DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

MOHAMED KHIDER UNIVERSITY OF BISKRA

FACULTY OF EXACT SCIENCES, SCIENCES OF NATURE AND LIFE

DEPARTMENT OF MATTER SCIENCES

DISSERTATION

Presented by

Saghiri Khadijah

To obtain the degree of

Doctorate in Chemistry

Option :
Theorical and Computational Chemistry

Entitled:

**Contribution to the Modeling of Biomolecules and Their Interactions:
Inhibition of Enzymes Involved in Cancer Diseases by a New Class of
Derivatives.**

Publicly defended on: 26/06/2024

In front of the jury composed of:

Mr. KENOUCHE Samir	MCA	University of Biskra	President
Mr. MELKEMI Nadjib	Prof.	University of Biskra	Supervisor
Mr. DAOUD Ismail	Prof.	University of Biskra	CO-Supervisor
MS. Laib Souhaila	MCA	University of Batna 2	Examiner
MS. Kamouli Saida	MCA	University of Biskra	Examiner

This dissertation is dedicated to my mother, who has always carried the weight of all my burdens, providing unwavering love and support throughout my life.

To my father, for his unwavering support.

To my brothers, sisters, uncles, and aunts, for their love and encouragement.

To all my friends and beloved...

Acknowledgments

All praise is due to Allah, the Most Gracious, the Most Merciful. I am deeply grateful for the strength, knowledge, ability, and opportunity to complete my degree.

I would like to express my deepest gratitude to the following individuals who have played a significant role in the completion of my degree

I express my sincere gratitude to Prof. MELKEMI Nadjib for his solid support, guidance, and patience throughout the entire process of this dissertation. His expertise and encouragement have been invaluable in shaping this work. I am truly appreciative of his consistent support and mentorship, which have played a pivotal role in my academic journey. His belief in my abilities and his unwavering dedication have motivated me to strive for excellence. I extend my heartfelt thanks to him for being an exceptional mentor.

I extend my heartfelt gratitude to Dr. Daoud Ismail for his invaluable guidance, which has greatly enriched my academic journey. I am truly appreciative of his willingness and interest to assist me in overcoming academic challenges. His kindness, humility, and encouragement have been a continuous source of inspiration and strength throughout my academic pursuit. I express my deepest appreciation to him for his exceptional guidance and support in his roles as both a mentor and co-supervisor.

Special thanks are extended to Dr. KENOUCHE Samir for his invaluable contribution to completing a part of this dissertation. His generosity in sharing his knowledge and valuable skills has been critical to my academic journey. I also want to express my gratitude for his willingness to serve as the president of the committee.

I would like to thank my committee members Dr. Laib Souhaila from Batna University, and Dr. Khamouli Saida from the University of Biskra for agreeing to judge this work.

I am deeply grateful to my friend, DJEBAILI Rachida, for her unwavering support and for always being there when I needed a helping hand.

I would like to express my sincere gratitude to my friends Imane Almi and Khelfaoui Hadjer for their help and encouragement. Thank you both for being exceptional colleagues.

I am also grateful to my colleagues Rania Kherachi, and Matai Merzaka for their encouragement over the years.

I express my gratitude to the Department of Matter Sciences and Mohamed Khider University of Biskra for providing me with the opportunity to pursue my Ph.D. studies.

Special thanks to my dearest friends Imane, Zahra, and Ibtissem, your support and

encouragement have been a source of inspiration throughout the ups and downs of this academic journey.

I also extend my heartfelt thanks to my mother and siblings for their unwavering support and encouragement. Your unwavering support and encouragement have been a constant source of strength and motivation throughout my academic journey. I am deeply grateful for your love and guidance

.

Abstract

This dissertation presents the results of two research studies aimed at improving our understanding of breast cancer treatment drugs and potential targets. The first study used PLS regression to create QSAR models for 54 analogs of 2-phenyl-1H-indole, known for their anti-proliferative activity on MDA MB231 and MCF-7 cancer cell lines. The dataset was split into training and testing datasets 10,000 times, with 75% of the molecules used for training and the rest for external validation. The best models were selected based on the highest probability of occurrence according to the Bayesian information criterion. As a result, the PLS regression equation derived explains 6.79% and 63% of the variability in anticancer activity around its mean for model 1(MDA MB231), and model 2 (MCF-7), respectively. The leave-one-out cross-validation R^2_{CV}, the bootstrapping correlation coefficients R^2_{boots}, and the predicted R^2_{pred} indicated a high predictive power for both models. This study was accompanied by molecular docking/dynamics simulations, revealing that ligands L39, L40, and L48 fit into the pocket of estrogen-α receptor (PDB:1A52), while ligand L47 showed affinity with progesterone receptor (PDB:1A28). This affinity was confirmed by high negative score values and the establishment of several non-covalent interactions with the active site residues of both receptors. Furthermore, drug-likeness and ADME prediction analyses showed favorable absorption and oral bioavailability characteristics for ligands L39 and L48, suggesting their potential as precursor compounds for breast cancer drug development.

The second study aims to identify the binding mechanism of Glutaminase C (GAC) as a potential target for triple-negative breast cancer (TNBC). Molecular docking was employed to explore the interaction of 26 Withangulatin A (WA) derivatives with the allosteric site of GAC. The molecular docking/dynamics simulation results revealed that compounds A5, A8, A13, and A18 show high affinity toward the allosteric pocket of the GAC (PDB:3UO9), as confirmed by the high negative score values. These compounds interact with the most important residues and suggest a similar binding mechanism to the native compound (BPTES) and the clinical trial drug (CB-839). The combination of MEP analysis and molecular docking/dynamics studies confirms the favorable reactive sites of these compounds. Finally, pharmacokinetics prediction showed that A8 and A13 present the best ADMET profile among the selected compounds.

Keywords: Breast cancer, 2-phenyl-1H-indole, Withangulatin A, allosteric site, molecular docking, molecular dynamic, MEP, QSAR, PLS, ADME.

تعرض هذه الأطروحة نتائج دراستين بحثيتين تهدفان إلى تحسين فهمنا لأدوية علاج سرطان الثدي والأهداف البيولوجية المحتملة. استخدمت الدراسة الأولى انحدار PLS لإنشاء نماذج العلاقة بين الهيكل والنشاط QSAR لـ 54 نظيرًا لـ 2-phenyl-1H-indole، المعروف بنشاطه المضاد للتكاثر على خطوط الخلايا السرطانية MDA MB231 وMCF-7. تم تقسيم مجموعة البيانات إلى مجموعات بيانات تدريب واختبار 10000 مرة، مع استخدام 75% من الجزيئات للتدريب والباقي للاختبار الخارجي. تم اختيار أفضل النماذج على أساس أعلى احتمالية لحدوث وفقا لمعيار المعلومات بايزي (BIC) . ونتيجة لذلك، تفسر معادلة انحدار PLS المشتقة 6.79% و63% من التباين في النشاط المضاد للسرطان للنموذج 1 (MDA MB231)، والنموذج 2 (MCF-7) حول متوسطه ،على التوالي. أشارت R^2_{cv} للتحقق من صحة الإجازة الواحدة، ومعاملات الارتباط التمهيدية R^2_{boots}،وR^2_{pred} المتوقعة إلى قوة تنبؤية عالية لكلا النموذجين. كانت هذه الدراسة مصحوبة بمحاكاة الالتحام الجزيئي/الديناميكيات، مما يكشف أن الروابط L39 وL40 وL48 تتلاءم مع جيب مستقبلات هرمون الاستروجين-α (PDB:1A52)، بينما أظهرت الروابط L47 تقاربًا مع مستقبلات البروجسترون (PDB:1A28). تم تأكيد هذا التقارب من خلال قيم الدرجات السلبية العالية وإنشاء العديد من التفاعلات غير التساهمية مع بقايا الموقع النشط لكلا المستقبلين. علاوة على ذلك، أظهرت تحليلات ملاءمة الدواء والتنبؤ بـ ADME خصائص مواتية للامتصاص والتوافر البيولوجي عن طريق الفم للروابط L39 و L48، مما يشير إلى إمكاناتها كمركبات أولية لتطوير أدوية سرطان الثدي.

تهدف الدراسة الثانية إلى تحديد آلية الارتباط للجلوتاميناز سي (GAC) كهدف محتمل لسرطان الثدي الثلاثي السلبي (TNBC). تم استخدام الالتحام الجزيئي لاستكشاف تفاعل 26 من مشتقات Withangulatin A (WA) مع الموقع الخيفي لـ GAC. كشفت نتائج محاكاة الالتحام الجزيئي/الديناميكيات أن المركبات A5 و A8 و A13 و A18 تظهر تقاربًا عاليًا تجاه الجيب الخيفي لـGAC (PDB:3UO9) ، كما تؤكده قيم الدرجات السلبية العالية. تتفاعل هذه المركبات مع البقايا الأكثر أهمية وتقترح آلية ربط مماثلة للمركب الأصلي (BPTES) ودواء التجارب السريرية (CB-839) يؤكد الجمع بين تحليل MEP ودراسات الالتحام الجزيئي/الديناميكيات على المواقع التفاعلية المفضلة لهذه المركبات. أخيرًا، أظهر تنبؤ الحرائك الدوائية أن A8 و A13 يمثلان أفضل ملف تعريف ADMET بين المركبات المختارة.

الكلمات المفتاحية: سرطان الثدي، 2-فينيل-H1-إندول، withangulatin a، موقع تفارغي، الالتحام الجزيئي، الديناميكية الجزيئية، MEP، QSAR، PLS، ADME.

TABLE OF CONTENTS

CHAPTER Ⅰ : Breast Cancer and Molecular Targets

<div align="center">

CHAPTER II:

Background on Computer-Aided- Drug Design (CADD) Methods

</div>

Chapter III:

QSAR Study, Molecular Docking/Dynamics Simulations and ADME Prediction of 2-Phenyl-1H-Indole Derivatives as Potential Breast Cancer Inhibitors

Chapter IV

Molecular Docking/Dynamics Simulations, MEP analysis, and Pharmacokinetics prediction of some Withangulatin A derivatives as Allosteric Glutaminase C Inhibitors in Breast Cancer

List of works

Publications

- **Saghiri Khadijah**, Ismail Daoud, Nadjib Melkemi, Mesli Fouzia. (2022). QSAR Study, Molecular Docking/Dynamics Simulations and ADME Prediction of 2-Phenyl-1H-Indole Derivatives as Potential Breast Cancer Inhibitors. Biointerface Research in Applied Chemistry. 13. 1-25. 10.33263/BRIAC132.154.

- **Saghiri, K˙**, Daoud, I., Melkemi, N., and Mesli, F. (2023). Molecular Docking/ Dynamics Simulations, MEP analysis, and Pharmacokinetics prediction of some Withangulatin A derivatives as Allosteric Glutaminase C Inhibitors in Breast Cancer. Chemical Data Collections, 101044

International conferences

- **Saghiri, k,** Melkemi, N, Daoud, S, and Kennouche, S. Quantitative structure-activity relationship (QSAR) study of 2-phenyl-1H-indole derivatives as an antitumor agent. 13th International Days of Theoretical and Computational Chemistry Biskra, Algeria, February 02-03, 2020 (poster presentation).

- **Saghiri, k,** Melkemi, N, Daoud, I. Abiza, F, Molecular Docking Study and Drug-Likeness of Benzimidazole Derivatives as Cytotoxic Agents. The 1st international Conference on Trends Methods in Analytical Chemistry" (TMAC2023) – Bechar, Algeria, March 2023 (oral presentation).

National conferences

- **Saghiri, k,** Melkemi, N. Quantitative structure-activity relationship (QSAR) study for a set of bioactive molecules of benzimidazole. 1st National Days of Chemistry and Its Applications., Batna, Algeria, November 27, 2019. (Poster presentation).

- **Saghiri, k**, Ben Akcha, R, Melkemi, N, Daoud, I, and Kennouche, S. Phytochemical study of the plant (marrubium vulgar L). Scientific national days on the biology of medicinal plants. Tébessa, Algeria, January 22, 2020. (Poster presentation).

- **Saghiri, k**, Melkemi, N, Daoud, I and Kennouche, S. Molecular Docking and drug-likeness study of indole derivatives as cytotoxic agents against MCF7 and MDA MB231 breast cancer cells. 2nd National Seminar of Chemistry-Biology Interfaces Sciences—Souk Ahras, Algeria, February 2021 (video-conference presentation).

- **Saghiri, k**, Melkemi, N, Daoud, I. Molecular docking/dynamics simulations and ADME prediction on aroyl hydrazones derivatives targeting hl-60 and molecular docking – Batna, Algeria, March 2022 (oral presentation).

- Elhadj D, B, Lebsir F, Bounaceur B, et **Saghiri, K**. Elucidation de Structure de Molecules Complexes Application aux Asphaltènes $1st$ evaluation of the biological activities of medicinal plants and molecular docking – Batna, Algeria, March 2022 (Poster presentation).

List of figures

Chapter I

Chapter II

Chapter III

Chapter IV

List of tables

List of main abbreviations

ADME: Absorption, Distribution, Metabolism, Elimination
B3LYP: Becke 3-Parameter Lee-Yang-Parr
BBB: Blood-Brain Barrier
BC: Breast Cancer
BIC: Bayesian Information Criterion
CADD: Computer Aided Drug Design
CNS: Central Nervous System
CV: Cross-Validation Method
CYP: cytochrome enzyme
DFT: Density Functional Theory
DM: Dipole Moment
GAC: Glutaminas C
GI: Gastrointestinal
GLS: Glutaminas
HBA: Hydrogen-Bond Acceptors
HBD: Hydrogen-Bond Donors
LOO: Leave-One-Out
MD: Molecular Dynamic
MEP: Molecular Electrostatic Potential
MOE: Molecular Operating Environment
MR: Molar Refractivity
n-ROTB: Number of Rotatable Bonds
NS: Negative Surface
OV: Overall Variance
PDB: Protein Data Bank
P-gp: Permeability Glycoprotein
PLS: Partial Least Squares Regression
PS: Positive Surface
QSAR: Quantitative Structure Relationship
RMSD: Root-Mean-Score Deviation
SDEP: Standard Deviation Error in Prediction
SF: Scoring Function
TPSA: Topological Polar Surface Area
U: Ootential E
nergy
WA: withangulatin A

General introduction

Breast cancer represents a global health concern affecting women worldwide from diverse socioeconomic backgrounds. According to the world health organization (WHO), it is estimated that there will be 22.3 million new cases and 685,000 deaths predicted in 2023. Women aged 40 to 60 are at higher risk, accounting for approximately 75% of all cases [1-3].

Research in breast cancer plays a critical role in various aspects of prevention, diagnosis, and therapy. Notably, research has contributed to a decrease in the number of women diagnosed with late-stage cancer. Furthermore, ongoing research efforts promote advancements in treatment techniques, leading to the development of more effective and less toxic treatments, as well as improved diagnostic and monitoring technologies [2, 4].

The aim of breast cancer research is to address global health disparities by increasing the inclusivity of clinical trials and improving access to care and treatment, particularly in low- and middle-income countries (LMICs). Advances in breast cancer prevention, detection, and treatment are made available to people worldwide [5].

Drug discovery is a comprehensive process aimed at identifying therapeutic compounds capable of curing and treating diseases. This process involves various stages, including candidate identification, synthesis, characterization, validation, optimization, screening, and assays for therapeutic efficacy. Once a compound shows promise in these investigations, it progresses to the drug development phase, which precedes clinical trials. The development of a new drug must undergo several stages to ensure it is safe, effective, and meets all regulatory requirements. The process of drug discovery and development is characterized by its lengthiness, complexity, and high costs. It involves the consideration of numerous biological targets for each new medicine ultimately approved for clinical use. New research tools may be required to investigate each new target effectively. From the initial discovery to the creation of a marketable medicine, this process takes approximately 10-15 years and requires an investment of around US$ 2-3 billion. On average, a million molecules are screened, but only one is explored in late-stage clinical trials and eventually made available to patients [6, 7].

In recent years, computer-aided drug design, also known as in silico methods, has emerged as a cost-effective approach to drug development. These methods involve using computer simulations to predict the efficacy of potential drugs, aiding researchers in identifying promising drug candidates and minimizing the need for expensive experiments. Several research endeavors have leveraged in silico methods to design novel bioactive compounds for addressing diverse cancer types, including breast cancer. Additionally, these methods have been instrumental in predicting the mechanism of action of specific bioactive agents against breast cancer. These computational approaches can help researchers make informed decisions about which compounds to advance to the next stage of development, ultimately saving resources and accelerating the drug discovery process [7-10].

In this work, we utilized in silico methods to predict the mechanism of actions of novel compounds against breast cancer. These techniques involved quantitative structure-activity relationship (QSAR) methods, molecular docking/dynamics simulations, quantitative molecular electrostatic potential (MEP) analysis, and pharmacokinetic/drug-likeness prediction. This strategy allowed us to conserve effort, time, and resources that might have been expended on unproductive in vitro and in vivo experiments. The outcomes of this research could elucidate the mechanisms of action of these compounds and affirm their reported therapeutic potential against breast cancer.

This dissertation is structured into four chapters:

Chapter 1: This chapter provides a bibliographic overview of the biological aspects related to breast cancer, covering signs and symptoms, etiology, risk factors, diagnosis, and treatment options. It also explores the histological classifications of breast cancer and its developmental process, alongside various treatment modalities.

Chapter 2: The second chapter is dedicated to discussing the fundamental methodologies and techniques utilized in our research, which include ligand-based drug design and structure-based drug design. This encompasses the application of QSAR methods, molecular docking and dynamics, as well as ADMET predictions and MEP analysis. In the subsequent chapters, we will present and analyze the outcomes of our research.

Chapter 3: This chapter focuses on the application of Partial Least Squares (PLS) regression to establish robust quantitative structure-activity relationships (QSAR) models for a series of 2-phenyl-1H-indole derivatives targeting breast cancer cells. The Bayesian Information Criterion (BIC) method was employed to identify the optimal model. Additionally, molecular docking/dynamics simulations were conducted to evaluate the binding modes of these compounds within the active sites of selected targets. Finally, ADME/Pharmacokinetics properties were assessed to identify the most promising candidate molecules with minimal oral administration problems.

Chapter 4: In this chapter, we aim to understand the binding mechanism and intermolecular interactions of novel Withangulatin A derivatives with the allosteric site of GAC using in silico analyses. Molecular docking/dynamics simulations were conducted to assess the binding mode of the compounds into the allosteric site of the target. MEP analyses were further used to predict the electron density distribution and understand the sites of electrophilic and nucleophilic attack. Finally, pharmacokinetic properties were evaluated to predict the best candidate compound with low oral administration problems.

References

1.   Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review. Cancers. 2021;13(17):4287.

2.   Breast cancer [Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

3.   Wilkinson L, Gathani T. Understanding breast cancer as a global health concern. The British Journal of Radiology. 2022;95(1130):20211033.

4.   Jallah JK, Dweh TJ, Anjankar A, Palma O, Dweh III TJ. A Review of the Advancements in Targeted Therapies for Breast Cancer. Cureus. 2023;15(10).

5.   Anderson BO, Cazap E, El Saghir NS, Yip C-H, Khaled HM, Otero IV, et al. Optimisation of breast cancer management in low-resource and middle-resource countries: executive summary of the Breast Health Global Initiative consensus, 2010. The lancet oncology. 2011;12(4):387-98.

6.   Deore AB, Dhumane JR, Wagh R, Sonawane R. The stages of drug discovery and development process. Asian Journal of Pharmaceutical Research and Development. 2019;7(6):62-7.

7.   Malik JA, Jan R, Ahmed S, Anwar S. Breast cancer drug repurposing a tool for challenging disease. Drug Repurposing: Molecular Aspect and Therapeutic Applications; Intech Open: London, UK. 2022:121-30.

8.   Muhammad A, Katsayal BS, Forcados GE, Malami I, Abubakar IB, Kandi AI, et al. In silico predictions on the possible mechanism of action of selected bioactive compounds against breast cancer. In Silico Pharmacology. 2020;8:1-13.

9.   Rajagopal K, Kalusalingam A, Bharathidasan AR, Sivaprakash A, Shanmugam K, Sundaramoorthy M, et al. In Silico Drug Design of Anti-Breast Cancer Agents. Molecules. 2023;28(10):4175.

10.   Rahman MM, Karim MR, Ahsan MQ, Khalipha ABR, Chowdhury MR, Saifuzzaman M. Use of computer in drug design and drug discovery: A review. International Journal of Pharmaceutical and Life Sciences. 2012;1(2).

CHAPTER I:

BREAST CANCER

AND

Molecular TARGETS

1. Introduction

Breast cancer (BC) represents a significant global health challenge as it is the most frequently diagnosed cancer worldwide. According to the GLOBOCAN database, over 2.3 million cases were reported in 2020 and it an estimated that by the year 2040, the incidence of newly diagnosed breast cancers will increase by more than 40%, reaching approximately 3 million cases annually [1]. BC is characterized by its heterogeneity, starting in breast tissue when the cells undergo uncontrolled growth and multiplication, eventually forming a tumor. It is noteworthy that breast cancer can affect both men and women, although it is more frequently diagnosed in women and stands as the leading cause of cancer-related deaths among females [2].

Understanding breast anatomy and hormonal physiology is critical in comprehending breast cancer development and all the factors that may contribute to its initiation. The breast is an ectodermal organ, conventionally found in mammals, and a highly specialized type of soporiferous gland in humans. Breasts are located on the front part of the chest and are also called mammary glands (Figure I.1). Both men and women have breasts that are mostly constituted of fatty adipose tissues; however, female breasts contain more glandular tissues as compared to males. The posterior surface of the breast is located on the pectoralis major fasciae segment, rectus abdominis muscles, external oblique abdominal muscles, and anterior serratus. It is attached to the skin by the Cooper suspensory ligaments and is separated from the investing fascia of the pectoralis major muscles by the retro-mammary bursa. The loose areolar tissue occupies the retromammary bursa, allowing the breast to move freely against the thoracic wall with the assistance of suspensory ligaments. The breast tissue and the stroma are made up of epithelial parenchymal cells containing lobules with a glandular structure that produces milk in females when activated. The breast contains lobes, which are further split into lobules made of tubuloalveolar glands. The content of lobules is conveyed to the nipple through the connecting channels called ducts. The space between glandular tissue and ducts contains fatty tissue and connective tissue [3-5].

Breast development occurs in several stages, and this process is controlled by hormones and growth factors such as estrogen, progesterone, prolactin, and growth hormones. Numerous

changes occur in the breast during puberty, pregnancy, lactation, and the menstrual cycle due to fluctuations in the levels of these hormones [6].

During puberty, the mammary gland develops in response to the hormones of estrogen and progesterone. Estrogen stimulates the growth and branching of the canal system and the proliferation of adipose tissue, while progesterone works with estrogen to differentiate the lobular structures. During pregnancy, the mammary gland matures and becomes capable of milk production. This occurs in preparation for breastfeeding. At menopause, the mammary gland undergoes involution, which is the regression of lobular cells, leading to decreased secretory activity in the lobules [7].
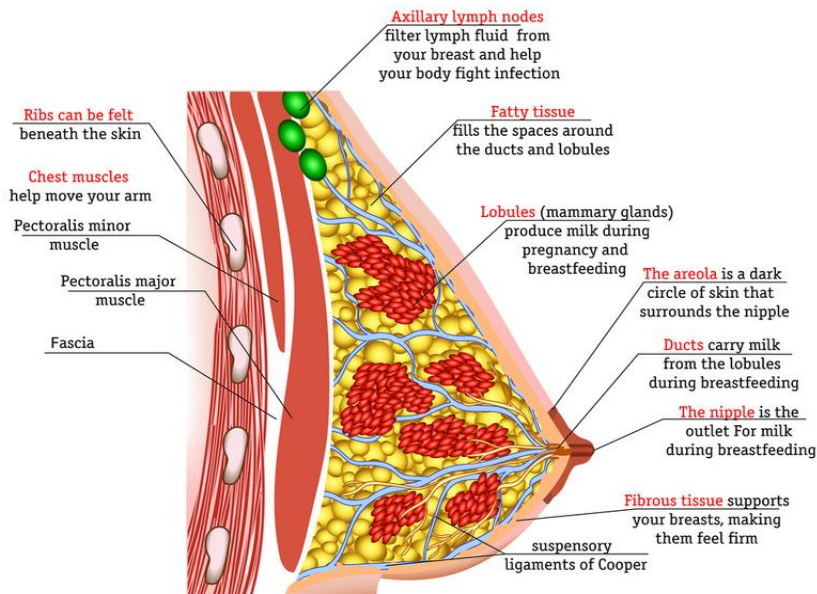
Figure I.1. Anatomy of female breast [8].

2. Signs and Symptoms

Breast cancer, a prevalent and potentially life-threatening condition, often presents various signs and symptoms that require prompt attention and medical evaluation for effective treatment and improved outcomes. Recognizing these signs and symptoms is imperative for early detection and timely intervention in breast cancer management [9].

➢ Early detection is crucial, and the most common breast complaint leading to medical advice is the detection of a breast mass.

➢ Approximately 90 percent of all breast masses are benign lesions, with smooth and rubbery masses typically associated with fibroadenoma in women in their 20s and 30s or cysts in women in their 30s and 40s.

➢ Malignancies may manifest as erythema, edema, and skin or nipple retraction.

➢ Nipple discharge, particularly spontaneous and bloody, linked to amass and confined to a single duct in one breast, is a noteworthy symptom of breast cancer.

➢ Suspicious breast masses associated with cancer are often solitary, discrete, and firm, potentially fixed to the skin or muscle.

➢ Unilateral and non-tender characteristics are common in such masses, and sometimes, thickened areas not forming discrete masses may also signal breast cancer.

➢ Bilateral occurrences of breast cancer at the initial diagnosis are rare.

3. Etiology and Risk Factors

The etiology of breast cancer is highly intricate, involving various endogenous and exogenous factors, along with the influence of genetic factors [10]. Detecting elements linked to a higher occurrence of breast cancer is essential for women's overall health screening. Breast cancer risk factors can be classified into seven major categories [11, 12]:

- **Gender:** Breast cancer is a disease that is unique to women and is a rare malignancy in men, accounting for less than 1% of all cases of cancer. Breast cancer occurs more often in older adult males who have had hormonal imbalances, exposure to radiation, and a family history of breast cancer, and the most common risk factor for this disease among men is the mutation of the BRCA2 gene [13].

- **Age:** Women over the age of 40 have a higher incidence of breast cancer. However, breast tumors in younger women are often larger, diagnosed at advanced stages, involve positive lymph nodes, and have weaker survival rates. The majority of breast cancers are detected in women aged 40 and older [14].

- **Genetic and family history:** Women with a first-degree relative who has had breast cancer are at an increased risk of developing the disease themselves. The risk is further

elevated if the affected family member is diagnosed at a younger age. For instance, a woman with a first-degree relative diagnosed before the age of 40 has a higher risk of being diagnosed with breast cancer compared to a woman of the same age without a family history of breast cancer. BRCA1 and BRCA2 genes are associated with hereditary breast cancer, approximately 40% of hereditary breast cancer cases are due to mutations in these genes, which are inherited dominantly through the autosomal method. These BRCA mutations are especially relevant to breast cancer occurring in younger premenopausal women [15].

- **Hormonal exposure:** Hormonal factors, including contraceptive methods, ovulation-stimulating drugs, and postmenopausal hormone therapy, can influence the risk of developing breast cancer. Some studies suggest that the use of oral contraceptive pills is associated with an increased risk of breast cancer [16], while others found no significant association [17]. Ovulation-stimulating medications for more than 6 months may increase the risk, but this finding is not consistent across all studies. Postmenopausal hormone therapy, especially with estrogen-progestin combinations, is linked to an elevated risk of breast cancer, but the risk diminishes after discontinuation. Factors such as early puberty, late menopause, nulliparity, and late first pregnancy may also contribute to increased breast cancer risk, while breastfeeding and multiparity may act as protective factors.

- **Lifestyle:** Several lifestyle factors have been associated with an increased risk of breast cancer, including but not limited to, obesity, alcohol consumption, smoking, and lower amount of physical activity [5].

4. Hormone receptors

Receptors represent a distinct group of proteins that operate by attaching to a particular ligand molecule. Upon the binding of a ligand to its receptor, the receptor changes shape, initiating a signal within the cell [18]. Hormone receptor-positive breast cancers are characterized by the presence of estrogen and progesterone receptors on the cancer cells. These receptors promote cancer cell growth when bound to their respective hormones [19]. Approximately 70% of breast cancers fall into the hormone receptor-positive category, characterized by positive expression of estrogen receptor (ER) and/or progesterone receptor (PR). The presence of these receptors is closely linked to the growth and spread of cancer

cells. Specifically, estrogen, when interacting with its receptor ER, plays a pivotal role in the progression of breast cancer [20].

4.1. Estrogen receptors

Estrogen is a steroid compound and one of the most important female sex hormones, primarily synthesized by the ovaries and placenta in females. Estrogens play crucial roles in modulating physiological and pathophysiological processes, predominantly through interactions with estrogen receptors (ERs). ERs are proteins located within cells that play a crucial role in the growth and development of breast tissue, serving as key biomarkers in breast cancer [21]. There are two different ER isoforms, alpha (ERα) and beta (ERβ), encoded by two separate genes, ESR1 and ESR2, respectively [22]. These receptors are present in both normal and cancerous breast cells. In normal breast tissue, estrogen binds to these receptors, triggering a series of cellular responses. These responses hold the promotion of cell growth and division, assist in the development and maintenance of female sexual characteristics, and influence bone healing processes [23].

4.2. Progesterone receptors

The progesterone receptor (PR) is a cellular protein triggered by the hormone progesterone, regulating gene expression, and controlling the hormone's actions in a receptor-dependent manner. PR proteins are present in numerous human tissues, including the mammary gland, the uterus, bone, brain, ovary, testes, and lower urinary tract tissues [24]. The classification of cancer cells hinges on the existence or nonexistence of PR, a pivotal element in comprehending the hormone receptor status essential for efficacious treatment. This differentiation is crucial as it impacts how cancer cells react to progesterone. Tailored strategies for successful breast cancer treatment, like hormone therapy, are designed in alignment with the presence or absence of progesterone receptors. Hormone therapy aims to either decrease the body's estrogen levels or disrupt the normal functioning of cancer cells, depending on the PR status [24].

5. Histological types of breast cancer

Breast cancer is divided into several categories based on various factors such as the location of the tumor, its size, whether it has spread to nearby lymph nodes or other parts of the body, and specific characteristics of the cancer cells. It is commonly classified into two

main types: invasive (infiltrating breast cancer) and noninvasive (in situ breast cancer) as represented in Figure I.2. These classifications indicate whether cancer has progressed to other places of the body or is restricted to the site of genesis [2].

5.1. Invasive (infiltrating breast cancer)

Based on the tissue and cell types involved, invasive breast cancers are divided into the following two types [25]:

a) Invasive Lobular Carcinoma (ILC):

Lobular carcinoma (LC) is the second most frequent type of breast cancer, comprising around 10-15% of all breast cancer cases. While ILC can develop at any age, it is more prevalent among older women. Compared to invasive ductal carcinoma (IDC), ILC typically occurs later in life, with onset often observed in the early 60s, as opposed to the mid-to-late 50s for IDC.

b) Invasive Ductal Carcinoma (IDC):

Ductal carcinoma (DC) is the prevalent form of breast cancer, accounting for approximately 80% of all breast cancer cases. The classification of invasive ductal carcinomas (IDC) comprises various subtypes, such as tubular carcinoma, medullary carcinoma, mucinous carcinoma, papillary carcinoma, and cribriform carcinoma of the breast.

5.2. Noninvasive (in situ breast cancer)

The cancer cells that are inside the milk ducts have not yet spread, either through the ducts into surrounding breast tissue or to other parts of the body (known as 'in situ'). There are two primary types of noninvasive breast cancer [2]:

a) Lobular Carcinoma in situ (LCIS)

LCIS is marked by a collection of small irregular cells with diminutive nuclei, distinct individual acini, and a lack of cellular cohesion. This condition entails the presence of abnormal cells within the milk-producing lobules of the breast. LCIS is considered a non-malignant high-risk lesion of breast cancer, suggesting that it serves as an indicator that a woman faces an increased risk of developing breast cancer in the future [26].

b) Ductal carcinoma in situ (DCIS)

Ductal carcinoma in situ (DCIS) involves the uncontrolled growth of malignant epithelial cells within the mammary ducts without spreading into the surrounding breast tissue [27]. If DCIS is not treated, the cells may develop the ability to spread and become invasive breast cancer [28].
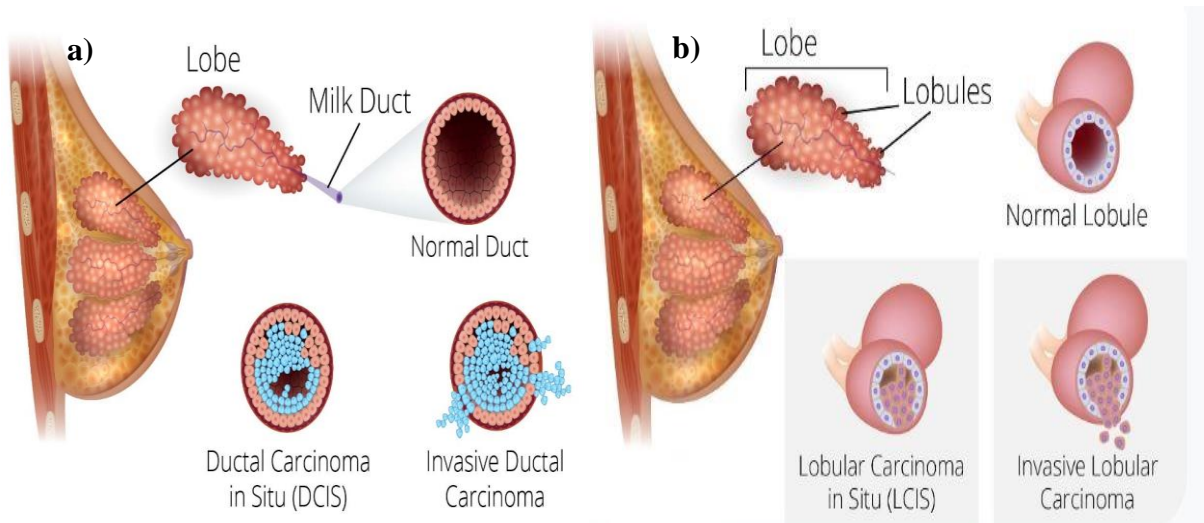
Figure I.2. In situ and invasive breast cancer (a) ductal carcinoma and (b) lobular [29].

6. Breast cancer molecular subgroups

Breast cancer is not a single disease but rather a group of distinct subtypes characterized by different molecular profiles. Understanding these molecular subgroups is critical for tailoring treatment strategies and predicting patient outcomes.

In clinical practice, the most commonly used molecular classification of BC is based on the expression of an estrogen receptor (ER), progesterone receptor (PR), a receptor for human epidermal growth factor 2 (HER2) in cancer cells, and Ki67 protein, which reflects proliferation levels. Currently, four molecular types of BC are distinguished: luminal type A, luminal type B, HER2-positive non-luminal, and triple-negative BC (TNBC) [30, 31]. These molecular subtypes might respond differently to particular treatments, which has unique clinical implications.

6.1. Luminal A

Luminal A breast cancers are generally associated with low metastatic potential and good prognosis. However, there is a proportion of patients who present with metastases in lymph nodes. It is recognized as hormone receptor-positive (ER +, PR +, HER2-, low Ki67), and has low proliferation rates. They are often associated with a good prognosis and respond well to hormone therapy [32].

6.2. Luminal B

Luminal B is a subtype of breast cancer within the luminal category, which is characterized by the presence of hormone receptors (ER+ and/or PR+). Luminal B cancers are more aggressive than Luminal A and may require more intensive treatment, often including chemotherapy. They are further divided into two subcategories [6]:

6.2.1. HER2- luminal type B

Luminal B (HER2-) tumors may exhibit high levels of Ki-67, a marker for cell proliferation, indicating that these tumors are actively dividing and growing [33].

6.2.2. Luminal B/HER2 +

Luminal B/HER2+ breast cancer represents a subtype that is characterized by the presence of both estrogen receptors (ER+) and overexpression of the HER2 protein (HER2+). The treatment of Luminal B/HER2+ breast cancer typically involves targeting both the hormonal receptors (ER) and the HER2, which may include hormone therapy and HER2-targeted therapies such as trastuzumab and pertuzumab [34].

6.3. HER2- positive (non-luminal)

HER2-positive breast cancer, often referred to as HER2+ breast cancer, is a specific molecular subtype of breast cancer characterized by the overexpression or amplification of the HER2 (Human Epidermal Growth Factor Receptor 2) protein. This subtype represents a distinct category of breast cancer with unique biological features and, in the past, was associated with a more aggressive disease course. However, advances in targeted therapies have significantly improved the prognosis for HER2-positive breast cancer patients [35]. HER2-positive breast cancer is not a homogeneous group. It can further be classified into subtypes based on hormone receptor status (ER and PR). These subtypes include:

➢ HER2+/ER+ (HER2-positive and estrogen receptor-positive)

➤ HER2+/ER- (HER2-positive and estrogen receptor-negative)

The treatment for HER2-positive breast cancer typically includes HER2-targeted therapies, such as trastuzumab, in combination with chemotherapy. Hormone therapy may also be considered for HER2+/ER+ cases. Chemotherapy and radiation therapy may be used depending on the stage and extent of the disease [36].

6.4. Basal-like (Triple-Negative)

Basal-like breast cancers are typically hormone receptor-negative (ER-/PR-) and HER2-negative (HER2-), thus referred to as triple-negative breast cancer (TNBC). They are characterized by a gene expression profile similar to basal cells in the mammary gland. TNBC breast cancers tend to be aggressive, are more likely to metastasize, and may have a worse prognosis. Treatment for this subtype often involves chemotherapy [37].

7. Development of breast cancer

Breast cancer originates from a disruption in the normal growth and division of cells in the body, where cells undergo controlled replication and eventually die after several cycles. The beginning of breast cancer is marked by abnormalities in the growth of cells within the breast, leading to alterations in the texture of breast tissues [38]. Typically, this abnormal growth occurs in the inner lining of milk ducts or lobules. Figure I.3 illustrates a schematic of breast cancer progression. As can be seen, the standard unit of breast ductal lobules, known as the terminal ductal lobular unit (TDLU), consists of lobules and ducts with luminal and myoepithelial cells in a bi-layered epithelium. It highlights atypical ductal hyperplasia (ADH) as a premalignant lesion with abnormal cell layers, serving as the precursor to ductal carcinoma in situ (DCIS), a non-invasive lesion containing abnormal cells. The figure emphasizes the increasing risk of developing malignant or invasive breast cancer (IBC) at each stage, with DCIS potentially progressing to IBC. However, predicting the progression of lesions remains unclear. Once cells invade, the risk of metastasis significantly rises, with lymph nodes identified as the primary site for breast cancer metastasis (MET). The schematic illustrates the loss of control over cell proliferation, survival, differentiation, and migration during the multistage process of breast cancer progression.

Abnormal interactions between tumor and stromal cells facilitate the formation of metastases, which involves a series of steps such as invading through the basement membrane, entering the vasculature (intravasation), surviving without adhesion, exiting the vasculature (extravasation), and establishing a new tumor in a foreign microenvironment. Furthermore, Figure I.3 highlights the proposed parallels between normal breast stem or progenitor cells and cancer cells, suggesting similarities in dormancy, self-renewal, and differentiation capabilities. This leads researchers to propose that cancer cells with stem cell-like characteristics, often termed 'cancer stem cells' or 'tumor-initiating cells,' drive breast cancer initiation, progression, and recurrence. The figure shows this hypothesis as involving epigenetic and genetic alterations in different stem or progenitor cells, including long-term (LT), short-term (ST), and luminal or basal (myoepithelial) progenitors. These alterations give rise to different tumor subtypes consisting of various cell types (mixed, luminal, or basal lineage) with distinct gene-expression profiles and prognoses [39].
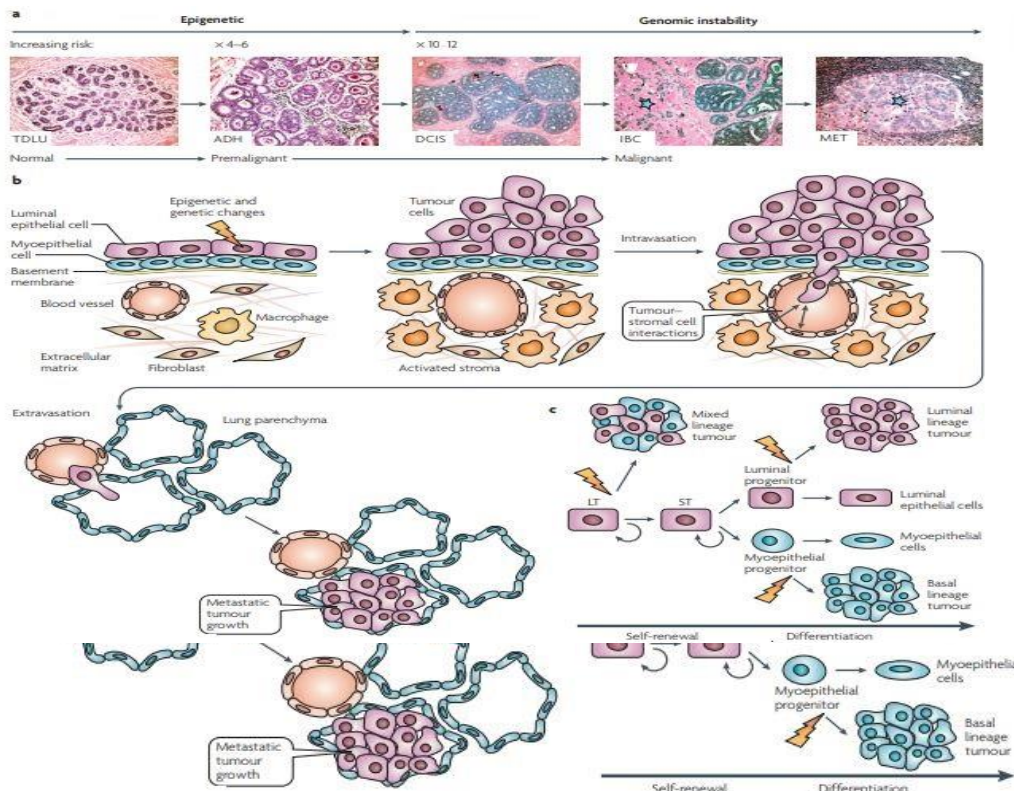
Figure I.3. Schematic representation of breast cancer development [39].

8. Breast cancer treatment

The treatment of breast cancer presents a challenge because of its heterogeneous nature. While various treatments exist, they may vary based on factors such as the type and stage of the cancer, hormone sensitivity, and patient age. Mostly, the main types of treatment for breast cancer include surgery, radiation therapy, hormone therapy, and, targeted therapy [40, 41]. In the next subsection, we will briefly focus on hormone therapy, chemotherapy, and targeted therapy.

As seen in table I.1, it outlines various breast cancer subtypes, providing their respective definitions and types of treatment. Each subtype is associated with specific characteristics and is recommended for different types of treatment, including endocrine therapy, cytotoxic, and anti-HER2 therapy, depending on the molecular profile of breast cancer.

Table I.1 Breast cancer subtypes, definition, and type of treatment [42].

Breast cancer subtype	Definition	Type of treatment
Luminal A	ER+ and PR+ HER2- low Ki-67	Endocrine (hormonal) therapy
Luminal B (HER2+	Luminal B (HER2-) ER+ HER2- and at least one of Ki-67 'high' PR 'Negative or low'	Endocrine (hormonal) therapy ± cytotoxic
	Luminal B-like) ER+ HER2 overexpressed or amplified; any Ki-67 any PR	Endocrine (hormonal) therapy + cytotoxic + anti HER2 therapy
HER2 overexpressing	HER2 overexpressed or amplified; ER and PR absent	Cytotoxic + anti-HER2 therapy
Triple-negative	ER- and PR-, HER2-	Cytotoxic

8.1. Endocrine (hormonal) therapy

Hormonal therapy is a commonly employed approach in the treatment of breast cancer, particularly in cases where tumors express estrogen receptor (ER) and progesterone receptor (PR). This type of therapy aims to block the growth of cancer cells that rely on hormones for their development. The most common drugs used in hormone therapy for breast cancer are

aromatase inhibitors, and antiestrogens, which work by blocking or decreasing the body's production of hormones [40, 43].

> **Aromatase inhibitors**

Aromatase inhibitors are a group of drugs that block the action of enzymes that are responsible for making estrogen in the ovaries and other tissues referred to as "aromatase." These inhibitors are used exclusively in postmenopausal patients with hormone-sensitive tumors. This is because, in postmenopausal women, the main source of estrogen is the conversion of androgens since the ovaries no longer produce it. Blocking aromatase lowers the amount of estrogen made by the body, which may stop the growth of cancer cells that need estrogen to grow. Examples of aromatase inhibitors are anastrozole, letrozole, and exemestane [44-46].

> **Antiestrogens**

Antiestrogens are drugs that can act as antagonists towards the actions of estrogens in the treatment of ER-positive breast cancers [47, 48]. Selective estrogen receptor modulators (SERMs) and selective estrogen receptor downregulators (SERDs) are the main types of antiestrogens. Tamoxifen, a well-known SERM, blocks estrogen from binding to breast cancer cells. Its efficacy extends beyond lowering the risk of recurrence in the affected breast to lowering the danger of developing cancer in the opposite breast as well as the possibility of distant recurrence. Tamoxifen is also approved for decreasing the risk of breast cancer in high-risk women and lowering the chance of local recurrence in women with ductal carcinoma in situ (DCIS) who had a lumpectomy [40]. However, it can result in notable side effects, including endometrial cancer and thromboembolic events [49, 50].

8.2. Chemotherapy

Chemotherapy for breast cancer involves the use of cytotoxic drugs and small molecules to specifically target and destroy breast cancer cells. Typically, these drugs are administered orally or through direct injection. The primary mechanism of cytotoxic agents is to induce DNA damage and restrict the replication of DNA in proliferating cells [51]. The cytotoxic drugs used in breast cancer treatment include carboplatin, cisplatin, doxorubicin, methotrexate, paclitaxel, and others. Patients may be given a single medicine or a combination of many drugs at the same time [40].

8.3. Targeted therapy (anti-HER2)

The human epidermal growth factor receptor (HER) family of receptors is important in the development of numerous human cancers. This family is made up of four major members: HER-1, HER-2, HER-3, and HER-4. They control cell growth, survival, and differentiation through many signal transduction pathways. Human epidermal growth factor receptor 2 (HER2) is a member of the epidermal growth factor receptor family with tyrosine kinase activity. Approximately 15-30% of breast cancers have HER2 amplification or overexpression [52]. HER2 has been effectively addressed in breast cancer and other cancers, including intestinal and ovarian cancers. It is being studied as a potential therapeutic target, with various approaches available to target HER2. These include the use of monoclonal antibodies such as trastuzumab and pertuzumab; tyrosine kinase inhibitors such as lapatinib and neratinib, and; antibody-drug conjugates such as ado-trastuzumab emtansine (TDM1) and trastuzumab (Herceptin), which have shown significant clinical activity in various trials [53].

