Pharmaceutical and Biomedical Sciences Journal
Volume 7(2), 2025, 192-201

QSAR Analysis and ADMET Prediction of Tamoxifen Derivatives Using LFER Hansch Model

Andzar Fikranus Shofa*, Rawwdha Luthfi Nur Muhammad, Norma Adityaningsih Prameswari, Sendi Handika Putra, Aisyah Afifah Setiawan, Dhia Salsabila Rizki
Departement of Pharmacy, Faculty of Health and Sciences, Syarif Hidayatullah State Islamic University, South Tanggerang, Indonesia 15412

**Corresponding author: andzar@apps.uinjkt.ac.id*

Received: 02 December 2025; Accepted 26 December 2025

Abstract: Tamoxifen is a pharmaceutical compound that can be widely used in the therapeutic regimen for breast cancer, particularly for postmenopausal women. Nevertheless, this clinical efficacy is frequently diminished and limited due to the emergence of drug resistance and heterogeneous therapeutic impacts with different degrees of severity. This has led to the discovery of structurally connected compounds that are more effective than tamoxifen. The current research study describes QSAR modeling and predicting the ADMET parameters of tamoxifen derivatives for the treatment of breast cancer. SPSS software was used for analysis of multiple linear regression and found the pIC50 = -0.059CLogP + 0.759LUMO - 0.011MR + 3.444. Model validation yielded R = 0.921, R^2 = 0.848 and Q^2 = 0.651, which suggests high predictability. The most important characteristic being LUMO energy, the second most important descriptor was followed by CLogP and MR. The ADMET prediction showed high intestinal absorption values (HIA > 90%) and satisfactory permeability over skin. Water solubility was impaired but also low. The metabolism of compounds seemed to predominantly occur via CYP3A4 enzyme. However, LD50 values were of acceptable size, ranging from 324 to 3000 mg/kg (within the safety profile). The findings of this work will thus show help in designing anticancer tamoxifen derivative products with the least toxicity. Additional structural optimization is advocated to achieve maximal therapeutic benefit and least toxicity.

Keywords: ADMET, breast cancer; drug design, QSAR, Tamoxifen

DOI: https://doi.org/10.15408/pbsj.v7i2.49191

1. INTRODUCTION

Breast cancer remains one of the most prevalent malignancies affecting women globally, accounting for a significant proportion of cancer-related morbidity and mortality (Mulyani & Nasution, 2023). Despite advancements in early detection and therapeutic interventions, resistance to conventional treatment options, including chemotherapy and hormone therapy, continues to pose a critical challenge (Chen et al., 2020). Among the therapeutic agents used for breast cancer, Tamoxifen, a selective estrogen receptor modulator (SERM), has been extensively employed due to its ability to modulate estrogen receptor activity (Shagufta & Ahmad, 2018). However, variations in individual metabolic response to Tamoxifen, primarily influenced by the CYP2D6 enzyme, result in diverse therapeutic outcomes (Chen et al., 2020). The need to optimize the therapeutic efficacy of Tamoxifen has driven researchers to explore its structural derivatives, aiming to enhance anti-cancer potency while minimizing adverse effects.

Quantitative Structure-Activity Relationship (QSAR) is considered to be a key method in predicting chemical compounds biological activities from molecular descriptors (Zekri et al., 2020). The computational approach can be used to predict potential drug candidates initially at the early stage of drug discovery, thus minimizing the time- and resource-consuming experimental validation process (Lipinski, 2011). QSAR is used for screening Tamoxifen and its derivatives to elucidate correlation between chemical structure and biological activity and guide synthesis of more effective analogs (Shagufta & Ahmad, 2018). However, the accuracy of the QSAR models is affected by the quality and size of training data and complexity of biological endpoints, which may result in conflicting predictions

from different QSAR methods (Zekri et al., 2020). As a solution to these limitations, several validation strategies such as internal validation, external validation, and Y-randomization are suggested to make the QSAR models more robust and powerful for predictive purposes (Lipinski, 2011).

Although QSAR models can be of great importance, they do face many challenges like the inherent variation of biological systems, especially in predicting complex endpoints like pharmacokinetics-toxicity, and offer a comprehensive strategy to evaluate the safety and efficacy of Tamoxifen derivatives (Zekri et al., 2020). As we discussed earlier, QSAR was utilized for dose optimization for Tamoxifen and its analogs (such as Z-endoxifen), a metabolite of this compound, which also plays a key role in therapeutics (Chen et al., 2020). It has been found that the efficacy of Tamoxifen may be highly dependent on the patient's genetic polymorphism in CYP2D6, and treatment must therefore be adjusted according to dose (Shagufta & Ahmad, 2018). Moreover, the hepatotoxic potential and mutagenicity of Tamoxifen derivatives require meticulous characterization of their ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) profiles, which is where QSAR modeling is useful (Lipinski, 2011).

