitory activity against KATO-III gastric cancer cells. In addition, the compound was cytotoxic to 60 cell lines from the National Cancer Institute [3], [7-9].

- ✓ **Ladanein from Marrubium Vulgare**

   M. Vulgare (Lamiaceae) is known as "**merriwa**", "**ifzi**", "amarriw" or "ifza", "marrîwet" in Arabic and as "horehound" in English, this plant (**fig II.3**) is widely used as a treatment for several ailments including cancer. [10-13]. Ladanien is the major compound isolated from M.vulgare, It has been reported to have cytotoxic effects on DA1-3b/M2 (IC50 = 10.4 μM), K562 (IC50=25.1 μM), K562R, and 697 (IC50 = 38 μM) cell lines [14].

Figure II.3 Nigella sativa, Marrubium vulgare, and Trigonella fenugreek plants.

II.2 COMPUTATIONAL DETAILS

The theoretical quantum chemical calculations were performed on a computer of a Processor: Intel(R) Core (TM) i7-4500U CPU @ 1.80GHz 2.40 GHz and Installed memory (RAM): 8.00 GB and System type: 64-bit Operating System, with windows 10 platform. by the mean Gaussian 09, ChemDraw, and MOE software program.

II.3 MOLECULE LIBRARY PREPARATION

II.3.1 The 2D and 3D structures

The 2D structures of Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien molecules shown in (**fig II.2**) were drawn by using ChemDraw Ultra 12.0 (**fig II.4**) then converted to 3D and pre-optimized by ChemDraw 3D pro 12.0 software [15]. (**fig II.5**). The 3D ligand structures were saved in MDL format and then imported to the Gaussian09 program to do the optimization and global reactivity calculation.

II.3.2 Optimization and global reactivity calculations

The structures of the compounds involved in the current study were optimized by using the frequency and optimization (opt+frq) parameters in Gaussian 09 program [16] (**figu II.6**). The density functional theory (DFT) method was used in this study to predict the various physicochemical properties. As a result, the compounds were re-optimized at the DFT/B3LYP/6-311G [17]. level of theory by using the Gaussian 09 program.

The calculated parameters were the total energy (E_T), the highest occupied molecular orbital energy (E_{HOMO}), the lowest unoccupied molecular orbital energy (E_{LUMO}), the energy gap ($\Delta E = E_{LUMO} - E_{HOMO}$) the global electrophilicity index ($\omega = \mu^2/2\eta$), the chemical potential ($\mu = [E_{LUMO} + E_{HOMO}]/2$), the chemical hardness ($\mu = [E_{LUMO} - E_{HOMO}]/2$), the chemical softness ($s = 1/2\eta$), and the nucleophilicity ($N = E_{HOMO}$ (Nucleophile) $- E_{HOMO}$ (tetracyanoethylene, TCE)) [18].
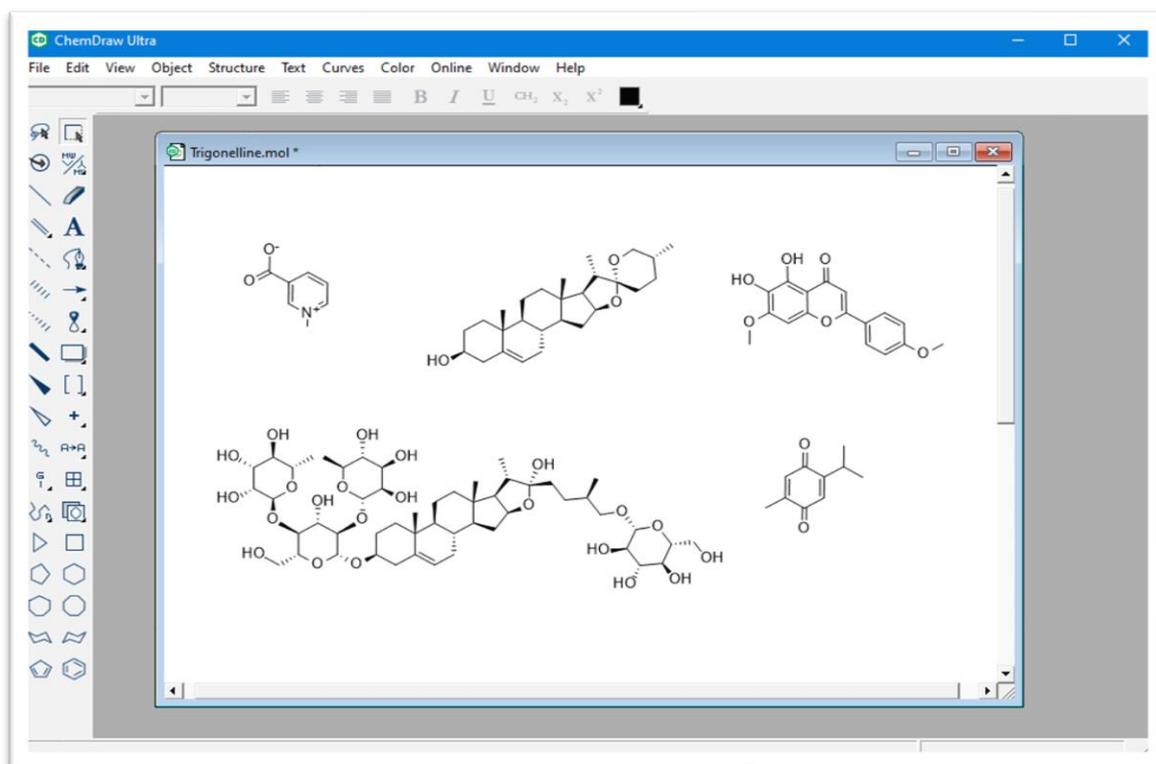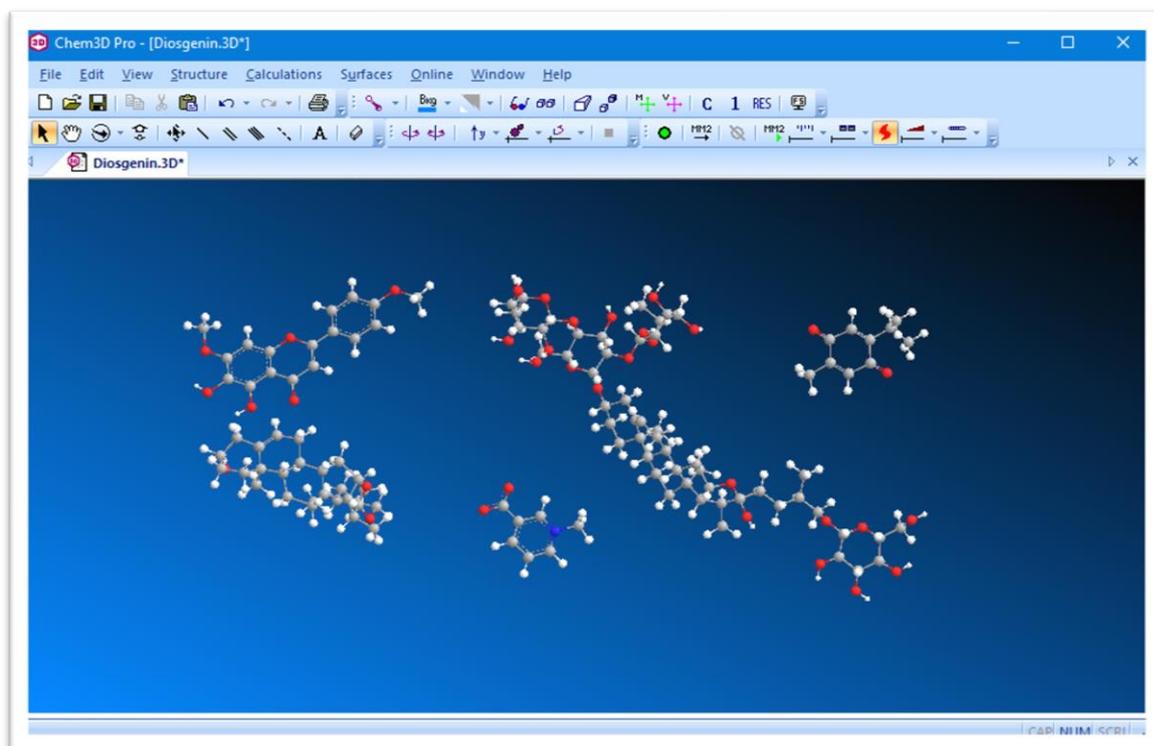
Figure II.4 ChemDraw 2D Ultra

Figure II.5 ChemDraw 3D Pro
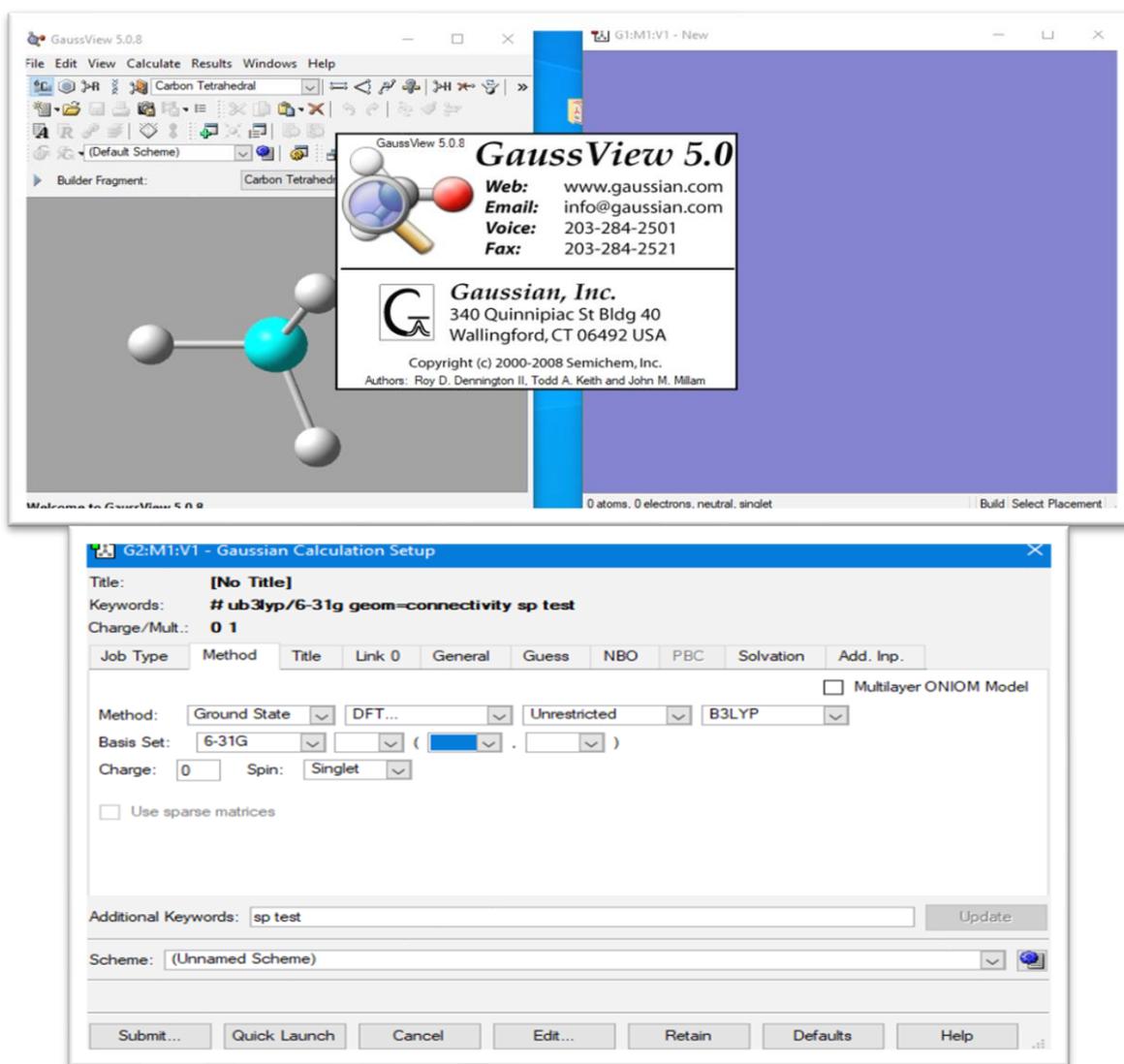
Figure II.6 Gaussian program.

II.3.3 ADME properties

ADMET drug properties, absorption, distribution, metabolism, elimination, and toxicity are critical parameters that are typically used to advance to clinical trials and select drug applicants for development [19]. Lipinski's rule of five, the Physicochemical Properties (Formula, Molecular weight (MW), the number of rotatable bonds, H-bond acceptors, H-bond donors the polar surface area TPSA), the Lipophilicity Properties (iLOGP), and the pharmacokinetic properties (Gl-absorption, BBB permeant, CYP1A2, CYP2C19, CYP2C9, CYP2D6, CYP3A4 inhibitors) were calculated by using SwissADME calculation Tools (**fig II.7**). the AMES, CL, LD50, and $T_{1/2}$ parameters were calculated by using ADMETlab web tools calculations (**fig II.8**), These parameters are used to assess the top-scoring molecule's oral bioavailability in a docking study.

Figure II.7 SwissADME calculation tool.

Figure II.8 ADMETlab calculation tool.

II.4 RECEPTOR PREPARATION

The protein's three-dimensional (3D) crystal structures of EGFR L858R mutant (PDB ID: 2itv) [20] and EGFR T790M mutant (PDB ID: 5hg5) [21] and EGFR WT (PDB ID: 1xkk) [22] **(fig II.10)** were downloaded from the protein data bank (PDB) [23] (**fig II.9**), The Protein Data Bank archive contains thousand protein structures obtained either by crystallography X-ray or by NMR. The proteins were imported into MOE software [24] (**fig II.11**) for visualizing the binding domain of these three complexes and identifying the amino acids in the binding pocket. The enzymes were prepared by removing the cofactors-phosphate ion for "1xkk" and sulfate ion and glycerol for "5hg5". The receptor protein was prepared by leaving water molecules within the active site to ensure the formation of a hydrogen bond between the ligand and the target using MOE software, The missing bonds in the protein structure, which were broken through X-ray diffraction were corrected, and the hydrogen atoms were added. The residues of each protein's active site were found using a site finder and are shown in **Table II.1**. The proteins were subjected to energy minimization by applying the assisted model building and energy refinement (Amber 10): Extended Hückel Theory (EHT) force field.
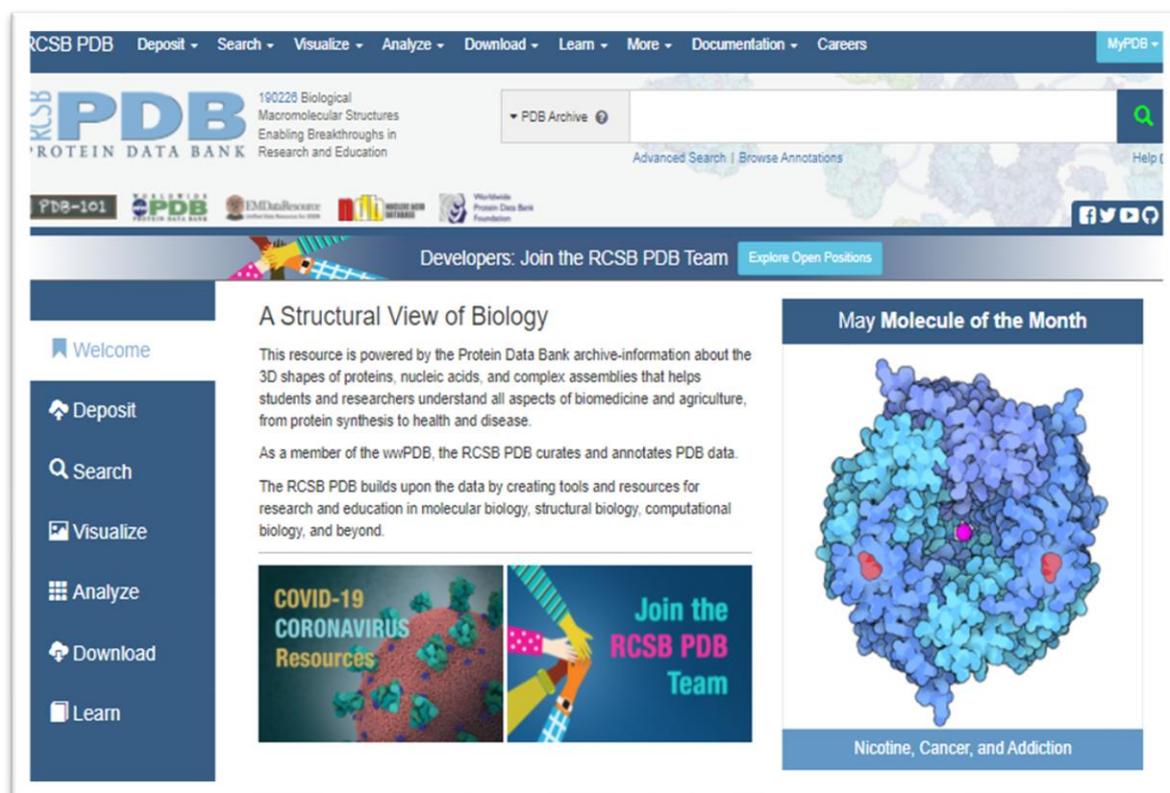
Figure II.9 Protein data bank PDB.
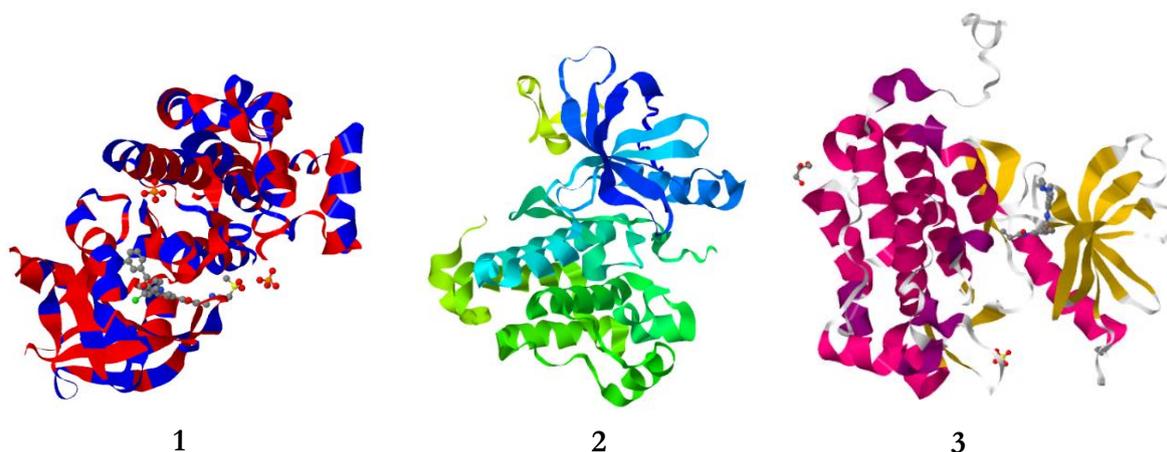
Figure II.10. The Crystal's structure of EGFR kinase domain **1**: (1xkk); **2**: (2itv); **3**: (5hg5).