Trastuzumab and pertuzumab are monoclonal antibodies that inhibit HER2 activity by binding to the extracellular domain of the HER2 receptor. Additionally, pertuzumab prevents the dimerization of HER2-HER3 receptors. In contrast, the small-molecule inhibitors lapatinib and neratinib bind to the intracellular tyrosine kinase domains of HER2 and other members of the HER family (Figure I.4) [43].

Despite the significant therapeutic benefits of anti-HER2 therapies, some HER2-positive patients may develop primary resistance, leading to disease progression of the disease and emphasizing the necessity for innovative therapeutic alternatives.
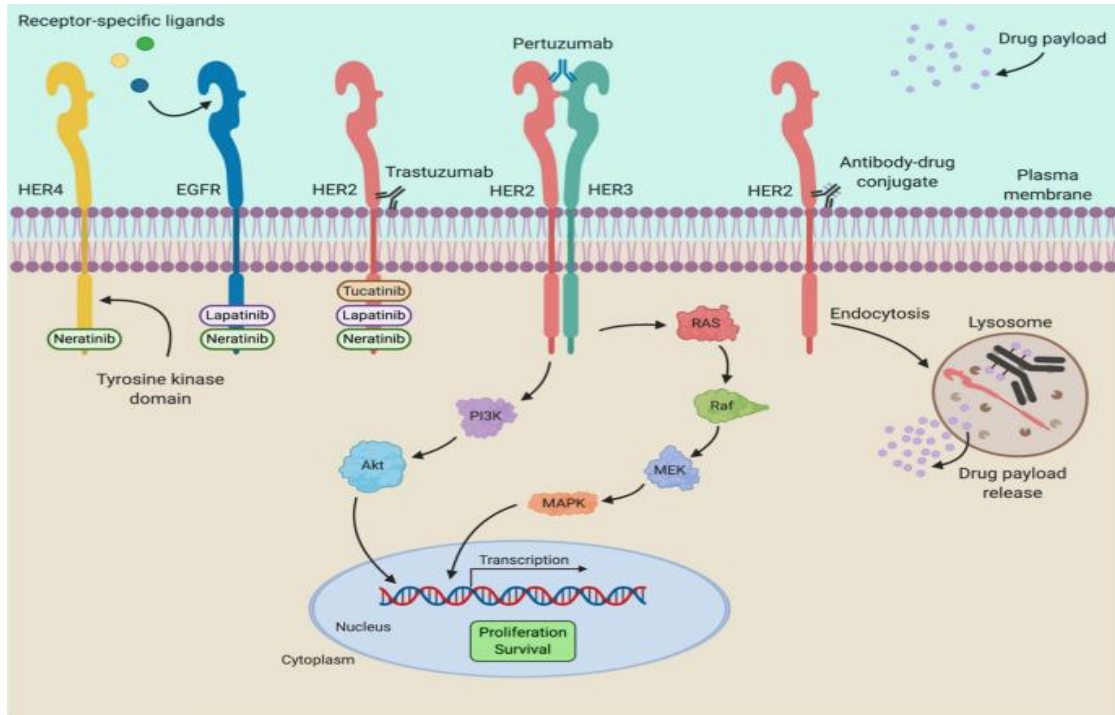
Figure I.4. Mechanism of action of agents targeting HER2 [53].

8.4. Glutaminase as a potential therapeutic target for therapeutic intervention

Glutaminase (GLS) is an essential mitochondrial enzyme crucial for breaking down glutamine (Gln), the most important amino acid in the human body. This enzymatic activity transforms Gln into glutamate, promoting cancer cell proliferation. Typically, differentiated cells primarily use mitochondrial oxidative phosphorylation to produce the necessary energy for cellular processes. However, most cancer cells undergo a significant reprogramming of cellular energy metabolism, adopting a modified form of aerobic glycolysis known as the "Warburg effect." This metabolic shift in tumor cells results in around 200 times more compared to normal cells. A significant amount of glucose carbon undergoes the Warburg effect and is excreted as lactate, which prevents it from being used in the mitochondrial tricarboxylic acid (TCA) cycle. As a result of this metabolic change, cancer cells become abnormally dependent on Gln [54, 55]. The figure below (Figure I.5) provides a summary of the metabolism of Gln in cancer cells.
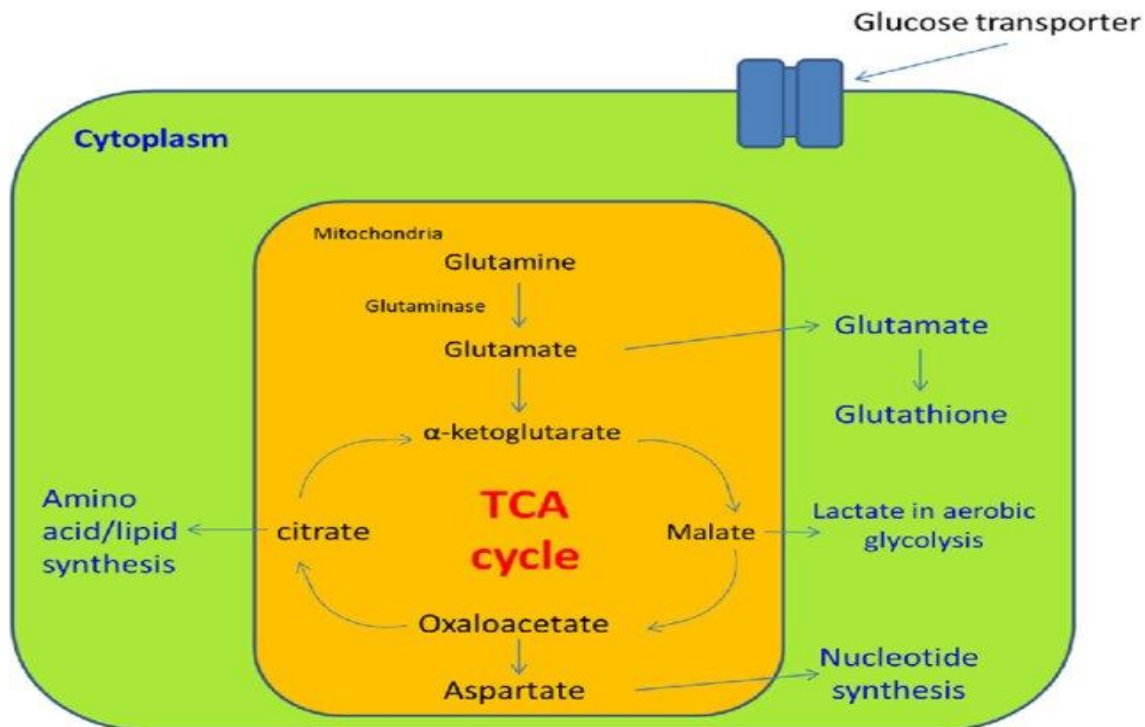
Figure I.5. The metabolism of Gln in cancer cells.

Cells undergo a series of metabolic processes involving Gln. Initially, Gln enters the cells via the Gln transporter. Once inside the mitochondria, glutaminase catalyzes its breakdown into glutamate. The resulting glutamate has two pathways, it can either be moved out to the cytoplasm or converted into glutathione. Within the mitochondria, glutamate is further converted into α-ketoglutarate and then enters the tricarboxylic acid (TCA) cycle. Malate, a product of the TCA cycle, is transported to the cytoplasm, where it is eventually converted into lactate through aerobic glycolysis, releasing energy. On the other hand, malate within the mitochondria can be converted into oxaloacetate, which can further transform into aspartate or citrate. Aspartate is transported out to the cytoplasm for nucleotide synthesis, while citrate formed from malate is transported out to the cytoplasm to contribute to amino acid and lipid synthesis [56].

GLS exhibits overexpression in numerous primary tumors including lung, breast, kidney, prostate, and colorectal. Consequently, the effectiveness of treating various tumor models with a potent GLS inhibitor or employing genetic silencing techniques to target GLS

has been affirmed as a viable therapeutic strategy for cancer [54, 57]. Indeed, GLS has two isoforms, kidney-type glutaminase (GLS1) and liver-type glutaminase (GLS2). On the other hand, GLS1 exists in two different isoforms: kidney-type (KGA/GLS1), and a spliced variation of kidney type known as glutaminase C or GAC [58]. Despite the functional distinctions between the two splicing variants of GLS remain unclear, it appears that the GAC variant of GLS plays a pivotal role in the modified metabolic characteristics observed in numerous rapidly dividing cells. Notably, GAC is expressed in acute myeloid leukemia, non-small cell lung cancer, and human breast cancer cell lines [58, 59].

The crystal structure of glutaminase C (GAC) is a tetramer, consisting of four subunits, with each subunit containing a Gln-binding site and one allosteric site per dimer-dimer [60]. Figure I.5 shows the crystal structure of GAC, illustrating an example when it is attached to the BPTES inhibitor.
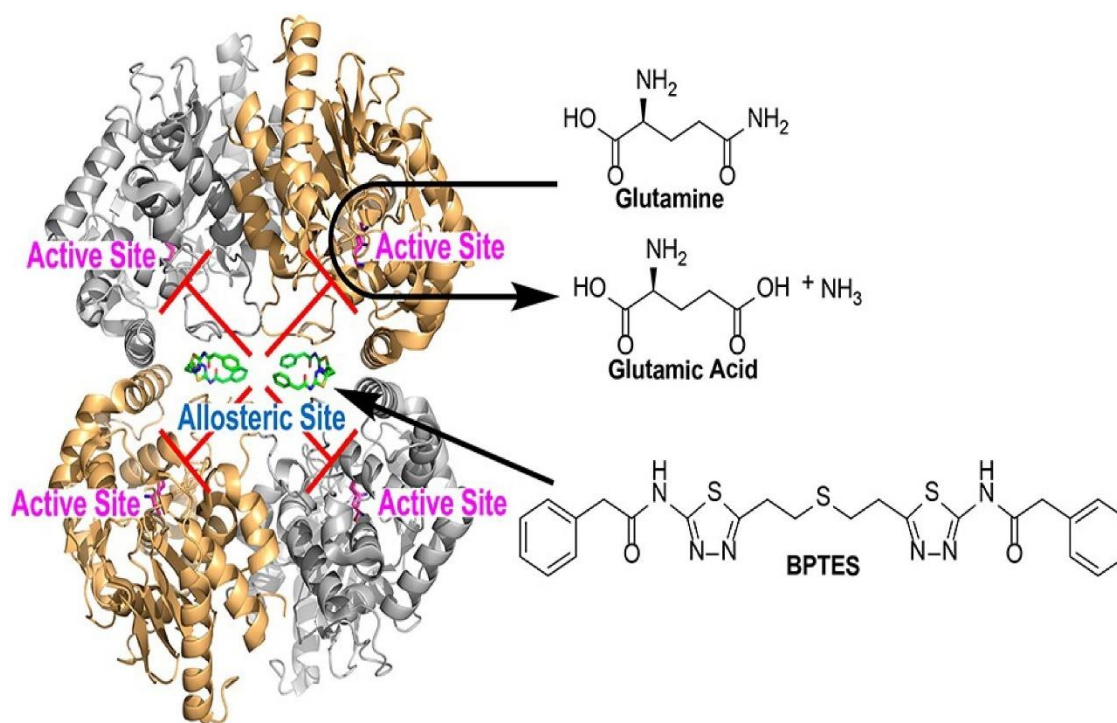
Figure I.5. The crystal structure of GLS in complex with glutamate and BPTES [61].

8.4.1. Allosteric inhibitors and inhibition strategies

Allosteric inhibitors specifically bind to the allosteric site of GAC, a site distinct from the catalytic site, inducing modifications in the enzyme's conformation and function, thereby modulating its activity (Figure I.6).

In recent years, GAC inhibitors have gathered significant attention due to their observed anti-proliferative effects and their ability to induce apoptosis in breast cancer, particularly in aggressive cancers such as TNBC [50]. A distinctive characteristic of TNBC metabolism, known as "glutamine addiction," has been uncovered, revealing its reliance on Gln to facilitate cell proliferation and metastasis. TNBC, in contrast to other breast cancer subtypes, exhibits elevated Gln demand and heightened susceptibility to Gln depletion. This finding raises the possibility that targeting Gln metabolism could be an effective strategy to improve the treatment of TNBC [51-54].

In the class of selective small molecule inhibitors targeting GAC, acivicin, CK, DON, Compound 968, BPTES, CB-839, and Ebselen are notable inhibitors that have been reported (Figure I.7) [62, 63].
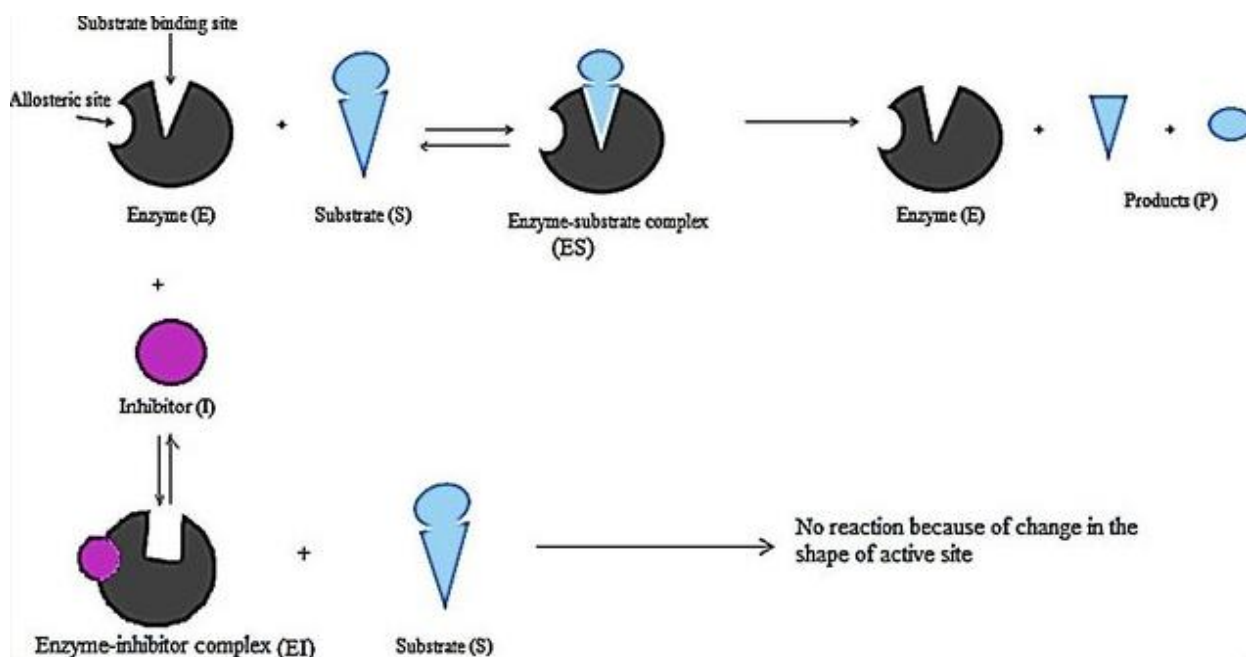
Figure I.6. Mechanism of allosteric inhibitors [64].

Acivicin and DON are Gln analogs, that bind to the active site of KGA/GAC but exhibit high drug toxicity due to their lack of selectivity against other Gln-binding enzymes. Compound 968, BPTES, and CB-839 are identified as an allosteric inhibitor of GAC (Figure I.7) [63]. Blocking GAC using these compounds results in decreased proliferation of tumor cells both in vitro and in vivo. Notably, studies have indicated that these drugs do not bind directly to the GAC active site but to distinct allosteric sites of both GLS1 isoforms, inducing conformational changes in the enzyme [65]. These changes can affect catalytic activity, and cause significant conformational changes in the tetramer, rendering it inactive [66, 67].
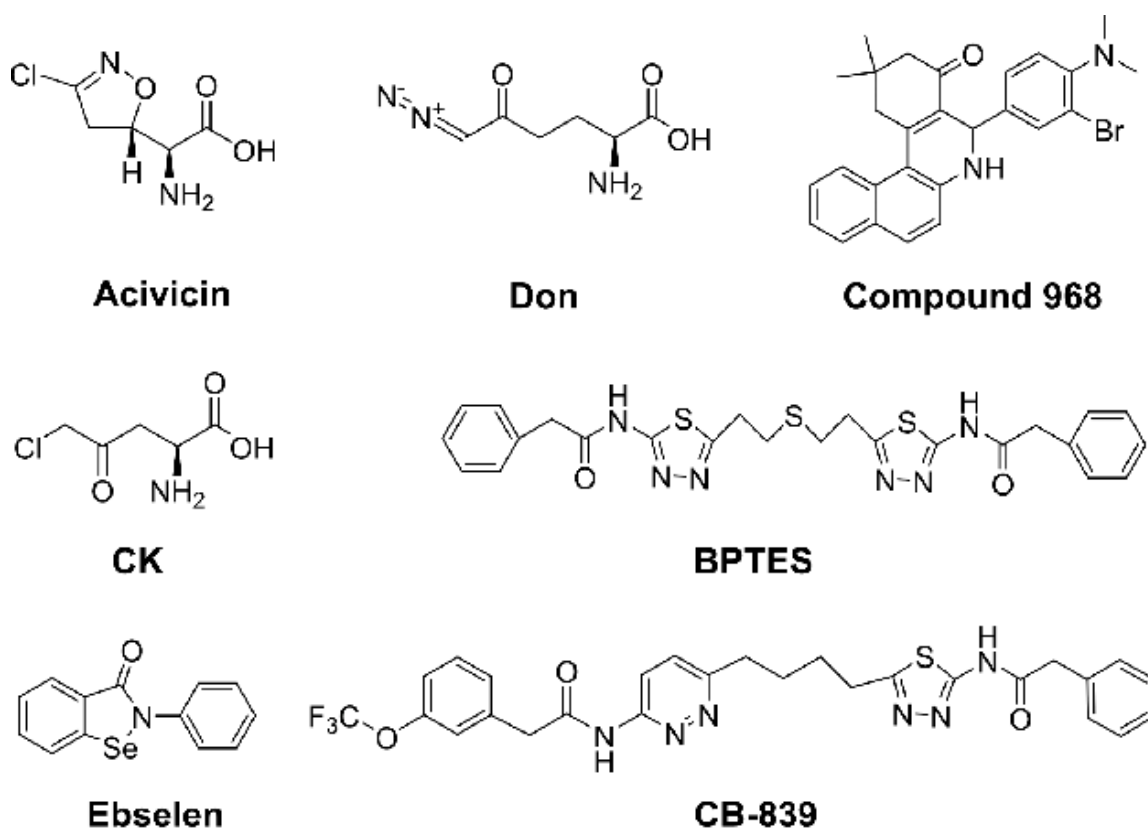
Figure I.7. Structures of glutaminase inhibitors [68].

REFERENCES

1. Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. The Breast. 2022;66:15-23.

2. Feng Y, Spezia M, Huang S, Yuan C, Zeng Z, Zhang L, et al. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes & diseases. 2018;5(2):77-106.

3. Javed A, Lteif A, editors. Development of the human breast. Seminars in plastic surgery; 2013: Thieme Medical Publishers.

4. Russo J, Russo IH. Development of the human breast. Maturitas. 2004;49(1):2-15.

5. Lawrence RA. Anatomy of the Breast. Breastfeeding: Elsevier; 2022. p. 38-57.

6. Gajbe B, Das Kurmi B, Kenwat R, Paliwal R, Paliwal SR. Chapter 1 - Breast cancer: introduction. In: Paliwal SR, Paliwal R, editors. Targeted Nanomedicine for Breast Cancer Therapy: Academic Press; 2022. p. 3-26.

7. Arendt LM, Kuperwasser C. Form and function: how estrogen and progesterone regulate the mammary epithelial hierarchy. Journal of mammary gland biology and neoplasia. 2015;20:9-25.

8. Breast Cancer [Available from: http://screenme.org/.

9. Apantaku LM. Breast cancer diagnosis and screening. American Family Physician. 2000;62(3):596-602.

10. Iles KE, Dickinson DA. Carcinogens, Environmental. In: Quah SR, editor. International Encyclopedia of Public Health (Second Edition). Oxford: Academic Press; 2017. p. 422-37.

11. Momenimovahed Z, Salehiniya H. Epidemiological characteristics of and risk factors for breast cancer in the world. Breast Cancer: Targets and Therapy. 2019;11:151-64.

12. Alkabban FM, Ferguson T. Continuing Education Activity.

13. Yalaza M, İnan A, Bozer M. Male breast cancer. The journal of breast health. 2016;12(1):1.

14. Shoemaker ML, White MC, Wu M, Weir HK, Romieu I. Differences in breast cancer incidence among young women aged 20–49 years by stage and tumor characteristics, age, race, and ethnicity, 2004–2013. Breast cancer research and treatment. 2018;169:595-606.

15.   Vogel VG. Epidemiology of breast cancer. The breast: Elsevier; 2018. p. 207-18. e4.

16.   Hunter DJ, Colditz GA, Hankinson SE, Malspeis S, Spiegelman D, Chen W, et al. Oral contraceptive use and breast cancer: a prospective study of young women. Cancer Epidemiology, Biomarkers & Prevention. 2010;19(10):2496-502.

17.   Zhu H, Lei X Fau - Feng J, Feng J Fau - Wang Y, Wang Y. Oral contraceptive use and risk of breast cancer: a meta-analysis of prospective cohort studies. (1473-0782 (Electronic)).

18.   Casem ML. Chapter 9 - Endocytosis. In: Casem ML, editor. Case Studies in Cell Biology. Boston: Academic Press; 2016. p. 217-40.

19.   Del Río JP, Alliende MI, Molina N, Serrano FG, Molina S, Vigil P. Steroid hormones and their action in women's brains: the importance of hormonal balance. Frontiers in public health. 2018;6:141.

20.   Li Z, Wei H, Li S, Wu P, Mao X. The role of progesterone receptors in breast cancer. Drug Design, Development and Therapy. 2022:305-14.

21.   Chen P, Li B, Ou-Yang L. Role of estrogen receptors in health and disease. Frontiers in Endocrinology. 2022;13.

22.   Bécourt S, Espié M. Hormonal Treatment of Breast Cancer. 2019.

23.   Lumachi F, Brunello A, Maruzzo M, Basso U, Mm Basso S. Treatment of estrogen receptor-positive breast cancer. Current medicinal chemistry. 2013;20(5):596-604.

24.   Scarpin KM, Graham JD, Mote PA, Clarke CL. Progesterone action in human tissues: regulation by progesterone receptor (PR) isoform expression, nuclear positioning and coregulator expression. Nuclear receptor signaling. 2009;7(1):nrs. 07009.

25.   Waks AG, Winer EP. Breast Cancer Treatment: A Review. JAMA. 2019;321(3):288-300.

26.   Braasch MC, Amin AL, Balanoff CR, Wagner JL, Larson KE. Prognostic Significance of Lobular Carcinoma In-Situ (LCIS) Diagnosed Alongside Invasive Breast Cancer. Breast Cancer: Basic and Clinical Research. 2022;16:11782234211070217.

27.   Nakhlis F, Morrow M. Ductal carcinoma in situ. Surgical Clinics. 2003;83(4):821-39.

28.   Snoj N, Dinh P, Bedard P, Sotiriou C. Molecular biology of breast cancer. Molecular Pathology: Elsevier; 2009. p. 501-17.

29.   [Available from: https://www.rockymountaincancercenters.com/breast-cancer/types-hormone-receptors.

30. Glajcar A, Łazarczyk A, Tyrak KE, Hodorowicz-Zaniewska D, Streb J, Okoń K, et al. Nodal status in luminal A invasive breast cancer: relationships with cytotoxic CD8 + and regulatory FOXP3 + cells tumor-associated infiltrate and other prognostic factors. Virchows Archiv. 2021;479(5):871-82.

31. Orrantia-Borunda E, Anchondo-Nuñez P, Acuña-Aguilar LE, Gómez-Valles FO, Ramírez-Valdespino CA. Subtypes of breast cancer. Breast Cancer [Internet]. 2022.

32. Weinberg F, Peckys DB, de Jonge N. EGFR expression in HER2-driven breast cancer cells. International journal of molecular sciences. 2020;21(23):9008.

33. Schlam I, Swain SM. HER2-positive breast cancer and tyrosine kinase inhibitors: The time is now. NPJ Breast Cancer. 2021;7(1):1-12.

34. Wilkinson L, Gathani T. Understanding breast cancer as a global health concern. The British Journal of Radiology. 2022;95(1130):20211033.

35. Curtit E, Nerich V, Mansi L, Chaigneau L, Cals L, Villanueva C, et al. Discordances in estrogen receptor status, progesterone receptor status, and HER2 status between primary breast cancer and metastasis. The oncologist. 2013;18(6):667.

36. Vrbic S, Pejcic I, Filipovic S, Kocic B, Vrbic M. Current and future anti-HER2 therapy in breast cancer. J buon. 2013;18(1):4-16.

37. Zagami P, Carey LA. Triple negative breast cancer: Pitfalls and progress. npj Breast Cancer. 2022;8(1):95.

38. Polyak K. Breast cancer: origins and evolution. The Journal of clinical investigation. 2007;117(11):3155-63.

39. Vargo-Gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. Nature Reviews Cancer. 2007;7(9):659-72.

40. Somaira N, Khaled A, Asef A, Alexandros GG. Breast Cancer- It's All in the DNA. In: Mehmet G, editor. A Concise Review of Molecular Pathology of Breast Cancer. Rijeka: IntechOpen; 2015. p. Ch. 1.

41. Ikhuoria EB, Bach C. Introduction to breast carcinogenesis–symptoms, risks factors, treatment and management. European Journal of Engineering and Technology Research. 2018;3(7):58-66.

42. Yip C-H, Rhodes A. Estrogen and progesterone receptors in breast cancer. Future oncology. 2014;10(14):2293-301.

43.   National Cancer Institute. "Hormone Therapy for Breast Cancer." Cancer.gov [Available from: https://www.cancer.gov/types/breast/breast-hormone-therapy-fact-sheet.

44.   Santen R, Harvey H. Use of aromatase inhibitors in breast carcinoma. Endocrine-related cancer. 1999;6(1):75-92.

45.   Avvaru SP, Noolvi MN, Aminbhavi TM, Chkraborty S, Dash A, Shukla SS. Aromatase inhibitors evolution as potential class of drugs in the treatment of postmenopausal breast cancer women. Mini Reviews in Medicinal Chemistry. 2018;18(7):609-21.

46.   Sayyad NB, Sabale PM, Umare MD, Bajaj KK. Aromatase inhibitors: development and current perspectives. Indian J Pharm Educ Res. 2022;56:311-20.

47.   Goldstein S. Selective estrogen receptor modulators and bone health. Climacteric. 2022;25(1):56-9.

48.   Chan HJ, Petrossian K, Chen S. Structural and functional characterization of aromatase, estrogen receptor, and their genes in endocrine-responsive and–resistant breast cancer cells. The Journal of steroid biochemistry and molecular biology. 2016;161:73-83.

49.   Dhingra K. Antiestrogens–tamoxifen, SERMs and beyond. Investigational new drugs. 1999;17:285-311.

50.   Ermiah A, Absheenah M, Alghamoudi B, Ermiah E. Benefits and Risks of Hormonal therapy (Tamoxifen) in Women with Positive Hormones Receptors Breast Cancer. Libyan International Journal of Oncology. 2022:7-13.

51.   Munzone E, Colleoni M. Clinical overview of metronomic chemotherapy in breast cancer. Nature reviews Clinical oncology. 2015;12(11):631-44.

52.   Iqbal N, Iqbal N. Human epidermal growth factor receptor 2 (HER2) in cancers: overexpression and therapeutic implications. Molecular biology international. 2014;2014.

53.   Tesch ME, Gelmon KA. Targeting HER2 in breast cancer: latest developments on treatment sequencing and the introduction of biosimilars. Drugs. 2020;80(17):1811-30.

54.   Lee EJ, Duggirala KB, Lee Y, Yun MR, Jang J, Cyriac R, et al. Novel allosteric glutaminase 1 inhibitors with macrocyclic structure activity relationship analysis. Bioorganic & Medicinal Chemistry Letters. 2022;75:128956.

55.   Warburg O. On the origin of cancer cells. Science. 1956;123(3191):309-14.

56.   Fung MKL, Chan GC-F. Drug-induced amino acid deprivation as strategy for cancer therapy. Journal of hematology & oncology. 2017;10(1):1-18.

57.   Yousaf R, Navid A, Azam SS. Discovery of novel Glutaminase allosteric inhibitors through drug repurposing and comparative MMGB/PBSA and molecular dynamics simulation. Computers in Biology and Medicine. 2022;146:105669.

58.   Thangavelu K, Chong QY, Low BC, Sivaraman J. Structural Basis for the Active Site Inhibition Mechanism of Human Kidney-Type Glutaminase (KGA). Scientific Reports. 2014;4(1):3827.

59.   Cyriac R, Lee K. Glutaminase inhibition as potential cancer therapeutics: current status and future applications. Journal of Enzyme Inhibition and Medicinal Chemistry. 2024;39(1):2290911.

60.   Stalnecker CA, Ulrich SM, Li Y, Ramachandran S, McBrayer MK, DeBerardinis RJ, et al. Mechanism by which a recently discovered allosteric inhibitor blocks glutamine metabolism in transformed cells. Proceedings of the National Academy of Sciences. 2015;112(2):394-9.

61.   Cassago A, Ferreira APS, Ferreira IM, Fornezari C, Gomes ERM, Greene KS, et al. Mitochondrial localization and structure-based phosphate activation mechanism of Glutaminase C with implications for cancer metabolism. Proceedings of the National Academy of Sciences. 2012;109(4):1092-7.

62.   Chen L, Cui H. Targeting glutamine induces apoptosis: a cancer therapy approach. International journal of molecular sciences. 2015;16(9):22830-55.

63.   Zhu M, Fang J, Zhang J, Zhang Z, Xie J, Yu Y, et al. Biomolecular interaction assays identified dual inhibitors of glutaminase and glutamate dehydrogenase that disrupt mitochondrial function and prevent growth of cancer cells. Analytical chemistry. 2017;89(3):1689-96.

64.   Vii VX, Muñoz-Torrero D, Riu M, Feliu C, editors. Recent Advances in Pharmaceutical Sciences2019.

65.   McDermott LA, Iyer P, Vernetti L, Rimer S, Sun J, Boby M, et al. Design and evaluation of novel glutaminase inhibitors. Bioorganic & medicinal chemistry. 2016;24(8):1819-39.

66.   Milano SK, Huang Q, Nguyen T-TT, Ramachandran S, Finke A, Kriksunov I, et al. New insights into the molecular mechanisms of glutaminase C inhibitors in cancer cells using

serial room temperature crystallography. Journal of Biological Chemistry. 2022;298(2):101535.

67.   Huang Q, Stalnecker C, Zhang C, McDermott LA, Iyer P, O'Neill J, et al. Characterization of the interactions of potent allosteric inhibitors with glutaminase C, a key enzyme in cancer cell glutamine metabolism. Journal of Biological Chemistry. 2018;293(10):3535-45.

68.   Zimmermann SC, Wolf EF, Luu A, Thomas AG, Stathis M, Poore B, et al. Allosteric Glutaminase Inhibitors Based on a 1,4-Di(5-amino-1,3,4-thiadiazol-2-yl)butane Scaffold. ACS Medicinal Chemistry Letters. 2016;7(5):520-4.

CHAPTER II:

Background on Computer-Aided- Drug Design (CADD) Methods

1. Brief overview of computer-aided drug design (CADD)

The process of drug discovery is both resource-intensive and time-consuming, with the ultimate goal of finding new potential drug candidates. To accelerate this intricate procedure, computational methods have been integrated into the pre-clinical stages of drug discovery. Computer-aided drug design (CADD), also expressed as in silico, constitutes a set of computational techniques employed for the exploration, advancement, and examination of drugs and bioactive compounds that share similar biochemical characteristics [1]. CADD combines various approaches, including computational, experimental, translational, and clinical models. This process not only demands a significant amount of time and resources but also relies on substantial funding [2].

A "lead" is a chemical substance that interacts with therapeutic targets to demonstrate possible biological action. The search for a lead compound is the most crucial step in the drug development process. Such lead chemicals can be found via computational screening (virtual screening) and high-throughput screening (an experimental technique) [3]. Virtual screening is a computational method that aids in finding a lead from a chemical library against the chosen target protein [4]. As shown in Figure II.1, in silico methods, particularly through virtual screening, play a crucial role in significantly reducing the time and resource inputs involved with chemical production and biological testing. Overall, the goal of CADD is to improve the drug discovery and development process by using the full potential of computational tools and data-driven strategies. This involves the following key aspects [4]:

1. Leveraging computational methods: applying computational methods to simplify the drug discovery and development procedure

2. Harnessing chemical and biological insights: utilizing chemical and biological information concerning ligands and/or targets for the identification and enhancement of novel drugs.

3. Implementing virtual screening filters: CADD designs and implements virtual filters that systematically screen compounds, identifying those with unfavorable characteristics such as low activity or inadequate ADMET properties. This step aims to select the most promising and viable candidates for further development.
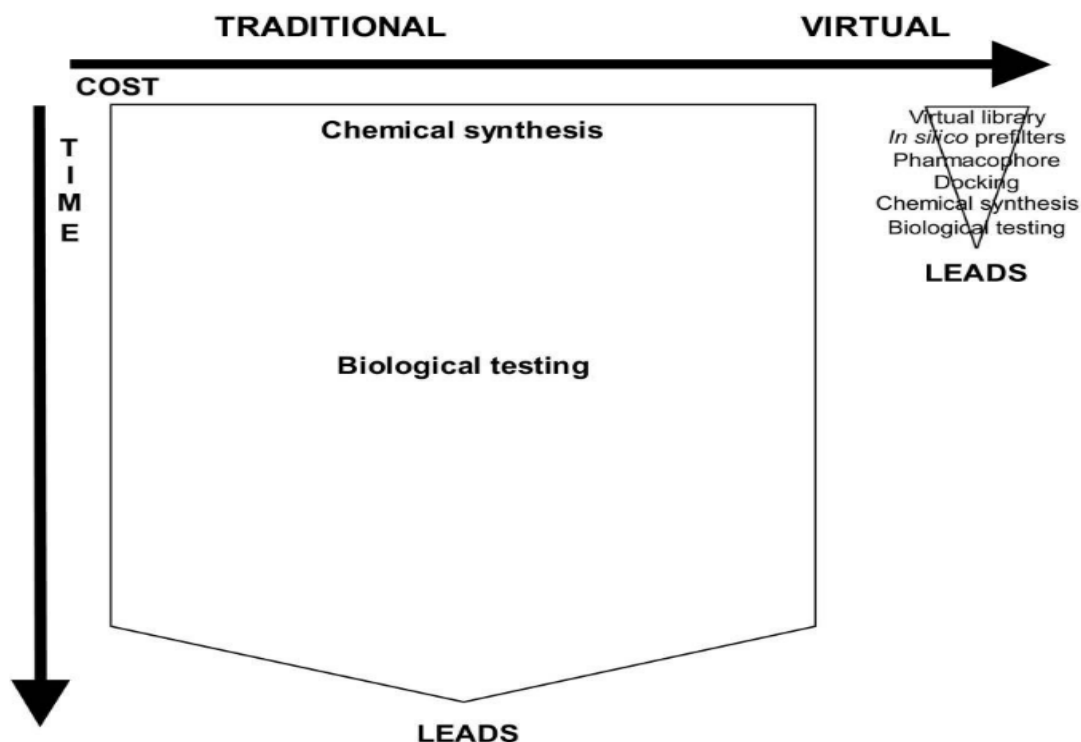
Figure II.1. Comparison of traditional and virtual screening in terms of expected cost and time requirements [4].

2. In silico virtual screening

CADD is categorized into two primary techniques: ligand-based drug design (LBDD) and structure-based drug design (SBDD). Both of these approaches are used in silico virtual screening for the identification and optimization of lead compounds [5-7]. Structure-based virtual screening becomes possible when the three-dimensional structure of the chosen target is known, often available in the Protein Data Bank (PDB), and can be done using molecular docking, de novo design, and molecular dynamics (MD). In contrast, LBDD doesn't require the three-dimensional structure of the target protein but relies on ligand data and often includes techniques such as quantitative structure-activity relationship (QSAR) modeling and pharmacophore modeling [2, 8, 9]. The flowchart in Figure II.2 highlights the various approaches to molecular modeling in CADD.
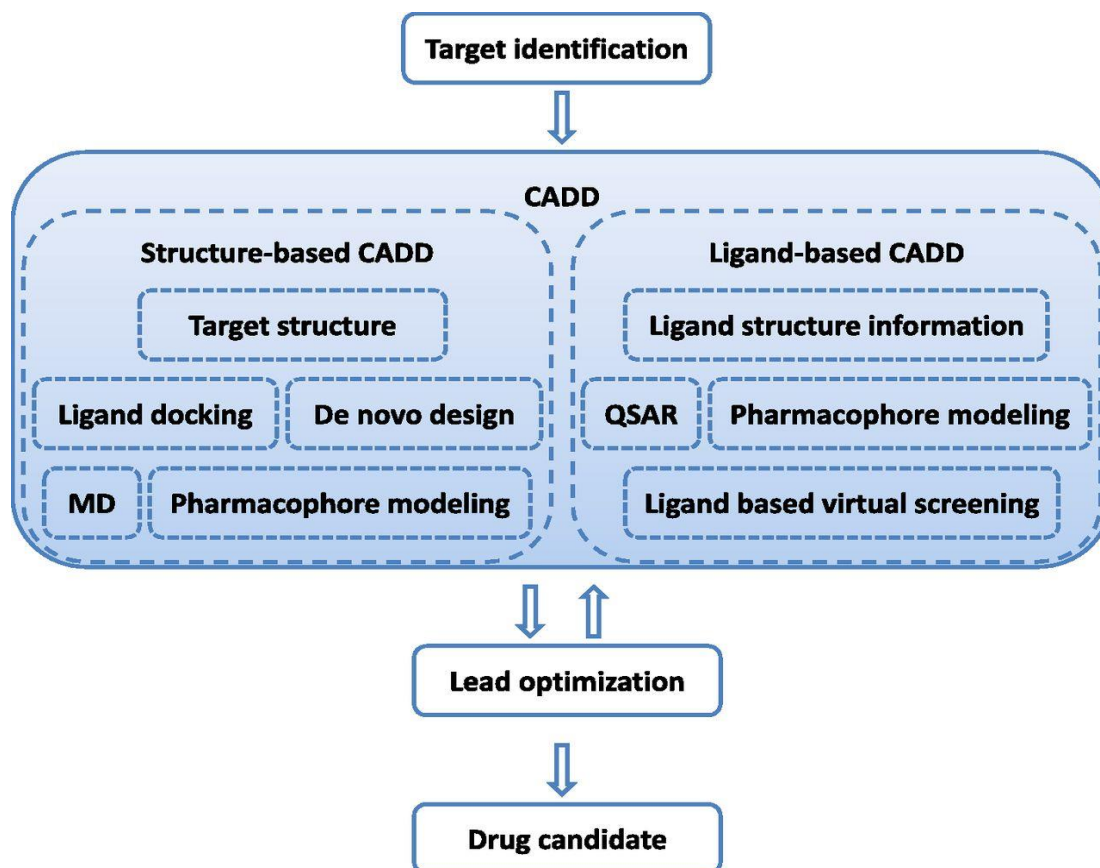
Figure II.2. Flowchart of the application of Computer-Aided Drug Design (CADD) used in the drug design process [12].

2.1. Ligand-based drug design

2.1.1. Quantitative structure-activity relationships (QSAR)

Over 50 years ago, Hansch and Fujita [10] developed a method known as Quantitative Structure-Activity Relationship (QSAR) analysis. Since then, QSAR has undergone significant changes, including advancements in the way we describe molecules, shifting from one-dimensional to multi-dimensional descriptors, as well as the adoption of various techniques to establish correlations between chemical structures and biological properties. These advances are made possible through the application of machine learning techniques. QSAR analysis remains a valuable approach for creating mathematical models that determine the quantitative correlation between structural molecular properties (referred to as descriptors) and specific properties, whether they are continuous values like pIC50, pEC50, Ki, or categorical distinctions like active, inactive, toxic, or nontoxic [11].

The process of QSAR is conducted through a general workflow (Figure II.3). This process is generally divided into multiple steps, including dataset collection and preparation, selection and generation of molecular descriptors, using a suitable statistical or mathematical model, model training and validation using a training dataset, and model testing using a test set [13].
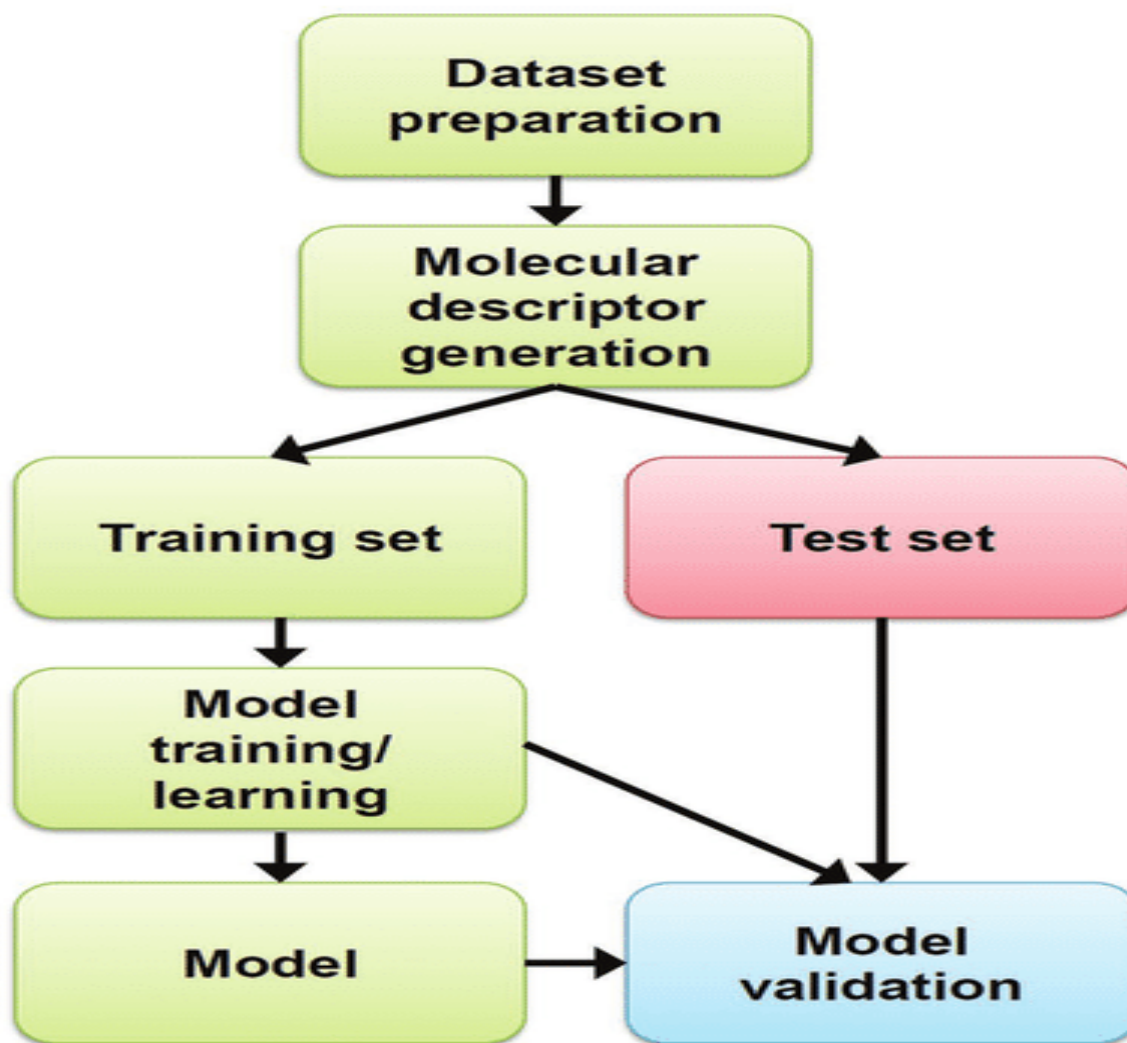
Figure II.3. Diagram illustrating the general workflow of the QSAR model [14].

2.1.1.1. Data collection, preparation, and curation

Selecting the right dataset is a crucial step when creating a QSAR model. Starting with a well-prepared dataset and a clear understanding of the chemical compounds and their associated properties provides a good foundation to move on to the critical steps of data analysis and model development. Firstly, the data must originate from reliable bioassay methods, and it is better to gather and use data from a single laboratory or source using the same experimental method to prevent discrepancies and variations between various sources. Additionally, the dataset should include a sufficient number of compounds to ensure the statistical reliability of the QSAR model, and the bioactivity data should include a wide range of values [13]. Suitable datasets for the QSAR model can be found through diverse sources, including research articles and electronic databases. These databases are accessible to the public, such as PubChem, BindingDB, ChEMBL, DSSTox, and the NIMH Psychoactive Drug Screening (PDSP), or commercially such as Wombat or MDDR [15].

2.1.1.2. Descriptor selection and generation

Molecular descriptors are quantitative attributes that characterize the features of molecular structures. These descriptors find extensive applications in chemoinformatic studies, particularly in QSAR analysis. The process of selecting and generating these descriptors plays a crucial role as it establishes a strong correlation with the biological or chemical activity of compounds. This correlation is instrumental in tasks like the analysis and categorization of chemical compounds, as well as in capturing structural variations essential for constructing robust QSAR models.

2.1.1.2.1. Types of descriptors

Molecular descriptors can take the form of either experimental data or computed physicochemical properties of molecules. Typically, they are diverse and can be categorized into different dimensions, each dimension provides unique perspectives on the chemical characteristics of molecules. These categories include [16-19]:

➢ 0D and 1D Descriptors: These provide fundamental information and mostly depend on the molecular formula, such as molecular weight and the count of constituent elements. The net charge is an example of a 1D descriptor.

➢ 2D Descriptors (Topological Descriptors): These descriptors are computed from the structural formula of the molecule. They reflect the connections within the molecular structure. 2D descriptors include constitutional, topology, total polar surface area, electrostatic and quantum chemical, geometrical, and molecular fingerprint properties of the chemical compound.

➢ 3D Descriptors: Calculated from the three-dimensional coordinates of the atoms within a molecule, 3D descriptors provide detailed spatial information about the compound. Common methods for calculating 3D descriptors such as Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA).

