Proceedings of the Indian National Science Academy (2023) 89:957–966
https://doi.org/10.1007/s43538-023-00210-9

RESEARCH PAPER

Virtual screening of flavonoids from Blighia sapida against ERK5 involved in breast cancer

Damilola S. Bodun[1] · Damilola A. Omoboyowa[1] · Joshua F. Adedara[2] · Ezekiel A. Olugbogi[1] · Favour O. Oluwamoroti[2] · Nkechi H. Atasie[3] · Isaac O. Oluwafemi[4]

Received: 31 January 2023 / Accepted: 5 September 2023 / Published online: 2 October 2023

Abstract
Breast cancer is a major public health concern worldwide, with a need for effective treatments that have minimal side effects. Flavonoids have been shown to have potential as anticancer agents. In this study, the inhibitory effect of flavonoids from Blighia sapida on the breast cancer target, Extracellular signal-regulated kinase 5 (ERK5) was evaluated. A pharmacophore hypothesis was developed based on the interactions between the target protein and its co-crystallized ligand and then applied to the screening of flavonoids present in B. sapida. The top-ranked flavonoids were identified using molecular docking, MMGBSA, and pharmacokinetics modeling, and their inhibitory activity was determined using density functional theory (DFT) and their pIC50 values. The results showed that Quercetin, Kaempferol, and (+)-Catechin had higher docking scores and obeyed the Lipinski rule of five, indicating their potential as lead molecules for ERK5 inhibition in breast cancer therapy. The study highlights the potential of flavonoids from B. sapida as promising inhibitors of ERK5 for breast cancer management.

Keywords ERK5 · Breast cancer · Flavonoids · Pharmacophore · Density functional theory · Molecular docking · AutoQsar

Abbreviations

ERK5	Extracellular signal-regulated kinase 5
ADMET	Absorption, Distribution, Metabolism, Excretion, and Toxicity.
QSAR	Quantitative Structural Activity Relationship

Introduction

Cancer is a condition in which some cells in the body grow out of control and spread to other parts of the body (NCI 2007). Breast cancer is the most prevalent cancer diagnosed in women in the United States, and it is the second largest cause of death in women, following lung cancer (American Cancer Society). This disease is most common in women and has a greater rate in women in industrialized countries, although it can also affect men (Mustafa et al. 2016). Extracellular signal-regulated kinase 5 (ERK5) is a member of the mitogen-activated protein kinase (MAPK) family, which includes highly conserved enzymes found in all eukaryotic cells and is involved in a variety of biological responses such as cell survival, proliferation, migration, and differentiation (Stecca and Rovida 2019). As ERK5 is more abundantly expressed in breast cancer cells and tumor tissue compared to normal breast cells and tissue, accumulating lines of evidence suggest a critical involvement of ERK5 in the initiation and advancement of different types of cancer in recent years (Hoang, et al. 2020).

ERK5 has a substantially bigger C-terminus than other MAP kinases, and it comprises both auto-inhibitory and nuclear shuttling activities (Drew et al. 2012). This structure lends itself to the development of specific therapeutic targets that are unlikely to interact with the functioning of other MAP kinases. Because of this specificity, tailored therapies may be able to block this pathway while reducing the inhibition of other MAP kinases that are important for healthy cell survival (Drew et al. 2012). Signaling substances such as growth factors, cytokines, neurotransmitters, hormones, or various cell stressors activate MAPKs by sending signals

✉ Damilola S. Bodun
dbodun56@gmail.com

[1] Department of Biochemistry, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria

[2] Department of Physiology, Ekiti State University, Ado-Ekiti, Ekiti State, Nigeria

[3] Nigerian Correctional Service, Enugu Custodial Centre, Enugu, Nigeria

[4] Department of Chemistry, Adekunle Ajasin University, Akungba Akoko, Nigeria

through tyrosine kinase, G-coupled proteins, or hormone receptors (Flaherty et al. 2010).

Flavonoids have been demonstrated to have a wide range of anticancer properties, including modulating reactive oxygen species (ROS)-scavenging enzyme activities, participating in cell cycle arrest, inducing apoptosis and autophagy, and suppressing cancer cell proliferation and invasiveness. Flavonoids have a dual role in ROS homeostasis: they operate as antioxidants in healthy cells and are powerful prooxidants in cancer cells, inducing apoptosis and down-regulating pro-inflammatory signaling pathways (Kopustinskiene et al. 2020).

Blighia sapida is a member of the Sapindaceae family and Sapindales order, commonly known as ackee or breadfruit (Olayinka et al. 2021). It is distributed across nations around the world, including Spain, Guatemala, Panama, Mexico, Venezuela, Colombia, Costa Rica, Portugal, France and Surinam (Olayinka et al. 2021; Sinmisola et al. 2019). Traditional medicine claims that the ackee fruit has medicinal characteristics that can be used to treat or relieve symptoms such as fever, constipation, skin infections, and diarrhea. It is also said to contain phytochemicals with anti-cancer and anti-diabetic properties (Xiao et al. 2016; Omoboyowa et al. 2021). In this study, the ligand-docking technique, pharmacophore hypothesis screening, and quantum calculations were used to predict ERK5 inhibitors from *B. sapida* flavonoids for breast cancer management.

Methods

All computational experiments, such as E-pharmacophore hypothesis generation, virtual screening of the ligands with the hypothesis, molecular docking, MMGBSA, AutoQsar, and visualization, were done with Schrodinger Inc.'s Maestro version 11.1 software. Supplementary Fig. 1 illustrates the entire process workflow.

Protein preparation

The protein (PDB ID: 5O7I) was used as the 3-dimensional crystal structure of ERK5. Bond orders were assigned to the target protein (ERK5) and hydrogen atoms were added. Water molecules within 5 Å distance of the ligand were eliminated from the structure of the protein–ligand complex using Protein Preparation Wizard The prime tool was also used to patch up missing loops and side chains (Jacobson et al., 2004). ERK5 was further improved by creating tautomeric states at a neutral pH and using the OPLS3 force field to constrain minimization. For molecular docking, the minimized ERK5 was chosen.

Receptor grid generation

For protein–ligand docking, receptor grid generation determines the binding orientation and size of the active site. Using the receptor grid generation module of Schrödinger Maestro 11.1 (Omoboyowa et al., 2020), the scoring coordinates of the ERK5 binding pocket were computed based on the co-crystallized ligand. The coordinates for the x, y, and z grids are 43.47, − 1.06, and 11.01, respectively.