RECEPTORS	RESIDUES
2itv	1:(GLY696 GLU697 ALA698 PRO699 ASN700 GLN701 ALA702 LEU718 GLY719 SER720 ALA722 PHE723 GLY724 THR725 VAL726 ALA743 ILE744 LYS745 LEU747 ALA755 LYS757 GLU758 ILE759 ASP761 GLU762 TYR764 VAL765 MET766 ALA767 SER768 VAL769 ASP770 CYS775 LEU788 ILE789 THR790 GLN791 LEU792 MET793 PRO794 GLY796 CYS797 ASP800 TYR827 ASP830 ARG831 ARG832 LEU833 ARG836 ASP837 LEU844 THR854 ASP855 PHE856 GLY857 ARG858 ALA859 LYS860 LEU861 ALA864 GLU866 ACE875 VAL876)
1xkk	1:(LEU718 GLY719 SER720 GLY721 VAL726 ALA743 ILE744 LYS745 MET766 CYS775 ARG776 LEU777 LEU788 THR790 GLN791 LEU792 MET793 PHE795 GLY796 CYS797 ASP800 TYR801 GLU804 ARG841 ASN842 LEU844 ILE853 THR854 ASP855 PHE856 LEU858 PHE997 TYR998 LEU1001 MET1002)
5hg5	1:(LEU718 GLY719 SER720 GLY721 ALA722 PHE723 VAL726 LYS728 LYS745 LEU747 ARG748 GLU749 ALA750 SER752 PRO753 LYS754 ILE759 GLU762 ALA763 VAL765 MET766 LEU777 ILE780 SER784 THR785 VAL786 LEU788 LEU792 MET793 PRO794 PHE795 GLY796 CYS797 LEU833 ARG836 ASP837 ARG841 ASN842 LEU844 ASP855 PHE856 GLY857 ARG858 ALA859 LYS860 LEU861 TYR869 ALA871 GLU872 GLY873 GLY874 LYS875 VAL876 TYR891)

Table II.1. Receptor grid construction using binding site residues during Induced Fit Docking

II.5 MOLECULAR DOCKING

The fundamental goal in molecular docking was the capacity to estimate the scoring function and analyze protein-ligand interactions to anticipate the binding affinity and activity of the ligand molecule. Molecular docking and scoring calculations studies are carried out using MOE. The crystal structures of EGFR proteins (PDB ID: 1xkk), (PDB ID: 2itv), and (PDB ID: 5hg5) were at resolutions (2.40 Å), (2.47 Å), and (1.52 Å) respectively, while a resolution between 1.5 and 2.5 Å is considered good for docking studies [25-27]. the type of docking that we used in this study is "Rigid receptor – flexible ligand" Docking, where molecules are allowed to rotate around certain bonds. The best RMSD score is close to 2Å with an energy score of around -7 kcal/mol; these two values are frequently used as a criterion to certify molecular docking results [28-30].
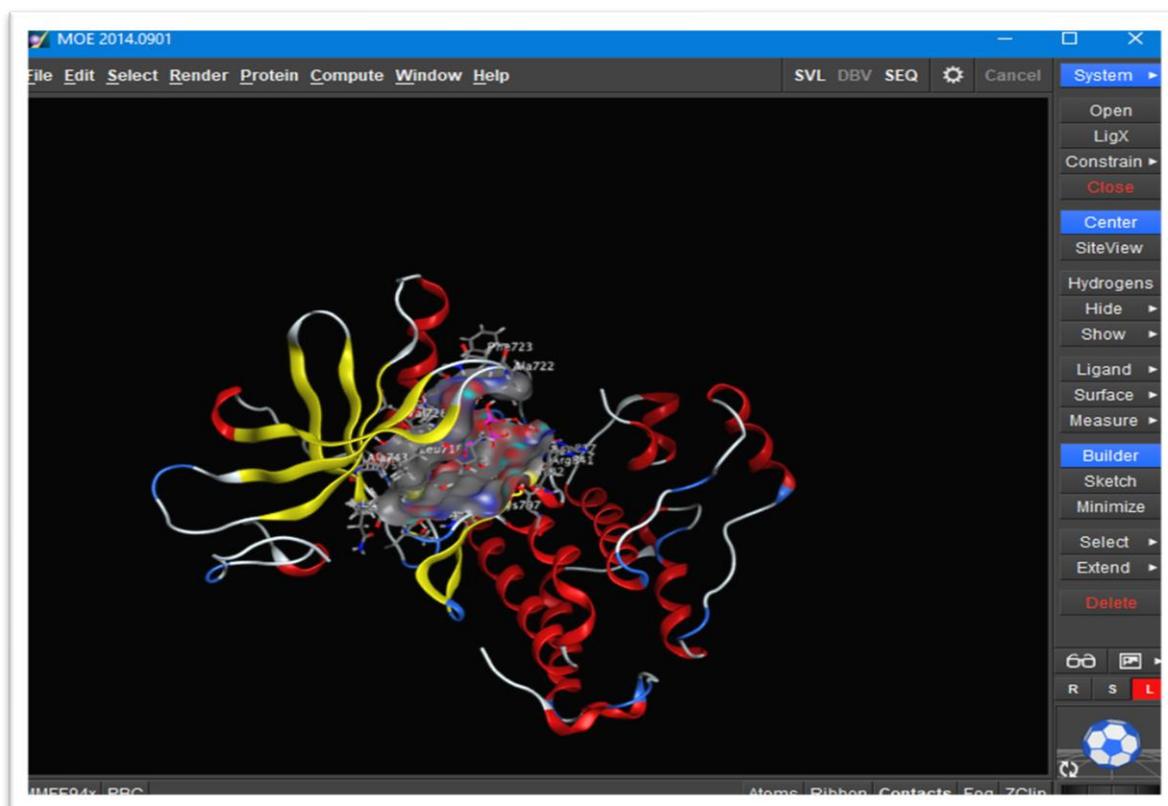
Figure II.11. MOE software

Chapter II References

[1]   A. Chebat *et al.*, "Étude de prévalence des effets indésirables liés à l'utilisation des plantes médicinales par les patients de l'Institut National d'Oncologie, Rabat," *Phytothérapie*, vol. 12, no. 1, pp. 25–32, 2014.

[2]   D.-Q. Wei, Y. Ma, W. C. S. Cho, Q. Xu, and F. Zhou, *Translational bioinformatics and its application*. Springer, 2017.

[3]   J. M. Alves-Silva, A. Romane, T. Efferth, and L. Salgueiro, "North African medicinal plants traditionally used in cancer therapy," *Front. Pharmacol.*, vol. 8, p. 383, 2017.

[4]   M. A. Randhawa and M. S. Al-Ghamdi, "A review of the pharmaco-therapeutic effects of Nigella sativa," *Pak J Med Res*, vol. 41, no. 2, pp. 77–83, 2002.

[5]   U. C. S. Yadav and N. Z. Baquer, "Pharmacological effects of Trigonella foenum-graecum L. in health and disease," *Pharm. Biol.*, vol. 52, no. 2, pp. 243–254, 2014.

[6]   K. El Bairi, M. Ouzir, N. Agnieszka, and L. Khalki, "Anticancer potential of Trigonella foenum graecum: cellular and molecular targets," *Biomed. Pharmacother.*, vol. 90, pp. 479–491, 2017.

[7]   J. Raju, J. M. R. Patlolla, M. V Swamy, and C. V Rao, "Diosgenin, a steroid saponin of Trigonella foenum graecum (Fenugreek), inhibits azoxymethane-induced aberrant crypt foci formation in F344 rats and induces apoptosis in HT-29 human colon cancer cells," *Cancer Epidemiol. Prev. Biomarkers*, vol. 13, no. 8, pp. 1392–1398, 2004.

[8]   S. Shabbeer *et al.*, "Fenugreek: a naturally occurring edible spice as an anticancer agent," *Cancer Biol. Ther.*, vol. 8, no. 3, pp. 272–278, 2009.

[9]   H. Hibasami *et al.*, "Protodioscin isolated from fenugreek (Trigonella foenumgraecum L.) induces cell death and morphological change indicative of apoptosis in leukemic cell line H-60, but not in gastric cancer cell line KATO III," *Int. J. Mol. Med.*, vol. 11, no. 1, pp. 23–26, 2003.

[10]   A. Tahraoui, J. El-Hilaly, Z. H. Israili, and B. Lyoussi, "Ethnopharmacological survey of plants used in the traditional treatment of hypertension and diabetes in south-eastern Morocco (Errachidia province)," *J. Ethnopharmacol.*, vol. 110, no. 1, pp. 105–117, 2007.

[11]   I. Teixidor-Toneu, G. J. Martin, A. Ouhammou, R. K. Puri, and J. A. Hawkins, "An ethnomedicinal survey of a Tashelhit-speaking community in the High Atlas, Morocco," *J. Ethnopharmacol.*, vol. 188, pp. 96–110, 2016.

[12] A. Merzouki, F. Ed-Derfoufi, and J. M. Mesa, "Contribution to the knowledge of Rifian traditional medicine. II: Folk medicine in Ksar Lakbir district (NW Morocco)," *Fitoterapia*, vol. 71, no. 3, pp. 278–307, 2000.

[13] F. Kabbaj, B. Meddah, Y. Cherrah, and E. Faouzi, "Ethnopharmacological profile of traditional plants used in Morocco by cancer patients as herbal therapeutics," *Phytopharmacology*, vol. 2, no. 2, pp. 243–256, 2012.

[14] R. Alkhatib *et al.*, "Activity of ladanein on leukemia cell lines and its occurrence in Marrubium vulgare," *Planta Med.*, vol. 76, no. 01, pp. 86–87, 2010.

[15] K. R. Cousins, "Computer review of ChemDraw ultra 12.0." ACS Publications, 2011.

[16] H. Khelfaoui, D. Harkati, and B. A. Saleh, "Molecular docking, molecular dynamics simulations and reactivity, studies on approved drugs library targeting ACE2 and SARS-CoV-2 binding with ACE2," *J. Biomol. Struct. Dyn.*, vol. 39, no. 18, pp. 7246–7262, 2021.

[17] A. D. Becke, "Density-functional thermochemistry. I. The effect of the exchange-only gradient correction," *J. Chem. Phys.*, vol. 96, no. 3, pp. 2155–2160, 1992.

[18] H. A. Abdulhassan, B. A. Saleh, D. Harkati, H. Khelfaoui, N. L. Hewitt, and G. A. El-Hiti, "In Silico Pesticide Discovery for New Anti-Tobacco Mosaic Virus Agents: Reactivity, Molecular Docking, and Molecular Dynamics Simulations," *Appl. Sci.*, vol. 12, no. 6, p. 2818, 2022.

[19] A. Zekri, D. Harkati, S. Kenouche, and B. A. Saleh, "QSAR modeling, docking, ADME and reactivity of indazole derivatives as antagonizes of estrogen receptor alpha (ER-α) positive in breast cancer," *J. Mol. Struct.*, vol. 1217, p. 128442, 2020.

[20] C.-H. Yun *et al.*, "Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity," *Cancer Cell*, vol. 11, no. 3, pp. 217–227, 2007.

[21] H. Cheng *et al.*, "Discovery of 1-{(3 R, 4 R)-3-[({5-Chloro-2-[(1-methyl-1 H-pyrazol-4-yl) amino]-7 H-pyrrolo [2, 3-d] pyrimidin-4-yl} oxy) methyl]-4-methoxypyrrolidin-1-yl} prop-2-en-1-one (PF-06459988), a Potent, WT Sparing, Irreversible Inhibitor of T790M-Containing EGFR Mutants," *J. Med. Chem.*, vol. 59, no. 5, pp. 2005–2024, 2016.

[22] E. R. Wood *et al.*, "A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells," *Cancer Res.*, vol. 64, no. 18, pp. 6652–6659, 2004.

[23] H. M. Berman *et al.*, "The protein data bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000.

[24] A. Furuhama, T. I. Hayashi, and N. Tatarazako, "Acute to chronic estimation of Daphnia magna toxicity within the QSAAR framework," *SAR QSAR Environ. Res.*, vol. 27, no. 10, pp. 833–850, 2016.

[25] C. Venugopal, C. M. Demos, K. S. Jagannatha Rao, M. A. Pappolla, and K. Sambamurti, "Beta-secretase: structure, function, and evolution," *CNS Neurol. Disord. Targets (Formerly Curr. Drug Targets-CNS Neurol. Disord.*, vol. 7, no. 3, pp. 278–294, 2008.

[26] A. Lesk, *Introduction to protein science: architecture, function, and genomics*. Oxford university press, 2010.

[27] C. W. Murray, C. A. Baxter, and A. D. Frenkel, "The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase," *J. Comput. Aided. Mol. Des.*, vol. 13, no. 6, pp. 547–562, 1999.

[28] T. C. Ramalho, M. S. Caetano, E. F. F. da Cunha, T. C. S. Souza, and M. V. J. Rocha, "Construction and assessment of reaction models of class I EPSP synthase: molecular docking and density functional theoretical calculations," *J. Biomol. Struct. Dyn.*, vol. 27, no. 2, pp. 195–207, 2009.

[29] E. Kellenberger, J. Rodrigo, P. Muller, and D. Rognan, "Comparative evaluation of eight docking tools for docking and virtual screening accuracy," *Proteins Struct. Funct. Bioinforma.*, vol. 57, no. 2, pp. 225–242, 2004.

[30] C. P. Vianna and W. F. de Azevedo, "Identification of new potential Mycobacterium tuberculosis shikimate kinase inhibitors through molecular docking simulations," *J. Mol. Model.*, vol. 18, no. 2, pp. 755–764, 2012.

Chapter III

Results and discussion

Chapter III.   Results and discussion

III.1 THE OPTIMIZATION AND GEOMETRIC STRUCTURE

The Optimized structures and numbering of atoms of Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien molecules are shown graphically in the **fig III.1** obtained at B3LYP/6-31G DFT method. **Table III.1** illustrates their geometrical parameters such as the calculated total energies, the RMS, and the dipole moments. The global minimum energies are found to be (-14655.37 eV), (-12952.10 eV), and (-35835.66 eV) for Thymoquinone, Trigonelline, Ladanien, and (-35885.49 eV), (-99229.05 eV) for Diosgenin and Protodioscin respectively. The RMS Cartesian force values are equal to $1.026.10^{-5}$, $3.096.10^{-5}$, $2.02.10^{-6}$ in Thymoquinone, Trigonelline, and Ladanien, while it equal to 0.00 in Diosgenin and Protodioscin. A molecule's dipole moment is represented by a three-dimensional vector that reflects the molecular charge distribution. As a result, it can be used as a descriptor to represent charge movement within a molecule. As a result of DFT/B3LYP/6-31G calculations, the highest dipole moment was observed for the Trigonelline (~14.4811Debye) whereas the smallest one was observed for the Thymoquinone (~0.1732 Debye).