Previous studies have explored the use of various Tamoxifen derivatives, including Raloxifene, Droloxifene, and Idoxifene, which have shown potential anti-cancer activities with different pharmacological profiles (Shagufta & Ahmad, 2018). These derivatives are often evaluated using QSAR techniques to understand their interaction with estrogen receptors and predict their pharmacokinetic behavior. However, conflicting results regarding their efficacy and toxicity profiles persist, primarily due to differences in experimental conditions, computational methods, and sample sizes (Lipinski, 2011). Addressing these discrepancies requires further research, especially in validating QSAR models across diverse datasets and experimental frameworks to achieve more consistent and reliable predictions (Zekri et al., 2020).

Given these research gaps, this study aims to further explore the potential of QSAR modeling in predicting the pharmacokinetic and toxicity profiles of Tamoxifen derivatives. By analyzing a broader set of molecular descriptors and integrating comprehensive ADMET predictions, this research seeks to identify novel derivatives with enhanced therapeutic potential and minimized adverse effects. The novelty of this study lies in its approach to combine computational modeling with extensive dataset validation, aiming to refine the predictive power of QSAR models and contribute to the advancement of precision medicine in breast cancer treatment to provide a more accurate, efficient, and personalized approach to breast cancer therapy through the development of optimized Tamoxifen analogs.

2. MATERIAL AND METHODS

2.1 Materials

Table 1. Chemical Structures and IC_{50} Values of Tamoxifen and Its Derivatives Againts Breast Cancer

No	Compoud	Chemical Formula	IC_{50} (µM)
1	Tamoxifen	$C_{26}H_{29}NO$	20.5
2	Raloxifene	$C_{28}H_{27}NO_4S$	13.7
3	Tamoxifene A	$C_{23}H_{23}NO$	16.9
4	Tamoxifene B	$C_{27}H_{29}NO_3$	2.9
5	Idoxifene	$C_{28}H_{30}INO$	6.5
6	Toremifene	$C_{26}H_{28}ClNO$	18.9
7	Ospemifene	$C_{24}H_{23}ClO_2$	12.6
8	4-hydroxy tamoxifene	$C_{26}H_{29}NO_2$	11.3
9	Lasofoxifene	$C_{28}H_{31}NO_2$	0.00288
10	Endoxifen	$C_{25}H_{27}NO_2$	0.675

The research utilized the following tools and materials: a laptop with an Asus brand, a Windows 10 operating system, a 64-bit AMD Quad Core A8-7410 processor, a CPU @ 2.5 GHz, and 4.00 GB of

RAM. The software employed included Chem Draw Ultra 12.0 and Chem Bio 3D Ultra (CambridgeSoft), as well as IBM SPSS Statistics 27.0.1. The data on tamoxifen-derived compounds and the IC50 values of breast anticancer activity were obtained from research conducted by Brauch and Jordan (2009), as shown in Table 1 .

2.2 Methods

a. Research Design

The study performs a computational and in silico study to characterize structure-activity relationships and pharmacokinetic properties of tamoxifen derivatives. Computational methods are favored due to their cost-efficiency, reduced experimental hazards, and the ability to screen multiple compounds simultaneously (Zekri et al., 2020). This design is exploratory in nature by correlating molecular descriptors to biological activities. The QSAR analysis follows a supervised learning approach through multilinear regression (MLR) which is suitable for the analysis of complex data involving multiple predictors (Frimayanti et al., 2011). ADMET predictions necessary to understand drug-like properties were performed with pkCSM and Protox III and were accessible through their respective online platforms (Pires et al., 2015).

b. Data Collection

The first phase of the study encompassed collecting applicable chemical and biological information from PubChem, a trusted, open-access chemical database. A text notation of chemical structures known as SMILES was obtained for each tamoxifen derivative. These SMILES notations were entered into ChemDraw and Chem3D to compute molecular descriptors: lipophilicity (CLogP), electronic properties (LUMO and HOMO), steric parameters (molecular refractivity, MR), polar surface area (TPSA). Furthermore, pkCSM was also used for ADMET property prediction (absorption, distribution, metabolism, excretion, and toxicity) (Pires et al., 2015). The retrieved data were collated into organized datasets for statistical analysis to ensure data integrity and completeness. Descriptors were standardized and normalized to mitigate bias caused by differences in scale in order to enhance reliability of the data. Data preprocessing also took into account missing values, the detection of outliers, and the consistency of descriptor calculation. The descriptors with high multicollinearity (VIF > 5) were removed in order to avoid redundancy and improve interpretation of models (Nystrom & Sanchez, 2011).

c. QSAR Model Development

The QSAR modeling process follows these steps: (1) data acquisition and preprocessing, (2) descriptor calculation using ChemDraw and Chem3D, (3) statistical analysis using SPSS, and (4) model validation through internal and external techniques. Internal validation employs leave-one-out (LOO) cross-validation, while external validation uses independent datasets (Frimayanti et al., 2011; Putra et al., 2023). The evaluation parameters include correlation coefficient (R), determination coefficient (R^2), and cross-validated R^2 (Q^2), adhering to the threshold values specified in previous QSAR studies..

