Research Article

Journal of Chemistry Letters
journal homepage: www.jchemlett.com
ISSN (online) 2717-1892 (print) 2821-0123

QSAR, docking and pharmacokinetic studies of 2,4-diphenyl indenol [1,2-B] pyridinol derivatives targeting breast cancer receptors

Auwal Salisu Isa, Adamu Uzairu, umar mele Umar, Muhammad Tukur Ibrahim, Abdullahi Bello Umar [2]

[1]*Department of Chemistry faculty of science Yobe state University*
[2]*Department of Chemistry, Ahmadu Bello University, Zaria*
[3]*chemistry department, faculty of science. yobe state university.*

ARTICLE INFO

Article history:
Received 12 November 2023
Received in revised form 8 January 2024
Accepted 8 January 2024
Available online 8 January 2024

Keywords:
Breast cancer (BC), Protein Data Bank, Molecular descriptor, Lee-Yang-Parr hybrid (B3LYP), Concordant correlation coefficient (CCC)

ABSTRACT

This study presents a computational approach for designing potent compounds against breast cancer. A robust quantitative structure-activity relationship (QSAR) model, developed using genetic algorithms and multilinear regression analysis, predicts chemical activity (pGI50) against breast cancer receptors. The model's reliability is validated with external metrics, emphasizing precision and strong relationships. Molecular docking investigations explore interactions between 2,4-diphenyl indenol [1,2-b] pyridinol derivatives and breast cancer receptors (2RMJ, 4OAR, 4RDH, 3ERT). Remarkable binding patterns are observed, insinuating at potential DNA gyrase inhibition. The compound's molecular properties and descriptors offer valuable insights into physicochemical characteristics, druglikeness, and potential pharmacological behavior. These findings contribute to drug design and development for personalized breast cancer therapy.
This research integrates computational methodologies with experimental data, paving the way for effective and targeted breast cancer treatments. The study emphasizes the potential of computational analysis to enhance precision and efficacy in breast cancer treatment strategies.

Introduction

Breast cancer is a disease characterized by the uncontrolled proliferation of cells. Various types of breast cancer exist [1]. The type of breast cells that develop into cancer determines the specific subtype of breast cancer [2]. The most common cancer among women is breast cancer, which exhibits diverse molecular characteristics [3]. To address this diversity, treatment strategies have evolved over the past decade, emphasizing biologically-directed medications and minimizing side effects [4]. Although certain characteristics, such as the impact of tumor load or metastatic patterns, influence treatment decisions, modern approaches primarily consider molecular heterogeneity [5]. Multimodal treatment advancements have increased the likelihood of successful therapy in 70–80% of patients [2]. However, effective treatments for advanced (metastatic) breast cancer remain elusive [6]. The primary goals for treating advanced breast cancer are prolonging life, symptom control, and minimizing medication risk to enhance the overall quality of life [7].

Breast cancer (BC) is the leading cancer among women in the United States, both in terms of diagnosis and cancer-related deaths. Despite improvements in patient care and screening, it remains a significant health concern [8]. In 2019, 268,600 new BC cases were diagnosed in women, leading to 41 deaths [9]. BC is a complex illness with various subtypes, with most responding well to hormonal and targeted treatments, resulting in favorable survival rates [10]. A notable subtype is triple-negative breast cancer (TNBC), accounting for 15-20% of annual BC diagnoses [11]. TNBC is distinct due to its aggressive behavior, poor prognosis, higher recurrence rates, and reduced survival rates [12–13]. Despite progress, the FDA has not approved specific remedies for TNBC. Chemotherapy remains the primary option, despite resistance development and adverse effects [14].

∗ Corresponding author: salisuisayaro1994@gmail.com
https://doi.org/ 10.22034/jchemlett.2024.424800.1146

Chemotherapy's long-term effects include infertility, osteoporosis, heart damage, and, rarely, leukemia, impacting survivors' quality of life [5].

The study's primary objective was to investigate the QSARs of synthesized compounds regarding their topoisomerase inhibitory and antiproliferative activities [15]. For many years, topoisomerase IIα has stood out as a significant focus in the fight against cancer due to its vital role in rapidly dividing cancer cells [16]. Among drugs aimed at inhibiting Topoisomerase IIα, those known as Topoisomerase II poisons are regularly employed in clinical settings [16]. Nevertheless, the application of topoisomerase II poisons leads to significant ramifications stemming from the DNA toxicity they cause through their mechanism of action. Hence, there is a compelling need to create a model for Topoisomerase IIα using an alternative computational method approach for regulation. The investigation explored the effects of hydroxyls in the indenopyridine ring and assessed the biological impact of introducing phenyl, phenol, and halophenyl moieties. The compounds with phenyl and phenol moieties exhibited remarkable improvements in their actions against the T47D cancer cell line. Some compounds displayed strong antiproliferative activity within the nanomolar range [17].

The research aims to develop a robust QSAR model with minimal side effects, an affordable cost, and efficacy against four breast receptors using scientific data from published works. The synthesized drug was evaluated under a structure-activity relationship, showing promise as a preclinical candidate for developing a robust chemotherapeutic agent for breast cancer [15]. The receptors that have been chosen each have a unique role in the development and treatment of breast cancer. Receptor 2RMJ is a DNA gyrase and has implications for preventing the growth of cancer cells. Receptor 4OAR is involved in cell signaling and suggests that it can help regulate the progression of breast cancer [17]. Receptor 4RDH is linked to anti-cancer compounds, which provides insights into potential therapeutic pathways. Receptor 3ERT is significant as an estrogen receptor and highlights its importance in breast cancer therapy, and it gives us more information about the effects of hormones [18]. By studying these receptors' interactions, we can improve our understanding of compound mechanisms and their potential as treatments for breast cancer.

Material and Methods:

Data Collection: This study utilized a statistical dataset of 84 anticancer synthesis compounds acquired from the Journal of Medicinal Chemistry [15]. The Genetic Algorithm was combined with Friedman's Multivariate Adaptive Regression Splines (MARS) and Holland's Genetic Algorithm [13]. GFA differs from other statistical strategies in that it generates a population of models instead of a single model [19]. The model underwent both internal and external validation to determine predictability and reliability [20].

Geometry Optimization

Quantum chemical descriptors were generated using Spartan 14 software (Spartan 14v1.14) and the 6-31G* basis set [21]. The compounds were optimized using Density Functional Theory (DFT) with the Lee-Yang-Parr hybrid functional (B3LYP) [22].

Descriptor Calculation

For each molecule shown in Table 1 (Sup. M), various physicochemical descriptors were determined [22, 38]. The "PaDel-Descriptor Model 2.20" program was used to generate an additional set of molecular descriptors. These descriptors were then combined with the quantum chemical descriptors obtained from the low-energy conformers [23].

Data Pre-Treatment/Feature Selection: Molecular descriptors were pre-processed by removing descriptors with constant values and pairs of variables with correlation coefficients greater than 0.7. This process was conducted using the "Data Pre-Treatment GUI 1.2" tool, employing the V-WSP method [24].

Dataset Division: The dataset of 84 chemical structures was divided into training and test sets using the Kennard Stone algorithm approach, implemented through the "Dataset Division GUI 1.2" program [25]. This application tool facilitates the selection of reasonable training and testing sets from a data source [26].