Table III.1 Calculated total energies (E), RMS Cartesian force, and dipole moments

B3LYP/6-31G			
Compd	E_T (eV)	RMS Cartesian force	dipole moments (D)
Thymoquinone	-14655.37	$1.026.10^{-5}$	0.1732
Trigonelline	-12952.10	$3.096.10^{-5}$	14.4811
Ladanein	-35835.66	$2.02.10^{-6}$	6.8578
Diosgenin	-35885.49	0.00	1.4549
Protodioscin	-99229.05	0.00	5.5381
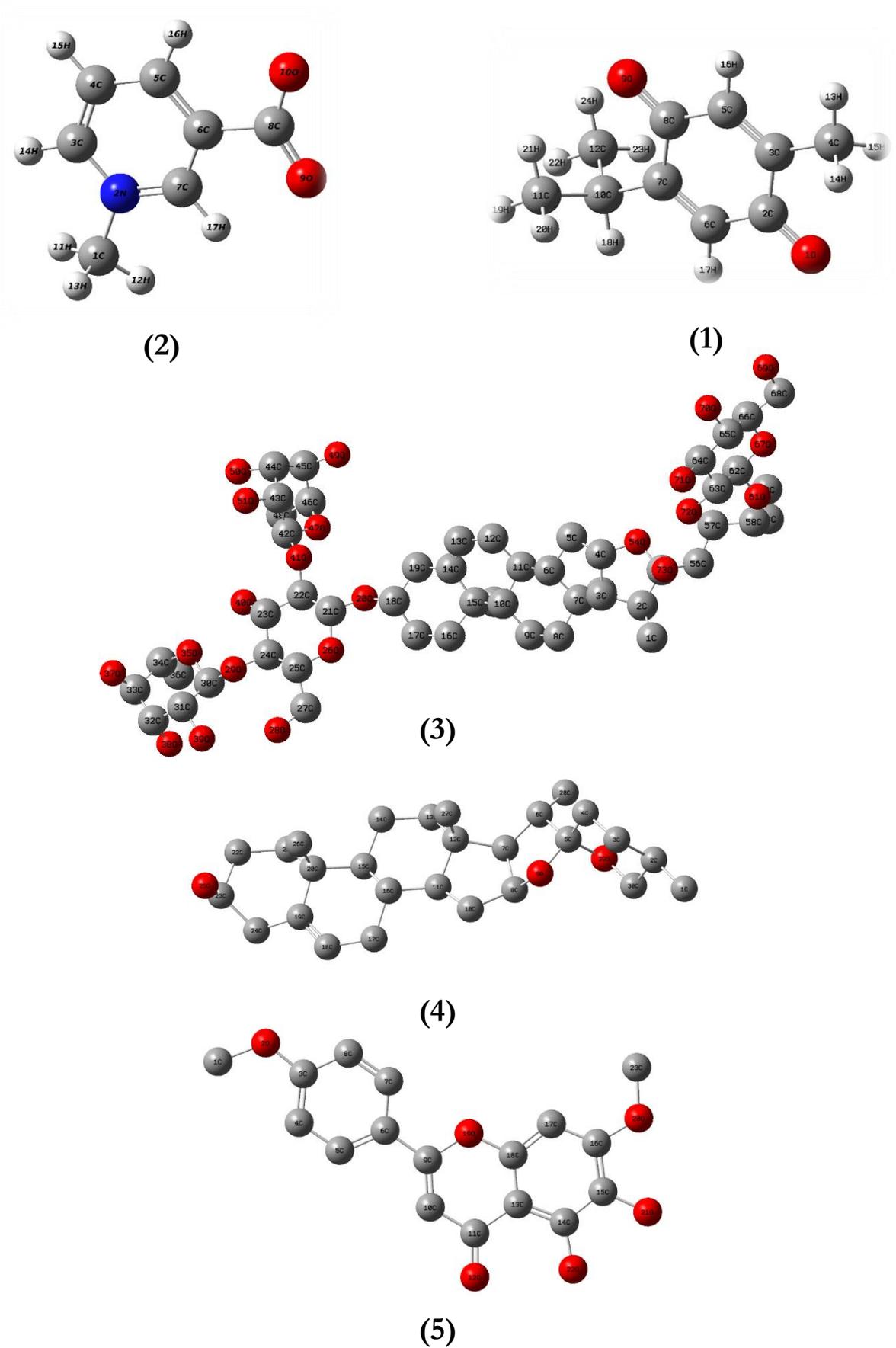
Figure III.1 Optimized structure of Thymoquinone (1), Trigonelline (2), Protodioscin (3), Diosgenin (4), and Ladanein (5).

III.2 ELECTRONIC PROPERTIES

III.2.1 Frontier Molecular Orbitals (FMO)

The energies of the frontier orbitals HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital), also known as frontier molecular orbitals (FMOs), play an important role in molecular systems, they reflect the reactivity of a molecule. LUMO is directly related to electron affinity, while HOMO is related to ionization potential. Higher HOMO energy indicates a molecule's affinity to react as a nucleophile, whereas a lower LUMO energy indicates a molecule's electrophilic nature.[1] These orbitals help in understanding the chemical stability and reactivity of molecules, as well as how the molecules interact with other species. [2,3], The energy gap between the HOMO and LUMO is important in influencing a molecule's electrical characteristics, kinetic stability, optical polarizability, and chemical reactivity descriptors including hardness and softness. The FMOs in the electronic transitions and their energy differences ΔE are determined to predict the energetic behaviors and reactivity of the Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien molecules against cancer. the molecular chemical reactivity is defined as a gap between two energy states. The energies of the four important FMOs (HOMO, HOMO-1, LUMO, and LUMO+1) were computed using the DFT method with the B3LYP/6-31G bases set, [4] Their 3D plots are illustrated in **figs III.2-6:**
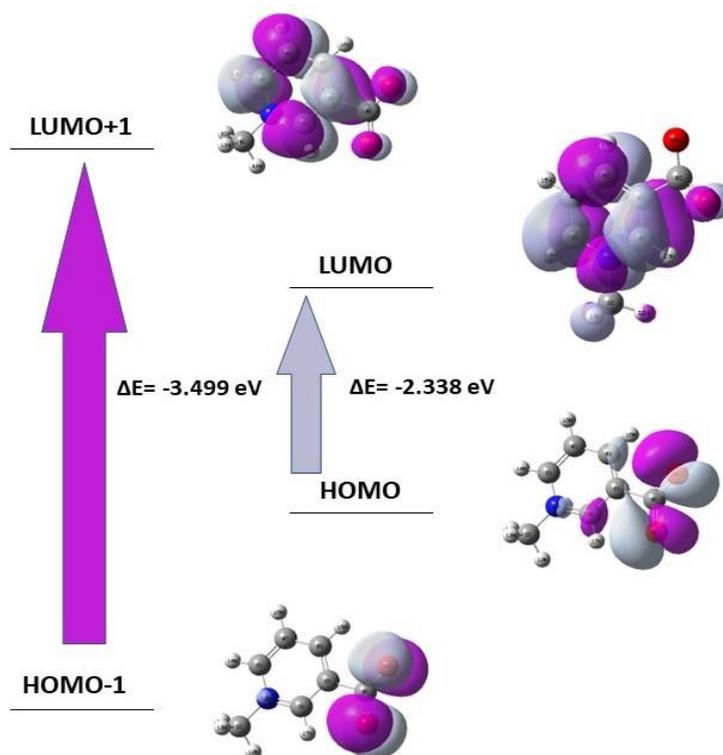
Figure III.2 Frontier molecular orbitals of trigonelline.
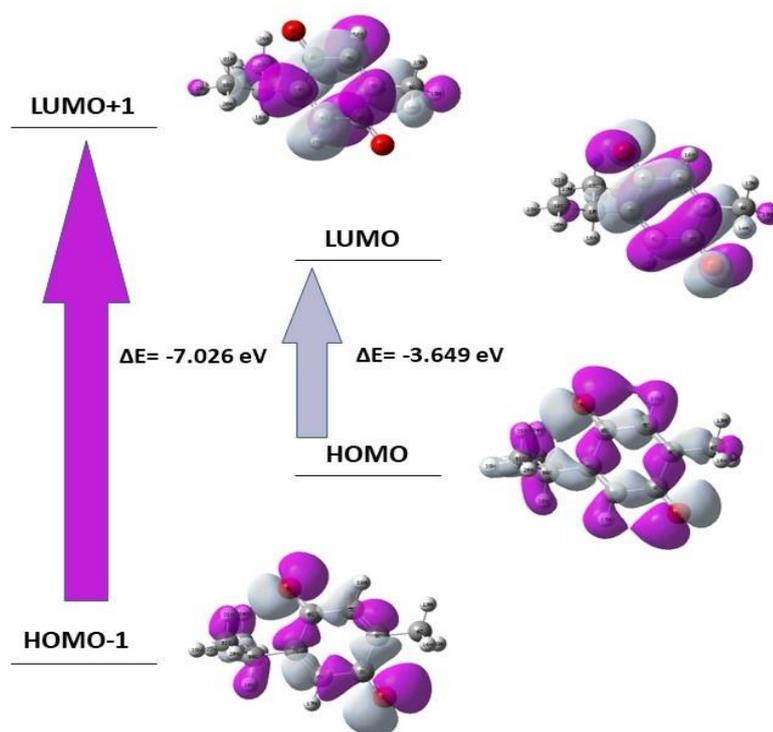
Figure III.3 Frontier molecular orbitals of Thymoquinone
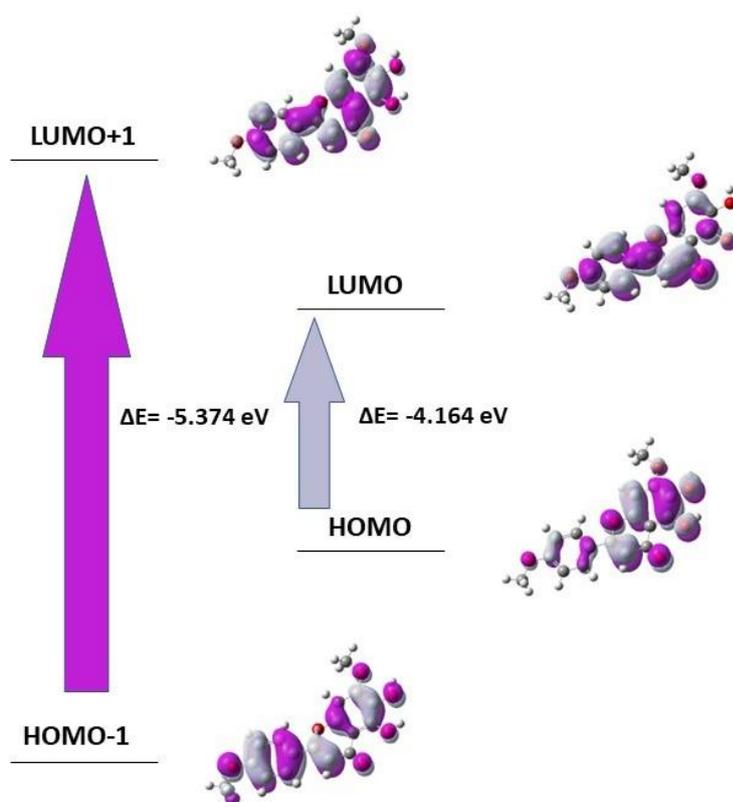
Figure III.4 Frontier molecular orbitals of ladanein.
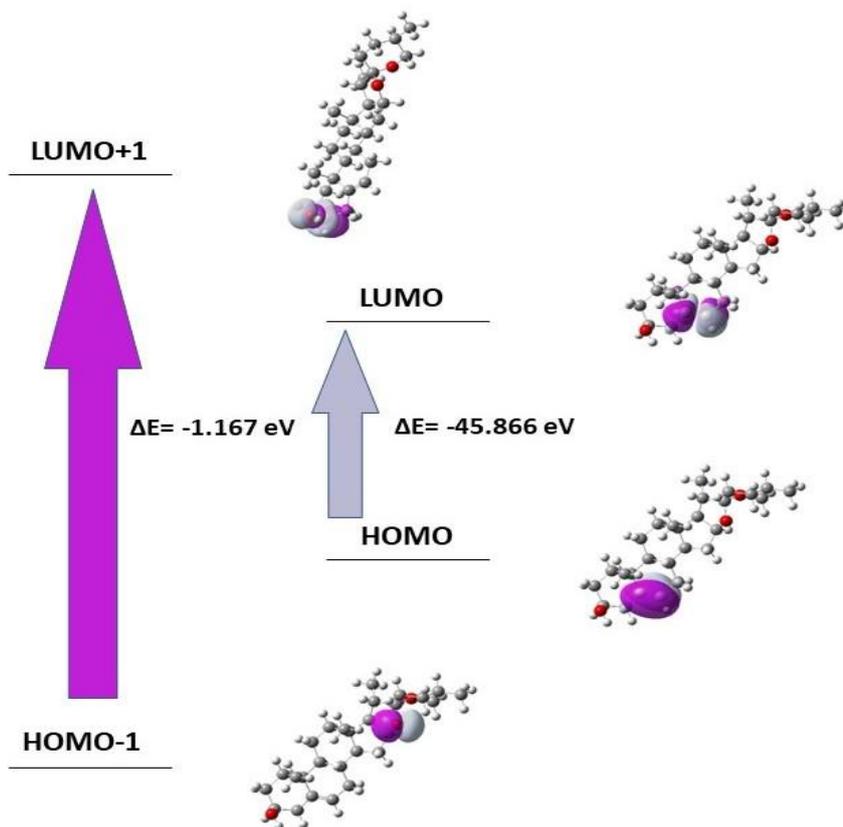
Figure III.5 Frontier molecular orbitals of diosgenin.
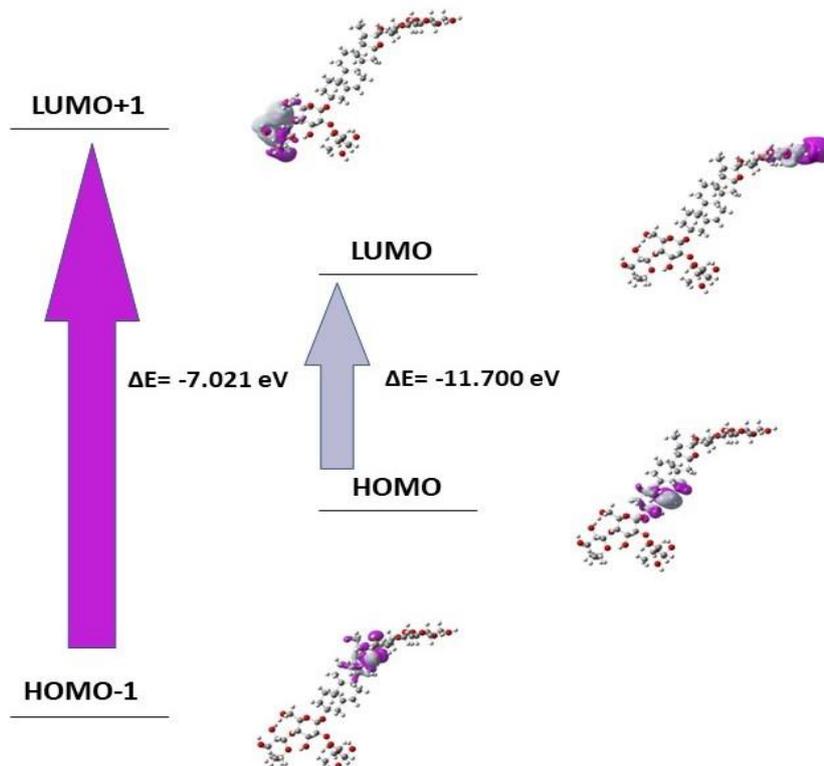
Figure III.6 Frontier molecular orbitals of Protodioscin.

The results of the FMOs energy analysis revealed that the energies levels of the HOMO – LUMO gap orbitals (ΔE) of Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien compounds are about (-3.649 eV), (-45.866 eV), (-11.700 eV), (-2.338 eV) and (-4.164 eV) respectively, while the energy gap between HOMO-1 and LUMO+1 for these compounds are closed to (-7.026 eV), (-1.167 eV), (-7.021 eV), (-3.499 eV), and (-5.374 eV) respectively.

It is clear from the figures of FMOs orbitals that the white-colored lobes represent the negative phase (nucleophilic behavior) while the purple-colored lobes correspond to the positive phase (electrophilic behavior).

The level of the energy gaps is different for all investigated compounds. Consequently, the energy gap (ΔE) of compounds Trigonelline and Thymoquinone, and Ladanien are higher compared with the other compounds Protodioscin and Diosgenin. the energy gap of studied compounds is in the order of Trigonelline **>** Thymoquinone **>** Ladanien **>** Protodioscin **>** Diosgenin. The HOMO-LUMO gap energy of the Diosgenin is equal to **-45.866** eV. This low energy value promotes the transfer of electrons in the Diosgenin molecule. It is noticed that the energy gap of the Diosgenin and Protodioscin compounds is smaller which makes these molecules soft and more reactive compared to the other molecules.

Interestingly, the Diosgenin compound with the energy gap ΔE = - 45866 eV are several hydrophilic interactions that could facilitate the binding with the receptors. This suggests that such hydrophilic interactions considerably impact the binding affinity of such small drugs to the receptors.

III.2.2 Molecular Electrostatic Potential (MEP)

The molecular electrostatic potential (MEP) is the potential energy of a proton at a particular location near a molecule, the MEP has been discovered to be a valuable descriptor for predicting electrophilic and nucleophilic attack reactions, as well as hydrogen bonding interactions.

the MEP is an electrostatic potential plot mapped onto a constant electron density surface that is used primarily for predicting sites and relative reactivity to electrophilic and nucleophilic attacks, in studies of biological recognition and interactions between the same molecules or other molecules and the correlation and prediction of a wide range of macroscopic properties.

the molecular electrostatic potential must be determined to validate the evidence of the molecule's responsiveness as an inhibitor, Although the MEP indicates the molecular size and shape of the positive, negative, and neutral electrostatic potentials in terms of color grading and is an indicator in the research of molecular structure properties **[5,6].** The red color represents the most electronegative electrostatic potential (high density of electrons) while the blue color indicates the most electropositive potential (minimal concentration of electrons) and the Regions where the potentials are zero are denoted by the green color.

