Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

QSAR model for the evaluation of biological active molecules on estrogen positive breast cancer

Itzel Mercado-Sánchez
Universidad de Guanajuato
Ismael Vargas-Rodríguez
Universidad de Guanajuato
Rogelio Chávez-Rocha
Universidad de Guanajuato
Cristina Fonseca-Yepez
Universidad de Guanajuato
Esthela Paola García-Tejada
Universidad de Guanajuato
Miguel Angel Vázquez
Universidad de Guanajuato
Adan Bazán-Jiménez
Universidad de Guanajuato
Juvencio Robles
Universidad de Guanajuato
Marco Antonio García-Revilla
magarcia@ugto.mx
Universidad de Guanajuato

Research Article

Keywords: QSAR, breast cancer, MCF-7, Trace Quadrupolar moment, molecular dynamics

Posted Date: August 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9952743/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

QSAR model for the evaluation of biological active molecules on estrogen positive breast cancer

Itzel Mercado-Sánchez[1], Ismael Várgas-Rodríguez[1], Rogelio Chávez-Rocha[1], Cristina Foseca-Yepez[1], Esthela P. García-Tejada[1], Miguel A. Vázquez[1], Adán Bazán-Jiménez[1], Juvencio Robles[1], Marco A. García-Revilla[1]*

[1]Department of Chemistry, Universidad de Guanajuato, Nora Alta,

S/N, Guanajuato, 3650, Guanajuato, México.

***Corresponding author(s). E-mail(s): magarcia@ugto.mx;**

Abstract

Context Breast cancer is a global health problem, and a key challenge in its treatment is the resistance to anticancer hormone therapy, which is the first-line treatment because of the over-expression of hormone receptors in the cells. For this reason, the understanding of mechanisms of biological activities of anticancer agents in positive-estrogen breast cancer for finding or repurposing drugs is an urgent issue. One of the main targets for this is the study of DNA intercalating molecules and inhibitors of topoisomerase IIα. In the present contribution, a quantitative structure–activity relationship (QSAR) model was developed to evaluate molecules with anticancer activity, with a particular focus on estrogen-positive breast cancer. The resulting model satisfied the established validation criteria ($R^2 > 0.6$, $Q^2 > 0.5$, $Q^2ext > 0.6$, $CCC > 0.8$) and incorporated a quantum-mechanical descriptor, the trace of the quadrupole moment, together with two-dimensional non-quantum-mechanical descriptors, revealing the influence of the quadrupole moment on the proposed mechanism of action of the molecules studied. Molecules **4a**, **5a**, **5b**, and **CID2948** were predicted to exhibit enhanced pharmacokinetic properties relative to the other compounds, and the stability of the Topoisomerase IIα-DNA-ligand complex was confirmed for the best-performing ligand and the reference drug.

Methods The training and test set molecules were selected using a pharmacophore model built from DNA-intercalating and Topoisomerase IIα-inhibiting compounds with antiproliferative activity against breast cancer. Their geometries were optimized at the M06-2X/6-311+G(d,p) level of theory using Gaussian 09. The QSAR model was built with QSARINS and combined quantum-chemical reactivity descriptors obtained at the M06-2X/6-311+G(d,p) level, including the trace of the quadrupole moment, with two-dimensional classical descriptors calculated using PaDEL. ADMET properties were predicted with SwissADME, ProTox 3.0, and T.E.S.T. Molecular docking was performed using the AutoDock Vina scoring function, and the stability of the Topoisomerase IIα-DNA-ligand complex was evaluated through 500 ns molecular dynamic simulations carried out in GROMACS with the CHARMM36 force field.

Keyword: QSAR, breast cancer, MCF-7, Trace Quadrupolar moment, molecular dynamics

1 Introduction

A multitude of factors influence the pathology of breast cancer, including age, lymph node status, tumor size, histopathologic features (lymphovascular invasion), hormone receptor status, and HER2 status[1, 2]. The molecular classification of breast cancer types is based on the expression of hormone receptors, which are classified into four main types: luminal A and B, HER2, and basal-like, in which the cells do not express hormone receptors [1]. The first-line treatment for breast cancer that is most utilized is the hormonal therapy [2], given the association between the most prevalent form of breast cancer and the over-expression of hormone receptors in the cells. In 1980, tamoxifen, an estrogen receptor modulator, was introduced as a standard treatment. Consequently, the development of resistance in cancer cells is a natural consequence. In contrast, *in vitro* studies have demonstrated that tamoxifen blocks estrogen-stimulated cell proliferation [3].

The impact of estrogen on carcinogenesis has been a subject of study for over a century. The development of effective therapeutic interventions has been a major area of research in this field. Anti-estrogens, a class of

pharmaceutical agents designed to counteract the effects of estrogen, have emerged as a significant therapeutic modality in the treatment of breast cancer. The MCF-7 cell line, distinguished by its expression of estrogen receptor (ER) and progesterone receptor (PgR), has been the focus of extensive research due to its functional receptor characteristics and favorable prognosis. It is noteworthy that the tamoxifen resistance model observed in MCF-7 cells is a distinctive feature that has garnered significant attention in the field. These culture cells are utilized in the therapeutic approach of anti-estrogen, serving as an alternative treatment modality [3].

The application of rational design as an integral part of computer-aided drug design provides useful insights into binding affinity and molecular interactions between a target protein and a ligand [4]. A pharmacophore model is defined as the set of steric and electronic features that are necessary to ensure optimal supramolecular interactions with a specific biological structure and to trigger (or block) its biological response [5]. The alignment of electronic and steric features is simply a (rather costly and prone to parameterization artifacts) way to delineate this consensus subset of conserved features, while still collaterally providing an intuitive description of the hypothesized interaction mode. The conventional approach to pharmacophore elucidation that does not involve alignment is based on quantitative structure-activity relationship (QSAR) [6]. A QSAR model is a mathematical statistical algorithm that establishes a quantitative relationship between molecular structure and properties. of the aforementioned relationship, the properties of drug candidates can be calculated from the model [7].

In the context of QSAR models, the structural relationship is conventionally derived from the molecular descriptors of the training set. These molecular descriptors have been previously described and classified as one-dimensional (molecular formula), two-dimensional (structural formula), three-dimensional (spatial considerations), and four-dimensional (assembly of conformations and interactions between the ligand with the biological target [8]. Nonetheless, over two decades ago, the utilization of reactivity descriptors derived from quantum chemistry [9] was initiated; these descriptors are referred to as quantum descriptors [7].

The quantum descriptors (QC) can be calculated using Conceptual Density Functional Theory (CDFT). These descriptors are contingent on alterations in energy with respect to the number of electrons, while maintaining a constant external potential [10], which essentially reflects the capability of gaining/losing electrons. Such quantities can be related to the Lewis acid/base concept, and, consequently, to the Hard and soft acids and bases (HSAB) theory [11], which is extremely useful in chemical reactivity and for chemical interactions characterization. The quantities in question can be calculated straightforwardly by employing mathematical formulas. These formulas involve the ionization potential (I) and the electronic affinity (A), which can be determined by finite differences [10, 12].

The employment of QC descriptors within QSAR models confers a notable advantage, as these descriptors are characterized by a discernible physical interpretation and are supported by fundamental chemical principles. This characteristic facilitates the comprehension of the mechanistic underpinnings of QSAR models. Consequently, QSAR models that integrate conventional 2D molecular descriptors with QC descriptors derived from quantum mechanics have been shown to demonstrate enhanced performance in comparison to individual models. This approach has been demonstrated to enhance the efficacy of the identification and selection of novel catalysts [12].