QSAR Model Development and Validation:

The study employed both quantum chemical and molecular descriptors as independent variables and response variables (pGI_{50} and pLC_{50}) as dependent variables. The Genetic Function Approximation (GFA) strategy was employed for both response variables through a multiple regression approach using the Material Studio program [20].

$$LOF = \frac{SSE}{(1-\frac{c+dp}{m})^2} \ldots\ldots\ldots\ldots Eqn1$$

GFA utilizes the Friedman Lack-of-Fit (LOF) metric to assess a model's fitness during the evolution process. LOF (Eqn 1) is calculated in Materials Studio using the formula provided by [27]: $LOF = SSE/((1 - (c + dp)/m)2)$ Here, SSE represents the sum of squared errors, c denotes the number of terms in the model that are not constant, and dp represents the total number of descriptors in all model terms (excluding the constant term), [28]

Internal Model Validation:

The internal cross-validation was conducted using the leave-one-out (LOO) method to validate the constructed models [29]. In this approach, the model is built using the remaining compounds after randomly excluding one compound from the dataset in each iteration [30]. The developed model is then applied to predict the activity of the excluded compound. This procedure is performed for each chemical and repeated accordingly. The cross-validated squared correlation coefficient, R^2cv (Q^2), was calculated using the following formula [31& 38].

$$Q^2 = 1 - \frac{\sum(Y_{OBS} - Y_{Pred})^2}{\sum(Y_{OBS} - \hat{Y})^2} \quad \text{…Eqn 2}$$

$$PRESS = \sum_{i=1}^{N}(y_{pred,i} - y_i)^2 \quad \ldots\ldots Eqn\ 3$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \ldots\ldots\ldots Eqn\ 4$$

Y_{pred} represents the projected activity of the practice compounds, Y_{OBS} denotes the observed activity of the training set compounds, and Y signifies the mean observed activity of the instructional set compounds, [29]. PRESS is a metric used to evaluate how effectively a regression model predicts new, unseen data points. It is frequently employed in cross-validation to assess a model's ability to generalize. The formula calculates the sum of the squared differences between the model's predicted values and the actual observed values for each data point. Essentially, PRESS Equation (3) measures the overall accuracy of predictions by considering the magnitude of errors. A lower PRESS indicates superior predictive performance, suggesting that the model is capturing the underlying patterns in the data. MAE Equation (4) is a straightforward metric used to quantify the average magnitude of errors between a predictive model's estimated values and the actual observed values. It provides a simple and interpretable measure of how closely the model's predictions align with the true values. The formula calculates the average absolute difference between predictions and actual values. Similar to PRESS, a lower MAE is desirable, signifying better predictive accuracy. MAE is easy to understand and is useful for comparing different models or assessing a model's performance over multiple predictions.

External Model Validation

To evaluate the predictive capability of the generated model, external validation was employed to calculate the predictive R^2 (R^2_{pred}) Equation (5) value and to apply the model for predicting activity values in the assessment group [23 & 38].

$$R^2_{pred} = 1 - \frac{\sum(Y_{pred(test)} - Y_{(Test)})^2}{\sum(Y_{(test)} - \widehat{Y_{training}})^2} \quad \ldots\ldots\ldots Eqn\ 5$$

$Y_{pred\ (test)}$ and $Y_{(test)}$ represent the anticipated and observed activity levels of the test compounds, respectively, while $Y_{(training)}$ denotes the mean activity value of the training set. The predicted correlation coefficient, R^2_{pred} (Eqn 5), is determined by extrapolating the expected activity of each component in the test set. However, it has been recognized that R^2_{pred} may not be sufficient to fully represent the external predictability of a model, as its value is influenced by $(Y_{(test)}\text{-}Y_{(training)})^2$ [31].

Docking: We obtained crystal structures of the central enzymes (DNA gyrase) from the Protein Data Bank (PDB). The structures have PDB codes 2RMJ, 4OAR, 4RDH, and 3ERT. Docking simulations were conducted using Auto Dock 4.2 within the PyRx software. To establish accurate binding interactions among the enzymes (proteins), and ligands (molecules), all water molecules, ligands, and cofactors associated with the enzymes were eliminated using Discovery Studio Visualizer software. Subsequently, the enzyme protein was saved in PDB format, recognized by the PyRx software, and converted into a macromolecule [32-33]. The docking interaction between the central enzyme and the protein was then computed to evaluate binding affinities. Various interaction types, such as hydrogen bonding, electrostatic interactions, and hydrophobic associations, were visualized and analyzed. Following that, all optimized ligands were saved in PDB format, which was also recognized by the PyRx software [26].

Pharmacokinetics study

PkCSM, an online server (http://structure.bioc.cam.ac.uk/pkcsm), and SwissADME (http://www.swissadme.ch/index.php) are available web engines designed to assess the ADMET properties and drug-likeness of small molecules [34]. One of the most crucial parameters in the pre-clinical stage of drug discovery is Lipinski's Rule of Five. It proposes that for a chemical compound to be permeable or easily absorbed into the body's system, it

shouldn't violate more than 2 of these criteria: molecular weight < 500, number of hydrogen bond donors < 5, number of hydrogen bond acceptors < 10, calculated Log P < 5, and polar surface area (PSA) < 100. Additionally, the bioavailability score (BAS) should be > 0.5 [35].

Result

Result of Model development and validation

Table 1 summarizes QSAR validation parameters, while Figure 1 compares experimental and predicted values. Figure 2 displays predicted values against residuals, and Figure 4 shows predicted values against leverage values. Additionally, Figure 5 illustrates the relationship between the first principal component (PC1) and the second principal component (PC2). The QSAR analysis was performed using regression. The robustness of the generated QSAR model was illustrated through an interactive activity graph (Figs. 1 and 2), which depicted the relationship between predicted and experimental activity. The QSAR model's equation is represented as follows:

pIC_{50}=5.4003−4.5609·ATS6i+4.4266·ATS6s−0.6453·MATS5v−1.2466·MATS4i−0.9855·GATS7m−0.2981·SpMax3_Bhm+2.2067·Dp.

The equation represents the relationship between the pIC_{50} value and the various variables in the model. The coefficients indicate the impact of each variable on the pIC_{50} value, and the equation allows for the prediction of pIC_{50} based on the values of these variables. The training set's characteristics were as follows: $N_{(Train)}$ = 64, R^2_{Train} = 0.8186, $R^2_{Adjusted}$ (train) = 0.7959, F = 36.0902, Q2(LOO) = 0.7654, R^2-R^2_{ADJ} = 0.0227, LOF = 0.0484, $RMS._{tr}$ = 0.1719. For the test set: $N_{(test)}$ = 20, R^2_{test} = 0.5901, $RMSE_{test}$ = 0.2081, $PRESS\_{EXT}$ = 0.8662, MAE_{EXT} = 0.1723.

The model's performance was also evaluated by comparing the predicted and actual pGI50 values in Fig. 1. The model equation was used to predict the values, as depicted in a scatter plot (Fig. 1), indicating strong model performance. Examining residual quantities and analyzing the standard residual against credit are shown in various figures (Fig. 2, & Fig. 3). The model exhibited excellent agreement with the test set data, as depicted in Table 1 and SM Table 1. Notably, no significant relative or systematic error was observed, indicating the model's reliability. The applicability domain of the model was established, with most compounds within the specified range except for a few outliers due to dissimilarity in chemical structures. Normalized residuals showed no unusual compounds > 3D for the dataset. The model's strength was further supported by statistical parameters such as MSE, RMSE, the slope of the models, and their coefficients. The predictive and descriptive potential of the model was evident, demonstrating its robustness in predicting anti-cancer activity for the breast tumor cell line. The model, based on the training set, successfully forecasted the activity of the compound validation set, and lower residual values from both the training and test sets indicated a strong correlation between activity and shape, as shown in Figures 1, 2, 3, 4, and 5. The generated QSAR model exhibited a high potential for accurately predicting anti-cancer activity based on the breast tumor cell line.