The molecular electrostatic potential of the studied molecules is calculated by the same method under the same base sets and the 3D plots are illustrated in **figs III.7-11.**
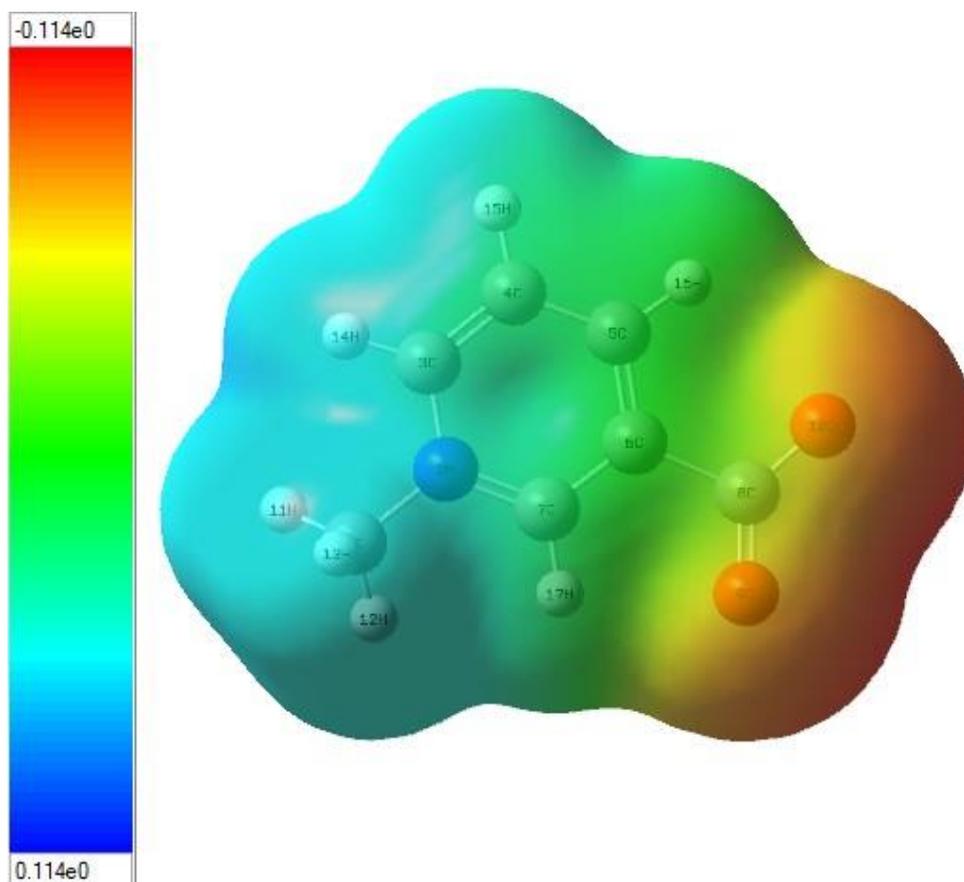
Figure III.7 Molecular electrostatic potential surface of trigonelline.

Figure III.8 Molecular electrostatic potential surface of Thymoquinone.

Figure III.9 Molecular electrostatic potential surface of Ladanein.
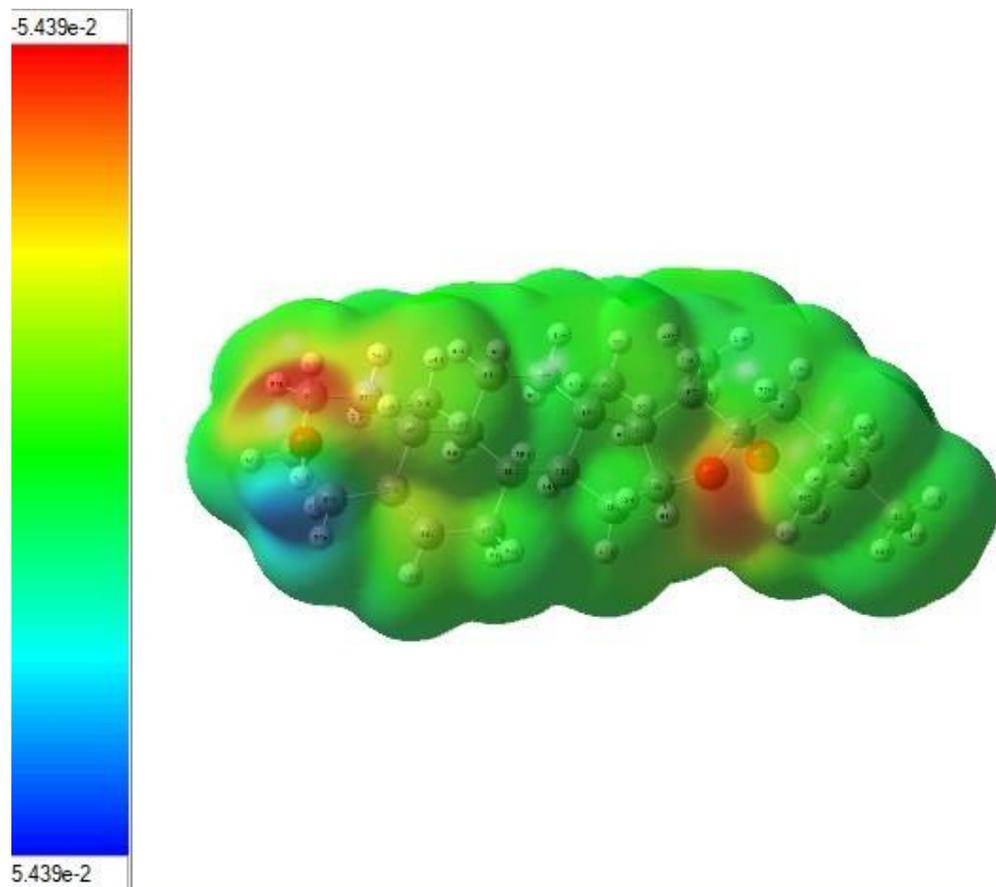
Figure III.10 Molecular electrostatic potential surface of Diosgenin.
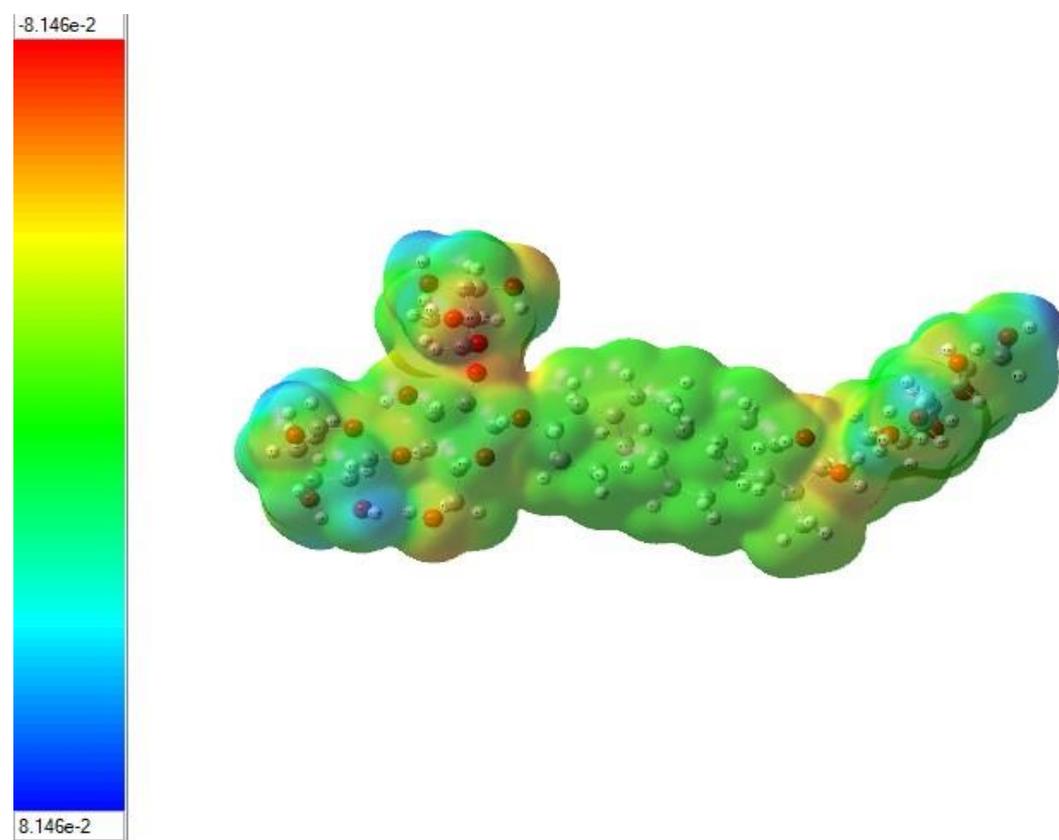
Figure III.11 Molecular electrostatic potential surface of Protodioscin.

As a result, MEP surfaces varies between -0.114 a.u (deepest red) to 0.114 a.u (deepest blue) for the trigonelline compound and between $-4.984 \ 0.10^{-2}$ a.u (deepest red) to $4.984 \ 0.10^{-2}$ a.u (deepest blue)) for the Thymoquinone compound, As for the third compound Ladanein the MEP surface vary is between -0.102 a.u deepest red and 0.102 a.u deepest blue, and between $-5.439 \ 0.10^{-2}$ a.u to $5.439 \ 0.10^{-2}$ a.u, and $-8.146 \ 0.10^{-2}$ a.u and $8146 \ 0.20^{-2}$ a.u for the Diosgenin and Protodioscin molecules respectively.

As can be seen, the MEPs for trigonelline molecule (**fig III.7**) indicated regions with the maximum positive potential (red color) localized on the oxygen atoms O_{10} and O_9 atom that refers to a negative potential (electrophilic attack), in addition to that, another region characterized by a blue color around the three atoms of hydrogen H_{11}, H_{14}, and H_{15} that refers to a positive potential (nucleophilic attack). the green color on this molecule is referred to as a neutral electrostatic potential (The zero potential sites).

In (**fig III.8**) it is clear that the MEP surface of the thymoquinone molecule shows a region characterized by a red color around the oxygen atoms O_1 and O_9 which indicates a possible site for an electrophilic attack, and another region characterized by a blue color around the oxygen atoms H_{13} and H_{16} that indicates the minimal concentration of electrons, while the potentials are zero are denoted by the green color.

The MEP surface of the third molecule Ladanein (**fig III.9**) shows a region of the most electronegative electrostatic potential characterized by a red color around the oxygen atoms O_{22} and O_{12}, while the most electropositive potential region is around the hydrogen atoms H_{32}, H_{33}, H_{35}, H_{36}, and H_{37} and the carbon atom C_{23}.

In the figures (**fig III.10** and **III.11**) of the other molecules Diosgenin and protodioscin, the positive potential (blue and light blue) sites are found around the region of hydrogen atoms H_{61} and H_{59} for diosgenin and around H_{15}, H_{16}, H_{17}, H_{28}, and H_{153} For protodioscin molecule. While the region of red color is around the oxygen atoms O_{90} and O_{25} for diosgenin and around O_{41}, O_{49}, O_{59}, O_{28}, and O_{73} for the other molecule. That is, atoms in this region tend to attract electrons (electrophilic).

III.3 REACTIVITY ANALYSIS

III.3.1 Global Reactivity Descriptors

The density functional theory (DFT) defines many important concepts of chemical reactivity by utilizing the electron density of the chemical system under consideration, The analysis of density functional theory descriptors provides more information about the compound's stability, electrophilicity, and nucleophilicity.[8] The E_{HOMO} and E_{LUMO} are indicators for predicting a molecule's ionization potential ($I= -E_{HOMO}$) and electron affinity ($A=-E_{LUMO}$) where I is the ionization potential and A is the electron affinity of the molecule. The concept of hardness (η) and softness (s) is related to a compound's reactivity and is a property that measures the extent of chemical reactivity to which the addition of a charge stabilizes the system. The chemical potential (μ) provides a global reactivity index and is related to charge transfer from a higher chemical potential system to a lower chemical potential system, Electronegativity (χ) is the ability to attract electrons and is directly related to all of the properties mentioned previously.[9] All these properties are calculated according to the following equations [10,11]:

$$\mu = \frac{-(I + A)}{2}$$

$$\eta = \frac{(I - A)}{2}$$

$$\chi = \frac{(I + A)}{2}$$

A large HOMO-LUMO gap indicates a hard molecule and is associated with more stable molecules, whereas a small gap indicates a soft molecule and is associated with a more reactive molecule. The electrophilicity index (ω), which measures the energy loss due to charge transfer, is another important descriptor, this is how the electrophilicity index is defined [12-14]:

$$\omega = \frac{(\mu 2)}{2\eta}$$

the global reactivity descriptors were calculated for the 5 compounds using the DFT method with the B3LYP/6-31G level. The results obtained are summarized in Table III.2.

Table III.2 The Global Reactivity Descriptors (HOMO, LUMO, Δ**E, s, χ, ω, μ, η**) of the 5

Compd	HOMO (eV)	LUMO (eV)	ΔE (eV)	η (eV)	S (eV)	μ (eV)	ω (eV)	χ (eV)
Thymoquinone	-7.19	-3.55	3.64	1.82	0.27	-5.37	7.90	5.37
Trigonelline	-5.13	-2.79	2.34	1.17	0.43	-3.96	6.71	3.96
Ladanein	-5.72	-1.55	4.16	2.08	0.24	-3.63	3.17	3.63
Diosgenin	-6.07	0.82	6.89	3.45	0.14	-2.62	1.00	2.62
Protodioscin	-5.97	0.57	6.54	3.27	0.15	-2.69	1.11	2.69

Using the energies of FMOs, we calculated the reactivity descriptors of Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien molecules. The results of The Global Reactivity Descriptors energy analysis revealed that the energies levels of the HOMOs orbitals of the studied compounds are about (-7.19 eV), (-5.13 eV), (-5.72 eV), (-6.07 eV), and (-5.97 eV) for Thymoquinone, Trigonelline, Ladanein, Diosgenin, and protodioscin respectively. While the LUMOs energies are about (-3.55 eV), (-2.79 eV), and (-1.55 eV) for the three molecules Thymoquinone, Trigonelline, Ladanein, and about (0.82 eV), (0.57 eV) for Diosgenin, and protodioscin respectively. The energies of E_{HOMO} and E_{LUMO} and their neighboring orbitals for the three molecules Thymoquinone, Trigonelline, and Ladanein, are all negative, which indicates that all these molecules are stable.

The values reported in **table III.2** show that the chemical hardness (η) is found to be 1.82, 1.17, 2.08, 3.45, and 3.27 eV for Thymoquinone, Trigonelline, Ladanein, Diosgenin, and protodioscin respectively. The small value of hardness (η) (1.17 eV) reflects high polarizability showing Trigonelline as a soft molecule with faster reactions since the electrons are further from the nucleus. The chemical potential (μ) value of trigonelline (-3.96 eV) and thymoquinone (-5.37 eV) and the Electronegativity (3.96 eV) and (5.37 eV) for these two molecules indicate that these molecules have a significative attractive electron power.

The global softness (*S*) characterizes the ability of a molecule to accept electrons, the chemical softness values in the studied compounds are similar with a small significant change, it has been computed and found to be 0.27, 0.43, 0.24, 0.14, and 0.15 eV for the five studied compounds respectively. Moreover, the electrophilicity index (ω) is about 7.90 eV for Thymoquinone, 6.71 eV for Trigonelline, and 3.17 eV for Ladanein, based on the value found in the electrophilicity index, we can conclude that the Thymoquinone is a good electrophile better than the other molecules.

Therefore, it can accept an electron doublet to form bonds with another reagent which is necessarily a nucleophile.

III.3.2 Local Reactivity Descriptors

The concepts of local and global reactivity descriptors have been widely used to understand the chemical reactivity and site selectivity, to analyze molecular site selectivity, Parr and Yang [15] define local descriptors such as Fukui functions. Thus, calculating Fukui functions can enable us to determine the active sites of a molecule, based on the electronic density changes experienced by the molecule during a reaction. Fukui functions $f^+(r)$, $f^-(r)$ and $f^0(r)$ are calculated for three chemical situations, using the following equations [16-18]:

$$f^-(r) = q_k(N) - q_k(N - 1), \text{ for electrophilic attack.}$$

$$f^+(r) = q_k(N+ 1) - q_k(N), \text{ for nucleophilic attack.}$$

$$f^0(r) = \frac{1}{2}\left[q_k(N+1) - q_k(N-1)\right], \text{ for Radical attack.}$$

where $q_k(N)$ is the atomic population on the k_{th} atom for the neutral molecule, while $q_k(N+1)$ and $q_k(N-1)$ are the atomic populations on the k_{th} atom for its anionic and cationic species, respectively. In addition to the information concerning the electrophilic and nucleophilic capacity of a given atomic site in the molecule, there is another Dual descriptor $(\Delta f(r))$ which is given by [19]:

$$\Delta f(r) = f^+(r) - f^-(r)$$

where $\Delta f(r)$ is defined as the difference between the nucleophilic and electrophilic Fukui function. Two situations need to be considered: if $\Delta f(r) > 0$, then the site is favored for a nucleophilic attack, whereas if $\Delta f(r) < 0$, then the site may be favored for an electrophilic attack.

The Fukui functions indices and Dual descriptor of the five molecules (Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien) were calculated at the DFT level of theory with the basis set (B3LYP/6-31G) and they are presented in Tables III.3 to III.7:

Table III.3 Values of the Fukui functions and Dual descriptor of Thymoquinone.

Atoms	N	N-1	N+1	f⁻	f⁺	f⁰	Δf
1 O	-0.409	-0.558	-0.291	0.148	0.118	0.267	-0.030
2 C	0.254	0.214	0.259	0.039	0.005	0.044	-0.034
3 C	0.113	0.092	0.143	0.021	0.029	0.050	0.008
4 C	-0.460	-0.454	-0.468	-0.005	-0.007	-0.013	-0.002
5 C	-0.154	-0.194	-0.126	0.039	0.027	0.067	-0.012
6 C	-0.183	-0.225	-0.132	0.042	0.050	0.093	0.008
7 C	0.149	0.119	0.178	0.030	0.028	0.058	-0.002
8 C	0.251	0.210	0.262	0.040	0.010	0.051	-0.029
9 O	-0.415	-0.564	-0.188	0.149	0.227	0.376	0.078
10 C	-0.166	-0.139	-0.212	-0.026	-0.046	-0.072	-0.020
11 C	-0.409	-0.398	-0.417	-0.010	-0.008	-0.019	0.002
12 C	-0.409	-0.398	-0.417	-0.010	-0.008	-0.019	0.002

It can be concluded from the analysis of the results in **table III.3** that two carboxylic functions **O1 and O9** have the largest F+, F- and F0 values and will be the preferred site for the nucleophilic attack, electrophilic attack, and radical attack respectively.