It is an evident overexpression of the Topoisomerase IIα in cancer cells. For this reason, Topoisomerase IIα is a relevant biological target for the designing of cancer therapy [13]. In addition, DNA intercalants have been employed as anti-cancers, antibiotics, and antimycotic drugs [14], because its capability of show a cytotoxic effect, when stabilizing the DNA-Intercalator-Topoisomerase II complex[15]. The present study developed a quantitative structure-activity relationship (QSAR) model employing 2D and QC descriptors to determine the logarithm of the half maximal inhibitory concentration ($\log IC_{50}$) in estrogen-positive breast cancer. Our study focuses on the Furthermore, pharmacokinetic parameters for 29 the molecules were calculated; besides, we run molecular dynamic simulations for the reference drug and the candidate with the best performance.

2 Results

2.1 Pharmacophore

The pharmacophore model consists of three aromatic frames. This configuration is attributed to the significance of stacking interactions in the context of DNA intercalators. These were found between **CID443939**, **CID3885**, **4a**, **5a**, and **5b** and Topoisomerase IIα-DNA according to a previous study [16], in which these molecules were evaluated *in silico* and *in vitro*. The pharmacophore model was constructed based on these five molecules (See Table 5). The model incorporates two fragments of hydrogen acceptors and two hydrogen donors, which are crucial for stabilizing the complex ligand– DNA - topoisomerase IIα through hydrogen interactions, Fig. 1.

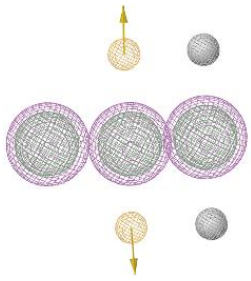


Fig. 1 Pharmacophore model of DNA intercalators, the elements of the model were represented as: aromatic framework, purple spheres, hydrophobic elements like green spheres, the hydrogen acceptor as gray spheres, and the hydrogen donor as orange spheres

Molecules matching the pharmacophore model, Fig. 1, were identified in the PubChem, Zinc, CHEMBL21, MCULE, MoltPort and LabNetwork databases (molecules labeled in the present work as 283, 2, 6, 1, 1 and 25).

Subsequently, the data were analyzed in accordance with the compound selection criteria, which consisted of antiproliferative activity against estrogen receptor-positive breast cancer (MCF-7 cell line) and an IC_{50} value of less than 100 μM, indicating that these are more potent compounds for this type of breast cancer. Moreover, during the selection of molecules, duplicate molecules and possible enantiomers were excluded, since these molecules had IC_{50} with unavailable experimental data.

2.2 Docking between the molecules set and Topoisomerase IIα-DNA

A docking simulation was executed between the molecules that fulfilled our pharmacophore model and the Topoisomerase IIα protein co-crystallized with a DNA sequence (PDB:5GWK) [17] and DNA sequence (PDB: 1DJD) [18]. A general observation of the interactions between the ligands and the biological target revealed that the predominant interaction was π-type stacking, followed by hydrogen-bonding interactions. It is noteworthy that intermolecular interactions are predominantly observed with the DNA sequence (see Table S1) and non-covalent interactions with Topoisomerase IIα residues outside the active site of the protein (see Table S2).

The docking results demonstrated that the ligand-DNA binding energies are higher compared with the co-crystallized ligand (PDB: 1DJD). Conversely, for the case of Topoisomerase IIα-DNA (PDB: 5GWK), the ligands exhibit a lower binding energy compared to that of the co-crystallized ligand, except for ligand **CID443939** (see Table 1). The binding energy between the tested molecules and Topoisomerase IIα-DNA ranged from -12.6 to -7.0 kcal/mol.

The molecule **CID443939** exhibited optimal binding energy (-12.6 kcal/mol) demonstrating the least favorable binding energy with Topoisomerase IIα-DNA. The molecules with binding energies that were not significantly different ($p > 0.05$) were **4a**, **CID2764**, **CID41867**, **CID42890**, **CID12567021**, **CID135494322**, **CID1382**, and **CID30323**. The rest of the molecules showed significant differences ($p < 0.05$) in binding energy compared with the reference molecule (Table 5). However, only **CID443939** was the molecule with a binding energy difference more negative than the reference, Table 1.

Regarding the ligand efficiency, it was observed that the ligand efficiencies between the analyzed molecules and DNA were lower than the reference (< -0.3 kcal/mol) for the following molecules: **5b**, **CID443939**, **CID57186217**, **CID2764**, **CID68897**, **CID381319**, **CID41867**, **CID42890**, **CID36462**, **CID135494322**, **CID139585971**, **CID2134**, **CID3663**, **CID65907**, **CID30323**, **CID2179**, and **CID4212.** However, for Topoisomerase IIα-DNA, all molecules exhibited ligand-binding energies comparable to or greater than the reference (> -0.2 kcal/mol), Table 1.

In general terms, the results of this experiment indicate that the binding energies between Topoisomerase IIα and DNA were more negative (ΔG<0) compared to the binding energy when only DNA was present, as well as higher ligand efficiencies.

Table 1 The average binding energy between the molecules being studied and the biological targets, DNA and Topoisomerase IIα-DNA. Ten experiments were carried out, each with the ten best poses.

Molecule	E_inter_ADN[a]	EL_ADN[b]	E_inter_TopoIIα[a]	EL_TopoIIα[b]
4a	-6.7±0.7*	-0.3	-8.5±06	-0.4
5a	-6.3±0.1*	-0.3	-8.0±0.5*	-0.3
5b	-5.6±0.9*	-0.2	-8.3±0.4*	-0.3
CID443939	-7.4±0.7*	-0.2	-12.6±0.6*	-0.3
CID3885	-5.6±0.9*	-0.3	-7.4±0.8*	-0.4
CID57186217	-6.7±0.6*	-0.2	-8.3±0.7	-0.3
CID2764	-6.5±0.5*	-0.2	-8.4±0.5*	-0.3
CID135908529	-7.8±0.7*	-0.3	-8.9±0.6*	-0.3
CID68897	-7.1±0.5*	-0.2	-8.6±0.5*	-0.2
CID137100894	-7.6±0.8*	-0.3	-8.9±0.6*	-0.3
CID381319	-7.7±0.7*	-0.2	-8.2±0.6*	-0.2
CID41867	-6.6±0.5*	-0.2	-8.1±0.6	-0.2
CID42890	-6.6±0.9*	-0.2	-8.1±0.5	-0.2
CID68063	-6.4±0.6*	-0.4	-7.7±0.5*	-0.4
CID36462	-6.1±0.3*	-0.1	-7.0±0.5*	-0.2
CID30623	-6.0±0.4*	-0.3	-7.5±0.5*	-0.4
CID12567021	-6.8±0.8*	-0.3	-8.6±0.8	-0.4
CID135494322	-6.7±0.6*	-0.2	-8.2±0.7	-0.3
CID3220	-6.5±0.7*	-0.3	-8.2±0.6*	-0.4
CID139585971	-6.1±0.6*	-0.2	-7.4±0.6*	-0.2
CID1382	-6.6±0.8*	-0.3	-7.9±0.6	-0.4
CID2948	-6.2±0.6*	-0.3	-7.6±0.7*	-0.4
CID2134	-6.0±0.6*	-0.2	-7.0±0.6*	-0.2
CID3663	-7.2±0.4*	-0.2	-9.7±0.4*	-0.3
CID65907	-7.1±0.8*	-0.2	-9.5±0.6*	-0.2
CID30323	-6.6±0.5*	-0.2	-8.4±0.6	-0.2
CID2179	-5.6±0.3*	-0.2	-7.4±0.6*	-0.3
CID3213	-6.3±0.5*	-0.3	-8.0±0.9*	-0.4
CID4212	-5.6±0.7*	-0.2	-8.3±0.5*	-0.3
Ligand co-crystallized	-6.8±1.0[1]	-0.3	-10.1±0.6[2]	-0.2

[a] Average bond energy ± standard deviation in kcal/mol, * p<0.05.