Ligand preparation

The 2D structures of the fifty-five (55) flavonoids from Blighia sapida and the reference compound were downloaded from the PubChem database after gathering from published literature (Sinmisola et al. 2019; Hamzah et al. 2013). The LigPrep module was utilized to prepare the ligands. Additionally, the ERK5 Co-ligand (4-(2-bromanyl-6-fluoranyl-phenyl)carbonyl-N-pyridin-3-yl-1H-pyrrole-2-carboxamide) was identified and retrieved from PubChem, after which it was prepared similarly to the other ligands and afterward utilized as a reference drug.

Generation of e-pharmacophore model

For virtual screening, a pharmacophore hypothesis based on the interactions of the co-crystallized ligand with the target protein was generated. The hypothesis was created using the Phase module. For the hypothesis generation settings, features that made interactions with the protein were chosen, and then a receptor-based excluded volume shell was created to mimic the receptor binding site, ignoring receptor atoms whose surfaces are within 2.00 Å of the ligand surface and limiting excluded volume shell thickness to 5.00 Å (SchrödingerTV 2016).

Virtual screening via pharmacophore hypothesis

Virtual screening was performed with the already prepared ligands via the Phase module (Dixon et al. 2006; Schrödinger and Phase 2020) to select a subset of ligands with the requisite characteristics for optimal binding to ERK5, as mapped by the hypothesis. The best hits were chosen based on fitness scores (> 1.0).

Virtual screening by Lipinski's rule and molecular docking

Forty-one (41) hits including the control ligand were screened through HTVS docking precision. The virtual screening by

Lipinski's rule of five to keep only oral bioavailable compounds was done via the QikProp module (Schrödinger 2020; Omoboyowa et al. 2022).

Extra precision (XP) docking was utilized to further screen the top five (5) lead ligands (Friesner et al. 2004).

Estimation of binding free energy

The MM-GBSA technique for calculating binding energy uses the energy characteristics of the free ligand, free receptor, and receptor-ligand complex to compute binding affinity (Omoboyowa 2022; Schrödinger 2018; Genheden and Ryde 2015). The MM-GBSA estimation was performed on the docked complexes using the formula:

$$\Delta G^{bind} = G^{complex}X - (G^{protein} + G^{Ligand}) \quad (1)$$

where ΔG^{bind} is the Protein–Ligand complex's binding free energy. $G^{complex}$ is the free energy of the complex, $G^{protein}$ is the free energy of the protein in the absence of the ligand, and G^{ligand} is the free energy of the ligand in the absence of the protein.

Density functional theory analysis

The physicochemical features of the top five (5) hit flavonoids from Blighia sapida were investigated using quantum chemical calculations using density functional theory (DFT) to anticipate molecules with noteworthy biological activity. To begin with, each bioactive molecule was subjected to a conformer distribution search, with the most stable conformer being chosen for comprehensive analysis. Many parameters can be produced as a result of the calculations conducted using this method. Highest occupied molecular orbital energy (E_{HOMO}), lowest unoccupied molecular orbital energy (E_{LUMO}), energy band gaps (E_g), ionization energy (I), electron affinity (A), chemical hardness, chemical softness, chemical potential, electronegativity, electronic energy, enthalpy, Gibb's free energy, and dipole moment are some of the parameters obtained from the calculations.

The difference between E_{LUMO} and E_{HOMO} was used to compute the energy bandgap (E_g)

$$Eg = ELUMO - EHOMO \quad (2)$$

Koopman's theorem (Koopmans, 1933) is used to connect electron affinity (A) and ionization potential (I) to E_{LUMO} and E_{HOMO}, as indicated in Eqs. (3) and (4), respectively.

$$I = -EHOMO \quad (3)$$

$$A = -ELUMO \quad (4)$$

Electronegativity (χ) and chemical hardness (η) Values of the compounds were calculated using equations derived by Parr and Pearson (Parr & Pearson, 1983).

$$\chi = \frac{I + A}{2} \quad (5)$$

$$\eta = \frac{I - A}{2} \quad (6)$$

Chemical softness (δ) is the reciprocal of chemical hardness.

$$\delta = \frac{1}{\eta} \quad (7)$$

Development of auto-QSAR model

The experimental data with pIC50 values of antagonists of ERK5 were retrieved from the CHEMBL website (https://www.ebi.ac.uk/chembl/) by blasting the FASTA sequence of the protein gotten from PDB. The list of antagonists was converted into.sdf format via the Datawarrior software after data cleaning using Microsoft Excel. The ligands were prepared and then used to create a QSAR model via the Auto-Qsar module. The best model, kpls_linear_14 was selected based on the ranking.

Evaluation of pharmacokinetics properties

Absorption. distribution, metabolism, and excretion (ADME) properties of the top five selected compounds were analyzed. The processes such as absorption, distribution, metabolism, and excretion are important ones in the development of drugs. These properties were predicted using the QikProp module.

Results and discussion

Virtual screening via pharmacophore model

In the study of how a compound interacts with the enzyme ERK5, we employed an E-pharmacophore model to identify the specific chemical groups or substructures of the co-crystallized ligand that were critical for optimal interaction with the enzyme (Omoboyowa, 2022). The model was then applied to screen the prepared phytochemicals, allowing us to filter the compounds down to a smaller set of 41. By comparing the features of these compounds to those of the co-crystallized ligand and evaluating their similarity, we were able to determine their fitness scores, which helped to further narrow down the selection (Fig. 1). Our analysis

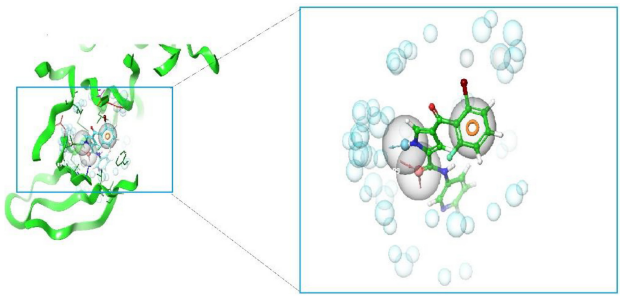

Fig. 1 The pharmacophore hypothesis generated in complex with the reference ligand

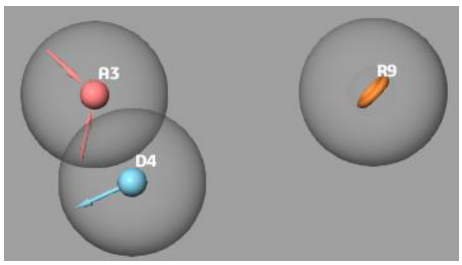

Fig. 2 The hydrogen bond acceptor (A3: red), hydrogen bond donor (D4: blue), and aromatic ring (R9: orange) of the pharmacophore hypothesis

revealed that the best hypothesis model for the optimal interaction with ERK5 consisted of one hydrogen bond acceptor, one hydrogen bond donor, and one aromatic ring. This finding is depicted in Fig. 2.