The table shows that positive **Δf** values are determined for the **C3, C6, O9, C11, and C12** atoms subject to nucleophilic attacks. On the other hand, **O1, C2, C4, C5, C7, C8, and C10** have negative **Δf** values This reveals that atoms are highly reactive concerning electrophilic attacks.

Table III.4 Values of the Fukui functions and Dual descriptor
of Trigonelline.

Atoms	N	N-1	N+1	f^-	f^+	f^0	Δf
1 C	-0.262	-0.225	-0.279	-0.036	-0.017	-0.054	0.018
2 N	-0.568	-0.582	-0.591	0.014	-0.023	-0.009	-0.037
3 C	0.150	0.069	0.192	0.081	0.042	0.123	-0.039
4 C	-0.155	-0.149	-0.162	-0.005	-0.007	-0.013	-0.002
5 C	-0.082	-0.178	-0.016	0.095	0.066	0.161	-0.029
6 C	-0.021	-0.003	-0.024	-0.017	-0.003	-0.020	0.014
7 C	0.139	0.050	0.230	0.089	0.091	0.180	0.001
8 C	0.455	0.397	0.437	0.058	-0.018	0.040	-0.076
9 O	-0.541	-0.600	-0.263	0.059	0.278	0.337	0.219
10 O	-0.524	-0.584	-0.240	0.059	0.284	0.343	0.224
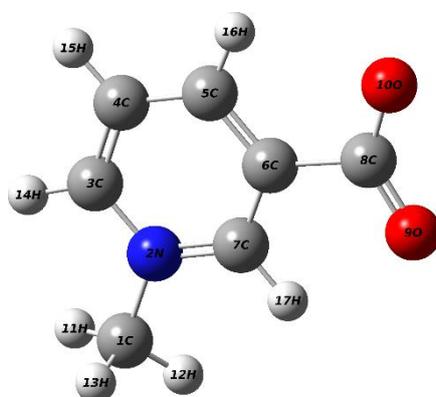
It can be concluded from the analysis of the results in **table III.4** that **O10 and O9** have the largest F+ and F0 values and will be the preferred site for the nucleophilic attack, and radicals attack respectively. While **C5** and **C7** have the largest F⁻ value will be the preferred site for electrophilic attack.

The table shows that positive **Δf** values are determined for the **C1, C6, C7, O9, and O10** atoms subject to nucleophilic attacks. On the other hand, **N2, C3, C4, C5, and C8** have negative Δf values This reveals that atoms are highly reactive concerning electrophilic attacks.

Table III.5 Values of the Fukui functions and Dual descriptor of Ladanien.

Atoms	N	N-1	N+1	f⁻	f⁺	f⁰	Δf
1 C	-0.173	-0.149	-0.204	-0.024	-0.031	-0.055	-0.007
2 O	-0.558	-0.582	-0.500	0.023	0.058	0.082	0.035
3 C	0.298	0.254	0.309	0.043	0.011	0.055	-0.032
4 C	-0.149	-0.158	-0.122	0.009	0.027	0.036	0.018
5 C	-0.133	-0.164	-0.115	0.031	0.018	0.049	-0.013
6 C	0.037	0.040	0.051	-0.003	0.014	0.011	0.017
7 C	-0.135	-0.164	-0.121	0.028	0.014	0.043	-0.014
8 C	-0.133	-0.142	-0.108	0.009	0.024	0.034	0.015
9 C	0.283	0.227	0.288	0.055	0.004	0.060	-0.051
10 C	-0.161	-0.186	-0.135	0.025	0.025	0.051	0
11 C	0.246	0.196	0.257	0.049	0.010	0.060	-0.039
12 O	-0.428	-0.529	-0.380	0.101	0.048	0.149	-0.053
13 C	0.138	0.142	0.155	-0.004	0.017	0.013	0.021
14 C	0.203	0.164	0.223	0.038	0.020	0.059	-0.018
15 C	0.216	0.197	0.251	0.019	0.034	0.054	0.015
16 C	0.312	0.281	0.342	0.030	0.029	0.060	-0.001
17 C	-0.131	-0.159	-0.102	0.027	0.029	0.056	0.002
18 C	0.212	0.211	0.234	0.001	0.022	0.023	0.021
19 O	-0.612	-0.637	-0.597	0.024	0.014	0.039	-0.01
20 O	-0.601	-0.619	-0.576	0.017	0.025	0.043	0.008
21 O	-0.658	-0.682	-0.587	0.024	0.071	0.095	0.047
22 O	-0.576	-0.593	-0.523	0.016	0.052	0.069	0.036
23 C	-0.167	-0.149	-0.188	-0.017	-0.021	-0.039	-0.004

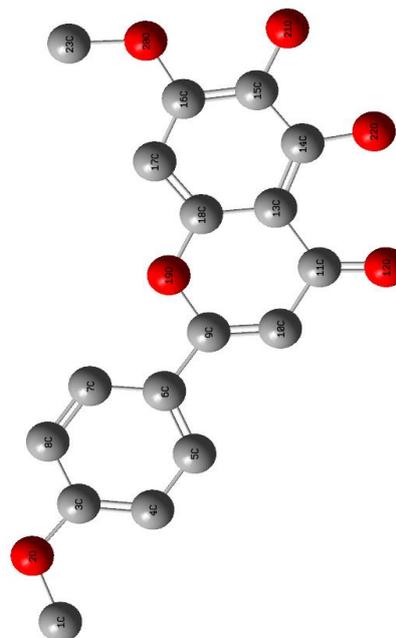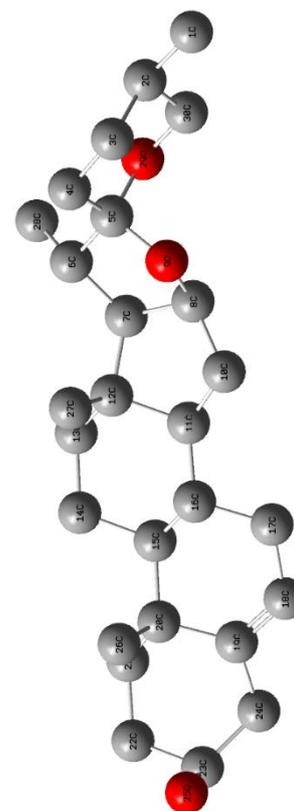It can be concluded from the analysis of the results in **table III.5** that **O21** and **O22** have the largest F+ value and will be the preferred site for nucleophilic attack. **O12** has the largest F⁻ and F0 values, the carboxylic function **O12** will be the preferred site for electrophilic attack and radical's attack.

The table shows that positive Δf values are determined for the O2, C4, C6, C8, C10, C13, C15, C17, C18, and O20-O22 atoms subject to nucleophilic attacks. On the other hand, C1, C3, C5, C7, C9, C11, C12, C14, C16, C19, and C23 have negative Δf values This reveals that atoms are highly reactive to electrophilic attacks.

Table III.6 Values of the Fukui functions and Dual descriptor of Diosgenin.

Atoms	N	N-1	N+1	f^-	f^+	f^0	Δf
1 C	-0.410	-0.403	-0.412	-0.007	-0.002	-0.009	0.005
2 C	-0.110	-0.094	-0.126	-0.016	-0.015	-0.032	0.001
3 C	-0.244	-0.227	-0.251	-0.016	-0.007	-0.024	0.009
4 C	-0.249	-0.234	-0.254	-0.015	-0.004	-0.020	0.011
5 C	0.410	0.411	0.380	-0.001	-0.029	-0.031	-0.028
6 C	-0.096	-0.086	-0.103	-0.009	-0.006	-0.016	0.003
7 C	-0.134	-0.125	-0.139	-0.009	-0.005	-0.014	0.004
8 C	0.113	0.113	0.098	-0.0004	-0.014	-0.015	-0.014
9 O	-0.527	-0.529	-0.491	0.002	0.035	0.037	0.033
10 C	-0.260	-0.247	-0.268	-0.012	-0.008	-0.021	0.004
11 C	-0.118	-0.102	-0.127	-0.016	-0.008	-0.024	0.008
12 C	0.057	0.064	0.051	-0.006	-0.006	-0.012	0.000
13 C	-0.247	-0.236	-0.253	-0.010	-0.006	-0.016	0.004
14 C	-0.267	-0.260	-0.272	-0.007	-0.005	-0.012	0.002
15 C	-0.083	-0.072	-0.098	-0.011	-0.015	-0.026	-0.004
16 C	-0.100	-0.097	-0.110	-0.004	-0.009	-0.013	-0.005
17 C	-0.298	-0.257	-0.332	-0.040	-0.033	-0.074	0.007
18 C	-0.164	-0.241	-0.100	0.076	0.064	0.141	-0.012
19 C	0.183	0.159	0.211	0.0246	0.027	0.052	0.002
20 C	-0.015	0.005	-0.041	-0.020	-0.026	-0.047	-0.006
21 C	-0.261	-0.245	-0.270	-0.015	-0.009	-0.025	0.006
22 C	-0.260	-0.247	-0.268	-0.013	-0.008	-0.021	0.005
23 C	0.110	0.133	0.097	-0.022	-0.012	-0.035	0.010
24 C	-0.351	-0.322	-0.375	-0.029	-0.024	-0.054	0.005
25 O	-0.606	-0.618	-0.582	0.011	0.024	0.035	0.013
26 C	-0.437	-0.424	-0.442	-0.013	-0.005	-0.019	0.008
27 C	-0.441	-0.431	-0.444	-0.009	-0.003	-0.012	0.006
28 C	-0.400	-0.392	-0.406	-0.008	-0.006	-0.014	0.002
29 O	-0.533	-0.534	-0.456	0.0005	0.077	0.078	0.077
30 C	-0.010	0.003	-0.041	-0.014	-0.030	-0.044	-0.016

It can be concluded from the analysis of the results in **table III.6** that **C18 and O29** have the largest F+ and F0 values and will be the preferred site for the nucleophilic attack, and radicals attack respectively. **While C18 and C19** have the largest F- values will be the preferred site for electrophilic attack. C18 was the preferred site for the three attacks nucleophilic, electrophilic, and radical attack.

The table shows that positive Δf values are determined for most atoms: **C1-C4, C6, C7, O9, and C10-C14, C17, C19, C21-C24, O25, C26-C28, and O29** atoms subject to nucleophilic attacks. On the other hand, **C5, C8, C15, C16, C18, C20, and C30** have negative Δf values This reveals that atoms are highly reactive to electrophilic attacks.

Table III.7 Values of the Fukui functions and Dual descriptor of Protodioscin.

Atoms	N	N-1	N+1	f^-	f^+	f^0	Δf
1 C	-0.398	-0.398	-0.401	-0.0001	-0.003	-0.003	-0.003
2 C	-0.080	-0.078	-0.084	-0.002	-0.003	-0.005	-0.002
3 C	-0.133	-0.132	-0.139	-0.0005	-0.005	-0.0064	-0.005
4 C	0.120	0.121	0.096	-0.0005	-0.024	-0.025	-0.024
5 C	-0.261	-0.259	-0.267	-0.001	-0.005	-0.007	-0.004
6 C	-0.110	-0.108	-0.115	-0.002	-0.005	-0.007	-0.004
7 C	0.057	0.058	0.052	-0.001	-0.004	-0.006	-0.004
8 C	-0.246	-0.245	-0.249	-0.001	-0.003	-0.004	-0.002
9 C	-0.268	-0.266	-0.273	-0.002	-0.004	-0.006	-0.003
10 C	-0.085	-0.083	-0.090	-0.001	-0.006	-0.007	-0.004
11 C	-0.107	-0.105	-0.114	-0.001	-0.007	-0.009	-0.006
12 C	-0.294	-0.287	-0.314	-0.007	-0.019	-0.027	-0.012
13 C	-0.171	-0.187	-0.133	0.015	0.038	0.054	0.023
14 C	0.159	0.151	0.176	0.008	0.017	0.025	0.009
15 C	-0.013	-0.007	-0.028	-0.005	-0.015	-0.021	-0.010
16 C	-0.257	-0.253	-0.265	-0.004	-0.007	-0.011	-0.003
17 C	-0.247	-0.245	-0.251	-0.001	-0.004	-0.006	-0.002
18 C	0.116	0.117	0.110	-0.0009	-0.006	-0.007	-0.005
19 C	-0.315	-0.310	-0.327	-0.005	-0.012	-0.017	-0.007
20 O	-0.517	-0.519	-0.504	0.001	0.013	0.014	0.011
21 C	0.333	0.334	0.329	-0.0007	-0.004	-0.005	-0.004
22 C	0.109	0.109	0.102	-0.0008	-0.006	-0.007	-0.006
23 C	0.121	0.123	0.114	-0.002	-0.006	-0.008	-0.004
24 C	0.068	0.071	0.069	-0.002	0.001	-0.0012	0.003
25 C	0.094	0.096	0.087	-0.001	-0.007	-0.008	-0.006
26 O	-0.522	-0.524	-0.518	0.002	0.003	0.005	0.002
27 C	-0.028	-0.029	-0.031	0.0007	-0.003	-0.0022	-0.004
28 O	-0.640	-0.637	-0.612	-0.002	0.027	0.025	0.029
29 O	-0.540	-0.536	-0.535	-0.004	0.005	0.0011	0.010
30 C	0.354	0.349	0.349	0.004	-0.004	0.0002	-0.010
31 C	0.057	0.055	0.055	0.001	-0.002	-0.00013	-0.004
32 C	0.112	0.110	0.111	0.002	-0.001	0.0012	-0.004
33 C	0.107	0.104	0.106	0.003	-0.001	0.0019	-0.004
34 C	0.081	0.094	0.076	-0.013	-0.004	-0.018	0.009
35 O	-0.547	-0.551	-0.554	0.003	-0.006	-0.0025	-0.010
36 C	-0.429	-0.427	-0.429	-0.001	-0.0007	-0.0019	0.001
37 O	-0.614	-0.606	-0.611	-0.007	0.002	-0.0049	0.010
38 O	-0.621	-0.622	-0.618	0.0004	0.003	0.0038	0.003
39 O	-0.625	-0.598	-0.623	-0.027	0.0015	-0.0255	0.029
40 O	-0.653	-0.657	-0.636	0.003	0.017	0.0213	0.014
41 O	-0.528	-0.529	-0.519	0.0009	0.009	0.0108	0.009
42 C	0.328	0.328	0.318	0.0001	-0.010	-0.0100	-0.010

Table III.7 Continued

43 C	**0.090**	**0.092**	**0.082**	**-0.002**	**-0.007**	**-0.0095**	-0.005
44 C	0.054	0.059	0.054	-0.004	-0.0004	-0.0049	0.004
45 C	0.096	0.098	0.087	-0.002	-0.008	-0.010	-0.006
46 C	0.065	0.069	0.051	-0.004	-0.014	-0.0184	-0.010
47 O	-0.521	-0.521	-0.498	0	0.022	0.023	0.022
48 C	-0.431	-0.427	-0.433	-0.003	-0.002	-0.006	0.002
49 O	-0.608	-0.614	-0.592	0.006	0.015	0.022	0.010
50 O	-0.629	-0.613	-0.626	-0.015	0.002	-0.012	0.019
51 O	-0.604	-0.608	-0.572	0.003	0.032	0.036	0.029
52 C	-0.418	-0.416	-0.420	-0.001	-0.002	-0.0041	-0.001
53 C	-0.440	-0.439	-0.441	0	-0.001	-0.0019	0.000
54 O	-0.526	-0.526	-0.461	0	0.064	0.064	0.066
55 C	0.384	0.386	0.360	-0.001	-0.024	-0.026	-0.022
56 C	-0.250	-0.249	-0.251	-0.0009	-0.0008	-0.0017	0.000
57 C	-0.260	-0.260	-0.266	-0.0001	-0.005	-0.0056	-0.005
58 C	-0.092	-0.090	-0.095	-0.002	-0.002	-0.0050	-0.001
59 C	-0.417	-0.416	-0.418	-0.0006	-0.001	-0.0017	0.000
60 C	-0.022	-0.013	-0.023	-0.008	-0.0013	-0.010	0.007
61 O	-0.514	-0.521	-0.515	0.006	-0.0004	0.0063	-0.007
62 C	0.338	0.331	0.339	0.006	0.0007	0.0074	-0.006
63 C	0.137	0.134	0.136	0.003	-0.0017	0.0014	-0.005
64 C	0.084	0.085	0.084	-0.001	0.0003	-0.0007	0.001
65 C	0.106	0.114	0.105	-0.007	-0.0012	-0.0089	0.006
66 C	0.081	0.068	0.077	0.013	-0.0039	0.0093	-0.017
67 O	-0.528	-0.525	-0.521	-0.003	0.006	0.0032	0.010
68 C	-0.023	-0.032	-0.022	0.008	0.0014	0.010	-0.007
69 O	-0.644	-0.605	-0.644	-0.038	-0.0005	-0.038	0.038
70 O	-0.652	-0.657	-0.645	0.004	0.006	0.011	0.002
71 O	-0.634	-0.638	-0.634	0.004	0.0003	0.0043	-0.004
72 O	-0.642	-0.645	-0.653	0.003	-0.011	-0.008	-0.014
73 O	-0.655	-0.657	-0.619	0.001	0.035	0.037	0.034
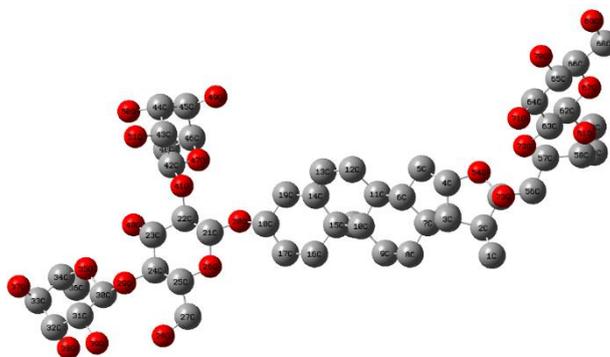
It can be concluded from the analysis of the results in **table III.7** that **O54** has the largest F+ and F0 values and will be the preferred site for nucleophilic attack and radical attack respectively. **While C13 and C66** have the largest F- values will be the preferred site for electrophilic attack.