[b] Ligand efficiency in kcal/mol.

[1] Ligand co-crystallized reference ligand with 1DJD.

[2] Ligand co-crystallized reference ligand with 5GWK.

As shown in Fig. 2, the molecules exhibiting the optimal binding energy with Topoisomerase IIα protein are identified. The analysis reveals that the predominant interaction is stacking, occurring predominantly between molecules that contain a DNA sequence. This observation underscores the significance of the aromatic rings present in the molecules. It is noteworthy that the findings of the QSAR model are consistent with those derived

from intermolecular interactions observed during docking process between the molecules and the biological target, Topoisomerase IIα-DNA complex. The QSAR model emphasizes the importance of incorporating the trace moment quadrupole, as it plays a pivotal role in the intermolecular interactions of aromatic systems. Consequently, the docking results align with the descriptors of the QSAR model.

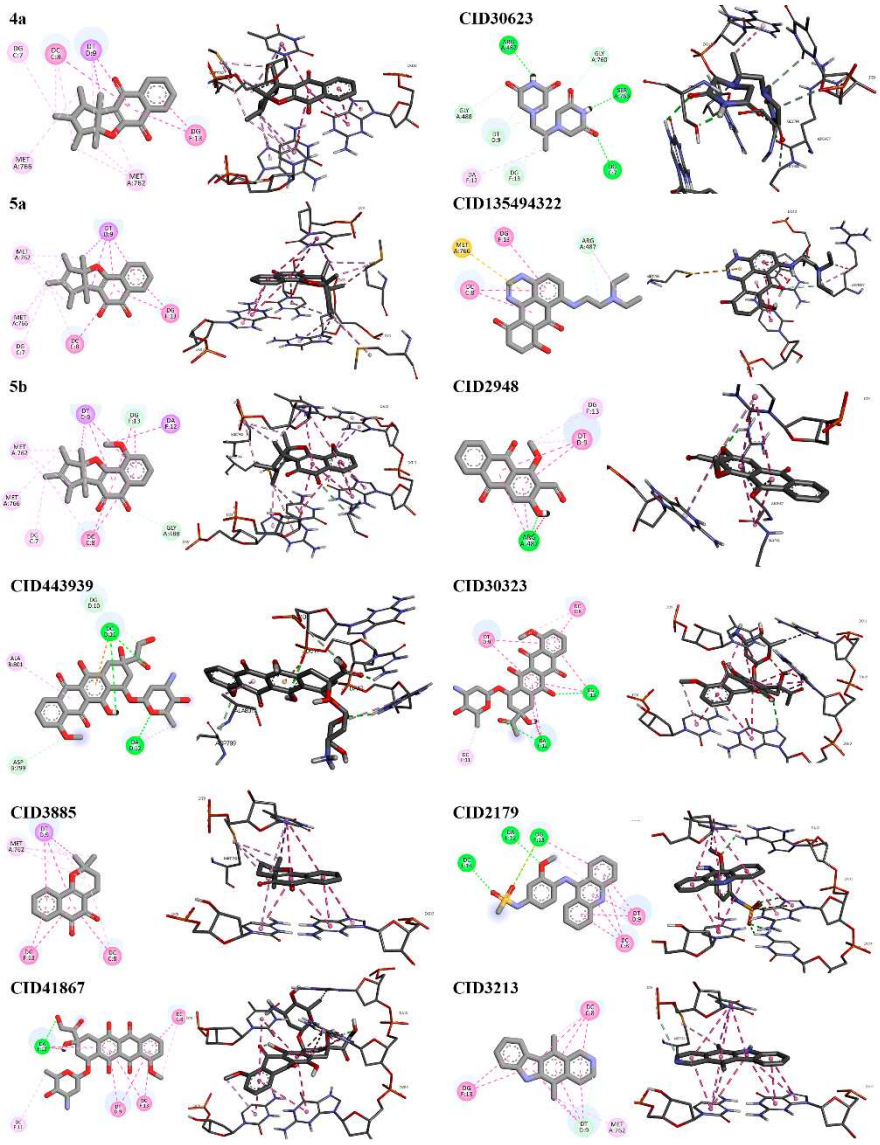


Fig. 2 Intermolecular interactions between molecules and biological targets in 2 and 3 dimensions; the pink color represents intermolecular interactions stacking-type, the green dotted line describes hydrogen bonding interactions, the orange-colored dotted line shows interactions sulfide-type, interactions π-alkyl, and Van der Waals interactions are shown as green circles without dotted lines

2.2 Chemical Descriptors

The molecular geometry of each molecule in the selected set was verified before proceeding with molecular geometry optimization using GaussView software. The molecular geometry was then optimized using the M06-2X/6-311+G(d,p) level of theory, as described in the methodology section. In this study, global descriptors were calculated (Fig. 3 and Table S3).

The results showed that the set of molecules had an ionization potential (I) greater than the electron affinity A. This means that the molecules tend to remain in a neutral state compared with their ability to form ions, which could imply greater stability under normal conditions.

Regarding hardness (η) and softness (S), the molecules exhibit larger values of η compared with S. This result suggests the molecules display a hard behavior and therefore are stable, resisting the removal [19].

The electrophilicity indices were determined through the calculation of donor electrophilic power ω^- and acceptor electrophilic power ω^+. The results showed higher values for ω^- than for ω^+. Consequently, it is plausible that the molecules under investigation can donate an electron more effectively to strong acceptor species.

Regarding the HOMO-LUMO gap, we confirmed that molecules with a small gap exhibited lower hardness, whilst those with a larger gap showed higher softness values. This is consistent with Pearson's acid-base principle[20], which connects a large HOMO-LUMO gap to a hard molecule and a small gap to a soft one. The HOMO-LUMO gap has been linked to toxicity[21]. In this sense, it can be inferred that a molecule with a narrower gap is more likely to react in a physiological environment, which could result in more significant adverse effects. In this regard, the molecule **CID443939** (doxorubicin) is one of the three molecules with the narrowest gap and is known to cause considerable adverse effects [22].

Another calculated descriptor was the trace of the quadrupole moment; as is well known, the quadrupole moment is a tensor that measures the distribution of the electron density relative to the Cartesian axes of a molecule. Therefore, if the distribution is symmetric, the trace of the quadrupole moment will consequently be zero. The results show that, in all cases of the analyzed molecules, the trace is non-zero, resulting in negative values of Q_{tz}, which could have implications for the preferential orientation of intermolecular interactions and molecular recognition processes.