Molecular Docking and MM-GBSA calculation

Molecular docking is a structure-based drug design method that simulates the molecular interactions and binding mode of target proteins and ligands (Fan et al., 2019). From Quercetin to Alloathyriol, Flavonoids from Blighia sapida demonstrated a favorable binding affinity in the active site of ERK5, with binding energies of − 9.257 kcal/mol to − 7.847 kcal/mol, respectively. A stronger binding is associated with lower binding energy. The molecular docking methodology explores the behavior of small molecules in the binding site of a target protein (Pagadala et al., 2017). The lead compounds listed in Table 1 bind inside the active region of ERK5 while creating interactions with the hydrophobic amino acid residues MET 140, LEU 189, LEU 139, TYR 66, LEU 137, ALA 82, ILE 115, and VAL 69 (Table 1; Fig. 3). These amino acid residues are critical for anticipating the ERK5 binding site and catalytic mechanism. Through H-bond formation with the hydroxyl group, the docked molecules interact with MET 140, ASP 138, GLU 102, ASN 187, and LYS 84 in the ERK5 binding pocket.

Inter and intra-molecular interactions such as hydrogen bonding, pi-pi stacking, pi-cation, and salt bridges occurred as a result of the binding between the flavonoids and the target protein. The best bioactive molecule of B. sapida was Quercetin, which had maximum binding energy of − 9.257 kcal/mol. Hydrogen bonds are formed between it and the hydrophobic and negatively charged amino acids MET 140 and ASP 138. With binding energy of − 8.788 kcal/mol, kaempferol is thought to interact mostly with negative-charged, positive-charged, and polar amino acids: LYS 84, ASN 187, and ASP 138. (+)-Catechin is found to completely occupy the ERK5 binding site with a binding energy of − 8.384 kcal/mol while interacting with negative, positive, and polar amino acids: LYS 84, ASP 138, and ASN 187 via hydrogen bonds. Actaealactone and Alloathyriol had binding energies of − 7.974 and − 7.847 kcal/mol, respectively. The top four bioactive flavonoids of B. sapida have higher binding energy than the pyrrole inhibitor (4-(2-bromanyl-6-fluoranyl-phenyl)

Table 1 Fitness Score and Interacting amino acids of lead flavonoids from Blighia sapida and control ligand

Compound name	PubChem ID	Fitness score	H-bond residues	Interacting active site hydrophobic amino acids
Quercetin	5,280,343	1.865	MET 140, ASP 138	LEU 106, PHE 201, LEU 189, ILE 115, LEU 137, LEU 139, MET 140, ILE 61, ALA 82, TYR 66, VAL 69
Kaempferol	5,280,863	1.498	LYS 84, ASN 187, ASP 138	TYR 66, VAL 69, ALA 82, ILE 115, LEU 137, LEU 139, MET 140, LEU 189
(+)-catechin	9064	1.705	LYS 84, ASP 138, ASN 187	TYR 66, LEU 189, ILE 115, LEU 137, LEU 139, MET 140, ALA 82, VAL 69
Actaealactone	11,537,736	1.809	MET 140, ASP 143, LYS 84	TYR 66, VAL 69, ALA 82, LEU 137, LEU 139, MET 140, LEU 189
Co-crystallized ligand	118,959,080	2.428	MET 140, ASP 138. OTHER INTERACTIONS: TYR 66(PI-PI STACKING)	VAL 69, TYR 66, ALA 82, MET 140, LEU 139, LEU 137, ILE 115, LEU 189, ILE 61
Alloathyriol	44,575,387	1.52	MET 140	TYR 66, VAL 69, ILE 61, LEU 106, PHE 201, ILE 115, LEU 137, LEU 139, MET 140, ILE 61, ALA 82, ILE 189

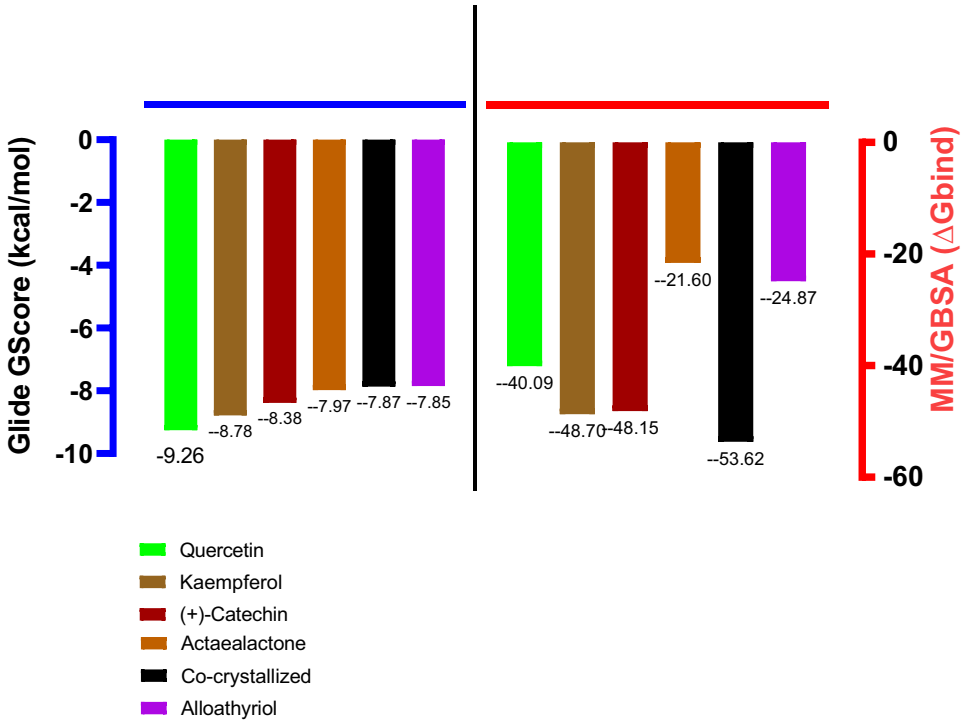

Fig. 3 Graphical representation of the docking score and MM-GBSA binding energy (ΔGbind) of Blighia sapida lead flavonoids and the control ligand. The left frame (blue) signifies the docking score, while the right frame (red) shows the MM-GBSA binding energy