➢ 4D, 5D, and 6D Descriptors: These are multidimensional descriptors that incorporate various parameters related to ligand topology, receptor-binding site flexibility, and molecular structure. For instance, 4D descriptors are based on reference grids and molecular dynamic simulations, while 5D descriptors involve multiple conformations, orientations, protonation states, and isosteriomers of the ligand. Solvation terms are considered in 6D descriptors.

2.1.1.2.2. Calculation of molecular descriptors

Calculation of various molecular descriptors for each compound in the datasets can be performed using different software, each with its unique capabilities. Table II.1 gives a summary of some software options that can be used for calculating molecular descriptors:

Table II.1. Examples of software options for calculating molecular descriptors. Adapted from [17, 19].

Software Name	Description and Features	Availability	Ref.
MOE (Molecular Operating Environment)	Allows for the computation of a wide array of descriptors.	Commercial	[20]
PaDEL-Descriptor	Software for calculating various molecular descriptors including 2D and 3D descriptors.	Free	[21]
Dragon	Offers an extensive set of molecular descriptors including topological and geometrical descriptors	Commercial	[22]
ChemAxon Calculator	Includes a range of physicochemical property calculations.	Free for academic use	[23]
RDKit	An open-source toolkit for cheminformatics, including descriptor calculation.	Open Source	[24]

| Gaussian 09W | An advanced computational chemistry software suite, capable of performing some descriptor calculations and more. | Commercial | [25] |

2.1.1.3. Regression Analysis and model development

The model development process involves establishing the relationship between molecular descriptors, known as independent variables (X-variables), and a biological response, known as dependent variable (Y-variable), using mathematical and statistical techniques. To achieve this, researchers can use a suitable QSAR model, which can be constructed using various mathematical and statistical methods. These methods include linear methods such as multiple linear regression (MLR), partial least squares (PLS), and principal component regression (PCR), as well as non-linear methods such as neural networks and support vector machines (SVMs) [17]. In general, the standard equation of the QSAR model is represented as described in Equation 1. The biological activity or response (Y) is represented as a function of molecular descriptors or variables (X1, X2, . . . , XP) [26].

$$Y = \beta 0 + \beta 1X1 + \cdots + \beta pXp + \epsilon \qquad (1)$$

2.1.1.3.1. Partial Least Squares Regression (PLS)

Partial Least Squares (PLS) regression is a recent technique that simplifies, and merges features from principal component analysis and multiple regression. It provides a reduced solution, which is statistically more robust than multiple linear regression (MLR). It is particularly useful when there are more variables than compounds in the datasets and when the variables being studied are associated [18, 27].

2.1.1.4. Variable selection methods

Variable selection involves the process of choosing the most relevant variables to include in a model while eliminating irrelevant or redundant ones. Its goal is to create a model that offers the best possible fit to ensure correct predictions. Typically, when faced with numerous candidate variables and uncertainty about which ones to use, variable selection methods are employed to select the final model. Stepwise regression is a variable selection method among a variety of ways of feature selection aimed at creating a simplified and easily

interpretable model by systematically choosing significant variables. There are two major criteria for the variable selection methods: stepwise regression selection and all possible subset selection [27].

2.1.1.4.1. Stepwise regression methods

The stepwise regression methods include three techniques for variable selection namely, backward elimination, forward selection, and stepwise selection. Each of these methods has its advantages and disadvantages [28].

2.1.1.4.1.1. Backward elimination

The backward elimination method stands out as the simplest among variable selection techniques, particularly, when dealing with situations where the sample size exceeds the number of variables. It starts with a first model that includes all the variables under study. Consecutively, variables are systematically removed one by one from the first model until the remaining variables are considered to have a significant impact on the model's outcome. The variable with the smallest test statistic (a measure of the variable's contribution to the model), either falling below the required cut-off value or having the highest p-value greater than the cut-off value (indicating it is the least significant), is initially removed from the model. Afterward, the model is reconfigured without the removed variable, leading to a recalculation of test statistics or p-values. This iterative process continues until each remaining variable attains statistical significance at the prescribed cut-off value. It is important to note that the cut-off value associated with the p-value, commonly known as 'p-to-remove,' does not necessarily have to be set strictly at the 0.05 if prediction performance [29].

2.1.1.4.1.2. Forward selection

Forward selection begins with an empty model and systematically adds variables one by one until no additional variable significantly contributes to the model's outcome. Unlike backward elimination, forward selection initiates with a minimal model, possibly just an intercept, and then evaluates all one-variable extensions of the model. It systematically adds variables that align with predefined criteria such as the "lowest p-value," "highest adjusted R2," "lowest Mallow's Cp," "lowest AIC," "lowest score under cross-validation," and more. This iterative procedure continues successively including one variable at a time until the chosen criterion no longer exhibits improvement. In this process, forward selection aims to

optimize various criteria, selecting variables based on their statistical significance and contribution to the model's explanatory power. Once a variable is added to the forward selection, it remains part of the final model configuration. This method is effective in simplifying models, especially when dealing with a larger number of variables relative to the sample size, although it may not guarantee the identification of the absolute best subset of variables in all situations [28-30].

2.1.1.4.1.3.   Stepwise selection

Stepwise regression is a method that combines both forward and backward selection. It enables the introduction and removal of variables at various stages of modeling. In this approach, the procedure begins similarly to forward selection, where variables are added to the model based on their statistical significance. However, what distinguishes stepwise regression is its continuous assessment of the importance of variables that are already included in the model. After adding a new variable, the approach calculates F-statistics for all variables currently included in the model as if they were the most recent input. If a variable's F-statistic becomes non-significant at this point, it is deleted from the model, indicating that its regression coefficient is regarded as zero in the context of the current model. This feature identifies variables that have become redundant as a result of the later inclusion of new variables, hence helping to simplify the model [28, 31].

2.1.1.4.2.   All possible subset selection

All possible subset selection systematically evaluates every possible combination of independent variables when building predictive models. In this process, models ranging from single-variable to P-variable are built to determine which one is the best according to specific criteria. This approach begins by generating all 2^p possible models, where 'p' represents the number of independent variables. It takes a comprehensive approach by evaluating specific statistical criteria such as the highest adjusted R^2 or lowest AIC for every conceivable model of every possible size. It is important to note that this method can be computationally intensive, as it requires the creation and evaluation of all possible regression models. This computational demand becomes particularly challenging as the number of predictors, 'k,' increases. Consequently, all possible subset selection regression may not always be feasible or practical for large datasets with numerous predictors [28, 32, 33].

2.1.1.4.3.   Stopping rule and selection criteria

In all stepwise selection methods, including all subset selection, it is essential to establish a stopping rule or selection criteria for choosing the final model. Typically, a standard significance level for hypothesis testing is employed as the stopping rule [28]. These rules help decide whether a variable should enter the model (forward selection), remain in the model (backward elimination), or continue with the selection process (stepwise selection).

The application of these rules often relies on F-tests, significance levels, or similar criteria as the variable selection process.

The Fisher test (F-test) can be defined as follows:

$$F = \frac{\frac{SSE_r - SSE_f}{df_r - df_f}}{\frac{SSE_f}{df_f}} \qquad (2)$$

➢   Where SSE is the sum of squared errors, defined as:

$$SSE = \sum_{i=1}^{n_0}(y_i - \hat{y}_i)^2 \qquad (3)$$

In forward selection, the common criterion for stopping is the ratio of the reduction in the residual sum of squares (RSS) caused by adding the next candidate variable to the model. This ratio is typically expressed in terms of an "F-to-enter" statistic or a "significance level to enter" (SLE). The F-to-enter statistic is calculated based on the F-test of the partial sum of squares of the variable under consideration. Forward selection stops when no variable outside the current model meets the criteria to enter. It is important to note that these F-tests should be seen as stopping rules rather than classical tests of statistical significance, as the selection process itself can introduce biases [30].

In backward elimination, the stopping rule is based on the "F-test" of the smallest partial sum of squares among the variables that remain in the model. Like forward selection, this criterion can be expressed as an "F-to-stay" statistic or a "significance level to stay" (SLS). Backward elimination stops when all variables remaining in the model meet the criteria to stay [30].

Stepwise selection combines both forward and backward elimination criteria. The variable selection process in stepwise selection stops when all variables in the model meet the criteria to stay, and no variables outside the model meet the criteria to enter. It is worth noting that the criterion for a variable to enter the model does not have to be the same as the criterion for it to stay. In some cases, using a more relaxed criterion for entry can force the selection process to consider a larger number of variable subsets [30].

However, researchers also often utilize other criteria to determine when to stop the selection process. These alternative stopping rules may include metrics such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or Mallows' Cp statistic [28, 33]. In particular, the Bayesian Information Criterion (BIC), also known as the Schwarz Criterion, is a statistical criterion used for model selection when choosing among a limited set of models[34]. However, it shares similarities with Cp and AIC. In the context of a least squares model with d predictors, the BIC can be expressed as follows [35]:

$$\text{BIC} = \frac{1}{n}(RSS + \log(n)d\hat{\sigma}^2) \tag{4}$$

Where, $\hat{\sigma}$ is an estimate of the variance of the error associated with each response measurement in the standard linear model, RSS is the residual sum of squares (equation 11).

Indeed, both the AIC and BIC are valuable tools for evaluating statistical models. Researchers often use these criteria to select models based on the dataset and research aims. If the primary goal is to achieve maximum predictive accuracy, AIC may be preferred. If the emphasis is on selecting a model that is theoretically sound and consistent with the proper model, BIC may be the better choice. If the emphasis is on selecting a model that is theoretically sound and consistent with the proper model, BIC may be the better choice [28].

2.1.1.5.   Outlier detection

A common challenge when constructing a QSAR (Quantitative Structure-Activity Relationship) model is the existence of outliers. Outliers are data points that significantly differ from the majority of observations in a dataset, indicating unusual behavior. In the context of developing QSAR models, outliers are compounds with unexpected biological

activity that does not fit the model. They help define the limits of the model's applicability and expose experimental constraints. A good QSAR model has few or no outliers, while a bad one has many. Analyzing outliers can provide insights into compounds that might operate through different mechanisms or interact with receptors differently [36, 37].

2.1.1.5.1. Types of outliers

In QSAR analysis three types of outliers can affect the quality of the model [38]:

➤ **Outliers in dependent variable (y-direction):**

These outliers significantly deviate from the normal distribution of the dependent variable (y), leading to a large error sum of squares and affecting the model's accuracy. For xample: Point 1 in Figure II.4 is an outlier in the y-direction. Robust regression methods can handle a few outliers without masking effects.

➤ **Outliers in predictor or independent variable (X-direction):**

These outliers are distant from the main body of samples in the independent variable (X). Points 2 and 3 in the example (Figure II.4) are X outliers or leverage points due to their outlying x values. Point 2 is a good leverage point, causing minimal impact on the error sum of squares, while point 3 is a bad leverage point, leading to significant fluctuations in the model. QSAR/QSPR data contaminated by these leverage points can result in drastic model variations even with negligible changes.

➤ **Outliers towards the model:**

These outliers exhibit a different relationship between X and y after building the regression model, they represent diverse molecular structures in QSAR/QSPR studies. For instance, points marked as 4 in Figure II.4 are outliers toward the model, being outliers in both y and X.

These three types of outliers often coexist in a model, making their simultaneous identification crucial.
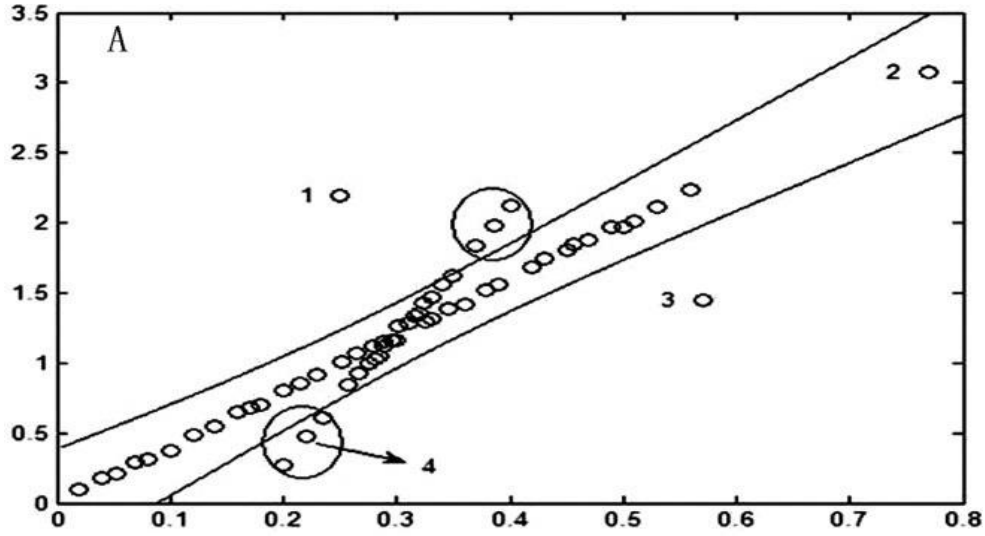
Figure II.4. Types of outliers in simple regression [38].

2.1.1.5.2.   Studentized deleted residual and leverage values

Studentized deleted residuals (r_i) and leverage values (h_{ii}) are employed in the process of outlier detection in multiple regression models. These statistical measures assist in identifying variables, both independent and dependent, that exhibit significant deviations from the expected patterns within the context of multiple regression analysis [39].

In simple regression analysis, the (h_{ii}) measure leverage as follows:

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (5)$$

Where x_i, is the ith variable, $\bar{x}$ is the sample mean.

The h_{ii} values are calculated based on the distance of each observation from the mean x-value. A leverage value greater than 2(k+1=n) is an indicator of outlying cases concerning their X values [30, 40].

The studentized deleted residuals (r_i) can be determined in the following formula:

$$r_i = \frac{y_i - \hat{y}_i}{S_{(i)}\sqrt{1 + h_{(ii)}}} \qquad (6)$$

where $\hat{y}_i$ is the predicted value from the estimates of the regression denote the ith observation, $S_{(i)}$ is the least squares obtained in a regression denote the ith observation, and $h_{(ii)}$ is the leverage value.

2.1.1.6.   Model validation methods

Validation is a crucial step in building a QSAR model to ensure its reliability when predicting the biological responses of new compounds. Typically, this involves dividing the dataset into a training set and a test set. Some methods can be used to select these subsets, including random selection, activity-based sampling (Y-response), and compound similarity-based selection (X-response). Moreover, methods such as k-means clustering and Kohonen's self-organizing map (SOM) are frequently applied. The size of the training and test sets is also important, and the optimal size will vary depending on the specific data set [41]. In QSAR analysis, there are two validation methods to verify the reliability of the selected models: internal validation, which assesses the goodness of fit, and external validation, which evaluates the model's predictive accuracy [42]. Figure II.5 illustrates the general workflow of model validation.
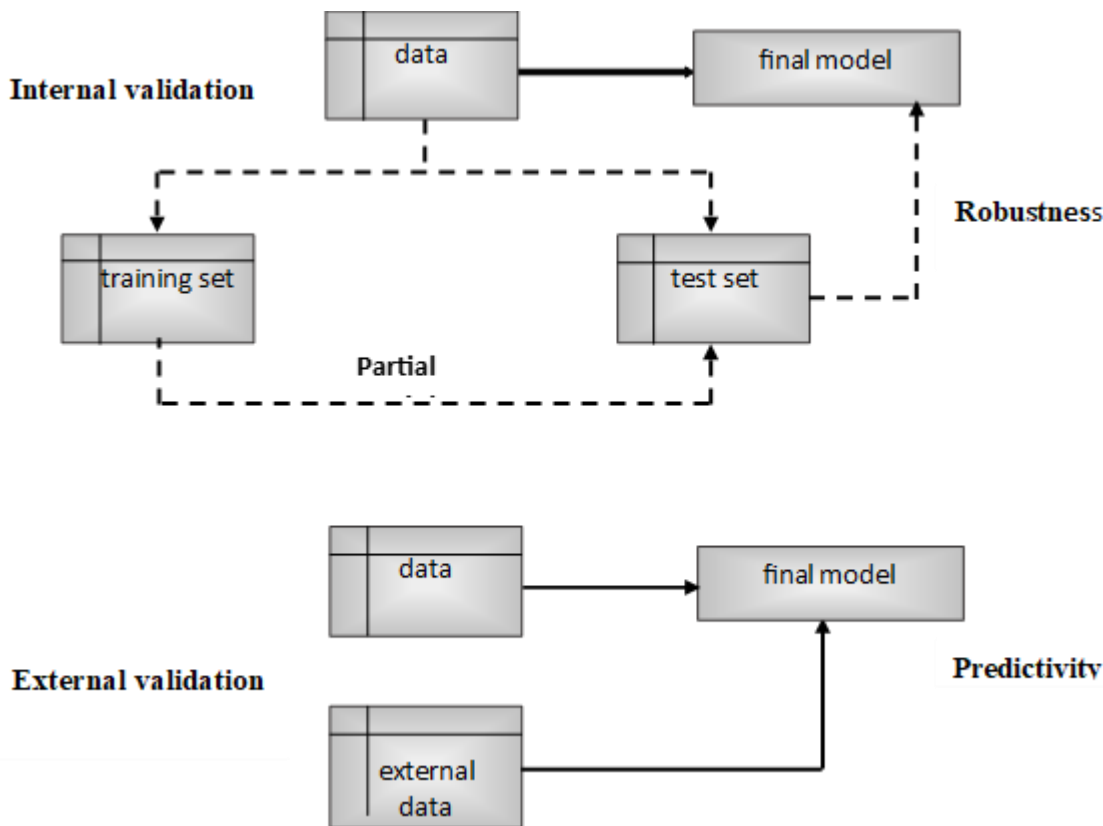
Figure II.5. General workflow of model validation: Internal and external validation. Adapted from [43]

2.1.1.6.1. Internal validation

Internal validation methods rely on using the same dataset that was employed to create the model. Various internal validation methods include least squares fit (R^2), cross-validation (Q^2), adjusted R^2 (R^2_{adj}), root mean squared error (RMSE), and bootstrapping. These methods help set up the reliability and robustness of QSAR models. The selection of training and test sets is important for the development of a statistically significant QSAR model.

➢ **Leave-One-Out (LOO) cross-validation**

LOO cross-validation evaluates the model's predictive power by systematically excluding one data point (compound) from the training dataset, and the QSAR model is rebuilt using the remaining molecules and the selected descriptors. The activity of the excluded compound is then predicted using the generated QSAR equation. This process is repeated for each compound in the training set, and the predicted activity values are used to calculate internal validation parameters. Finally, the model predictivity is judged using the predicted

residual sum of squares (PRESS) and cross-validated R^2 (Q^2) for the model while the value of standard deviation of error of prediction (SDEP) is calculated from PRESS [44].

$$PRESS = \Sigma(Y_{obs} - Y_{pred})^2 \tag{7}$$

PRESS measured quantifies the difference between the predicted and observed values for each excluded compound. The variables Y_{obs} and Y_{pred} represent the observed and LOO-predicted activity values, respectively.

$$SDEP = \sqrt{\frac{PRESS}{n}} \tag{8}$$

SDEP measures the spread or variability of prediction errors, and it is derived from the PRESS values. The variable n indicates the number of observations.

The Cross-Validated R^2(Q^2) is determined by the formula:

$$Q^2 = 1 - \frac{\Sigma(Y_{obs(train)} - Y_{pred(train)})^2}{\Sigma(Y_{obs(train)} - \overline{Y}_{training})^2} = 1 - \frac{PRESS}{\Sigma(Y_{obs(train)} - \overline{Y}_{training})^2} \tag{9}$$

Where, $Y_{obs(train)}$ refers to the observed activity, $Y_{pred(train)}$ represents the predicted activity of the training set molecules, calculated using the LOO method, and $\overline{Y}_{training}$ is the mean of the response values in the training set. Additionally, it is important to note that a threshold value of Q^2 is specified as 0.5. This threshold is often used to assess the goodness of fit of the model, and values above 0.5 indicate that the model provides a reasonable fit to the data [44].

➤ **Coefficient of determination (R²)**

The coefficient of determination R^2, often called the "R-squared" value, quantifies the proportion of the variance in the dependent variable. The value of R^2 indicates the coefficient of indicates the goodness of fit of a model, expressing the percentage of variation defined by the regression equation between the observed response and the independent variables. R^2 values vary from 0 to 1, with higher values indicating a better fit. In general, A significant QSAR model typically expects an R^2 value greater than or equal to 0.6, as closer values to 1 signify a stronger model fit. R² is calculated as follows [30, 45]:

$$R^2 = 1 - \frac{SSR}{SST} \tag{10}$$

Where:

SSR is the sum of squared residuals (the differences between observed and predicted values).

SST is the total sum of squares (a measure of the total variance in the dependent variable).

Note that *SSR* and *SST* are described as follows [45]:

$$RSS = \sum(y_i - \hat{y}_i)^2 \tag{11}$$

$$TSS = \sum(y_i - \tilde{y})^2 \tag{12}$$

y_i represents the observed values of the dependent variable. And $\hat{y}_i$ represents the predicted values of the dependent variable based on the regression model.:

$\tilde{y}$ represents the mean (average) of the observed values of the dependent variable.

➤ **Adjusted coefficient of determination (R_{adj}^2)**

Despite R^2 being used to evaluate how effectively a regression model explains the variance in the dependent variable, it comes with a limitation. This limitation becomes evident when additional predictor variables are added to the model, as the R^2 value tends to increase when additional predictor variables are added to the model, even if these added variables do not significantly contribute to reducing the unexplained variance in the dependent variable. In other words, R^2 can sometimes be misleading in assessing the value of additional predictors. To overcome this limitation, researchers frequently use (R^2_{adj}) as an alternative. The R^2_{adj} is a modified form of R^2 that takes the number of predictor factors and degrees of freedom into consideration. The R^2 value is adjusted by dividing the residual sum of squares (RSS) and total sum of squares (TSS) by their degrees of freedom. The resulting change penalizes the addition of unnecessary variables in the equation. If an additional variable fails to successfully reduce unexplained variance, R^2_{adj} decreases, offering a more realistic assessment of the model's goodness of fit.

the adjusted (R_{adj}^2) is calculated as follows [43]:

$$R_{adj}^2 = 1 - \frac{SSR/(n-p-1)}{SST/(n-1)} \tag{13}$$

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

Where:

n = number of observations

p = number of predictor variables

➢ **F-statistics:**

The F-statistic, also known as the Fisher test, is indeed a valuable tool

for assessing the overall significance level of a regression. [46].

It is used to judge the overall significance of a regression model. It assesses whether the model as a whole is statistically significant in explaining the variance in the dependent variable. The F-statistic can be calculated using the following formula. [47]:

$$F = \frac{\left[\frac{\Sigma\left(Y_{iexp} - Y_{Ts}\right)^2}{n}\right]}{\left[\frac{\Sigma\left(Y_{iexp} - Y_{ipred}\right)^2}{N-P-1}\right]} \qquad (14)$$

Where:

n is the number of molecules in the Training Set (TS).

P is the number of descriptors in the model.

Y_{iexp} and Y_{ipred} are the experimental and predicted values of biological activity for such molecule i.

$\bar{Y}_{Ts}$ is the average value of biological activity for the training set.

➢ **Bootstrapping**

Bootstrapping is an internal validation method that involves random sampling from the dataset. Instead of analyzing subsets of the data repeatedly, this approach involves analyzing sub-samples of the data multiple times. Each sub-sample is generated randomly with replacement from the original dataset. In a typical bootstrap validation, K groups of size n are created through multiple random selections of n objects from the original dataset. Certain

objects may be included in the same random sample multiple times, while others may never be selected. The model derived from n randomly selected objects is used to predict the target properties for the excluded samples. A high average R^2_{boot} value in bootstrap validation indicates the robustness of the model [41, 48].

2.1.1.6.2. External validation

External validation is a critical step towards finding the true predictive power of a QSAR model. In order to precisely evaluate the model's predictive power, the predicted activities must be compared to the observed activities of an external test set of compounds that were not used in the model development. Researchers proposed several statistical characteristics to evaluate the predictive power of a QSAR model using the test set. These include: R^2_{pred}. The value of R^2_{pred} reflects the degree of correlation between the observed and predicted activity data as shown in Eq (16) [49, 50]

$$R^2_{pred} = 1 - \frac{\Sigma(Y_{obs(test)} - Y_{pred(test)})^2}{\Sigma(Y_{obs(test)} - \bar{Y}_{trainig})^2} \tag{15}$$

Where $Y_{obs(test)}$ and $Y_{pred(test)}$ representing the observed and predicted activity values for the test set and $\bar{Y}_{trainig}$ is the average value for the dependent variable for the training set. It was mentioned that when $R^2_{pred} > 0.60$, it is considered a good indicator of strong external predictability [49].

2.2. Structure-based drug design (SBDD)

2.2.1. Molecular docking

Researchers proposed a concept in 1894 that has become a fundamental principle in biochemistry. This concept simplifies the intricate process of enzyme-substrate interactions by likening it to a 'key in a lock.' In this concept, enzymes represent the 'lock,' while substrates or other small molecule ligands are the 'key.' The success of this interaction relies on the precise fit of the 'key' (substrate) into the 'keyhole' (active site or binding pocket) of the 'lock' (enzyme or receptor). Only when the fit is perfect can productive biochemical reactions occur. Keys that are too small, too large, or poorly configured will not engage effectively with the lock (Figure II.6) [51]. Likewise, molecular docking is a computational method employed to design drugs by identifying potential candidates that can specifically target proteins. These candidates are identified using a docking algorithm, which attempts to determine how a small

molecule, known as a ligand, binds to a large macromolecular target at its active site. The active site is typically a region within an enzyme, which is a large protein responsible for catalyzing chemical reactions. This active site is often situated within a cleft or pocket in the enzyme's three-dimensional structure [52]. In recent years, a variety of docking software programs have been developed for both academic and commercial purposes. Notable examples include [53] AutoDock [54], AutoDock Vina [55], GOLD [56], and MOE-Dock [57].
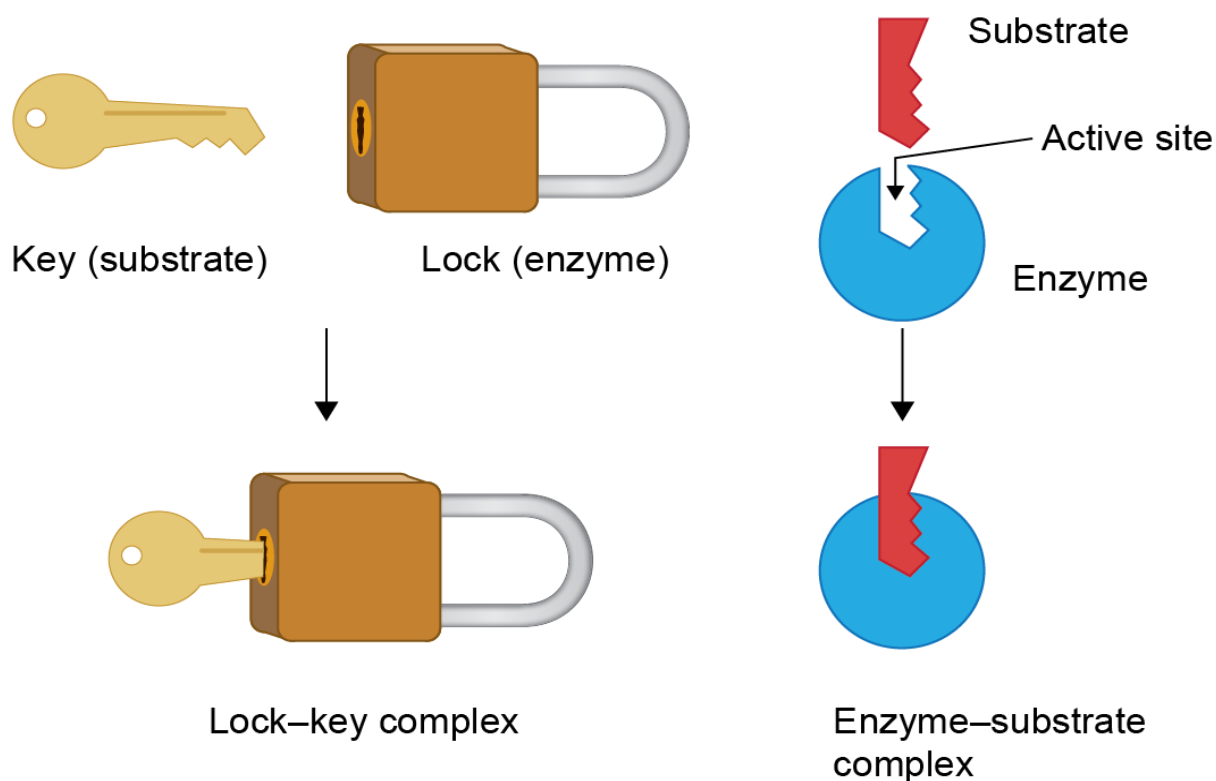
Figure II.6. The lock-and-key model of enzyme action [58].

2.2.1.1. Types of molecular docking

➢ **Rigid docking**

In rigid docking, both the receptor (protein) and the small molecule (ligand) are considered rigid structures. In this approach, the ligand is subjected to rotations and translations in three-dimensional space during the docking process, but its internal geometry or conformation remains constant and unaltered. To perform rigid docking, a pre-computed

library of all possible 3D conformations for each ligand representation is required to identify at least one potentially accurate conformation. Rigid docking offers faster computational speed but does not account for the flexibility of the ligand within the binding pocket, which may yield false-positive and false-negative results [59, 60].

➢ **Semi-flexible docking**

In semi-flexible docking, the ligand is considered flexible, while the receptor (protein) remains rigid. This approach involves sampling the six degrees of freedom of the ligand (rotational and translational). It is particularly useful when studying ligands that can adopt multiple conformations or when there is a need to account for the flexibility of the ligand during the binding process. It allows for a more realistic representation of the binding interactions between the ligand and the protein, which can lead to better predictions of binding affinities and binding modes [59, 61].

➢ **Flexible-Flexible Docking:**

This method is based on the idea that both the protein and the ligand can adopt flexible conformations during the binding process. It can involve either the use of an induced fit model or conformational selection [59]. These methods address the dynamic conformational changes in both the target and ligand, offering a more accurate representation of intermolecular binding interactions. Flexible docking methods employ techniques like fast Monte Carlo simulations, distance geometry, genetic algorithms, and tabu search. However, they require more computational resources due to the increased degrees of freedom, resulting in a larger search space for examining protein-ligand binding modes [62, 63].

2.2.1.2. Molecular docking basics

Molecular docking aimed to predict the structure of a ligand-receptor complex. This process involves two connected steps: firstly, the searching algorithm to explore various conformations of the ligand within the protein's active site, and secondly, the ranking of these conformations using a scoring function. Ideally, the searching algorithms should be capable of replicating the experimentally observed binding mode, while the scoring function should assign the highest rank to this mode among all generated conformations [64]. Figure II.7 summarizes the two main components of molecular docking: the search algorithm and the scoring function.

2.2.1.2.1. Searching algorithm

➢ **Systematic search techniques**

Systematic search algorithms are particularly useful for flexible-ligand docking, where they explore all possible binding conformations of the ligand by manipulating its degrees of freedom. There are three primary types of systematic search methods: exhaustive search, fragmentation, and conformational ensemble. Exhaustive search methods involve rotating all possible rotatable bonds in the ligand systematically, which provides comprehensive conformation sampling. However, initial screening applies constraints to manage complexity. Glide and FRED are examples of such methods. Fragmentation methods divide the ligand into rigid fragments, incrementally building the ligand's binding conformation by placing fragments in the binding site. Conformational ensemble methods represent ligand flexibility by docking pre-generated ligand conformations. Different docking runs produce a variety of binding modes, ranked by energy scores. Programs like FLOG and Q-Dock use conformational ensemble methods [65].

➢ **Stochastic algorithms**

Stochastic or random search methods introduce randomness rather than following a systematic approach to modifying the degrees of freedom in the system. These methods are advantageous for their speed, as they can potentially find optimal solutions quickly. However, they do not ensure a thorough exploration of conformational space, which may lead to missing the true solution. To address this lack of convergence, increasing the number of algorithm iterations is a common strategy.

Examples of stochastic algorithms include Monte Carlo methods, Tabu search methods, Evolutionary Algorithms (such as Genetic Algorithms), and Swarm optimization (SO) methods. These algorithms use different acceptance criteria based on probability to make favorable changes and explore a broad range of the energy landscape. While stochastic algorithms offer speed, the computational cost associated with these methods can be a limiting factor [66, 67].

➢ **Simulation methods**

Simulation methods typically start with a known initial state and aim to reach a lower energy state. These methods include Energy Minimization (EM), Molecular Dynamics (MD),

and Simulated Annealing. The predominant simulation approach for molecular docking is Molecular Dynamics Simulation, which involves calculating the system's trajectory through the application of Newtonian mechanics [66].
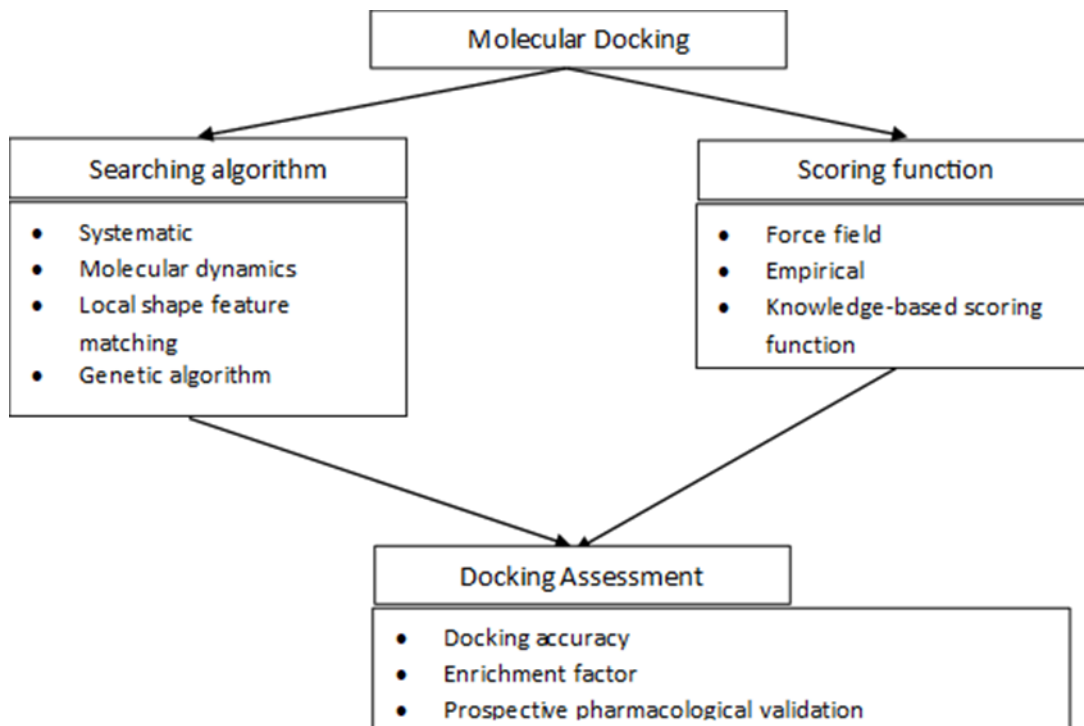
Figure II.7. Schematics illustrate the methods used for protein-ligand docking. Edit from [61].

2.2.1.2.2. Scoring functions (SFs)

SFs are mathematical methods commonly employed to predict the strength of interactions or binding affinity between molecules after docking. Scoring functions can be classified into two main types: classical scoring functions and machine learning-based scoring functions. Classical scoring functions include physics-based, empirical, and knowledge-based approaches. Figure II.8 describes the different scoring functions and some examples of their formulae [68].

Classical SFs

$$p = \sum_{m=1}^{M} w_m x_m$$

DOCK (force-field SF):

$$E_{bind} \equiv \overbrace{\sum_{k=1}^{K} \sum_{l=1}^{L} \left(\frac{A_{kl}}{d_{kl}^{12}} - \frac{B_{kl}}{d_{kl}^{6}} \right)}^{x_0} + \overbrace{\sum_{k=1}^{K} \sum_{l=1}^{L} \left(332 \frac{q_k q_l}{\varepsilon(d_{kl}) d_{kl}} \right)}^{x_1}$$

PMF (knowledge-based SF):

$$PMF \equiv \overbrace{\sum_{i} \sum_{j} \sum_{k=1}^{K_j} \sum_{l=1}^{L_i} H_{kl}(d_{kl})}^{x_m}$$

X-Score (empirical SF):

$$\Delta G_{bind} \equiv w_0 + w_1 \Delta G_{vdW} + w_2 \Delta G_{HBonds} + w_3 \Delta G_{rotor} + w_4 \Delta G_{hydrophobic}$$

with the braced terms labelled x_1, x_2, x_3, x_4 respectively.

Machine-learning SFs

$$p = f_{RF}(x_m) \qquad x_{ij} \equiv \sum_{k=1}^{K_j} \sum_{l=1}^{L_i} \Theta(d_{cutoff} - d_{kl})$$
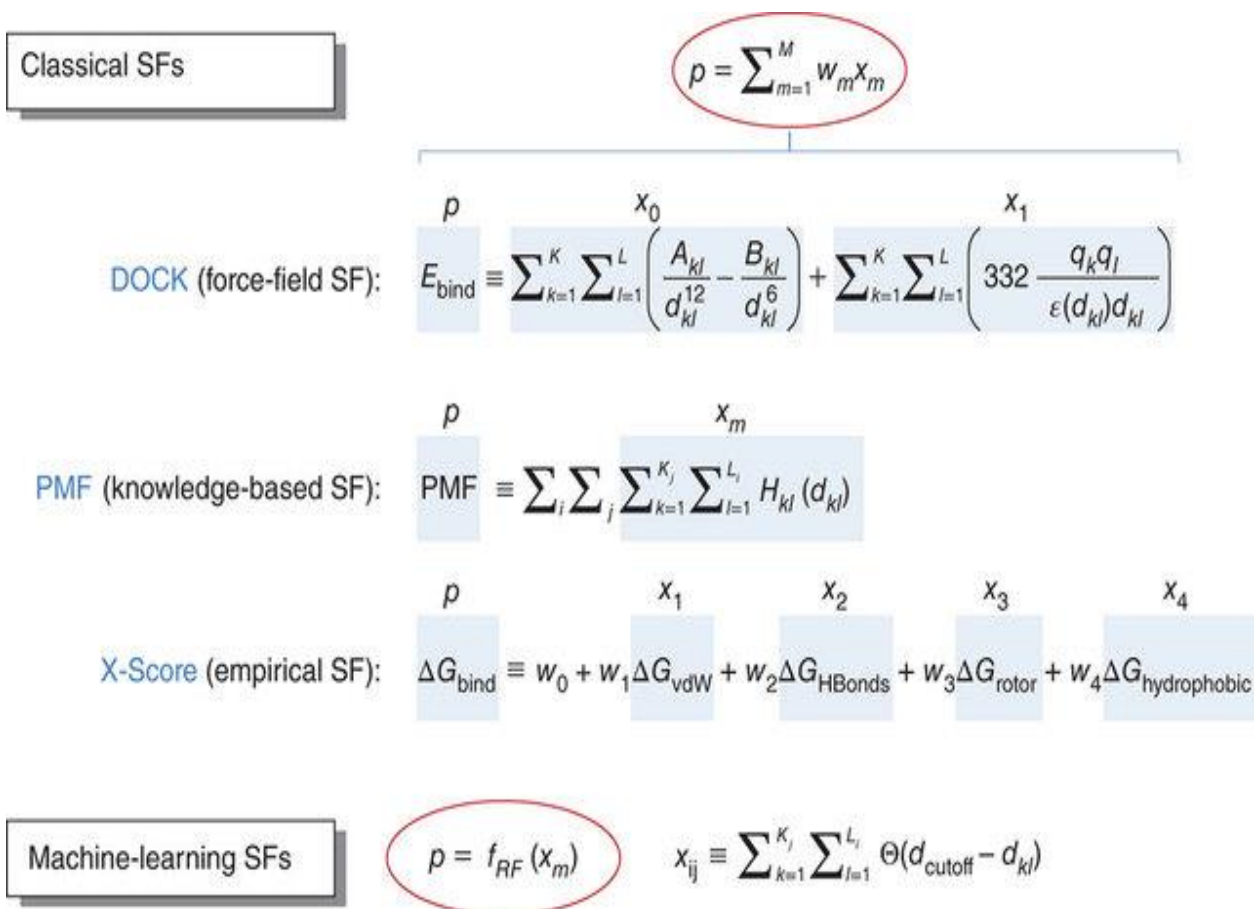
Figure II.8. Scoring functions in molecular docking [69].

➢ **Physics-based (force-field based) scoring functions**

Physics-based or force-field-based SFs use energy terms from the molecular mechanics force field. These force fields have parameters that are adjusted to reproduce experimental observations or ab initio quantum mechanical calculations. These SFs evaluate protein-ligand interactions by assessing the non-covalent interaction energy between atoms within the protein and ligand. In general, this interaction energy includes van der Waals and electrostatic interactions. In their basic forms, these pairwise interactions are represented using a Lennard-Jones potential and Coulomb interaction between point charges. Different physics-based SFs employ various potentials to describe van der Waals and electrostatic interactions, which depend on the underlying force field design. For instance, the dielectric constant can be distance-dependent to consider electrostatic screening due to the solvent and the lower dielectric constant in protein-ligand binding sites. In addition to van der Waals and

electrostatic terms, physics-based SFs often include shorter-range and sometimes directional terms to account for hydrogen bonding as well as solvation energy. Typically, several SFs based on force fields are widely used in molecular docking studies, such as GoldScore, AutoDock, and Generalised-Born Volume Integral/Weighted Surface Area (GBVI/WSA) [68, 70, 71].

> **Empirical scoring functions**

Empirical SFs are used to estimate the binding affinity of a protein-ligand complex by summing various energetic factors involved in the binding process, including hydrogen bonds, hydrophobic effects, and potential clashes between the protein and ligand (Figure 7). These factors are assigned coefficients that are figured out through multiple linear regression analyses conducted on a training set of known protein-ligand complexes with their established binding affinities. Unlike physics-based methods, empirical SFs are much faster because of their simplified energy term calculations. However, their accuracy relies heavily on the quality and diversity of the dataset used for model development [68].

> **Knowledge-based scoring functions:**

Knowledge-based SFs rely on statistical analyses of ligand-protein complexes to extract information about interatomic contact frequencies and distances between ligands and proteins. These functions operate on the principle that more favorable interactions occur more frequently. To derive these functions, collected frequency data is transformed into pairwise potentials based on atom types, often using methods like the Boltzmann law. The final score is calculated by choosing favorable contacts and penalizing repulsive interactions between ligand and protein atoms within a specific cutoff distance. Knowledge-based SFs offer distinct advantages due to their simplicity, as they do not depend on ab initio calculations or the replication of binding affinities. Moreover, they can model less common interactions like sulphur-aromatic or cation-π interactions, which are often challenging for empirical methods. Examples of knowledge-based functions include DrugScore and GOLD/ASP [64, 67, 72].

> **Machine-Learning-Based Scoring Functions**

Machine-learning-based SFs differ from classical SFs in their approach. While classical SFs assume a specific mathematical functional form, machine-learning-based SFs leverage various machine-learning algorithms like support vector machines, random forests, neural

networks, deep learning, and more. Although machine-learning-based SFs have demonstrated superior performance compared to classical SFs, they are not commonly integrated directly into docking software. Instead, they are frequently used for re-scoring existing results. This is because machine-learning-based scoring functions depend heavily on the quality of the training dataset. When a protein-ligand complex is initially docked using classical docking software and then the resulting structure is rescored using machine-learning-based scoring functions, the accuracy of the predictions is enhanced. This approach leverages the strengths of both classical and machine-learning-based methods, leading to more accurate results in molecular docking studies [71].

2.2.1.3. Steps involved in molecular docking

In general, molecular docking steps can be performed using different steps. The following flowchart (Figure II.9) represents the general steps involved in molecular docking.

2.2.1.3.1. Protein preparation

Protein preparation involves various essential steps that should be executed before starting the docking run. The starting step is downloading the 3D structure of the protein target as a PDB file which can be found from the Protein Data Bank (https://www.rcsb.org/). These structural data are typically generated through experimental techniques like X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, or electron microscopy. However, it's worth noting that most PDB files lack coordinates for hydrogen atoms. Consequently, adding the missing hydrogen atoms is a crucial aspect of this process. Furthermore, these preparatory steps include adding the missing side chains and missing bonds, addition of charge, and removal of unnecessary water molecules and cofactors. Subsequently, the 3D structure of the target protein undergoes minimization to optimize its conformation for docking studies [68].

Protein Preparation
- Obtain PDB file of protein target crystal structures
- If no crystal structure, create homology model
- Visualize protein in (e.g. PyMOL)
- Protonate amino acid side chains at physiologic pH
- Energy minimization of protonated protein

Grid Preparation
- Identify protein target binding site
- If PDB contains reference ligand, use centroid of reference as center for grid
- If no reference ligand, use centroid of amino acids lining binding pocket
- Set edge lengths of grid box

Ligand Preparation
- Obtain SDF file of ligand database (e.g. ZINC)
- Convert 2D to 3D structures with correct protonation state(s) and tautomers at physiologic pH (e.g. Omega, Ligprep)

Perform Docking
- Set docking software parameters (e.g. Autodock)
- Docking software will sample ligand poses in binding pocket within confines of the prepared grid

Analysis
- Rank ligands based on docking score
- Visualize ligand poses
- Filter results based on investigator preference

Figure II.9. Flowchart of the general protocol of molecular docking [60].

2.2.1.3.2. Binding site detection

Binding sites in the target macromolecular can be experimentally found by site-directed mutagenesis or X-ray crystallography. Additionally, a vast of literature exists on proteins that have been co-crystallized with their substrates or known inhibitors, which can offer valuable insights into binding sites [73]. In cases where data is limited, such as the absence of a

protein-ligand complex structure obtained through X-ray crystallography or the exploration of allosteric sites, various cavity detection algorithms or online servers have been developed for this purpose. Some notable programs and servers used for identifying putative active sites [64] such as POCKET [74], PASS [75], GRID [76], and MOE-Site-Finder [77]. Moreover, there are also several web servers used for prediction of allosteric sites, such as PASSer [78], AlloPred [79], and PARS [80].

2.2.1.3.3. Ligand preparation

Ligands can be prepared in various ways, such as by drawing their two-dimensional (2D)structure using chemical drawing software like Avogadro [81] and ChemDraw [82]. Otherwise, they can be obtained from vendors or public databases such as ZINC [83] and PubChem [84]. These ligand structures are represented in different file formats, like SMILES, SDF, MAE, and MOL2, each compatible with different molecular modeling software. For molecular docking simulations, it is essential to convert the 2D ligand structures into 3D PDB files. This conversion ensures that the ligand's spatial arrangement is accurately represented for docking simulations [60].