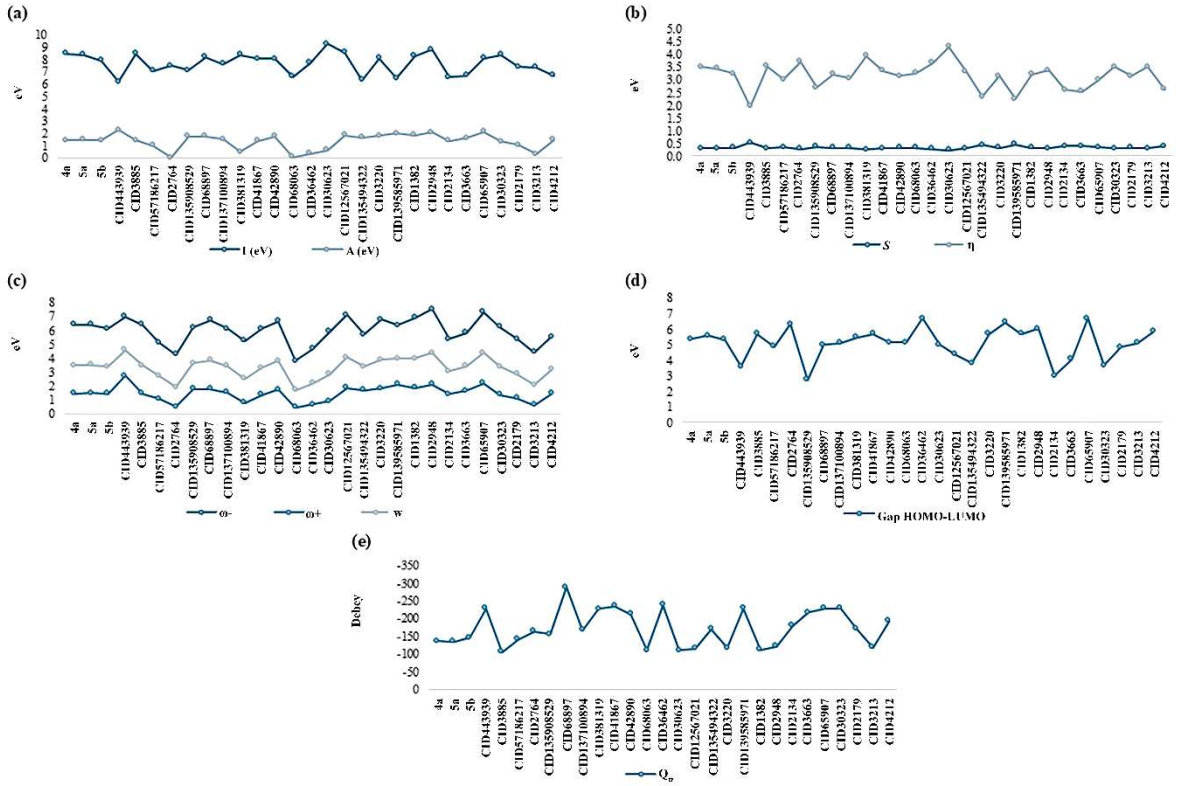


Fig. 3 Global descriptors of reactivity of studied molecules, the level theory was M06-2X/6-311+G** in the gas phase

On the other hand, the classical 1D, 2D and 3D molecular descriptors were determined using the PaDEL software. A total of 1,875 descriptors were obtained.

Finally, the QSAR model was constructed based on calculated chemical reactivity indices and classical 1D, 2D, and 3D molecular descriptors. We have included the energy bonding extracted from the docking.

2.2 QSAR model

The QSARINS software was employed to identify the QSAR model of molecules exhibiting antiproliferative activity in breast cancer cells with estrogen-positive (MCF-7) characteristics. The total molecular set was divided into two sets: the training set, which comprises 80 % of the total, and the test set, which comprises the remaining 20%, see Table 1. The obtained QSAR model, which correlates antiproliferative activity in the MCF-7 cell line with hybrid-type descriptors (QC descriptor combined with 2D non-QC descriptors), is displayed in equation 1.

$$\log IC_{50} = 11.5524 - 0.2204\ \mathbf{Qtz} - 3.1456\ \mathbf{ATSC5s} + 5.2668\ \mathbf{GATS2p} - 2.5531\ \mathbf{GATS3s} + 2.4545\ \mathbf{Sp_Max2_Bhi} \quad (1)$$

The scoring criteria of the QSAR model are as follows: R^2 value was found to be 0.81, the R^2_{adj} was 0.76 and the Root Mean Squared Error of the training set ($RMSE_{tr}$) was 0.31, Table 2. The QSAR model was trained using a set of 24 molecules and validated with a set of 5 molecules from Table 2. The internal validation criteria for the cross-validation coefficient leave-one-out (Q^2_{LOO}), Root Mean Squared of Errors (RMSE), cross-validation coefficient leave-many-out 30% (Q^2_{LMO}), R^2- Q^2_{LOO} and concordance Correlation Coefficient (CCC) were 0.69, 0.50, 0.64, 0.12 and 0.83, respectively. The values of external validation R^2_{ext}, Q^2 - $F3$, CCC_{ext} and Root Mean Squared of Errors external ($RMSE_{ext}$) were 0.94, 0.73, 0.83 and 0.47, respectively (see Table 3).

The graphs depicting the internal validation of the QSAR model are presented in Fig. 4A-D. This finding suggests that the model is not merely a product of a chance correlation [23]. As shown in Fig. 4A of the external validation graphs, a direct correlation is observed between the experimental and predicted $\log IC_{50}$ values, indicating a linear relationship between these variables. The assessment of model stability was calculated, 30% of the data points were excluded and the model was recalculated (LMO); the model is therefore considered stable if the Q^2_{LMO} values are close to those of the Q^2 model, which was indeed observed in Fig. 4B. The applicability domain is indicative of the robustness, significance, and validation of the QSAR model. Consequently, the model is expected to reliably predict across the entire chemical space [24, 25].

Furthermore, the applicability domain plot for the QSAR model indicates the presence of outliers in the response (Y-outliers) and in the structures (X-outliers) [24]. However, this is deemed acceptable, as the plot shows only a single outlier. Regarding the Insubria graph (see Fig. 4C), which was used to assess the applicability domain [13], it quantifies the discrepancy between leverage values ($h_{i/i}$) and $pLogIC_{50}$. Additionally, a thorough examination of William's graph reveals the absence of any outliers, exceeding 2.5 times the standard deviation (Fig. 4D). As demonstrated in Fig. 1, the molecules were found to be within the structural domain of the pharmacophore model. Consequently, the prediction's reliability can be considered high. According to the QSAR model, as illustrated in Table 1, the $\log IC_{50}$ values for the test set molecules are predicted, along with the values from the training set.

Table 2 The result of $plogIC_{50}$ using the QSAR model equation.

ID	Molecule	$\log IC_{50}$ (μM)	$plogIC_{50}$ (μM)
1	**4a**[a]	0.100	-0.457
2	**5a**[a]	-0.921	-0.555
3	**5b**[a]	-0.721	-1.066
4	**CID443939**[b]	-1.372	-1.418
5	**CID3885**[a]	-0.337	-0.200
6	**CID5718621**[a]	-2.155	-1.756
7	**CID2764**[a]	1.061	0.537

8	**CID135908529[b]**	0.591	-0.128
9	**CID68897[a]**	-0.495	-0.545
10	**CID137100894[a]**	-1.699	-0.617
11	**CID381319[a]**	1.176	1.107
12	**CID41867[b]**	-0.097	-0.820
13	**CID42890[a]**	-0.538	-1.123
14	**CID68063[a]**	-1.097	-1.140
15	**CID36462[a]**	-0.444	-0.527
16	**CID30623[a]**	-0.469	-0.351
17	**CID12567021[a]**	-0.143	-0.967
18	**CID135494322[b]**	-2.000	-1.816
19	**CID3220[a]**	-1.569	-1.542
20	**CID139585971[a]**	0.710	0.956
21	**CID1382[a]**	-1.538	-1.717
22	**CID2948[a]**	-2.000	-1.741
23	**CID2134[a]**	0.079	0.159
24	**CID3663[a]**	0.083	0.445
25	**CID65907[a]**	-1.699	-1.401
26	**CID30323[a]**	-1.770	-1.645
27	**CID2179[b]**	-0.337	-0.518
28	**CID3213[a]**	-0.921	-1.059
29	**CID4212[a]**	-0.456	-0.555

[a]Training set

[b]Test set

Table 3 Statistical parameters for the QSAR model.