carbonyl-N-pyridin-3-yl-1H-pyrrole-2-carboxamide) that was utilized as a positive control for the ligands. The binding energy of the pyrrole inhibitor was − 7.864 kcal/mol, indicating that B. sapida bioactive flavonoids have a strong potential to bind ERK5 for the treatment of various cancers, particularly breast cancer.

The MM-GBSA module, which is integrated with the Schrodinger suite's prime program, was used to compute the binding free energy (ΔG^{bind}) of the flavonoids binding to ERK5. Following the docking analysis, the module was used to calculate the binding free energy for the screened compounds. The MM-GBSA technique is a reliable post-docking method for estimating the binding position of docked complexes (Sgobba et al. 2012). Quercetin, Kaempferol, (+)-Catechin, Actaealactone, and Alloathyriol had binding energies of − 36.10, − 53.04, − 44.67, − 29.59, and − 24.16 kcal/mol, respectively, according to the MM-GBSA output (Fig. 3). The control ligand (− 53.62 kcal/mol) had the best MM-GBSA value (Fig. 4).

Pharmacokinetics and drug-likeness screening

To aid drug discovery, accurate data on the pharmacokinetic features of the molecule must be obtained as soon as possible, as this ultimately adds to the compound's success or failure (Ranjith and Ravikumar, 2019). Medication similarity evaluates the likelihood of a chemical becoming an oral drug with high bioavailability. Because the flavonoids were virtually screened for violations of the Lipinski Rule of Five, none of the top-ranked flavonoids broke this principle, implying that these molecules are medication prospects (Table 2).

It is important to know the pharmacokinetics of phytochemicals before their development into new pharmaceutical drugs. Using the QikProp module of Schrodinger, pharmacokinetic properties of the lead compounds were predicted using binding to human serum albumin ($QPlog^{Khsa}$), Madin-Darby canine kidney cell permeability (QPP^{MDCK}), blood/brain partition coefficient ($QPlog^{BB}$), IC50 value for the blockage of HERG K + channels ($QPlog^{HERG}$), and Caco-2 cell permeability (QPP^{Caco}). All of the lead compounds fell into the range of values for $QPlog^{KHSA}$, therefore they should all be able to bind to human serum albumin. All of the lead compounds except Kaempferol have been predicted to have good values for $QPlog^{HERG}$. Again, all of the lead compounds fell within the range of expected values for the brain/blood partition coefficient (Table 3). The results also showed that only Quercetin and Actaealactone have poor Caco-2 cell permeability.

Density functional theory analysis

DFT is a quantum-based computation that is used to predict the electronic structure and characteristics of prospective medicinal compounds (Burke and Wagner 2013). It is a tool that is utilized in tandem with all other drug discovery technologies.