The regression coefficient (R^2) value close to 1 or 0 indicated substantial descriptor variation capture, contributing to a reliable model. The 0.5901 value further indicated the model's effectiveness. When assessing the accuracy and reliability of a model based on external data, external validation metrics are utilized. The $RMSE_{Ext}$. score of 0.2081 represents the level of prediction error, with lower values indicating higher accuracy. The MAE_{Ext} score of 0.1723 provides information on prediction accuracy. $PRESS_{Ext}$, which measures squared differences, has a score of 0.8662, while Next, which shows predictable variance, has a score of 0.5901. Q2-F_1, Q^2-F_2, and Q^2-F_3 are cross-validated correlations for subsets. The CCC_{Ext} score of 0.8711 evaluates precision, r^2m aver. (0.4454) gauges the strength of the relationship between variables, and r^2m delta (0.1965) highlights variability. These metrics combined assess the model's predictive accuracy, generalization, and consistency with external data, revealing its reliability and suitability for predicting activity across various contexts.

The results in Table 1 show that the model passes all validation parameters. Figure 1 displays a scatter plot comparing experimental and predicted activity values using a model equation. While most compounds closely match the predictions, Figure 2 displays a scatter plot of predicted values versus residual values. Figure 3 shows the connection between standard residuals (on the Y-axis) and leverage values (on the X-axis) for the HAT i/i (h* = 0.375) criterion. This graph can help spot any data points that could potentially have a significant impact on the model. Figure 4 displays a scatter graph that shows how predicted values from the model equation relate to leverage values using the HAT i/i (h* = 0.375) criterion. This graph can help identify data points that might have a big impact on how well the model performs. You can get a better idea of how to refine the model and make sure it's strong enough for real-world use. Figure 5 presents a scatter plot that displays Principal Component 2 (PC2) with 20.34% explained variance and Principal Component 1 (PC1) with 40.77% explained variance. This graph is useful for visualizing data patterns and relationships, which can help in comprehending intricate multivariate data.

Table 1. Accepted QSAR Tool for Model Validation

Validation tool	Interpretation	Acceptable value	Develop model value	Remarks
R^2 Trian	Co-efficient of determination	≥ 0.5	0.8186	Pass
Q^2CV	Cross validation co-efficient	>0.5	0.7654	Pass
R^2adj	Adjusted co-efficient of determination	>0.5	0.7959	Pass
R^2Q2cv	Different between R^2 and Q^2cv	≥ 0.03	0.0532	Pass
Next/test set	Minimum number of external tests set	> 5	20	Pass
R^2test set	Co-efficient of determination of external and test set	≥0.5	0.5901	Pass
Delta rm2		<0.2	0.1965	Pass
CCC	Concordant correlation coefficient	≥ 0.8	0.8711	Pass
RMSE	Root means square error	< 0.3	0.2081	Pass
LOF			0.0484	Pass
F	Fisher statistic	>10	36.0902	Pass

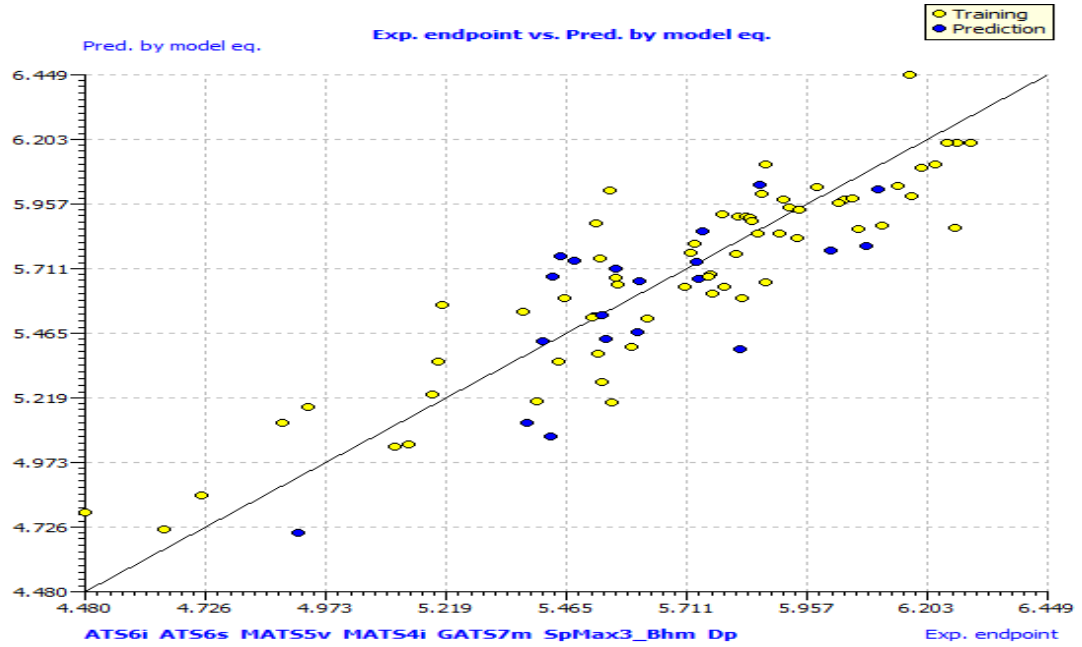

Fig. 1. Graph of experimental values against predicted values

.