	Fitting criteria			**Internal validation**						**External validation**				
Threshold values	R^2 > 0.6	R^2_{adj} >0.6	$RMSE_{tr}$	CCC_{tr} ≥ 0.8	Q^2_{LOO} > 0.5	$RMSE_{cv}$	Q^2_{LMO} > 0.6	CCC_{cv} ≥ 0.8	R^2-Q^2_{loo} <0.1	R^2_{ext} > 0.6	$RMSE_{ext}$t	R^2_{ext} >0.6	Q^2-F3 >0.7	CCC_{ext} ≥ 0.8
Model	0.81	0.76	0.31	0.83	0.69	0.50	0.64	0.83	0.12	0.6875	0.47	0.94	0.73	0.83

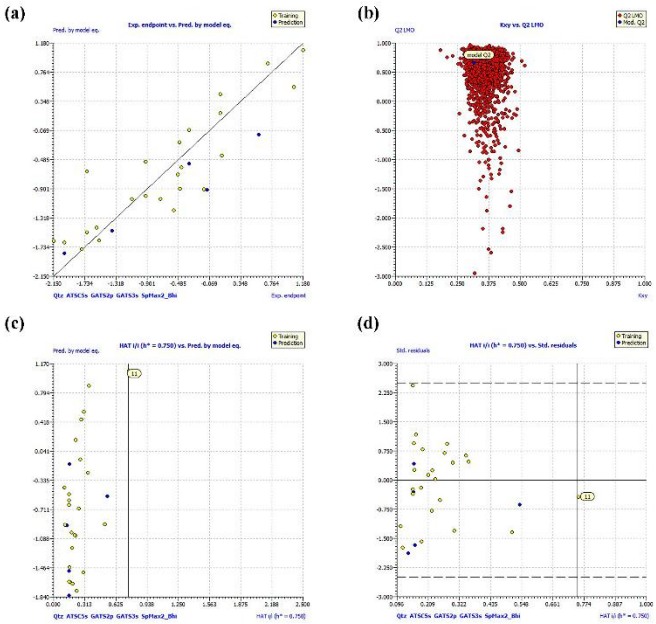


Fig. 4 Graphs of internal and external validation. **(a)** QSAR model graph experimental values versus logIC_{50} predicted values, **(b)** Scatter plot of the LMO models compared to the original QSAR model, **(c)** applicability domain of Insubria graph, and **(d)** William's graph, the yellow dots represent the training set, and the blue dots represent the test set; the number 11 corresponds to the molecule **CID381319**.

2.3 ADMET Properties

The determination of the pharmacokinetic properties of the studied molecules was performed with the free server SwissADME [15], T.E.S.T [16], and ProTox 3.0 [26]. Regarding the lipophilicity descriptor, the octanol/water partition coefficient (Log P) was calculated for the molecules under investigation. The desirable SwissADME range established for Log P is -0.7 to 5, and all the molecules that were the focus of this study are preferentially lipophilic (Log P values range from 0.61 to 5.57), except for **CID30623**, which has a Log P value of -1.40. Furthermore, the water solubility (mol/L) was calculated, and most of the molecules being found to be soluble or moderately soluble. However, **CID381319** and **CID139585971** exhibit poor solubility in water, while **CID30623** demonstrates high solubility. These observations are consistent with the lipophilicity indices, as illustrated in Table 4. Furthermore, molecules with moderate solubility exhibiting log P values between 3.84 and 5.0, as well as water-soluble molecules with Log P values between 0.61 and 2.63, demonstrated high gastrointestinal absorption. Low absorption was determined for molecules outside of these ranges, as well as for molecules with poor and high-water solubility (Log P values of 4.35 and 1.4, respectively). These findings indicate low water solubility and high lipophilicity, which consequently result in low gastrointestinal absorption. Furthermore, compounds with the potential to permeate the blood-brain barrier were qualitatively determined. Consequently, molecules with high Log P values (2.15–5.0) demonstrated the capability to permeate the BBB.

Regarding the prediction of P-glycoprotein (P-gp) substrates, molecules **4a**, **5a**, **5b**, **CID3885**, **CID57186217, CID381319**, **CID68063**, **CID30623**, **CID12567021**, **CID3220**, **CID1382**, **CID2948**, **CID2134**, and C**ID2179** are not predicted to be substrates of P-gp. Consequently, it can be deduced that these molecules may not be associated with Multi Drug Resistance (MDR), a phenomenon that is mediated by this efflux pump. The following molecules have been identified as substrates of P-gp: **CID443939**, **CID137100894**, **CID139585971**, **CID65907**, **CID30323**, **CID2764**, **CID135908529**, **CID68897**, **CID41867**, **CID42890**, **CID36462**, **CID135494322**, **CID3663**, **CID3213**, and **CID4212**, indicating susceptibility to MDR.

The Environmental Protection Agency (EPA) has assessed the toxicity of the molecules in question, and the results indicate that all the molecules studied have the potential to induce teratogenesis in humans and/or animals. **CID3663** and **CID443939** were identified as the molecules with the highest toxicity development

index value (1.00), while **CID68063** and **CID12567021** were identified as the molecules with the lowest toxicity index, at 0.53 and 0.57, respectively.

In the context of mutagenesis prediction, the Ames test was employed. Such a test utilizes histidine-dependent *Salmonella* strains to detect frameshift mutations or base pair substitutions by exposing them to a test compound. Exposure of *S. typhimurium* cells to the mutagenic compound results in the reversal of the mutation and the restoration of the bacteria's capability to synthesize histidine. This, in turn, enables colony growth on histidine-deficient medium [27]. The positive result in Ames indicates that it induces the growth of mutated bacteria in at least one of the five strains. The molecules identified as positive for mutagenicity are: **CID443939**, **CID3885**, **CID57186217**, **CID139585971**, **CID65907**, **CID30323**, **CID2764**, **CID135908529**, **CID68897**, **CID137100894**, and **CID381319**. **CID41867**, **CID42890**, **CID68063**, **CID12567021**, **CID3220**, **CID1382**, **CID3663**, **CID2179**, **CID3213**, and **CID4212**.

Furthermore, the potential inhibition of enzymes involved in drug metabolism (CYP1A2, CYP2C19, CYP3A4, CYP2D6, and CYP3A4) was calculated, with most molecules inhibiting at least one of the enzymes evaluated. Except for the molecules **CID65907**, **CID30323**, **CID135908529**, **CID68897**, **CID41867**, **CID42890**, **CID30623**, **CID139585971**, **CID3663**, and **CID4212**, which have been predicted to exhibit no inhibition of CYP450 family enzymes, the remaining molecules have been identified as those that do exhibit such inhibition.

The metabolic pathway by which a molecule could be metabolized by cytochrome P450 (CYP450) enzymes was determined by ProTox 3.0. The predicted metabolism indicates that several could be metabolized by specific enzymes in this pathway, namely CYP1A2, CYP2C19, CYP2D6, CYP3A4, and CYP2E1. However, a comparison of SwissADME predictions of CYP450 enzyme inhibition with the ProTox 3.0-determined CYP450 metabolic pathway revealed the potential of bioaccumulation in the body of **CID3885**, **CID57186217**, **CID381319**, **CID36462**, **CID135494322**, and **CID2179**. This outcome can be attributed to the predicted inhibition of CYP450 enzymes, which are critical for the metabolic transformation of these molecules (see Fig. 5).

Finally, SwissADME utilizes a bioavailability scoring function, to indicate the degree to which the oral administration of a pharmaceutical agent is likely to be efficacious. The molecules **CID443939**, **CID65907**, **CID30323**, **CID68897**, **CID381319**, **CID41867**, **CID36462**, and **CID3663** are characterized by the lowest bioavailability function (0.17 and 0.11), indicating a lower bioavailability by an oral administration. Conversely, **4a**, **5a**, **5b**, and **CID3885** have emerged as promising candidates for oral administration, with a bioavailability of 0.85.

