ROYAUME DU MAROC

Université Mohammed V - Rabat

Faculté de Médecine et de Pharmacie

Year 2021

N°:MM044202

MASTER'S THESIS

MASTER of « Biotechnologie médicale »

OPTION: « biomédicale »

titled

Virtual screening by molecular docking of biochemical molecules from SANCDB database against breast cancer

supported by :

Laila Rahel

Before the jury composed of:

Pr.Anniz Tarik, Faculty of Medicine and Pharmacy of Rabat, PresidentDr.

Charoute Hicham, Institut Pasteur du Maroc, Supervisor

Pr Kartti Souad, Faculty of Medicine and Pharmacy Rabat, examiner

FOREWORD

This work is carried out within the framework of the Medical Biotechnology Master, coordinated by Professor Mouna Ouadghiri. under the direction and supervision of Professor Hicham Charout, bio-informatician at the Institut Pasteur's Laboratory for Genomics and Human Genetics.

Dedication

I dedicate this work:

to the soul of my father whom I have wished he would be with me in these times, to my mother for her sacrifices, love, tenderness and prayer throughout my studies. This work represents the culmination of your support and encouragement. May Allah grant you long life by my side to taste the fruit of this work.

To my dear brothers

Words fail me to express to you my attachment, love and affection for you. I implore Allah to give you a better future.

To my dear friends for their support and encouragement.

To all my comrades, for all the pleasant and unforgettable moments we have spent together, and For your precious advice and your help. Your knowledge was an honor for me.

To all the people I have omitted to mention, you find here the expression of all my affection and gratitude

Acknowledgement

First of all, I thank Almighty God for giving me the power and the will to complete this work.

I extend my sincere thanks and gratitude to **Professor Azeddine Ibrahimi** , Director of the Medical Biotechnology Laboratory, for his support, availability, efforts and encouragement, and for the time he was able to devote to training our promotion. I extend my heartfelt thanks to Professor **Mouna Ouadghiri** , Associate of the Master of Medical Biotechnology. at the IPM Laboratory of Genomics and Human Genetics, first to open the doors of the laboratory in which this work was carried out, and for its guidance, valuable advice and availability throughout this work. I would also like to thank **Pr. Anniz Tarik** for accepting being president jury of this work.

I would like to thank my mentor, **Dr. Hicham CHAROUTE**, a bio-informatician at the Research Unit of Epidemiology, Biostatistics and Bioinformatics at Pasteur Institut of Morocco, who has contributed greatly to this work by sharing his vast knowledge with me. Thank you for your support, availability, expert advice, shared knowledge and professionalism. I thank you for the constant attention with which you have progressed my work despite your countless occupations.

I would also like to thank **Dr. Kartti Souad** for accepting the examination of this work.

A big thank also to all the teachers of our Master.

Titre : Dépistage virtuel par arrimage moléculaire de molécules biochimiques de la base de données SANCDB contre le cancer du sein

Nom : Rahel laila

Résumé

Le cancer du sein est l'un des plus grands dilemmes mondiaux et son traitement actuel est de cibler les récepteurs hormonaux par l'utilisation d'agonistes/antagonistes partiels. Les médicaments puissants pour le traitement du cancer du sein sont le tamoxifène, le trastuzumab, le paclitaxel, etc. qui présentent des effets indésirables et une résistance chez les patientes. L'objectif de l'étude est d'identifier les substances biochimiques avec des actions puissantes sur l'inhibition ERα, PR, EGFR et mTOR. L'étude actuelle est réalisée à l'aide d'approches d'amarrage moléculaire, car les interactions protéine-ligand jouent un rôle vital dans la conception des médicaments. Les structures 3D de ERα, PR, EGFR et mTOR ont été obtenues à partir de la banque de données de protéines et ancrées avec 1017 structures 3D ligand de la base de données SANCDB et nous choisissons les cinq meilleurs ligands qui ont un score d'amarrage élevé en utilisant Autodock vina . La propriété de ressemblance de médicament a été vérifiée en appliquant la règle de Lipinski de cinq sur le 1017 pour évaluer l'activité anti-cancer du sein. Les résultats confrme que Octahydroeuclin a le meilleur score d'amarrage pour le cancer du sein suivi par Kaurenoic, Hamiltonine E Neolistine , Epicatechin gallate . Cette étude suggère que les molécules sélectionnées peuvent être étudiées et évaluées davantage pour le traitement du cancer du sein et les stratégies de gestion.

Mots clés : cancer du sein, bioinformatique, amarrage moléculaire, SANCDB, Autodock vina

Tittle: Virtual screening by molecular docking of biochemicalmolecules from SANCDB database against breast cancer

Name : Rahel Laila

Abstract

Breast cancer is one of the biggest global dilemmas and its current therapy is to target the hormone receptors by the use of partial agonists/antagonists. Potent drugs for breast cancer treatment are Tamoxifen, Trastuzumab, Paclitaxel, etc. which show adverse efects and resistance in patients. The aim of the study has to identify biochemicals with potent actions on ERα, PR, EGFR and mTOR inhibition. The current study is performed using molecular docking approaches as protein-ligand interactions play a vital role in drug design. The 3D structures of ERα, PR, EGFR and mTOR were obtained from the protein data bank and docked with 1017 ligand 3D structures from SANCDB database and we choose the top five ligands that have a high docking score using Autodock vina . Drug-likeness property was checked by applying the Lipinski's rule of five on the 1017 to evaluate anti-breast cancer activity. The results confrm that Octahydroeuclin has the best docking score for breast cancer followed by Kaurenoic, Hamiltonin E Neolistine , Epicatechin gallate . This study suggests that the selected molecules can be further investigated and evaluated for breast cancer treatment and management strategies.

Keywords: breast cancer, bioinformatics, molecular docking, SANCDB, Autodock vina.

العنوان : فحص افتراضي عن طريق الإرساء الجزيئي للجزيئات الكيميائية الحيوية من قاعدة بيانات س.أ.ن.س.د.ب لمكافحة سرطان الثدي

الاسم : الراحل ليلى

ملخص

سرطان الثدي هو واحد من أكبر المعضلات العالمية وعلاجه الحالي هو استهداف مستقبلات الهرمونات ، Tamoxifen عن طريق استخدام جزئيين/عداء. الأدوية الفعالة لعلاج سرطان الثدي هي ، وما إلى ذلك ، والتي تظهر آفات معاكسة ومقاومة في المرضى. Paclitaxel، Trastuzumab والهدف من هذه الدراسة هو تحديد الكيماويات الحيوية التي لها إجراءات فعالة في مجال التخفيف من حدة الانبعاثات ، وتخفيف حدة الفقر ، والحد من الفقر ، والحد من الفقر. وتجري الدراسة الحالية باستخدام نهج رسو الجزيئات حيث تلعب تفاعلات البروتين والليغاند دورا حيويا في تصميم العقاقير. وقد تم الحصول من مصرف بيانات البروتين ورسومها بـ mTOR و EGFR و PR و EREA لـ Dعلى الهياكل 3 ونحن نختار أعلى خمس سلاسل لها درجة عالية من SANCDB من قاعدة البيانات 1017D هيكل 3 وتم التحقق من الممتلكات الشبيهة بالعقاقير بتطبيق قاعدة الخمس .Autodock vina الإرساء باستخدام التي وضعها ليبينسكي على 1017 لتقييم نشاط مكافحة سرطان الثدي. النتائج تشير إلى أن Hamiltonin E، Kaurenoic أوكتاهيدروكلين لديه أفضل نتيجة لرسو السفن لسرطان الثدي تليها وتشير هذه الدراسة إلى أنه يمكن مواصلة دراسة وتقييم .Epicatechin gallate، Neolistine الجزيئات المختارة من أجل استراتيجيات علاج وإدارة سرطان الثدي.

الكلمات الرئيسية: سرطان الثدي ، المعلوماتية الحيوية ، رسو الجزيئات ، ، أوتودوك فينا,

Abbreviation list

BC: Breast cancer

SANCDB : South African natural compound database

PR: progesterone receptor

ER: Estrogen Receptor

mTOR : mammalian target of rapamycin

EGFR: Epidermal growth factor Receptor

ERE: estrogen response element

DNA: deoxyribonucleic acid

TGFα : Transforming growth factor alpha

HB-EGF : heparin-binding EGF-like growth factor

SDS-PAGE: sodium dodecyl sulphate–polyacrylamide **gel** electrophoresis

TNBC: Triple-negative breast cance

AFs: activation function domains

TORC1: target of rapamycin kinase complex 1

TORC2: target of rapamycin kinase complex 2

ADME: absorption, distribution, metabolism, and excretion

PI3K : Phosphoinositide 3-kinases

PDB :Protein Data Bank

HBD: Number of hydrogen bond do nors

HBA: Number of hydrogen bond acceptors

M.W : molecular weight

S6K **:** S6 kinase

List of Table

List of Figures

Table of contents

Introduction

Breast cancer (BC) is the most frequently diagnosed cancer among women worldwide and one of the leading cause of cancer-related deaths (Zardavas et al. 2015). The majority of BCs arise from epithelial cells, in either the ducts or lobules, as the result of genetic and epigenetic alterations, which lead to aberrant growth control and disruption of intracellular signaling. Because of this, BC is considered a heterogeneous disease with multiple sub-types, with cells of distinct origin and function (Masuda et al. 2012).

Overexpression of **Estrogen Receptor α**, **Progesteron Receptor**, **Epidermal Growth Factor Receptor** and **mammalian target of rapamycin**, during Breast Cancer development, has a significant impact. ERs are transcription factors that, upon binding to the estrogen steroid hormone, migrate to the nucleus, driving specific gene expression that intercept cell growth control pathways, Therefore, inhibiting ER α, PR, EGFR and mTOR , particularly ERalpha, through the use of hormone-based therapeutics such as tamoxifen is a primary goal of endocrine therapy in BC (Tryfonidis et al. 2016). The current study is performed by the use of molecular docking to identify inhibitor molecules against ER, PR, EGFR and mTOR. The 3D structures of the four receptors were obtained from the protein data bank and docked with 1017 3D PubChem structures from a **South African natural compound database**, named **SANCDB.**

These breakthroughs are the result of the close collaboration between the Biologist, Chemist and computer scientist. Our work on cancer has been divided into two parts :

A first part : a biological part that consists in understanding the structure of the 4 receptors « ERα, PR, EGFR and mTOR » , their mechanisms of action and their roles in the development of breast cancer

A second part : A computer part that looked in the first part on the download of the 4 proteins from PDB database and the 1017 ligands downloaded from the

database « **SANCDB** » to detect the correct ligand-protein configuration based on molecular docking approaches.

Molecular docking is a computer-based structure-based method widely used in drug discovery. Docking can identify new compounds with therapeutic significance, predict ligand-target interactions at the molecular level, or describe structure-activity relationships (SAR).There is no need to know the chemical structure of other target modifiers in advance. Although it was originally developed to help understand the molecular recognition mechanism between small and large molecules, the use and application of docking in drug discovery has changed dramatically in the past few years.(Pinzi and Rastelli 2019)

Our work can contribute to scientific research in the field of targeted therapy against breast tumor pathology.