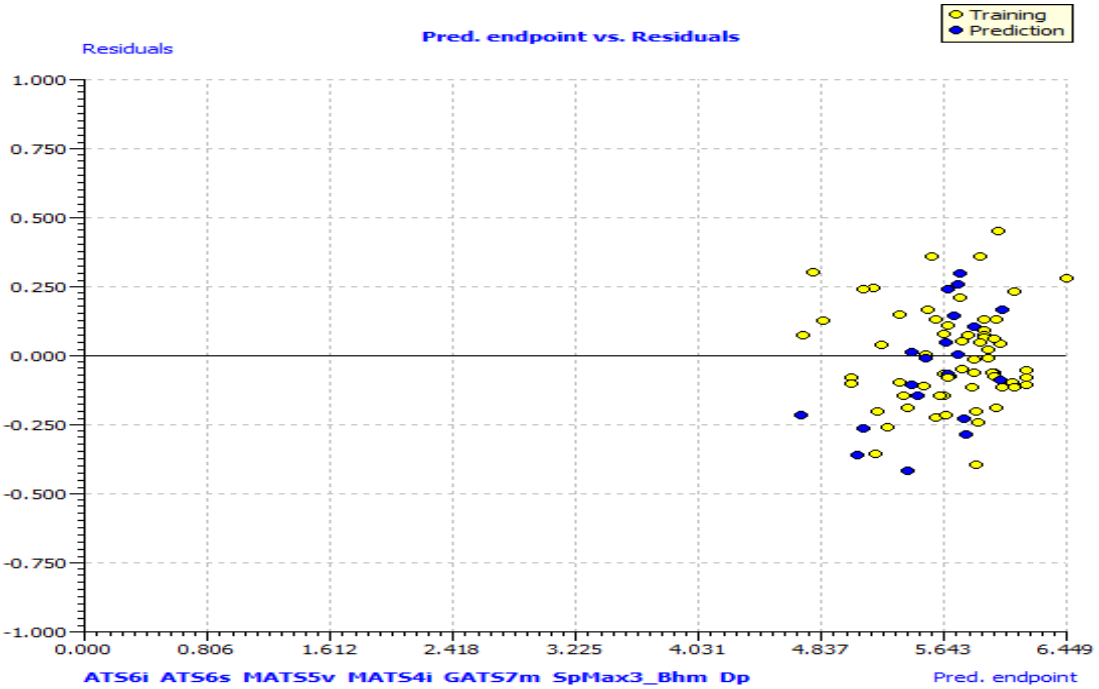

Fig. 2. Graph of predicted value against Residual values

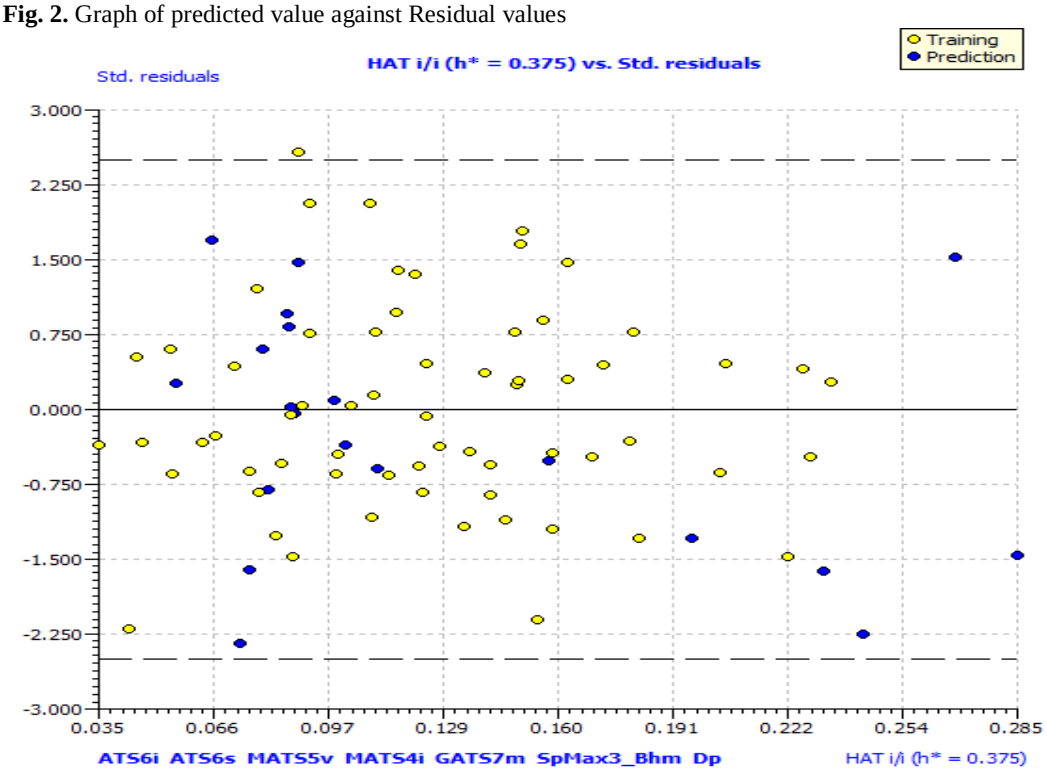

Fig. 3. Graph of standard residual against Leverage values

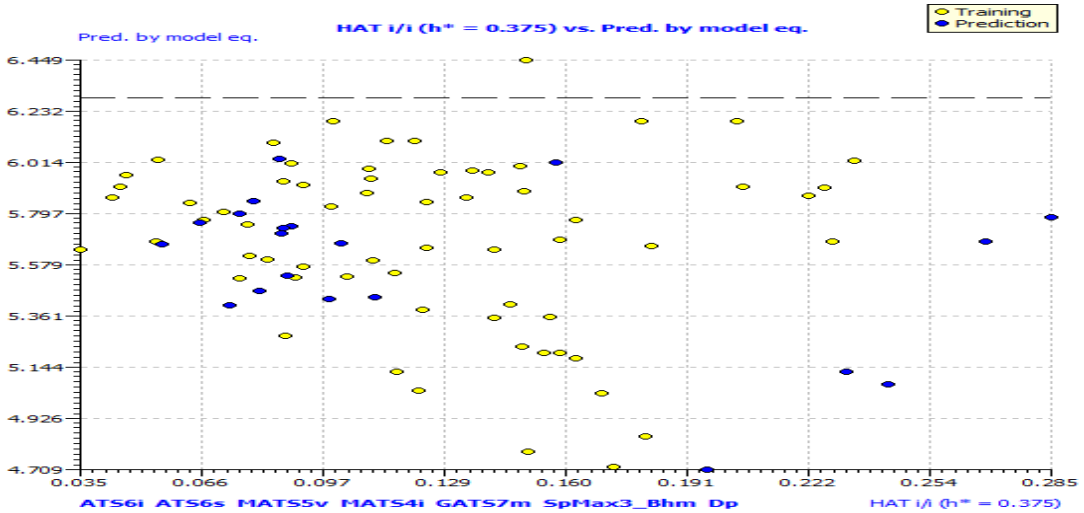

Fig. 4. Scattered Graph of predicted by Model equation against Leverage Values

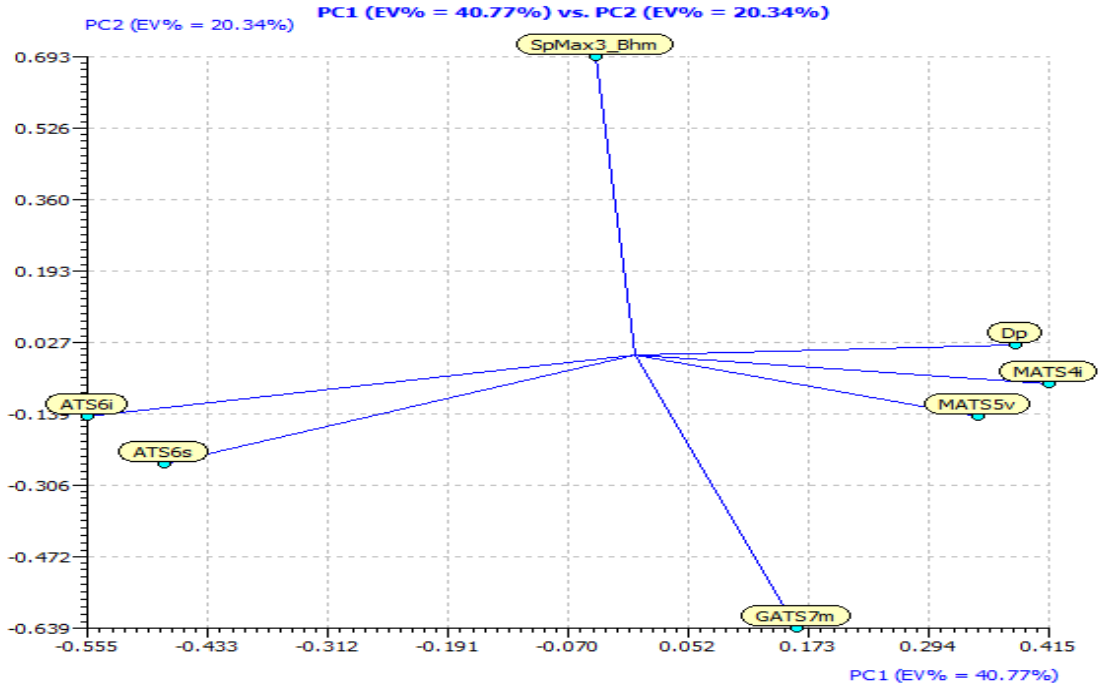

Fig. 5. graph of first principal component (PC1) against second principal component (PC2)

Docking

Analyzing the molecular interactions that lead to the binding of ligands to their target receptors provides valuable insights. To conduct this analysis, four different breast cancer receptor cell lines were considered: 2RMJ, 4OAR, 4RDH, and 3ERT (Figure 6-9). These receptor structures were downloaded from the protein data bank (PDB) and prepared for docking analysis, as shown in Figures 6 to 9. The results of the receptor-ligand interactions with the lowest binding or affinity energy are presented in Figures 10, 11, 12, and 13. Table 2 presents docking results showing the four receptors and their amino acid residue interactions with the ligands. The strength of the interaction between the ligands and the receptor binding pocket is measured by these binding affinities. The analysis shows that different ligands have varying binding affinities with different receptors.