As indicated by the preceding data, molecules **CID3663** and **CID2948** exhibit enhanced pharmaceutical properties and reduced toxicity levels, as determined by SwissADME, ProTox 3.0 and T.E.S.T. calculations.

Table 4 ADMET properties of the molecules.

Molecule	Log P	Solubility Class	GI absorption	BBB permeant	Pgp substrate	DTQ[1]	DT[2]	Mutagenicity	Bioavailability Score
4a	4.42	Moderately	High	Yes	No	0.79	Yes	Negative	0.85
5a	4.42	Moderately	High	Yes	No	0.84	Yes	Negative	0.85
5b	2.63	Soluble	High	Yes	No	0.87	Yes	Negative	0.85
CID443939	1.27	Soluble	Low	No	Yes	1.00	Yes	Positive	0.17
CID3885	2.20	Soluble	High	Yes	No	0.69	Yes	Positive	0.85
CID57186217	3.62	Moderately	High	No	No	0.93	Yes	Positive	0.55
CID2764	0.61	Soluble	High	No	Yes	0.83	Yes	Positive	0.55
CID135908529	5.00	Moderately	High	Yes	Yes	0.64	Yes	Positive	0.55
CID68897	2.34	Moderately	Low	No	Yes	0.79	Yes	Positive	0.17
CID137100894	2.52	Soluble	High	No	Yes	0.90	Yes	Positive	0.55
CID381319	4.35	Poorly	Low	No	No	0.81	Yes	Positive	0.17
CID41867	1.27	Soluble	Low	No	Yes	0.91	Yes	Positive	0.17
CID42890	1.86	Moderately	Low	No	Yes	0.95	Yes	Positive	0.55
CID68063	3.84	Moderately	High	No	No	0.53	Yes	Positive	0.55

CID36462	0.60	Soluble	Low	No	Yes	0.95	Yes	Negative	0.17
CID30623	-1.40	Very soluble	Low	No	No	0.67	Yes	Negative	0.55
CID12567021	2.15	Soluble	High	Yes	No	0.57	Yes	Positive	0.55
CID135494322	1.75	Soluble	High	No	Yes	0.76	Yes	Negative	0.55
CID3220	2.72	Soluble	High	No	No	0.80	Yes	Positive	0.55
CID139585971	5.57	Poorly	Low	No	Yes	0.85	Yes	Negative	0.55
CID1382	3.79	Moderately	High	Yes	No	0.83	Yes	Positive	0.55
CID2948	2.50	Soluble	High	No	No	0.94	Yes	Negative	0.55
CID2134	2.50	Soluble	High	No	No	0.66	Yes	Positive	0.55
CID3663	3.60	Moderately	Low	No	Yes	1.00	Yes	Positive	0.11
CID65907	1.75	Moderately	Low	No	Yes	0.78	Yes	Positive	0.17
CID30323	1.83	Moderately	Low	No	Yes	0.95	Yes	Positive	0.17
CID2179	4.01	Moderately	High	No	No	0.81	Yes	Positive	0.55
CID3213	4.80	Moderately	High	Yes	Yes	0.81	Yes	Positive	0.55
CID4212	1.00	Soluble	Low	No	Yes	0.75	Yes	Positive	0.55

[1] Developmental Toxicity Quantitative

[2] Developmental Toxicity

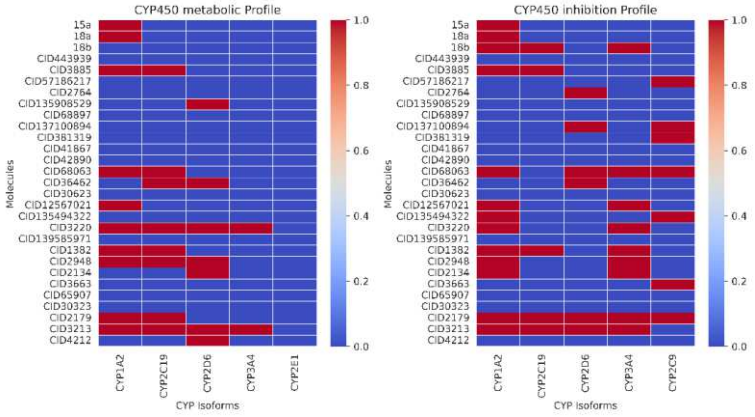


Fig. 5 Pharmacokinetic properties of molecules in CYP450 enzyme metabolic and inhibition, the left-hand heatmap shows the metabolic profile and the right-hand heatmap shows the inhibition profile of CYP450 enzymes and blue represents inactive and red represents active

2.5 Molecular Dynamics simulation

A molecular dynamics simulation was performed to predict the stability of the Topoisomerase IIα-DNA-ligand complex. This simulation was performed over a period of 500 ns. The ligands that were subjected to testing were **CID443939** (the reference ligand) and **CID2948** (the test ligand). The positional changes of the complex over time were measured using the root mean square deviation (RMSD) criterion during molecular dynamics experiments.

The stability analysis of the protein during simulation was assessed using the backbone RMSD. The presence of **CID443939** resulted in an RMSD value of 3.20 ± 0.55 Å for the protein backbone. Similarly, when **CID2948** was present, the protein backbone exhibited an RMSD value of 3.20 ± 0.55 Å. These observations indicate that the protein maintained its stability during the 500-ns molecular dynamics simulation, as shown in Fig. 6.

Regarding the fluctuations observed in the molecular dynamics simulations, several relevant facts that must be considered. Topoisomerase IIα is a protein that is expressed at a high level in cancer cells [18]. Furthermore,

the protein is a dimer, composed of chains A and B, and its primary structure consists of three sections: the N-terminal ATPase domain (residues 1–420); the active site, which splits into the primase topoisomerase (TOPRIM) domain, the winged helix domain (WHD), and the Tower domain (residues 421–1193); and the C-terminal domain (residues 1202–1429) [19, 20].

The primary fluctuations in the root mean square fluctuation (RMSF) were observed around the TOPRIM and WHD domains, as illustrated in Fig. 7. This phenomenon can be attributed to the presence of an alpha helix and three beta sheets within the TOPRIM domain (for residues from 526 to 608). This behavior is consistent with that described in the literature [28].

The radius of gyration of the system was found to be 3.730 ± 0.012 nm and 3.710 ± 0.013 nm for **CID443939** and **CID2948**, respectively. These values are quite close to one another, suggesting that both **CID2948** and **CID443939** exhibit a similar degree of structural compaction in the system, Fig 8.

The solvent-accessible surface area was calculated by the protein in the presence of each ligand, Fig. 9. The results showed that the value of SASA in the presence of **CID443939** was 702.04 ± 6.98 nm^2, whereas that of **CID2948** was 691.50 ± 7.99 nm^2. These results suggest that the SASA difference between **CID443939** and **CIS2948** is negligible, which is in alignment consistent with the radius of gyration.

In consideration of the results, it can be posited that the protein maintains its structural integrity throughout the 500 nanosecond (ns) molecular dynamics simulation. The graphs illustrating the structural stability analysis of the apo protein Topoisomerase IIα from molecular dynamics over a period of 500 ns are shown in Figure S1.