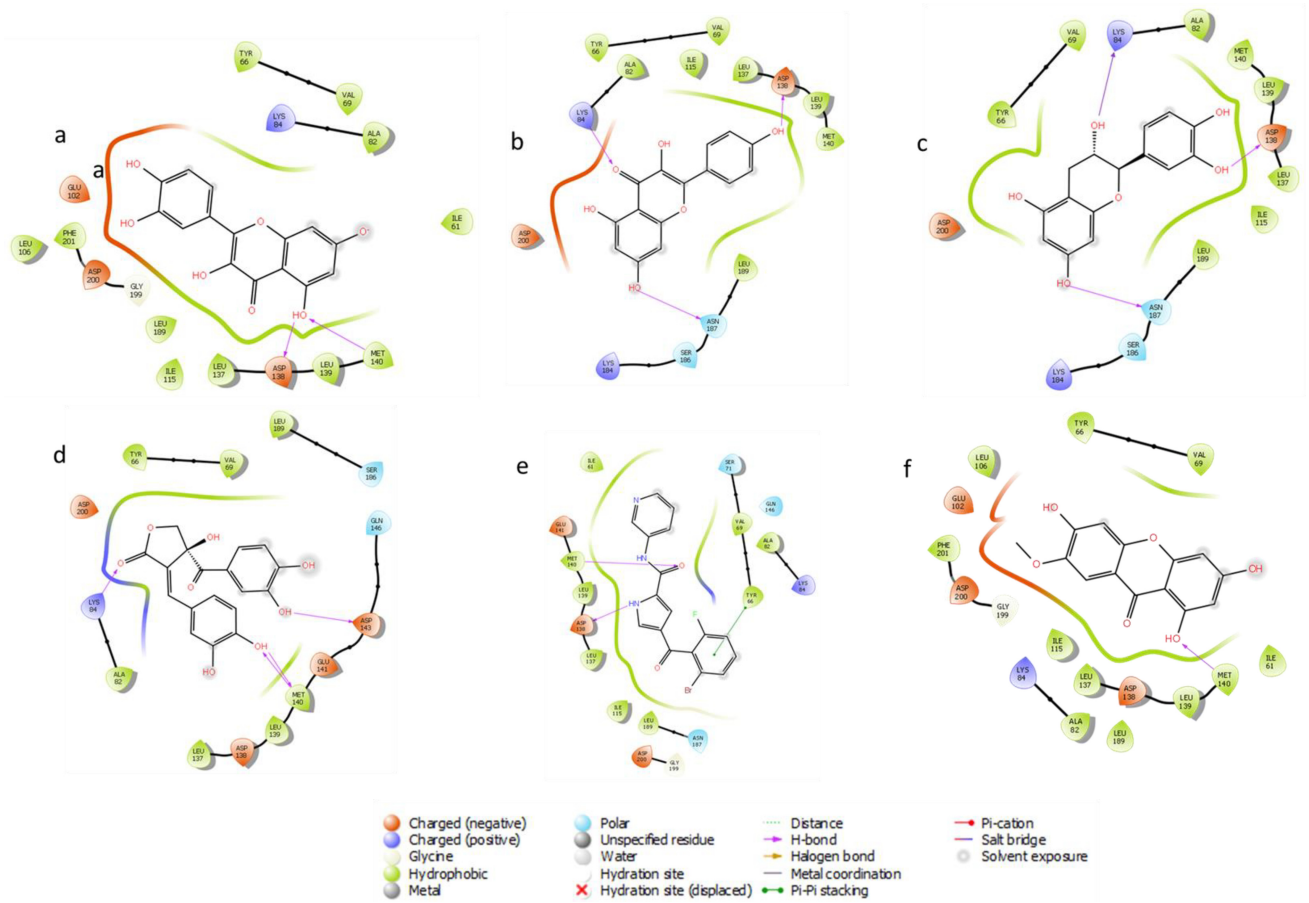

Fig. 4 2D molecular contacts profiling of hit compounds with amino acid residues at the active site of ERK5: **a** Quercetin **b** kaempferol, **c** (+) – Catechin **d** Actaelactone **e** Co-crystallized Ligand and **f** Alloathyriol

Table 2 Drug Likeness properties of lead compounds and control ligand

Entry name	mol MW	Hbond acceptors	Hbond donors	ALogP	Polar surface area	Rule of five
Quercetin	302.24	7	5	2.531	131.36	0
Kaempferol	286.24	6	4	2.7984	111.13	0
(+)-Catechin	290.272	6	5	1.9202	110.38	0
Actaealactone	358.304	7	4	0.9343	147.35	0
Co-crystallized Ligand	388.195	4	1	2.7951	71.95	0
Alloathyriol	274.229	5	2	1.7616	102.96	0

Thermochemical analysis

Thermodynamic characteristics are important factors in determining the spontaneity and chemical stability of a chemical reaction. Gibbs free energy is a thermodynamic parameter used to characterize the interaction between ligands and receptors (Chaudhary et al. 2022). It shows the likelihood of biomolecular processes taking place. A positive free energy value suggests that binding will not occur until additional external energy is added, whereas a negative free energy value indicates that binding will occur spontaneously. The degree of the negative free energy determines the extent of the ligand's interaction with the receptor. Enthalpy is also a measure of a thermodynamic system's total energy. When the ligand attaches to the receptor, the binding enthalpy indicates the energy change in the system. The computed Gibb's free energy for the investigated compounds is shown in Supplementary Table 2. Because all of the molecules have negative free energy, they can attach to the target receptor without requiring any external energy.

Table 3 Pharmacokinetics profile of lead compounds of Blighia sapida

Model	Quercetin	Kaempferol	(+)-Catechin	Actaealactone	Alloathyriol
QPlogHERG	− 4.875	− 5.054	− 4.897	− 4.223	− 4.476
QPPCaco	19.934	54.93	49.958	10.054	218.753
QPlogBB	− 2.305	− 1.824	− 1.95	− 2.673	− 1.186
QPPMDCK	7.185	21.49	19.395	3.429	95.701
QPlogKhsa	− 0.354	− 0.201	− 0.427	− 0.597	− 0.257
% Oral Absorption	52.196	64.083	59.884	42.501	76.2

QPlogKhsa: Binding to human serum albumin (− 1.5 to + 1.5)

QPlogHERG: IC50 value for blockage of HERG K + channels (below –5)

QPPMDCK: Apparent Madin-Darby canine kidney cell permeability innm/sec

QPlogBB: Brain/blood partition coefficient (− 3.0 to 1.2)

QPPCaco: Apparent Caco-2 cell permeability in nm/sec (< 25 poor, > 500 great)

The top hit chemical, Quercetin, as well as Acetalactone, had the highest free energy (− 1103.987 and − 1295.839 Hartree, respectively), implying that Acetalactone will interact more than the other compounds investigated. The dipole moment reveals the polarity of a compound as well as the distribution of electrons inside it (Fleming 1982). It improves the receptor protein's binding affinity, nonbonded interactions, and hydrogen bond formation. Acetalactone (7.00 debye) also has the largest dipole moment, as demonstrated in Supplementary Table 2.

Frontier molecular orbital (FMOs)

The most important orbitals in the molecule are the FMOs, HOMOs, and LUMOs. They are important in optical, electric, and UV–Vis spectral chemistry, as well as quantum chemistry. The FMOs describe how a molecule interacts with other molecules and provide information on electron transport in a molecule, as well as a molecule's chemical reactivity and stability (Sengupta et al. 2021). The HOMO energy defines the molecule's propensity to donate electrons; greater E_{HOMO} values imply a stronger inclination for the molecule to donate electrons. The E_{LUMO} determines a molecule's ability to receive an electron; a lower E_{LUMO} number improves the likelihood of taking electrons. Therefore, higher values of E_{HOMO} and lower values of E_{LUMO} are responsible for the low stability and high reactivity of a molecule. The E_{HOMO} values for the compounds tested rise in order as shown in Table 4: Quercetin > Kaempferol > (+)—Catechin > Alloathyriol > Acetalactone. The greatest value of E_{HOMO} is found in quercetin (− 5.48 eV), indicating that it has a higher tendency to donate an electron to the target receptor than other substances. In addition, the calculated E_{LUMO} values are provided in Table 4, Quercetin has the lowest E_{LUMO} value, indicating that it has a lower ability to receive electrons than the other compounds investigated. Furthermore, as shown in Supplementary Table 1, both the E_{HOMO} and E_{LUMO} are completely dispersed across the molecule structure, implying considerable HOMO–LUMO overlapping, resulting in robust charge transfer behavior. The band gap energy between the E_{HOMO} and E_{LUMO} is crucial for forecasting a molecule's chemical reactivity. The chemical reactivity and stability of a molecule are reflected in the band gap energy levels. The molecule becomes tougher, more stable, and less reactive as the band gap energy increases. High reactivity and low stability are associated with a narrowing of the energy band gap. The energy band gap values are as follows: Quercetin < Kaempferol < Alloathyriol < Acetalactone < (+)–Catechin. Among the isolated chemicals, Quercetin has the smallest band gap, indicating that it is more reactive toward the target receptor than the others.