The ligand 28 with the receptor 2RMJ (Table 2 and figure 10) forms hydrogen bonds with LYS 745, LEU 788 and PRO 794, forming a unique interaction driven by electrostatics and hydrogen bonds. It also establishes pi-alkyl bonds with ALA 743 and LEU 718, and a pi-sigma interaction with VAL 726. This complex binding mechanism generates a strong affinity, potentially guiding the development of targeted anti-cancer drugs.

The enzyme 4RDH (Table 2 and Figure 11) exhibits a robust ligand binding affinity of -10 kcal/mol, indicating a strong interaction. This binding is characterized by specific amino acid interactions within the receptor's binding site, including alkyl bonding with LYS 88A, pi-alkyl interaction with ALA 95A, and a pi-sigma bond with ILE 92A (Figure 11). These interactions signify a harmonious fit between the receptor and ligand, crucial for stable binding and potential biological effects.

The enzyme 4OAR and ligand interaction shown in Figure 12 exhibits high binding affinity with amino acids ASN 719 and GLY 722, aided by hydrophobic interactions with MET 759, LEU 726, TRP 735, and LEU 763. Complexity is increased by unusual interactions like pi-sigma with LEU 718 and amide pi-stacked with PHE 778, highlighting its potential therapeutic outcome.

Enzyme 3ERT displayed in Table 2 shows a robust binding affinity of -10 kcal/mol and engages with specific amino acid residues: LEU 354 A through alkyl interaction, TRP 383 via pi-pi stacking, LYS 529 involving pi-cation/anion interaction, LEU 536 through pi-alkyl interaction, and GLY 380 through van der Waals forces as shown in figure 13. These diverse interactions highlight the intricate nature of ligand-receptor binding and suggest potential mechanisms for influencing the receptor's function in the context of breast cancer treatment.