2.2.1.3.4. Molecular docking validation

After performing docking with the selection of an appropriate scoring function, it is essential to assess the accuracy of pose prediction. The most effective way to evaluate the docking algorithm is to compare the predicted binding orientation of the ligand with the position of the reference ligand in the experimentally determined structure, if available. The structural comparison between the two sets of coordinates is quantified using the root mean squared deviation (RMSD) Eq. (17) [85], measured in angstroms (Å). In practice, this value should be less than or equal to 2 Å for a successful docking result. It's worth noting that while RMSD calculations are straightforward, this metric is not normalized to the number of atoms and should not be regarded as an absolute measure [86].

$$\boldsymbol{the\ RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{ai}-x_{bi})^2 + (y_{ai}-y_{bi})^2 + (z_{ai}-z_{bi})^2} \qquad (16)$$

Where:

N: The total number of atoms being compared.

(x_{ai}, y_{ai}, z_{ai}): The coordinates (x, y, and z) of the i-th atom in the experimental (crystallographic) structure.

(x_{bi}, y_{bi}, z_{bi}): The coordinates (x, y, and z) of the i-th atom in the simulated (docking) structure.

2.2.2. Molecular dynamic simulation (MD)

Scientists face a common challenge in understanding how a protein or other biomolecules work. Although having an atomic-level structure is highly beneficial and offers significant insights into the biomolecule's workings, the constant motion of atoms within these structures introduces complexity. Molecular function and interactions among molecules are heavily influenced by the dynamics of these molecules. The goal is not just to capture a static image, it involves observing biomolecules in motion, manipulating them at the atomic level, and studying their responses. However, directly monitoring the movements of individual atoms and precisely manipulating them is a formidable task. A compelling alternative, researchers turn to atomic-level computer simulations of the relevant biomolecules [87].

Molecular dynamics (MD) is a computational technique employed for studying biomolecules within a virtual environment. In this approach, each constituent atom is represented as a particle, creating a multi-particle mechanical system that is analyzed through a simulation during MD analysis. The potential energies of the atoms are described by a mathematical expression involving various forces and spatial parameters [88].

In MD simulation, time-dependent interactions between biomolecules (protein-protein, protein-nucleic acid, protein-ligand, etc.) will be recorded and analyzed. Many researchers conduct molecular docking to predict the protein/protein or ligand or nucleic acid interactions, followed by MD simulations to predict the stability of these interactions in a dynamic environment. Nowadays, MD is being used in diverse fields, including drug design, nanobiotechnology, and various other fields of science [89]. In general, the system's trajectory of MD is computed using Newtonian mechanics. It calculates forces acting on each atom based on changes in potential energy between current and new positions. These forces, combined with the masses of the atoms, determine how atom positions change over successive short-time steps, following Newton's second law of motion [66, 90].

$$F_i = m_i \frac{d^2 r_i}{dt^2} \tag{17}$$

In equation (18), m_i represents the mass of atom i, r_i denotes the position of atom i at time t, and F_i represents the net force exerted on atom i.

2.3. Quantitative molecular electrostatic potential analysis

The molecular electrostatic potential (MEP) is a theoretical concept within computational chemistry that characterizes the distribution of electrostatic charge within a molecule. It is depicted as a three-dimensional representation, highlighting the variation in electrostatic potential across the surface of the molecule.

In a molecular system, the electrostatic potential (ESP) is expressed as [91]:

$$V(r) = \sum_A \frac{Z_A}{|r - R_A|} - \int \frac{\rho(r\prime)}{|r - r\prime|} \, dr' \tag{18}$$

Where Z_A represents the charge on nucleus A located at R_A and $\rho(r')$ denotes the total electronic density.

2.4. ADMET prediction

Predicted ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties play a crucial role in early drug discovery by facilitating the screening and identification of compounds with unfavorable characteristics [92]. This early screening based on predicted ADMET properties, assists researchers in prioritizing and focusing on drug candidates with a higher probability of success in further preclinical and clinical evaluations. It reduces the likelihood of investing in compounds that may be poorly absorbed, quickly metabolized or associated with toxicity issues [93].

In summary, the utilization of computational models and in silico predictions enables researchers to efficiently identify compounds that could face challenges concerning absorption, distribution, metabolism, excretion, and toxicity. This approach results in significant time and resource savings throughout the drug development process. While several notable in silico tools are available in commercial software packages based on proprietary datasets, there is a growing trend toward open-source software and web services in this field

[93]. Numerous popular web services and tools have emerged, such as ADMETLab [94], SwissADME [95], ADMETSar, ProTox [96], and pkCSM [97].

2.4.1. Oral administration

Orally administered drugs undergo disintegration in the gastrointestinal (GI) tract. Dissolved drug molecules can be absorbed through the gut wall, while precipitated drug particles are excreted through the GI tract. As illustrated in Figure II.10, when the drug molecules pass through the gut wall, they can be expelled by transporters or metabolized by enzymes. Drugs that successfully pass the gut wall will reach the liver, where numerous enzymes metabolize foreign compounds. Reactions are classified into phase I and phase II based on the enzymes involved. These reactions increase the hydrophilicity of xenobiotics, making them easily excretable through the kidney. Drugs that are not chemically modified during the metabolism process enter the systemic circulation. However, some drugs are unable to reach the target organ, tissue, or cell because they are bound by proteins in the blood. The unbound free form of drugs and metabolites can reach the target cell and biomolecules, with some swiftly eliminated through the kidney. Therapeutic effects occur when the drug concentration is adequate at the site of action after interacting with the physiological system. This intricate journey of orally administered drugs underlines the importance of understanding absorption, distribution, metabolism, excretion, and the role of binding proteins in pharmacokinetics [98].

In the following subsection, we will present a brief explanation of the definitions of absorption, distribution, metabolism, excretion, and toxicity. This detailed insight will facilitate a better understanding of ADME-TOX prediction results and their interpretation.

2.4.1.1. Absorption

Absorption is a fundamental process in pharmacology and toxicology that plays a crucial role in determining how substances, whether drugs or toxicants, enter the body and affect its various tissues and organs. Substances can enter the body through various ways, including the gastrointestinal (GI) tract, skin, lungs, ocular (eye), mammary gland, uterine, and injection sites. One of the primary factors affecting absorption is solubility. Solubility refers to a substance's ability to dissolve in a particular solvent, such as water. Substances that are poorly soluble or insoluble in water or other body fluids often have limited absorption,

particularly relevant for salts and ionized compounds, which tend to be poorly absorbed [99]. On the other hand, lipid-soluble (hydrophobic) substances are readily absorbed. Additionally, factors like Caco-2 permeability, human intestinal absorption, skin permeability, and the interaction with P-glycoprotein as a substrate or inhibitor are crucial when assessing and predicting compound absorption levels [100].
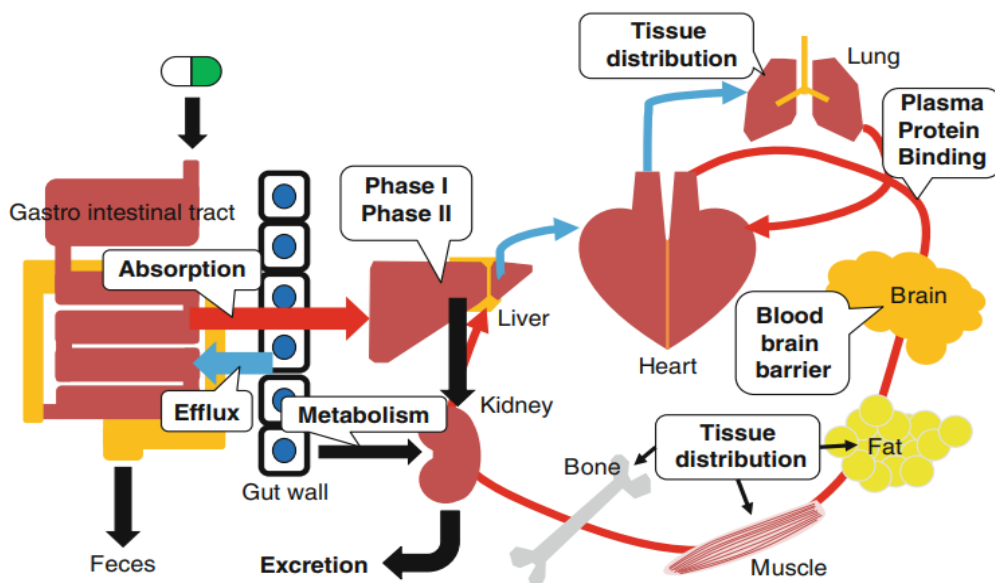
Figure II.10. ADME for oral administration of drugs [98].

2.4.1.2. Distribution

Distribution refers to the process by which a substance, often a drug, is transported throughout the body once it has been absorbed. After entering the bloodstream, a drug is distributed to various tissues and organs. The distribution process is influenced by several factors, including the drug's chemical properties, molecular size, and interactions with proteins in the blood. Some drugs may readily cross barriers like the blood-brain barrier (BBB), while others may not. Understanding the distribution of a drug is crucial for determining its concentration at target sites and evaluating its effectiveness in specific tissues or organs. The prediction of drug distribution in the body involves three main areas of study: BBB permeability (for brain penetration), volume of distribution (for dispersal within the body), and plasma protein binding (for drug-protein interactions) [98, 101].

2.4.1.3. Metabolism

Metabolism, also known as biotransformation, is the process by which the body chemically alters drugs and other foreign compounds. The primary location for drug metabolism is the liver, which contains numerous enzymes that catalyze these chemical transformations. Reactions during metabolism are typically classified into two phases: phase I and phase II. During these processes, the hydrophilicity (water-solubility) of xenobiotics (foreign substances) is increased, facilitating their excretion through the kidneys. The cytochrome P450 (CYP) enzymes play crucial roles and contributing significantly to drug-drug interactions. Notably, enzymes like CYP3A4, CYP2C9, and CYP2D6 collectively catalyze the hepatic metabolism of approximately 50% of drugs, highlighting the importance of the CYP450 superfamily. Additionally, the use of homology modeling for different isoforms of the CYP enzyme has proven valuable in refining these prediction models [98, 102].

2.4.1.4. Excretion

Excretion is the process of eliminating substances from the body, and clearance (Cl) serves as a vital parameter for quantifying how drugs are eliminated. The primary organ responsible for excretion is the kidney. During excretion, these substances are filtered from the blood and expelled from the body through urine. These metabolites can then be excreted in the urine. In general, hydrophobic drugs tend to undergo metabolic transformations in the liver, converting them into more polar, water-soluble compounds that are easily eliminated from the body. The rate of excretion can significantly affect a drug's duration of action in the body. Other routes of excretion include through feces, breath, sweat, or breast milk. Additionally, Some substances may also undergo enterohepatic recirculation, where they are excreted in bile, reabsorbed in the intestines, and re-enter the bloodstream [101].

2.4.1.5. Toxicity

Toxicity refers to the degree to which a substance, often a drug or chemical compound, can harm an organism. Toxicity can be categorized into various types, including acute toxicity, subacute toxicity, and chronic toxicity. Factors influencing a compound's toxicity include its chemical structure, dose, route of administration, and duration of exposure [103].

2.4.2. Drug-likeness and rule-of-five

Drug-likeness is a concept used in the early stages of drug discovery to assess whether a chemical compound has properties that make it a suitable candidate for further development as a drug. The "rule of five" (Ro5) is a set of guidelines designed to evaluate the drug-likeness of a compound, developed by C.A. Lipinski [104]. It is based on the analysis of the properties of many orally active drugs. According to the Ro5, a compound is considered drug-like if it meets certain criteria related to its molecular weight, lipophilicity, hydrogen bonding, and polar surface area. The Ro5 helps identify compounds with characteristics that are more likely to result in successful drug candidates.

The Ro5 indicates that the probability of poor absorption or permeation increases when [104] :

- ➢ The total number of H-bond donors (sum of OHs and NHs) exceeds 5.
- ➢ The molecular weight (MWT) is beyond 500.
- ➢ The Log P is higher than 5 (or MLogP is above 4.15).
- ➢ The total number of H-bond acceptors (sum of Ns and Os) exceeds 10.

Compounds that violate these rules may face challenges in terms of absorption and bioavailability. However, a recent study conducted by Hartung et al, [105] proposed that the (Ro5) may have limitations and should not be strictly adhered to in the optimization of oral exposure. Hartung's analysis suggests that the number 5, serving as a threshold in the Ro5, may not consistently align with low oral exposure, challenging the rigid application of this rule. The Ro5, designed to predict poor absorption or permeation based on specific criteria, relies on the analysis of drug candidates that achieved sufficient systemic exposure in early clinical studies. Despite being a valuable alert tool for newly synthesized compounds, the Ro5 should be approached with caution, considering the varied molecular properties observed in approved oral drugs compared to a pre-1997 dataset.

\

REFERENCES

1.   Sabe VT, Ntombela T, Jhamba LA, Maguire GEM, Govender T, Naicker T, et al. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. European Journal of Medicinal Chemistry. 2021;224:113705.

2.   Arya H, Coumar MS, Bhatt TK. Introduction of structural bioinformatics with respect to drug discovery. The design & development of novel drugs and vaccines: Elsevier; 2021. p. 3-9.

3.   Arya H, Coumar MS. Chapter 4 - Lead identification and optimization. In: Bhatt TK, Nimesh S, editors. The Design & Development of Novel Drugs and Vaccines: Academic Press; 2021. p. 31-63.

4.   Kapetanovic I. Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. Chemico-biological interactions. 2008;171(2):165-76.

5.   Shaker B, Ahmad S, Lee J, Jung C, Na D. In silico methods and tools for drug discovery. Computers in Biology and Medicine. 2021;137:104851.

6.   Yu W, MacKerell AD. Computer-aided drug design methods. Antibiotics: methods and protocols. 2017:85-106.

7.   Ankur G, Swatantra K, Vimal KM, Bipin P, Shailendra KS. High-Throughput Screening for Drug Discovery toward Infectious Diseases: Options and Challenges. In: Shailendra KS, editor. High-Throughput Screening for Drug Discovery. Rijeka: IntechOpen; 2022. p. Ch. 2.

8.   Sharma V, Wakode S, Kumar H. Structure-and ligand-based drug design: Concepts, approaches, and challenges. Chemoinformatics and bioinformatics in the pharmaceutical sciences. 2021:27-53.

9.   Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular Docking and Structure-Based Drug Design Strategies. Molecules. 2015;20(7):13384-421.

10.   Hansch C, Fujita T. p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. Journal of the American Chemical Society. 1964;86(8):1616-26.

11.   Neves BJ, Braga RC, Melo-Filho CC, Moreira-Filho JT, Muratov EN, Andrade CH. QSAR-based virtual screening: advances and applications in drug discovery. Frontiers in pharmacology. 2018;9:1275.

12.   Sliwoski G, Kothiwale S, Meiler J, Edward W. Lowe J. Computational Methods in Drug Discovery. Pharmacological Reviews. 2014;66(1):334-95.

13.   Myint KZ, Xie X-Q. Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. International journal of molecular sciences. 2010;11(10):3846-66.

14.   Sahlgren C, Meinander A, Zhang H, Cheng F, Preis M, Xu C, et al. Tailored approaches in drug development and diagnostics: from molecular design to biological model systems. Advanced Healthcare Materials. 2017;6(21):1700258.

15.   Golbraikh A, Wang XS, Zhu H, Tropsha A. Predictive QSAR modeling: methods and applications in drug discovery and chemical risk assessment. Handbook of computational chemistry. 2012:1309-42.

16.   Faulon J-L, Bender A. Handbook of chemoinformatics algorithms: CRC press; 2010.

17.   Moriwaki H, Tian Y-S, Kawashita N, Takagi T. Mordred: a molecular descriptor calculator. Journal of Cheminformatics. 2018;10(1):4.

18.   Peter S, Dhanjal J, Malik V, Radhakrishnan N, Jayakanthan M, Sundar D. Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications. Reference Module in Life Sciences. Elsevier; 2018.

19.   G Damale M, N Harke S, A Kalam Khan F, B Shinde D, N Sangshetti J. Recent advances in multidimensional QSAR (4D-6D): a critical review. Mini reviews in medicinal chemistry. 2014;14(1):35-55.

20.   Molecular Operating Environment (MOE) (2014) 2014.09. Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.

21.   Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. Journal of computational chemistry. 2011;32(7):1466-74.

22.   Mauri A, Consonni V, Pavan M, Todeschini R. Dragon software: An easy approach to molecular descriptor calculations. Match. 2006;56(2):237-48.

23.   ChemAxon.   JChem   Base   [Available   from: https://www.chemaxon.com/products/jchem-base/.

24.   G. L. RDKit: open-source cheminformatics. [Available from: http://www.rdkit.org.

25. Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Revision A.02, ed: Gaussian, Inc., Wallingford CT, 2016.

26. Peter SC, Dhanjal JK, Malik V, Radhakrishnan N, Jayakanthan M, Sundar D. Quantitative structure-activity relationship (QSAR): modeling approaches to biological applications. 2019.

27. Kar S, Roy K. Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs. 2011:111-22.

28. Mohammad Ziaul Islam C, Tanvir CT. Variable selection strategies and its importance in clinical prediction modelling. Family Medicine and Community Health. 2020;8(1):e000262.

29. Musa KI, Mansor WNAW, Hanis TM. Data Analysis in Medicine and Health using R: CRC Press; 2023.

30. Rawlings JO, Pantula SG, Dickey DA. Applied regression analysis: a research tool. Second Edi ed. New York: Springer New York, NY; 1998.

31. Thompson ML. Selection of variables in multiple regression: Part I. A review and evaluation. International Statistical Review/Revue Internationale de Statistique. 1978:1-19.

32. lecture-26 : Variable Selection. 1 December 2015. p. 36-401.

33. Witten D, James G. An introduction to statistical learning with applications in R: springer publication; 2013.

34. Murtaugh PA. Methods of variable selection in regression modeling. Communications in Statistics-Simulation and Computation. 1998;27(3):711-34.

35.   James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R: Springer; 2013.

36.   Verma RP, Hansch C, Brik A, Wu C-Y, Best MD, Wong C-H, et al. An approach toward the problem of outliers in QSAR pp 4597–4621. Bioorganic & Medicinal Chemistry. 2005;13(15).

37.   Walfish S. Statistical Outliers in the Laboratory Setting. AMERICAN LABORATORY. 2014;46(2):22-3.

38.   Cao DS, Liang YZ, Xu QS, Li HD, Chen X. A new strategy of outlier detection for QSAR/QSPR. Journal of computational chemistry. 2010;31(3):592-602.

39.   Coenders G, Saez M. Collinearity, heteroscedasticity and outlier diagnostics in regression. Do they always offer what they claim. New approaches in applied statistics. 2000;16(1):79-94.

40.   Kenouch S, Harkati D, Ghamri M, Chikhaoui AR, Melkemi N. Predictive QSAR model and clustering analysis of some Benzothiazole derivatives as cytotoxic inhibitors. SDRP Journal of Computational Chemistry & Molecular Modelling. 2018;2(3):1-8.

41.   Roy K, Kar S, Das RN. Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment: Academic press; 2015.

42.   Király P, Kiss R, Kovács D, Ballaj A, Tóth G. The Relevance of Goodness-of-fit, Robustness and Prediction Validation Categories of OECD-QSAR Principles with Respect to Sample Size and Model Type. Molecular Informatics. 2022;41(11):2200072.

43.   OECD. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models2014.

44.   Roy K, Kar S, Das RN. A Primer on QSAR-QSPR Modeling Fundamental Concepts.pdf>2015.

45.   Gramatica P. Principles of QSAR models validation: internal and external. QSAR & combinatorial science. 2007;26(5):694-701.

46.   Muhammad U, Uzairu A, Ebuka Arthur D. Review on: quantitative structure activity relationship (QSAR) modeling. J Anal Pharm Res. 2018;7(2):240-2.

47.   Traoré Y, Koné MG-R, Ouattara O, Ziao N. QSAR approach to estimating the analgesic activity of a series of tri-substituted pyrimidine derivatives. Journal of Computational Chemistry & Molecular Modelling. 2018;2(4):1-14.

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

49.   Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR models-strategies and importance. Int J Drug Des Discov. 2011;3:511-9.

50.   Tropsha A. Best practices for QSAR model development, validation, and exploitation. Molecular informatics. 2010;29(6-7):476-88.

51.   Tripathi A, Bankaitis VA. Molecular docking: from lock and key to combination lock. Journal of molecular medicine and clinical applications. 2017;2(1).

52.   Mukhopadhyay M. A brief survey on bio inspired optimization algorithms for molecular docking. International Journal of Advances in Engineering & Technology. 2014;7(3):868.

53.   Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophysical reviews. 2017;9:91-102.

54.   Österberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins: Structure, Function, and Bioinformatics. 2002;46(1):34-40.

55.   Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry. 2010;31(2):455-61.

56.   Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. Journal of molecular biology. 1997;267(3):727-48.

57.   Corbeil CR, Williams CI, Labute P. Variability in docking success rates due to dataset preparation. Journal of computer-aided molecular design. 2012;26(6):775-86.

58.   [Available   from: https://learning.eupati.eu/mod/book/tool/print/index.php?id=229&chapterid=59.

59.   SHARMA A, KUNWAR S, VAISHALI VA, SINGH C, DEV M, SHARMA NC. MOLECULAR DOCKING: AN EXPLANATORY APPROACH IN STRUCTURE-BASED DRUG DESIGNING AND DISCOVERY. databases. 2021;11:12.

60.   Issa NT, Badiavas EV, Schürer S. Research techniques made simple: Molecular docking in dermatology-A foray into in silico drug discovery. Journal of investigative dermatology. 2019;139(12):2400-8. e1.

61.   M. Venkata Saileela MVR, Venkata Rao Vutla. A REVIEW ON APPLICATIONS OF MOLECULAR DOCKING IN DRUG DESIGNING. 2017.

62.   Stoilov A, Yurukov B, Milanov P, editors. Analysis of docking algorithms by HPC methods generated in bioinformatics studies. ITM Web of Conferences; 2018: EDP Sciences.

63.   de Azevedo Junior WF, Dias R, Macedo Timmers LFS, Pauli I, Caceres RA, Pereira Soares MB. Bioinformatics tools for screening of antiparasitic drugs. Current drug targets. 2009;10(3):232-9.

64.   Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Current computer-aided drug design. 2011;7(2):146-57.

65.   Huang S-Y, Zou X. Advances and challenges in protein-ligand docking. International journal of molecular sciences. 2010;11(8):3016-34.

66.   Yadava U. Search algorithms and scoring methods in protein-ligand docking. Endocrinol Int J. 2018;6(6):359-67.

67.   Salmaso V, Moro S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. Frontiers in Pharmacology. 2018;9.

68.   Stanzione F, Giangreco I, Cole JC. Use of molecular docking computational tools in drug discovery. Progress in Medicinal Chemistry. 2021;60:273-343.

69.   Ain QU, Aleksandrova A, Roessler FD, Ballester PJ. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2015;5(6):405-24.

70.   Meli R, Morris GM, Biggin PC. Scoring functions for protein-ligand binding affinity prediction using structure-based deep learning: A review. Frontiers in bioinformatics. 2022;2:57.

71.   Li J, Fu A, Zhang L. An overview of scoring functions used for protein–ligand interactions in molecular docking. Interdisciplinary Sciences: Computational Life Sciences. 2019;11:320-8.

72.   Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH. Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. The AAPS Journal. 2012;14(1):133-41.

73.   Wang X, Song K, Li L, Chen L. Structure-based drug design strategies and challenges. Current Topics in Medicinal Chemistry. 2018;18(12):998-1006.

74.   Levitt DG, Banaszak LJ. POCKET: a computer graphies method for identifying and displaying protein cavities and their surrounding amino acids. Journal of molecular graphics. 1992;10(4):229-34.

75.   Brady GP, Stouten PF. Fast prediction and visualization of protein binding pockets with PASS. Journal of computer-aided molecular design. 2000;14:383-401.

76.   Kastenholz MA, Pastor M, Cruciani G, Haaksma EE, Fox T. GRID/CPCA: a new computational tool to design selective ligands. Journal of Medicinal Chemistry. 2000;43(16):3033-44.

77.   Labute P, Santavy M. Locating binding sites in protein structures. Journal of Chemical Computing Group. 2007.

78.   Tian H, Jiang X, Tao P. PASSer: Prediction of allosteric sites server. Machine learning: science and technology. 2021;2(3):035015.

79.   Greener JG, Sternberg MJ. AlloPred: prediction of allosteric pockets on proteins using normal mode perturbation analysis. BMC bioinformatics. 2015;16(1):1-7.

80.   Panjkovich A, Daura X. PARS: a web server for the prediction of protein allosteric and regulatory sites. Bioinformatics. 2014;30(9):1314-5.

81.   Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of cheminformatics. 2012;4(1):1-17.

82.   Brown T. ChemDraw. The Science Teacher. 2014;81(2):67.

83.   Sterling T, Irwin JJ. ZINC 15–ligand discovery for everyone. Journal of chemical information and modeling. 2015;55(11):2324-37.

84.   Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. Nucleic acids research. 2019;47(D1):D1102-D9.

85.   Dias R, de Azevedo J, Walter F. Molecular docking algorithms. Current drug targets. 2008;9(12):1040-7.

86.   Feyza MS, Selin S, Ece AS. Fundamentals of Molecular Docking and Comparative Analysis of Protein–Small-Molecule Docking Approaches.

87.   Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018;99(6):1129-43.

88.   Kumari I, Sandhu P, Ahmed M, Akhter Y. Molecular dynamics simulations, challenges and opportunities: a Biologist's prospective. Current Protein and Peptide Science. 2017;18(11):1163-79.

89.   Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. Journal of Computer-Aided Molecular Design. 2012;26(1):15-26.

90.   Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel drug discovery. Expert opinion on drug discovery. 2018;13(1):23-37.

91.   Lu T, Chen F. Quantitative analysis of molecular surface based on improved Marching Tetrahedra algorithm. Journal of Molecular Graphics and Modelling. 2012;38:314-23.

92.   Agu PC, Afiukwa CA, Orji OU, Ezeh EM, Ofoke IH, Ogbu CO, et al. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. Scientific Reports. 2023;13(1):13398.

93.   Venkatraman V. FP-ADMET: a compendium of fingerprint-based ADMET prediction models. Journal of Cheminformatics. 2021;13(1):75.

94.   Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Research. 2021;49(W1):W5-W14.

95.   Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific reports. 2017;7(1):1-13.

96.   Banerjee P, Eckert AO, Schrey AK. ProTox-II: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res. 2018;46.

97.   Pires DE, Blundell TL, Ascher DB. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J Med Chem. 2015;58.

98.   Shin HK, Kang Y-M, No KT. Predicting ADME properties of chemicals. Handbook of computational chemistry. 2017;59:2265-301.

99.   Song N-N, Zhang S-Y, Liu C-X. Overview of factors affecting oral drug absorption. Asian J Drug Metab Pharmacokinet. 2004;4(3):167-76.

100.   Han Y, Zhang J, Hu CQ, Zhang X, Ma B, Zhang P. In silico ADME and toxicity prediction of ceftazidime and its impurities. Frontiers in pharmacology. 2019;10:434.

101.   Chandrasekaran B, Abed SN, Al-Attraqchi O, Kuche K, Tekade RK. Computer-aided prediction of pharmacokinetic (ADMET) properties. Dosage form design parameters: Elsevier; 2018. p. 731-55.

102.   Krüger A, Gonçalves Maltarollo V, Wrenger C, Kronenberger T. ADME profiling in drug discovery and a new path paved on silica. Drug discovery and development-new advances. 2019:1-30.

103.   Barreto EF, Larson TR, Koubek EJ. Drug Excretion. Reference Module in Biomedical Sciences: Elsevier; 2021.

104.   Lipinski C. a, Lombardo, F., Dominy, BW & Feeney, PJ Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001;46(1-3):3.

105.   Hartung IV, Huck BR, Crespo A. Rules were made to be broken. Nature Reviews Chemistry. 2023;7(1):3-4.

Chapter III

QSAR Study, Molecular Docking/Dynamics Simulations and ADME Prediction of 2-Phenyl-1H-Indole Derivatives as Potential Breast Cancer Inhibitors

1.   Introduction

Heterocyclic compounds play a vital role in the advancement of drugs for various conditions, including cancer due to their diverse biological activities. A majority of commonly used medications, including chlordiazepoxide, imipramine, guanethidine, and indapamide, along with various antibiotics like penicillin, cephalosporin, norfloxacin, and streptomycin, incorporate heterocyclic rings. These compounds constitute approximately 80% of therapeutic drugs and demonstrate distinct chemical reactivity. Heterocyclic compounds, especially those containing nitrogen, play a crucial role in drug design, constituting around 60% of FDA-approved drugs. They serve as integral components in modern drug discovery and are recognized as foundational elements in medicinal chemistry [1].

The indole core is a widely distributed heterocycle in natural and synthetic bioactive compounds, including anticancer agents, and is considered a privileged scaffold in the design of such agents. Its unique physical, chemical, and biological properties make it a valuable component in the development of anticancer drugs [1, 2]. Recent studies have emphasized the therapeutic potential of indole core molecules to bind to multiple receptors with high affinity, making it useful in the development of novel bioactive drugs [1, 3, 4].

Aromatic nitrogen-based heterogeneous indole is widely present in nature, occurring in plants, bacteria, fungi, coal tar, and other sources. Bicyclic indole has been identified as an intracellular signal in both Gram-positive and Gram-negative bacteria, playing a crucial role in regulating various physiological processes such as plasmid stability, biofilm formation, spore generation, virulence, and drug resistance. In addition, it serves as an intercellular signal in bacteria, influencing processes like spore formation, plasmid stability, drug resistance, biofilm formation, and virulence. Key bacterial species, including Bacillus alvei, Vibrio cholera, Escherichia coli, and Enterococcus faecalis, depend on the tryptophan biosynthesis pathway to produce indoles [3].

Heterocyclic indole (C8H7N) is composed of a six-membered benzene ring fused with a five-membered pyrrole moiety and therefore is named as benzopyrroles. It exists as a colorless crystalline solid with an aromatic fragrance at room temperature. The synthesis of indole and its nucleophilic substitution reactions under basic conditions was first described by Adolf Von Baeyer in 1866 [5]. Figure III.1 summarizes the physicochemical properties of indole.

Indole serves as a foundational structure present in a variety of naturally occurring compounds. Examples include skatole (found in feces), tryptophan (an amino acid), hereroauxin (plant hormones), serotonin (a vasoconstrictor hormone), bufotenine (found in certain toads and toxic mushrooms), as well as important pharmaceutical compounds like vincristine and vinblastine (anti-leukemic alkaloids), etc. Cruciferous vegetables containing indole-3-carbinol have been utilized in the treatment and management of various chemotherapy-related cancers such as colorectal, breast, and prostate cancers [5].

Molecular formula	C_8H_7N
Molecular weight	117.151 g/mol
Physical state	Solid
Melting point	52-54 °C
Solubility	0.19 g/100 ml water
Density	1.17 g/cm^3
Crystal structure	Pna2
XLog P3	2.1
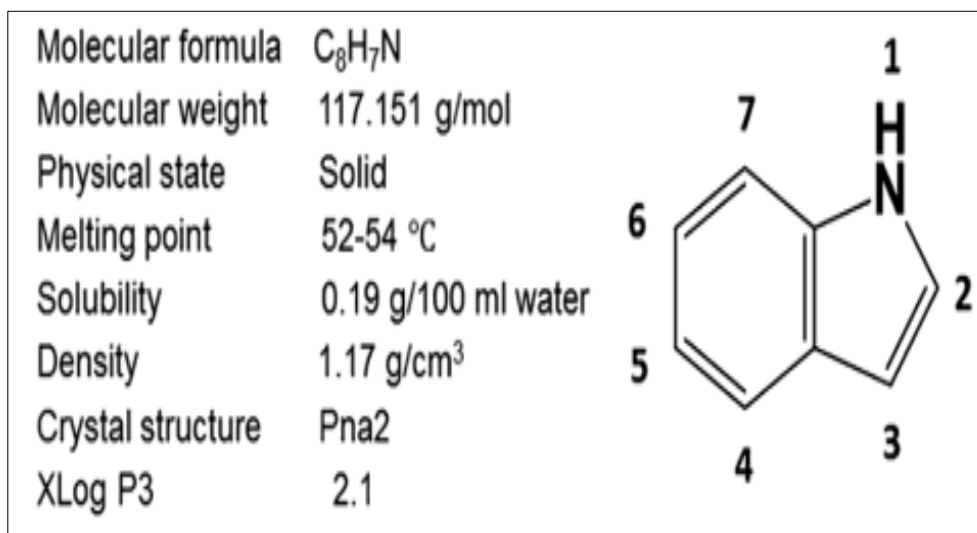
Figure III.1. Pysicochemichal property indole of indole core [5].

Few naturally occurring indole-based compounds are exemplified in Figure III.2. Few fungi such as Fusarium semitectum, Aspergillus fumigatus, Penicillium chrysogenum, and Claviceps purpurea (Ergot) are extensively explored for the synthesis of indole derivatives for the utilization in the healthcare sector. Figure III.2 illustrates some examples of naturally occurring compounds based on indole. Fungi like Fusarium semitectum, Aspergillus fumigatus, Penicillium chrysogenum, and Claviceps purpurea (Ergot) have been extensively studied for their ability to synthesize indole derivatives for application in the healthcare industry.
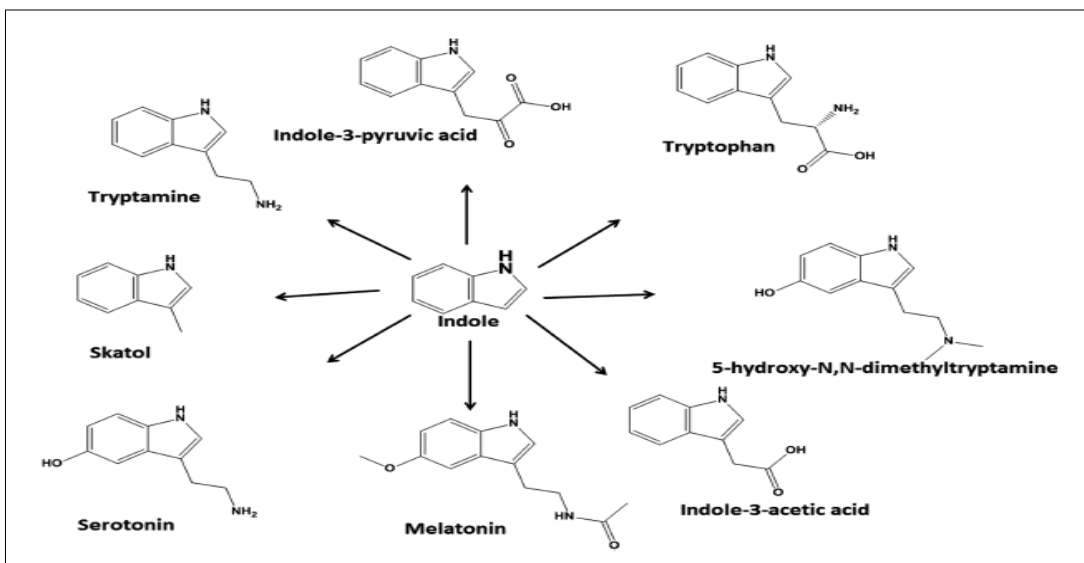
Figure III.2. Natural derivatives comprised indole parent core [5].

Some studies have investigated the antitumor activities of a series of derivatives of 2-phenyl-1H-indole on MDA-MB-231 and MCF-7 breast cancer cells [6, 7]. These derivatives have garnered attention due to their potential to target estrogen receptors associated with both estrogen-sensitive and metastatic breast cancer. Several publications [8-10] have reported that 2-phenylindole derivatives can bind these receptors associated with different types of breast cancer. Previous in silico studies have been carried out on 2-phenyl-1H-indole analogs. Among them are Liao et al [11], Halder et al [12, 13] and El-Nakkady [14]. However, the origins of such activity differences are still not understood; These knowledge gaps may pose challenges in the development of new drugs targeting breast cancer and in the quest for effective treatments for this significant health concern.

This chapter employed quantitative structure-activity relationships (QSAR) analysis on fifty-four molecules of 2-phenyl-1H-indole derivatives to establish new QSAR models utilizing partial least squares (PLS) regression. Subsequently, molecular docking techniques were utilized to predict potential binding interactions between our molecules and selected targets, specifically estrogen and progesterone receptors. Furthermore, molecular dynamics simulations were conducted on the complexes (Receptor-Ligand) exhibiting high negative score energy obtained post molecular docking to validate their stability based on potential energy. Lastly, drug-likeness and ADME parameters were assessed to identify effective and orally bioavailable anticancer compounds.

2. Materials and method

2.1. Biological data

In this chapter, we collected a series of 2-phenyl-1H-indole derivatives, from the literature [6, 7], specifically targeting two breast cancer cell lines: MDA-MB-231 and MCF-7. The reported anti-proliferative activity for these compounds was given as IC_{50} values. Following this, these activities were transformed into pIC_{50} ($pIC_{50} = \log(1/IC_{50})$), serving as the dependent variable for the prediction of a QSAR model. The reported experimental activities are listed in Table III.1.

Table III.1. Chemical structure and experimental antiproliferative activities of the studied ligands [6, 7].
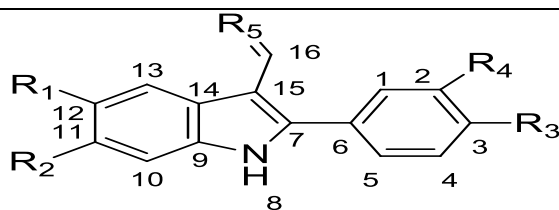
| Ligand | R1 | R2 | R3 | R4 | R5 | IC_{50} | | pIC_{50} | |
						MDA-MB-231	MCF-7	(MDA-MB231)	(MCF-7)
1	OMe	H	OMe	H	O	260	180	6.58	6.74
2	H	OMe	OMe	H	O	35	160	7.45	6.80
3	H	F	OMe	H	O	59	43	7.22	7.37
4	F	H	OMe	H	O	540	240	6.26	6.62
5	H	Cl	OMe	H	O	27	65	7.56	7.19
6	Me	Cl	OMe	H	O	26	62	7.58	7.21
7	Me	H	OMe	H	O	86	140	7.06	6.85
8	Pr	H	OMe	H	O	20	54	7.69	7.27
9	i-Pr	H	OMe	H	O	29	97	7.53	7.01
10	n-Bu	H	OMe	H	O	6.7	22	8.17	7.66

Table III.1. Continued

11	Sec-Bu	H	OMe	H	O	72	180	7.14	6.74
12	t-Bu	H	OMe	H	O	280	580	6.55	6.24
13	n-Pent	H	OMe	H	O	5.5	20	8.26	7.70
14	n-Hex	H	OMe	H	O	7.4	6	8.13	8.22
15	H	OMe	H	OMe	O	1030	390	5.98	6.41
16	H	OMe	OMe	OMe	O	270	20	6.56	7.70
17	H	OMe	OMe	OH	O	800	1650	6.09	5.78
18	H	OMe	Me	H	O	31	100	7.50	7.00
19	H	Cl	Me	H	O	7.8	37	8.10	7.43
20	Me	H	Me	H	O	48	165	7.31	6.78
21	n-Bu	H	Me	H	O	34	54	7.46	7.27
22	n-Bu	H	Et	H	O	27	58	7.56	7.24
23	Et	H	n-Bu	H	O	300	200	6.52	6.70
24	n-Bu	H	F	H	O	350	200	6.45	6.70
25	n-Bu	H	CF3	H	O	33	66	7.48	7.18
26	n-Pent	H	CF3	H	O	42	67	7.37	7.17
27	n-Hex	H	CF3	H	O	43	22	7.36	7.66
28	H	OMe	OMe	H	NMe	34	220	7.46	6.66
29	n-Bu	H	OMe	H	NMe	6	27	8.22	7.57
30	n-Pent	H	OMe	H	NMe	6	21	8.22	7.68
31	n-Bu	H	CF3	H	NMe	32	140	7.49	6.85
32	n-Bu	H	OMe	H	NOH	40	212	7.39	6.67
33	n-Bu	H	CF3	H	NOH	497	1660	6.304	5.78

Table III.1. Continued

34	H	H	H	H	C(CN)2	430	190	6.36	6.72
35	H	H	OMe	H	C(CN)2	720	770	6.14	6.11
36	OMe	H	OMe	H	C(CN)2	590	340	6.22	6.47
37	H	OMe	OMe	H	C(CN)2	260	300	6.58	6.52
38	F	H	OMe	H	C(CN)2	400	310	6.39	6.51
39	H	F	OMe	H	C(CN)2	280	220	6.55	6.66
40	H	OMe	Me	H	C(CN)2	180	250	6.74	6.60
41	Me	H	OMe	H	C(CN)2	280	200	6.55	6.70
42	Me	Cl	OMe	H	C(CN)2	75	200	7.12	6.70
43	n-Pr	H	OMe	H	C(CN)2	83	190	7.08	6.72
44	i-pr	H	OMe	H	C(CN)2	210	310	6.67	6.51
45	n-Bu	H	OMe	H	C(CN)2	26	13	7.58	7.89
46	n-Pent	H	OMe	H	C(CN)2	42	76	7.37	7.12
47	n-Hex	H	OMe	H	C(CN)2	46	25	7.33	7.60
48	n-Bu	H	Me	H	C(CN)2	65	100	7.18	7.00
49	n-Bu	H	Et	H	C(CN)2	76	200	7.11	6.70
50	n-Bu	H	CF3	H	C(CN)2	56	150	7.25	6.82
51	n-Pent	H	CF3	H	C(CN)2	78	100	7.10	7.00
52	n-Hex	H	CF3	H	C(CN)2	150	68	6.82	7.17
53	H	H	H	H	O	420	650	6.37	6.19
54	H	H	OMe	H	O	470	180	6.32	6.74

2.2. QSAR modeling

2.2.1. Molecular descriptor calculation

The 54 molecules under investigation were initially optimized using the Molecular Mechanics force field (MM+) in Hyperchem version 8.0.3 software[15]. Following this, the molecular structures obtained were further refined using the AM1 semi-empirical method [16], employing the Polak-Ribiere conjugate gradient algorithm with an RMS gradient of 0.1 Kcal/(Å.mol). Subsequently, the minimized molecular structures underwent full optimization in an aqueous solution utilizing the CPCM solvation model [17]. This optimization procedure employed the density functional theory (DFT) approach with the UB3LYP functional and the 6-31G(d,p) basis set, implemented in Gaussian 09 [18]. The results obtained from density functional theory (DFT) calculations were utilized to determine quantum chemical descriptors, which included the dipole moment (DM) and atomic charges (qC2, qC3, qC7, qC11, qC12, qC15, qC16, and qN8). The atomic charges were determined using the CHELPG method based on the electrostatic potentials of the molecule [19]. Additionally, the Marvin sketch 19.3.0 package was employed to compute other descriptors such as hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), rotatable bond (Rot-Bond), polarizability (Pol), surface area grid (SAG), molar refractivity (MR), and partition coefficient (logP). The values of these descriptors are listed in Table **III.2**

Table III.2. Selected molecular descriptors.