The table shows that positive Δf values are determined for the **C13, C14, O20, O24, O28, O29, C34, C36, O35-O41, C44, O47, C48, O49-O51, C53, O54, C56, C59, C60, C64, C65, O67 O69,**

O70 and O73 atoms are subject to nucleophilic attacks. On the other hand, **C1-C12, C15-C19, C21-C23, C25, C27, C30-C33, O35, C42, C43, C45, C46, C52, C55, C57, C58, O61, C62, C63, C66, C68, O71, and O72** have negative Δf values This reveals that atoms are highly reactive with respect to electrophilic attacks.

III.4 ADMET PROPERTIES

To highlight potential drug candidates, ADMET properties were developed for preliminary prediction of pharmacokinetic, physicochemical, and drug-like properties during drug discovery. In silico studies provide access to pharmacokinetic parameters (ADMET), one of the most discussed topics in medicinal chemistry courses is that high oral bioavailability is an important factor in optimizing bioactive molecules as therapeutics [20]. The results of ADMET prediction using SwissADME and ADMETlab are summarized in **table III.8:**

In general, compounds are adsorbed when they verify Lipinski and Viber rules, which means that a given molecule with partition coefficient iLogP less than 5, weight (MW) under 500 Da, a surface area less than 140 Å and rotatable bonds under 10 is easily adsorbed. All compounds passed Lipinski and Viber rules except the **Protodioscin** molecule**.**

the results of **table III.8** show that the Protodioscin molecule is a P-gp substrate and has low GI-absorption. While the other molecules Diosgenin, Ladanein, Thymoquinone, and Trigonelline are highly absorbed and are not P-gp substrates.

The blood-brain barrier acts as an additional boundary between the circulating blood and the extracellular space of the brain, only **Diosgenin,** and **Thymoquinone** are BBB permeants.

Cytochrome P450s is considered an important enzyme system for drug metabolism in liver. The two main subtypes of cytochrome P450 are CYP3A4 and CYP2D6. The results obtained from table 3-8 show that the compound **Ladanein** inhibits the two cytochrome P450 subtypes, which means that they can be metabolized in the liver. The other compounds Diosgenin, Protodioscin, Thymoquinone, and Trigonelline are not CYP3A4 CYP2D6 and Inhibitors.

Table III.8 ADME properties for the studied

Molecule	MW	#Rotatable Bond	#H-Bond acceptors	#H-Bond donors	MR	TPSA	iLOGP	GI absorption	BBB permeant	Pgp substrate	CYP2D6 inhibitor	CYP3A4 inhibitor	T	CL	hERG	Ames
Diosgenin	414.62	0	3	1	121.59	38.69	4.49	High	Yes	No	No	No	1.66	1.52	1	0
Ladanein	314.29	3	6	2	84.95	89.13	2.85	High	No	No	Yes	Yes	1.64	0.02	0	1
Protodioscin	1049.2	14	22	13	252.16	346.06	4.07	Low	No	Yes	No	No	2.76	0.34	1	0
Thymoquinone	164.2	1	2	0	47.52	34.14	1.99	High	Yes	No	No	No	1.53	1.29	0	0
Trigonelline	137.14	1	2	0	35.05	44.01	-3.11	High	No	No	No	No	1.09	1.43	0	0

III.5 MOLECULAR DOCKING STUDIES

III.5.1 The binding affinity of the studied ligands with wild type-EGFR:

Molecular docking studies are an efficient tool and an important step in drug design. It was used in this study to identify receptor-ligand interactions in the binding pocket of the wild type-EGFR protein **(PDB ID: 1xkk)**, **(PDB ID: 2itv)**, and **(PDB ID: 5hg5)**. The protein was thought to be rigid, whereas the ligands were thought to be flexible.

At first, the selected compounds Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien were docked into the binding site of the EGFR protein **(PDB ID: 1xkk)** to assess their abilities to inhibit the disease. **Table III.9** summarizes the results obtained. The results show that the molecular docking score of the five compounds are about (-5.5899, -5.9482, -6.9800, -4.7911, and -4.3359 kcal/ mol), and the RMSDs values are about (2.2915, 3.1618, 3.1772, 1.2649, and 2.1991 Å) for Diosgenin, Ladanien, Protodioscin, Thymoquinone, and Trigonelline respectively. The first three compounds Diosgenin, Ladanien, and Protodioscin showed the best docking scores of (-5.5899, -5.9482, and -6.9800 kcal/mol) respectively (**i.e.**, they have the highest inhibitory potential).

Because the best RMSD value is close to 2 Å we can say that Diosgenin, Thymoquinone and Trigonelline compounds have a good RMSD ranging between (1.2649 Å) and (2.2915Å).

The results from the **fig III.12 (A)** showed that the **Diosgenin** molecule interacts with one amino acid **HOH 22** with H- acceptor interaction bond between O-9 and O with a length of (3.20 Å)**,** while the **Ladanien** molecule forms three different interactions with two amino acids, two strong hydrogen bonds (3.18 Å), (2.98 Å) between O 21 and N and between O 22 and N respectively, of **MET 793** residues, and a **Pi-H** interaction between 6-ring and CD1 of **LEU 718** with a length of (3.99 Å) as shown in the **fig III.12 (B)**. **Protodioscin** molecule makes interactions with four amino acids ASP 916, LYS 913, HOH 22, and HOH 4, in five different interactions, H-donor interaction with ASP 916 between O-37 and OD2 (2.98 Å), and H-acceptor interactions with LYS 913 between O-40 and NZ (3.09 Å), and between O-51 and NZ (3.05 Å), another H-acceptor interaction with HOH 22 between O-54 and O (2.98 Å), and the last one is also an H-acceptor interaction with HOH 4 between   O-61 and O by a length of (3.15 Å), **fig III.12 (C)** represent these interactions. The **Thymoquinone** and **Trigonelline** molecules did not form any interactions with this receptor as shown in **fig III.12 (D, E)**.

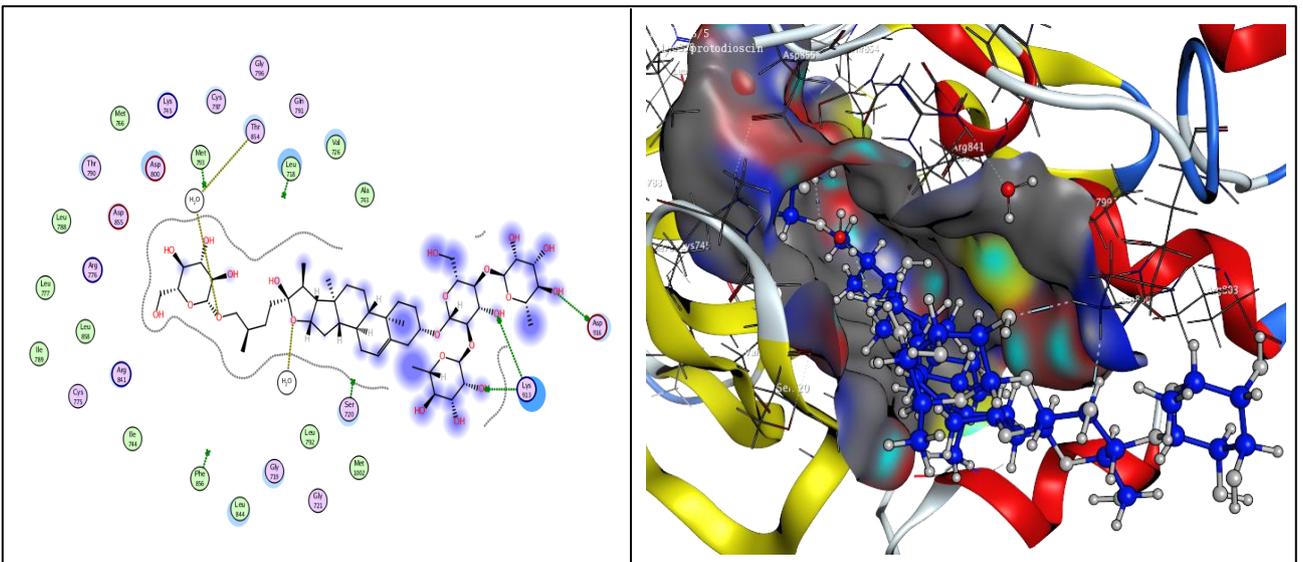Table III.9 The results obtained from docking of compounds with 1xkk.

| Ligands | S score (kcal/mol) | RMSD (Å) | Bonds between atoms of compounds and residues of the active site of 1xkk | | | | | |
			Atom of compound	Atom of receptor	Involved receptor residues	Type of interaction bond	Distance (Å)	E (kcal/mol)
Diosgenin	-5.5899	2.2915	O-9	O	HOH 22	H-acceptor	3.20	-0.8
Ladanein	-5.9482	3.1618	O-21	N	MET 793	H-acceptor	3.18	-0.9
			O-22	N	MET 793	H-acceptor	2.98	-0.7
			6-ring	CD1	LEU 718	Pi-H	3.99	-0.6
Protodioscin	-6.9800	3.1772	O-37	OD2	ASP 916	H-donor	2.98	-3.5
			O-40	NZ	LYS 913	H-acceptor	3.09	-3.2
			O-51	NZ	LYS 913	H-acceptor	3.05	-3.5
			O-54	O	HOH 22	H-acceptor	2.98	-1.1
			O-61	O	HOH 4	H-acceptor	3.15	-0.8
Thymoquinone	-4.7911	1.2649	—	—	—	—	—	—
Trigonelline	-4.3359	2.1991	—	—	—	—	—	—
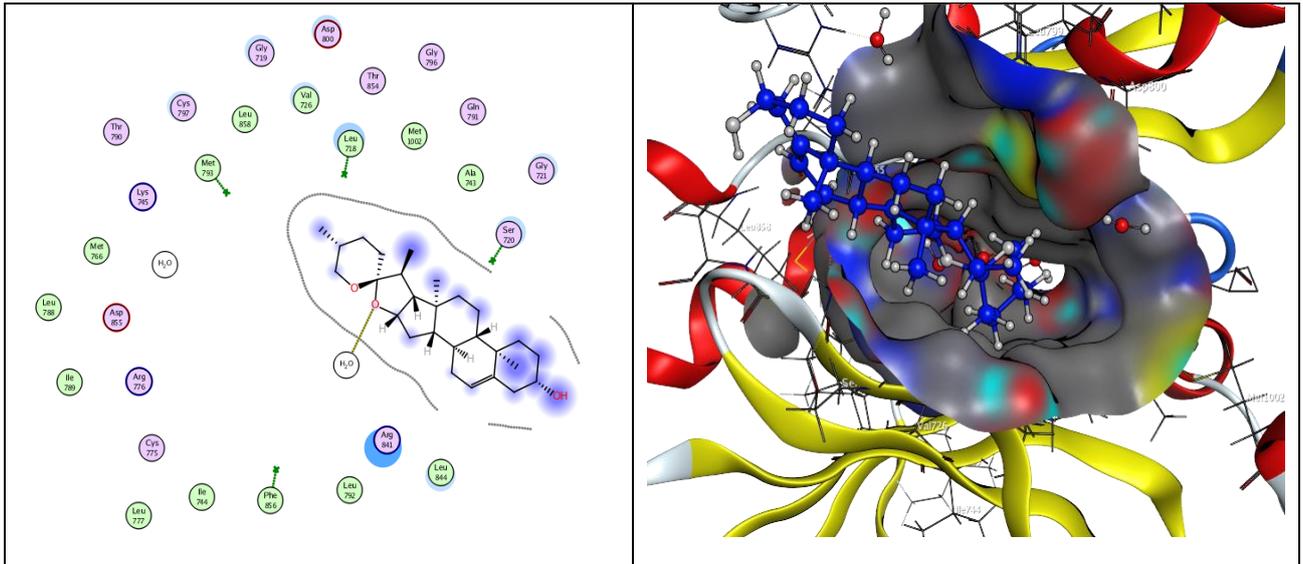
(A)
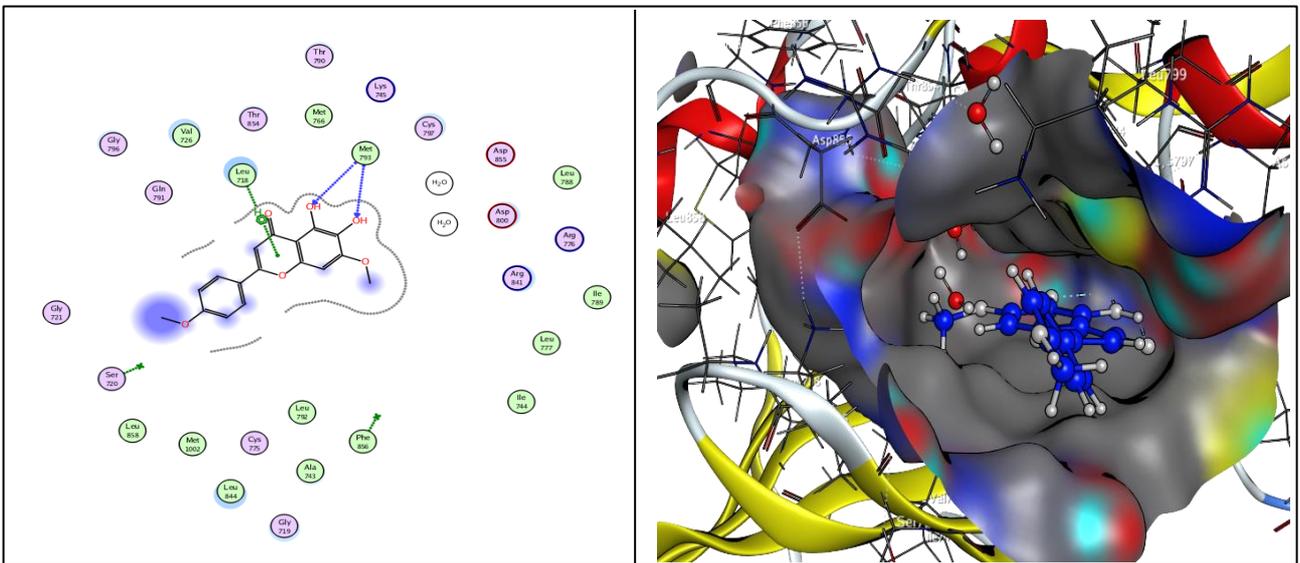
(B)