d. ADMET Predictions

ADMET predictions are critical for assessing the pharmacokinetic properties of the compounds. Absorption metrics include solubility, Caco-2 permeability, human intestinal absorption (HIA), and skin permeability. Distribution is assessed via VDss (volume of distribution) and BBB (blood-brain barrier) permeability. Metabolism focuses on enzyme interaction, specifically with CYP450 enzymes (CYP2D6 and CYP3A4). Excretion evaluates total clearance and renal OCT2 substrates, while toxicity considers hepatotoxicity and carcinogenicity using Protox III (Pires et al., 2015).

e. Statistical Analysis

Statistical analysis was performed to measure the relationships between the molecular characteristics and biological activity of tamoxifen derivatives in this study. Multilinear regression (MLR) and Pearson's correlation were used to conduct this analysis (using IBM SPSS Statistics 25). MLR

performed well in estimating how multiple independent variables contributed to a dependent variable, allowing a comprehensive understanding of how molecular descriptors influence biological activity (Nystrom & Sanchez, 2011). To examine the strength and direction of the relationship across molecular descriptors, Pearson's correlation was employed. A correlation matrix was created to identify the linear relationships among the variables, ensuring that multicollinearity did not influence the regression model. The degree of multicollinearity was determined with the variance inflation factor (VIF); a VIF value above 5 indicates multicollinearity issues that potentially disrupt the model's accuracy (Zekri et al., 2020). The significance parameters of predictor variables were evaluated by p-values (<0.05) indicating that only significant descriptors were included. Using coefficient of determination (R^2) and adjusted R^2 to model multiple predictors, goodness of fit of the regression model was analyzed. Finally, standardized coefficients (β) were examined in order to understand the relative importance of each descriptor in predicting biological activity. Residual analysis including inspection of standardized residuals and Cook's distance was performed for potential outliers or influential data points.

f. Model Validation

Model validation is crucial to ensure the credibility of QSAR predictions. Internal validation through LOO cross-validation was used to evaluate the model's stability. Additionally, external validation was performed using independent datasets to verify the model's predictability. The criteria for model acceptability included $R^2 > 0.6$, $Q^2 > 0.5$, and F-statistics surpassing the threshold (Hadaji et al., 2018).

3. RESULTS AND DISCUSSION

The QSAR analysis was performed on ten tamoxifen derivatives to establish the correlation between molecular descriptors and anticancer activity against breast cancer. The QSAR model was constructed using pIC50 as the dependent variable, while descriptors such as lipophilicity (CLogP, LogS), electronic parameters (HOMO, LUMO, pKa, tPSA), and steric factors (MR, MW, MV) were considered as independent variables (Table 2)

Table 2. Molecular Descriptors of Lipophilic, Electronic, and Steric Properties of Selected Tamoxifen Derivatives

No	Compoud	pIC50	CLogP	LogS	HOMO	LUMO	pKa	tPSA	MR	MW	MV
1	Tamoxifen	-1.31	5.77	-7.05	-9.48	-4.79	5.77	12.47	119.72	371.51	168.65
2	Raloxifene	-1.13	5.05	-6.88	-7.08	-3.60	8.40	98.24	141.21	473.58	202.40
3	Tamoxifene A	-1.23	4.77	-5.91	-9.97	-4.56	9.94	35.35	105.11	329.43	149.13
4	Tamoxifene B	-0.46	5.09	-6.81	-8.82	-4.43	9.22	69.56	127.64	414.52	177.07
5	Idoxifene	-0.81	6.71	-8.83	-9.33	-4.30	8.99	12.47	143.85	523.45	199.63
6	Toremifene	-1.28	5.98	-7.01	-9.48	-4.43	9.23	12.47	124.52	405.96	178.95
7	Ospemifene	-1.1	5.41	-6.43	-9.96	-4.42	14.18	29.46	113.17	378.89	165.28
8	4-hydroxy tamoxifene	-1.05	5.36	-6.74	-9.48	-4.19	9.17	32.70	121.75	387.51	173.44
9	Lasofoxifene	2.54	5.2	-6.79	-9.33	0.27	9.40	32.70	130.09	413.55	184.85
10	Endoxifen	0.17	5.09	-6.67	-9.99	-4.21	9.17	41.49	116.84	373.49	166.87

To refine the QSAR model, a Pearson correlation analysis was performed (**Table 3**). A significant positive correlation between LUMO and pIC50 (r = 0.916) indicates that molecules with higher LUMO energy exhibit increased anticancer potency. Conversely, CLogP exhibited a weak negative correlation (r = -0.208), suggesting that excessive lipophilicity could reduce biological activity. The MR descriptor also showed a minor positive correlation (r = 0.169), indicating its minor influence on potency.