I- Estrogene receptor

1- Definition

The estrogen receptor (ER) exists in two forms known as ERα and ERβ. Currently, a clinical role has only been established for ERα. The primary use of ERα in breast cancer is for predicting likely response to hormone treatment. Patients with breast cancers expressing ERα are seven to eight times more likely to benefit from endocrine therapy than ERα-negative patients. For the initial three to five years after primary diagnosis, ERα-positive patients generally have a better outcome than ERα-negative patients. Overall, however, the prognostic value of ERα is relatively weak and only of limited value in the clinically important subgroup of patients with lymph node-negative disease. Further work is required to establish if ERα has a clinical role in breast cancer.(Björnström and Sjöberg 2005)

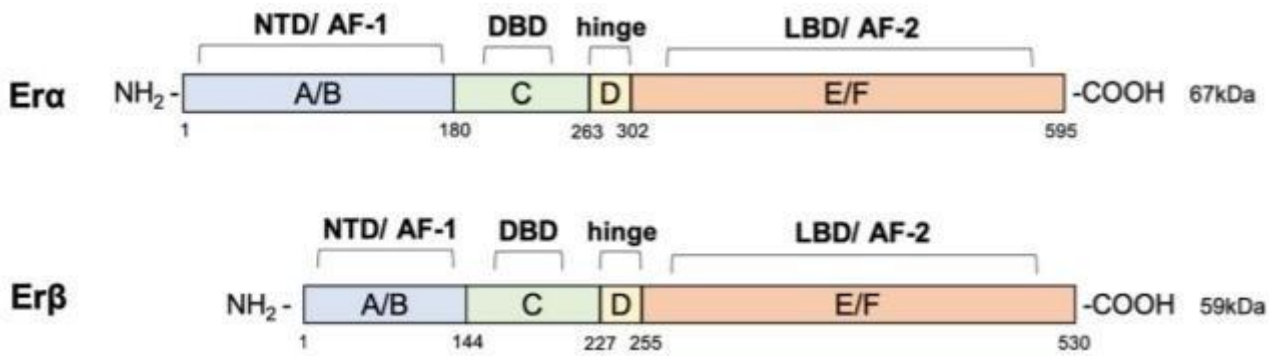


Figure 1: Structural organization of estrogen receptors

2- Mechanism of action :

Estrogen receptors (ERs) act by regulating the transcriptional process. The classic mechanism of ER action involves the binding of estrogen to the receptor in the nucleus, and then the receptor dimerizes and specifically targets the specific estrogen response element (ERE) at the promoter of the target gene. Join to the response element. However, ER can also regulate gene expression without binding directly to DNA. It is caused by

protein-protein interactions with other DNA-binding transcription factors in the nucleus. In addition, membrane-bound ER mediates the non-genomic effects of estrogen. This can lead to both altered protein function in the cytoplasm and regulation of gene expression. The latter two mechanisms of ER action allow the regulation of a wider range of genes than can be regulated by the classical mechanism of ER action alone (Björnströmand Sjöberg, 2005b).

3- Role of Estrogen and ER in Breast cancer formation :

(Saha Roy and Vadlamudi 2012) and (Mondal et al., 2019), believed that the changes in the expression of ERα and ERβ in normal breasts confirmed the direct correlation between ER nuclear expression and the properties of breast proliferation; the relative levels of ERβ and ERα in breast cancer are responsible for cell proliferation and endocrine therapy response Related to the activity of multiple signaling pathways (Madeira et al., 2013). A study showed that under the condition of co-expression of ERα and ERβ, expression is often found to be negative, indicating that there is an association between this special biomarker combination and the aggressiveness of breast cancer. The molecular interaction between ERα and ERβ and its clinical significance in breast cancer are controversial.

II- Epidermal growth factor receptor:

1- Definition :

Epidermal growth factor receptor (EGFR, ErbB1) is a transmembrane protein that exerts tyrosine kinase activity upon ligand induced activation. EGFR can be activated by binding EGF or at least six other structurally related protein ligands, including transforming growth factor α (TGFα), heparin-binding EGF-like growth factor (HB-EGF), betacellulin (BTC), amphiregulin, epiregulin, and epigen. The gene encoding it is localized on human chromosome 7p11.2. Recombinant soluble human EGFR is a 621 amino acid glycoprotein comprising

the extracellular domain of EGFR, and migrates at an apparent MW of 97.5 kDa by SDS-PAGE analysis under reducing conditions.(Ferguson 2008)

2- Mechanism of action :

In Figure an overall model is presented that combines structural information for EGFR on the outside and on the inside of the membrane. EGFR is shown with its extracellular region in the tethered configuration and its kinase domain in the inactive form. Ligand binding to the extracellular region induces receptor-mediated dimerization that brings the intracellular domains into close proximity, and promotes the association of the kinase domains in an asymmetric dimer. In the asymmetric EGFR-TK dimer, one molecule is activated through interaction of its N-lobe with the C-lobe of the cyclin-like activator (shown in the inactive conformation). It is thought that the activated kinase phosphorylates the Cterminal tail of the activator (cyclin-like) receptor. In a subsequent step ,it is proposed that the roles of the two receptors switch, such that both intracellular domains can become trans-autophosphorylated. There is structural information for all but the most C-terminal half dozen amino acids of the extracellular region, which link to the presumably helical Transmembranaire domain. The TM helices of ErbB receptors self- and hetero-associate in membranes (Mendrola et al. 2002), The **ErbB** family of proteins contains four receptor tyrosine kinases, structurally related to the **epidermal growth factor receptor** (**EGFR**) . Although TM interactions of this sort may aid in stabilizing the dimer, or in orienting its components, mutations that disrupt TM domain association do not influence receptor signaling (Biotechnology and Sabatini 1988). On the intracellular side of the membrane, several key pieces of information remain missing—as implied in (**Figure 2**). There is no reliable structural information for the first ~30 amino acids of the intracellular JM region. By analogy with other RTKs, this region may play a regulatory role—possibly contributing to autoinhibition (Hubbard 2004)

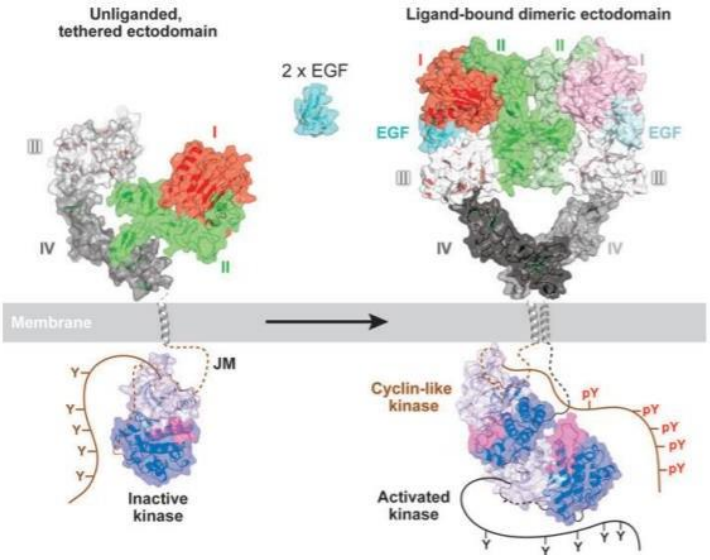


Figure 2: Mechanism of EGFR activation. The crystal structures from EGFR are placed so as to provide a framework to consider the mechanism **of** activation of EGFR at the cell membrane

3- role of EGFR in breast cancer developement :

(Engelman 2009)The human EGFR family contains 4 closely related receptors, which are transmembrane glycoproteins, including extracellular ligand binding domains and intracellular receptor tyrosine kinase domains. Important signaling pathways stimulated by EGFR receptors are PI3 kinase, Ras-Raf-MAPK, JNK, and PLCc (they are protein chains in cells that transmit signals from cell surface receptors to DNA in the nucleus) and cause excessive The biological function (Downward et al., 1984). At the cellular level, ligands not only induce cell proliferation, but also change adhesion and movement and prevent cell apoptosis; at the physiological level, ligands promote invasion and angiogenesis (Eccles, 2011Activation of members of the EGFR family promotes scattering and infiltration of breast epithelial cells in 3D cultures. This is a culture environment that allows cells to grow and interact with the surrounding extracellular framework in three dimensions, such as loss of cell polarity. Characteristics of epithelial differentiation (Wang et al., 1998). In vitro, any of these effects can lead to a malignant phenotype. Disregulation of the EGFR pathway through overexpression or constitutive activation can promote tumor processes, including angiogenesis and metastasis, and is associated with the poor prognosis of many human malignancies (Martinazzi et al., 1993). In addition to cross-interactions between members of the EGFR family, there is evidence of significant interactions between members of the EGFR family and other receptor tyrosine kinases such as c-MET and IGF-1R, and it is possible that such alternative signaling pathways are associated with resistance to EGFR-targeted

therapies (Jin and Esteva, 2008). Expression of both EGFR and HER2 has been reported to be inversely correlated with estrogen receptor (ER) status, and EGFR-HER2 heterodimers have been shown to increase the metastatic potential of breast cancer cell lines (Masuda et al., 2012). The overexpression rate of EGFR is high in TNBC. TNBC is a type of breast cancer that does not have any common receptors. The negative effect of EGFR overexpression is particularly obvious in TNBC. Therefore, EGFR has the potential as a therapeutic target for TNBC, and there is currently no specific targeted therapy. One of the mechanisms of EGFR overexpression is the amplification of the EGFR gene, which has been found in oligodendroglioma (Fallon et al., 2004), glioblastoma, gastric cancer, and breast cancer (Reis-Filho et al., 2005). describe. Overall, EGFR gene amplification is not common in breast cancer: previous studies have shown that EGFR gene amplification is present in 0.8-14% of tumors (Nishikawa et al., 1994). However, gene amplification has been shown in *25% of metaplastic breast cancer cases, which is a specific phenotype of TNBC (Yatabe et al., 2005). In breast cancer, as previously done in lung cancer (with an inframe deletion of exon 19 and a point mutation of exon 21) (Lynch et al., 2004), using the identification of the EGFR mutation, Responds to EGFR-targeted treatments. In breast cancer, EGFR expression levels or genetic mutation status are increasingly being used to select patients for specific treatments. However, it has not yet been proven whether EGFR is really a

predictive biomarker.

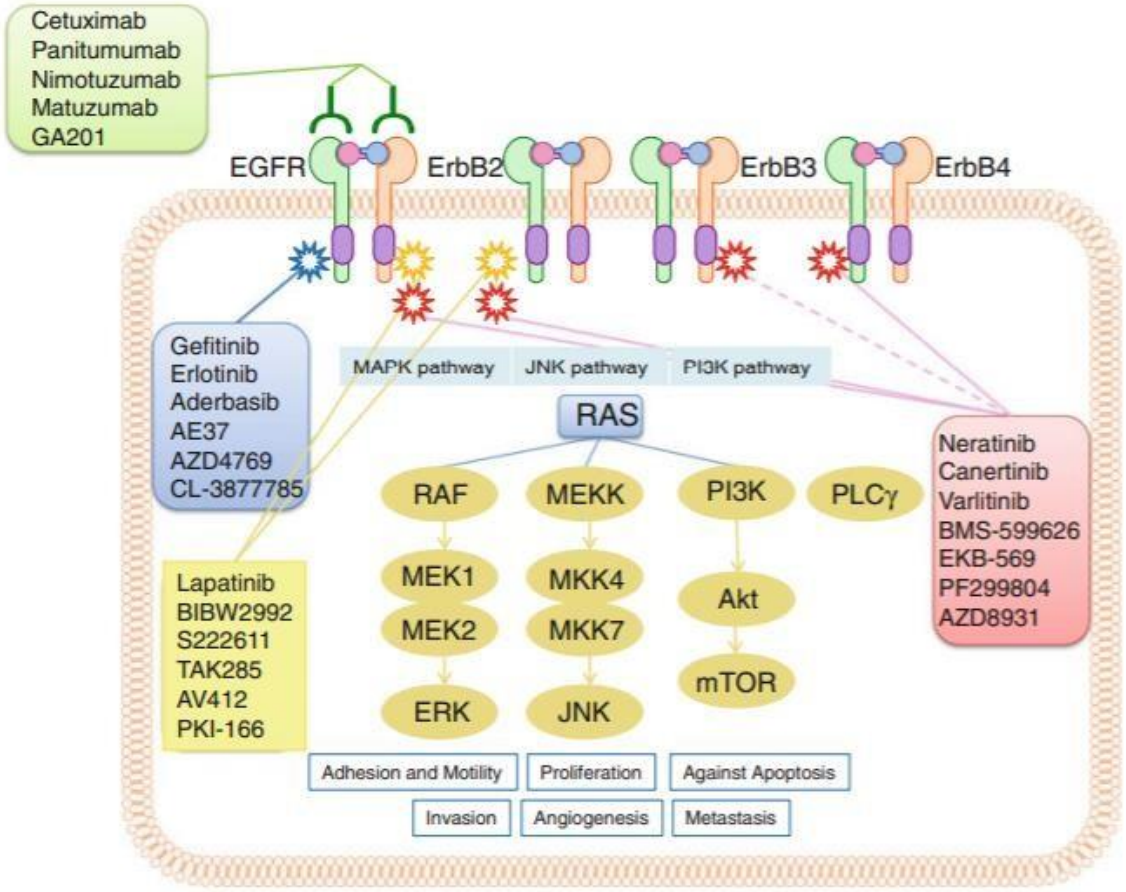


Figure 3: The important signaling pathways stimulated by EGFR receptors

III- Progesterone Receptor :