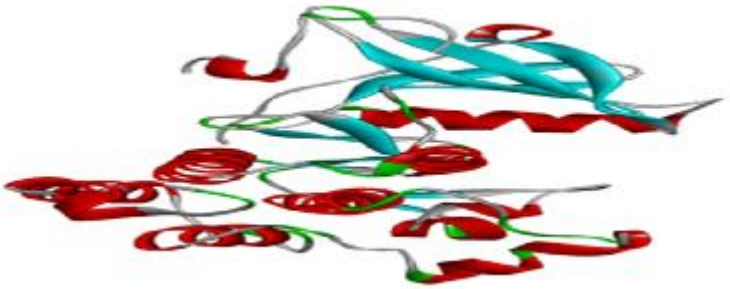

Fig. 5. 3D Prepared Crystals Receptor of 2RMJ

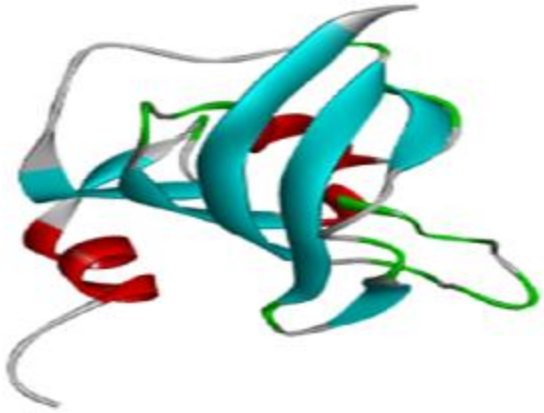

Fig. 6. 3D Prepared Crystal receptor 4RHD

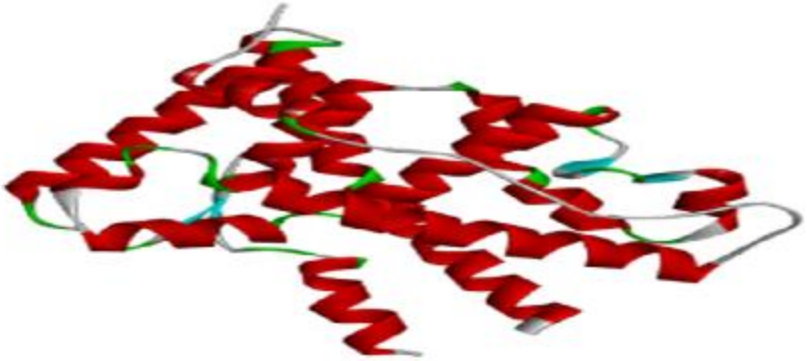

Fig. 7. 3D Prepared Receptor 4AOR

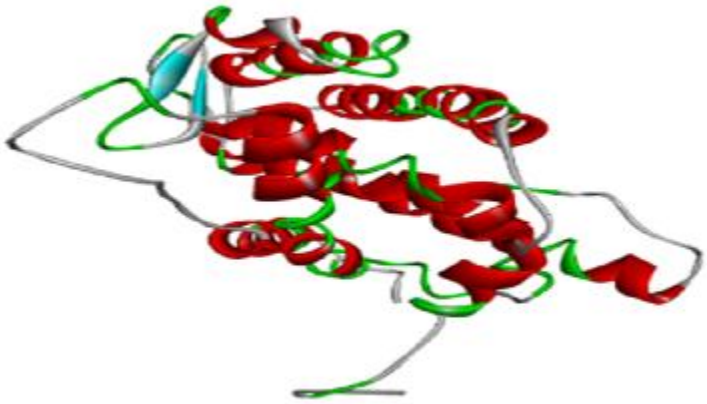

Fig. 8. 3D prepared crystals Receptor 3ERT

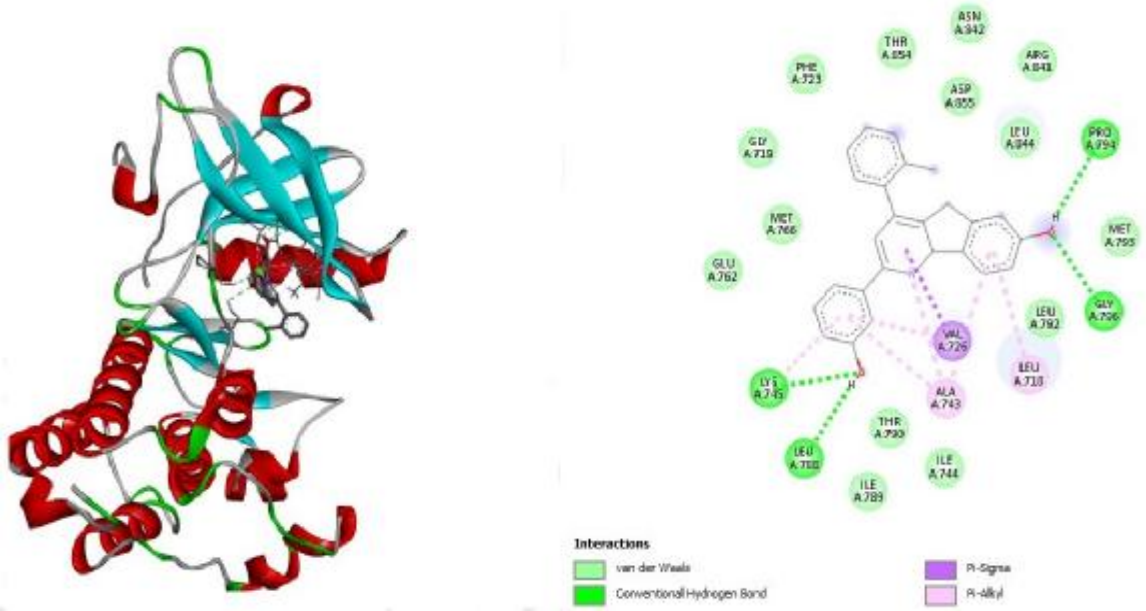

Fig. 9. 3D and 2D ligand protein (2RMJ) interaction with amino acid residues

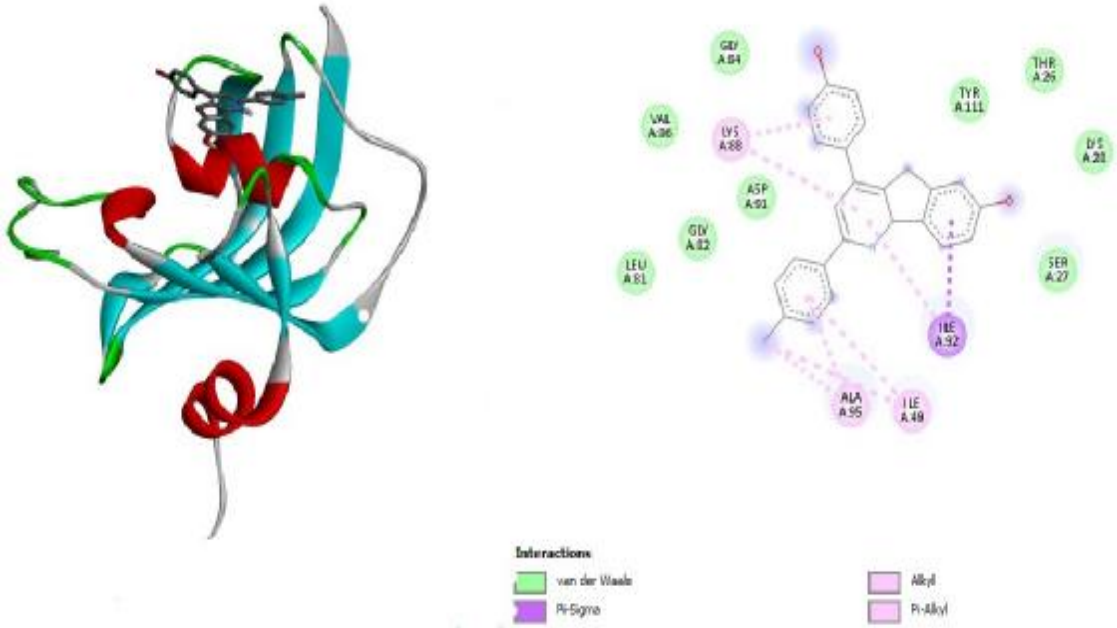

Fig. 10. 3D and 2D Ligand protein(4RDH) interaction with amino acid residues

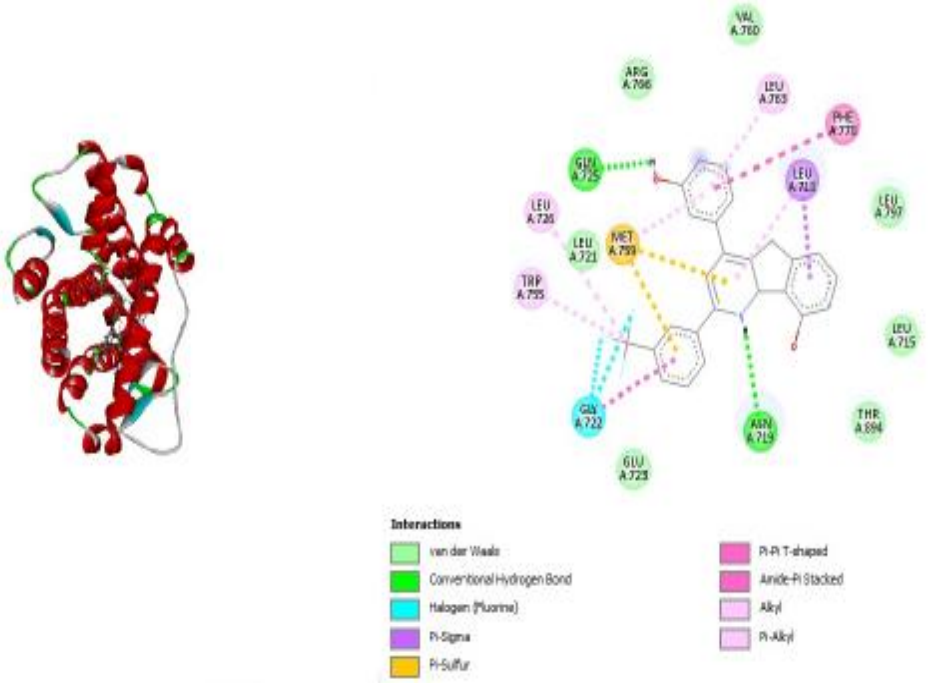

Fig. 11. 3D and 2D ligand protein (4OAR) interaction with amino acid residues

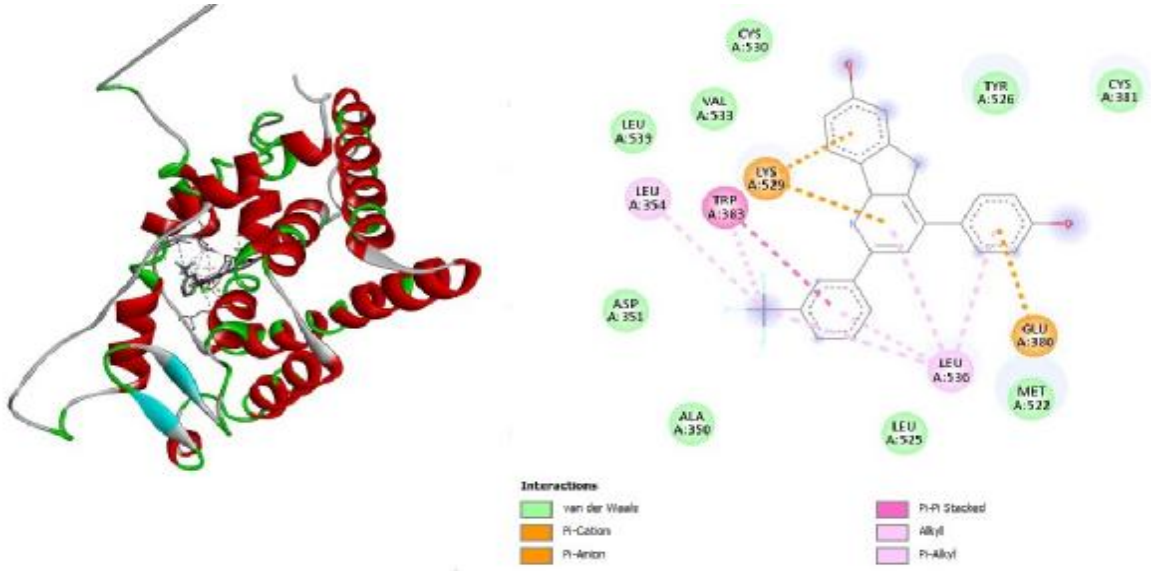

Fig. 12. 3D and 2D ligand protein (3ERT Receptor) interaction

Table 2. Docking result showing the four receptors and their amino acid residue types of interactions with the ligands