Among the molecular descriptors, LUMO (Lowest Unoccupied Molecular Orbital) demonstrates the strongest positive correlation (0.916) with pIC50, emphasizing its significant role in predicting anticancer activity. A higher LUMO value suggests increased electron affinity, enhancing the potential for nucleophilic attacks, which is critical for interaction with estrogen receptors (Zhang et al., 2010). This supports the hypothesis that tamoxifen derivatives with enhanced LUMO energy may achieve stronger binding to receptor sites, thus exhibiting superior anticancer activity (Chen et al., 2020).

Table 3. Pearson Correlation Matrix between pIC_{50} and Selected Physicochemical and Electronic Descriptors of Tamoxifen Derivatives.

QSAR Descriptor	pIC_{50}	CLogP	LogS	HOMO	LUMO	pKa	tPSA	MR	MW	MV
pIC_{50}	1									
CLogP	**-0.208**	1								
LogS	0.023	-0.883	1							
HOMO	-0.072	-0.113	-0.167	1						
LUMO	**0.916**	-0.189	0.037	0.142	1					
pKa	0.024	-0.168	0.29	-0.311	0.022	1				
tPSA	0.025	-0.618	0.286	0.781	0.09	-0.01	1			
MR	**0.169**	0.493	-0.77	0.704	0.282	-0.318	0.285	1		
MW	0.055	0.603	-0.848	0.588	0.149	-0.147	0.191	0.96	1	
MV	0.164	0.462	-0.714	0.732	0.309	-0.246	0.312	0.987	0.945	1

Conversely, the negative correlation of CLogP (-0.208) suggests that excessive lipophilicity may compromise drug bioavailability and increase toxicity risks. An optimal balance of hydrophilicity and lipophilicity is crucial for effective membrane permeability and pharmacokinetic performance (Megawati et al., 2023). These findings are consistent with previous research indicating that over-lipophilicity often leads to poor pharmacokinetics and off-target interactions, potentially elevating toxicity (Mulyani & Nasution, 2023). Additionally, MR (Molar Refractivity) demonstrates a weaker correlation (0.169), indicating a less prominent impact of steric hindrance in influencing biological activity. However, MR's contribution cannot be disregarded, as slight modifications in steric properties can still impact receptor binding and overall molecular stability (Putra et al., 2023).

The QSAR model was validated using the Leave-One-Out (LOO) cross-validation technique, yielding a Q^2 value of 0.651, meeting the acceptance threshold ($Q^2 > 0.5$). Additionally, the statistical significance of the model was verified with an F-value of 11.137, surpassing the F-critical value of 4.45 at a 99% confidence level. The strong correlation and predictive reliability of the model confirm its applicability in designing more effective tamoxifen derivatives.

The validated QSAR equation obtained was:

$$pIC50 = -0.059CLogP + 0.759LUMO - 0.011MR + 3.444$$
$$n = 10,\ r = 0.921,\ R^2 = 0.848,\ F = 11.137,\ Q^2 = 0.651$$

The validation through the Leave-One-Out (LOO) cross-validation method, yielding $Q^2 = 0.651$, demonstrates the model's reliability while acknowledging its limitations. The moderate Q^2 value implies that other influential factors, potentially unaddressed molecular descriptors or complex biological interactions, may contribute to the variations in anticancer activity. These findings align with existing literature highlighting the intricate nature of QSAR modeling for complex biological responses (Frimayanti et al., 2011; Male et al., 2019). Future studies should consider integrating additional descriptors, such as quantum chemical or three-dimensional structural descriptors, to refine predictive power.

The model showed strong statistical validation with $n = 10$, $r = 0.921$, $R^2 = 0.848$, $F = 11.137$, and $Q^2 = 0.651$, indicating high predictive power and stability. Among the descriptors, LUMO energy demonstrated the highest correlation with pIC50 ($r = 0.916$), followed by CLogP (-0.208) and MR (0.169). indicates the complex interplay between lipophilicity (CLogP), electronic properties (LUMO), and steric factors (MR) in determining biological activity. The correlation coefficient ($R^2 = 0.848$) suggests a robust model, capable of predicting the anticancer efficacy of tamoxifen derivatives with considerable accuracy (Arba et al., 2018). These findings suggest that increasing LUMO energy while reducing CLogP and MR values may enhance anticancer activity.

The LUMO energy has the largest positive coefficient, indicating that the electronic features of the molecules are central in determining inhibitory potency. Higher LUMO energy, reflecting a lower tendency of the molecule to accept electrons, may lead to enhanced specificity and stability in the ligand–receptor interactions, especially in targets where charge transfer or electrostatic complementarity is crucial. This is in line with earlier QSAR and quantum chemical analyses indicating that the frontier molecular orbitals, particularly LUMO, are a major determinant in the ligand reactivity and binding affinity in the enzyme–inhibitor and receptor–ligand systems (Parr & Yang, 1989; Cherkasov et al., 2014).