1-definition :

The progesterone receptor. (PR) is a member of the nuclear receptor superfamily, which

specifically regulates the expression of target genes in response to the hormonal stimulus. The

PR exists as two isoforms in most rodents and humans, PR-A and PR-B, which are produced

from a single gene by translation initiation at two distinct start codons under the control of

separate promoters. The difference between PR-A and PR-B is that PR-A is a truncated form

of PR- B. (Richer *et al.*, 2002). In humans, the N-terminal 164 amino acids of PR-B are

missing in isoform PR-A. Molecular dissection has identified two distinct activation function

domains (AFs) within both PRs: AF-1, which is located in the N-terminal region, is ligand

independent; AF-2, which is ligand dependent, is contained in the ligand-binding domain that

is located in the C-terminal region.(Gao and Nawaz, 2002)

2- Mechanism of action:

In response to the binding of its related steroid hormone progesterone, PR regulates the expression of gene networks. In order to control the development, differentiation and proliferation of target tissues as well as pathological processes in endocrine cancers, PR in the absence of ligands is largely restricted in its interaction with chromatin. The binding of hormones leads to a nuclear localization in which the receptor attacks the genome embedded in chromatin to bind specific DNA recognition sequences. DNA-linked receptor recruits transcriptional corepressor complexes through protein-protein interactions, which subsequently modify the chromatin environment to activate or suppress the target genes.(Cui et al. 2005)

3-progesterone receptor and Breast cancer progression

In breast cancer cell models and clinical studies, the ratio of PR-A to PR-B is a critical determinant of a biological or physiological response to progesterone (Mote, Graham, and Clarke 2007). In normal tissues, PR-A and PR-B are usually found in the ratio 1: 1. However, imbalanced expression of PR-A and PR-B is

observed in the normal breasts of high-risk women breast cancer develops, while altered ratios in breast tumors are associated with endocrine resistance (Pa et al. 1999). Differential signaling and transcriptional activities of isoforms, as well as altered ability of PR-A to trans-repress other steroid receptors, are likely to contribute to breast pathology(Abdel-Hafiz et al. 2002).

IV- mTOR :

1- definition :

mTOR is a serine/threonine kinase composed of the following substances two different protein complexes: the multiprotein complex (TORC1) sensitive to rapamycin and nutrition and the complex (TORC2) which is sensitive to growth factors but not sensitive to nutrition and rapamycin. TORC1 responds to amino acids, stress, oxygen, energy requirements and growth factors, and promotes cell growth and cell cycle progression. TORC2 responds to growth factors and regulates cell survival and metabolism, as well as the cytoskeleton.(Hare and Harvey 2017)

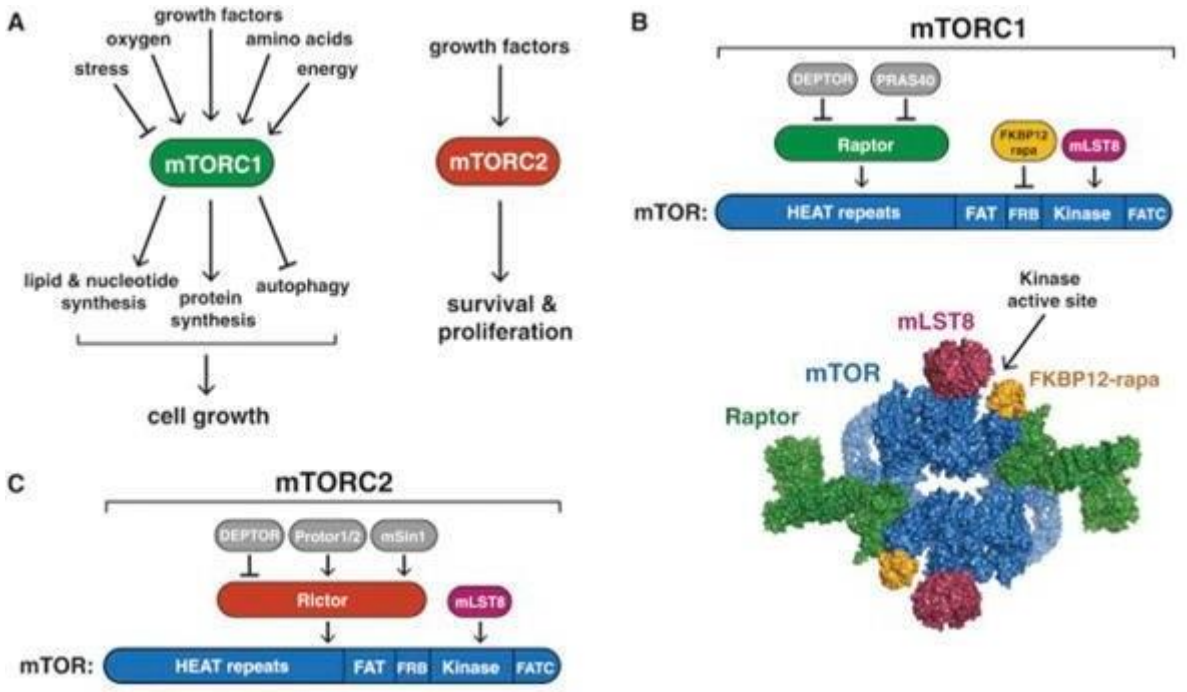


Figure 4: **mTORC1 et mTORC2** (Saxton and Sabatini 2017)

(A) Les voies de signalisation mTORC1 et mTORC2.

(B) sous-unités mTORC1 et sites de **liaison** respectifs sur mTOR. La FKBP12-rapamycine. La structure cryo-EM de 5,9 Â de mTORC1 (sans DEPTOR et PRAS40, PDB ID : 5FLC) de est représentée comme un modèle de remplissage d'espace et colorée par sous-unité.

(C) sous-unités mTORC2 et sites de liaison respectifs sur mTOR.

2- mTOR mechanism:

mTOR regulates a variety of cell activities through mTORC1 and mTORC2. mTOR can respond to extracellular stimuli, such as cytokines, growth factors, and DNA stress. It mainly regulates cell growth, cell cycle and other physiological activities through the PI3K/Akt/mTOR pathway.mTORC1 is activated by the PI3K / AKT pathway and is inhibited by the TSC1 / TSC2 complex (tumor suppressor gene). It is a major regulator of ribosome biosynthesis and protein synthesis through phosphorylation and activation of S6K **(S6 kinase),** as well as phosphorylation and inactivation.

3-mTOR and breast cancer:

The PI3K / AKT / mTOR signaling pathway plays a central role in the cell Physiology through transmission of signal transduction events in response to extracellular stimuli. This pathway controls many cellular functions such as proliferation, growth, survival, motility and metabolism (Engelman, Luo, and Cantley 2006). Mutations in this signaling pathway are often found in cancer, particularly frequently in breast cancer, where around 60% of tumors have genetic changes that hyperactivity the PI3K / AKT / mTOR signaling pathway (Engelman 2009). Preclinical studies have shown that these changes are oncogenic drivers and thus therapeutic goals. Multiple drugs against PI3K, mTOR and AKT are in clinical development.

Materials and Methods:

Molecular modelling is intended to predict the structure and reactivity of molecules or systems of molecules with a precision comparable to the best

results obtained experimental. In addition, protein modelling is the only way to obtain structural information if experimental techniques fail. In the field of molecular modeling, molecular docking, which is also known under the name of molecular docking, is a computer procedure that attempts to predict the conformation (relative position and orientation) of two molecules (a ligand and a receptor) interacting and forming a stable complex. what is considerably easier to implement, che(O'Boyle et al. 2011)aper and faster than using the in vitro experimental methods (Trouillas, 2011).

I - Collection and preparation of targets 3D structures:

1 - Collection of targets 3D structures:

Protein Data Bank (PDB) (https://www.rcsb.org) is a database founded in 1971 by the Brookhaven National Laboratory to archive crystallographic, three-dimensional structures of biological macromolecules (proteins, nucleic acids, etc.). As of July 27, 2021, the PDB contains more than 180,000 structures of which 52732 human sequence structures and 13046 amino acid-containing structures are found. To cope with this vast influx of data, a file format capable of storing standardized and optimized all the structural information available. This is the PDB format "original format of the bank" thus named in reference to the bank from which it originated and whose extension is "pdb". We used this database to download the crystalline structure of 4 proteins "ER, PR, EGFR,mTOR whose PDB identifiers are 3ERT, 4OAR, 2J6M, and 4DRH respectively.

Figure 5 : Portail de la base de données PDB) (https://www.rcsb.org)

2 - Protein 3D structures cleaning using Pymol

Pymol is a software for visualizing and analyzing proteins 3D structures. After downloading the 4 targeted proteins in PDB format, Pymol will allow us to visualize them with their amino acid sequence.We used this software for visualization and preparation of the four receptors for virtual screening.

During this stage of preparation, we removed the molecules from the water that come from crystallography and are incorporated into the crystalline structure, to facilitate calculations and clear the bond pocket (Wong & Lightstone, 2011)

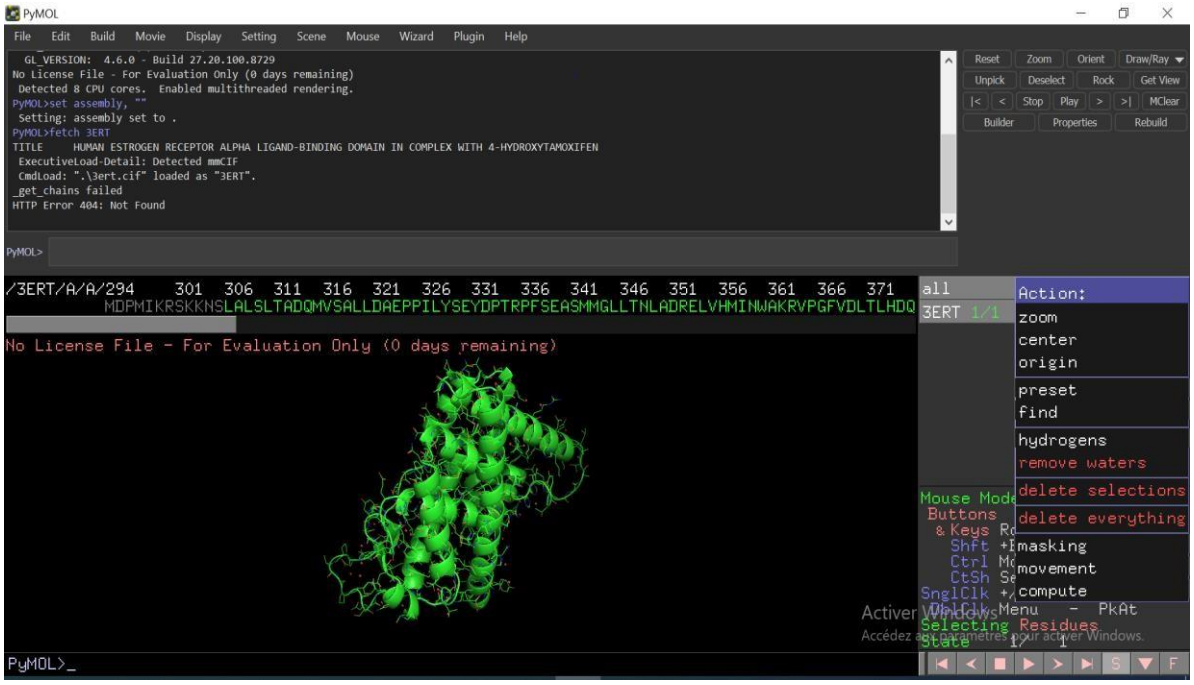


Figure 6: Green ER visualized by Pymol software for preparation

3 - Identification of active sites:

➢ According to PDBsum :

Based on **PDBsum**, ERα, PR, EGFR, and mTOR were selected as targets for breast cancer. Te X-ray crystal structure of ERα and co-crystallized ligand (PDB ID: 3ERT), PR and co-crystallized ligand (PDB ID: 4OAR), EGFR and co-crystallized ligand (PDB ID: 2J6M), and mTOR and co-crystallized ligand (PDB ID: 4DRH) were availed from Protein Data Bank.