Compound ID	Receptor(S)	Types of interaction	Amino-acid residue	Binding score (kcal/mol)
		Hydrogen Bond	LYS745	
		Hydrogen Bond	LEU788	
		Pi-Alkyl	ALA743	
		Pi-Sigma	VAL726	

28	2RMJ	Pi-Alkyl	LEU718	-10.9
		Hydrogen Bond	GLY796	
		Hydrogen Bond	PRO794	
		Alkyl	LEU354	
		Pi-Pi stacked	TRP383	
83	3ERT	Pi-cation/anion	LYS529	-10.0
		Alkyl	LEU536	
		Van der Waal	GLU380	
		Alkyl	LYS88	
83	4RDH	Pi-Alkyl	ALA95	-10.0
		Pi-sigma	ILE92	
		Hydrogen Bond	GLN725	
		Pi-sulfur	MET754	
		Alkyl/ Pi-Alkyl	LEU726	-10.2
84	4OAR	Alkyl/ Pi-Alkyl	TRP735	
		Halogen (fluorine)	GLY722	
		Hydrogen Bond	ASN719	
		Pi-sigma	LEU718	
		Amide Pi-stacked	PHE773	
		Alky	Leu763	

4.3 Pharmacokinetics result

The result of pharmacokinetic in Table 3 of the three compounds reveals crucial observations about their potential pharmaceutical properties and suitability for drug development. However, the table shows the Bioavailability score, molecular weight, gastro intestinal absorption, Lipinski's rule of five, and synthetic accessibility.

Compound ID 84 ($C_{25}H_{16}F_3NO_2$) has a molecular weight of 419.40 g/mol and predominantly consists of 31 aromatic heavy atoms. It exhibits moderate hydrophobicity but displays poor water solubility, suggesting low gastrointestinal absorption. It also shows potential as a P-glycoprotein (P-gp) substrate and interactions with specific cytochrome P450 enzymes. This compound is unlikely to penetrate the blood-brain barrier. Although it adheres to Lipinski's rule of five, it violates Ghose's and Egan's rules. Its bioavailability score is 0.55, with no structural liabilities.

Secondly, Compound ID 28 ($C_{24}H_{16}ClNO_2$) has a molecular weight of 385.84 g/mol. It is categorized as "poorly soluble" and exhibits variable lipophilicity. Pharmacokinetically, it indicates high potential for gastrointestinal absorption but low permeability of the blood-brain barrier. It also interacts with specific cytochrome P450 enzymes and shows potential as a P-gp substrate. While this compound aligns with some druglikeness rules, its lipophilicity varies.

Lastly, Compound ID 83 possesses a moderate molecular weight and a fraction of sp3 hybridized carbon atoms, limiting its rotational flexibility. It features several hydrogen bond acceptors and donors, indicating relevance to binding interactions. However, its lipophilicity suggests solubility challenges. In terms of pharmacokinetics, it demonstrates poor blood-brain barrier permeability, potential as a P-gp substrate, and interactions with specific cytochrome P450 enzymes. Its evaluation against druglikeness rules is mixed, while bioavailability and synthetic accessibility scores guide its potential as a drug candidate..

Table 3 Pharmacokinetics studies table of Bioavailability score, molecular weight

Formula	MW	TPSA	MLOGP	GI	Pgp	CYP2C19	CYP2D6	Ghose	Egan	Muegge	BAS	Synthetic Accessibility
$C_{25}H_{16}F_3NO_2$	419.4	53.35	4.32	Low	Yes	Yes	Yes	1	1	1	0.6	3.21
$C_{24}H_{16}ClNO_2$	385.84	53.35	4.02	High	Yes	Yes	Yes	1	1	1	0.6	3.12
$C_{25}H_{16}F_3NO_2$	419.4	53.35	4.32	Low	Yes	Yes	Yes	1	1	1	0.6	3.25

Conclusion

This research aims to develop efficient treatments for diverse breast cancer types using QSAR modeling and molecular docking. The validated QSAR model suggests promising anti-cancer potential for the proposed compounds. The study signifies a significant advancement in oncology, offering insights for innovative therapies. Ongoing research may enhance outcomes and quality of life for breast cancer patients. Pharmacokinetic profiles provide crucial drug development insights.

In contrast to other research, this research differentiates itself by adopting a computational approach for designing anti-cancer compounds against breast cancer. It incorporates a more extensive dataset and places significant emphasis on the utilization of a QSAR model, along with conducting molecular docking and pharmacokinetics investigations.

Reference

[1] Burstein HJ, Curigliano G, Thürlimann B, Weber WP, Poortmans P, Regan MM, Senn HJ, Winer EP, Gnant M, Aebi S, André F. Customizing local and systemic therapies for women with early breast cancer: the St. Gallen International Consensus Guidelines for treatment of early breast cancer 2021. Ann Oncol. 2021 Oct;32(10):1216-1235.

[2] Wu J, Hicks C. Breast cancer type classification using machine learning. J Pers Med. 2021;11(2):1–12. https://doi.org/10.3390/jpm11020061

[3] Trujillo M, Kharbanda A, Corley C, Simmons P, Allen AR. Tocotrienols as an anti-breast cancer agent. Antioxidants. 2021;10(9):https://doi.org/10.3390/antiox10091383

[4] Wang L, Zhang S, Wang X. The Metabolic Mechanisms of Breast Cancer Metastasis. Front Oncol. 2021;10:https://doi.org/10.3389/fonc.2020.602416

[5] Akram M, Iqbal M, Daniyal M, Khan AU. Awareness and current knowledge of breast cancer. Biol Res. 2017;50(1):https://doi.org/10.1186/s40659-017-0140-9

[6] Tsang JYS, Tse GM. Molecular Classification of Breast Cancer. Adv Anat Pathol. 2020;27(1):https://doi.org/10.1097/PAP.0000000000000232

[7] Shimozuma K. Breast cancer. Jpn J Cancer Chemother. 2019.

[8] Fenichel M. American Cancer Society changes breast cancer screening guidelines to reflect analysis of benefits and harms. JAMA. [published online ahead of print, November 11, 2022]. doi:10.1001/jama.2022.18452.

[9] Smith RA, Saslow D, Sawyer KA, Burke W, Costanza ME, Evans III WP, Foster Jr RS, Hendrick E, Eyre HJ, Sener S. American Cancer Society guidelines for breast cancer screening: update 2003. CA Cancer J Clin. 2003 May;53(3):141-169.

[10] Iwamitsu Y, Shimoda K, Abe H, Tani T, Okawa M, Buck R. The relation between negative emotional suppression and emotional distress in breast cancer diagnosis and treatment. Health Commun. 2005 Nov;18(3):201-215.

[11] Burstein HJ, Curigliano G, Thürlimann B, Weber WP, Poortmans P, Regan MM, Senn HJ, Winer EP, Gnant M, Aebi S, André F. Customizing local and systemic therapies for women with early breast cancer: the St. Gallen International Consensus Guidelines for treatment of early breast cancer 2021. Annals of oncology. 2021 Oct 1;32(10):1216-35.