(C)
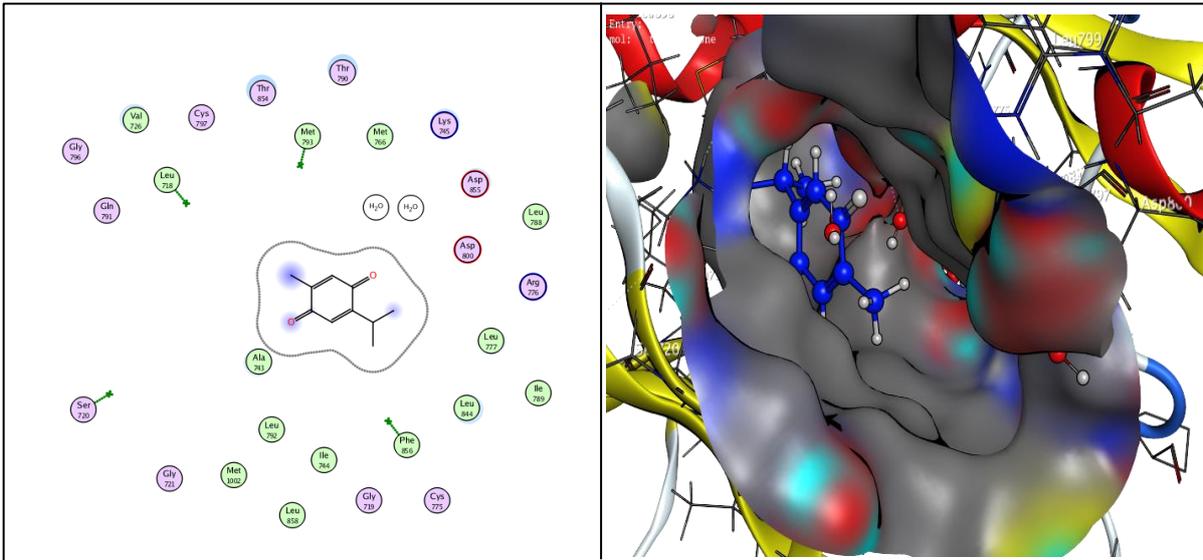
(D)
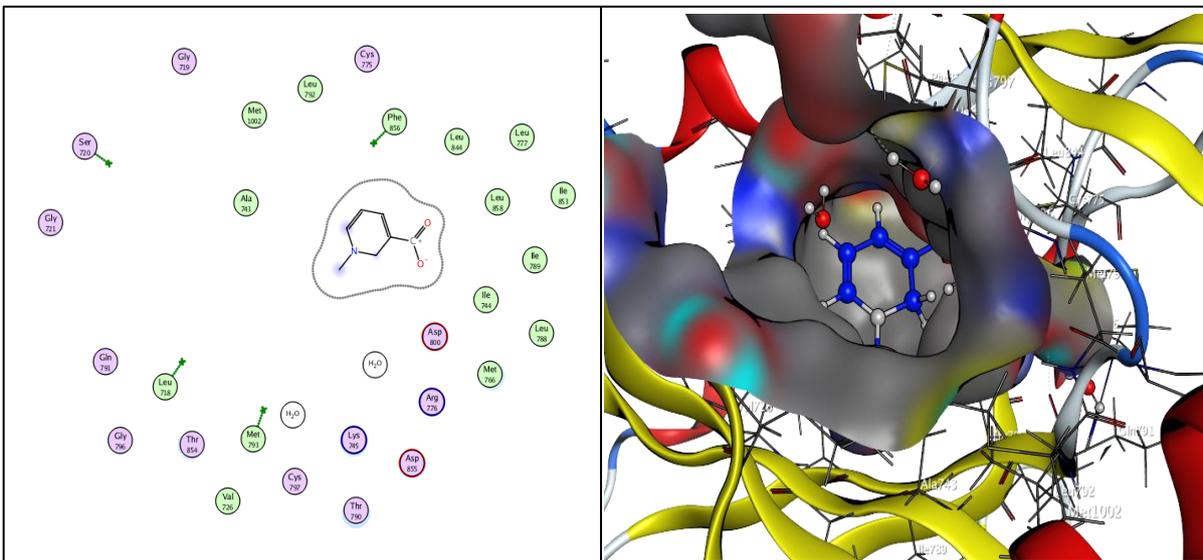
(D)
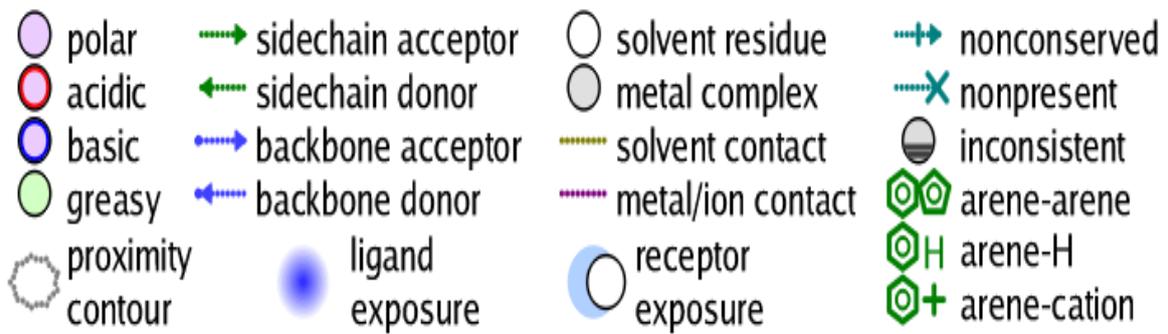
Figure III.12 interactions between (1-5) molecules with 1xkk receptor.

III.5.2 The binding affinity of the studied ligands with EGFR L858R mutation :

To know the selectivity of our compounds for EGFR (PDB ID: 2itv) receptor, molecular docking was carried out using MOE software, Therefore, we aim to study the molecular interactions implicated between the active binding sites of the target protein, with the selected compounds Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien. The Docking results are summarized in **table III.10.**

Table III.10 summarizes the binding score and RMSD value of docking results of the selected compounds in (2itv) pockets, the results obtained from the table show that the binding score of the five molecules Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien are about (**-4.3372, -5.4591, -7.6220, -3.6003, and -5.1731** kcal/mol), and the RMSDs values are about (2.6311, 2.6253, 1.9304, 2.4824, and 2.6586 Å) respectively.

The results from the **fig III.13** show that all compounds form interactions with the protein (2itv), the **Diosgenin** molecule forming only one H-acceptor interaction with amino acid LYS 745 between O-25 and NZ with a length (2.94 Å), as shown in the **fig III.13 (A),** while **Ladanein** makes two interactions, an H-donor interaction with amino acid ASP 855 between O-22 and OD1 with a length of (3.10 Å), and a Pi-H interaction with amino acid ARG 841 between 6-ring and CB with a length (3.99 Å) as shown in the **fig III.13 (B).**

The results from the **fig III.13 (C)** show that the **protodioscin** molecule forms four interactions with four different amino acids, an H-donor's interactions with GLU 804, PRO 794, and ASP 803 between (O-38...OE2), (O-49…O), and (O-71…OD2), with a length (3.01 Å), (2.72 Å), and (3.32 Å) respectively, as well as H-acceptor interaction with LYS 745 between O-73 and NZ, with a length of (2.71 Å).

The results from **fig III.13 (D)** showed that the **thymoquinone** compound had formed two interactions, one hydrogen bond (H-acceptor) with LYS 745 (O-1…NZ) (3.29 Å), and a Pi-H interaction with VAL 726 (6-ring…CG2) (4.23 Å). Whereas **fig III.13 (E)** showed that the **Trigonelline** compound forms three interactions. The first one is a Hydrogen bond with LYS 745 between (O-9…NZ), and the second and the third one is an ionic interaction with LYS 745 between O-9 and NZ, and between O-10…NZ, with a length (3.19 Å) and (3.49 Å) respectively.

Table III.10 The results obtained from docking of compounds with 2itv.

| Ligands | S score (kcal/mol) | RMSD (Å) | Bonds between atoms of compounds and residues of the active site of 2itv | | | | | |
			Atom of compound	Atom of receptor	Involved receptor residues	Type of interaction bond	Distance (Å)	E (kcal/mol)
Diosgenin	-5.4591	2.6253	O-25	NZ	LYS 745	H-acceptor	2.94	-3.8
Ladanein	-5.1731	2.6586	O-22	OD1	ASP 855	H-donor	3.10	-1.5
			6-ring	CB	ARG 841	Pi-H	3.99	-0.8
Protodioscin	-7.6220	1.9304	O-38	OE2	GLU 804	H-donor	3.01	-1.8
			O-49	O	PRO 794	H-donor	2.72	-2.5
			O-71	OD2	ASP 837	H-donor	3.32	-1.1
			O-73	NZ	LYS 745	H-acceptor	2.71	-3.1
Thymoquinone	-4.3372	2.6311	O-1	NZ	LYS 745	H-acceptor	3.29	-3.7
			6-ring	CG2	VAL 726	Pi-H	4.23	-0.6
Trigonelline	-3.6003	2.4824	O-9	NZ	LYS 745	H-acceptor	3.19	-2.6
			O-9	NZ	LYS 745	Ionic	3.19	-3.3
			O-10	NZ	LYS 745	ionic	3.49	-1.9

(A)

(B)

(C)
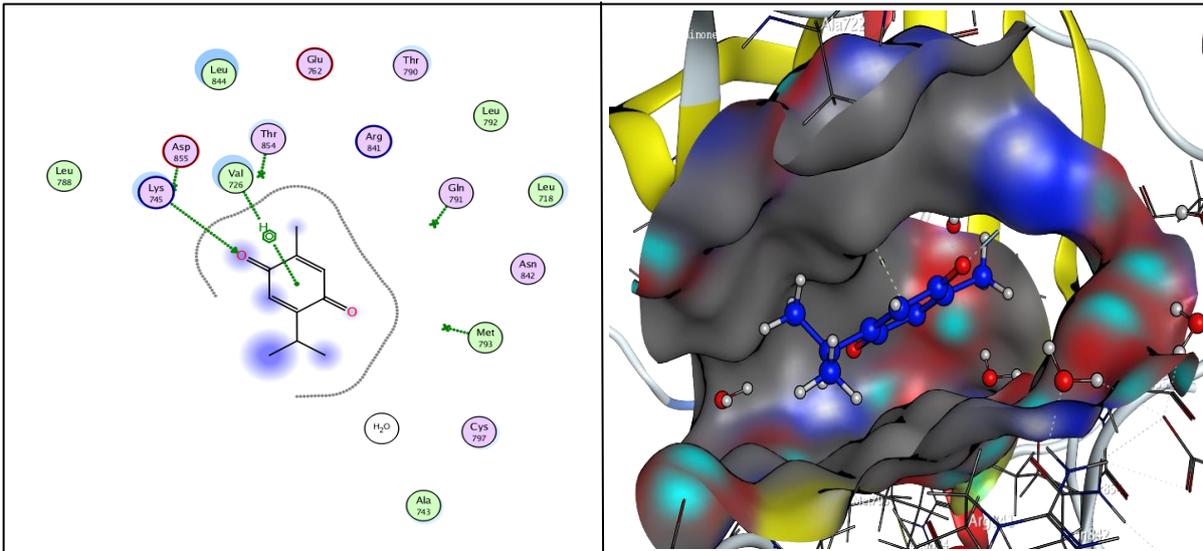
(D)
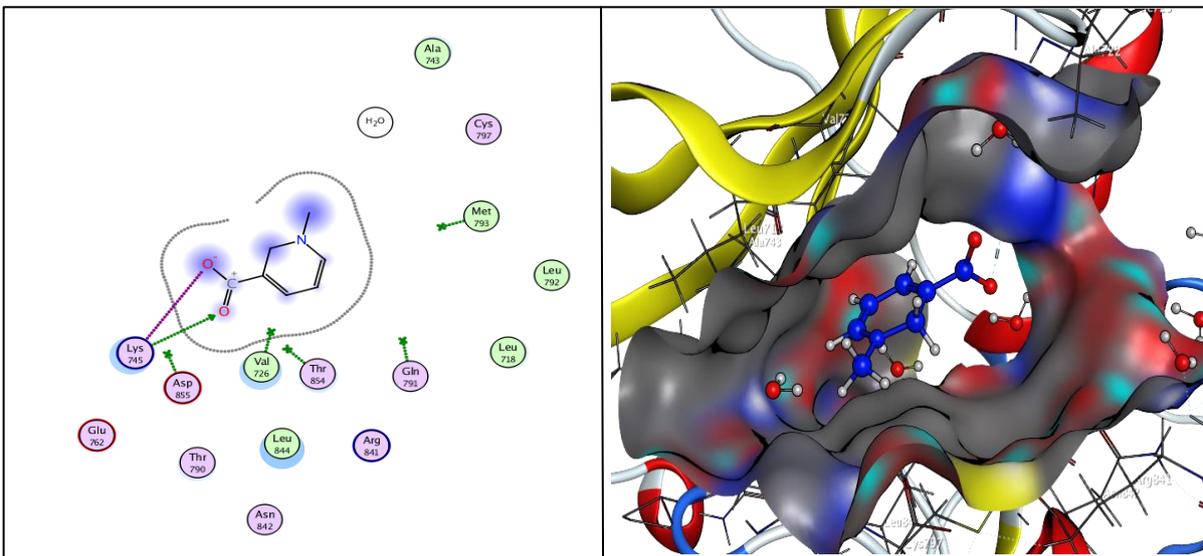
(E)

Figure III.13 interactions between (1-5) molecules with 2itv receptor.

III.5.3 The binding affinity of the studied ligands with T790M mutation EGFR:

The selected compounds Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien were docked into the binding site of the EGFR protein **(PDB ID 5hg5).**

Table III.11 summarizes the binding score and RMSD value of docking results of the selected compounds in (5hg5) pockets, the results obtained from the table show that the binding score of the five molecules Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien are about (**-5.2982, -5.3151, -8.6786, -3.7021, and -6.2684** kcal/mol), and the RMSDs values are about (1.2197, 2.7084, 3.1972, 1.4729, and 1.9689 Å) respectively. It is clear from the docking results of compounds with the (5hg5) receptor that the ligands Ladanein, Diosgenin, and thymoquinone gives a good RMSD value and energy score of less than -7 kcal/mol. On the other hand, protodioscin had an RMSD of more than 3 Å with the best energy score (-8.6786 kcal/mol). meanwhile, the ligand Trigonelline has an RMSD of less than 1.5 Å.

The results from the **fig III.14 (A)** and **fig III.14 (E)** show that the Diosgenin and Trigonelline compounds did not form any interaction in 5hg5 pockets.

Fig III.14 (D) show that thymoquinone molecule interact with MET 793 by forming an H-acceptor bond between O-9 and N with a length 3.25 Å.

The Ladanein molecule formed two Pi-H interactions with two different amino acids LEU 718 (6-ring…CB) with a length (4.00 Å) and VAL 726 (6-ring…CG2) with a length (4.20 Å) as shown in the **fig III.14 (B).**

Fig III.14 (C) shows that the protodioscin compound interacts with five amino acid residues ASP 800, HOH 9388, HOH 9245, LYS 875, and HOH 9141 different interactions, an H-donor interaction with ASP 800 between O-50 and OD1 with a length of (3.48 Å), as well as four H-acceptor interactions with HOH 9388 (O-39…O) with a length 2.90 Å, HOH 9245 (O-40…O) with a length 2.94 Å, LYS 875 (O-69…NZ) with a length 2.77 Å, and HOH 9141 (O-72…O) with a length 2.84 Å.

Table III.11 The results obtained from docking of compounds with 5hg5.