⇒ The crystalline structure of ER has made it possible to identify Estrogen binding sites to the extracellular domain of ER. The site of this connection is located mainly on the surface of ER is formed by the amino acids Thr347, Gly521,leu525,Met343,Gly 420

⇒ The crystalline structure of PR has made it possible to identify Progesterone binding sites to the extracellular domain of PR. The site of this connection is located mainly on the surface of PR is formed by the amino acids Met 759,leu 721,leu 718, Asn 719

⇒ The crystalinne structure of EGFR has made it possible to identify EGF binding sites to the extracellular domain of EGF. The site of this

connection is located mainly on the surface of EGF is formed by the amino acids Asp855,Leu488,Ala743,Thr790, Gln791,Met793

⇒ The crystalline structure of mTOR has made it possible to identify binding sites to the extracellular domain of mTOR. The site of connection is located mainly on the surface of PR is formed by the amino acids Gln85,Arg203,Phe 2039, Phe37, Tyr57.

4 - Targets preparation for docking using AutoDockTools:

After the elimination of water with Pymol, we use Autodocktools for the preaparation of our proteins. During this preparation step, we added loads that are missing in each protein and hydrogen atoms (figure 7), the addition of these molecules is essential for the next step which concerns the screening, then we determined the active site of the 4 molecules based on the data base "**PDB sum**" and using '**Grid Box'**, AutoDock requires pre-calculated grid maps, one for each type of atom present in the ligand being docked (figure 8) and (figure 9). This helps to make hosting calculations extremely fast. ... Each point in the grid map stores the potential energy of a 'probe' atom or a functional group due to all the atoms of the macromolecule, After these steps our target is well prepared, so you have to download it and save it as **pdbq**t and a file **'config.txt'** which includes the dimensions of the box is also uploaded.The calculation step takes into consideration the initial state of the ligand (position and orientation), and the dimensions of the grid, this calculation aims at the prediction of the position that the ligand can take in the active site of the target. Molecular docking involves placing the ligand in the active site of the protein and sampling possible conformations, positions and orientations, retaining only those that represent the most favorable.

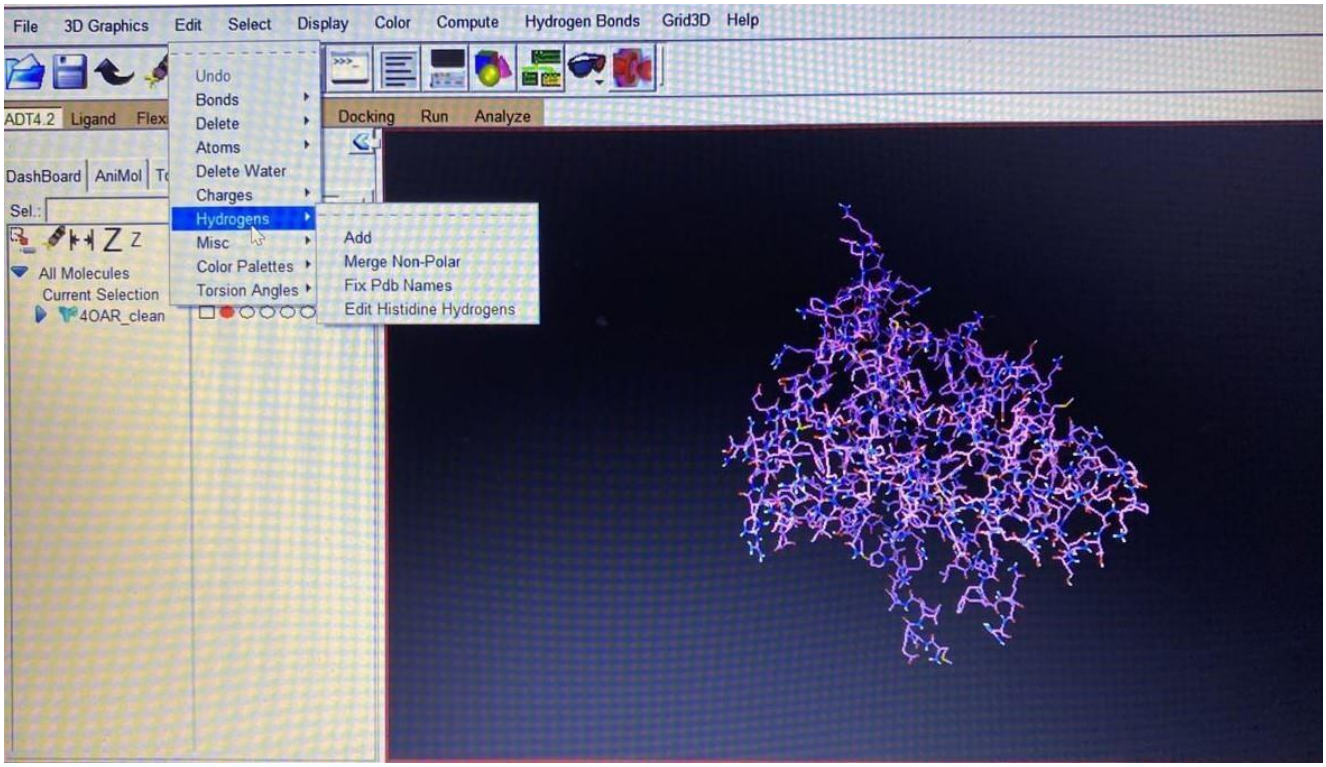


Figure 7 :Loads (Gasteiger loads) and polar hydrogen are added to the cleaned PR structure (PDB ID: 4OAR) using the AutoDock Tools

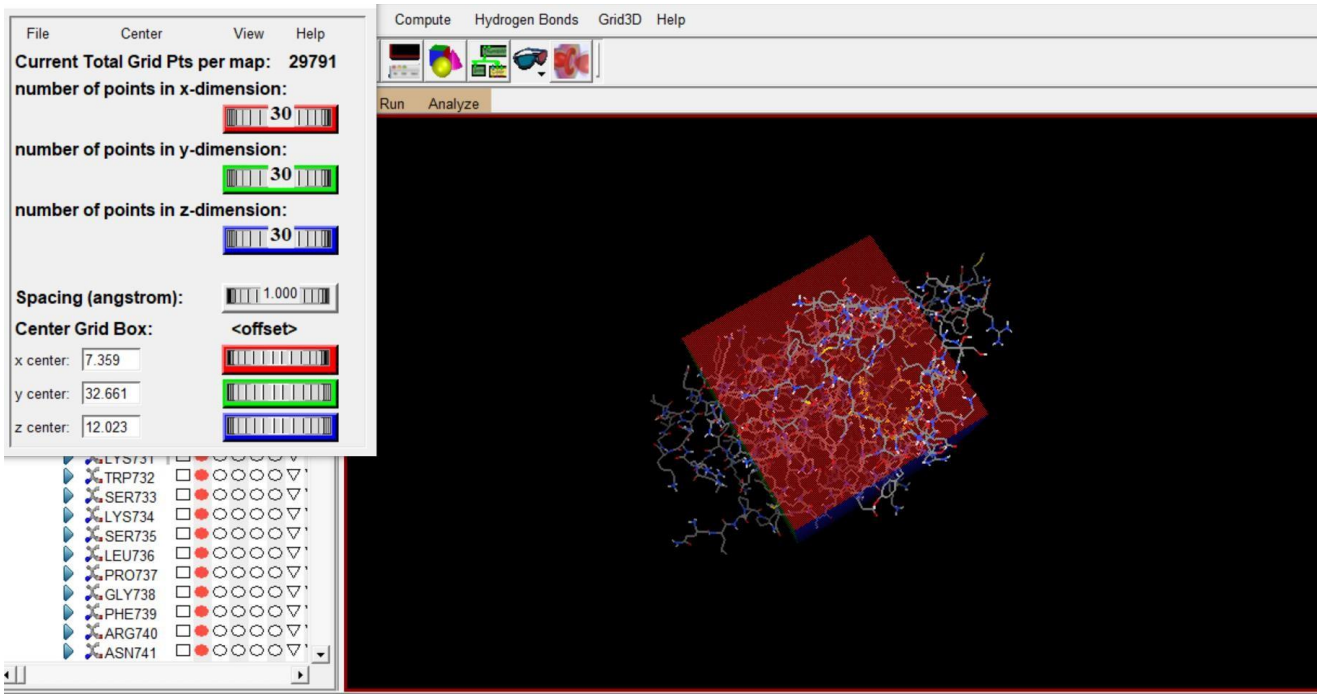


Figure 8: The specific area is defined in the PR receptor (PDB ID: 4OAR) using the Grid Box option in AutoDock Tools

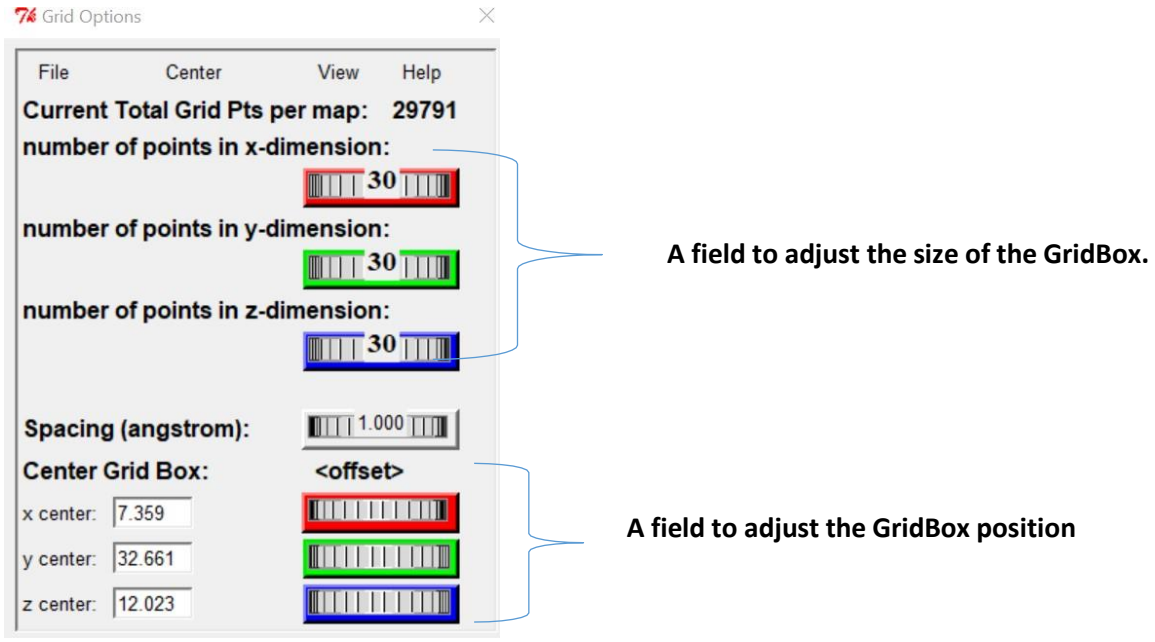


Figure 9 : Grid Box option in AutoDock Tools

5 - Ligands collection :

A number of 1017 molecules were downloaded in PDB format from SANCDB (the database of South African natural compounds) (https://sancdb.rubi.ru.ac.za/).