The CLogP reflects the lipophilicity and enters activity negatively, as evidenced from the negative regression coefficient. Lipophilicity is traditionally a major driver of membrane permeability and target engagement; however, excessive hydrophobicity can be deleterious by promoting nonspecific interactions with either biological membranes or plasma proteins, ultimately decreasing the effective concentration of free ligand at the target site. This negative influence of CLogP in this model suggests that a ligand with moderate or low hydrophobic character may be preferred in the binding site of the biological target, possibly because of the presence of polar or charged amino acid residues within the binding pocket. This behavior is in good agreement with the classical Hansch theory, which advocates an optimum lipophilicity window beyond which the biological activity decreases (Hansch & Fujita, 1964; Kubinyi, 1993).

The negative contribution to activity from the steric and polarizability effects is captured by the MR descriptor with a relatively small coefficient. MR reflects both molecular volume and electronic deformability; thus, bulkier and more polarizable molecules tend to show slightly reduced biological activity. This may imply that steric hindrance could restrict the ability of larger ligands to assume an optimal binding conformation within the active site, assuming that the pocket of the target is spatially constrained. Similar observations have been reported in QSAR studies where increases in molecular size adversely affected binding efficiency due to unfavorable steric clashes or reduced conformational flexibility (Hansch et al., 1995; Todeschini and Consonni, 2009). Taking into consideration this negative MR contribution supports the idea of compact molecular frameworks with controlled polarizability being preferable for achieving higher potency within this ligand series (OECD, 2014; Roy et al., 2015).

The ten tamoxifen derivatives studied based on ADMET properties: absorption, distribution, metabolism, excretion and toxicity have the best absorption of ten tamoxifen derivatives. Most of the compounds had good skin permeability (Log KP > -2.5), high intestinal absorption (HIA > 90%) and good skin permeability based on absorption analysis. Meanwhile all substances exhibited a low solubility aqueously, but the better solubility was achieved by raloxifene (-3.716 log mol/L). Most of the included compounds also showed high Human Intestinal Absorption (HIA > 90%), suggesting good absorption and possibility of effective oral bioavailability. Still, poor aqueous solubility continues as an important barrier, which may block systemic circulation and requirement for modifications for formulations (Pires et al., 2015). Notwithstanding effective skin penetration (Log KP > −2.5), transdermal administration might not be the ideal method for the systemic anticancer therapy (Nystrom & Sanchez, 2011). The Caco-2 permeability values varied greatly across the derivatives, with ospemifene providing the highest permeability (1.178 x 10^{-6} cm/s). The derivatives identified as P-glycoprotein (P-gp) substrates and inhibitors and multidrug resistance mechanisms. Distribution of VDss varied widely with idoxifene with the largest VDss (1.056 log L/kg), confirming wide tissue distribution. The BBB permeability (log BB) ranged from -1.039 (raloxifene) to 1.329 (tamoxifen) indicating that certain derivatives were sufficient to cross the blood-brain barrier, as per the latter of which the CNS may take influence. Yet, CNS permeability (log PS) values were all < -2, suggesting that none of the compounds can penetrate the CNS effectively. All derivatives were metabolized by CYP3A4, and most were CYP1A2 inhibitors, suggesting possible drug-drug interactions. Interestingly, raloxifene and ospemifene were not CYP2D6 inhibitors, which suggests a lower likely effect of CYP2D6-metabolized medications on CYP2D6-metabolized drug. The total values of clearance ranged from 0.192 (ospemifene) to 0.81 log ml/min/kg (lasofoxifene), suggesting differential elimination rates. None of these compounds were kidney OCT2 substrates, indicating that they showed a slight interaction

with renal excretion.

Metabolically, most tamoxifen derivatives serve as substrates for CYP3A4, underlining the enzyme's pivotal role in their metabolic clearance. Notably, Raloxifen and Ospemifen, which do not inhibit CYP2D6, present a reduced risk of adverse drug interactions, highlighting their potential for safer use in polypharmacy settings (Chen et al., 2020). These findings are consistent with prior studies that emphasize the role of CYP3A4 and CYP2D6 in the metabolism of selective estrogen receptor modulators (SERMs) (Brøsen & Naranjo, 2001). The varying clearance rates, particularly Lasofoxifen's high clearance (0.81 log ml/min/kg), imply a narrower therapeutic window, necessitating careful therapeutic monitoring and dosage adjustments (Krihariyani et al., 2020).

Toxicity predictions revealed that all derivatives were non-carcinogenic, but some exhibited hepatotoxicity (tamoxifen B, toremifene, 4-hydroxy tamoxifen, and tamoxifen A). The LD50 values ranged from 324 mg/kg (lasofoxifene) to 3000 mg/kg (tamoxifen B), indicating a generally acceptable safety profile. Most compounds fell within toxicity level IV, with tamoxifen B categorized as level V, suggesting the highest safety margin. Toxicological assessments indicate minimal carcinogenic risks across the compounds, aligning with their current clinical applications. However, hepatotoxicity observed in Toremifen and 4-hydroxy tamoxifen necessitates hepatic function monitoring during therapeutic use. The wide LD50 range (324–3000 mg/kg) implies an overall acceptable safety profile, yet further in vivo validation is essential for clinical translation (Banerjee et al., 2024). The identification of hepatotoxic risk factors also encourages structural modifications to reduce off-target toxicity while maintaining therapeutic efficacy (Widiyanti et al., 2021).