Table 4 Global reactivity descriptors of Blighia sapida lead compounds obtained via DFT

Compounds	E_{HOMO} (eV)	E_{LUMO} (eV)	Eg (eV)	I((eV)	A(eV)	η (eV)	δ (eV-1)	χ (eV)	μ(eV-1)
Quercetin	− 5.48	− 1.84	3.64	5.48	1.84	1.82	0.549451	3.66	− 3.66
Kaempferol	− 5.53	− 1.81	3.72	5.53	1.81	1.86	0.537634	3.67	− 3.67
(+) –Catechin	− 5.63	− 0.08	5.71	5.63	− 0.08	2.855	0.350263	2.775	− 2.775
Acetalactone	− 6.02	− 1.78	4.24	6.02	1.78	2.12	0.471698	3.9	− 3.9
Alloathyriol	− 5.74	− 1.55	4.19	5.74	1.55	2.095	0.477327	3.645	− 3.645

Eg energy band gaps, E_{HOMO} highest occupied molecular orbital energy, E_{LUMO} lowest unoccupied molecular orbital energy, *I* ionization energy, *δ* chemical softness, *A* electron affinity, *η* chemical hardness, *μ* chemical potential, *χ* electronegativity

Global reactivity descriptors

To gain a thorough understanding of the chemical stability and reactivity of bioactive chemicals toward the target receptor, global reactivity descriptors (GRD) were developed. Ionization energy, electron affinity, chemical hardness, chemical softness, chemical potential, and electronegativity are the GRDs that are calculated. The ionization energy (I) of a molecule describes its chemical reactivity and stability. It's the amount of energy it takes to remove an electron from a molecule. High ionization energy suggests chemical inertness and stability, whereas low ionization energy indicates high reactivity and chemical inertness (Chakraborty et al., 2020). The ionization energy of quercetin (5.48 eV) is the lowest, making it the most reactive chemical toward the target receptor, ERK5. The energy released when an electron is added to a neutral molecule is known as electron affinity (A). A molecule with a high electron affinity is more likely than one with a low electron affinity to receive electrons easily (Geerlings and Proft 2002). The most reactive chemicals include kaempferol and quercetin, which have the highest electron affinity.

Understanding the reactivity of a chemical system requires an understanding of chemical hardness and softness. Chemical hardness describes a molecule's resistance to electron cloud deformation (Mortier et al. 1985). The band gap energy of a hard molecule is enormous, whereas the band gap energy of a soft molecule is tiny. The soft molecule will polarize more quickly and easily than the hard molecule (Obot et al. 2015). Catechin has the greatest hardness value (2.86 eV) in Table 4, suggesting that it is the hardest molecule. The softest chemical is quercetin, which has the lowest softness value (0.54 eV). The ability of a molecule to draw electrons.

Electronegativity is the tendency of a compound to attract electrons to itself. Table 4 shows that acetalactone is the most electronegative of all the chemicals (3.9 eV).

Analysing of auto-QSAR model

Quantitative structure-activity relationship (QSAR) modeling is a statistical tool used in computational biology and drug design to predict the biological activity of chemical compounds based on their molecular structure. Machine learning approaches, such as decision trees, neural networks, or support vector machines, are often used to train the models using a dataset of chemically described compounds. AutoQsar, a Schrodinger module, automatically divides datasets into 25% test and 75% train sets (Supplementary Table 3). Standard Deviation (S.D) (0.4079), Coefficient of Determination (R2) (0.7563), Root Mean Squared Error (RMSE) (0.4044), and Cross-Validation (Q2) (0.7518) were the parameters for the best model, Kpls_linear_14 (Table 5).

A low RMSE value indicates that the model's predictions are close to the actual activity, a high Q2 value indicates that the model is robust and has the good predictive ability, and a low standard deviation value indicates that the model's predictions are consistent and have little spread or variability (Veerasamy et al. 2011; Gramatica and Sangion 2016). These values suggest that the model has moderate to good predictive ability and stability.

The scatter plot of the train and test data sets observed and expected activity is depicted in Fig. 5. The QSAR model was used to predict the pIC50 values of the top-hit flavonoids and the reference medication. The top-ranked flavonoids performed less than the reference medication in terms of activity prediction (Table 6).

Table 5 Parameters of the QSAR model

Model Code	S.D	R^2	RMSE	Q^2
Kpls_linear_14	0.4079	0.7563	0.4044	0.7518

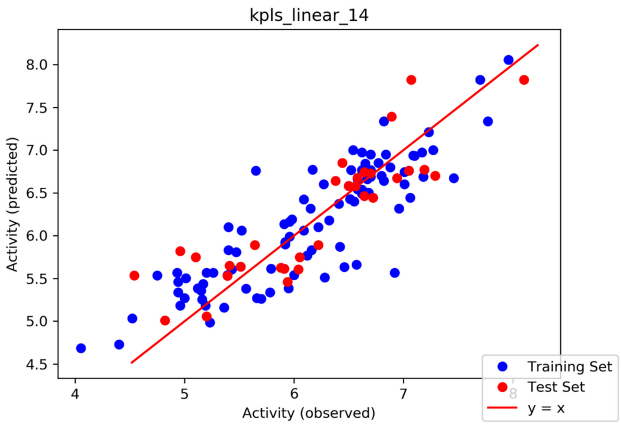

Fig. 5 Analysis of the scatter plot for the QSAR model

Table 6 Predicted bioactivity of the lead compounds and reference ligand

PubChem ID	Compound name	pIC50(μM)
9064	(+)-Catechin	5.648
11,537,736	Actaealactone	5.741
44,575,387	Alloathyriol	5.556
118,959,080	Co-crystallized ligand	6.062
5,280,863	Kaempferol	5.71
5,280,343	Quercetin	5.764

Conclusion

ERK5 is a protein kinase with cancer and inflammation therapeutic potential, making it a drug development target. This work investigated flavonoids from Blighia sapida to identify new lead compounds as ERK5 inhibitors for breast cancer therapy. The results demonstrated promising in silico inhibitory efficacy, drug-likeness, and future research promise. The use of virtual screening methods in conjunction with pharmacophore hypotheses aided in the identification of novel possible ERK5 inhibitors. Using the QSAR model, Quercetin and Actaealactone showed the highest-ranked predicted activity, implying that these chemicals are ERK5 inhibitors. However, additional experimental validation is required to demonstrate the therapeutic potential of these compounds as medication candidates in breast cancer therapy.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43538-023-00210-9.