SANCDB is currently the only web-based NP database in Africa. It aims to provide a useful resource for the in silico screening of South African NPs for drug discovery purposes.This work consists of testing natural products with the aim of designing their anti-cancerous effects, we have used the database SANCDB,this database integrates all the necessary information on compounds such as their chemical structures, Smilles, 3D Structure, biological effects and other information.

6- Ligands preparation using OpenBabel :

As a first step of virtual screening, we must convert the collected files from SANCDB, The collected ligands are of PDB formats, we must convert them to PDBqt Format supported by the program Autodock vina, the conversion is performed by open Babel. Open Babel is a free, open-source software program designed to perform search, analysis, storage, and interconversion between numerous file formats used in molecular modeling, computational chemistry, and related fields. The software provides both ready-to-use programs and a comprehensive and scalable programmer toolkit for chemical software development (O'Boyle et al. 2011)

7- Molecular docking using AutoDock vina :

After the preparation of the four receptors, it is only necessary to proceed with the molecular docking. To do this we used AutoDock Vina which is one of the fastest, most used, and most efficient open source host engines. The software is part of the AutoDock suite, Each position of each ligand in each receptor leads to a possible complex. AutoDock Vina then calculates an approximation of the free enthalpy of formation of this complex through a scoring function (Lipinski et al. 1997). The best score corresponds to the answer given by the software.In vina the orders, for example the location and the name of the file or the software will deposit the result, and the orientations are passed to the command line, The screening consists in creating a program by a LINUX system that will repeat the commands by changing each time the protein. LINUX is a complete, free,

multitasking, multi-platform, multi-user UNIX-like operating system often referred to as the Linux kernel. By extension, Linux commonly refers to the free operating system combining the kernel and a set of system utilities (Techno-Science.net, 2021)

8 - Analysis of Protein-ligand interactions using Discovery studio:

The biovia discovery studio visualizer (https://discover.3ds.com/discovery-studio-visualizer-download) will be used to better understand and visualize the bonding and restored interaction between the target and ligand.

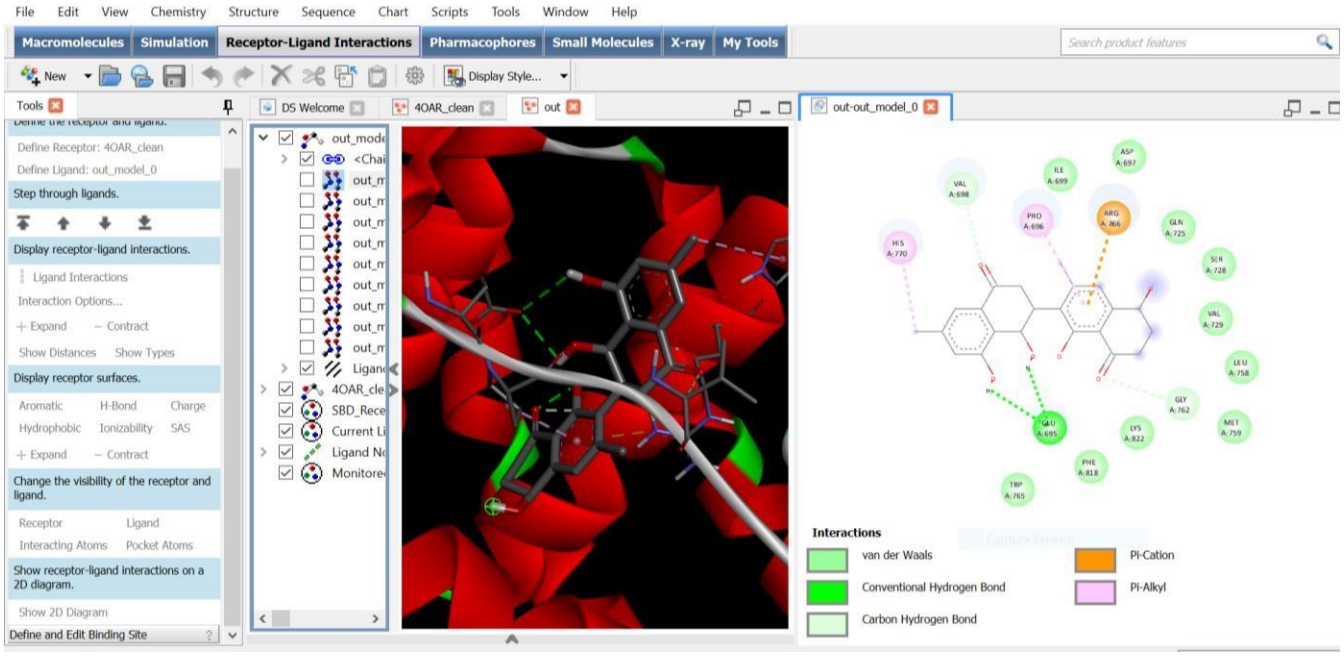


Figure 10 : 2D and 3D interaction between SANC00517 and PR viewed by DISCOVERY STUDIO

Results and discussion :

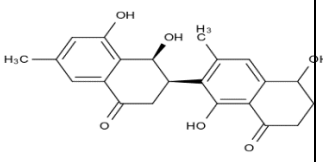
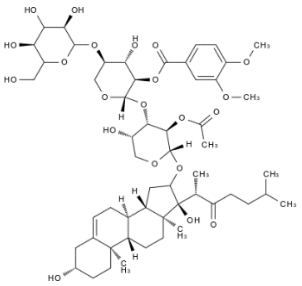
Molecular docking result :

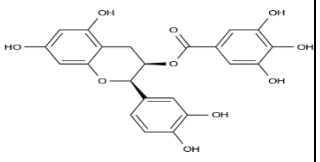
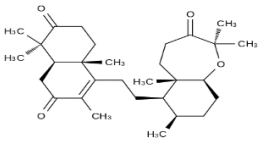
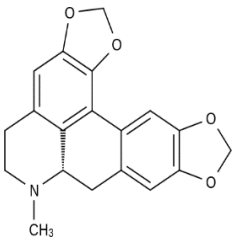
1- Progesterone receptor :

The first five ligands selected against progesteron receptor are Octahydroeuclein with a docking score of -10 kcal/mol,3 β , 17 α- Dihydroxy- 16 β-[(O- β- D-glucopyranos yl-(1→4)- O-(2 -O- 3,4 -dimethoxybenzoyl- β -D-xylopyranos yl) -(1→3)- 2 -O- acetyl- α- L-

arabinopyranos yl) oxy] cholest - 5 - en- 22 - onewith a docking score of 9 , 7 kcal/ mol, Epicatechin gallate with a docking score of - 9 , 9 kcal/ mol, Sodwanone R with a docking score of - 10 , 3 kcal/ mol, and the Neolits ine with a docking score of - 10 , 2 kcal/ mol. (Tableau 1)

Tableau 1 : The first five ligands with the vina docking score (progesterone receptor)

Protein	Ligand code	Ligand code	Structure	Docking score (kcal/mol)
Progesteronreceptor	SANC00517	Octahydroeuclein		-10
	SANC00500	3β, 17 α- Dihydroxy- 16 β- [(O- β- D- glucopyranos yl- (1→4) - O- (2 - O- 3,4- dimethoxybenzoyl		-9,7

		-β-D-xylopyranosyl)-(1→3)-2-O-acetyl-α-L-arabinopyranosyl)oxy]cholest-5-en-22-one		
	SANC00942	Epicatechin gallate		-9,9
	SANC00421	Sodwanone R		-10,3
	SANC01094	Neolitsine		-10,2

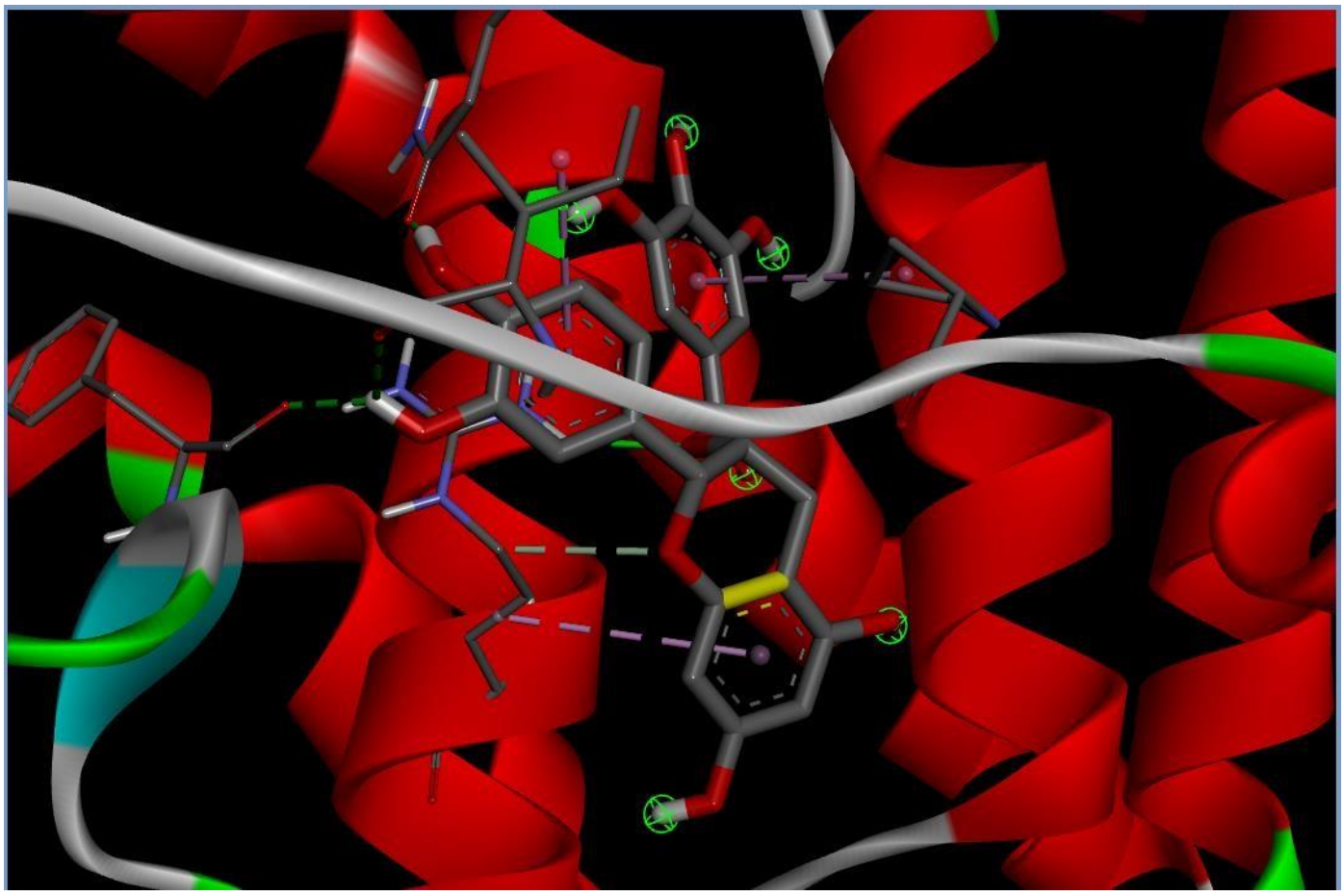


Figure 11: Visualization of PR receptor interaction (in red) with ligand SANC01094 (in grey) pa DISCOVERY STUDIO Visualization of PR receptor interaction (in red) with ligand SANC01094 (in grey) pa DISCOVERY STUDIO

```
mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1        -10.0      0.000      0.000
   2        -10.0      1.764      2.927
   3         -9.6      1.617      3.060
   4         -9.2      1.309      7.566
   5         -9.1      1.883      7.858
   6         -8.8      1.925      3.777
   7         -8.7      1.816      8.002
   8         -8.4      2.089      7.992
   9         -8.3      2.038      8.442
Writing output ... done.
```

Figure 12: the nine possible ligand conformations SANC00507

Octahydroeuclein is extracted from Euclea natalensis , while the source organism of the 3β,17α-Dihydroxy-16β-[(O-β-D-glucopyranosyl-(1→4)-O-(2-O-3,4-dimethoxybenzoyl-β-D-xylopyranosyl)-(1→3)-2-O-acetyl-α-L-arabinopyranosyl)oxy]cholest-5-en-22-one is Ornithogalum saundersiae and Epicatechin gallate is isolated from Ceratonia siliqua, the Sodwanone R is extracted from Axinella weltneri and the Neolitisine is isolated from Cissampelos capensis.