Ligand	HBA	HBD	Rot-Bond	Pol (Å³)	SAG (Å²)	MR (Å³)	qC2	qC3	qC7	qN9	qC11	qC12	qC15	qC16	DM (Debye)	Log P
1	3	1	4	33.5	499.4	81.67	-0.20	0.38	0.14	-0.44	-0.26	0.35	-0.25	0.44	7.14	3.0
2	3	1	4	33.5	494.2	81.67	-0.19	0.37	0.09	-0.39	0.37	-0.29	-0.22	0.44	9.13	3.0
3	2	1	3	30.6	455.8	75.42	-0.25	0.39	0.16	-0.39	0.30	-0.25	-0.27	0.50	8.54	3.3
4	2	1	3	30.6	464.0	75.42	-0.21	0.39	0.13	-0.41	-0.23	0.28	-0.27	0.45	10.01	3.3
5	2	1	3	32.8	474.5	80.01	-0.22	0.40	0.11	-0.42	0.07	-0.09	-0.25	0.46	9.49	3.8
6	2	1	3	34.4	496.2	85.05	-0.20	0.39	0.12	-0.43	-0.00	0.13	-0.24	0.45	8.84	4.3
7	2	1	3	32.8	491.3	80.25	-0.26	0.40	0.16	-0.38	-0.22	0.15	-0.29	0.50	7.75	3.7
8	2	1	5	36.5	537.1	89.45	-0.26	0.40	0.13	-0.40	-0.17	0.05	-0.27	0.49	7.82	4.6
9	2	1	4	36.5	550.2	89.40	-0.26	0.40	0.18	-0.38	-0.23	0.10	-0.31	0.50	7.85	4.4
10	2	1	6	38.3	579.1	94.05	-0.25	0.39	0.14	-0.38	-0.21	0.12	-0.27	0.50	7.80	5.0

Table III.2. Continued

11	2	1	5	38.3	566.4	94.00	-0.25	0.38	0.15	-0.38	-0.18	0.11	-0.31	0.51	7.81	4.8
12	2	1	4	38.3	537.9	93.87	-0.22	0.40	0.21	-0.49	-0.19	0.00	-0.31	0.47	7.66	4.7
13	2	1	7	40.2	616.7	98.65	-0.25	0.39	0.16	-0.40	-0.18	0.06	-0.29	0.51	7.82	5.4
14	2	1	8	42	651.3	103.3	-0.24	0.39	0.15	-0.38	-0.18	0.07	-0.28	0.50	7.80	5.9
15	3	1	4	33.5	491.0	81.67	0.39	-0.27	0.13	-0.33	0.38	-0.29	-0.28	0.51	8.10	3.0
16	4	1	5	36.0	528.3	88.13	0.23	0.25	0.14	-0.43	0.40	-0.32	-0.24	0.46	7.60	2.8
17	4	2	4	34.0	501.4	83.65	0.28	0.29	0.17	-0.47	0.40	-0.30	-0.26	0.46	7.52	2.7
18	2	1	3	32.8	470.0	80.25	-0.14	0.17	0.11	-0.42	0.39	-0.29	-0.25	0.46	9.82	3.7
19	1	1	3	32.1	457.6	78.59	-0.15	0.16	0.12	-0.44	0.08	-0.11	-0.24	0.45	9.33	4.4
20	1	1	2	32.1	462.8	78.82	-0.15	0.16	0.09	-0.41	-0.21	0.16	-0.25	0.46	8.24	4.3
21	1	1	5	37.6	551.6	92.63	-0.21	0.18	0.14	-0.38	-0.21	0.12	-0.28	0.50	8.32	5.7
22	1	1	6	39.5	593.1	97.23	-0.15	0.05	0.08	-0.38	-0.21	0.10	-0.24	0.49	8.20	6.1
23	1	1	6	38.5	579.1	97.23	-0.18	0.15	0.12	-0.37	-0.18	0.02	-0.29	0.51	8.22	6.1
24	1	1	5	35.4	530.6	87.8	-0.22	0.31	0.14	-0.4	-0.21	0.10	-0.27	0.49	7.13	5.3
25	1	1	6	36.5	567.7	93.56	-0.09	-0.02	0.12	-0.39	-0.21	0.11	-0.26	0.49	6.76	6.0
26	1	1	7	38.4	618.5	98.16	-0.09	-0.02	0.13	-0.38	-0.19	0.08	-0.28	0.51	6.52	6.5
27	1	1	8	40.2	618.8	102.80	-0.09	-0.03	0.09	-0.37	-0.17	0.09	-0.26	0.50	6.74	6.9
28	3	1	4	36	521.9	88.62	-0.22	0.37	0.09	-0.44	0.37	-0.30	-0.18	0.24	6.74	3.2
29	2	1	6	40.9	628.2	101.00	-0.24	0.38	0.12	-0.41	-0.22	0.07	-0.25	0.32	5.71	5.2
30	2	1	7	42.7	637.4	105.60	-0.26	0.39	0.13	-0.41	-0.22	0.07	-0.25	0.32	5.69	5.7
31	1	1	6	39.0	604.7	100.50	-0.11	-0.01	0.08	-0.38	-0.23	0.12	-0.25	0.33	3.16	6.3
32	3	2	6	39.6	584.8	97.87	-0.25	0.37	0.14	-0.40	-0.20	0.08	-0.29	0.33	6.19	5.0
33	2	2	6	37.9	583.2	97.38	-0.09	-0.03	0.1	-0.41	-0.20	0.09	-0.28	0.35	3.91	6.0
34	2	1	2	33.5	462.6	83.19	-0.05	-0.08	0.07	-0.41	-0.10	-0.09	-0.11	0.10	14.58	3.8
35	3	1	3	36.0	507.8	89.66	-0.22	0.40	0.09	-0.41	-0.12	-0.06	-0.12	0.10	17.27	3.6
36	4	1	4	38.4	535.1	96.12	-0.28	0.39	0.09	-0.31	-0.24	0.35	-0.09	0.10	14.67	3.5
37	4	1	4	38.4	548.3	96.12	-0.20	0.38	0.08	-0.35	0.39	-0.27	-0.09	0.09	18.88	3.5
38	3	1	3	35.5	512.9	89.87	-0.21	0.41	0.10	-0.38	-0.23	0.32	-0.14	0.12	18.55	3.8
39	3	1	3	35.5	513.1	89.87	-0.28	0.40	0.07	-0.33	0.32	-0.25	-0.06	0.08	15.57	3.8
40	3	1	3	37.7	532.9	94.70	-0.15	0.19	0.02	-0.36	0.40	-0.27	-0.04	0.05	16.66	4.1
41	3	1	3	37.7	529.6	94.70	-0.21	0.40	0.10	-0.39	-0.22	0.19	-0.14	0.13	17.06	4.1
42	3	1	3	39.5	549.3	99.50	-0.22	0.41	0.04	-0.34	-0.02	0.20	-0.06	0.08	16.32	4.8
43	3	1	5	41.4	591.6	103.90	-0.27	0.39	0.12	-0.36	-0.19	0.11	-0.11	0.11	15.90	5.0
44	3	1	4	41.4	577.9	103.50	-0.28	0.41	0.12	-0.34	-0.19	0.08	-0.13	0.12	15.88	4.9
45	3	1	6	43.2	632.3	108.50	-0.28	0.42	0.12	-0.35	-0.17	0.11	-0.10	0.11	13.32	5.5
46	3	1	7	45.1	649.0	113.10	-0.29	0.40	0.07	-0.32	-0.16	0.08	-0.08	0.10	15.80	5.9

Table III.2. Continued

47	3	1	8	46.9	702.8	117.70	-0.27	0.41	0.09	-0.33	-0.21	0.13	-0.10	0.11	13.34	6.3
48	2	1	5	42.5	613.8	107.08	-0.22	0.19	0.10	-0.34	-0.20	0.15	-0.09	0.10	14.13	6.2
49	2	1	6	44.4	642.7	111.68	-0.16	0.06	0.10	-0.34	-0.21	0.15	-0.10	0.12	14.12	6.6
50	2	1	6	41.5	631.4	108.01	-0.09	-0.03	0.04	-0.31	-0.19	0.15	-0.07	0.10	12.67	6.5
51	2	1	7	43.3	661.2	112.61	-0.10	-0.01	0.07	-0.34	-0.16	0.09	-0.09	0.11	12.68	7.0
52	2	1	8	45.2	665.5	117.21	-0.13	-0.00	0.04	-0.33	-0.16	0.09	-0.07	0.10	12.50	7.4
53	1	1	2	28.5	406.1	68.74	-0.03	-0.09	0.17	-0.51	-0.07	-0.13	-0.28	0.47	7.94	3.3
54	2	1	3	31.0	450.4	75.21	-0.21	0.40	0.19	-0.51	-0.08	-0.13	-0.29	0.46	8.06	3.1

2.2.2. Regression analysis

To elucidate a mathematical model correlating the biological activities of 2-phenyl-1H-indole derivatives with their molecular descriptors, quantitative structure-activity relationships (QSAR) analysis was conducted using partial least squares (PLS) regression. All statistical calculations for this analysis were performed using Matlab software [20].

2.2.3. Molecular descriptors selection

The proposed model should be explicitly correlated with the physicochemical, biological, and toxicological properties of molecules under investigation [21, 22]. The selection of molecular descriptors has a crucial effect on the high accuracy of affinity prediction [23]. To obtain a valid QSAR model, several descriptors were selected and then used as independent variables. Firstly, Pearson's correlation matrix was performed between each parameter and the biological response (activity). Using a correlation significance test, the Pearson's p values were determined to test the null hypothesis against the alternative that there was a nonzero correlation. The indices with a very weak correlation coefficient r with the response were eliminated, whereas descriptors with high dependence on the response were selected.

Variable selection by the Stepwise regression method is used to identify the best subset of the molecular descriptors. This combination of backward and forward selections uses the minimum number of descriptors to develop a good predictive model. Thus, we must select good subsets of descriptors. However, it should be noted that the subset of molecular descriptors that do the best at meeting well-defined objective criteria can be highly varied depending on precisely which observations are included in the training set. In addition, the best

training model does not necessarily guarantee a better quality of prediction. This depends on the training and test sets obtained from the original dataset; for this reason, we conducted a statistical simulation for which 10000 splits were performed, resulting in 10000 training and test sets. Firstly, the best model was selected for each training set resulting in 10000 best training models following the Bayesian Information Criterion (BIC) [24, 25].

2.2.4. Partial least square

The Partial least square (PLS) method is an optimum choice when there are many intercorrelated descriptors for a particular number of a data set [51]. In this work, regression diagnoses for identifying possible outliers were performed by computing the leverage values (hii) and the studentized deleted residuals (ri). The diagonal elements of the hat matrix, which are indicated by the leverage values, are used to define the outlying observations of X. Leverage values greater than 2(p+1= n) are considered to indicate outlying cases concerning their X values. On the other side, the magnitude of the student residues is used to classify the outlying Y observation [24, 26].

2.2.5. Validation for QSAR models.

2.2.5.1. Internal validation

The QSAR model must be validated properly for better biological activity prediction; the model's validity can be assessed using internal validation, including least-squares fit (R^2), adjusted R2 (R^2_{adj}), the Fisher test (F), and cross-validation coefficient (Q^2_{cv}) for the significance of the regression equation [27, 28].

2.2.5.2. External validation

External validation of the test set was carried out using standard deviation error in prediction (SDEP), squared correlation coefficient (R^2_{pred}), and squared bootstrapping correlation coefficient (R^2_{boot})[27, 29].

2.3. Molecular docking protocol

Molecular docking was carried out to study the possible interactions between the analyzed ligands and active sites of receptors and to see the best binding modes with the high affinity of receptor-ligands (i.e., low energy score). The ligands were considered flexible structures, whereas the protein receptors were rigid. All the docking calculations were performed using MOE software [30], passing by two major steps:

2.3.1. Preparation of ligands

The optimized molecular structures of the relevant ligands were converted and saved as 3D protein structure database (PDB) files, which were then employed as input files for MOE-docking.

2.3.2. Selection and preparation of target

Kaufmann *et al.* have reported the existence of cytotoxicity in breast cancer cell line (MCF-7) with estrogen receptor status ER+ and progesteron receptor status PR+ [6]. Consequently, X-ray crystal structures of both hormone receptors were selected. The first one is the estrogen receptor alpha (ER-α) complexed with the estradiol (PDB ID:1A52) [31], and the second is the human progesterone receptor (PR) (PDB ID:1A28) [32]. These structures were downloaded from the RCSB Protein Data Bank (http://www.rcsb.org/pdb). Both hormone receptors, 1A52 and 1A28, are present in hormone-positive breast cancer cells, with each receptor comprising 2 chains (A) and (B) and having resolutions of 2.80 Å and 1.80 Å, respectively. The molecular structures of both receptors were prepared using the following steps:

One of the co-crystallized ligands, one sequence of chains, and all contained ions were removed. Several studies have found that water molecules play an important role in the biological environment of proteins and ligand binding [33-36]. Therefore, water molecules have been kept in our study. The analysis was performed on-chain (A) for both enzymes. The selected parameters and the docking protocol were set according to previous studies reported by Daoud and co-workers [37-39].

To validate the docking method, the native ligands: estradiol and progesterone were re-docked into the binding site pocket of ER-α and progesterone receptors, respectively. The root-mean-square deviation (RMSD) values were calculated to justify the accuracy of this method. As presented in table III.3, the score energies of the complexes 1A52-estradiol and 1A28-progesterone are -7.2805 Kcal/mol and -10.6014 Kcal/mol, while the RMSD values are: 1.0286 Å and 1.1595 Å, respectively, which belong to the optimal RMSD value (less than 2 Å) [40]. After enzymes validation, all the ligands were docked into the binding site pocket of both ER-α and PR.

Table III.3. Some information of both proteins ER-α and PR.

Receptor	Resolution(Å)	Docking score (kcal/mol)	Native ligand	Chain	Residues	RMSD(Å)
1A52	2.80	-7.280	Estradiol	A,B	258	1.028
1A28	1.80	-10.601	Progesterone	A,B	256	1.159

2.4. Molecular dynamics simulation (MD)

Molecular dynamics (MD) simulation is a powerful computational tool for better understanding the biological, macromolecular, and dynamic behavior of proteins at different times [41].

In this study, MD simulations using MOE software [30] were conducted to determine the stability of molecular interactions between receptor and ligands over time since the best conformations of both complexes ER-α and PR resulted from docking have been considered. Nosé-Poincaré-Andersen (NPA) [42] algorithm was explored in these simulations thanks to its high accuracy and satisfied sensitivity [43]. The energy was minimized using MMFF94X potential energy function with a gradient of 0.1 RMS Kcal/mol.Å. In MOE protocol; different steps were selected: for a total of 1000 Ps the equilibrium and production were set to 100 Ps and 900 Ps, respectively at 300 K. The obtained MD results were presented in graphs to interpret the potential energy variations as a function of time, these graphs were plotted with Matplotlib in python [44].

2.5. ADME drug-likeness and pharmacokinetics

The prediction and pharmacokinetics of ADME (absorption, distribution, metabolism, and excretion) are important criteria in drug design and discovery [45, 46]. SwissADME online calculations (http://www.swissadme.ch) were used to evaluate the drug-likeness and pharmacokinetic properties.

To estimate the ability of a molecule to have oral bioavailability, the drug-likeness prediction based on different rules, namely Lipinski[47], Veber [48], and Egan [49], rules have been tested for each candidate.

The pharmacokinetics parameters predicted in this study are gastrointestinal absorption, P-gp (P-glycoprotein) substrate, and blood-brain barrier were reported in this study. P-gp is a transmembrane responsible for the efflux of many xenobiotics and toxic substances [50]. It can also efflux many drugs out of the cells, which reduces the activity of many drugs [46]. Since drugs can inhibit or induce cytochrome P450 enzymes, they play an important role in drug metabolism. CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4 enzymes are essential in the human body; indeed, this class contains over 50 enzymes six of them metabolize 90% of drugs, notably CYP3A4 and CYP2D6 are the two most significant enzymes [51, 52].

In this study, SwissADME [53] served as a valuable tool for predicting the effects of the studied molecules on these enzyme types. Furthermore, the BOILED-Egg [54] method was employed to explore passive human gastrointestinal absorption (HIA) and blood-brain barrier (BBB) permeation.

3. Results and discussion

3.1. QSAR modeling

A QSAR study was carried out on dataset comprises a total of 54 compounds utilized for constructing the PLS regression model and assessing its performance. Firstly, to verify a model's predictive ability, the dataset was divided into a training dataset and a testing dataset 10000 times. 75 % of the molecules under probe were used in the training set to create the QSAR model, while the rest of the molecules were the subject of the test set for external validation. These sets are randomly extracted from a larger pool and distributed differently concerning the model response domain.

Figures III.3(A) and III.4(A) show the results of the simulation obtained for the first antiproliferative activity (MDA- MB 231). Figures III.3(B) and III.4(B) are for the second antiproliferative activity (MCF-7). According to Figure III.3, the best model is the one that contains four descriptors; it is the size of the subset having the lowest value of the BIC criterion. This last criterion was chosen because it penalizes larger models more heavily and tends to select a smaller subset of descriptors compared to other criteria [55]. Each box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are points that

have values beyond the ends of the whiskers. The best choice of descriptors will balance fit with model size. The next step is determining the molecular descriptors retained in this subset. Indeed, according to Figure III.4, the best subsets of molecular descriptors are those with the highest probability of occurrence according to the BIC criterion.

Initially, both HBA and MR descriptors were eliminated because they are highly correlated, and the correlation coefficient r with the response is very weak. Consequently, we used fourteen molecular descriptors in the QSAR study for both biological activities.

Figure III.4 (A) reveals that the following variables: X2 (Rot - Bond), X13 (MD), X14 (LogP), and X5 (qC2) have the highest frequency for the first antiproliferative activity against cancer lines MDA-MB231. While In figure III.4 (B), X2 (Rot-Bond), X1 (HBD), X5 (qC2), X13 (DM), and X14 (Log P) show high frequencies for the second antiproliferative activity
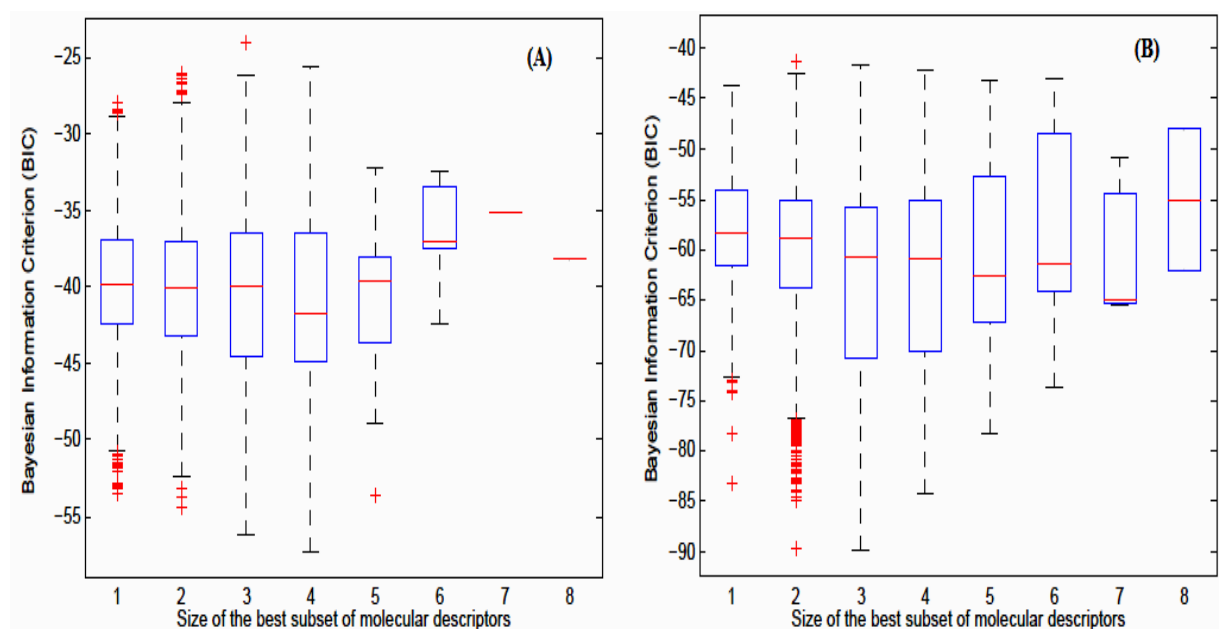
against cancer lines MCF-7.

Figure III.3. Box plots of the BIC values calculated from 10000 random training sets from the antiproliferative activity against the breast cancer cells line: (A) MDA- MB231, (B) MCF-7.
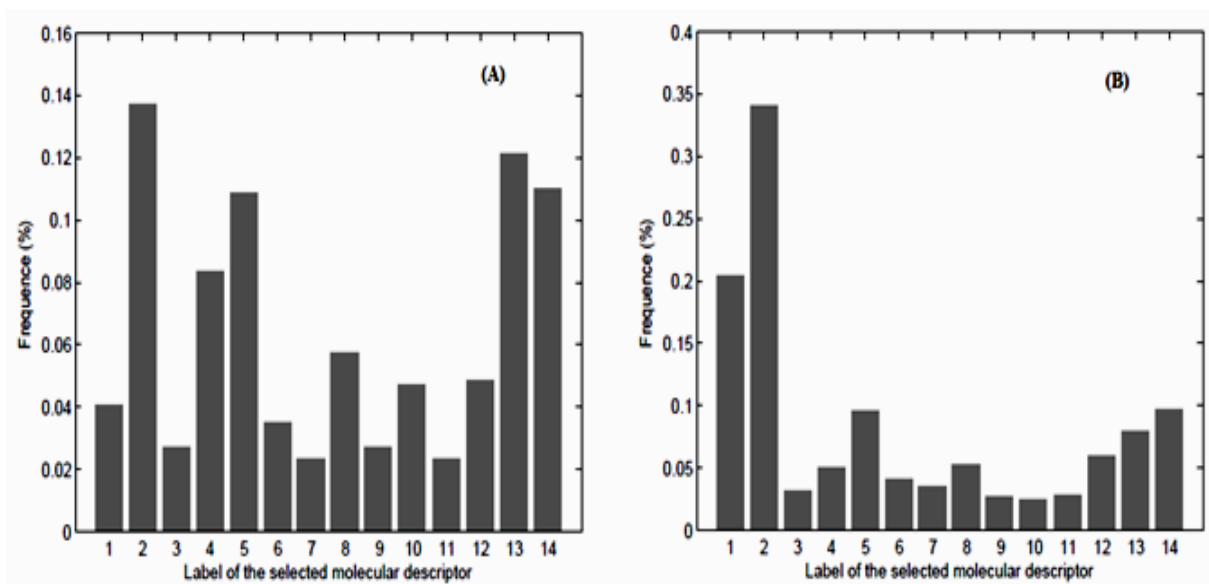
Figure III.4. Frequency of occurrence of molecular descriptors from the antiproliferative activity against the breast cancer cells line: (A) MDA-MB231, (B) MCF-7.

The values of the observed antiproliferative activities $y = pIC_{50}/Sy$ and those determined by the PLS regression are collected in Tables III.4 and III.5. A strong correlation between the experimental (y_{exp}) and the predicted values (y_{fit}). For the first biological activity displayed in Table III.4, we observed from the studentized deleted residual (r_i) values that all the observations of training set subsets have studentized deleted residuals between -2 and 2, except for the values of the observations 24, 12, and 4 which are greater than the threshold |2|. On the other hand, the observations of test set subsets have studentized deleted residual (r_i) values between -1 and 1, except for observation 19. The observations 24, 12, 4, and 19 are considered outlying data points; these points were not considered during the predictive model determination. For the second biological activity (MCF-7), observations 12, 16, 33, and 8 are outlying data points. To highlight the weight of each molecular descriptor, the PLS regression models are written with scaled variables. The standardized regression coefficient value of each descriptor highlights the relative importance of the descriptors in the determination of the biological activity of the ligands under study.

Table III.4. Diagnostic statistics for regression of biological activity of MDA-MB231.

Ligand	Y_{exp}	Y_{fit}	h_{ii}	r_i	Ligand	Y_{exp}	Y_{fit}	h_{ii}	r_i
Training set					**Training set**				
L2	12.348	11.941	0.133	0.885	L40	11.170	10.742	0.094	0.912
L5	12.535	12.009	0.082	1.120	L41	10.852	10.928	0.096	-0.158
L6	12.561	12.109	0.094	0.965	L42	11.800	11.134	0.109	1.457
L8	12.750	12.690	0.071	0.126	L43	11.727	11.612	0.077	0.239
L9	12.484	12.580	0.082	-0.201	L44	11.059	11.553	0.075	-1.045
L10	13.537	12.817	0.077	1.556	L46	12.217	11.986	0.156	0.505
L14	13.466	13.088	0.158	0.833	L47	12.151	12.423	0.180	-0.605
L15	9.915	9.878	0.375	0.094	L48	11.902	11.837	0.101	0.138
L16	10.879	10.621	0.283	0.612	L49	11.790	11.760	0.129	0.063
L17	10.097	10.323	0.277	-0.533	L50	12.010	11.693	0.133	0.685
L18	12.436	11.640	0.063	1.719	L51	11.771	11.872	0.167	-0.221
L20	12.121	11.915	0.201	0.462	L53	10.561	11.383	0.140	-1.870
L21	12.369	12.566	0.068	-0.409	**Test set**				
L22	12.535	12.514	0.073	0.043	L1	11.118	11.733	0.625	-1.881
L25	12.389	12.494	0.091	-0.220	L3	12.206	11.482	0.214	1.397
L26	12.217	12.684	0.115	-1.009	L7	11.930	11.528	0.370	0.792
L28	12.369	12.442	0.148	-0.157	L11	12.060	12.374	0.223	-0.542
L29	13.616	13.124	0.092	1.052	L13	13.946	14.101	0.497	1.0153
L30	13.616	13.331	0.119	0.612	L23	11.013	11.666	0.251	-1.260
L31	12.412	13.119	0.173	-1.618	L27	12.439	11.685	0.405	1.802
L32	12.252	13.072	0.098	-1.815	L33	10.644	11.051	0.499	-0.153
L34	10.544	10.498	0.135	0.098	L38	10.802	10.554	0.787	0.848
L35	10.173	10.851	0.108	-1.486	L45	12.807	13.071	0.441	-0.538
L36	10.316	11.530	0.125	-2.926	L52	11.522	11.790	0.432	-0.542
L37	10.905	10.587	0.188	0.711	L54	10.684	11.073	0.2513	-0.693
L39	10.852	11.359	0.095	-1.085					

Table III.5. Diagnostic statistics for regression of the biological activity MCF-7.

Ligand	Y_{exp}	Y_{FIT}	h_{ii}	r_i	Ligand	Y_{exp}	Y_{FIT}	h_{ii}	r_i
Training set					**Training set**				
L2	14.412	14.704	0.158	-0.519	**L40**	13.988	13.670	0.093	0.545
L4	14.030	14.377	0.088	-0.592	**L41**	14.200	13.786	0.096	0.713
L5	15.239	14.434	0.078	1.404	**L42**	14.200	13.851	0.118	0.608
L6	15.281	14.443	0.090	1.476	**L43**	14.242	14.601	0.077	-0.609
L9	14.857	14.956	0.0797	-0.167	**L44**	13.797	14.335	0.080	-0.923
L10	16.235	15.531	0.085	1.223	**L46**	15.090	15.232	0.156	-0.251
L14	17.422	16.101	0.178	2.048	**L47**	16.108	15.693	0.184	0.751
L15	13.585	13.438	0.591	0.376	**L48**	14.836	14.622	0.110	0.368
L17	12.250	12.223	0.630	0.073	**L49**	14.200	14.778	0.129	-1.023
L18	14.836	14.219	0.063	1.053	**L50**	14.454	14.735	0.131	-0.491
L20	14.370	14.071	0.200	0.544	**L51**	14.836	15.046	0.164	-0.374
L21	15.408	15.076	0.067	0.5622	**L53**	13.119	13.840	0.1414	-1.297
L22	15.345	15.245	0.071	0.168	**Test set**				
L24	14.200	15.202	0.065	-1.765	**L1**	14.321	14.603	0.673	-0.644
L25	15.217	15.231	0.092	-0.022	**L3**	15.660	14.317	0.983	0.911
L26	15.196	15.547	0.119	-0.611	**L7**	14.555	14.640	0.386	-0.374
L28	14.115	14.973	0.163	-1.584	**L11**	14.321	14.225	0.248	-1.166
L29	16.044	15.683	0.097	0.621	**L13**	16.361	15.531	0.393	0.630
L30	16.277	16.012	0.128	0.463	**L19**	15.787	14.625	0.539	2.022
L31	14.518	15.567	0.170	-1.987	**L23**	14.236	14.453	0.274	-1.479
L32	14.136	14.164	0.630	-0.073	**L27**	16.276	15.024	0.538	0.776
L34	14.242	13.297	0.135	1.729	**L38**	13.832	14.050	0.784	-1.127
L35	12.950	13.788	0.106	-1.492	**L45**	16.765	14.930	0.474	1.684
L36	13.713	14.444	0.127	-1.306	**L52**	15.235	14.641	0.455	-0.550
L37	13.819	13.896	0.196	-0.140	**L54**	14.321	14.326	0.248	-0.307
L39	14.115	14.074	0.093	0.070					

The PLS regression equations obtained are expressed as follows:

$$y'_{(\text{MDA MB231})} = 11.86(\pm 0.19) + 0.16(\pm 0.11)x'_2 - 0.59(\pm 0.09)x'_5 -$$
$$0.63(\pm 0.09)x'_{13} + 0.16(\pm 0.11)x'_{14} \qquad\qquad \text{Eq. (1)}$$

$$y'_{(\text{MCF}-7)} = 14.41(\pm 0.42) - 0.33(\pm 0.11)x'_1 + 0.56(\pm 0.17)x'_2 - 0.29(\pm 0.10)x'_5 -$$
$$0.30(\pm 0.10)x'_{13} - 0.05(\pm 0.18)x'_{14} \qquad\qquad \text{Eq. (2)}$$

Where $y' = pIC_{50}/S_y$ and $x' = X_j/S_{xj}$, S_y and S_{xj} are the standard deviations corresponding to the biological response and the *jth* descriptor, respectively.

The selected molecular descriptors are X_1 = HBD, X_2 = Rot-Bond, X_5 = qC2, X_{13} = DM and X_{14} = Log P.

According to the goodness of fit statistics summarized in Table III.6 , 79% and 63% of the variability in anticancer activity around its mean are explained by the PLS regression equations (1) and (2), respectively. The F-statistics reveal the significance of the PLS regression equations. F-values were calculated with different numbers of degrees of freedom; therefore, the models cannot be compared. The numbers in parentheses are the corresponding degrees of freedom. High values of the F indicate that model is statistically significant.

In order to test the validity of the predictive power of the developed models, the leave-one-out (LOO) technique was used. The obtained values of R^2cv are greater than 0.5 [29, 56] which indicates that the best qualification of the QSAR models 1 and 2. Furthermore, it is assumed that a value of   is greater than 0.50 may be taken as an indicator of good external predictability [56] (see Table III.6).

Equations (1) and (2) exhibit a high value of and a value of greater than 0.50; this result confirms that the resulting QSAR models have good external predictability and robustness (Table III.6). Considering the predictive performance of the two PLS models, however, the predictive power of equation (1) appears to be higher compared to equation (2).

Table III.6. Statistical features of the obtained models.

Models	Goodness of fit				Goodness of prediction		
	R^2	R^2_{adj}	R^2_{cv}	F^b_{obs}	R^2_{pred}	R^2_{boot}	SDEP^a
Models 1 (n_{tr} = 38; n_{ts} =12)	0.79	0.76	0.78	29.97(3;33)	0.69	0.70	0.465
Models 2 (n_{tr} = 38; n_{ts}=12)	0.63	0.59	0.63	11.31(5;33)	0.61	0.60	0.536

From equations (1) and (2), the positive coefficient of X_2 (Rot-Bond) indicates that any increase in the number of rotatable bonds (an increase in the flexibility of molecules) causes an increase in anti-proliferative activity against two cancer cells lines MDA MB231 and MCF-7. According to Kaufmann and colleagues [6] who reported that the extension of the alkyl chain from methyl to pentyl decreased the IC_{50}, which confirmed that these theoretical results are in full agreement with the experimental findings. In equations (1), the positive sign of X14 (log P) means that highly hydrophobic groups are required to enhance the anti-proliferative activity against cancer lines MDA MB231. On the other hand, we note that from equation (2), log P has a negligible role in anticancer activity because of its low coefficient (0.05).

The negative values of descriptors X5 (qC2) and X13 (DM) showed that increasing of these parameters is detrimental to the anti-proliferative activity against both cancer cells lines.

However, the electron-withdrawing substitutes in position C2 of the phenyl ring in 2-phenyl-1H-indole derivatives decrease the anti-proliferative activity against two cancer cells lines MDA-MB231 and MCF-7.

In equation (1), the negative coefficient of X1 (HBD) explains that any increase in this parameter causes a decrease in anti-proliferative activity against the cancer cells lines MDA-MB231.

3.2. Molecular docking

3.2.1. Identification of the active site of ER and PR

The key active site residues of ER-α pocket are His524 (A), Glu353 (A), Arg394 (A), Leu387(A), and Met388(A) [57]. The catalytic amino acid triads are His524(A), Arg394(A), and Glu353(A) [58, 59]. Several researchers [31, 58-61] have reported that ER is special among the steroid receptors in its capacity to grasp a wide variety of non-steroidal compounds, The size of the ligand-binding domain cavity (450 A°3), which is about twice compared to estradiol's molecular volume (245 A°3). The "Site Finder" option in MOE was utilized to identify the essential residues constituting the active site of ER-α (PDB: 1A52) complexed with estradiol. These residues include Glu323(A), Leu346(A), Glu353(A), Leu384(A), Leu387(A), Met388(A), His524(A), and Leu525(A). On the other hand, for PR (PDB: 1A28), the active site residues identified are Gln725(A), Leu715(A), Leu718(A), Met801(A), Met756(A), Met759(A), Leu763(A), Arg766(A), Phe778(A), Met801(A), Asn719(A), Gly722(A) and Cys891(A).

The co-crystallized ligand (estradiol) of ER-α formed two interactions with residues, the first one hydrogen bond which established between the oxygen atom of the hydroxyl group of estradiol and the Glu353(A), the second type is pi-H interacted between the aromatic ring of estradiol and the Leu387(A) (Figure III.5.a). The co-crystallized ligand (progesterone) of PR interacted with a ketone group at C3 position with NH2 group of Gln725(A) forming a hydrogen bond. Another hydrogen bond was observed by the interaction of the NH2 group of Arg766(A) (Figure III.5.b).

3.2.2. Interaction between ligands and both receptors

Docking calculation results of the energy score and interactions between the best ligands and the active site residues of ER-α and PR are listed in Tables III.7 and III.8, respectively. Likewise, the results of the rest 2-phenyl-1H-indole derivatives are provided in Appendix A and Appendix B.

The optimal pose of each ligand could be achieved at the lowest energy score and the interaction of the ligands with the amino acid residues of the receptor. The molecular

interactions of the ligands with the active site could be visualized by using: 'Ligand Interaction' option implemented in MOE software.
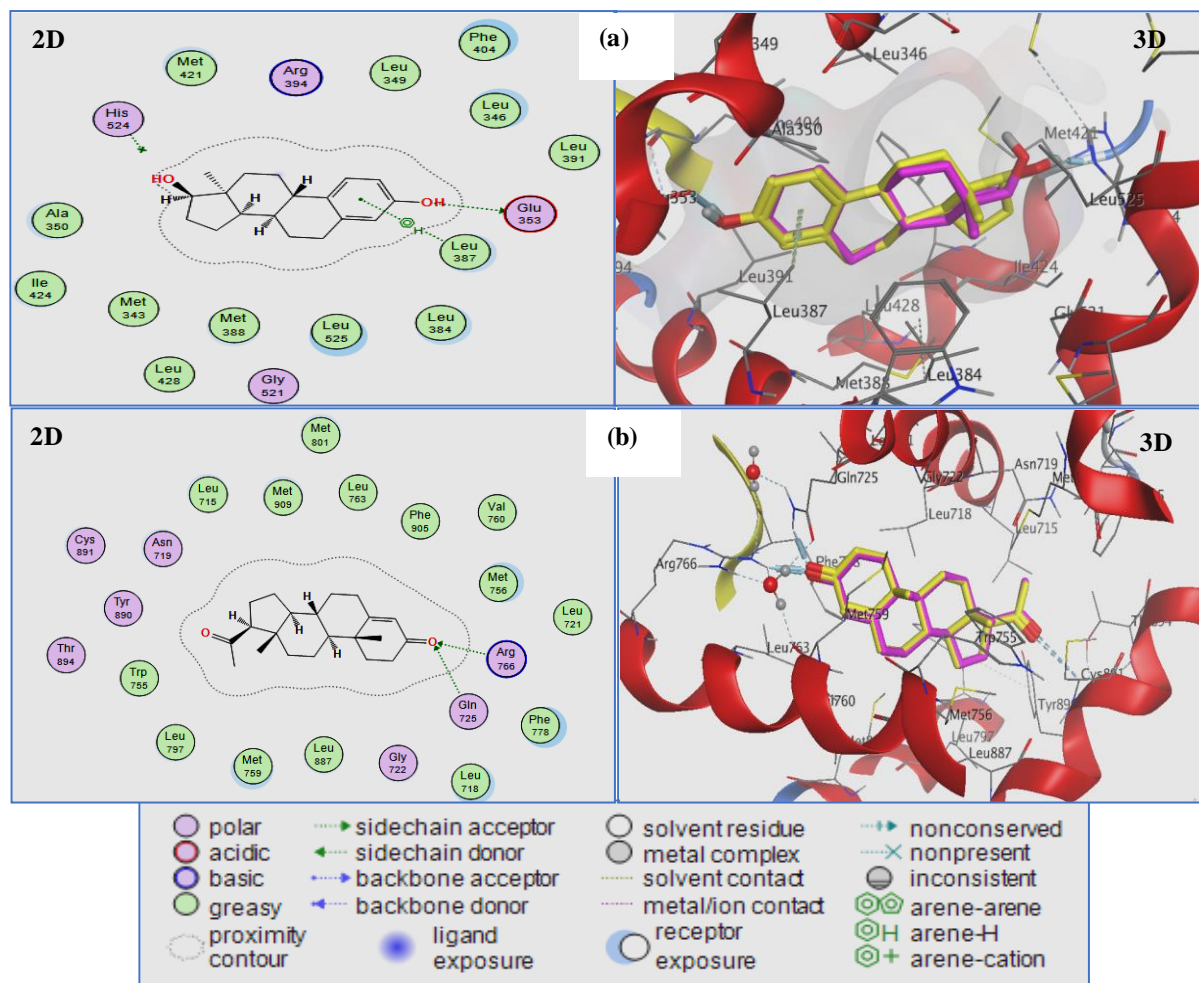
Figure III.5. 2D and 3D visualization of the best poses of ER-α with estradiol (a) and the PR with progesterone (b).

a) **Interactions between ligands and ER-α target**

According to the literature, A.Imberty et al [62] revealed that if the distance between hydrogen bond belongs to the interval between 2.5 and 3.1 Å, are considered strong. However, the interactions between 3.1 and 3.5 Å are supposed weak.

The molecular docking results summarized in Appendix A showed that almost all tested ligands had a binding score ranging between -5.825 and -7.408 kcal/mol. Besides, two main

interactions such as hydrogen bond and hydrophobic interaction (H-Pi) appeared between the ligands and ER- α target compared to the co-crystallized ligand.

Table III.7. Docking score and interactions between best ligands and the active site residues of ER.

| Ligand | S-score (kcal/ mol) | Bonds between ligands and of the active site residues | | | | | |
		Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic Distance (Å)	Bond energy (Kcal/mol)
L37	-7.307	5-ring	CE2	PHE404(A)	Pi-H	4.15	-0.6
L39	-7.368	/	/	/	/	/	/
L40	-7.349	N28	OG1	THR347(A)	H-Acceptor	2.90	-1.0
		6-ring	CD1	LEU387(A)	Pi-H	3.90	-0.8
L48	-7.408	/	/	/	/	/	/
Lref	-7.280	O 36	OE1	GLU353(A)	H-donor	2.84	-5.1
Estrdiol		6-ring	CD1	LEU387(A)	pi-H	4.14	-0.6

As observed in Table III.7, the ligands L37 and L40 have the best docking score compared to the native ligand (docking score of estradiol = -7.280) and their values are -7.307and -7.349 Kcal/mol, respectively. We note that ligand L40 interacts with two active site residues by forming a strong [62] hydrogen bond acceptor with the Thr347(A) at distance: 2.9Å and Pi-H bond with residue Leu387(A). Knowing that the native ligand established the same interactions with the same residue. Similarly, the ligand L37 established the hydrophobic interaction Pi-H with the residue Phe404(A) (see Figure III.6).

Besides, we note clearly that the ligands L39 and L48 exhibited low score energy values: -7.368 and -7.408 Kcal/mol, respectively, compared to all other ligands and the co-crystallized ligand, but they are not involved in making interactions with the active site residues of ER-α. In addition, according to the literature ligand L48 has a low value of IC50 (65 nM) (see Table III.7). For this reason, it is usually recommended to perform MD simulations between this ligand and the ER-α pocket in order to evaluate the stability of docked complexes [63].
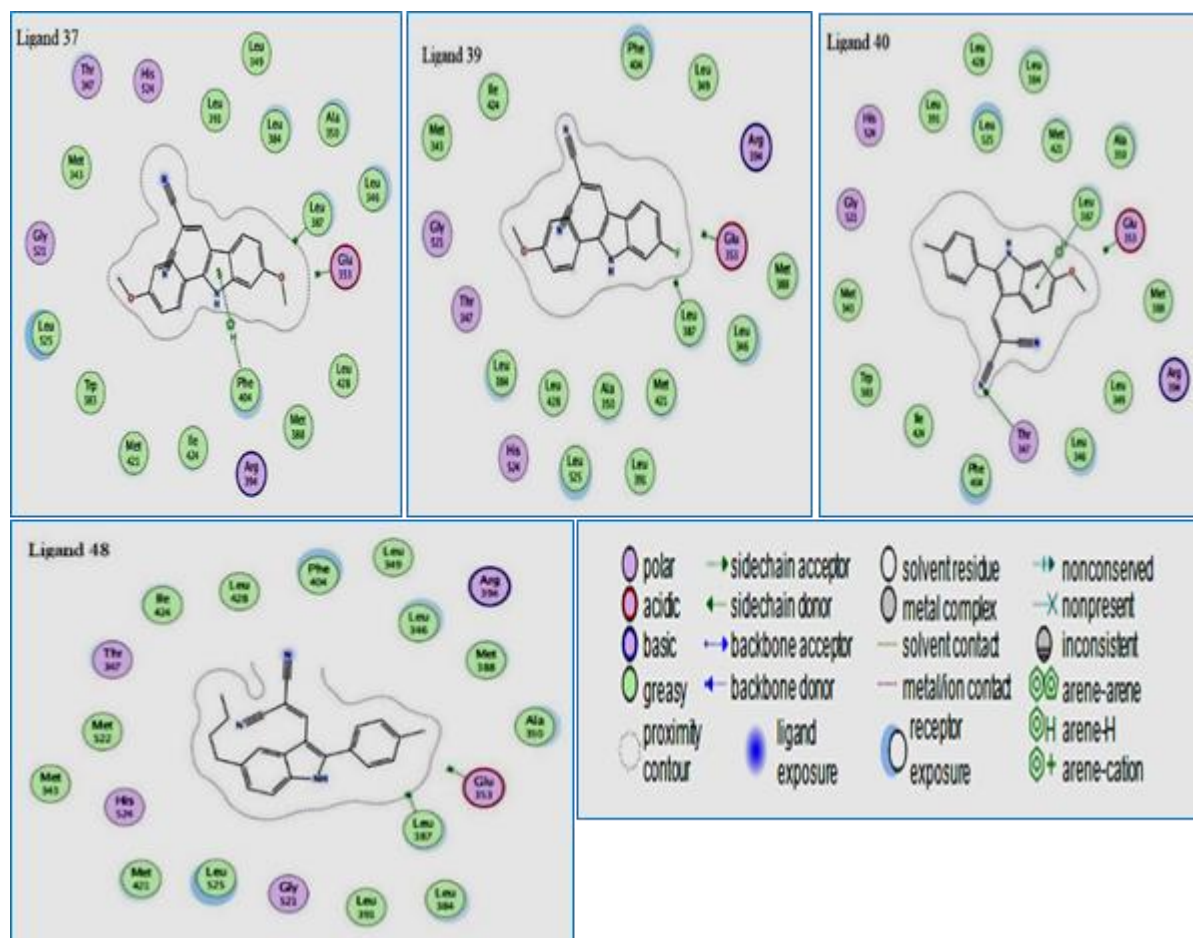
Figure III.6. 2D visualization of the interactions between the best ligands L37, L39, L40 and L48 with ER-α.

b) **Interactions between ligands and PR target**

As presented in Appendix B, most of the complexes give score energy varies from -7.7667 to -9.9522 kcal/mol. It can be seen that H-bond and H-Pi are the main interaction modes between ligands and the active site residues of PR.

Table III.8. Docking score and interactions between the best ligands and the active site residues of PR.

Ligand	S-score (kcal/mol)	Bond between ligand and of the active site residues					
		Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic Distance (Å)	Bond energy (Kcal/mol)
L22	-9.952	N 9	SD	MET801(A)	H-donor	3.73	-2.7
L23	-9.261	N 9	SD	MET801(A)	H-donor	3.75	-3.3
		5-ring	CE2	PHE778(A)	pi-H	4.47	-0.9
L26	-9.053	N 9	SD	MET801(A)	H-donor	4.17	-1.0
L47	-9.356	N 27	CA	MET756(A)	H-acceptor	3.65	-0.7
		N 28	CA	VAL760(A)	H-acceptor	3.29	-1.1
L$_{ref}$ progesterone	-10.601	O38	NE2	GLN725(A)	H- acceptor	3.28	-1.4
		O38	NH2	ARG766(A)	H-acceptor	2.79	-4.2

It is evident that the ligands L22, L23, L26, and L47 have the highest negative scores - 9.952, -9.261, -9.053 and -9.356 Kcal/mol, respectively. Ligands L22 and L47 have the lowest score energy values compared to all other ligands, and they established the weak [62] hydrogen bond with the active site residues of PR. Ligand L22 formed hydrogen bond (H-donor) between N9 atom of this ligand and SD of Met801(A) residues at distance equal to 3.73 Å. In addition, the ligand L47 makes two weak [62] hydrogen bonds (H-acceptor) with Met756(A) and Val760(A) at distances equal to 3.65 and 3.29 Å, respectively. The ligands L23 and L26 are also involved in making a weak [62] hydrogen bond (H-donor), which is formed between the N9 atom of these ligands and SD of Met801(A) residues at distance equal to 3.75 and 4.17 Å respectively. On the other hand, ligand L23 is making pi-H interaction between the 5-ring of this ligand and CE2 of Phe778 (A). (See Figures III.7).
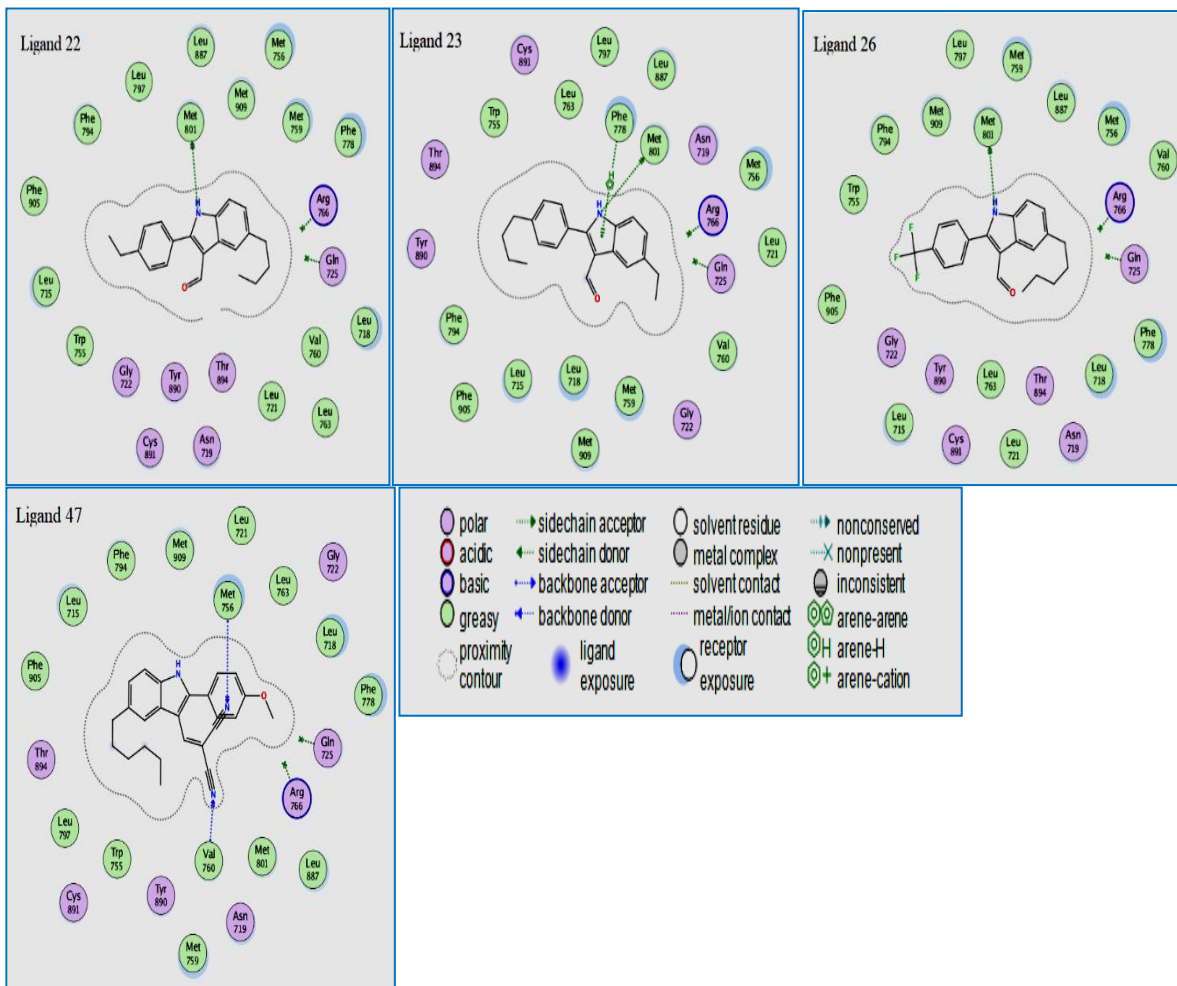
Figure III.7. 2D visualization of the interactions between the best ligands L22, L23, L26 and L47 with PR.