Acknowledgements The authors would like to thank Dr. Oluwatoba E. Oyeneyin of the Department of Chemistry at Adekunle Ajasin University, Akungba-Akoko for showing them how to use the Spartan software.

Authors' contributions BDS, IOO, JFA, and NHA curated, and interpreted the data, and wrote the first draft of the manuscript. BDS, FOO and EAO wrote the methodology. ODA Supervised and Proofread the manuscript. All authors read and approved the final manuscript.

Funding No funding was received.

Data availability Not applicable.

Declarations

Competing interests All authors certify that they have no affiliations with or involvement in any organization with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethical approval Not Applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Avruch, J.: MAP kinase pathways: the first twenty years. Biochim Biophys Acta. **1773**(8), 1150–1160 (2007)

Becke, A.D.: Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. **98**, 5648–5652 (1993)

Breast Cancer Facts & Figures | American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/breast-cancer-facts-figures.html

Breast Cancer Fighting Foods (and What to Avoid). (n.d.). Retrieved May 27, 2022, from https://www.healthline.com/nutrition/breast-cancer-foods.

Burke, K., Wagner, L.O.: DFT in a nutshell. Int. J. Quantum Chem. **113**(2), 96–101 (2013)

What Is Cancer? - NCI (nciglobal,ncienterprise). (2007, September 17). (CgvArticle). https://www.cancer.gov/about-cancer/understanding/what-is-cancer.

Carmell, N., Rominiyi, O., Myers, K.N., McGarrity-Cottrell, C., Vanderlinden, A., Lad, N., Perroux-David, E., El-Khamisy, S.F., Fernando, M., Finegan, K.G., Brown, S., Collis, S.J.: Identification and validation of ERK5 as a DNA damage modulating drug target in glioblastoma. Cancers **13**(5), 944 (2021). https://doi.org/10.3390/cancers13050944

Chakraborty, T., Gazi, K., Ghosh, D.: C, Computational of the atomic radii through the conjoint action of the effective nuclear charge and ionization energy. Mol Phys. **108**, 2081–2092 (2020)

Chang, L., Karin, M.: Mammalian MAP kinase signalling cascades. Nature **410**(6824), 37–40 (2001)

Chaudhary, A., Kumar, M., Kumar, N., Agarwal, N.K.: Synthesis, spectroscopic characterization, biocidal evaluation molecular docking & DFT investigation of 16–18 membered macrocyclic complexes of cobalt (II). J. Chem. Sci. **134**(4), 113 (2022)

Cook, S.J., Tucker, J.A., Lockhead, P.A.: Small molecule ERK5 kinase inhibitors paradoxically activate ERK5 signalling be careful what you wish for…. Biochem Soc Trans. **48**(5), 1859–1875 (2020)

Dixon, S.L., Smondyrev, A.M., Knoll, E.H., Rao, S.N., Shaw, D.E., Friesner, R.A.: PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J. Comput. Aided Mol. Des. **20**, 647–671 (2006)

Drew, B.A., Burow, M.E., Beckman, B.S.: MEK5/ERK5 pathway: the first fifteen years. Biochem. Biophys. Acta. **1825**(1), 37–48 (2012). https://doi.org/10.1016/j.bbcan.2011.10.002

Fan, J., Fu, A., Zhang, L.: Progress in molecular docking. Quantitat Biol. **7**, 83–89 (2019)

Ferguson, F.M., Gray, N.S.: Kinase inhibitors: The road ahead. Nat. Rev. Drug Discov. **17**(5), 353–377 (2018). https://doi.org/10.1038/nrd.2018.21

Flaherty, P.T., Chopra, I., Jain, P., Yi, S., Allen, E., Cavanaugh, J.: Identification of benzimidazole-based inhibitors of the mitogen activated kinase-5 signaling pathway. Bioorg Med. Chem. Lett. **20**(9), 2892–2896 (2010)

Fleming, I.: Frontier Orbitals and Organic Chemical Reactions 1976. Willey and Sons, London (1982)

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., et al.: Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. **47**, 1739–1749 (2004)

Friesner, R.A., Murphy, R.B., Repasky, M.P., Frye, L.L., Greenwood, J.R., Halgren, T.A., et al.: Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for proteinligand complexes. J. Med. Chem. **49**, 6177–6196 (2006)

Geerlings, P., De Proft, F.: Chemical reactivity as described by quantum chemical methods. Int. J. Mol. Sci. **3**, 276–306 (2002)

Genheden, S., Ryde, U.: The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. **10**, 449–461 (2015)

Gramatica, P., Sangion, A.: A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology. J. Chem. Inf. Model. **56**(6), 1127–1131 (2016)

Hamzah, R.U., Egwim, E.C., Kabiru, A.Y., Muazu, M.B.: Phytochemical and invitro antioxidant properties of the methanolic extract of fruits of Blighia Sapida, Vitellaria Paradoxa and Vitex Doniana. Oxid. Antioxid. Med Sci. **2**, 217 (2013)

Hayashi, M., Fearns, C., Eliceiri, B., Yang, Y., Lee, J.D.: Big mitogen-activated protein kinase 1/extracellular signal-regulated kinase 5 signaling pathway is essential for tumor-associated angiogenesis. Cancer Res. **65**(17), 7699–7706 (2005)

Hoang, V.T., et al.: ERK5 is required for tumor growth and maintenance through regulation of the extracellular matrix in triple negative breast cancer. Front. Oncol. (2020). https://doi.org/10.3389/fonc.2020.01164

Imran, M., Salehi, B., Sharifi-Rad, J., Gondal, T.A., Saeed, F., Imran, A., Shahbaz, M., Fokou, P.V.T., Arshad, M.U., Khan, H., Guerreiro, S.G., Martins, N., Estevinho, L.M.: Kaempferol: A key emphasis to its anticancer potential. Molecules **24**(12), 2277 (2019). https://doi.org/10.3390/molecules24122277

Jensen, F.: Polarization consistent basis sets: Principles. J. Chem. Phys. **115**(20), 9113–9125 (2001)

Kato, Y., Kravchenko, V.V., Tapping, R.I., Han, J., Ulevitch, R.J., Lee, J.D.: BMK1/ERK5 regulates serum-induced early gene expression through transcription factor MEF2C. Embo J. **16**(23), 7054 (1997a)