Table 4. In Silico Prediction of Absorption-Distribution Profiles for Tamoxifen and Its Derivatives

	Absorption							Distribution			
Compound	**Water Solubility (log mol/L)**	**Caco2 Permeability (log Papp in 10^{-6} cm/s)**	**HIA**	**Skin Permeability (Log KP)**	**P-gp substrate**	**P-gp I Inhibitor**	**P-gp II Inhibitor**	**VDss (log L/kg)**	**Unbound Fraction (fu)**	**BBB Permeability (log BB)**	**CNS Permeability (log PS)**
Tamoxifen	-5.929	1.065	96.885	-2.737	yes	yes	yes	0.83	0.093	1.329	-1.473
Raloxifene	-3.716	0.77	93.522	-2.735	yes	yes	yes	-1.492	0.114	-1.039	-1.932
Tamoxifene A	-5.816	1.058	97.095	-2.737	yes	yes	yes	0.786	0.095	1.072	-1.499
Tamoxifene B	-5.043	0.872	95.553	-2.753	yes	yes	yes	0.632	0.124	-0.16	-0.914
Idoxifene	-6.165	1.058	94.996	-2.727	yes	yes	yes	1.056	0.064	1.295	-1.945
Toremifene	-6.226	0.971	96.48	-2.735	yes	yes	yes	0.36	0.123	1.273	-1.342
Ospemifene	-6.527	1.178	95.648	-2.736	yes	yes	yes	-0.132	0.131	1.07	-1.418
4-hydroxy tamoxifene	-5.099	1.026	93.541	-2.735	yes	yes	yes	0.305	0.008	-0.292	-1.278
Lasofoxifene	-5.471	0.895	96.001	-2.737	yes	yes	yes	-0.308	0.078	-0.302	-1.539
Endoxifen	-5.012	1.02	93.751	-2.735	yes	yes	yes	0.258	0.01	-0.401	-1.303

The rational design of next-generation tamoxifen analogs necessitates an early, integrated ADMET strategy to balance absorption, distribution, metabolism, excretion, and toxicity. This strategy involves fine-tuning key physicochemical and electronic properties to maximize efficacy and safety (Xiao et al.,

2024; Wang, Clark & Ma'ayan, 2016). In accordance with Lipinski's rule of five, it is imperative to maintain calculated lipophilicity (cLogP) below 5 to ensure adequate membrane permeability and oral bioavailability (Lipinski et al., 2001). Furthermore, it is essential to optimize molar refractivity (MR) between 40 and 130 to support favorable van der Waals interactions without exacerbating hydrophobic toxicity. The hepatotoxicity of tamoxifen, which has been extensively documented, encompasses a range of adverse effects, including steatosis, cholestasis, and, in rare instances, peliosis hepatis.

Table 5. In silico prediction of metabolism-excretion profiles for tamoxifen and its derivatives

Compound	Metabolism						Excretion	
	CYP2D6 Substrate	CYP3A4 Substrate	CYP1A2 Inhibitor	CYP2C9 Inhibitor	CYP2D6 Inhibitor	CYP3A4 Inhibitor	Total Clearence (log ml/min/kg)	Renal Substrate OCT2
Tamoxifen	No	Yes	Yes	No	Yes	No	0.556	No
Raloxifene	No	Yes	Yes	Yes	No	Yes	0.746	No
Tamoxifene A	No	Yes	Yes	No	Yes	Yes	0.661	No
Tamoxifene B	No	Yes	Yes	No	Yes	No	0.685	No
Idoxifene	No	Yes	Yes	No	Yes	No	0.726	No
Toremifene	No	Yes	Yes	No	Yes	No	0.587	No
Ospemifene	No	Yes	Yes	No	No	Yes	0.192	No
4-hydroxy tamoxifene	No	Yes	Yes	No	Yes	No	0.594	No
Lasofoxifene	No	Yes	Yes	No	Yes	No	0.81	No
Endoxifen	No	Yes	Yes	No	Yes	No	0.7	No

The mechanism underlying this toxicity is primarily attributed to the generation of reactive metabolites by CYP2D6, which induce mitochondrial dysfunction, oxidative stress, and hepatocyte apoptosis (Chitturi & Farrell, 2013; Zhao et al., 2014). At the molecular-orbital level, the energy of the lowest unoccupied molecular orbital (LUMO) is reduced by electron-withdrawing substituents, thereby enhancing the stacking of pi-pi bonds and the formation of hydrogen bonds with aromatic residues in the ERα binding pocket. This, in turn, strengthens the binding of antagonists while preserving cLogP and MR within optimal ranges through iterative in silico docking and ADMET modeling (Zhang et al., 2023; Yuan et al., 2023).