Tableau 2: Classification and source organism of each ligand (PR)

Protein	Ligand code	Ligand name	Source	Classification
	SANC00517	Octahydroeuclein	Euclea natalensis	-Naphthoquinone

Progesteron receptor	SANC00500	3β, 17α- Dihydroxy- 16β -[(O-β-D-glucopyranosyl-(1→4)-O-(2-O-3,4-dimethoxybenzoyl- β- D-xylopyranosyl)-(1→3)-2-O-acetyl-α-L-arabinopyranosyl) oxy] cholest-5-en-22-one	Ornithogalum saundersiae	-Cholestane -Glycoside -Steroids and -steroid derivatives (Classyfire)
	SANC00942	Epicatechin gallate	Ceratonia siliqua	Flavonoids Flavonoids (Classyfire)
	SANC00421	Sodwanone R	Axinella weltneri	□Terpenoid □Triterpene □Phenol lipids (Classyfire)
	SANC01094	Neolitsine	Cissampelos capensis	□Alkaloid □Aporphines (Classyfire)

Major interactions between Octahydroeuclein and PR involve the following active site residues: Gln 725 and Met 759, while interactions between 3β,17α-Dihydroxy-16β-[(O-β-D- glucopyranosyl-(1 4)-O-(2-O-3,4-dimethoxybenzoyl-β-D-xylopyranosyl)-(1 3)-2-O-acetyl-α-L-arabinopyranosyl)oxy]cholest-5-en-22-one and PR involved the following active site residues: Asn719,Gly722,Cys891,Leu718,TRP755-Leu797-Met801.while major

interactions between Epicatechin gallate and PR involved the following active site residues: Arg766,Phe778,Gln725,Leu721-Met759,while major interactions between Sodwanone R and PR Leu721,GLN725.interactions between Neolitsine and PR Leu721 and GLN725 .(table)

Tableau 3: Hydrogenic and hydrophobic interactions, for each ligand with the PR receptor, determined by discovery studio (active site residues are indicated in Bold)

Protein	Code de ligand	Residues involved in hydrogen bonds.	Residues involved in hydrophobic interactions.
Progesterone receptor	SANC00517	ILE699-Asp697-**GLN725**-Ser728-Val729-Leu758-**Met759**-Glu 695-Val 698- Gly 762-Lys822-Phe818-TRP765-Val698	PRO 696- His 770
	SANC00500	**Asn719**-Val760-**Gly722**-Tyr890-**Cys891**-**Leu718**-**Leu721**-Phe778-Val729-Ala779-Tyr777-PRO780-PRO696-Asp697-His770-TRP765-Leu819-Lys822-**TRP755**-Phe818-	**Leu 797**- Met 756- **Met 801**- Val 698

		Leu758-Phe761-Leu887-	
	SANC00942	**Arg766**- ILE 699-**PHE 778**- **GLN725**-His 770- Lys 769-Val 698- PRO 780-ALA 779- **leu 721**-TRP 765- GLN 815-Phe 818- leu 819-Glu 695- Glu 762-Lys 758- Val 729-**Met 759**	PRO 696
	SANC 00421	Ala779-PRO780-**Leu721**-Phe761-**GLN725**-Leu758-Val729-Gly762-Lys822-Glu695-Asp697-GLN815-Lys769-Ser728-His770	TRP765-Phe818-PRO696
	SANC01094	**Arg 766-GLN725-Leu721-Gly722-TRP755-Phe778**-Leu763-Val760-**Met801-Leu797**	Met756-**Met759**

Octahydroeuclein is a compound belonging to the family naphtoquinone, or 1,4-naphtoquinone, is a quinone derived from naphthalene of which many derivatives are pharmacologically active, being generally cytotoxic and possessing antibacterial effects as appropriate, antimycosics, antivirals,

insecticides, anti-inflammatories and/or antipyretics. Based on our results, we find that Octahydroeuclein binds to PR with a relatively high docking score of -10.0kcal/mol. The study by H. Peter T et al., 2021, which consists of carrying out a molecular docking between 712 molecules, only 12 molecules have a strong affinity, in this study Octohydroeuclein had an affinity of -8.3 kcal/mol and one of the most harmful and promising effects was suggested as an antiviral agent against SARS-CoV-2

• Epicatechin gallate (ECG) is a flavan-3-ol, a type of flavonoid, present in green tee, and protects HBMVEC from ischemia/reperfusion lesions by improving apoptosis and autophagy and promoting neovascularization. Based on our results we find that Epicatechin gallate binds to PR with a relatively high docking score of 9.9 kcal/mol. A study by P.Maiti et al., 2020, showed that Epicatechin gallate has an excellent inhibition of kinase EGFR L858R kinase against lung cancer with a high docking score of 9.1 kcal/mol.

• Sodwanone R is a compound belonging to the family Terpenoide and triterpene, being generally cytotoxic and possessing an anticancer action such as Inhibition of activation of inducible factor 1 hypoxia (HIF-1) in tumor cells of the breast and prostate, Based on our results we note that Sodwanone R has a strong affinity to the PR receptor with a docking score of -10.3 kcal/mol. According to the study by Habib et al.,2020, Sodwanone R has a docking score of -8.0 kcal/mol and an EGFR inhibition against lung cancer.

• Neolitsin is a compound belonging to the family Alkaloid aporphines, According to several studies, The cytotoxic potential of aporphine alkaloids against four human cancer cell lines (Hep-2, MCF-7, B16-F10 and 786-0) was also evaluated. According to our results Neolitsine binds to the PR with a high docking score of - 10.3.

•3β,17α-Dihydroxy-16β-[(O-β-D-glucopyranosyl-(14)-O-(2-O-3,4-dimethoxybenzoyl-β-D-xylopyranosyl)-(1 3)-2-O-acetyl-α-L-arabinopyranosyl)oxy]cholest-5-in-22-one is a steroidal glycosial isolated from

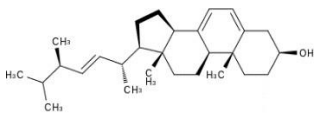
the bulbs of Ornithogalum saundersiae by Kubo et al in 1992.9 several findings have revealed that3β,17α-Dihydroxy-16β-[(O-β-D-glucopyranosyl-(14)-O-(2-O-3,4-dimethoxybenzoyl-β-D-xylopyranosyl)-(1 3)-2-O-acetyl-α-L-arabinopyranosyl)oxy]cholest-5-en-22-one could kill various cancer cells, such as colon cancer cells, hepatocellular carcinoma, leukemia, and so on. According to our studies, this molecule has a high docking score of -9.7 kcal/mol.

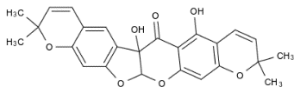
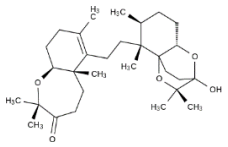
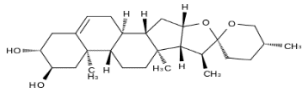
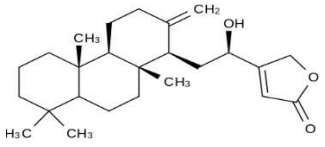
We find that the five molecules have a high docking score against Progesteron receptor. We can then suggest that these molecules can be used to develop new compounds that are effective against breast cancer by blocking the binding site of the PR receptor.

2- Epidermal growth factor receptor:

the first five ligands selected against **epidmal growth factor receptor** are : **Ergosterol** with a docking score of -8.7 kcal / mol, **Kraussianone 4** with a docking score of -8,8kcal/mol, **Sodwanone O** with a docking score of -8,9kcal/mol, **Yuccagenin** with a docking score of - 8 , 8 kcal/ mol, **Hamiltonin E** with a docking score of -8,7 kcal/mol (tableau 4)

Tableau 4: The first five ligands with the vina docking score (EGFR)

Protein	Ligand code	Ligand name	Structure	docking score (kcal/mol)
EGFR	SANC00832	Ergosterol		-8,7

	SANC00347	Kraussianone 4		-8,8
	SANC00419	Sodwanone O		-8,9
	SANC00746	*Yuccagenin*		-8,8
	SANC00189	Hamiltonin E		-8,7

We note that Sodwanone O extracted from Axinella weltneri shows the highest score against EGFR compared to other compounds: Kraussianone 4 which is a molecule belongs to the family Terpenoide and it is extracted from Eriosema kraussianum and Ergosterol which is a compound belongs to the family of steroids and it is isolated from Fusarium proliferatum, and Yuccagenin which is a steroidal compound extracted from Agapanthus africanus and Hamiltonin E which is a compound of the family Hamiltonin and extracted from Chromodoris hamiltoni.

Tableau 5: Classification and source organism of each ligand (EGFR)

Protein	Ligand code	Ligand name	Source	Classification
EGFR	SANC00832	**Ergos terol**	Fusarium proliferatum	□Steroids □Steroids and steroid derivatives (Classyfire)
	SANC00347	**Kraussianone 4**	Eriosema kraussianum	□Pyrano-isoflavone □ Isoflavonoids (Classyfire)
	SANC00419	**Sodwanone O**	Axinella weltneri	□Terpenoid □Triterpene □ Oxepanes (Classyfire)
	SANC00746	**Yuccagenin**	Agapanthus africanus	□Steroidal sapogenin □ Phenol lipids (Classyfire)
	SANC00189	Hamiltonin E	Chromodoris hamiltoni	□Hamiltonin □Sesquiterpene □Terpenoid □ Phenol lipids (Classyfire)

Major interactions between EGFR and Ergosterol involve the following active site residues: THR6790, Met793, Leu844, Asp855. We find that the following active site residues: Asp 855, THR790, Ala 743, Met793, Gly 796, Leu844 are involved in major interactions between Kraussianone 4 and EGFR, while Gly 796, Asp855, Leu844, Met793, Ala743 are involved in the interaction between Sodwanone O and EGFR and Ala743, Gly796, Leu844, PRO 794 are involved between Yuccagenin and EGFR. Hamiltonin E and EGFR contain the following residues from the active site: THR790, Asp855, Ala743, Gly796, Met793, GLN791, Leu844.