[12] Momenimovahed Z, Salehiniya H. Epidemiological characteristics of and risk factors for breast cancer in the world. In: Breast Cancer: Targets and Therapy. Vol 11. Published online. doi:10.2147/BCTT.S176070

[13] Umar AB, Uzairu A, Shallangwa GA, Uba S. Ligand-based drug design and molecular docking simulation studies of some novel anticancer compounds on MALME-3M melanoma cell line. Egyptian J Med Hum Genet. Published online 2021. doi:10.1186/s43042-020-00126-9

[14] Ginsburg O, Yip CH, Brooks A, Cabanes A, Caleffi M, Yataco JAD, Gyawali B, McCormack V, de Anderson MML, Mehrotra R, Mohar A, Murillo R, Pace LE, Paskett ED, Romanoff A, Rositch AF, Scheel JR, Schneidman M, Unger-Saldaña K, … Anderson BO. Breast Cancer Early Detection: A Phased Approach to Implementation. Cancer. 2020;126(S10):https://doi.org/10.1002/cncr.32887

[15] Kadayat TM, Park S, Shrestha A, Jo H, Hwang SY, Katila P, Shrestha R, Nepal MR, Noh K, Kim SK, Koh WS. Discovery and biological evaluations of halogenated 2, 4-diphenyl Indeno [1, 2-b] pyridinol derivatives as potent topoisomerase IIα-targeted

chemotherapeutic agents for breast cancer. J Med Chem. 2019 Aug 9;62(17):8194-8234.

[16] Goldberg J, Pastorello RG, Vallius T, Davis J, Cui YX, Agudo J, Waks AG, Keenan T, McAllister SS, Tolaney SM, Mittendorf EA, Guerriero JL. The Immunology of Hormone Receptor Positive Breast Cancer. Front Immunol. 2021;https://doi.org/10.3389/fimmu.2021.674192

[17] Conzen SD. Nuclear receptors and breast cancer. Mol Endocrinol. 2008;https://doi.org/10.1210/me.2007-0421

[18] The American Cancer Society. Breast Cancer Hormone Receptor Status. https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-hormone-receptor-status.html. Published 2021.

[19] Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. QSAR modelling and docking analysis of some thiazole analogues as alfa-glucosidase inhibitors. J Eng Exact Sci. 2019;https://doi.org/10.18540/jcecvl5iss3pp0257-0270

[20] Umar BA, Uzairu A, Shallangwa GA, Uba S. Rational drug design of potent v600e-braf kinase inhibitors through molecular docking simulation. J Eng Exact Sci. 2019;https://doi.org/10.18540/jcecvl5iss5pp0469-0481

[21] Abdullahi SH, Uzairu A, Ibrahim MT, Umar AB. Chemo-informatics activity prediction, ligand based drug design, Molecular docking and pharmacokinetics studies of some series of 4, 6-diaryl-2-pyrimidinamine derivatives as anti-cancer agents. Bull Natl Res Cent. 2021;45(1):https://doi.org/10.1186/s42269-021-00631-w

[22] Umar AB, Uzairu A, Shallangwa GA, Uba S. Docking-based strategy to design novel flavone-based arylamides as potent V600E-BRAF inhibitors with prediction of their drug-likeness and ADMET properties. Bull Natl Res Cent. 2020;https://doi.org/10.1186/s42269-020-00432-7

[23] Ibrahim MT, Uzairu A, Umar BA, Sadiq BA, Isyaku Y. Molecular modelling, docking and pharmacokinetic studies of n-arylidenequinoline-3-carbohydrazides analogs as novel β-glucuronidase inhibitors. J Mex Chem Soc. 2020;https://doi.org/10.29356/jmcs.v64i1.1025

[24] Arthur DE, Uzairu A, Mamza P, Abechi SE, Shallangwa G. Activity and toxicity modelling of some NCI selected compounds against leukemia P388ADR cell line using genetic algorithm-multiple linear regressions. J King Saud Univ - Sci. 2020;32(1):https://doi.org/10.1016/j.jksus.2018.05.023

[25] Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. Molecular docking investigation and pharmacokinetic properties prediction of some anilinopyrimidines analogues as egfr t790m tyrosine kinase inhibitors. Egyptian J Basic Appl Sci. 2021;https://doi.org/10.1080/2314808X.2021.1946650

[26] Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A. In-silico studies of some oxadiazoles derivatives as anti-diabetic compounds. J King Saud Univ - Sci. 2020;32(1):https://doi.org/10.1016/j.jksus.2018.06.006

[27] Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. In-silico activity prediction and docking studies of some 2, 9-disubstituted 8-phenylthio/phenylsulfinyl-9h-purine derivatives as Anti-proliferative agents. Heliyon. 2020;https://doi.org/10.1016/j.heliyon.2020.e03158

[28] Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. Structure-based design of some quinazoline derivatives as epidermal growth factor receptor inhibitors. Egyptian J Med Hum Genet. 2020;https://doi.org/10.1186/s43042-020-00107-y

[29] Umar AB, Uzairu A, Shallangwa GA, Uba S. In silico evaluation of some 4-(quinolin-2-yl)pyrimidin-2-amine derivatives as potent V600E-BRAF inhibitors with pharmacokinetics ADMET and drug-likeness predictions. Future J Pharm Sci. 2020;https://doi.org/10.1186/s43094-020-00084-4

[30] Abdullahi SH, Uzairu A, Shallangwa GA, Uba S, Umar AB. In-silico activity prediction, structure-based drug design, molecular docking and pharmacokinetic studies of selected quinazoline derivatives for their antiproliferative activity against triple negative breast cancer (MDA-MB231) cell line. Bull Natl Res Cent. 2022;https://doi.org/10.1186/s42269-021-00690-z

[31] Umar AB, Uzairu A, Shallangwa GA, Uba S. Ligand-based drug design and molecular docking simulation studies of some novel anticancer compounds on MALME-3M melanoma cell line. Egyptian J Med Hum Genet. 2020;https://doi.org/10.1186/s43042-020-00126-9

[32] Umar BA, Uzairu A, Shallangwa GA. Quantum modeling and molecular dynamic simulation of some amino acids and related compounds on their corrosion inhibition of steel in acidic media. Portugaliae Electrochimica Acta. 2020;https://doi.org/10.4152/pea.202005313

[33] Adeniji SE, Arthur DE, Oluwaseye A. Computational modeling of 4-Phenoxynicotinamide and 4-Phenoxypyrimidine-5-carboxamide derivatives as potent anti-diabetic agent against TGR5 receptor. J King Saud Univ - Sci. 2020;32(1):https://doi.org/10.1016/j.jksus.2018.03.007

[34] Adawara SN, Shallangwa GA, Mamza PA, Abdulkadir I. Chemoinformatic design and profiling of some derivatives of 1, 2, 4-oxadiazole as potential dengue virus NS-5 inhibitors. Bull Natl Res Cent. 2022;46(1):https://doi.org/10.1186/s42269-022-00755-7

[35] Ibrahim MT, Uzairu A, Uba S, Shallangwa GA. Design of more potent quinazoline derivatives as EGFRWT inhibitors for the treatment of NSCLC: a computational approach. Future J Pharm Sci.

2021;7(1):https://doi.org/10.1186/s43094-021-00279-3

[36] Adhikari N, Amin SA, Saha A, Jha T. "Combating breast cancer with non-steroidal aromatase inhibitors (NSAIs): Understanding the chemico-biological interactions through comparative SAR/QSAR study." European Journal of Medicinal Chemistry. 2017 Sep 8;137:365-438.

[37] Speck-Planche A, Kleandrova VV, Luan F, Cordeiro MN. Chemoinformatics in anti-cancer chemotherapy: multi-target QSAR model for the in silico discovery of anti-breast cancer agents. Eur J Pharm Sci. 2012 Aug 30;47(1):273-279.

[38] Wang K, Si H. Prediction of Activities of Halogenated 2, 4-Diphenyl Indeno [1, 2b] pyridinol Derivatives Using QSAR Model Against Breast Cancer. JA