3.3. MD simulation

Molecular dynamics simulation is a theoretical approach based on solving Newton's equation of motion for an atomic system. This type of calculation is very beneficial for better understanding the behavior and the stability of the complexes Receptor-ligand. The best binding poses of complexes were obtained from molecular docking based on a high negative docking score. Four ligands (L37, L39, L40 and L48) interacted with the pocket of ER−α and four ligands (L22, L23, L26, and L47) with the pocket of PR. Figure III.8 represented the evolution of potential energy as a function of time corresponding to the best complexes.

It can be seen from figure III.8-b, the potential energy of complex 1A52-L40, changes brutally from 3599.457 to 1853.135 Kcal/mol during the first 200 Ps after that, this parameter starts to stabilize with a value around 1771.805 Kcal/mol.

We have mentioned above that ligands L39 and L48 have given no interactions with the active site residue of ER-α. However, these complexes formed new interactions after MD simulations indicating that they are deeply interacting with the pocket of ER- α. (see table III.9, Figure III.9.(a) and 9.(c).

The potential energy related to the complex 1A52-L39 changes from 3584.170 to 1927.078 Kcal/mol during the first 200 Ps and this complex starts to reach stability from 200 to 1000 Ps (see Figure. III.8-a).
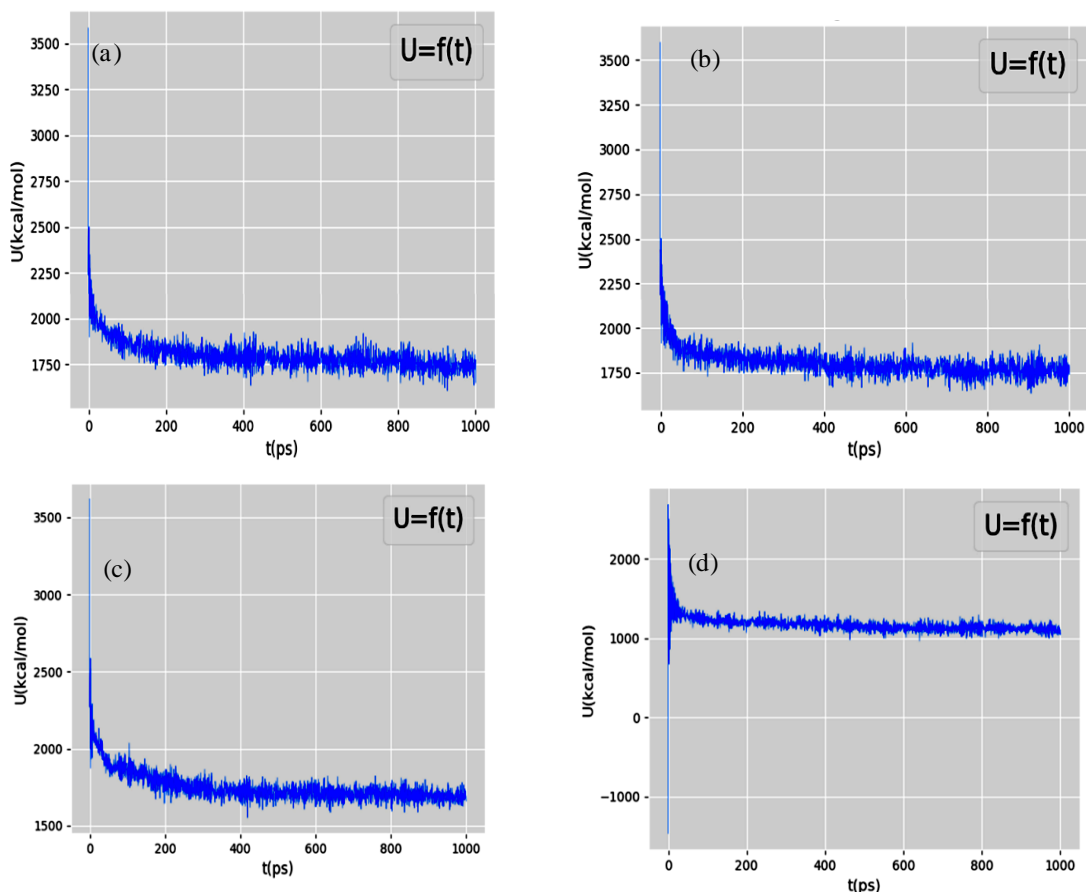
Figure III.8. The variation of the potential energy as function of time for complexes: (a) 1A52-L39, (b) 1A52-L40, (c) 1A52-L48, (d) 1A28-L47.

Concerning the potential energy of the complex 1A52-L48, which varies from 3619.667 to 1757.572 Kcal/mol during the first 200 Ps and then this energy, begins to stabilize around 1658.511 Kcal/mol (see Figure. III.8-c). Regarding the complex 1A28-L47, the potential energy changes from 1468.31 to 1293.96 Kcal/mol at the beginning of the first 35 Ps and it slightly changes between 1293.96 and 1246.44 Kcal/mol, in the second part the complex begins to achieve the equilibrium stability during the first 141 Ps (see Figure III.8-d). However, the complexes: 1A52-L37, 1A28-L22, 1A28-L23 and 1A28-L26 did not display any interactions after MD simulation, which means that these ligands are not stable and might be moving away from their original positions.

According to the MD simulation results, the ligands (L39, L40 and L48), with the first target ER–α and ligand L47 with the second target (PR), formed stable complexes with the active site residues by forming hydrogen bonds (H-acceptor and H-donor) (see Tables III.9 and III.10). The results obtained indicate that the ligands (L39, L40, L47 and L48) can be stabilized with high affinity through a hydrogen bond formed by either the nitrogen atom of five-membered ring or with the nitrile group of these ligands interacting with the key residues Lys529(A), Thr347(A), Leu387 (A) with ER–α and MET765 with PR (Figure III.9). We note here that our results are in good agreement with the experimental results [6], which reported the importance of nitrogen atom of the indole for the anti-proliferative activity. Furthermore, the importance of the nitrile group in small molecules has been also described in several papers based on protein–ligand interactions [64-67].
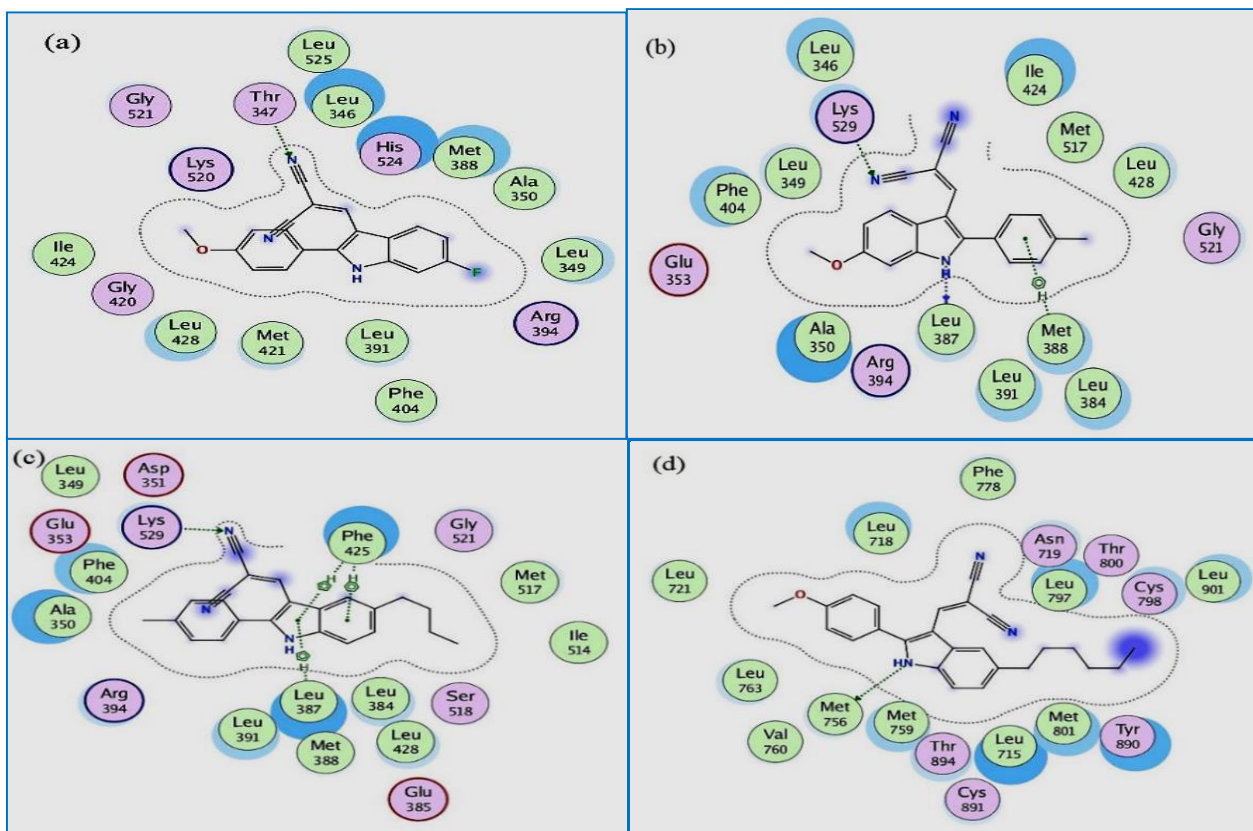
Figure III.9. 2D visualization of the best pose for the complexes: (a) 1A52-L39, (b) 1A52-L40, (c) 1A52-L48, (d) 1A28-L47 after the MD simulations.

Table III.9. Interactions between ligands and the active site residues of ER after MD simulation.

Ligand	Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic Distance (Å)	Bond energy (Kcal/mol)
			Bonds between ligands and of the active site residues			
L_{39}	N 28	OG1	THR347(A)	H-acceptor	2.96	-2.4
	N 9	O	LEU387(A)	H-donor	2.91	-2.9
L_{40}	N 27	NZ	LYS529(A)	H-acceptor	3.22	-8.7
	6-ring	CG	MET388(A)	pi-H	4.16	-0.7
	N 28	NZ	LYS529 (A)	H-acceptor	2.90	-11.9
L_{48}	5-ring	CB	LEU387 (A)	pi-H	3.99	-1.0
	5-ring	CZ	PHE425 (A)	pi-H	4.12	-0.9
	6-ring	CZ	PHE425 (A)	pi-H	4.08	-0.8

Table III.10. Interactions between ligands and the active site residues of PR after MD simulation.

Ligand	Bonds between ligands and of the active site residues					
	Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic Distance (Å)	Bond energy (Kcal/mol)
L$_{47}$	N 9	SD	MET 756 (A)	H-donor	4.33	-0.8

3.4. ADME/Pharmacokinetics predictions

Physicochemical properties and ADME prediction of the best ligands obtained by molecular docking/dynamics simulations are calculated by using SwissADME web tool (http://www.swissadme.ch). The obtained results are given in Table III.11 and Table III.12.

According to Table III.11, we can observe that all these selected ligands have a high probability of gastro-intestinal absorption (GI absorption), which have passively absorbed by the gastrointestinal tract, but they do have not a blood brain barrier feature (BBB) for the central nervous system (CNS) penetration except the molecules L40. On the other side, several previous papers [68, 69] reported that drugs such as tamoxifen can easily pass the blood-brain barrier, but this feature might affect the processes throughout the central nervous system (CNS). As well, we found that all these selected ligands could act as CYP1A2, CYP2C9, CYP2C19, and CYP3A4 inhibitors and non-inhibitors of CYP2D6 (see table III.11).

According to the parameters calculated (Table III.12), we can note that all ligands have several hydrogen bond donors <7 (n-HD: (0~7)) and hydrogen bond acceptors <12 (n-HA : (0~12)). The Molecular weight of these ligands belongs to the interval: 100~600 g/mol and the MLogP, and WLogP values are <5. In addition, we note that the number of nROTB for all ligands is <11. The results presented in Table III.12 indicated that all ligands have appeared no violations, and they complied with Lipinski, Veber, and Egan, except the ligand L47 rejected by Egan rules.

Table III.11. ADME prediction of the best ligands obtained after molecular docking and dynamic simulations.

	Pharmacokinetics							
	GI-ABS	BBB	P-gp substrate	CYP1A2 inhibitor	CYP2C19 Inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor
L39	High	No	No	Yes	Yes	Yes	No	Yes
L40	High	Yes	No	Yes	Yes	Yes	No	Yes
L47	High	No	Yes	Yes	Yes	Yes	No	Yes
L48	High	No	No	Yes	Yes	Yes	No	Yes

Table III.12. Physicochemical Property and Drug likeliness of the best ligands obtained after molecular docking and dynamic simulations.

Ligands	Physicochemical Property						Drug Likeliness		
	TPSA ($Å^2$) (0~140)	n-ROT (0~11)	MW (g/mol) (100~600)	MLog P WLogP (0~5)	n-HA (0~12)	n-HD (0~7)	Lipinski	Veber	Egan
L39	72.60	3	317.32	2.05 / 4.72	4	1	Accepted	Accepted	Accepted
L40	72.60	3	313.35	1.89 / 4.47	3	1	Accepted	Accepted	Accepted
L47	72.60	8	383.49	2.97 / 6.29	3	1	Accepted	Accepted	Not Accepted 1 violation: WLOGP>5.88
L48	63.37	5	339.43	3.13 / 5.81	2	1	Accepted	Accepted	Accepted

TPSA: Topological Polar Surface Area, **n-ROT:** Number of Rotatable, **MW:** Molecular Weight, **Log P:** Logarithm of partition coefficient of compound between n-octanol and water, **n-HA:** Number of hydrogen bond acceptors, **n-HD:** Number of hydrogen bonds donors

Figure III.10 explains the Egan BOILED-Egg method used to predict gastrointestinal absorption and brain penetration or accessibility of these inhibitors. From Figure III.10, it can be observed that the ligand L47 is PGP+ so it is a substrates of P-gp and could reduce the activity as we mentioned before, while the rest of the ligands are PGP- and they are not substrates of P-gp.
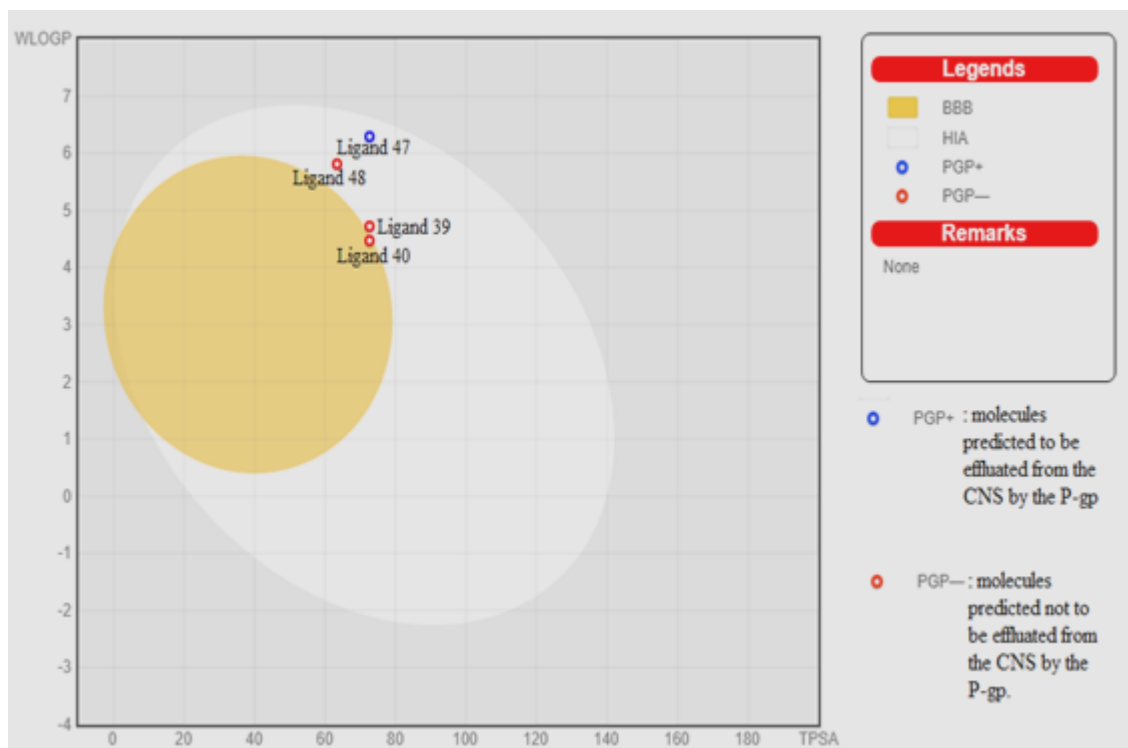
Figure III.10. The Egan BOILED-Egg plot. The point located in the yellow region (yolk) is the molecules predicted to be passively permeated through the blood-brain barrier (BBB), while others in the white area are the molecules predicted to be passively absorbed by the gastrointestinal tract (HIA).

In the light of the above discussions, we can observe that all these studied ligands tend to be a high pharmacologically active and they have good absorption, oral bioavailability, and permeability, except the ligand L47 which is predicted to exhibit an oral administration problem confirmed by one violation in Egan rules.

4. Conclusions

In this present study, the QSAR study confirmed that the two models developed are robust and could successfully predict the pIC_{50} values associated with anti-proliferative activities, which are validated by the two coefficients of bootstrapping correlation coefficient (R^2_{boot}) and leave-one-out cross-validation coefficient (Q^2_{CV}). In addition, this study shows that flexibility, lipophilicity, and the type of groups substituted on the C2 carbon of the molecules studied have a direct effect on antiproliferative activities.

Molecular docking/dynamics simulations results show that the ligands (L39, L40 and L48 bind to ER and ligand L47 binds with PR have a strong affinity with the active site residues this is confirmed by the formation of several non-covalent bonds namely: hydrogen bonds and hydrophobic interactions with the binding pocket of both receptors.

The Pharmacokinetics and drug-likeness studies revealed that the ligands L39 and L48 are orally bioavailable. Finally, the ADME prediction allowed us to suggest that these ligands could be the best drug candidates against breast cancer.

REFERENCES

1.	Sachdeva H, Mathur J, Guleria A. Indole derivatives as potential anticancer agents: A review. Journal of the Chilean Chemical Society. 2020;65(3):4900-7.

2.	Dorababu A. Indole–a promising pharmacophore in recent antiviral drug discovery. RSC Medicinal Chemistry. 2020;11(12):1335-53.

3.	Dadashpour S, Emami S. Indole in the target-based design of anticancer agents: A versatile scaffold with diverse mechanisms. European journal of medicinal chemistry. 2018;150:9-29.

4.	Wan Y, Li Y, Yan C, Yan M, Tang Z. Indole: A privileged scaffold for the design of anti-cancer agents. European journal of medicinal chemistry. 2019;183:111691.

5.	Banik BK, Banerjee B. Heterocyclic Anticancer Agents: Walter de Gruyter GmbH & Co KG; 2022.

6.	Kaufmann D, Pojarova M, Vogel S, Liebl R, Gastpar R, Gross D, et al. Antimitotic activities of 2-phenylindole-3-carbaldehydes in human breast cancer cells. Bioorg Med Chem. 2007;15(15):5122-36.

7.	Pojarová M, Kaufmann D, Gastpar R, Nishino T, Reszka P, Bednarski PJ, et al. [(2-Phenylindol-3-yl) methylene] propanedinitriles inhibit the growth of breast cancer cells by cell cycle arrest in G2/M phase and apoptosis. Bioorg Med Chem. 2007;15(23):7368-79.

8.	Gaikwad R, Bobde Y, Ganesh R, Patel T, Rathore A, Ghosh B, et al. 2-Phenylindole derivatives as anticancer agents: synthesis and screening against murine melanoma, human lung and breast cancer cell lines. Synthetic Communications. 2019;49(17):2258-69.

9.	Kelly PM, Bright SA, Fayne D, Pollock JK, Zisterer DM, Williams DC, et al. Synthesis, antiproliferative and pro-apoptotic activity of 2-phenylindoles. Bioorg Med Chem. 2016;24(18):4075-99.

10.	Gaikwad R, Ghorai S, Amin SA, Adhikari N, Patel T, Das K, et al. Monte Carlo based modelling approach for designing and predicting cytotoxicity of 2-phenylindole derivatives against breast cancer cell line MCF7. Toxicology in Vitro. 2018;52:23-32.

11.	Liao SY, Qian L, Miao TF, Lu HL, Zheng KC. CoMFA and docking studies of 2-phenylindole derivatives with anticancer activity. Eur J Med Chem. 2009;44(7):2822-7.

12.	Adhikari N, Halder AK, Saha A, Das Saha K, Jha T. Structural findings of phenylindoles as cytotoxic antimitotic agents in human breast cancer cell lines through

multiple validated QSAR studies. Toxicology in vitro : an international journal published in association with BIBRA. 2015;29(7):1392-404.

13.   Halder AK, Adhikari N, Jha T. Comparative QSAR modelling of 2-phenylindole-3-carbaldehyde derivatives as potential antimitotic agents. Bioorg Med Chem Lett. 2009;19(6):1737-9.

14.   El-Nakkady SS, Hanna MM, Roaiah HM, Ghannam IA. Synthesis, molecular docking study and antitumor activity of novel 2-phenylindole derivatives. Eur J Med Chem. 2012;47(1):387-98.

15.   HyperChem. molecular modelling system, Hypercube Inc. v8 ed. 1115 NW 4th Street, Gainesville, FL 32601, USA, 2009

Hypercube Inc.

16.   Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. Journal of the American Chemical Society. 1985;107(13):3902-9.

17.   Barone V, Cossi M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. The Journal of Physical Chemistry A. 1998;102(11):1995-2001.

18.   Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Revision A.02, ed: Gaussian, Inc., Wallingford CT, 2016.

19.   Breneman CM, Wiberg KB. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. Journal of Computational Chemistry. 1990;11(3):361-73.

20.   MATLAB 8.0 and Statistics Toolbox 8.1. Natick, Massachusetts, United States.: The MathWorks, Inc.

21.   Winter R, Montanari F, Noé F, Clevert D-A. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. Chemical science. 2019;10(6):1692-701.

22.   Roy K, Kar S, Das RN. A Primer on QSAR-QSPR Modeling Fundamental Concepts.pdf>2015.

23.   Parthasarathi R, Dhawan A. Chapter 5 - In Silico Approaches for Predictive Toxicology. In: Dhawan A, Kwon S, editors. In Vitro Toxicology: Academic Press; 2018. p. 91-109.

24.   Kenouch S, Harkati D, Ghamri M, Chikhaoui AR, Melkemi N. Predictive QSAR model and clustering analysis of some Benzothiazole derivatives as cytotoxic inhibitors. SDRP Journal of Computational Chemistry & Molecular Modelling. 2018;2(3):1-8.

25.   James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R: Springer; 2013.

26.   Rawlings JO, Pantula SG, Dickey DA. Applied regression analysis: a research tool. Second Edi ed. New York: Springer New York, NY; 1998.

27.   Roy K. On some aspects of validation of predictive quantitative structure-activity relationship models. Expert opinion on drug discovery. 2007;2(12):1567-77.

28.   Srivastava V, Selvaraj C, Singh SK. Chemoinformatics and QSAR. Advances in Bioinformatics. 2021:183-212.

29.   Kar S, Roy K. Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs. 2011:111-22.

30.   Environment MO. Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.; 2019.

31.   Tanenbaum DM, Wang Y, Williams SP, Sigler PB. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. Proceedings of the National Academy of Sciences. 1998;95(11):5998-6003.

32.   Williams SP, Sigler PB. Atomic structure of progesterone complexed with its receptor. Nature. 1998;393(6683):392-6.

33.   Ball P. Water is an active matrix of life for cell and molecular biology. Proceedings of the National Academy of Sciences. 2017;114(51):13327-35.

34.   Pantsar T, Poso A. Binding Affinity via Docking: Fact and Fiction. Molecules. 2018;23(8):1899.

35.   Stanzione F, Giangreco I, Cole JC. Use of molecular docking computational tools in drug discovery. Progress in Medicinal Chemistry. 2021;60:273-343.

36.   Kim KH. Outliers in SAR and QSAR: 3. Importance of considering the role of water molecules in protein–ligand interactions and quantitative structure–activity relationship studies. Journal of Computer-Aided Molecular Design. 2021;35(3):371-96.

37.   Daoud I, Melkemi N, Salah T, Ghalem S. Combined QSAR, molecular docking and molecular dynamics study on new Acetylcholinesterase and Butyrylcholinesterase inhibitors. Computational biology and Chemistry. 2018;74:304-26.

38.   Chenafa H, Mesli F, Daoud I, Achiri R, Ghalem S, Neghra A. In silico design of enzyme α-amylase and α-glucosidase inhibitors using molecular docking, molecular dynamic, conceptual DFT investigation and pharmacophore modelling. J Biomol Struct Dyn. 2021.

39.   Daoud I, Mesli F, Melkemi N, Ghalem S, Salah T. Discovery of potential SARS-CoV 3CL protease inhibitors from approved antiviral drugs using: virtual screening, molecular docking, pharmacophore mapping evaluation and dynamics simulation. Journal of Biomolecular Structure and Dynamics. 2022;40(23):12574-91.

40.   Cole JC, Murray CW, Nissink JWM, Taylor RD, Taylor R. Comparing protein–ligand docking programs is difficult. Proteins: Structure, Function, and Bioinformatics. 2005;60(3):325-32.

41.   Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Medicinal research reviews. 2006;26(5):531-68.

42.   Bond SD, Leimkuhler BJ, Laird BB. The Nosé–Poincaré method for constant temperature molecular dynamics. Journal of Computational Physics. 1999;151(1):114-34.

43.   Sturgeon JB, Laird BB. Symplectic algorithm for constant-pressure molecular dynamics using a Nosé–Poincaré thermostat. The Journal of Chemical Physics. 2000;112(8):3474-82.

44.   Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. 2007;9(3):90-5.

45.   Feinberg EN, Joshi E, Pande VS, Cheng AC. Improvement in ADMET prediction with multitask deep featurization. Journal of medicinal chemistry. 2020;63(16):8835-48.

46.   Chandrasekaran B, Abed SN, Al-Attraqchi O, Kuche K, Tekade RK. Computer-aided prediction of pharmacokinetic (ADMET) properties. Dosage form design parameters: Elsevier; 2018. p. 731-55.

47.   Lipinski C. a, Lombardo, F., Dominy, BW & Feeney, PJ Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001;46(1-3):3.

48.   Veber D, Johnson S, Cheng H, Smith B, Ward K, Kopple K. om c.(2002). Molecular properties that influence the oral bioavailability of drug candidates Journal of Medicinal Chemistry.45(12):2615-23.

49.   Egan WJ, Merz KM, Baldwin JJ. Prediction of drug absorption using multivariate statistics. Journal of medicinal chemistry. 2000;43(21):3867-77.

50.   Amin ML. P-glycoprotein inhibition for optimal drug delivery. Drug target insights. 2013;7:DTI. S12519.

51.   Lynch T, Price AL. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. American family physician. 2007;76(3):391-6.

52.   Tsaioun K, Kates SA. ADMET for medicinal chemists: a practical guide: John Wiley & Sons; 2011.

53.   Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific reports. 2017;7(1):1-13.

54.   Daina A, Zoete V. A boiled-egg to predict gastrointestinal absorption and brain penetration of small molecules. ChemMedChem. 2016;11(11):1117.

55.   Schwarz G. Estimating the dimension of a model. The annals of statistics. 1978:461-4.

56. Lee W, Park S-J, Hwang J-Y, Hur K-H, Lee YS, Kim J, et al. QSAR Model for Predicting the Cannabinoid Receptor 1 Binding Affinity and Dependence Potential of Synthetic Cannabinoids. Molecules. 2020;25(24):6057.

57. Lee H, Lee S, Baek JY, Seo C-S, Yun H, Kim S-N, et al. Estrogenic activity of ethyl gallate and its potential use in hormone replacement therapy. Bioorganic & medicinal chemistry letters. 2021;40:127919.

58. Swamy V, Deshpande J, Pathageri J, Kothekar V, Padhye S. Molecular Docking Studies on Estrogen Receptor-A and Chalcone Derivatives. International Journal of Advances in Pharmacy and Biological Sciences. 2011;1(3):87-93.

59. Gao LM, Maldonado W, Narváez-Pita X, Carmona-Negrón JA, Olivero-Verbel J, Meléndez E. Steroid-Functionalized Titanocenes: Docking Studies with Estrogen Receptor Alpha. Inorganics. 2016;4(4):38.

60. Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engström O, et al. Molecular basis of agonism and antagonism in the oestrogen receptor. Nature. 1997;389(6652):753-8.

61. Essen LO. Protein Crystallography in Drug Discovery.(Series: Methods and Principles in Medicinal Chemistry, Vol. 20.). Edited by Robert E. Babine and Sherin S. Abdel-Meguid. Angewandte Chemie (International Ed in English). 2004;43(47):6408.

62. Imberty A, Hardman KD, Carver JP, Perez S. Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A. Glycobiology. 1991;1(6):631-42.

63. Ganesan A, Coote ML, Barakat K. Molecular dynamics-driven drug discovery: leaping forward with confidence. Drug discovery today. 2017;22(2):249-69.

64. Wang Y, Du Y, Huang N. A survey of the role of nitrile groups in protein–ligand interactions. Future Med Chem. 2018;10(23):2713-28.

65. Sun J, Baudry J, Katzenellenbogen JA, Katzenellenbogen BS. Molecular basis for the subtype discrimination of the estrogen receptor-β-selective ligand, diarylpropionitrile. Molecular Endocrinology. 2003;17(2):247-58.

66. Fleming FF, Yao L, Ravikumar P, Funk L, Shook BC. Nitrile-containing pharmaceuticals: efficacious roles of the nitrile pharmacophore. Journal of medicinal chemistry. 2010;53(22):7902-17.

67.   Wang X, Wang Y, Li X, Yu Z, Song C, Du Y. Nitrile-containing pharmaceuticals: target, mechanism of action, and their SAR studies. RSC Medicinal Chemistry. 2021;12(10):1650-71.

68.   Rotheneichner P, Romanelli P, Bieler L, Pagitsch S, Zaunmair P, Kreutzer C, et al. Tamoxifen Activation of Cre-Recombinase Has No Persisting Effects on Adult Neurogenesis or Learning and Anxiety. Frontiers in neuroscience. 2017;11(27).

69.   Paganini-Hill A, Clark LJ. Preliminary assessment of cognitive function in breast cancer patients treated with tamoxifen. Breast cancer research and treatment. 2000;64(2):165-76.

Chapter IV

Molecular Docking/Dynamics Simulations, MEP analysis, and Pharmacokinetics prediction of some Withangulatin A derivatives as Allosteric Glutaminase C Inhibitors in Breast Cancer
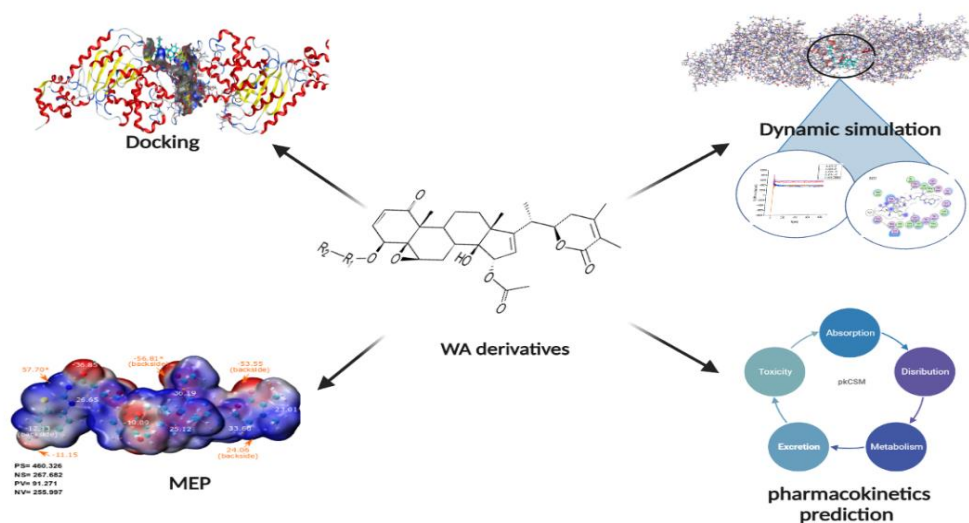
1. Introduction

In recent years, there has been a notable interest in natural products and their derivatives as potential sources for developing anticancer drugs. Physalis angulata L., a traditional Chinese medicinal plant, has a long history of use in treating various ailments including tumors, hepatitis, rheumatism, and nephronia. From this plant species, numerous withanolides with notable bioactivities have been isolated. Among them, Withangulatin A (WA) stands out for its potent antitumor, anti-inflammatory, and immunosuppressive properties. Despite substantial research efforts directed towards understanding WA's anticancer mechanisms, such as its impact on cellular morphology, inhibition of topoisomerase II activities, and modulation of general protein synthesis, the specific molecular targets, and detailed mechanisms of action of WA remain elusive. This lack of clarity has impeded further advancements in utilizing WA for cancer therapy [1,2]. As we mentioned in chapter I, triple-negative breast cancer (TNBC) is an aggressive subtype that lacks the expression of hormone receptors: estrogen (ER), progesterone (PR), and human epidermal growth factor receptor-2 (HER2) [3,4]. It's worth mentioning that many cancer cells have an increased need for glutamine, referred to as "glutamine addiction" [5], which helps them grow by primarily assisting in the formation of tricarboxylic acid (TCA) cycle intermediates [6,7]. Once glutamine enters the mitochondria via the glutamine transporter, glutaminase (GLS) enzyme converts it to stoichiometric amounts of ammonia and glutamate [8]. The Glutaminase (GLS) is an enzyme identified as one of the primary metabolic pathways supporting the proliferation of several cancer cells and is considered a potential therapeutic target for cancer treatment [9, 10]. It has two distinct isoforms: GLS-1, also known as kidney-type, and GLS-2, also known as liver-type [11]. GLS1 kidney-type can occur in two variant forms, kidney glutaminase (KGA),and glutaminase C (GAC) [11]. It has been reported that the GAC variant is linked with high-grade and metastatic breast cancer [12, 13].Currently, the most frequently used selective GLS1 inhibitors are Telaglenastat (CB-839) and BPTES (bis-2-(5-phenylacetamido-1,2,4-thiadiazol-2-yl)ethyl sulfide), which are noncompetitive inhibitors that bind through the allosteric site located between the two dimers of GLS1, leading to the formation of an inactive tetramer and potentially blocking glutamine hydrolysis [12, 14]. However, only a few GLS1 inhibitors have been identified [15]. As a result, researchers considered that inhibiting GLS1 is

a promising therapeutic and may provide new targets for cancer treatment [11]. Discovering natural products and their derivatives has become increasingly important in drug development [16]. They play a crucial role as oral drugs that exceed Lipinski rule of five. This is evident from the rise in the molecular mass of oral drugs, despite this, using natural products in traditional medicine provides valuable information about their effectiveness and safety. The pool of natural products containing bioactive compounds covers a wider area of chemical space compared to small-molecule libraries [17]. Withanolides are a group of compounds that belong to the natural steroid lactones [18]. Interestingly, they have attracted attention because of numerous reports of their cytotoxic properties [19, 20]. Recently, withangulatin A (WA) derivatives, have been reported to have potent antiproliferative activity against TNBC [21]. Additionally, some WA derivatives were investigated as potential GLS1 inhibitors and found to inhibit cell growth in TNBC by binding to an allosteric site on GAC [22].

In this study, Molecular docking was performed on twenty-six WA derivatives with CB-839 as a reference compound toward the allosteric site of GAC. The top four scored compounds, together with CB-839, were subjected to molecular dynamic simulations to validate their stability at the binding site. The molecular electrostatic potential (MEP) analysis was then applied to the chosen compounds to investigate intermolecular interactions. Furthermore, pharmacokinetics and toxicology were used to predict the properties of the studied compounds.

2. Materials and methods

2.1. Biological data

In this study, a total of twenty-three cytotoxic withangulatin A (WA) derivatives with reported IC_{50} values against breast cancer cell line MDA-MB-231, were collected from the literature [21]. Table IV.1 shows the chemical structure of withangulatin A (WA) derivatives as well as the reported in vitro results.
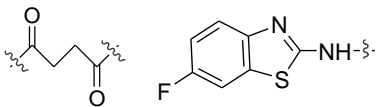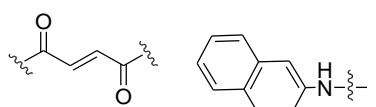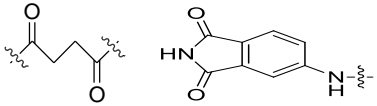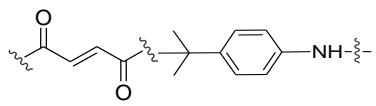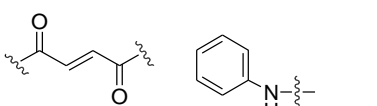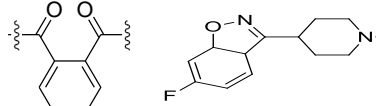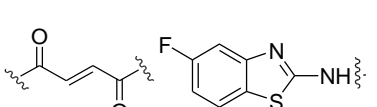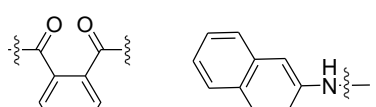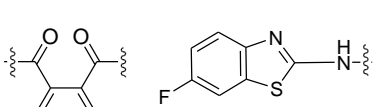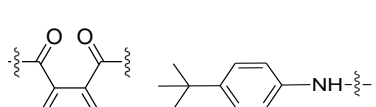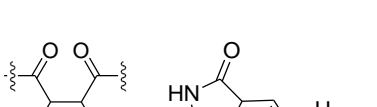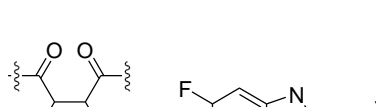
Table IV.1. Chemical structure and experimental antiproliferative activities of WA derivatives under study.

Compound	R₁	R₂	$IC_{50}(\mu M)$ [20] (MDA-MB231)	Comp	R₁	R₂	$IC_{50}(\mu M)$ [21] (MDA-MB231)
A1			1.05	A10			0.07
A2			19.88	A11			8.99
A3			0.35	A12			>50
A4			0.49	A13			1.35
A5			0.61	A14			0.49
A6			0.23	A15			4.87
A7			>50	A16			0.22

117

Table IV. Continued

A8		0.94	**A17**		1.01	
A9		12.76	**A18**		0.70	
A19		0.63	**A23**		0.39	
A20		1.88	**A24**		3.29	
A21		2.28	**A25**		28.67	
A22		>50	**A26**		2.22	

2.2. Molecular docking

2.2.1. Ligands preparation

The two-dimensional (2D) structures of cytotoxic withangulatin A (WA) derivatives were drawn in ChemDraw Ultra 12.0 and then converted to three-dimensional (3D) structures in MDL (.mol) format to determine a set of coordinates indicating the lowest energy conformation for the given structure. Hyperchem version 8.0.3 software [23] was used for geometry optimization using the Molecular Mechanics force field (MM+) with the Polak-Ribiere conjugate gradient algorithm, with the root mean square (RMS) gradient of 0.1 kcal/mol.

The optimized molecular structures of WA derivatives were generated as a 3D protein structure database (PDB files) and used as input files for docking analysis carried out with MOE software [24].

2.2.2. Target selection and preparation

The X-ray crystal structure of human glutaminase C (GAC) complexed with bis-2-(5-phenylacetimido-1,2,4-thiadiazol-2-yl) ethyl sulfide (BPTES), (PDB ID: 3UO9) [25], which has a good resolution (2.30 Å) [46], was downloaded from the RCSB protein data bank (http://www.rcsb.org/pdb). The GAC structure is a tetramer with four glutamate-binding pockets and one allosteric site per GAC dimer that binds BPTES [25] (see Figure IV.1).
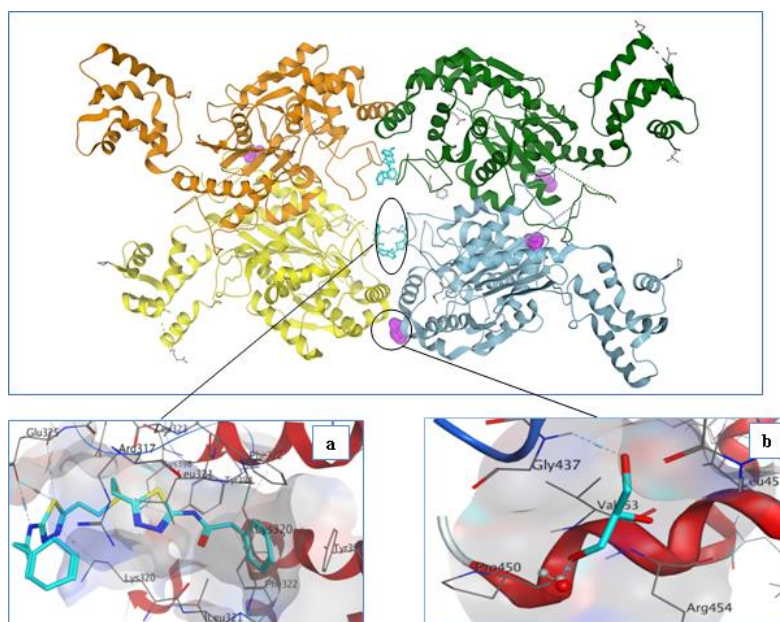
Figure IV.1. Tetramer crystal structure of human GAC bound to BPTES. Glutamate binding pockets are shown as pink spheres and BPTES molecules are shown as the cyan stick. (a) The allosteric binding pocket of BPTES. (b) Glutamate binding pocket.

The X-ray structure was prepared by removing the duplication chains (C and D) and keeping the two dimer chains A and B. All water molecules in the binding pocket have been kept because they have a key role in target-ligand interactions [26-28]. Next, the protein structure was corrected by adding the missing hydrogen atoms using MOE software [24]. The site finder tool in MOE software was used to identify the many active site cavities. Afterward,

the Web server PASSer: Protein Allosteric Sites [29] was further applied for the prediction of the possible amino acids that build up the allosteric site of the selected target. The top-ranked pocket was selected as a potential binding site.

Herein, a semi-rigid docking method was conducted in which the ligand is flexible, but the protein structure is rigid. The Triangle Matcher combined with the London dG scoring function was chosen as the initial scoring methodology and the GBVI/WSA dG scoring function was chosen for the final refinement. Docking processes were set following the default methodology described by Daoud et al. [30, 31]. Knowing that the native ligand (BPTES) is located between the subunits A and B. The validation and justify the accuracy of the docking method was carried out by re-docking of the native ligand BPTES into the same allosteric site and calculate the root-mean-square deviation (RMSD) value (see Table IV.2).

Table IV.2. Information of the selected target (3UO9).

Receptor	Docked Chains	Docking score (kcal/mol)	Native ligand	RMSD(Å)
3UO9	A and B	-10.90	BPTES	1.82
	C and D	-9.616	BPTES	2.80

As shown in Table IV.2, the complex GAC-BPTES has an RMSD value of 1.829 Å, which is less than 2 Å and belongs to the desired range [32]. After enzyme validation, all the WA derivatives were docked into the GAC binding site pocket and several parameters were obtained, including, the interaction type, bond-length, and bond energy. The binding interaction and energy score were compared with the native ligand (BPTES). Meanwhile, a clinical compound, CB-839 (Telaglenastat), was also docked with the aim to predict its affinity for the GAC and compare it with the WA derivatives. Likewise, molecular docking was also performed on the GAC domain binding sites located between the subunits (C and D).

2.3. Molecular dynamics simulation (MD)

Molecular dynamics is based on Newton's equations of motion to simulate the behaviour of a molecular system over time [33]. Previous studies have used MD simulations to evaluate the stability of the complex (protein-ligand) using potential energy as a function of time [30,

31, 34]. Therefore, the best score of four complexes and the reference drug CB-839 obtained from molecular docking were considered for MD simulations to confirm their stability based on potential energy variation as a function of time using MOE software [24].

The energy was minimized using the MMFF94X force field with a gradient of 0.1 RMS kcal/mol1. Å-1, and the Nosé-Poincaré-Andersen (NPA) algorithm [35] was applied. The MD simulations were carried out in the following steps: 1) production MD runs were set for 700 ps, 2) the system was equilibrated for 100 ps at 300 K with a time step of 0.002 ps. Finally, to evaluate the stability of complexes, the plot of the potential energy variations as a function of time (ps) was done with OriginePro [36].

2.4. Molecular electrostatic potential

Molecular electrostatic potential (MEP) was conducted to predict the reactive sites for electrophilic and nucleophilic attacks on the selected structures. It is frequently made by mapping the electrostatic potential on the molecule's electron density surface, which allows us to demonstrate the distribution of the electronic charge throughout the structure. Currently, this approach is being used to better understand the molecular environment, hydrogen bond interactions, and biological recognition processes.

The top four structures were fully optimized in Gaussian 09 Software [37], using the density functional theory (DFT) with the hybrid functional UB3LYP combining with the 6-31 G(d,p) basis set in an aqueous solution with the CPCM solvation model [38]. Multiwfn 3.7 [39] in conjunction with VMD 1.9.1 software [40] was used for the quantitative analysis of electrostatic potential on the van der Waals surface.

2.5. Pharmacokinetics (PK) and toxicity prediction

In silico methods are commonly used as a preliminary step in analyzing new chemical substances to predict lead candidates and various pharmacokinetic properties [41]. The web server pkcsm [42] was used to estimate the most important pharmacological and toxicological parameters, including Caco-2 permeability (colon adenocarcinoma), human intestinal absorption (HIA), skin permeability, P-glycoprotein (P-gp) substrate, P-gp I inhibitor, P-gp II inhibitor, steady-state volume of distribution (VDss) and blood-brain barrier permeability. In addition, Cytochrome P450 inhibitors were evaluated for CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4. It was also used to predict toxicological properties such as AMES

toxicity, maximum tolerated dose, human ether a-go-go gene (hERG) I and II inhibitors, Oral Rat Acute Toxicity (LD50), hepatotoxicity, and skin sensitization.

3. Results and discussion

3.1. Molecular docking study

3.1.1. Binding site residues of the target

Table IV.3. The predicted top 3 binding pockets residues of the allosteric site.