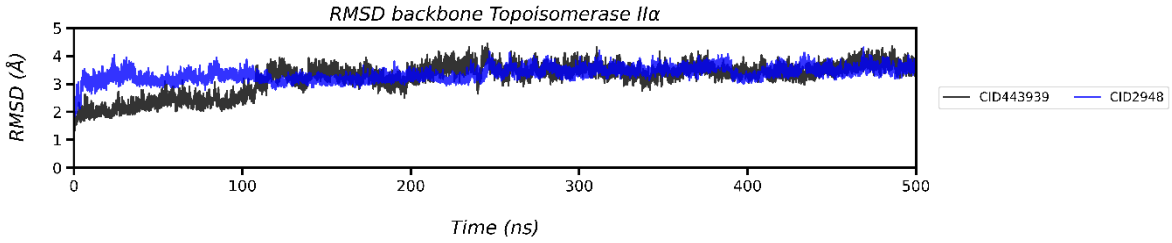


Fig.6 RMSD of backbone of Topoisomerase IIα over a period of 500 ns

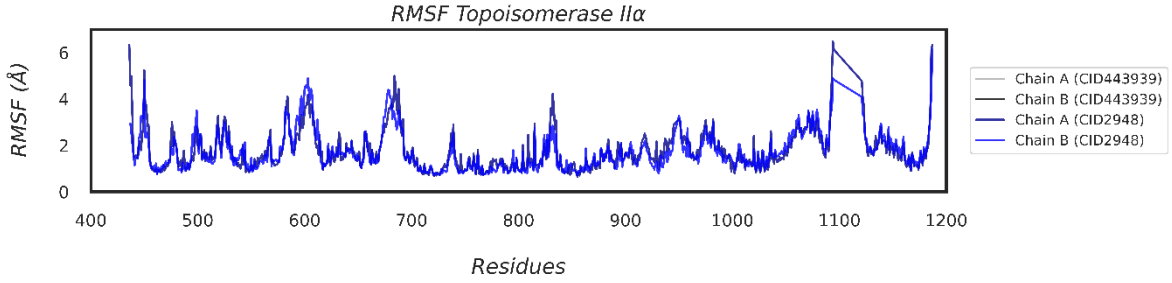


Fig. 7 RMSF of residues of Topoisomerase IIα in molecular dynamics simulation over a period of 500 ns

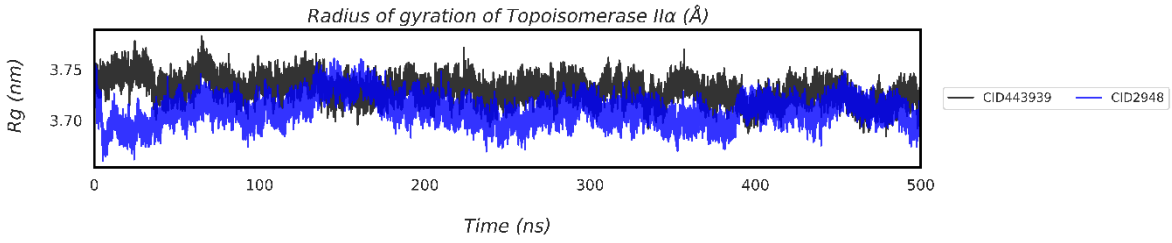


Fig. 8 Radius of gyration (Rg) of Topoisomerase IIα during molecular dynamic simulations over a period of 500 ns

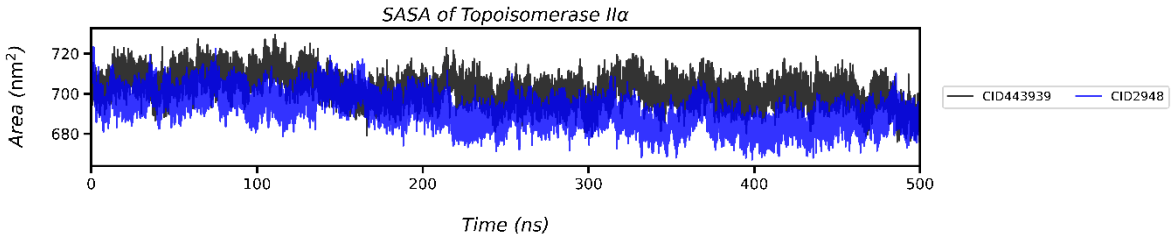


Fig. 9 The solvent-accessible surface area (SASA) of Topoisomerase IIα over a period of 500 ns

This enabled the study of the protein-ligand complex's stabilization during simulation. The complex stability between **CID443939**, **CID2948**, and Topoisomerase IIα with a chain of DNA was tested using molecular dynamics at 500 ns. The results demonstrated that both complexes exhibited consistent stability throughout the simulation, as evidenced by root-mean-square deviation (RMSD) values of 3.88 ± 1.05 and 2.63 ± 0.67 Å, respectively, as depicted in Fig. 10. In both cases, the stability of the ligands was maintained during the molecular dynamics simulation.

Furthermore, the ligands were found to form hydrogen bonds with Topoisomerase IIα. **CID443939** formed a maximum of six hydrogen bonds over the course of the simulation, whereas **CID2948** formed a maximum of three, Fig. 11. This is probably because the position of **CID443939** was closer to the protein, and therefore, had more contact with it than **CID2948.** Regarding the hydrogen bonds between the ligands and DNA, two hydrogen bond interactions were observed with **CID2948** throughout the entire molecular dynamic simulation (see Fig. 12). In contrast, **CID443939** exhibited only two hydrogen bond interactions with DNA at specific times (138–139 ns), with no interactions observed during the rest of the simulation. In contrast, hydrogen-bond interactions were observed in the docking experiments for **CID443939**.

Additionally, with respect to the stacking interactions between **CID443939** and the topoisomerase II–DNA complex, the ligand exhibited the capacity to form stacking interactions with topoisomerase IIα, Fig. 13. However, no stacking interactions are observed after 300 ns. This behavior could be attributed to the conformational change observed in the protein–ligand RMSD. In contrast, no stacking interactions were observed with DNA during the simulation. Consequently, the intermolecular interaction of **CID443939** occurs fundamentally with the protein. Regarding the stacking interaction between **CID2948** and the topoisomerase II-DNA complex, it was observed that these interactions predominantly occurred with the DNA, as **CID2948** had up to ten interactions over 500 ns, Fig. 14.

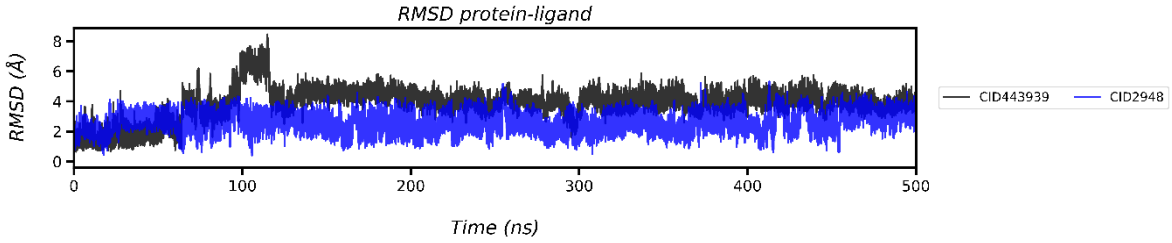


Fig. 10 RMSD of protein-ligand-DNA over a period of 500 ns

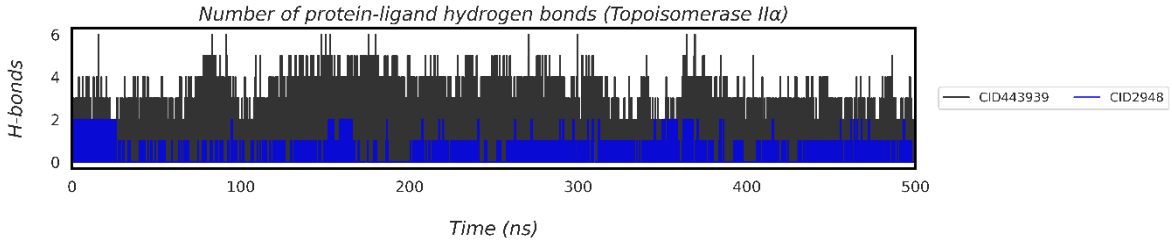


Fig. 11 Number of hydrogen bonds by each ligand with the protein Topoisomerase IIα calculated during the molecular dynamics simulation of 500 ns.