Tableau 6: Hydrogenic and hydrophobic interactions, for each ligand with the EGFR receptor, determined by discovery studio (active site residues are indicated in Bold)

Protein	ligand code	Residues involved in hydrogen bonds	Residues involved in hydrophobic interactions
EGFR	SANC00832	**THR790**-Leu788-Met766-Glu762-THR854-Gly719-**Met793**-**Leu844**-Leu792-ILE744-**Asp855**-Phe723	Leu 718- Val 726- ALA 743- Lys 745
	SANC00347	Leu788- Glu762-Met 766-THR854-**Asp855**-**THR790**-**ALA743**-**Met793**-**Gly796**-Asp800-Leu718-**Leu844**-Gly719**.**	Lys797-Arg841

	SANC00419	**Gly796**-**Asp855**-Cys797-Lys745-THR854-**Leu844**-**Met793**-**Ala743**-Leu792- Leu718-PHE 723- Gly 721-Gly719-Asp800	Val 726
	SANC00746	Ala722-Gly719-Leu718-Val726-**Ala743**-Gly721-ARG841-Cys797-Asp800-**Gly796**-**Leu844**-Leu792-**PRO794**	PHE723
	SANC00189	Gly719-Gly721-PHE723-Lys745-**Asp855**-Glu762-**THR790**-**Ala743**-**Gly796**-Leu792-**Met793**-**GLN791**-**leu844**-Cys797-Leu718-THR854.	Val 726

Ergosterol is a steroid compound, according to our results Ergosterol has a high docking score of -8.7 kcal/mol, the latter has antifungal activity.

Kraussianone 4: is classified in the flavonoid family, and has a docking score of -8.8 kcal/mol . According to the study by (Awouafack et al. 2015), Kraussianone 4 has several important antimicrobial, cytotoxicity, anti-mycobacterial, antioxidant, erectile dysfunction, vasodilator and hypoglycemic pharmacological properties.

Sodwanone O is a compound of the family of terpenoid and triterpene, in our results this compound has a docking score of -8.9 kcal/mol.

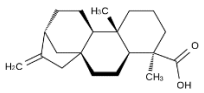
Yuccagenin is a compound of the Steroidal sapogenin family and according to our results it has a significant docking score of -8.8 kcal/mol.

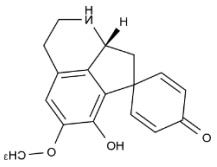
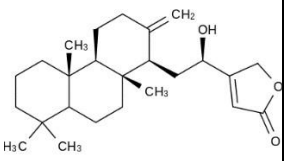
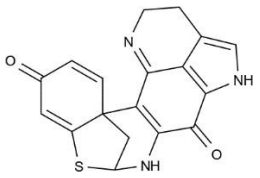
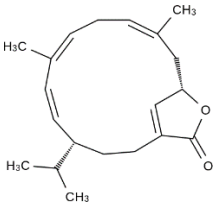
Hamiltonin E is ranked in the Hamiltonin family and has a docking score of -8.7 kcal/mol.

3- Estrogen receptor:

The first five ligands selected against Estrogen receptor are : Kaurenoic acid with a docking score -10,2 kcal/mol, Crotsparine with a docking score -9,3kcal/mol, Hamiltonin E with a docking score -10,0kcal/mol, 11H-8,13a-Methanopyrrolo[4',3',2':4,5]quino[7,8-d][1,3]benzothiazepine-6,11(2H)-dione, 3,5,7,8-tetrahydro-, (8R,13aR) with a docking score -9,3 kcal/mol, (+)-(1R*,10R*)-Cembra-2E,4E,7E,11Z-tetraen-20,10-olide with a docking score -9,4kcal/mol.(table 7)

Tableau 7: The first five ligands with the vina docking score (Estrogen receptor)

Protein	Ligand code	Ligand name	Structure	dockingscore (kcal/mol)
Estrogen Receptor	SANC00898	Kaurenoic acid		-10,2

	SANC00382	Crotsparine		-9,3
	SANC00189	Hamiltonin E		-10,0
	SANC00131	11H-8,13a-Methanopyrrolo[4',3',2':4,5]quino[7,8-d][1,3]benzothiazepine-6,11(2H)-dione, 3,5,7,8-tetrahydro-, (8R,13aR)		-9,3
	SANC00701	(+)-(1R*,10R*)-Cembra-2E,4E,7E,11Z-tetraen-20,10-olide		-9,4

Kaurenoic acid est extrait de Helichrysum kraussii Schefflera umbellifera, tandis que Crotsparine est extrait de Cissampelos capensis, Hamiltonin E est isolé de Chromodoris hamiltoni, 11H-8,13a-Methanopyrrolo[4',3',2':4,5]quino[7,8-d][1,3]benzothiazepine-6,11(2H)-dione,3,5,7,8-tetrahydro-, (8R,13aR) is extracted from Latrunculia lorii, (+)-(1R*,10R*)-Cembra-2E,4E,7E,11Z-tetraen-20,10-olid is isolated Croton gratissimus (Table 8)

Tableau 8: Classification and source organism of each ligand (ER)

Protein	Ligand code	Ligand name	Source	Classification
Estrogen receptor	SANC00898	Kaurenoic acid	Helichrysum kraussii Schefflera umbellifera	Phenol lipids (Classyfire)
	SANC00382	Crotsparine	Cissampelos capensis	Alkaloid Aporphine Proaporphines (Classyfire)
	SANC00189	Hamiltonin E	Chromodoris hamiltoni	Hamiltonin Sesquiterpene Terpenoid Phenol lipids (Classyfire)
	SANC00131	11H-8,13a-Methanopyrrolo[4',3',2':4,5]quino[7,8-d][1,3]benzothiazepine-6,11(2H)-dione, 3,5,7,8-	Latrunculia lorii	Alkaloid Pyrroloiminoquinone Phenanthrolines (Classyfire)

		tetrahydro-, (8R,13aR)		
	SANC00701	(+)-(1R*,10R*)-Cembra-2E,4E,7E,11Z-tetraen-20,10-olide	Croton gratissimus	Cembranolide Phenol lipids (Classyefire)

Major interactions between ER and Kaurenoic acid involved the following active site residues: Leu387, Leu525, TRP347, Glu353, Gly521, Glu419, Leu346.While Crotsparine involved Ala350-Leu525-TRP383, Met343-Gly521, Arg394-Leu387, Met421, Glu353 in major interactions with ER.While the major interactions between Hamiltonin E and ER involved the following active site residues: Leu387-Leu525, Arg394-Glu353-Ala350-PHE 404-Met343-TRP383-Leu428-Met421-Gly521.

11H-8,13a-Methanopyrrolo[4',3',2':4,5]quino[7,8-d][1,3]benzothiazepine-6,11(2H)-dione,3,5,7,8-tetrahydro-, (8R,13aR) involves temporal interactions with ER at: Leu387-Ala350- Leu428-Met421- Gly521-Met343- Leu346-Glu353-PHE404. (+) -(1R*,10R*)-Cembra-2E,4E,7E,11Z-tetraen-20,10-olide and ER involved major interactions at: Met421-TRP383-Leu525- Leu 346--Ala350

Tableau 9: Hydrogenic and hydrophobic interactions, for each ligand with the ER receptor, determined by Disvovery studio (residue from active site indicated in Bold)

Protein	Code ligand	Residus involved in hydrophobic interactions	Residus involved in hydrogenated interactions

Estrogen Receptor	SANC00898	**Leu387**-Met388	Val418-His524-**Leu525**-**TRP383**-**THR347**-Leu384-ILE424-Leu428-**Glu353**-Leu349-**Gly521**-**Glu419**-**Leu346**-Arg394
	SANC00382	**Ala350**-**Leu525**-**TRP383**	**Glu353**-Leu349-**Arg394**-**Leu387**-Leu391-Phe404-**Met421**-ILE424-**Met343**-**Gly521**-THR347
	SANC00189	Leu391-Met388-**Leu387**-**Leu525**-Leu384	**Arg394**-**Glu353**-**Ala350**-**PHE 404**-**Met343**-**TRP383**-**Leu428**-**Met421**-ILE424-**Gly521**-His524
	SANC00131	**Leu387**-**Ala350**-Leu384	**Leu428**-**Met421**-Met388-ILE424-**Gly521**-**Met343**-THR347-**Leu346**-**Glu353**-Leu349-**PHE404**-Leu391
	SANC00701	**Met421**-ILE424-**Leu 346**-His524-leu384-**Ala350**-**TRP383**-**Leu525**	**Arg394**-**Leu387**-**PHE404**-Met388-Leu391-Leu349-**Glu353**-**Leu428**-**Gly420**-**Gly521**-**Met343**

- Kaurenoic acid is a compound of the family Phenol lipids, it is used in traditional medicine as a treatment of pain and as an antidote against poison from poisonous animals (Rosa mariana et al.,2020). We note that in our results this compound has a high docking score of -10.2kcal/mol.

- Crotsparin is composed of the Alkaloid family, the study by BURGOS A et al., shows that Crotsparin has antibacterial activity, and according to our results

this element has a significant docking score of -9.3 kcal/mol. other studies have shown that alkaloids have anticancer activity.(Mondal et al. 2019)

- Hamiltonin E is a new sesterpene belonging to the terpene family, in our results we find that Hamiltonin E has a high docking score of -10.0 kcal/mol.

- 11H-8,13a-Methanopyrrolo[4',3',2':4,5]quino[7,8-d][1,3]benzothiazepine-6,11(2H)-dione, 3,5,7,8-tetrahydro-, (8R,13aR) is a compound of the alkaloid family and our results show a high docking score of -9,3 kcal/mol.

- (+)-(1R*,10R*)-Cembra-2E,4E,7E,11Z-tetraen-20,10-olide is a Cembranolide, according to our results it has a significant docking score of -9.9 kcal/mol. According to (Mfotie Njoya, Eloff, and McGaw 2018) this molecule inhibits the growth of cancer cells by inducing the activation of caspase 3/7 with additional anti-inflammatory and antioxidant activities.

We find that these five molecules have a high docking score and very close against ER, We can suggest then that these molecules can be

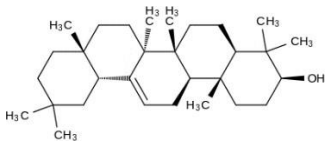
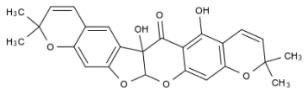
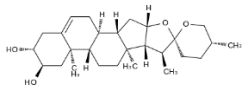
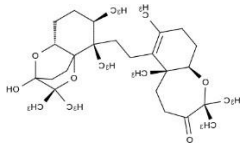
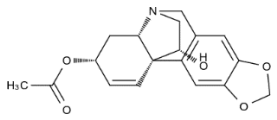
Used to develop new compounds with anti-cancer efficacy by blocking the binding site of the ER receptor.

4- mTOR :

the first five ligands selected against progesterone are: [beta]-amyrin, Kraussianone 4, Yuccagenin, Sodwanone O, 3-O-Acetylhamayne.

Tableau 10 :The first three ligands with the vina docking score (mTOR)

Protein	Code	Name	Structure	Score de docking (kcal/mol)

mTOR	SANC01037	[beta]-amyrin		-7,8
	SANC00347	Kraussianone 4		-7,9
	SANC00746	Yuccagenin		-7,9
	SANC00419	Sodwanone O		-7,8
	SANC00321	3-O-Acetylhamayne		-7,8

[beta]-amyrin est extrait de Croton pseudopulchellus, tandis que Kraussianone 4 est isolé de Eriosema kraussianum, tandis que les organismes Agapanthus africanus, Axinella weltneri et Crinum bulbispermum Crinum moorei sont la source de Yuccagenin, Sodwanone O, 3-O-Acetylhamayne respectively.