Site Number	Probability	Residues
Site 1	74.16%	Chain A: LEU321, PHE322, TYR394, LYS320 and ASP327. Chain B: LEU321, PHE322, LYS320, TYR394, LEU323, LYS320 and ASP327
Site 2	29.06%	Chain B: LEU321, PHE322, SER462, MET333, PRO313, LYS481, ASN335, ARG387, HIS330, ALA473, ILE391, TYR466, ASP467, SER469, CYS463, SER314, GLY470, VAL334, LYS320, ALA336, HIS461
Site 3	27.31%	Chain A :LEU524, GLU545, LYS539, CYS525, ASP541, PRO542, PHE227, PHE536, GLY546, ARG544, HIS230, ASP223, SER226

We identified the allosteric site residues of the studied target using the PASSer online server.

According to the results of Table IV.3 above, we can notice that site 1 is the most favourable (74.16%) among the three sites obtained. Nevertheless, further docking analysis was carried out on the other two sites, as illustrated in Appendix D, and their interactions were analyzed and presented in Appendix E and F

3.1.2. Receptor-compounds interactions

All molecular docking results of site 1 between A and B chains are listed in Table IV.4. Meanwhile, the docked poses for site 1 between C and D chains are listed in Appendix (see Appendix C.).

Table IV.4. Docking score, distances bonds and bond energy of the withangulatin derivatives with GAC receptor (Site 1 between A and B chains).

Compound	S-score (kcal/mol)	Bonds between ligands and of the active site residues					
		Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic Distance (Å)	Bond energy (Kcal/mol)
A1	-6.40	S 93	OE2	GLU325(B)	H-donor	4.16	-0.4
A2	-7.21	O 6	NZ	LYS398(B)	H-acceptor	2.97	-1.1
		5-ring	CA	GLU325(B)	Pi-H	4.04	-0.8
A3	-7.44	O 18	CE	LYS320(A)	H-acceptor	2.97	-0.8
		O 89	OH	TYR394(B)	H-acceptor	2.86	-2.1
		N 105	N	PHE322(B)	H-acceptor	3.11	-1.0
A4	-7.31	C 15	OE2	GLU325(B)	H-donor	3.43	-0.7
		C 77	OD1	ASP327(B)	H-donor	3.49	-0.7
		O 90	N	LEU323(B)	H-acceptor	2.96	-3.5
A5	-8.49	C 13	O	GLU325(B)	H-donor	3.38	-0.8
		C 77	OD1	ASP327(B)	H-donor	3.52	-0.7
A6	-7.45	C 13	O	GLU325(B)	H-donor	3.39	-0.7
		C 15	OE2	GLU325(B)	H-donor	3.24	-0.8
		O 90	N	LEU323(B)	H-acceptor	3.18	-1.9
		6-ring	CA	LEU321(B)	Pi-H	4.66	-0.7
A7	-7.39	O 88	N	LEU323(B)	H-acceptor	3.11	-2.1
		5-ring	N	PHE322(B)	Pi-H	4.12	-1.4
		6-ring	6-ring	PHE322(B)	Pi-Pi	3.93	-0.0
A8	-9.41	S 84	O	LYS320(A)	H-donor	2.90	-2.7
		O 6	NZ	LYS320(A)	H-acceptor	3.26	-0.8
		O 18	NZ	LYS398(B)	H-acceptor	3.17	-0.8
		O 77	N	PHE322(B)	H-acceptor	3.00	-1.8
		N 81	N	PHE322(B)	H-acceptor	3.66	-0.7
A9	-6.79	O 47	OE2	GLU325(B)	H-donor	2.70	-2.2
		N 78	O	LEU316(B)	H-donor	3.10	-2.6

Table IV.4. Continued

A10	-6.23	/	/	/	/	/	/
A11	-7.08	O 65	NZ	LYS399(B)	H-acceptor	3.07	-3.6
		6-ring	CD	ARG317(A)	Pi-H	4.02	-1.1
		6-ring	NH1	ARG317(A)	Pi-cation	3.59	-0.7
		6-ring	NH2	ARG317(B)	Pi-cation	3.59	-0.6
A12	-7.64	6-ring	CA	LEU321(B)	Pi-H	4.42	-0.7
		6-ring	N	PHE322(B)	Pi-H	3.91	-1.5
A13	-8.58	O 38	NZ	LYS398(B)	H-acceptor	3.14	-3.9
		O 77	N	LEU323(B)	H-acceptor	2.95	-3.0
		N 89	N	PHE322(B)	H-acceptor	3.10	-3.6
A14	-6.59	/	/	/	/	/	/
A15	-7.53	O 18	CE	LYS320(A)	H-acceptor	3.09	-1.3
		O 83	OH	TYR394(B)	H-acceptor	3.20	-0.7
		O 99	N	LEU323(A)	H-acceptor	3.10	-3.4
		N 78	6-ring	TYR394(B)	H-Pi	4.61	-0.7
A16	-6.09	/	/	/	/	/	/
A17	-5.47	C 15	OE2	GLU325(B)	H-donor	3.23	-1.4
		6-ring	CA	LEU321(B)	Pi-H	4.87	-0.6
A18	-8.07	O 36	NH1	ARG317(B)	H-acceptor	2.85	-1.4
		O 36	NH2	ARG317(B)	H-acceptor	2.82	-1.2
		O 42	NZ	LYS398(A)	H-acceptor	3.08	-1.0
		O 46	NZ	LYS398(A)	H-acceptor	2.52	-5.5
A19	-6.29	O 47	OE2	GLU325(B)	H-donor	3.38	-1.3
		O18	CE	LYS320(A)	H-acceptor	3.01	-1.2
A20	-5.64	5-ring	CG	LEU321(B)	Pi-H	4.15	-0.8
A21	-4.58	C 86	OE2	GLU325(B)	H-donor	3.50	-0.7
		N 90	O	LEU323(B)	H-donor	3.42	-1.0
		S 96	O	LEU323(B)	H-donor	3.54	-2.4
A22	-6.20	/	/	/	/	/	/
A23	-5.55	O 6	OH	TYR394(B)	H-acceptor	3.25	-1.0

Table IV.4. Continued

A24	-6.02	/	/	/	/	/	/
A25	-5.99	O 3	NH1	ARG317(B)	H-acceptor	3.05	-0.9
		O 3	NH2	ARG317(B)	H-acceptor	2.94	-4.1
A26	-5.27	O 67	NZ	LYS399(B)	H-acceptor	2.92	-7.3
Ref CB-839	-8.31	O 7	N	LEU323(A)	H-acceptor	3.11	-0.9
		NAX 49	O	LEU323(A)	H-donor	3.08	-1.5
		OAB 2	NZ	LYS398(A)	H-acceptor	2.95	-4.9
		NAT 44	N	PHE322(A)	H-acceptor	3.23	-3.6
Co-crystallized BPTES	-10.90	NAV 46	N	LEU323(A)	H-acceptor	3.06	-3.1
		SAY 51	CA	LEU321(A)	H-acceptor	3.76	-1.0
		SAY 51	N	PHE322(A)	H-acceptor	3.23	-2.1
		CAM 25	6-ring	PHE322(B)	H-Pi	4.17	-0.7

a) 3UO9- native ligands interactions

The co-crystallized ligand (BPTES) is predicted to have a high binding affinity with GAC, giving a total docking score of -10.90 Kcal/mol. The binding interactions of BPTES within the allosteric site formed six interactions with the active site residues of GAC, five hydrogen bonds acceptor with the: LYS398(A), PHE322(A), LEU323(A), LEU321(A), and PHE 322(A) at distances of: 2.95 Å, 3.23 Å, 3.06 Å, 3.76 Å, and 3.23 Å, and bonds energy of: -4.9, -3.6, -3.1, -1.0, and -2.1 7 kcal/mol, respectively. One hydrogen bond donor with LEU 323(A) at 3.08 Å and bond energy of -1.5 kcal/mol; and one hydrophobic interaction (H-pi) with PHE322(B) at a distance of 4.17 Å and bond energy -0.7 kcal/mol. The reference compound CB-839 is predicted to have a significant binding affinity toward GAC with a total docking score of -8.313 Kcal/mol. The binding interaction of CB-839 showed a hydrogen

bond of length (3.11 Å) and bond energy of -0.9 kcal/mol with the residue LEU323(A) (see Table IV.4 and Figure IV.2). These findings were in combination with the literature that defines the allosteric pocket in GAC. The authors identified the amino acids LEU323, PHE322, LEU321, and TYR394 as the key functions of the allosteric site for inhibitory activity [25, 43].
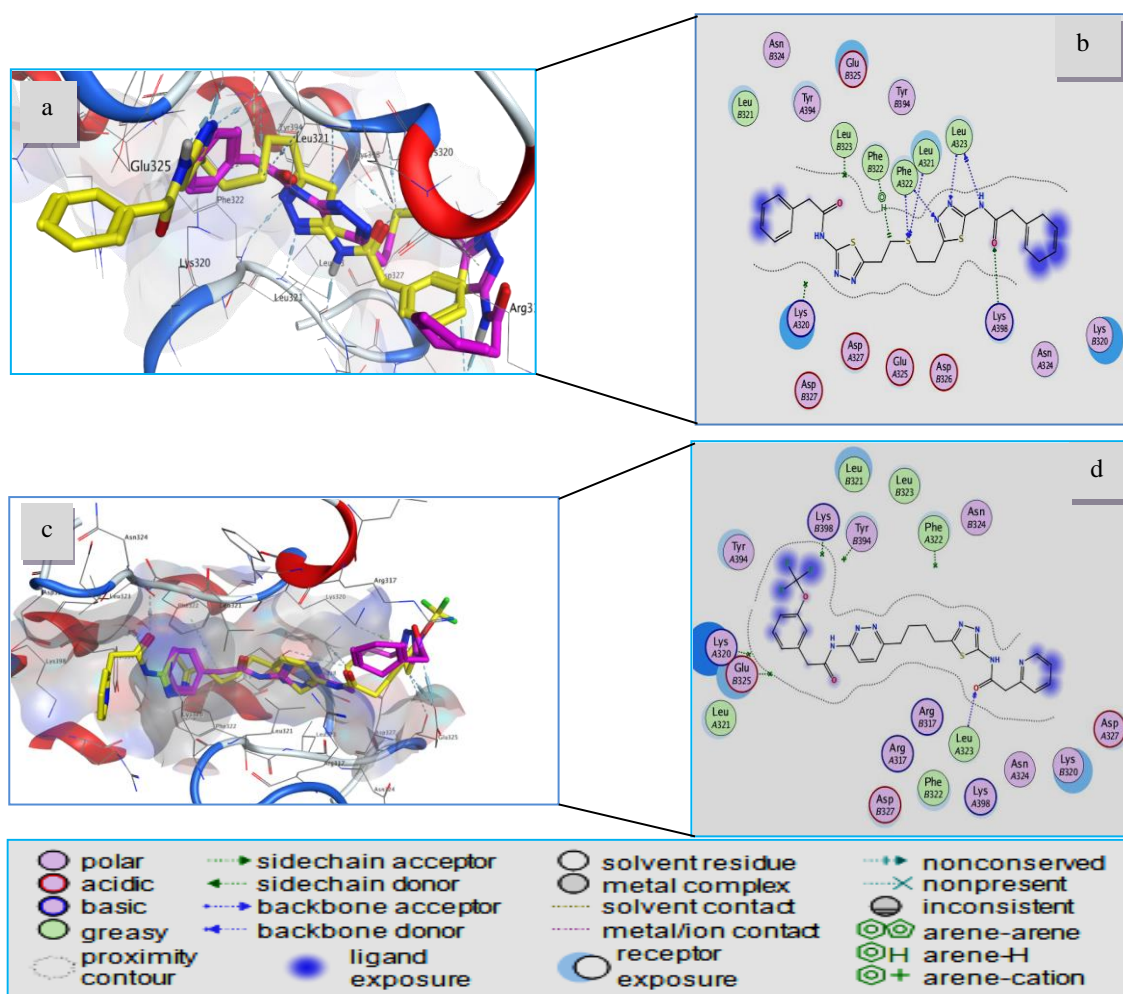
Figure IV.2. (a) 3D structure of the superimposition of native co-crystallized BPTES (yellow) and docked co-crystallized BPTES (pink); **(b)** 2D representation of the interactions of docked BPTES with the binding pocket of 3UO9; **(c)** 3D structure of the superimposition of cb-839 (yellow) and docked co-crystallized BPTES (pink); **(d)** 2D representation of the interactions of docked cb-839 with the binding pocket of 3UO9.

b) 3UO9- WA derivatives interactions

Top-scoring docked compounds were chosen based on their low energy and interaction with receptor active site residues. In MOE software, the "Ligand Interactions" tool was used to visualize the binding sites of 3UO9-WA derivatives.

As shown in Table IV.4, the binding scores of all complexes formed range between -5.47 and -9.41 kcal/mol. Furthermore, two significant interactions were observed: hydrogen bonds and hydrophobic interactions (H-Pi) between the compounds and the pocket of the GAC target. Finally, these results were compared with the co-crystallized ligand BPTES and the clinical compound CB-839.

All WA derivatives were found to bind at the same allosteric binding site as BPTES and CB-839. The four compounds A5, A8, A13, and A18 form complexes with high negative binding scores: -8.49, -9.41, -8.58, and -8.07 Kcal/mol, respectively.

It appears that the 3UO9-A8 complex has the highest negative score (-9.41 kcal/mol) toward GAC compared to the other compounds and the reference drug CB-839, which has a docking score of -8.31 kcal/mol (Table **IV.**4). Moreover, this compound has shown a low value of cytotoxicity with an IC50 of 0.94 μM against the TNBC cell line. Similarly, both compounds: A5 and A13 fit well in the binding pocket of GAC, and they have good binding affinity than the reference drug CB-839.

According to the study by A.Imberty et al. [43], which revealed that the interaction distance of hydrogen bonds belonging to the interval between 2.5 and 3.1Å is considered strong. While weak hydrogen bonds have a distance between 3.1 and 3.5 Å.

Table IV.4 shows that compound A5 established two hydrogen bond donors at distances of 3.38 Å and 3.52 Å, and bond energy of -0.8 and -0.7 kcal/mol with the residues GLU325(B) and ASP327(B), respectively. Compound A_8 forms three weak hydrogen bond acceptors with LYS320(A), LYS398(B), and PHE322(B) at distances 3.26 Å, 3.17 Å, and 3.66 Å, and bond energies of -0.8 kcal/mol, -0.8 kcal/mol, and -0.7 kcal/mol, respectively. It also exhibits two strong hydrogen bonds; one hydrogen bond donor with LYS320(A) at 2.90 Å and bond energy of -2.7 kcal/mol, the other hydrogen bond acceptor with PHE322(B) at 3.00 Å and bond energy of -1.8 kcal/mol (see Figure IV.3). As displayed in Figure IV.4, the compound A13 interacts with LYS398(B) and PHE322(B) via weak hydrogen bond acceptors

and LEU323(B) via a strong hydrogen bond acceptor with a distance of 3.14 Å, 3.10 Å, and 2.95 Å, and bond energies of -3.9 kcal/mol, -3.6 kcal/mol, and -3.0 kcal/mol respectively. On the other hand, compound A18 exhibits four hydrogen bonds. Two strong hydrogen bond acceptors were established with ARG317(B) and two further hydrogen bond acceptors with LYS 320 (A) (see Table IV.4 and Figure IV.4). Likewise, the binding score values of the 26 ligands docked in the allosteric site that located between C and D chains ranged between -5.33 and -9.16 kcal/mol (see Appendix A).

Based on the analysis of binding sites 2 and 3, their binding energies fell within the range of -5.68 to -9.09 and -5.70 to -6.83, respectively. The binding potential at these sites is comparatively weaker than that at site 1. Moreover, the analysis also showed that the majority of interacting residues were associated with ASP467(B) and GLY470(B) for site 2, and ARG534(A) and ASP541(A) for site 3 (see Appendix E and F). It is worth noting that no prior studies have suggested that these amino acids are associated with inhibitory activity at the allosteric site of GAC.

As a result, the analysis confirms that site 1 is the most potential binding site, as it fits well with most compounds and forms interactions with the most important amino acids.
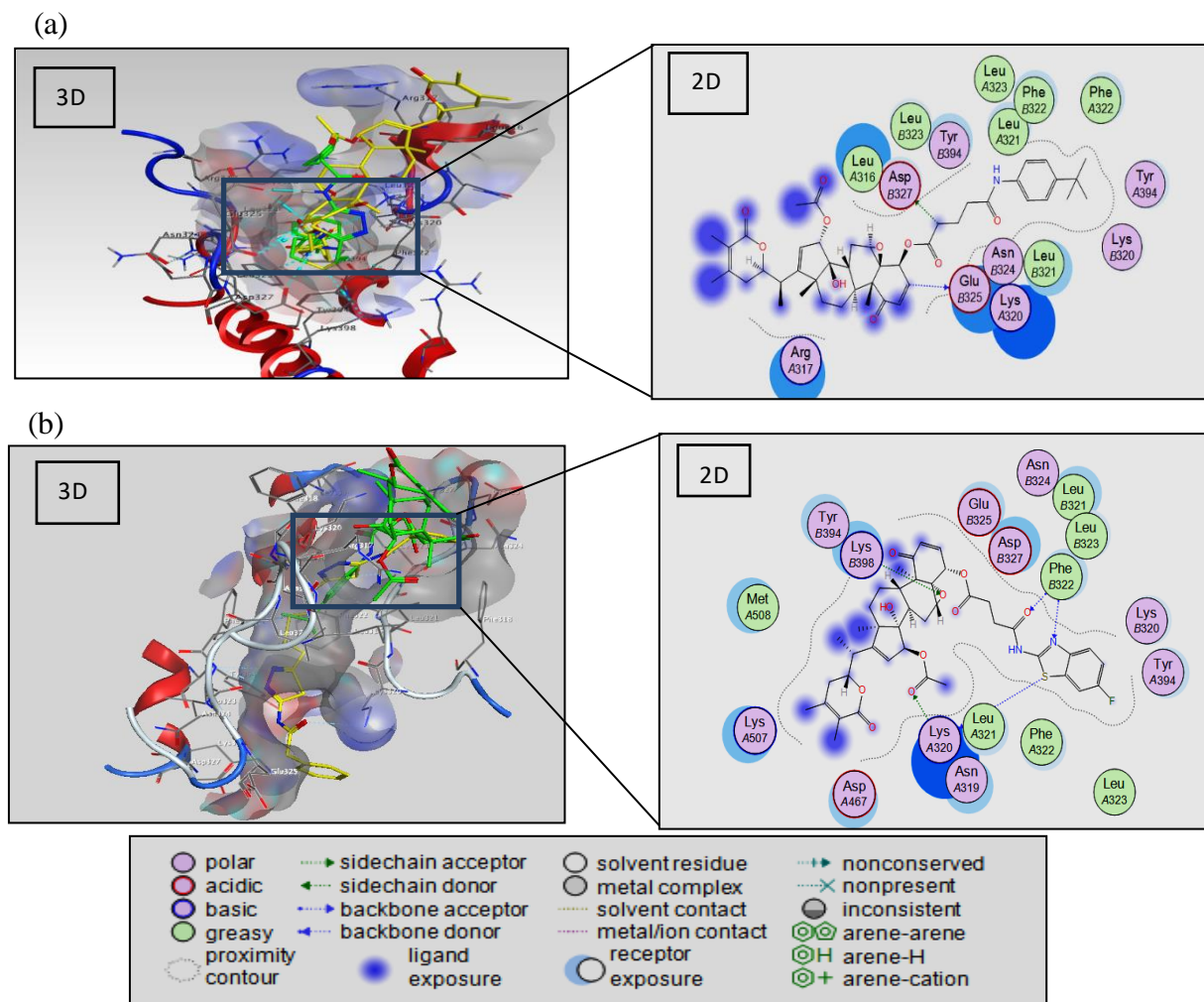
Figure IV.3. 3D crystal structure and 2D representation of interactions of a) A5 and b) A8 complexed with the allosteric site of 3UO9. The overlay of the co-crystalized ligand of BPTES is shown in green sticks and docked compound in yellow sticks.
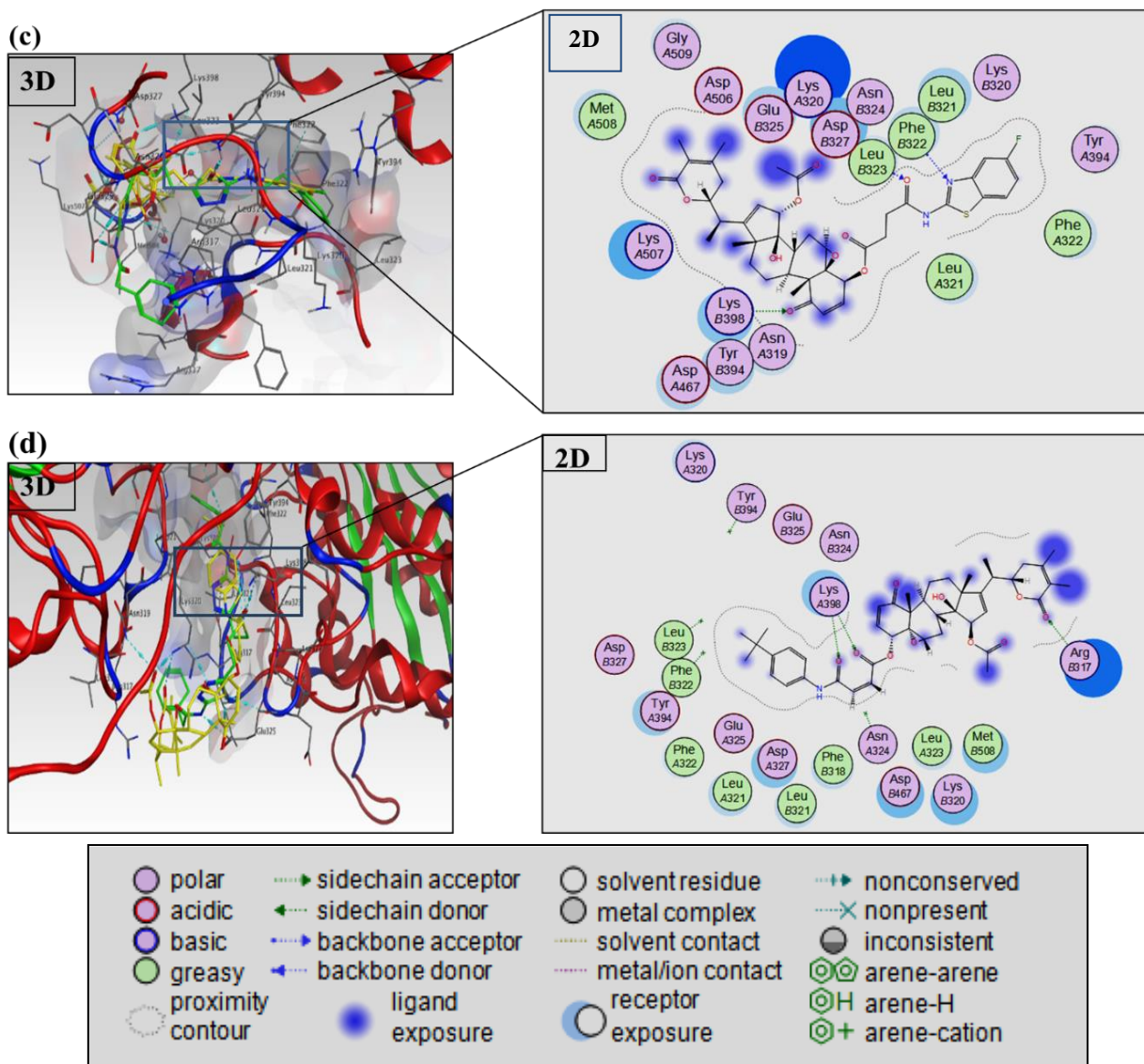
Figure IV.4. 3D crystal structure and 2D representation of interactions of c) A13 and d) A18 complexed with the allosteric site of 3UO9. The overlay of co-crystalized ligand of BPTES is shown in green sticks and docked compound in yellow sticks.

3.2. MD simulations

MD simulation allowed us to determine the binding interactions in a more accurate way than molecular. Thus, MD simulations were carried out for the best-selected compounds and the reference drug CB-839 within the target receptor (GAC) during the simulation time of 800

ps. According to the plots (Figure IV.5), which represent the evolution of potential energy U (Kcal/mol) as a function of time (ps) for the best complexes, it is apparent that the potential energy of the **3UO9-A5** complex fluctuates from -4961.13 to 8135.60 Kcal/mol at the first 10 ps. After that, the complex reached equilibrium at 14 ps with an energy potential value of 3600 kcal/mol. Likewise, the potential energy of the **3UO9-A8** complex varied from -5217.00 to 8180.03 Kcal/mol during the first 12 ps, and it begins to reach stability at around 5104.96 Kcal/mol (see Figure IV.5). Furthermore, the **3UO9-A13** complex potential energy fluctuates from -6710.87 to 6633.94 Kcal/mol during the first 10 ps and tends to stabilize during the MD simulation run. However, the potential energy of the **3UO9-A18** complex showed a significant fluctuation from 6962.46 to 4429.42 Kcal/mol at the beginning of the simulation. After that, it presents a slight variation during the next 200 ps until a stable state is observed in potential energy (see Figure IV.5). Regarding the **3UO9-CB839** complex, the potential energy keeps fluctuating from 5846.99 to 5416.91 kcal/mol during the 400 ps. After that, a slight variation in potential energy could be observed before stabilizing at around 5300 kcal/mol.
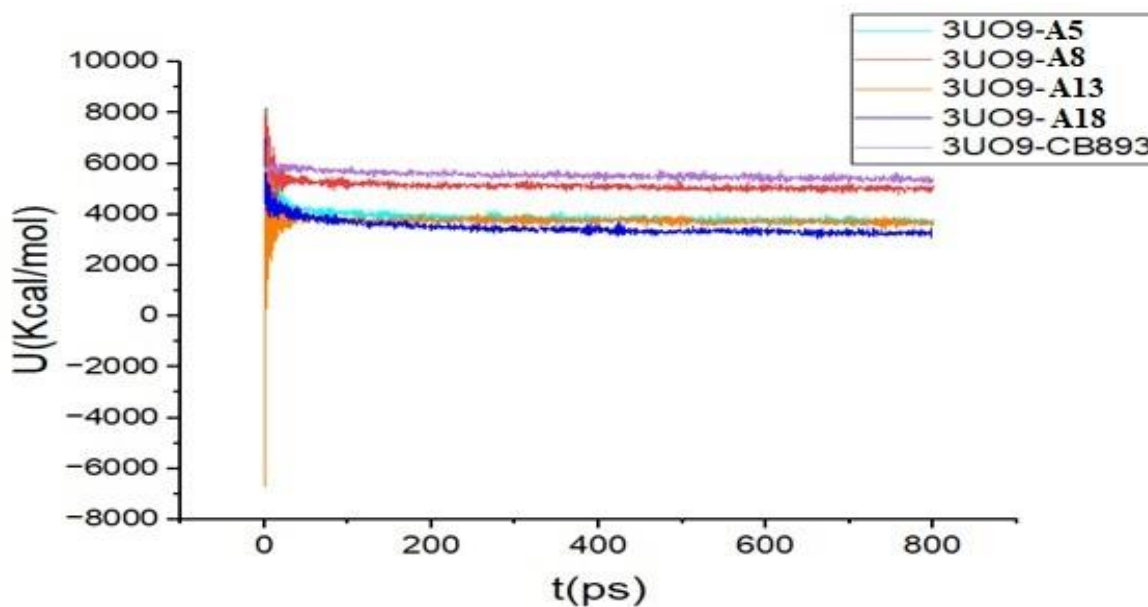
Figure IV.5. The variation of the potential energy as a function of time for the best complexes

3.2.1. Protein-ligand interactions after MD simulations

The MD simulation results (the binding interactions) are summarized in Table IV.5. Table IV.5 shows that the results at the allosteric site differ slightly from those obtained from

molecular docking simulation. It is apparent that compound A5 established three hydrogen bonds with the allosteric site of the 3UO9 target: a strong H-donor (2.72 Å) [44] and bond energy of -3.7 kcal/mol with ASP467(A) and two strong H-acceptors with ARG317(A) and ARG317(B), respectively (distances: 2.84 and 2.86 Å and bond energies : -6.0 kcal/mol and -3.2 kcal/mol) (Figure IV.6).

3UO9-A8 Complex maintained its original site by establishing a strong H-donor (distance :2.81 Å and bond energy: -4.8 kcal/mol) and a weak one (3.78 Å) [44] with GLU325(B), as well as a hydrophobic interaction H-Pi with PHE318(A) (Figure IV.6). The compound A13 shows two strong hydrogen bond acceptors with the same residue: ARG317(A) (distances: 3.07 and 2.77 Å, bond energies: -2.7 and -1.5 kcal/mol). PHE332(B), an essential amino acid, has also been involved in the interaction by forming a strong H-acceptor (distance: 2.85 Å and bond energy: -.1.6 kcal/mol). The water molecule is also represented in binding by the H-acceptor bond (2.37 Å), and the same interaction from molecular docking was retained by the residue LEU323(B) (see Figure IV.7).

The 3UO9-A18 complex maintained the same molecular docking interactions with the residue ARG317(B) via a strong H-acceptor bond (distance: 2.54 Å and energy: -3.1 kcal/mol) and two strong H-acceptor bonds (distances : 2.97 and 2.74 Å, and bond energies: -2.4 and -10.2 kcal/mol) with LYS398(A), another interaction with the residue ASP327(A) was observed via a strong H-donor bond (2.93 Å) (see Figure IV.7). CB-839 exhibited five binding interactions, including four weak hydrogen acceptor bonds [44] with TYR394(B), PHE322(A), ARG387(A), and LEU323(A) at a distance of 3.17, 3.11, 3.28 and, 3.11 Å, and bond energies of -1.1, -4.1, -1.3, and -2.4 kcal/mol, respectively. In addition, this compound formed a strong hydrogen bond donor with LEU323(A) with distance of 3.10 Å and bond energy of -2.4kcal/mol.

At the end, MD simulation results for these four complexes validate molecular docking experiments. Hence, compounds with hydrogen bonds are stronger throughout the MD run. Furthermore, the selected compounds formed the same binding interactions as CB-8393. Similarly, new interactions were formed with the key residues PHE322(A), LEU323(A), and ARG317(B). Knowing that these amino acids were reported to have a crucial role in the allosteric site of GAC [43, 45].

Table IV.5. MD simulation results of the best compounds and CB-839 with the allosteric site of GAC (pdb: 3UO9).

Compound	Atom of ligand	Bonds between ligands and of the active site residues				
		Involved receptor atoms	Involved receptor residues	Type of interaction	Distance (Å)	Bond Energy (Kcal/mol)
A5	O 47	OD1	ASP467(A)	H-donor	2.72	-3.7
	O 65	NH1	ARG317(A)	H-acceptor	2.84	-6.0
	O 86	NH1	ARG317(B)	H-acceptor	2.86	-3.2
A8	N 78	OE2	GLU325(B)	H-donor	2.81	-4.8
	S 84	OE2	GLU325(B)	H-donor	3.78	-0.7
	C 98	6-ring	PHE318(A)	H-Pi	3.92	-1.2
A13	O 6	NH1	ARG317(A)	H-acceptor	3.07	-2.7
	O 6	NH2	ARG317(A)	H-acceptor	2.77	-1.5
	O 65	O	HOH68(A)	H-acceptor	2.37	-11.3
	O 77	N	PHE322(B)	H-acceptor	2.85	-1.6
	O 77	N	LEU323(B)	H-acceptor	2.89	-4.1
A18	C 45	OD1	ASP327(A)	H-donor	2.93	-0.8
	O 36	NH1	ARG317(B)	H-acceptor	2.54	-3.1
	O 42	NZ	LYS398(A)	H-acceptor	2.97	-2.4
	O 46	NZ	LYS398(A)	H-acceptor	2.74	-10.2
CB-839	N 13	O	LEU323(A)	H-donor	3.10	-2.4
	N 8	OH	TYR394(B)	H-acceptor	3.17	-1.1
	N 9	N	PHE322(A)	H-acceptor	3.11	-4.1
	N 11	NH2	ARG387(A)	H-acceptor	3.28	-1.3
	N 12	N	LEU323(A)	H-acceptor	3.11	-2.4
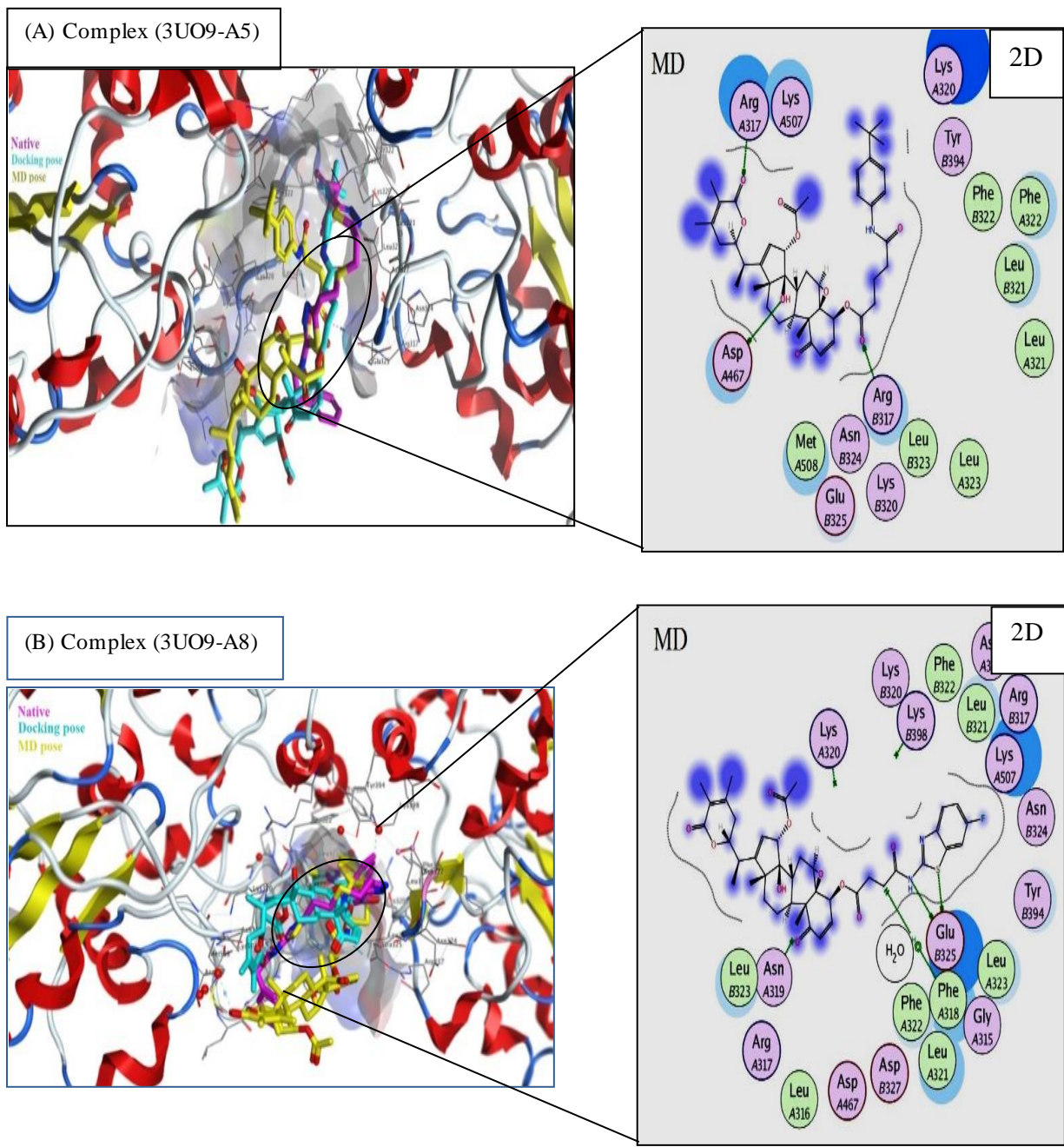
Figure IV.6. 3D and 2DStructure comparison after MD between the native ligand BPTES (pink), docking (cyan), and MD (yellow) poses of the studied compounds towards the allosteric sites of the target protein.
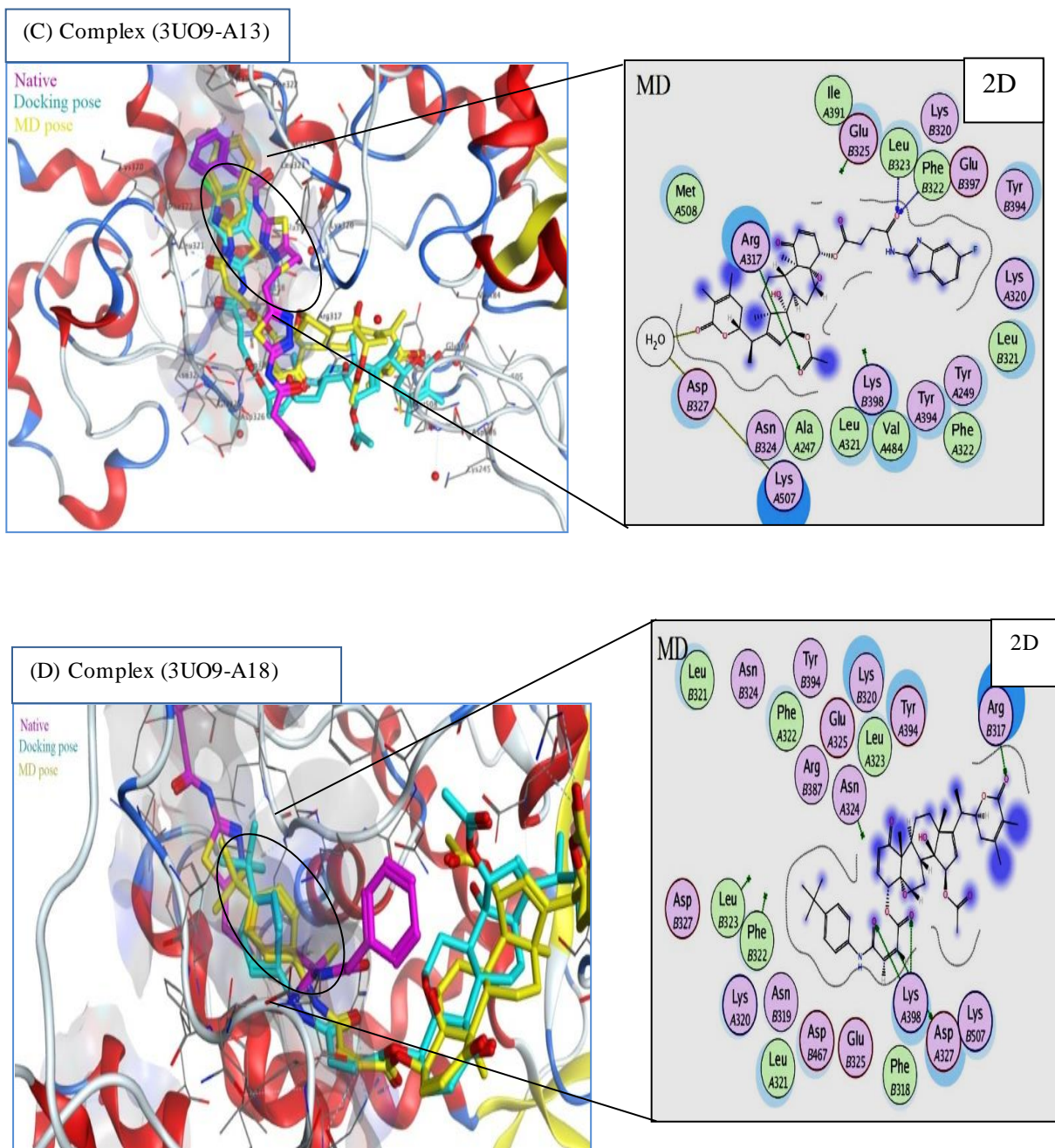
Figure IV.7. 3D and 2DStructure comparison after MD between the native ligand BPTES (pink), docking (cyan), and MD (yellow) poses of the studied compounds towards the allosteric sites of the target protein.

3.3. MEP analysis

The molecular electrostatic potential (MEP) is a useful tool for predicting the reactive sites of molecules. It helps determine the electron density distribution and understand the site of the electrophilic and nucleophilic attack and hydrogen bonding interactions [46]. Figure IV.8 displays the map of electrostatic potential for the selected compounds. Low electron densities are characteristic of regions with positive electrostatic potential, which confirms their electrophilic character, while negative electrostatic potential indicates high electron density and susceptibility to electrophilic attack. The map is colour-coded, the blue regions indicate a favourable site for nucleophilic attack and the red areas are sites for electrophilic attack. Orange and cyan spheres correspond to the positions of maxima and minima, respectively.

As can be seen in Figure IV.7, the global surface minimum of compound A5 with a value of -56.31 kcal/mol is observed near the O36 of the carbonyl group of the WA scaffold, indicating a high electronic density favourable for the electrophilic attack and demonstrating the possibility to form a hydrogen bond acceptor. The global surface maximum, which has a value of 38.48 kcal/mol, is found around the hydrogen atom of nitrogen (N47) of the amide group substituent. This demonstrates the poor electronic density in this region caused by the amide carbonyl withdrawing effect.

In compound A8, the carbonyl oxygen atom O3 of the acetoxy group (-OAc) displayed the minimum potential surface, indicating a high electron density and a potential hydrogen bond acceptor, which was observed in docking results through the formation of an H-acceptor with LYS398(B). Furthermore, the global surface maximum of A8 was detected around the hydrogen atom of N43 from the amide function with a value of 57.70 kcal/mol, indicating a potential hydrogen bond donor (see Figure IV.8).

The global minimum for compound A13 is close to the carbonyl oxygen atoms O36 of the WA scaffold, with an ESP value of -54.75 kcal/mol, making it susceptible to electrophilic attack. Besides, the global maximum is found near the hydrogen of N43, with a value of 63.97 kcal/mol.

The global surface minimum for A18 is also located close to the carbonyl oxygen atoms of the WA scaffold, making it susceptible to electrophilic attack. Besides, the global maximum is

found around the carbon atom (C9) of the WA scaffold near the ester substituent with an ESP value of 33.37 kcal/mol indicating the poor electron density at C9 caused by the tautomerization effect of α, β-unsaturated of the WA moiety and ester substitution, which susceptible for the nucleophilic attack(see Figure IV.8).

Significantly, The MEP analysis of the four compounds (A5, A8, A13, and A18) revealed that the carbonyl oxygen atoms of the WA scaffold are the most vulnerable site for electrophile attack, indicating the formation of a hydrogen bond acceptor, while the hydrogen atom of the amide group substituent represents a low electronic density area and the most susceptible site for the nucleophilic attack.

These findings were in agreement with MD simulation results.
Furthermore, the four compounds (A5, A8, A13, and A18) appear to occupy 63.1%, 63.23%, 53.17%, and 61.73% of the total surface, respectively, indicating that they have large positive surfaces. Consequently, they have an electrophilic character. Thus, nucleophilic reagents are more likely to attack these electron-deficient areas.

ESP map   Optimized structure

A5

PS= 495.856
NS= 290.449
PV= 52.558
NV= 257.657

A8

PS= 460.326
NS= 267.682
PV= 91.271
NV= 255.997

A13

PS= 435.124
NS= 300.149
PV= 101.958
NV= 196.923

A18

PS= 467.314
NS= 289.660
PV= 70.950
NV= 203.216
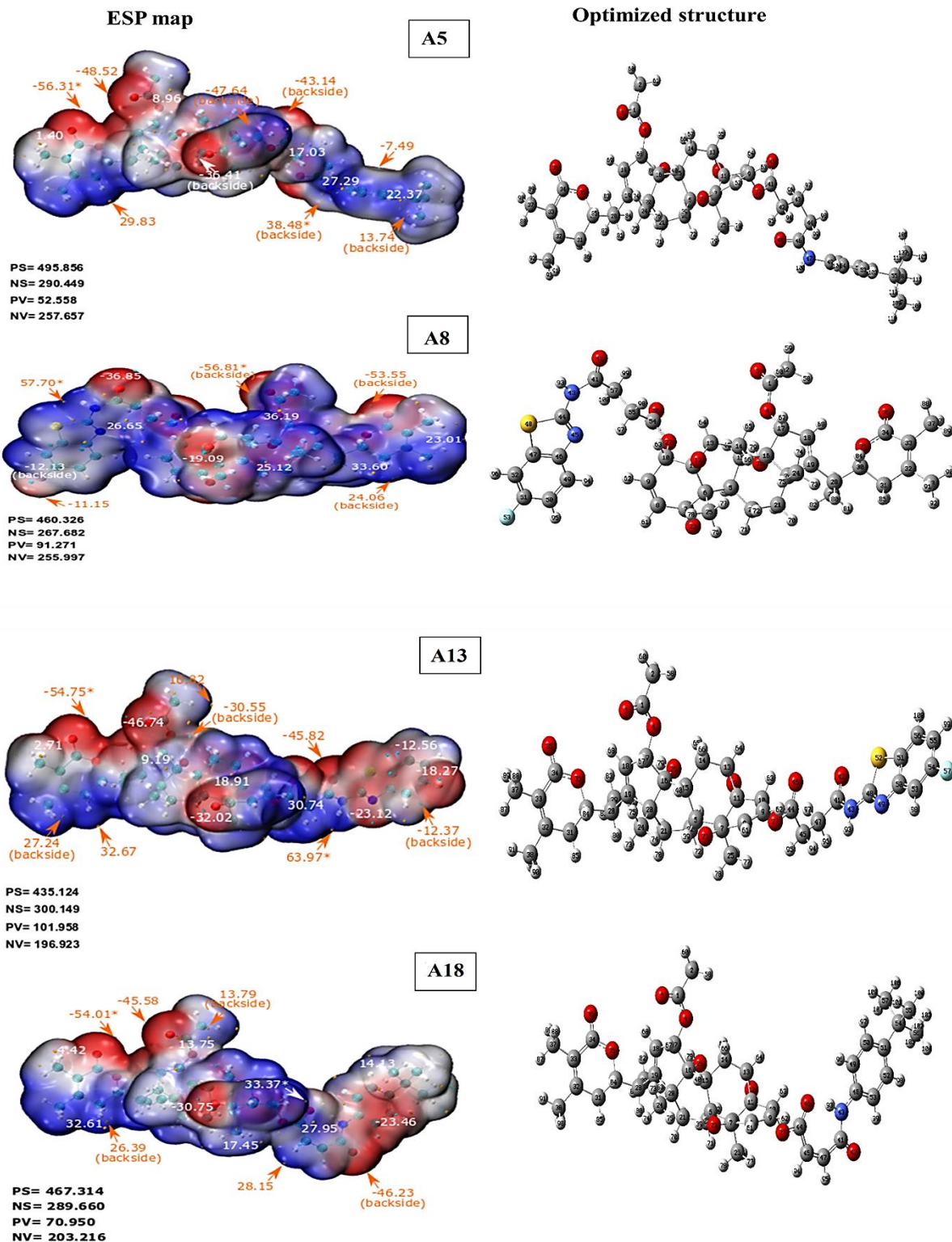
Figure IV.8. ESP-mapped (Kcal/mol) Van der Waals surfaces, with a color scale ranging from red for negative ESP through white for neutral ESP to blue for positive ESP The van der Waals surface represents the iso-surface of ρ =0.001 a.u. and the grid spacings were adjusted

to 0.2 Bohr. Values marked with a star indicate the global extremums. The positive surface area (PS), negative surface area (NS), positive variance (PV), and negative variance (NV) are represented in the bottom left corner by bold numerals with the units [Kcal/mol]2, (Å)2, respectively. The optimized structure and their numbering atomic labels at the UB3LYP/6-31 G (d, p) level theory are represented on the right side.