Ligands	S score (kcal/mol)	RMSD (Å)	Bonds between atoms of compounds and residues of the active site of 5hg5					
			Atom of compound	Atom of receptor	Involved receptor residues	Type of interaction bond	Distance (Å)	E (kcal/mol)
Diosgenin	-5.3151	2.7084	—	—	—	—	—	—
Ladanein	-6.2684	1.9689	6-ring	CB	LEU 718	Pi-H	4.00	-0.6
			6-ring	CG2	VAL 726	Pi-H	4.20	-0.7
Protodioscin	-8.6786	3.1972	O-50	OD1	ASP 800	H-donor	3.48	-0.9
			O-39	O	HOH 9388	H-acceptor	2.90	-0.6
			O-40	O	HOH 9245	H-acceptor	2.94	-0.8
			O-69	NZ	LYS 875	H-acceptor	2.77	-7.8
			O-72	O	HOH 9141	H-acceptor	2.84	-1.1
Thymoquinone	-5.2982	1.2197	O-9	N	MET 793	H-acceptor	3.25	-2.1
Trigonelline	-3.7021	1.4729	—	—	—	—	—	—
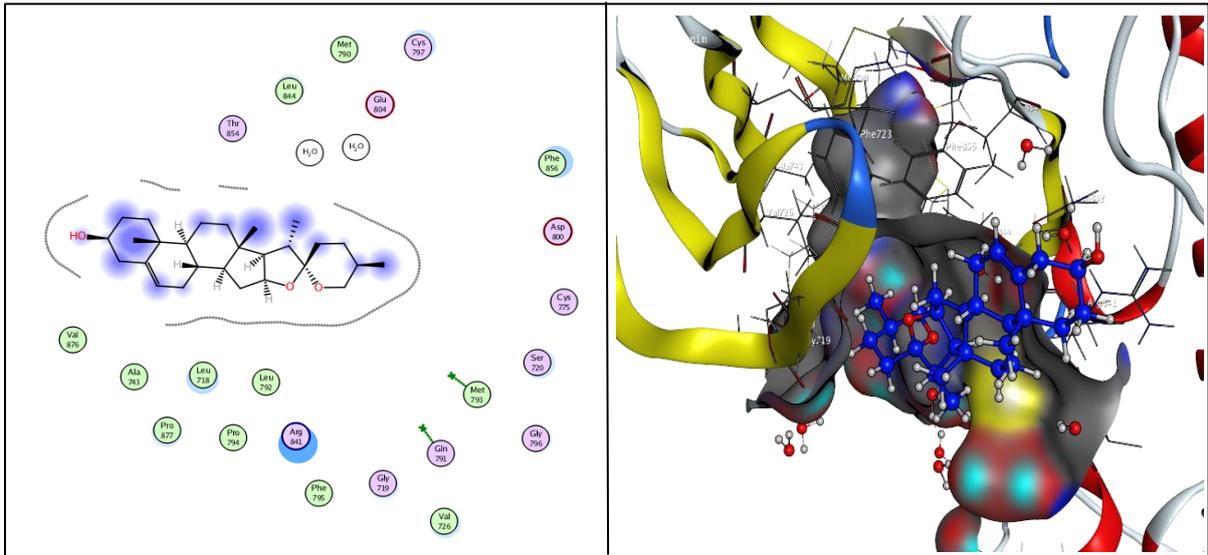
(A)
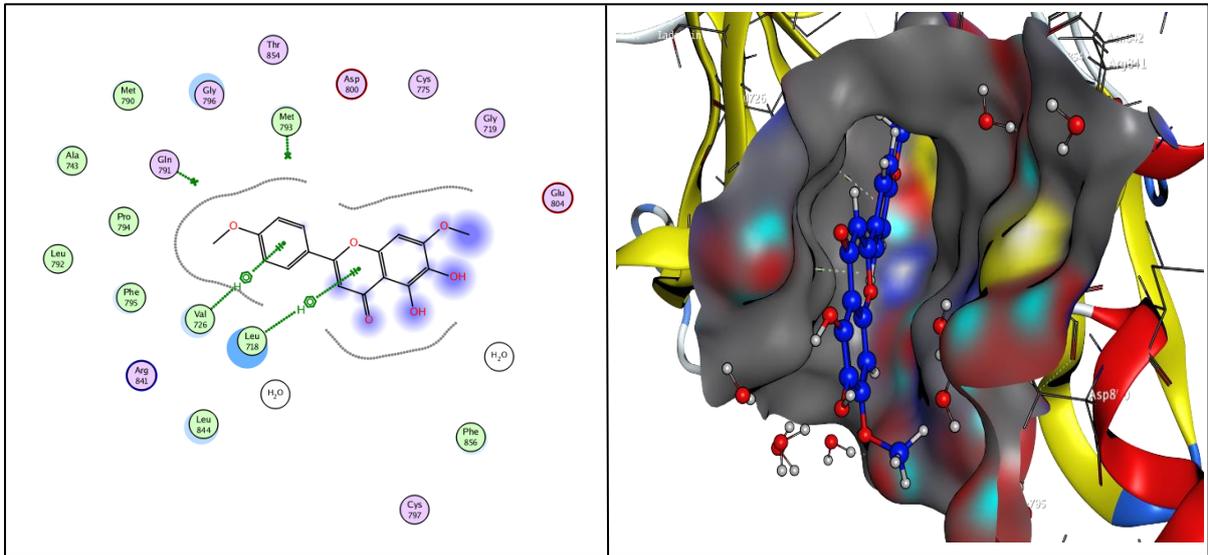
(B)
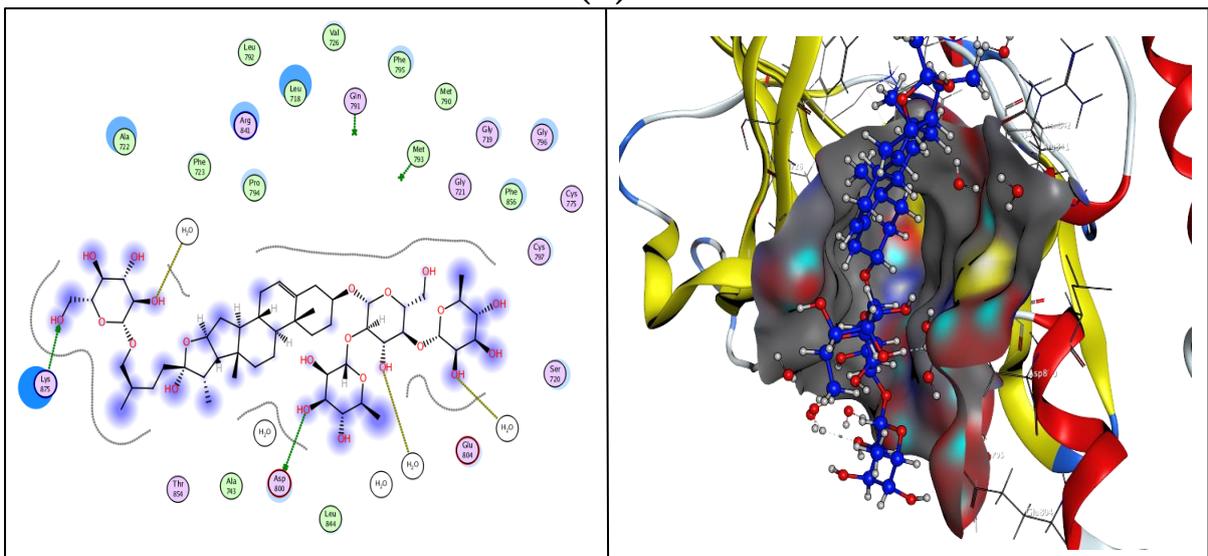
(C)
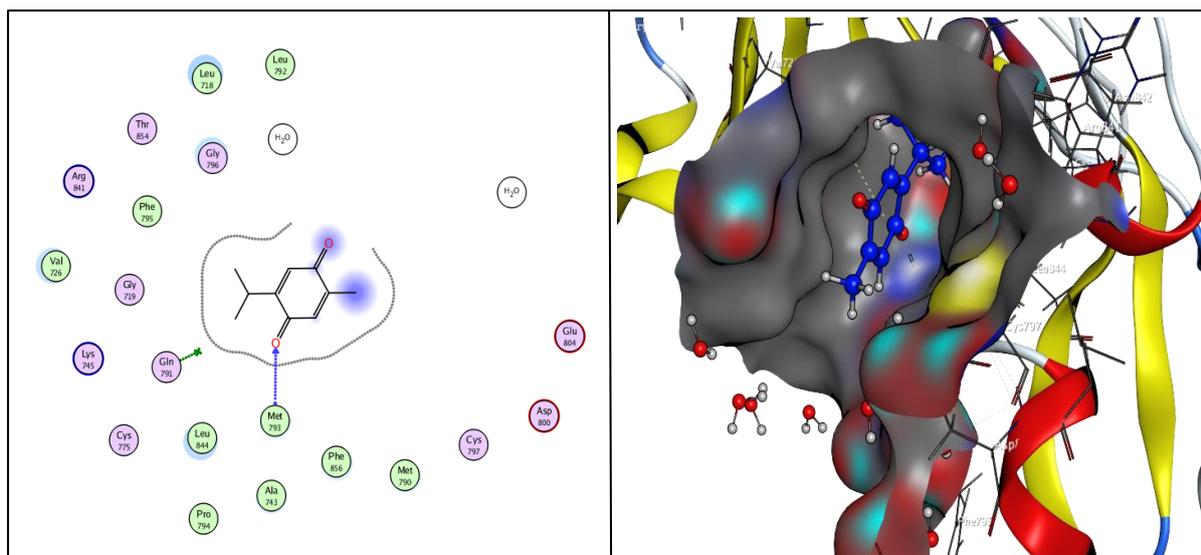
(D)
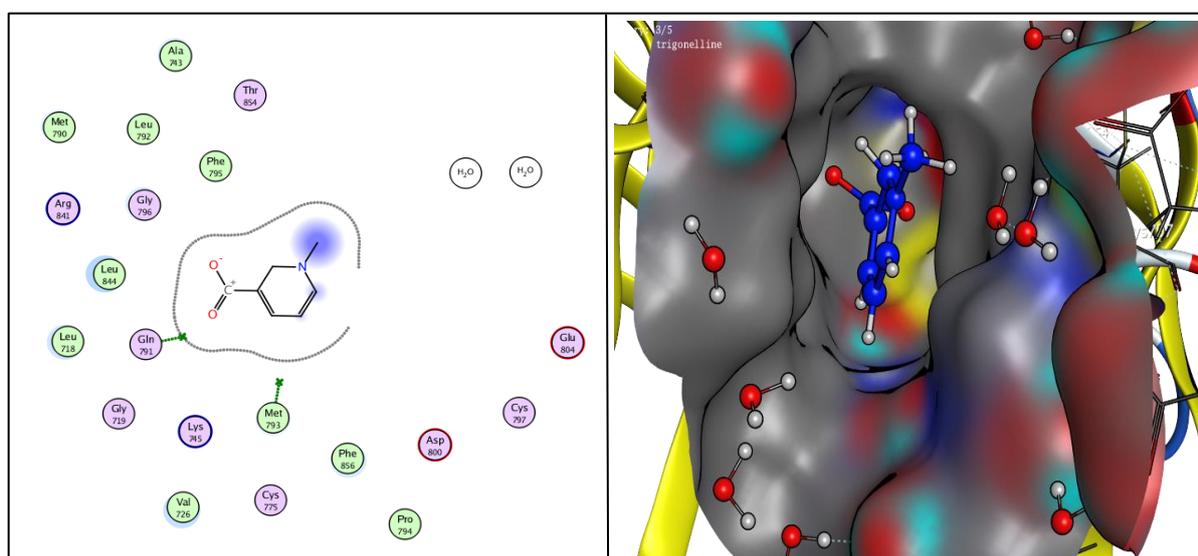
(E)
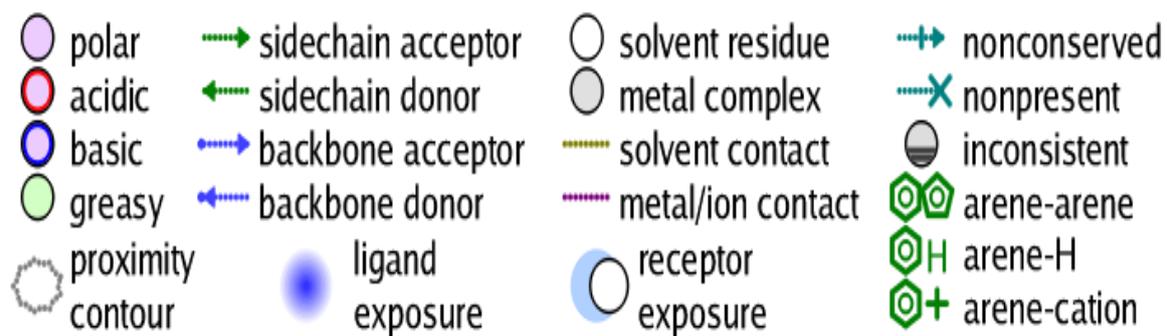
Figure III.14 interactions between (1-5) molecules with 5hg5 receptor.

Conclusion

Conclusion

T he present research aimed to apply a computational approach to develop therapeutic agents for Cancer Treatment. In this work, we studied the epidermal growth factor receptor tyrosine kinase mutation disease. This study was conducted in order to pinpoint the best drug candidates from the set of five compounds of the major compounds from selected North African plants used traditionally in cancer therapy. By using the DFT method, ADMET properties, and molecular docking calculations.

The optimized molecular structures of Thymoquinone, Diosgenin, Protodioscin, Trigonelline, and Ladanien have been carried out using the DFT/B3LYP/6-31G method and their geometrical parameters were also determined, then evaluated for their inhibitory activities towered three different EGFR mutations Wild-Type, L858R and T790M using molecular docking.

Molecular properties such as electrostatic Potential (MEP), frontiers orbitals (FMO), gap energies, and reactivity descriptors (global reactivity and local reactivity) have been discussed. The FMO results reveal that the energy gap of the trigonelline and thymoquinone compounds is smaller which makes these molecules soft and more reactive compared to the other molecules. meanwhile, the calculated MEP maps show that the positive potential sites are favorable for a nucleophilic attack, whereas the negative potential sites are favorable for the electrophilic attack.

The Docking results were discussed based on the different interactions between the ligands and 1xkk, 2itv, and 5hg5 proteins. The docking results show that all studied compounds interacted with L858R mutation, the protodioscin ligand gave the best energy score and good RMSD value followed by Diosgenin and Ladanein. In addition, the docking results with 1xkk showed that Thymoquinone and Trigonelline did not form any interactions while the other ligands interacted with this protein. On the other hand, the results of docking with 1xkk showed that only three of five ligands were interacted with this protein.

It is noticed that **Protodioscin** gives good results by interacting with the three receptors in the docking study (at least form four interactions with every receptor), but unfortunately, the ADME analysis results showed that the Protodioscin compound had not passed Lipinski and Viber rules and it is a P-gp substrate and had low GI- absorption. It means that the protodioscin compound had a good pharmacodynamics property, although the bad pharmacokinetics properties.
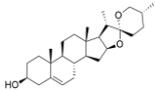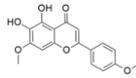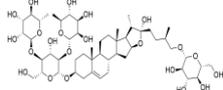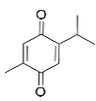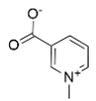
Chapter 3 References

[1]   M. Govindarajan, M. Karabacak, S. Periandy, and D. Tanuja, "Spectroscopic (FT-IR, FT-Raman, UV and NMR) investigation and NLO, HOMO–LUMO, nbo analysis of organic 2, 4, 5-trichloroaniline," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 97, pp. 231–245, 2012.

[2]   A. M. Asiri, M. Karabacak, M. Kurt, and K. A. Alamry, "Synthesis, molecular conformation, vibrational and electronic transition, isometric chemical shift, polarizability and hyperpolarizability analysis of 3-(4-Methoxy-phenyl)-2-(4-nitro-phenyl)-acrylonitrile: a combined experimental and theoretical analysis," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 82, no. 1, pp. 444–455, 2011.

[3]   M. Lecuit, "Chloroquine and COVID-19, where do we stand?," *Medecine et maladies infectieuses*, vol. 50, no. 3. Elsevier, pp. 229–230, 2020.

[4]   O. Noureddine, N. Issaoui, S. Gatfaoui, O. Al-Dossary, and H. Marouani, "J. King Saud Univ." Sci, 2021.

[5]   M. Hagar, H. A. Ahmed, G. Aljohani, and O. A. Alhaddad, "Investigation of some antiviral N-heterocycles as COVID 19 drug: molecular docking and DFT calculations," *Int. J. Mol. Sci.*, vol. 21, no. 11, p. 3922, 2020.

[6]   R. A. Costa *et al.*, "Spectroscopic investigation, vibrational assignments, HOMO-LUMO, NBO, MEP analysis and molecular docking studies of oxoaporphine alkaloid liriodenine," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 174, pp. 94–104, 2017.

[7]   E. D. Glendening, C. R. Landis, and F. Weinhold, "Natural bond orbital methods," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 2, no. 1, pp. 1–42, 2012.

[8]   R. G. Parr, "Density functional theory of atoms and molecules," in *Horizons of quantum chemistry*, Springer, 1980, pp. 5–15.

[9]   N. Özdemir, S. Dayan, O. Dayan, M. Dinçer, and N. Ö. Kalaycıoğlu, "Experimental and molecular modeling investigation of (E)-N-{2-[(2-hydroxybenzylidene) amino] phenyl} benzenesulfonamide," *Mol. Phys.*, vol. 111, no. 6, pp. 707–723, 2013.

[10]   R. G. Parr, L. v Szentpály, and S. Liu, "Electrophilicity index," *J. Am. Chem. Soc.*, vol. 121, no. 9, pp. 1922–1924, 1999.
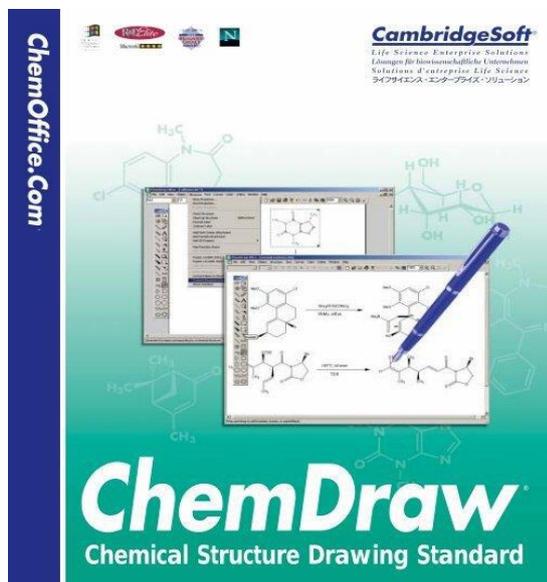
[11] R. G. Parr, "W. Yang Density functional theory of atoms and molecules," *Oxford Univ. Press*, vol. 1, p. 1989, 1989.

[12] J. F. Janak, "Proof that∂ E∂ n i= ε in density-functional theory," *Phys. Rev. B*, vol. 18, no. 12, p. 7165, 1978.

[13] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz Jr, "Density-functional theory for fractional particle number: derivative discontinuities of the energy," *Phys. Rev. Lett.*, vol. 49, no. 23, p. 1691, 1982.

[14] R. G. Parr and W. Yang, "Density functional approach to the frontier-electron theory of chemical reactivity," *J. Am. Chem. Soc.*, vol. 106, no. 14, pp. 4049–4050, 1984.

[15] R. G. Parr and R. G. Pearson, "Absolute hardness: companion parameter to absolute electronegativity," *J. Am. Chem. Soc.*, vol. 105, no. 26, pp. 7512–7516, 1983.

[16] P. K. Chattaraj and S. Giri, "Stability, reactivity, and aromaticity of compounds of a multivalent superatom," *J. Phys. Chem. A*, vol. 111, no. 43, pp. 11116–11121, 2007.

[17] R. R. Contreras, P. Fuentealba, M. Galván, and P. Pérez, "A direct evaluation of regional Fukui functions in molecules," *Chem. Phys. Lett.*, vol. 304, no. 5–6, pp. 405–413, 1999.

[18] C. Morell, A. Grand, and A. Toro-Labbé, "New dual descriptor for chemical reactivity," *J. Phys. Chem. A*, vol. 109, no. 1, pp. 205–212, 2005.

[19] M. Ouassaf, S. Belaidi, S. Chtita, T. Lanez, F. Abul Qais, and H. Md Amiruddin, "Combined molecular docking and dynamics simulations studies of natural compounds as potent inhibitors against SARS-CoV-2 main protease," *J. Biomol. Struct. Dyn.*, pp. 1–10, 2021.

Appendices

Appendix A: Chemical structure, molecular formula, and smiles of the selected compounds.

N.	Compound	Structure	Molecular formula	Smiles
1.	**Diosgenin**		$C_{27}H_{42}O_3$	CC1CCC2(C(C3C(O2)CC4C3(CCC5C4CC=C6C5(CCC(C6)O)C) C)C)OC1
2.	**Ladanein**		C17H14O6	COC1=CC=C(C=C1)C2=CC(=O)C3=C(C(=C(C=C3O2)OC)O)O
3.	**Protodioscin**		C51H84O22	CC1C2C(CC3C2(CCC4C3CC=C5C4(CCC(C5)OC6C(C(C(C(O6)CO)OC7C(C(C(C(O7)C)O)O)O)O)O)OC8C(C(C(C(O8)C)O)O)O)C)OC1(CCC(C)COC9C(C(C(C(O9)CO)O)O)O)O)O
4.	**Thymoquinone**		C10H12O2	CC1=CC(=O)C(=CC1=O)C(C)C
5.	**Trigonelline**		C7H7NO2	C[N+]1=CC=CC(=C1)C(=O)[O-]

Appendix B: some Computer-aided drug design software.