Tableau 11: Classification and source organism of each ligand (mTOR)

Protein	Ligand Code	Ligand Name	Source	Classification
mTOR	SANC01037	[beta]-amyrin	Croton pseudopulchellus	Phenollipids Phenol lipids (Classyfire)
	SANC00347	Kraussianone 4	Eriosema kraussianum	Pyrano-isoflavone Isoflavonoids (Classyfire)
	SANC00746	Yuccagenin	Agapanthus africanus	Steroidal sapogenin Phenol lipids (Classyfire)
	SANC00419	Sodwanone O	Axinella weltneri	Terpenoid Triterpene Oxepanes (Classyfire)
	SANC00321	3-O-Acetylhamayne	Crinum bulbispermum Crinum moorei	Alkaloid Amaryllidaceae Amaryllidaceae alkaloids (Classyfire)

The major interactions between mTOR and [beta]-amyrin involved the following active site residues: THR2098 TRP2101-Phe2039, while interactions between Kraussianone 4 and mTor and Yuccagenin and mTOR involved Phe2039.Major interactions between ER and Sodwanone O and between 3-O-Acetylhamayne and ER involved the following active site residues: THR2098-PHE2039.(Table20)

Tableau 12: Hydrogenic and hydrophobic interactions, for each ligand with the receptor mTOR, determined by discovery studio (residues from the active site are shown in Fat)

Protein	Code de ligand	Residues involved in hydrogen bonds.	Residues involved in hydrophobic interactions
mTOR	SANC01037	Arg2024-**THR2098**-**TRP2101**-**Phe2039**-Val2094-Tyr2038-Leu2097-Val2044-Tyr2088	------
	SANC00347	**Phe2039**-TYR2038-TYR2088-Asn2043-Lys2045	Arg2042
	SANC00746	Lys2045-Asn2043-Val2044-Val2094-Leu2097-Tyr2088	**Phe2039**-Arg2042-TYR2038
	SANC00419	Val2044-Leu2097-Tyr2088-Tyr2038-Val2094-**THR2098**-**PHE2039**	Arg2042
	SANC00321	**Phe2039**-Val2094-**THR2098**-Val2044	-----

Beta-amyrin is a pentacyclic triterpenoid that is an oleanan substituted in position 3beta by a hydroxy group and containing a double bond between positions 12 and 13. It is one of the most common triterpenoids in higher plants. It has a role as a plant metabolite and Aspergillus metabolite. It is a pentacyclic triterpenoid and a secondary alcohol.Studies show that it has antimicrobial anticancer activity. In our results, it has a docking score of -7.8 kcal/mol

Kraussianone 4: a pyrano-isoflavone isolated from the roots of Eriosema kraussianum, this compound showed good results, improving blood flow in the uterine artery, and resulting in improved fetal outcomes and decreased antiangiogenic factors in pregnant Sprague-Dawley rats treated with L-NAME.in our results Kraussianone 4 has a docking score of -7.9kcal/mol.

Yuccagenin is a yuccagenin belongs to the class of organic compounds called triterpenoids. These are terpenic molecules containing six isoprene units. Yuccagenin is an extremely weak basic compound, according to our results, it has a docking score of -7.9 kcal/mol.

Sodwanone O is a compound of the tritenes family and inhibits the activation of hypoxia-inducible factor 1 (HIF-1) in breast and prostate tumour cells. In our results Sodwanone expresses a docking score -7.8kcal/mol.

3-O-Acetylhamayne: is composed of the alkaloid family, studies have shown that 3-O-Acetylhamayne has anticancer activity and antibacterial activity, and in our results, 3-O-Acetylhamayne has a docking score of -7.8 kcal/mol.

I- Conclusion and perspectives:

In this report, 1017 compounds, found in South African database traditionally used against several diseases, were selected to determine their interactions with known proteins in the development of breast cancer. Molecular docking revealed that 20 of the 1017 compounds selected had the potential for further investigation as potential therapeutic agents for breast cancer. We noted that Octahydroeuclein has an inhibitory efficacy against cancer Progesterone receptor, the results also revealed that Epicatechin gallate isolated from Ceratonia siliqua, Sodwanone R from Axinella weltneri, and Cissampelos capensis Neolitsin showed high inhibitory efficacy against the PR receptor, as these compounds exhibit strong interactions with PR. Kaurenoic acid, Hamiltonin E, (+)-(1R*,10R*)-Cembra-2E,4E,7E,11Z-tetraen-20,10-olide were highly effective against ER. Our results could then prove useful for developing new natural compounds with an anti-

cancerous effect. In this context, it would be interesting to carry out molecular dynamics simulations to verify the stability of the selected molecular Docking positions. Molecular dynamics will allow access to the most stable conformation from a Docking position by assigning some flexibility to the ligand and the protein.

References :

Abdel-Hafiz, Hany, Glenn S. Takimoto, Lin Tung, and Kathryn B. Horwitz. 2002. “The Inhibitory Function in Human Progesterone Receptor N Termini Binds SUMO-1 Protein to Regulate Autoinhibition and Transrepression.” *The Journal of Biological Chemistry* 277(37): 33950–56.

Awouafack, Maurice Ducret, Pierre Tane, Michael Spiteller, and Jacobus Eloff. 2015. “Eriosema (Fabaceae) Species Represent a Rich Source of Flavonoids with Interesting Pharmacological Activities.” *Natural product communications* 10: 1325–30.

Biotechnology, Rorer, and Communicated David D Sabatini. 1988. “Ligand-Induced Stimulation of Epidermal Growth Factor Receptor Mutants with Altered Transmembrane Regions.” *Proc. Natl. Acad. Sci. USA*: 5.

Björnström, Linda, and Maria Sjöberg. 2005. “Mechanisms of Estrogen Receptor Signaling: Convergence of Genomic and Nongenomic Actions on Target Genes.” *Molecular Endocrinology* 19(4): 833–42. https://academic.oup.com/mend/article/19/4/833/2741274 (August 10, 2021).

Cui, Xiaojiang et al. 2005. “Biology of Progesterone Receptor Loss in Breast Cancer and Its Implications for Endocrine Therapy.” *Journal of Clinical Oncology* 23(30): 7721–35. https://ascopubs.org/doi/full/10.1200/JCO.2005.09.004 (September 8, 2021).

Engelman, Jeffrey A. 2009. “Targeting PI3K Signalling in Cancer: Opportunities, Challenges and Limitations.” *Nature Reviews. Cancer* 9(8): 550–62.

Engelman, Jeffrey A., Ji Luo, and Lewis C. Cantley. 2006. “The Evolution of Phosphatidylinositol 3-Kinases as Regulators of Growth and Metabolism.” *Nature Reviews. Genetics* 7(8): 606–19.

Ferguson, Kathryn M. 2008. “Structure-Based View of Epidermal Growth Factor Receptor Regulation.” *Annual Review of Biophysics* 37(1): 353–73. http://www.annualreviews.org/doi/10.1146/annurev.biophys.37.032807.125829 (September 3, 2021).

Gao, Xiuhua, and Zafar Nawaz. “Role of Steroid Receptor Coactivators and Corepressors of Progesterone Receptors in Breast Cancer.” 4(5): 5.

Hare, Stephen H., and Amanda J. Harvey. 2017. “MTOR Function and Therapeutic Targeting in Breast Cancer.” *American Journal of Cancer Research* 7(3): 383–404.

Hubbard, Stevan R. 2004. “Juxtamembrane Autoinhibition in Receptor Tyrosine Kinases.” *Nature Reviews Molecular Cell Biology* 5(6): 464–71. https://www.nature.com/articles/nrm1399 (September 3, 2021).

Lipinski, Christopher A. 2004. “Lead- and Drug-like Compounds: The Rule-of-Five Revolution.” *Drug Discovery Today. Technologies* 1(4): 337–41.

Lipinski, Christopher A., Franco Lombardo, Beryl W. Dominy, and Paul J. Feeney. 1997. “Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings.” *Advanced Drug Delivery Reviews* 23(1): 3–25. https://www.sciencedirect.com/science/article/pii/S0169409X96004231 (December 15, 2021).

Masuda, Hiroko et al. 2012. “Role of Epidermal Growth Factor Receptor in Breast Cancer.” *Breast cancer research and treatment* 136(2): 10.1007/s10549-012-2289–9. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3832208/ (September 6, 2021).

Mendrola, Jeannine M., Mitchell B. Berger, Megan C. King, and Mark A. Lemmon. 2002. “The Single Transmembrane Domains of ErbB Receptors Self-Associate in Cell Membranes.” *The Journal of Biological Chemistry* 277(7): 4704–12.

Mfotie Njoya, Emmanuel, Jacobus N. Eloff, and Lyndy J. McGaw. 2018. “Croton Gratissimus Leaf Extracts Inhibit Cancer Cell Growth by Inducing Caspase 3/7 Activation with Additional Anti-Inflammatory and Antioxidant Activities.” *BMC complementary and alternative medicine* 18(1): 305.

Mondal, Arijit et al. 2019. “Alkaloids for Cancer Prevention and Therapy: Current Progress and Future Perspectives.” *European Journal of Pharmacology* 858: 172472. https://linkinghub.elsevier.com/retrieve/pii/S0014299919304248 (December 14, 2021).

Mote, P. A., J. D. Graham, and C. L. Clarke. 2007. “Progesterone Receptor Isoforms in Normal and Malignant Breast.” *Ernst Schering Foundation Symposium Proceedings* (1): 77–107.

O’Boyle, Noel M. et al. 2011. “Open Babel: An Open Chemical Toolbox.” *Journal of Cheminformatics* 3: 33.

Pa, Mote, Balleine Rl, McGowan Em, and Clarke Cl. 1999. “Colocalization of Progesterone Receptors A and B by Dual Immunofluorescent Histochemistry in Human Endometrium during the Menstrual Cycle.” *The Journal of clinical endocrinology and metabolism* 84(8). https://pubmed.ncbi.nlm.nih.gov/10443705/ (December 12, 2021).

Pinzi, Luca, and Giulio Rastelli. 2019. “Molecular Docking: Shifting Paradigms in Drug Discovery.” *International Journal of Molecular Sciences* 20(18): E4331.

Saxton, Robert A., and David M. Sabatini. 2017. “MTOR Signaling in Growth, Metabolism, and Disease.” *Cell* 168(6): 960–76. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5394987/ (December 12, 2021).

Tryfonidis, Konstantinos, Dimitrios Zardavas, Benita S. Katzenellenbogen, and Martine Piccart. 2016. “Endocrine Treatment in Breast Cancer: Cure, Resistance and Beyond.” *Cancer Treatment Reviews* 50: 68–81. https://linkinghub.elsevier.com/retrieve/pii/S0305737216300792 (August 3, 2021).

Zardavas, Dimitrios, Alexandre Irrthum, Charles Swanton, and Martine Piccart. 2015. “Clinical Management of Breast Cancer Heterogeneity.” *Nature Reviews Clinical Oncology* 12(7): 381–94. http://www.nature.com/articles/nrclinonc.2015.73 (August 3, 2021).

Techno-Science.net. (2021). Linux : définition et explications. Techno-Science.net.

A field to adjust the GridBox position

Abstract

Breast cancer is one of the biggest global dilemmas and its current therapy is to target the hormone receptors by the use of partial agonists/antagonists. Potent drugs for breast cancer treatment are Tamoxifen, Trastuzumab, Paclitaxel, etc. which show adverse efects and resistance in patients. The aim of the study has to identify biochemicals with potent actions on ERα, PR, EGFR and mTOR inhibition. The current study is performed using molecular docking approaches as protein-ligand interactions play a vital role in drug design. The 3D structures of ERα, PR, EGFR and mTOR were obtained from the protein data bank and docked with 1017 ligand 3D structures from SANCDB database and we choose the top five ligands that have a high docking score using Autodock vina . Drug-likeness property was checked by applying the Lipinski's rule of five on the 1017 to evaluate anti-breast cancer activity. The results confrm that Octahydroeuclin has the best docking score for breast cancer followed by Kaurenoic, Hamiltonin E Neolistine , Epicatechin gallate . This study suggests that the selected molecules can be further investigated and evaluated for breast cancer treatment and management strategies.