3.4. Pharmacokinetics predictions

Table IV.6 displays the pharmacokinetic analysis of the top four compounds.

Table IV.6. The pharmacokinetics prediction profile of the best compounds.

Properties	Reference [42]	Compounds				
		A5	A8	A13	A18	CB-839
Absorption						
Intestinal absorption (human)**%**	<30% is poorly absorbed	94.09	98.51	98.44	92.57	80.60
Caco-2 permeability(cm /s)	>0.90 high permeability	0.59	1.06	1.00	0.62	1.36
Skin Permeability (log Kp)	> -2.5 is low	-2.73	-2.73	-2.73	-2.73	-2.73
P-glycoprotein substrate	-	Yes	Yes	Yes	Yes	Yes
P-glycoprotein I inhibitor	-	Yes	Yes	Yes	Yes	Yes
P-glycoprotein II inhibitor	-	Yes	Yes	Yes	Yes	Yes
Distribution						
VDss (human) (log L/kg)	low is < -0.15, High is > 0.45	0.16	0.36	0.38	0.13	-0.21
BBB permeability (Log BB)	logBB> 0.3 cross the BBB logBBB< -1 poorly distributed to brain	-0. 53	-1.13	-1.11	-0.49	-1.97
CNS permeability (log PS)	Penetrate: log PS >-2, Unpenetrate: log PS < -3	-2.37	-3.05	-3.05	-2.41	-3.9

Table IV.6. Continued

Metabolism						
CYP1A2 inhibitor	-	No	No	No	No	No
CYP2C19 inhibitor	-	No	No	No	No	No
CYP2C9 inhibitor	-	No	No	No	No	Yes
CYP2D6 inhibitor	-	No	No	No	No	No
CYP3A4 inhibitor	-	Yes	Yes	Yes	Yes	Yes
Excretion						
Total clearance (log ml/min/kg)	-	9.21	6.89	6.84	5.15	2.30
Renal OCT2 substrate	-	No	No	No	No	No
Toxicity						
AMES toxicity	No	No	No	No	No	No
Max. tolerated dose (human)	Low is ≤ 0.477, High is >0.477	-0.473	-0.441	-0.493	-0.81	0.521
Oral Rat Acute Toxicity (LD50) (mol/kg)	-	3.13	3.15	3.19	3.14	2.32
hERG I inhibitor	No	No	No	No	No	No
hERG II inhibitor	No	No	No	No	No	Yes
Hepatotoxicity	No	Yes	No	No	No	Yes
Skin Sensitisation	No	No	No	No	No	No

The results show that the top four candidates and the reference compound CB-839 have high intestinal absorption rates, with values ranging from 80.60%to 92.57%. The Caco-2 cell permeability reveals that the selected compounds and the reference CB-839, can pass through the intestinal epithelial cell barrier, except for compounds A5 and A18, which have Caco-2 permeability values less than 0.9.

The top four selected compounds and CB-839 are considered to have relatively moderate skin permeability with the same value of -2.73. It is also apparent that all compounds were found to be P-gp substrates, as well as P-gpI and P-gpII inhibitors.

Regarding distribution, the VDss (human) values of these compounds are predicted to be moderate, with values ranging from 0.13 to 0.38 log L/kg. On the other hand, CB-839 gave a low distribution with a value of -0.21 log L/kg.

Furthermore, all compounds are predicted to be unable to cross the blood-brain barrier (BBB), while compound A18 has moderate BBB permeability. Likewise, CB-839 showed poor BBB permeability with a logBB value of 1.97. However, these compounds and CB-839 are unable to penetrate the central nervous system (CNS).

The CYP450 isoform prediction indicates that all compounds are predicted to act as non-inhibitors of CYP1A2, CYP2C9, CYP2C19, and CYP2D6 except for the reference compound CB-839, which acts as an inhibitor of CYP2C9. Moreover, they were all found to act as CYP3A4 inhibitors.

The results show that these compounds have moderate clearance, with values ranging from 5.15 to 9.21 ml/min/kg, except for CB-839, which has a low clearance with a value of 2.30 ml/min/kg. In addition, it was observed that not all compounds are likely to be renal OCT2 substrates.

Based on the toxicity prediction, the selected compounds showed no risk of AMES toxicity, and no skin sensitivity was observed. CB-839 has a high maximum tolerated dose of 0.52, while other compounds displayed low values.

The predicted oral rat acute toxicity (LD50) values of the selected candidates varied from 3.13 to 3.19 mol/kg, which were higher than the reference compound CB-839 (2.39 mol/kg). Moreover, the selected compounds are neither hERG I nor hERG II inhibitors. However, the standard compound CB-839 acts as a HERG II inhibitor. CB-839 and compound A5 are predicted to have a potential risk of hepatotoxicity, in contrast to other compounds that showed no hepatotoxicity risk.

These findings suggest that only compounds: A8 and A13 are without hepatotoxicity, they are non-hERG inhibitors and do not induce skin sensitization. Additionally, they also exhibit relatively high Caco-2 and HIA absorption. They displayed the best ADMET profile compared to other compounds.

4. Conclusion

In this study, molecular docking was used to investigate the binding mechanism by which a set of WA derivatives interact with the allosteric site of GAC in a similar mode to the BPTES and CB-839. According to the results obtained, the best four compounds A5, A8, A13, and A18, were selected based on their score energies. These compounds fit well in the allosteric GAC pocket and occupy the same binding site as BPTES and CB-839. MD simulation was further used to validate the stability of the best complexes (3UO9-A5, 3UO9-A8, 3UO9-A13, and 3UO9-A18) and the reference compound CB-839. The results reveal that the best compounds establish the interactions involved with the key residues namely: PHE322(A), LEU323(A), and ARG317(B). Electrostatic potential analysis showed that the compounds have electrophilic characteristics, with the carbonyl oxygen atoms being the most vulnerable site for electrophile attack. Furthermore, MEP analyses were performed in combination with docking and dynamic simulations, which showed that the favorable reactive sites of these compounds formed hydrogen bond interactions with residues of the target. Finally, pharmacokinetics prediction showed that A8 and A13 had the best ADMET profile and share similar properties with the reference compound CB-839. These comparison results demonstrate that the chosen compounds might be used for further development as a novel class of allosteric GAC inhibitors for treating TNBC.

REFERENCES

1.   Zhu T, Chen C, Wang S, Zhang Y, Zhu D, Li L, et al. Cellular target identification of withangulatin A using fluorescent analogues and subsequent chemical proteomics. Chemical Communications. 2019;55(57):8231-4.

2.   Zhuang Y, Wang Y, Li N, Meng H, Li Z, Luo J, et al. Hydrolytic Metabolism of Withangulatin A Mediated by Serum Albumin Instead of Common Esterases in Plasma. European Journal of Drug Metabolism and Pharmacokinetics. 2023:1-14.

3.   Yin L, Duan J-J, Bian X-W, Yu S-c. Triple-negative breast cancer molecular subtyping and treatment progress. Breast Cancer Research. 2020;22(1):1-13.

4.   Zagami P, Carey LA. Triple negative breast cancer: Pitfalls and progress. npj Breast Cancer. 2022;8(1):95.

5.   Lampa M, Arlt H, He T, Ospina B, Reeves J, Zhang B, et al. Glutaminase is essential for the growth of triple-negative breast cancer cells with a deregulated glutamine metabolism pathway and its suppression synergizes with mTOR inhibition. PloS one. 2017;12(9):e0185092.

6.   Ren L, Ruiz-Rodado V, Dowdy T, Huang S, Issaq SH, Beck J, et al. Glutaminase-1 (GLS1) inhibition limits metastatic progression in osteosarcoma. Cancer & Metabolism. 2020;8(1):4.

7.   Altman BJ, Stine ZE, Dang CV. From Krebs to clinic: glutamine metabolism to cancer therapy. Nature Reviews Cancer. 2016;16(10):619-34.

8.   Song M, Kim S-H, Im CY, Hwang H-J. Recent development of small molecule glutaminase inhibitors. Current Topics in Medicinal Chemistry. 2018;18(6):432-43.

9.   Li T, Copeland C, Le A. Glutamine Metabolism in Cancer. In: Le A, editor. The Heterogeneity of Cancer Metabolism. Cham: Springer International Publishing; 2021. p. 17-38.

10.   Yoo HC, Yu YC, Sung Y, Han JM. Glutamine reliance in cell metabolism. Experimental & molecular medicine. 2020;52(9):1496-516.

11.   Soth MJ, Le K, Di Francesco ME, Hamilton MM, Liu G, Burke JP, et al. Discovery of IPN60090, a clinical stage selective glutaminase-1 (GLS-1) inhibitor with excellent pharmacokinetic and physicochemical properties. Journal of medicinal chemistry. 2020;63(21):12957-77.

12. Wu C, Chen L, Jin S, Li H. Glutaminase inhibitors: a patent review. Expert Opinion on Therapeutic Patents. 2018;28(11):823-35.

13. Cassago A, Ferreira AP, Ferreira IM, Fornezari C, Gomes ER, Greene KS, et al. Mitochondrial localization and structure-based phosphate activation mechanism of Glutaminase C with implications for cancer metabolism. Proceedings of the National Academy of Sciences. 2012;109(4):1092-7.

14. Shen Y-A, Chen C-L, Huang Y-H, Evans EE, Cheng C-C, Chuang Y-J, et al. Inhibition of glutaminolysis in combination with other therapies to improve cancer treatment. Current opinion in chemical biology. 2021;62:64-81.

15. Lee Y-Z, Yang C-W, Chang H-Y, Hsu H-Y, Chen I-S, Chang H-S, et al. Discovery of selective inhibitors of Glutaminase-2, which inhibit mTORC1, activate autophagy and inhibit proliferation in cancer cells. Oncotarget. 2014;5(15):6087.

16. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. Journal of Natural Products. 2020;83(3):770-803.

17. Atanasov AG, Zotchev SB, Dirsch VM, Orhan IE, Banach M, Rollinger JM, et al. Natural products in drug discovery: advances and opportunities. Nature Reviews Drug Discovery. 2021;20(3):200-16.

18. Singh A, Raza A, Amin S, Damodaran C, Sharma AK. Recent Advances in the Chemistry and Therapeutic Evaluation of Naturally Occurring and Synthetic Withanolides. Molecules. 2022;27(3):886.

19. Choudhary MI, Yousuf S, Atta ur R. Withanolides: Chemistry and Antitumor Activity. In: Ramawat KG, Mérillon J-M, editors. Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 3465-95.

20. Misico RI, Nicotra VE, Oberti JC, Barboza G, Gil RR, Burton G. Withanolides and related steroids. Progress in the Chemistry of Organic Natural Products Vol 94. 2011:127-229.

21. Zhou W-X, Chen C, Liu X-Q, Li Y, Kong L-Y, Luo J-G. Synthesis and biological evaluation of novel withangulatin A derivatives as potential anticancer agents. Bioorganic Chemistry. 2021;108:104690.

22.   Zhou W-X, Chen C, Liu X-Q, Li Y, Lin Y-L, Wu X-T, et al. Discovery and optimization of withangulatin A derivatives as novel glutaminase 1 inhibitors for the treatment of triple-negative breast cancer. European Journal of Medicinal Chemistry. 2021;210:112980.

23.   HyperChem. molecular modelling system, Hypercube Inc. v8 ed. 1115 NW 4th Street, Gainesville, FL 32601, USA, 2009

Hypercube Inc.

24.   Environment MO. Molecular Operating Environment (MOE), 2015.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.; 2015.

25.   DeLaBarre B, Gross S, Fang C, Gao Y, Jha A, Jiang F, et al. Full-Length Human Glutaminase in Complex with an Allosteric Inhibitor. Biochemistry. 2011;50(50):10764-70.

26.   Ball P. Water is an active matrix of life for cell and molecular biology. Proceedings of the National Academy of Sciences. 2017;114(51):13327-35.

27.   Stanzione F, Giangreco I, Cole JC. Chapter Four - Use of molecular docking computational tools in drug discovery. In: Witty DR, Cox B, editors. Progress in Medicinal Chemistry. 60: Elsevier; 2021. p. 273-343.

28.   Kim KH. Outliers in SAR and QSAR: 3. Importance of considering the role of water molecules in protein–ligand interactions and quantitative structure–activity relationship studies. Journal of Computer-Aided Molecular Design. 2021;35(3):371-96.

29.   Xiao S, Tian H, Tao P. PASSer2. 0: Accurate Prediction of Protein Allosteric Sites Through Automated Machine Learning. Frontiers in Molecular Biosciences. 2022:619.

30.   Daoud I, Melkemi N, Salah T, Ghalem S. Combined QSAR, molecular docking and molecular dynamics study on new Acetylcholinesterase and Butyrylcholinesterase inhibitors. Computational biology and Chemistry. 2018;74:304-26.

31.   Daoud I, Mesli F, Melkemi N, Ghalem S, Salah T. Discovery of potential SARS-CoV 3CL protease inhibitors from approved antiviral drugs using: virtual screening, molecular docking, pharmacophore mapping evaluation and dynamics simulation. Journal of Biomolecular Structure and Dynamics. 2021:1-18.

32.   Cole JC, Murray CW, Nissink JWM, Taylor RD, Taylor R. Comparing protein–ligand docking programs is difficult. Proteins: Structure, Function, and Bioinformatics. 2005;60(3):325-32.

33.   Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev. 2006;106(5):1589-615.

34.   Belkadi A, Kenouche S, Melkemi N, Daoud I, Djebaili R. Molecular docking/dynamic simulations, MEP, ADME-TOX-based analysis of xanthone derivatives as CHK1 inhibitors. Structural Chemistry. 2022;33(3):833-58.

35.   Bond SD, Leimkuhler BJ, Laird BB. The Nosé–Poincaré method for constant temperature molecular dynamics. Journal of Computational Physics. 1999;151(1):114-34.

36.   Origin. Origin(Pro). 2022 ed. Northampton, MA, USA.: OriginLab Corporation; 2022.

37.   Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Revision A.02, ed: Gaussian, Inc., Wallingford CT, 2016.

38.   Barone V, Cossi M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. The Journal of Physical Chemistry A. 1998;102(11):1995-2001.

39.   Lu T, Chen F. Multiwfn: A multifunctional wavefunction analyzer. Journal of Computational Chemistry. 2012;33(5):580-92.

40.   Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996;14(1):33-8.

41.   Pantaleão SQ, Fernandes PO, Gonçalves JE, Maltarollo VG, Honorio KM. Recent advances in the prediction of pharmacokinetics properties in drug design studies: a review. ChemMedChem. 2022;17(1):e202100542.

42.   Pires DE, Blundell TL, Ascher DB. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. J Med Chem. 2015;58(9):4066-72.

43.   Milano SK, Huang Q, Nguyen T-TT, Ramachandran S, Finke A, Kriksunov I, et al. New insights into the molecular mechanisms of glutaminase C inhibitors in cancer cells using serial room temperature crystallography. Journal of Biological Chemistry. 2022;298(2):101535.

44.   Imberty A, Hardman KD, Carver JP, Perez S. Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A. Glycobiology. 1991;1(6):631-42.

45.   Huang Q, Stalnecker C, Zhang C, McDermott LA, Iyer P, O'Neill J, et al. Characterization of the interactions of potent allosteric inhibitors with glutaminase C, a key enzyme in cancer cell glutamine metabolism. J Biol Chem. 2018;293(10):3535-45.

46.   Alcamí M, Mó O, Yáñez M. Modelling Intrinsic Basicities: The Use of the Electrostatic Potentials and the Atoms-in-Molecules Theory. In: Murray JS, Sen K, editors. Theoretical and Computational Chemistry. 3: Elsevier; 1996. p. 407-56.

GENERAL CONCLUSION

During the scope of PhD research, two investigations were conducted in this dissertation to enhance the understanding of breast cancer treatment drugs and potential targets. In conclusion, the first study successfully established robust QSAR models that accurately predicted the pIC_{50} values associated with antiproliferative activities. The study also highlighted the significant impact of flexibility, lipophilicity, and the type of groups substituted on the C2 carbon of the molecules on antiproliferative activities. Molecular docking and dynamics simulations further confirmed the strong affinity of certain ligands with the active site residues of both ER and PR receptors, suggesting their potential as effective drugs against breast cancer. Pharmacokinetics and drug-likeness studies indicated that these ligands could be orally bioavailable and promising drug candidates. The ADME prediction also supported their potential as effective drugs against breast cancer.

In the second study, molecular docking was utilized to investigate the binding mechanism of WA derivatives with the allosteric site of GAC, revealing their similarities with BPTES and CB-839. The best compounds, namely A5, A8, A13, and A18, were identified based on their score energies and were shown to establish stable interactions with key residues of the target. Electrostatic potential analysis indicated their electrophilic characteristics, particularly the carbonyl oxygen atoms, which are vulnerable sites for electrophile attack. Pharmacokinetics prediction highlighted A8 and A13 as having the best ADMET profile, similar to the reference compound CB-839. These findings suggest that the selected compounds could serve as novel allosteric GAC inhibitors for treating TNBC.

In summary, both studies contribute to the understanding of breast cancer treatment by identifying potential drug candidates and elucidating their mechanisms of action. The findings from these studies lay the groundwork for further research and development of effective and targeted therapies for breast cancer.

Appendices

Appendix A. Docking score and interactions between ligands and the active site residues of ER.

Ligand	S-score (kcal/mol)	Bonds between ligands and of the active site residues					
		Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic Distance (Å)	Bond energy (Kcal/mol)
L1	-5.978	N 8	O	LEU346(A)	H-donor	3.39	-0.9
		O 20	NH₂	ARG394(A)	H-acceptor	2.91	-0.9
L2	-6.706	/	/	/	/	/	/
L3	-6.729	/	/	/	/	/	/
L4	-6.594	/	/	/	/	/	/
L5	-5.825	6-ring	CD1	LEU387(A)	pi-H	4.03	-0.6
L6	-5.960	/	/	/	/	/	/
L7	-6.474	/	/	/	/	/	/
L8	-7.132	6-ring	CD1	LEU387(A)	pi-H	4.33	-0.7
L9	-6.045	/	/	/	/	/	/
L10	-6.521	/	/	/	/	/	/
L11	-5.645	/	/	/	/	/	/
L12	-4.990	/	/	/	/	/	/
L13	-6.620	/	/	/	/	/	/
L14	-7.039	6-ring	CD1	LEU 387(A)	pi-H	4.02	-0.6
L15	-6.709	5-ring	CE2	PHE 404(A)	pi-H	3.58	-0.6
L16	-6.963	C 37	OE2	GLU353(A)	H-donor	3.18	-0.9
L17	-6.365	O 31	OE1	GLU353(A)	H-donor	2.81	-3.1
L18	-5.805	/	/	/	/	/	/
L19	-6.264	/	/	/	/	/	/
L20	-5.952	/	/	/	/	/	/
L21	-6.838	6-ring	CD1	LEU525(A)	pi-H	3.8	-0.6
L22	-6.541	6-ring	CD1	LEU525(A)	Pi-H	3.69	-0.7
L23	-6.452	/	/	/	/	/	/
L24	-7.287	/	/	/	/	/	/
L25	-4.049	5-ring	CB	ALA350(A)	Pi-H	4.18	-0.6
L26	-6.786	F 40	NH2	ARG394(A)	H-Acceptor	2.72	-1.0
L27	-4.189	5-ring	CB	ALA350(A)	Pi-H	4.25	-0.7
L28	-7.042	/	/	/	/	/	/

Appendix A. Continued

L29	-6.861	5-ring	CB	LEU346(A)	Pi-H	4.00	-0.8
L30	-6.705	/	/	/	/	/	/
L31	-7.064	F 45	NH2	ARG394(A)	H-Acceptor	2.68	-1.0
		6-ring	CD1	LEU525(A)	Pi-H	3.62	-0.6
L32	-6.710	5-ring	CB	LEU346(A)	Pi-H	3.97	-0.9
L33	-4.259	/	/	/	/	/	/
L34	-7.047	N 29	OG1	THR347(A)	H-acceptor	2.88	-1.0
L35	-7.086	/	/	/	/	/	/
L36	-6.866	/	/	/	/	/	/
L37	-7.307	5-ring	CE2	PHE404(A)	Pi-H	4.15	-0.6
L38	-7.024	6-ring	CD1	LEU387(A)	Pi-H	4.17	-0.6
L39	-7.368	/	/	/	/	/	/
L40	-7.349	N28	OG1	THR347(A)	H-Acceptor	2.90	-1.0
L41	-6.860	N28	CA	THR347(A)	H-acceptor	3.56	-0.7
L42	-6.757	/	/	/	/	/	/
L43	-7.009	/	/	/	/	/	/
L44	-6.164	/	/	/	/	/	/
L45	-7.155	N 28	CA	THR347(A)	H-Acceptor	3.43	-0.9
L46	-6.837	/	/	/	/	/	/
L47	-6.900	/	/	/	/	/	/
L48	-7.408	/	/	/	/	/	/
L49	-6.481	/	/	/	/	/	/
L50	-5.333	/	/	/	/	/	/
L51	-5.409	/	/	/	/	/	/
L52	-5.943	N 27	OG1	THR347(A)	H-acceptor	2.99	-1.3
		F 44	NH$_2$	ARG394(A)	H-acceptor	2.58	-1.1
L53	-6.137	6-ring	CD1	LEU387(A)	Pi-H	4.17	-0.6
L54	-7.001	6-ring	CE2	PHE404(A)	Pi-H	3.83	-0.6
L$_{ref}$	-7.280	O 36	OE1	GLU353(A)	H-donor	2.84	-5.1
Estrdiol		6-ring	CD1	LEU387(A)	pi-H	4.14	-0.6

Appendix B. Docking score and interactions between ligands and the active site residues of PR.

| Ligand | S-score (kcal/mol) | Bond between ligand and of the active site residues | | | | | |
		Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic Distance (Å)	Bond energy (Kcal/mol)
L1	-8.688	N 8	SD	MET801(A)	H-donor	3.96	-2.4
L2	-8.420	/	/	/	/	/	/
L3	-7.752	/	/	/	/	/	/
L4	-7.707	F 32	NH2	ARG766(A)	H-acceptor	2.71	-1.4
		6-ring	CB	MET759(A)	pi-H	4.49	-0.6
L5	-7.712	6-ring	CE2	PHE 778(A)	pi-H	4.24	-0.6
L6	-7.600	/	/	/	/	/	/
L7	-7.900	/	/	/	/	/	/
L8	-8.382	N 9	SD	MET801(A)	H-donor	4.29	-2.1
L9	-8.141	/	/	/	/	/	/
L10	-8.128	N 9	SD	MET756(A)	H-donor	3.26	-1.9
L11	-8.000	N 9	SD	MET801(A)	H-donor	4.15	-2.5
L12	-8.956	/	/	/	/	/	/
L13	-8.646	N 9	SD	MET756(A)	H-donor	3.10	-3.4
L14	-8.216	O 17	SG	CYS891(A)	H-donor	3.61	-1.3
L15	-8.224	/	/	/	/	/	/
L16	-8.551	/	/	/	/	/	/
L17	-8.695	/	/	/	/	/	/
L18	-8.196	C 16	SD	MET801(A)	H-donor	3.78	-1.0
L19	-7.854	Cl 31	NH2	ARG766(A)	H-acceptor	3.18	-0.9
		6-ring	CE2	PHE778(A)	pi-H	3.74	-0.7
L20	-7.964	6-ring	CE2	PHE778(A)	pi-H	3.82	-0.6
L21	-8.753	/	/	/	/	/	/
L22	-9.952	N 9	SD	MET801(A)	H-donor	3.73	-2.7
L23	-9.261	N 9	SD	MET801(A)	H-donor	3.75	-3.3
		5-ring	CE2	PHE778(A)	pi-H	4.47	-0.9
L24	-8.797	N 9	SD	MET801(A)	H-donor	4.19	-1.2
		6-ring	CB	MET759(A)	pi-H	4.39	-0.6
L25	-9.030	/	/	/	/	/	/
L26	-9.053	N 9	SD	MET801(A)	H-donor	4.17	-1.0
L27	-8.575	N 9	SD	MET756(A)	H- donor	3.53	-0.9

L28	-9.188	/	/	/	/	/	/
L29	-8.249	/	/	/	/	/	/
L30	-8.410	/	/	/	/	/	/
L31	-8.784	/	/	/	/	/	/
L32	-9.151	N 9	SD	MET801(A)	H-donor	3.55	-2.8
L33	-8.719	N 9	SD	MET801(A)	H-donor	4.05	-1.0
		O 40	SD	MET759(A)	H-donor	3.28	-2.2
L34	-7.047	N 29	OG1	THR347(A)	H-acceptor	2.88	-1.0
L35	-8.909	N 9	O	LEU718(A)	H-donor	3.44	-0.8
		C 16	SD	MET756(A)	H-donor	3.73	-1.2
L36	-8.335	N 9	O	LEU718(A)	H-donor	2.84	-1.1
		5-ring	CB	LEU718(A)	pi-H	4.72	-0.6
L37	-9.355	/	/	/	/	/	/
L38	-8.915	N 9	O	LEU718(A)	H-donor	3.28	-1.1
		C 16	SD	MET756(A)	H-donor	3.76	-1.2
		F 35	NH$_2$	ARG766(A)	H-acceptor	3.09	-0.7
L39	-8.913	N 9	O	LEU718(A)	H-donor	3.44	-0.8
		C 16	SD	MET756(A)	H-donor	3.72	-1.2
L40	-8.839	N 27	SD	MET801(A)	H-donor	4.06	-0.3
L41	-8.951	N 27	SD	MET756(A)	H-donor	3.56	-0.1
L42	-8.338	N 9	OD1	ASN719(A)	H-donor	3.22	-1.5
		N 28	CA	VAL760(A)	H-acceptor	3.28	-0.8
L43	-8.429	/	/	/	/	/	/
L44	-7.272	N 9	SD	MET801(A)	H-donor	3.76	-3.3
		N 28	SD	MET759(A)	H-donor	3.10	2.4
L45	-8.906	N 9	SD	MET801(A)	H-donor	4.50	-1.4
		N27	CA	MET756(A)	H-acceptor	3.72	-0.6
L46	-7.897	/	/	/	/	/	/
L47	-9.356	N 27	CA	MET756(A)	H-acceptor	3.65	-0.7
		N 28	CA	VAL760(A)	H-acceptor	3.29	-1.1
L48	-8.741	N 9	O	LEU718(A)	H-donor	2.83	-1.5
		5-ring	CB	LEU718(A)	pi-H	4.67	-0.6
L49	-8.478	N 9	SD	MET801(A)	H-donor	3.31	-2.7
		N 27	SG	CYS891(A)	H-donor	3.29	-1.7
		N 28	SD	MET759(A)	H-donor	3.47	0.3

Appendix B. Continued

L50	-8.124	N 9	SD	MET801(A)	H-donor	4.32	-1.9
L51	-7.278	N 9	SD	MET801(A)	H-donor	3.66	-3.4
		N 28	SD	MET759(A)	H-donor	3.00	3.8
L52	-7.766	N 9	SD	MET801(A)	H-donor	3.84	-3.1
		N 27	SD	MET756(A)	H-donor	3.26	0.9
		N 28	SD	MET759(A)	H-donor	3.09	2.4
		F 44	NH2	ARG766(A)	H-acceptor	2.93	-1.1
L53	-6.857	6-ring	CB	MET759(A)	pi-H	4.26	-0.9
L54	-7.961	6-ring	CB	MET759(A)	pi-H	4.38	-0.7
Lref progesterone	-10.601	O38	NE2	GLN725(A)	H-acceptor	3.28	-1.4
		O38	NH2	ARG766(A)	H-acceptor	2.79	-4.2

Appendix C . Docking score, distances bonds, and bond energies of the withangulatin derivatives with GAC receptor (Site 1 between C and D chains).

Compound	S-score (kcal/mol)	Bonds between ligands and of the active site residues					
		Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic distance (Å)	Bond energy (Kcal/mol)
A1	-7.69	O 14	OH	TYR394(D)	H-acceptor	3.09	-0.7
		6-ring	N	PHE322 (C)	pi-H	4.33	-1.4
		5-ring	N	PHE322(D)	pi-H	4.54	-1.5
A2	-7.23	N 15	O	LYS 320(C)	H-donor	3.26	-1.0
		O 30	NZ	LYS320(C)	H-acceptor	3.11	-2.9
		O 63	NZ	LYS320(D)	H-acceptor	3.05	-3.8
		O 100	NZ	LYS398(C)	H-acceptor	3.35	-1.4
A3	-5.33	N 15	O	LYS320(C)	H-donor	4.33	-1.3
A4	-7.72	O 102	ND2	ASN319(C)	H-acceptor	3.09	-3.3
A5	-7.01	6-ring	CB	GLU325(C)	pi-H	3.75	-0.6
A6	-6.02	O 12	OH	TYR394(C)	H-acceptor	2.65	-2.0
		O 28	O	HOH 761(C)	H-acceptor	2.98	-1.3
		6-ring	N	PHE322(D)	pi-H	3.63	-1.7
		6-ring	N	LEU323(D)	pi-H	4.67	-1.0
A7	-7.52	O 14	OH	TYR394(C)	H-acceptor	3.46	-0.7
		N 19	NZ	LYS320(D)	H-acceptor	3.23	-1.2
		O 70	OH	TYR394(D)	H-acceptor	2.81	0.2

		O 11	NZ	LYS320(D)	H-acceptor	2.90	-9.2
A8	-8.69	O 58	CA	LEU321(C)	H-acceptor	3.37	-0.9
		O 67	OH	TYR394(D)	H-acceptor	2.88	-2.0
A9	-8.31	O 9	OH	TYR394(C)	H-acceptor	2.94	-2.7
		O 11	NZ	LYS320(D)	H-acceptor	3.17	-3.0
		O 97	N	LEU323(C)	H-acceptor	3.04	-0.9
A10	-7.07	O 84	N	LEU323(C)	H-acceptor	3.27	-0.5
A11	-7.42	O 94	NZ	LYS320(D)	H-acceptor	2.92	-7.1
		6-ring	CE	LYS398(D)	pi-H	4.18	-1.2
A12	-6.72	6-ring	OH	TYR394(D)	Pi-H	4.30	-0.7
A13	-9.16	O 9	OH	TYR394(C)	H-acceptor	3.06	-0.7
		N 16	CA	LEU321(D)	H-acceptor	3.56	-0.7
A14	-6.24	O 94	NZ	LYS320(D)	H-acceptor	3.17	-3.5
		5-ring	CA	LEU321(C)	Pi-H	4.55	-0.9
A15	-7.93	O 7	OH	TYR394(C)	H-acceptor	3.00	-0.8
		6-ring	N	PHE322(D)	pi-H	4.28	-1.2
		5-ring	OH	TYR394(D)	Pi-H	4.06	-1.4
A16	-7.32	O 9	N	LEU321(C)	H-acceptor	2.74	-1.0
A17	-5.97	6-ring	N	PHE322(D)	pi-H	4.10	-0.6
A18	-8.89	O 7	CE	LYS398(C)	H-acceptor	3.49	-0.7
		O 76	NZ	LYS320(D)	H-acceptor	3.10	-4.6
		6-ring	CA	LEU321(C)	pi-H	4.09	-0.7
		6-ring	N	PHE322(C)	pi-H	4.03	-2.0
A19	-7.14	O 22	N	LEU323(C)	H-acceptor	3.24	-0.7
		O 64	NZ	LYS398(C)	H-acceptor	3.18	-3.4
		6-ring	N	PHE322(C)	pi-H	4.30	-1.4
A20	-7.43	5-ring	CA	LEU321(D)	pi-H	4.16	-0.7
		S 18	O	LEU323(C)	H-donor	4.38	-0.7
A21	-7.38	O 17	NZ	LYS320(D)	H-acceptor	2.73	-2.1
		O 100	NZ	LYS398(C)	H-acceptor	3.06	-4.9
A22	-5.64	6-ring	CD	LYS320(D)	pi-H	4.04	-0.6
A23	-5.53	C 28	6-ring	TYR394(C)	H-pi	4.27	-0.6
A24	-6.57	/	/	/	/	/	/
A25	-6.48	O 17	NH1	ARG317(D)	H-acceptor	2.93	-1.0
		O 82	NZ	LYS320(D)	H-acceptor	2.99	-4.0
		6-ring	NZ	LYS320(D)	pi-cation	3.88	-1.0
A26	-5.90	N 15	O	GLU325(C)	H-donor	3.09	-3.7
		O 31	NZ	LYS398(C)	H-acceptor	3.08	-0.9
Ref CB-839	-8.56	O 7	NZ	LYS398(C)	H-acceptor	2.89	-3.2
		N 9	CE	LYS398(C)	H-acceptor	3.48	-0.9
		N 12	NZ	LYS320(D)	H-acceptor	3.12	-1.1

Appendix C. Continued

	NAW45	O	LEU323(C)	H-donor	3.30	-1.8
Co-	SAY 49	O	LYS320(C)	H-donor	3.71	-0.3
cryst	SAZ 50	O	LYS320(D)	H-donor	3.82	-0.6
alliz	NAS 41	N	PHE322(C)	H-acceptor	2.98	-2.2
ed -	NAU 43	N	LEU323(C)	H-acceptor	3.07	-4.3
BPT	NAV 44	N	LEU323(D)	H-acceptor	3.05	-2.5
ES	SAY 49	CA	LEU321(C)	H-acceptor	3.69	-1.2
	SAY 49	N	PHE322 (C)	H-acceptor	4.30	-1.2
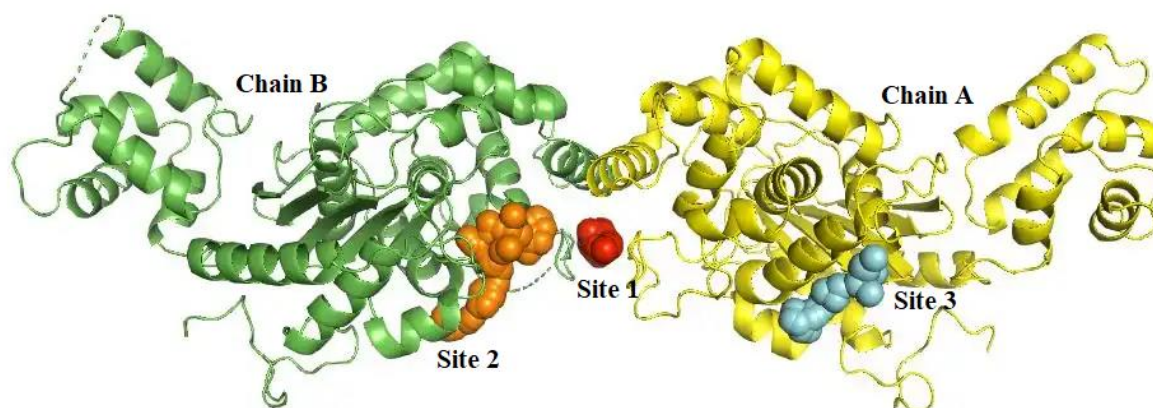
Appendix D. The three highest-ranking pockets of GAC generated by PASSer are depicted in ball-shaped, with colors red (pocket 1), orange (pocket 2) and cyan (pocket 3).

Appendix E. Docking score, distances bonds, and bond energies of the withangulatin derivatives with site 2 of GAC receptor.

Compound	S-score (kcal / mol)	Bonds between ligands and of the active site residues					
		Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic distance (Å)	Bond Energy (Kcal/mol)
A1	-7.10	C 72	6-ring	TYR249(B)	H-pi	4.54	-0.7
		5-ring	CB	ASP467(B)	pi-H	4.22	-0.8
A2	-7.51	6-ring	CB	LYS507(B)	pi-H	3.87	-0.6
A3	-6.58	/	/	/	/	/	/
A4	-6.28	/	/	/	/	/	/
A5	-5.68	O 108	O	HOH8(B)	H-acceptor	3.04	-1.5
		O 108	CA	GLY470(B)	H-acceptor	3.68	-1.2

A6	-5.71	O 27	O	HOH8(B)	H-acceptor	3.16	-1.2
		O 27	CA	GLY470(B)	H-acceptor	4.12	-0.7
A7	-6.46	/	/	/	/	/	/
A8	-6.65	O 67	OD2	ASP467(B)	H-donor	3.45	-2.7
A9	-8.88	O 69	CA	GLY470(B)	H-acceptor	3.58	-1.2
		O 86	N	GLY470(B)	H-acceptor	3.34	-4.1
A10	-8.10	N 12	OD2	ASP467(B)	H-donor	3.29	-1.1
		N 12	OD2	ASP467(B)	H-donor	3.29	-3.5
		O 112	O	HOH8(B)	H-acceptor	3.10	-1.0
A11	-8.41	/	/	/	/	/	/
A12	-6.95	/	/	/	/	/	/
A13	-9.09	O 68	OD2	ASP467(B)	H-donor	3.26	-0.7
		O 68	OD2	ASP467(B)	H-donor	3.26	-3.5
A14	-8.33	O 83	O	HOH667(B)	H-acceptor	2.83	-2.0
A15	-6.87	N 10	OD2	ASP467(B)	H-donor	3.47	-1.5
		N 10	OD2	ASP467(B)	H-donor	3.47	-7.1
A16	-7.71	C 60	6-ring	TYR249(B)	H-pi	4.42	-0.6
A17	-8.42	O 29	CA	GLY470(B)	H-acceptor	3.71	-0.9
		O 98	N	SER314(B)	H-acceptor	3.81	-3.1
A18	-9.01	/	/	/	/	/	/
A19	-8.49	O 22	ND2	ASN335(B)	H-acceptor	3.34	-1.5
		O 55	O	HOH713(B)	H-acceptor	2.98	-1.8
A20	-8.77	/	/	/	/	/	/
A21	-8.52	N 15	OD2	ASP467(B)	H-donor	3.60	-2.1
		O 31	ND2	ASN319(B)	H-acceptor	3.07	-0.8
		O 100	N	LYS 320(B)	H-acceptor	4.07	-0.9
		5-ring	CA	PHE 318(B)	pi-H	4.13	-1.3
A22	-6.88	O 31	NE2	HIS 461(B)	H-acceptor	3.20	-0.7
A23	-7.42	/	/	/	/	/	/
A24	-8.39	N 15	OD2	ASP467(B)	H-donor	3.71	-2.2
		O 93	O	HOH 8(B)	H-acceptor	3.14	-0.7
A25	-7.65	/	/	/	/	/	/

A26	-8.47	/	/	/	/	/	/
Lref CB-839	-8.88	N 10	OD2	ASP467(B)	H-donor	3.61	-1.0
		N 10	OD2	ASP467(B)	H-donor	3.61	-4.5
		6-ring	NE	ARG387(B)	pi-cation	3.71	-1.0
Co-crystallized BPTES	-8.01	NAX 49	O	ASP467(B)	H-donor	3.32	-1.8
		NAS 43	NE2	HIS461(B)	H-acceptor	3.35	-1.5

Appendix F. Docking score, distances bonds, and bond energies of the withangulatin derivatives with site 3 of GAC receptor.

Compound	S-score (kcal/mol)	Bonds between ligands and of the active site residues					
		Atom of ligand	Involved receptor atoms	Involved receptor residues	Type of interaction	Interatomic distance (Å)	Bond Energy (Kcal/mol)
A1	-6.23	O 22	N	ASP541(A)	H-acceptor	2.94	-4.0
A2	-6.06	O 14	N	ASP541(A)	H-acceptor	2.93	-2.6
		O 89	NH1	ARG544(A)	H-acceptor	3.02	-3.1
A3	-6.62	N 15	O	GLU545(A)	H-donor	3.32	-1.9
A4	-6.28	O 74	O	ARG534(A)	H-donor	3.06	-0.8
A5	-6.11	O 71	ND1	HIS535(A)	H-acceptor	3.13	-2.0
		6-ring	CD2	LEU540(A)	pi-H	3.80	-0.7
A6	-6.83	N 14	O	PHE536(A)	H-donor	3.03	-4.7
		6-ring	ND1	HIS535(A)	pi-H	3.92	- 0.7
A7	-6.33	O 87	ND1	HIS535(A)	H-acceptor	3.30	-1.6
		O 98	N	ASP541(A)	H-acceptor	2.99	-2.2
A8	-6.20	S 14	O	ASP541(A)	H-donor	3.50	-0.4
A9	-6.06	N 12	O	PHE536(A)	H-donor	3.07	-1.7
		O 86	NE2	HIS230(A)	H-acceptor	3.38	-1.4
		6-ring	CB	HIS535(A)	pi-H	3.83	-1.3
A10	-5.83	/	/	/	/	/	/
A11	-5.70	/	/	/	/	/	/
A12	-6.19	O 23	O	ARG534(A)	H-donor	2.81	-1.4
		6-ring	CD2	LEU540(A)	pi-H	3.86	-0.7
A13	-6.59	O 11	N	ASP541(A)	H-acceptor	3.17	-1.8
		N 16	N	ASP541(A)	H-acceptor	3.30	-2.0
		5-ring	CD2	LEU540(A)	pi-H	3.92	0.6
A14	-5.99	S12	O	ASP541(A)	H-donor	3.52	-0.7

Appendix F. Continued

A15	-6.42	O 26	OG	SER226(A)	H-acceptor	2.91	-1.5
		O 96	ND1	HIS535(A)	H-acceptor	3.49	-0.8
A16	-6.01	O 38	OG	SER226(A)	H-acceptor	3.05	-1.1
		O 70	NE2	HIS230(A)	H-acceptor	2.92	-3.2
A17	-6.06	O 29	OG	SER226(A)	H-acceptor	2.91	-1.5
		O 98	NH1	ARG217(A)	H-acceptor	2.94	-0.9
A18	-6.16	O 67	ND1	HIS535(A)	H-acceptor	2.97	-3.6
		O 93	NE	ARG534(A)	H-acceptor	2.97	-1.1
A19	-5.93	O 55	NE2	HIS230(A)	H-acceptor	2.88	-4.6
A20	-5.93	/	/	/	/	/	/
A21	-5.83	/	/	/	/	/	/
A22	-6.08	/	/	/	/	/	/
A23	-6.25	/	/	/	/	/	/
A24	-5.98	/	/	/	/	/	/
A25	-5.71	O 99	NH1	ARG544(A)	H-acceptor	3.10	-4.7
		6-ring	CD2	LEU540(A)	pi-H	3.88	-0.8
A26	-6.09	/	/	/	/	/	/
CB-839	-5.89	S 1	O	ARG534(A)	H-donor	3.71	-2.2
		S 1	O	PHE536(A)	H-donor	4.36	-0.7
		N 13	O	ARG534(A)	H-donor	2.96	-6.6
		6-ring	CG	ARG534(A)	pi-H	4.01	-1.1
Co-crystallized BPTES	-6.14	N 49	O	ARG 534(A)	H-donor	2.90	-2.0
		6-ring	OG	SER226(A)	pi-H	3.44	-0.6
		5-ring	CA	GLY546(A)	pi-H	4.48	-1.0

QSAR Study, Molecular Docking/Dynamics Simulations and ADME Prediction of 2-Phenyl-1*H*-Indole Derivatives as Potential Breast Cancer Inhibitors

Khadijah Saghiri [1] , Ismail Daoud [2,3,*] , Nadjib Melkemi [1] , Fouzia Mesli [3]

[1] Laboratory of Molecular Chemistry and Environment (LMCE), Department of Chemistry, University Mohamed Khider, Biskra, Algeria; khadijah.saghiri@univ-biskra.dz (K.S.), n.melkemi@univ-biskra.dz (N.M.)

[2] University Mohamed Khider, Department of Matter Sciences, BP 145 RP, (07000) Biskra, Algeria; i.daoud@univ-biskra.dz (I.D.)

[3] Laboratory of Natural Substances and Bioactive (LASNABIO), University of Abou- BakrBelkaid, 13000, Tlemcen, Algeria; meslifouzia2018@gmail.com (F.M.)

* Correspondence: i.daoud@univ-biskra.dz (I.D.);

Abstract: In the current investigation, QSAR studies have been performed on fifty-four analogs of the 2-phenyl-1*H*-indole analogs, which have antiproliferative activity against two cancer cells lines MDA MB231 and MCF-7.Partial least square (PLS) regression exhibited a good correlation between fitted and observed biological activities. The descriptive and predictive performances of QSAR models have been assessed and validated through internal and external validations. The leave-one-out cross-validation R^2_{CV}, the bootstrapping correlation coefficients R^2_{boot} and the predicted R^2_{pred} reveal a high predictive power for both models developed. Molecular docking/dynamics simulations results show that the ligands L39, L40, and L48 fit well in the pocket of the estrogen-α (PDB:1A52) and ligand L47 with the progesterone (PDB:1A28) confirmed by the high negative score values and by the several interactions (H-bonds and hydrophobic interactions) which established between these ligands and the active site residues of both receptors. Finally, Drug-likeness and ADME prediction have revealed that the ligands L39, L40, and L48 complied with Lipinsky, Veber, and Egan rules that presented a good absorption and oral bioavailability, except the ligand L47. Clearly, this finding shows that these ligands could be used as potential precursors to developing breast cancer drugs.

Data Article

Molecular docking/dynamics simulations, MEP analysis, and pharmacokinetics prediction of some withangulatin A derivatives as allosteric glutaminase C inhibitors in breast cancer

Khadijah Saghiri [a], Ismail Daoud [b c] 🗹, Nadjib Melkemi [a], Fouzia Mesli [c]

Show more ⌄

+ Add to Mendeley   ⛬ Share   � Cite

Get rights and content ↗

Abstract

The current study aims to understand the binding mechanism and intermolecular interactions of the novel Withangulatin A derivatives towards the allosteric site of Glutaminase C (GAC) using in silico analysis. The molecular docking and dynamics simulation results revealed that compounds L5, L8, L13, and L18 show high affinity toward the allosteric pocket of the GAC (PDB:3UO9), as confirmed by the high negative score values and hydrogen bonds. We found that these compounds interact with the