Kato, Y., Kravchenko, V.V., Tapping, R.I., Han, J., Ulevitch, R.J., Lee, J.D.: BMK1/ERK5 regulates serum-induced early gene expression through transcription factor MEF2C. EMBO J. **16**, 7054–7066 (1997b). https://doi.org/10.1093/emboj/16.23.7054

Koopmans, T.: Ordering of wave functions and Eigen energies to the individual electrons of an atom. Physica. **1**, 104–113 (1933)

Kopustinskiene, D.M., Jakstas, V., Savickas, A., Bernatoniene, J.: Flavonoids as anticancer agents. Nutrients **12**(2), 457 (2020). https://doi.org/10.3390/nu12020457

Mortier, W.J., Van Genechten, K., Gasteiger, J.: Electronegativity equalization: application and parametrization. J Am Chem Soc. **107**, 829–835 (1985)

Mulloy, R., Salinas, S., Philips, A., Hipskind, R.A.: Activation of cyclin D1 expression by the ERK5 cascade. Oncogene **22**(35), 5387 (2003)

Mustafa, M., et al.: Breast cancer: detection markers, prognosis, and prevention. IOSR J. Dental. Med. Sci. **15**, 73–80 (2016)

Obot, I.B., Kaya, S., Tuzum, B.: Theoretical evaluation of triazine derivatives as steel corrosion inhibitors: DFT and Monte Carlo simulation approaches. Res Chem. Intermediates. **42**, 4963–4983 (2015)

Olayinka, J.N., Ozolua, R.I., Akhigbemen, A.M.: Phytochemical screening of aqueous leaf extract of Blighia sapida K.D. Koenig (Sapindaceae) and its analgesic property in mice. J. Ethnopharmacol. (2021). https://doi.org/10.1016/j.jep.2021.113977

Omoboyowa, D.A.: Exploring molecular docking with E-pharmacophore and QSAR models to predict potent inhibitors of 14-α-demethylase protease from Moringa spp. Pharmacol. Res-Mod. Chin. Med. **4**, 100147 (2022)

Omoboyowa, D.A., Karigidi, K.O., Aribigbola, T.C.: Nephro-protective efficacy of Blighia sapida stem bark ether fractions on experimentally induced diabetes nephropathy. Comp. Clin. Pathol. (2021). https://doi.org/10.1007/s00580-020-03186-w

Omoboyowa, D.A., Balogun, T.A., Saibu, O.A., Chukwudozie, O.S., Alausa, A., Olubode, S.O., Aborode, A.T., Batiha, G.E., Bodun, D.S., Musa, S.O.: Structure-based discovery of selective CYP17A1 inhibitors for Castration-resistant prostate cancer treatment. Biol. Methods. Protocols. (2022). https://doi.org/10.1093/biomethods/bpab026

Pagadala, N.S., Syed, K., Tuszynski, J.: Software for molecular docking: a review. Biophys. Rev. **9**, 91–102 (2017)

Parr, R.G., Pearson, R.G.: Absolute hardness: companion parameter to absolute electronegativity. J. Am. Chem. Soc. **105**(26), 7512–7516 (1983)

Ranjith, D., Ravikumar, C.: SwissADME predictions of pharmacokinetics and drug-likeness properties of small molecules present in Ipomoea mauritiana Jacq. J. Pharmacognosy. Phytochem. **8**(5), 2063–2073 (2019)

Schrödinger Release 2018–4: Maestro Schrödinger. New York, NY: LLC, 2018

Schrödinger, LLC, Protein Preparation Wizard; Epik, 2020. Available at: https://www.schrodinger.com/protein-preparation-wizard.

Schrödinger, LLC, Glide, 2020. Available at: https://www.schrodinger.com/glide.

Schrödinger, LLC, LigPrep, 2020. Available at: https://www.schrodinger.com/ligprep

Schrödinger, LLC, Phase, 2020. Available at: https://www.schrodinger.com/phase.

Schrödinger, LLC, Prime, 2020. Available at: https://www.schrodinger.com/prime.

Schrödinger, LLC, QikProp, 2020. Available at: https://www.schrodinger.com/qikprop.

SchrödingerTV, 6 Dec 2016, Pharmacophore Hypothesis - With e-Pharmacophores from a Protein-Ligand Complex, YouTube, 3:33. https://www.youtube.com/watch?v=imEUJdTDplI.

Sekar, A., Soundhararajan, R., Srinivasan, H.: In silico analysis of quercetin and its analogues against targeted proteins biointerface research in applied chemistry. J. Chem. Phys. **98**, 5648–5652 (1993)

Sengupta, S., Murmu, M., Mandal, S., Hirani, H., Banerjee, P.: Competitive corrosion inhibition performance of alkyl/acyl substituted 2-(2-hydroxybenzylideneamino) phenol protecting mild steel used in adverse acidic medium: A dual approach analysis using FMOs/molecular dynamics simulation corroborated experimental findings. Colloids Surf., A **617**, 126314 (2021)

Sgobba, M., Caporuscio, F., Anighoro, A., Portioli, C., Rastelli, G.: Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations. Eur. J. Med. Chem. **58**, 431–440 (2012)

Sinmisola, A., Oluwasesan, B.M., Chukwuemeka, A.P.: Blighia sapida K.D. Koenig: A review on its phytochemistry, pharmacological and nutritional properties. J. Ethnopharmacol. (2019). https://doi.org/10.1016/j.jep.2019.01.017

Sinmisola, A., Oluwasesan, B.M., Chukwuemeka, A.P.: Blighia sapida KD Koenig: A review on its phytochemistry, pharmacological and nutritional properties. J. Ethnopharmacol. **235**, 446–459 (2019)

Stecca, B., Rovida, E.: Impact of ERK5 on the hallmarks of cancer. Int. J. Mol. Sci. **20**(6), 1426 (2019)

Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C.P., Agrawal, R.K.: Validation of QSAR models-strategies and importance. Int. J. Drug Des. Discov **3**, 511–519 (2011)

Xiao, H., Yao, Z., Peng, Q., Ni, F., Sun, Y., Zhang, C.X., Zhong, Z.X.: Extraction of squalene from camellia oil by silver ion complexation. Sep. Purif. Technol. **169**, 196–201 (2016)