Table 6. In silico toxicity assessment of tamoxifen and its Derivatives

Compoud	Toxicity			
	LD50 Prediction (mg/kg)	Toxicity level	Carcinogenicity	Hepatotoxicity
Tamoxifen	1190	IV	No	Yes
Raloxifene	1000	IV	No	No
Tamoxifene A	1190	IV	No	Yes
Tamoxifene B	3000	V	No	Yes
Idoxifene	1700	IV	Yes	No
Toremifene	1700	IV	No	Yes
Ospemifene	1700	IV	No	No
4-hydroxy tamoxifene	1190	IV	No	Yes
Lasofoxifene	324	IV	No	No
Endoxifen	1190	IV	No	No

4. CONCLUSION

The findings of this study reveal a significant relationship between molecular parameters and the anticancer activity of tamoxifen derivatives against breast cancer. The validated QSAR equation, pIC50

= -0.059CLogP + 0.759LUMO - 0.011MR + 3.444, highlights the influence of LUMO energy as the most impactful molecular descriptor, followed by CLogP and MR. Increasing LUMO while reducing CLogP and MR values may enhance the anticancer potential of these derivatives. ADMET predictions indicate that most derivatives exhibit high intestinal absorption (HIA > 90%), low water solubility, and adequate skin permeability. While metabolism primarily involves the CYP3A4 enzyme, raloxifene and ospemifene do not affect CYP2D6, minimizing potential drug interactions. The hepatotoxic potential of certain derivatives suggests the need for careful consideration in further drug development. Despite these risks, the LD50 values remain within a safe range (324–3000 mg/kg), suggesting acceptable safety profiles.

5. REFERENCES

Arba, M., Ruslin., Nursan., Maulidiyah., Tjahjono, D.H., 2018. Hubungan kuantitatif struktur-aktivitas (HKSA) dan penambatan molekul senyawa turunan benzamida sebagai inhibitor alosterik mitogen enhanced kinase (MEK). Jurnal Kimia VALENSI, 4(1), pp.42–51.

Banerjee, P., Kemmler, E., Dunkel, M. and Preissner, R., 2024. ProTox 3.0: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Research, 52(W1), pp.W513–W520. https://doi.org/10.1093/nar/gkae303

Brauch, H. and Jordan, V.C., 2009. Targeting of tamoxifen to enhance antitumour action for the treatment and prevention of breast cancer: the "personalised" approach?. European Journal of Cancer, 45(13), pp.2274–2283. https://doi.org/10.1016/j.ejca.2009.05.032.

Brøsen, K. and Naranjo, C.A., 2001. Review of pharmacokinetic and pharmacodynamic interaction studies with citalopram. European Neuropsychopharmacology, 11(4), pp.275–283.

Chander, S., Tang, C.R., Al-Maqtari, H.M., Jamalis, J., Penta, A., Hadda, T.B., et al., 2017. Synthesis and study of anti-HIV-1 RT activity of 5-benzoyl-4-methyl-1,3,4,5-tetrahydro-2H-1,5-benzodiazepin-2-one derivatives. Bioorganic Chemistry, 72, pp.74–79.

Chen, Q., et al., 2020. Targeting tamoxifen metabolites in the treatment of breast cancer. Journal of Medicinal Chemistry, 63(5), pp.2320–2330. https://doi:10.1021/jm20200767.

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R. and Consonni, V., 2014. QSAR modeling: where have you been? Where are you going to?. Journal of Medicinal Chemistry, 57(12), pp.4977–5010.

Chitturi, S. and Farrell, G.C., 2013. Estrogen receptor antagonists: adverse effects of hormones and hormone antagonists on the liver. In: Kaplowitz, N. and DeLeve, L.D., editors. Drug-Induced Liver Disease. 3rd edn. Elsevier, pp.610–612.

Frimayanti, N., Yam, M.L., Lee, H.B., Othman, R., Zain, S.M. and Rahman, N.A., 2011. Validation of quantitative structure-activity relationship (QSAR) model for photosensitizer activity prediction. International Journal of Molecular Sciences, 12(12), pp.8626–8644.

Hadaji, N., Kenouche, S. and Saleh, B.A., 2018. QSAR modeling, docking, ADME and reactivity of indazole derivatives as antagonizes of estrogen receptor alpha (ER-α) positive in breast cancer. Journal of Molecular Structure, 1217, p.128442.

Hansch, C. and Fujita, T., 1964. ρ–σ–π analysis. A method for the correlation of biological activity and chemical structure. Journal of the American Chemical Society, 86(8), pp.1616–1626.

Hansch, C., Leo, A. and Hoekman, D., 1995. Exploring QSAR: hydrophobic, electronic, and steric constants. Washington, DC: American Chemical Society.

Krihariyani, D., Sasongkowati, R. and Haryanto, E., 2020. Studi in siliko sifat farmakokinetik, toksisitas, dan aktivitas imunomodulator brazilein kayu secang terhadap enzim 3-chymotrypsin-like cysteine protease coronavirus. Journal of Indonesian Medical Laboratory and Science, 1(1), pp.76–90.

Kubinyi, H., 1993. QSAR: Hansch analysis and related approaches. Weinheim: VCH Publishers.

Lipinski, C.A., 2011. ADMET screen. In: Schwab, M., editor. Encyclopedia of Cancer. Springer Berlin Heidelberg, pp.80–84. https://doi:10.1007/978-3-642-16483-5_114.

Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 46(1–3), pp.3–26.

Male, Y., Sutapa, I.W. and Pusung, Y.A.D., 2018. Prediksi potensi antikanker senyawa turunan xanthon

menggunakan hubungan kuantitatif struktur dan aktivitas (HKSA). Chemistry Progress, 11(1), pp.28–37.

Megawati, S.D., Ekowati, J. and Siswandono, S., 2023. Quantitative structure-activity relationship (QSAR) of N-benzoyl-N'-naphtylthiourea derivative compounds by in silico as anticancer through inhibition of VEGFR2 receptors. ICGT 2022, pp.137–148.

Mulyani, F. and Nasution, H., 2023. Application of tamoxifen in breast cancer treatment. Jurnal Farmasi UMMAT. Available at: https://journal.ummat.ac.id/index.php/farmasi/article/view/22563

Nindita, D.L. and Sanjaya, I.G.M., 2014. Modeling hubungan kuantitatif struktur dan aktivitas (HKSA) pinocembrin dan turunannya sebagai anti kanker. UNESA Journal of Chemistry, 3(2), pp.26–34.

Nystrom, M.P. and Sanchez, M., 2011. Biomedical QSAR and its application. In: Atkinson, M.R.Q. and Wright, W.E.S., editors. Encyclopedia of Biopharmaceutics. Springer.

Organisation for Economic Co-operation and Development (OECD), 2014. Guidance document on the validation of (Quantitative) Structure–Activity Relationship [(Q)SAR] Models. Paris: OECD Publishing.

Parr, R.G. and Yang, W., 1989. Density-Functional Theory of Atoms and Molecules. New York: Oxford University Press.

Pires, D.E.V., Blundell, T.L. and Ascher, D.B., 2015. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. Journal of Medicinal Chemistry, 58(9), pp.4066–4072. https://doi:10.1021/acs.jmedchem.5b00104.

Putra, G.S., Sulistyowaty, M.I., Yuniarta, T.A., Yahmin, Y., Sumari, S., Saechan, C. and Yamauchi, T., 2023. QSAR study of benzylidene hydrazine benzamides derivatives with in vitro anticancer activity against human lung cancer cell line A459. Journal of Pharmacy & Pharmacognosy Research, 11(6), pp.1123–1136.

Roy, P.P., Paul, S., Mitra, I. and Roy, K., 2015. On two novel parameters for validation of predictive QSAR models. Molecules, 20(9), pp.1519–1534.

Shagufta and Ahmad, I., 2018. Tamoxifen a pioneering drug: an update on the therapeutic potential of tamoxifen derivatives. European Journal of Medicinal Chemistry, 143, pp.515–531. https://doi:10.1016/j.ejmech.2017.11.056.

Todeschini, R. and Consonni, V., 2009. Molecular descriptors for chemoinformatics. 2nd edn. Weinheim: Wiley-VCH.

Wang, Z., Clark, J. and Ma'ayan, A., 2016. Drug-induced liver injury prediction using machine learning. Chemical Research in Toxicology, 29(4), pp.571–581.

Widiyanti, H., Banon, C. and Adfa, M., 2021. Analisis hubungan kuantitatif struktur terhadap aktivitas turunan senyawa cubebin sebagai antikanker dengan metode Recife Model 1 (RM1). Bencoolen Journal of Pharmacy, 1(1), pp.46–58.

Xiao, Y., et al., 2024. ADMETlab 3.0: an updated comprehensive online ADMET prediction platform. Nucleic Acids Research, 52(W1), W422–W430.

Yuan, H., et al., 2023. In silico study of the "magic chloro" effect: electronic properties and bioactivities. RSC Advances, 13(45), pp.32156–32167.

Zekri, A., Harkati, D., Kenouche, S. and Saleh, B.A., 2020. QSAR modeling, docking, ADME and reactivity of indazole derivatives as antagonizes of estrogen receptor alpha (ER-α) positive in breast cancer. Journal of Molecular Structure, 1217, p.128442. https://doi:10.1016/j.molstruc.2020.128442.

Zhang, J., Lai, N. and Wei, D., 2023. Molecular docking and DFT studies on ERα antagonists: implicating frontier molecular orbitals in binding. Journal of Molecular Modeling, 29(9), p.253.

Zhang, T.-T., Jia, J.-F. and Wu, H.-S., 2010. Substituent and solvent effects on electronic structure and spectral property of ReCl(CO)3 (NΛN) (NΛN = glyoxime). The Journal of Physical Chemistry A, 114(46), pp.12251–12257.

Zhao, F., et al., 2014. The effect and mechanism of tamoxifen-induced hepatocyte steatosis in vitro. International Journal of Molecular Sciences, 15(3), pp.4019–4030.