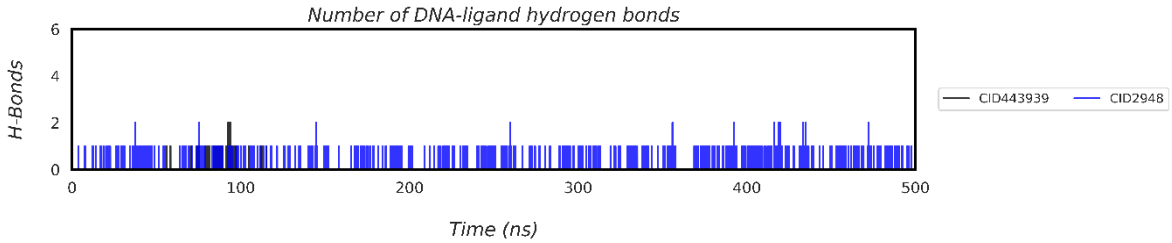


Fig. 12 Hydrogen bonds between ligands and DNA sequence over a period of 500 ns.

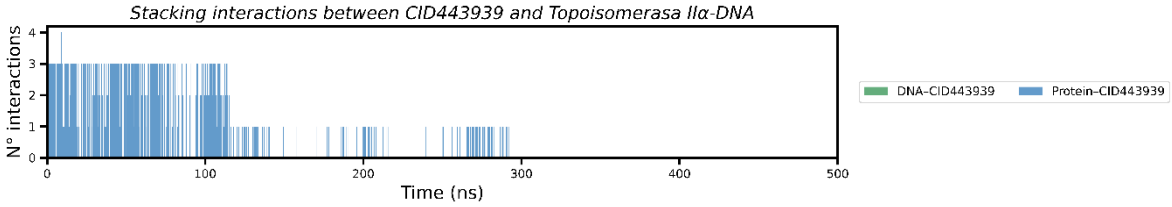


Fig. 13 Stacking interactions between **CID443939** and Topoisomerase II- DNA complex, the lines green exposed the stacking interactions between DNA and ligand, while the lines blue showed stacking interactions between Topoisomerase II and ligand.

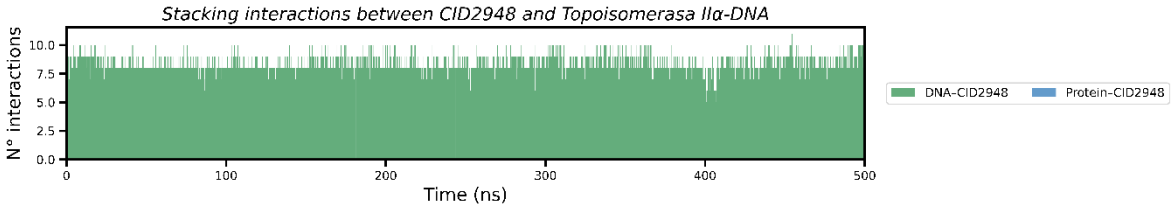


Fig. 14 Stacking interactions between **CID2948** and Topoisomerase II- DNA complex, the lines green exposed the stacking interactions between DNA and ligand, while the lines blue showed stacking interactions between Topoisomerase II and ligand.

3 Methodology

3.1 Pharmacophore model

The establishment of a pharmacophore model was based on DNA intercalating molecules and inhibitors of topoisomerase IIα, including doxorubicin (**CID443939**), lapachone (**CID3885**), and their analogues (**4a**, **5a**, **5b**). The pharmacophore model was developed using the ZincPharmer server [29]. The Pharmit software [30] was utilized in search of chemical structures corresponding to the specified model in the Pubchem [31], Zinc [29] and CHEMBL [32] databases.

3.3 Docking with Topoisomerase IIα and DNA

Molecular docking was performed with AutoDock Vina, with the homonym scoring function [33]. The maximum number of binding modes generated was 20, and 5 replicates were carried out.

The docking process was performed between molecules derived from the pharmacophore model and the Topoisomerase IIα protein and a DNA sequence.

The crystal of the Topoisomerase IIα PDB: 5GHK [17] was co-crystalized with a DNA sequence. The search space cartesian coordinates were (23.7, -31.81, -60.53); the number of points in x, y and z dimensions was 50, 24 and 16, respectively. Calibrating the docking process using the reference compound etoposide, which was

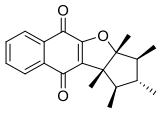
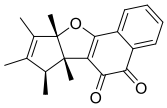
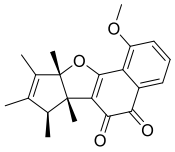
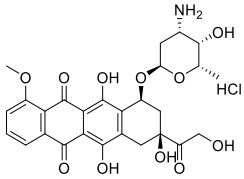
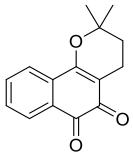
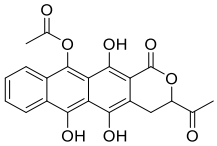
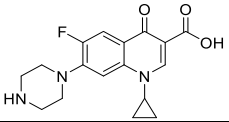
co-crystallized with the protein and calculated the RMSD between the experimental ligand and the pose ligand (0.1 ± 0.3 Å).

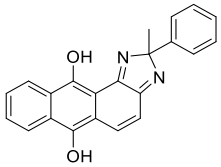
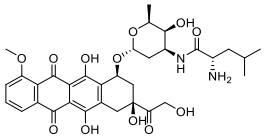
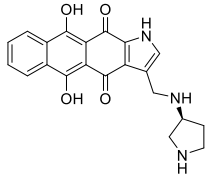
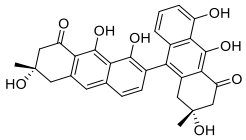
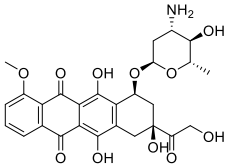
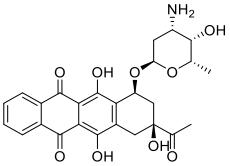
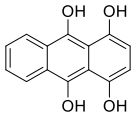
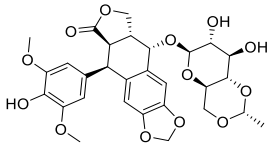
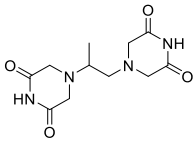
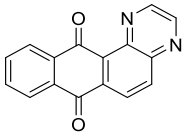
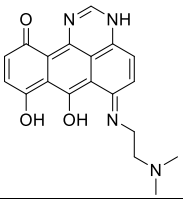
On the other hand interactions between the molecules with only one DNA sequence (PDB:1DJD [18]) were evaluated. The docking experiment had the number of points in x, y, and z dimensions of 66, 126, and 76, respectively; the space cartesian coordinates were -4.095, -2.511, and 16.889, respectively. Finally, the RMSD of the co-crystalized ligand was calculated (1.67 ± 0.44 Å).

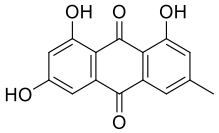
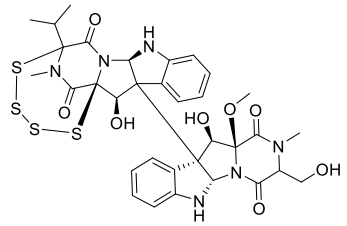
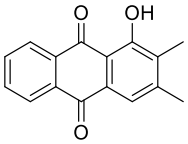
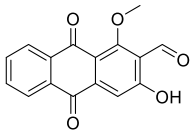
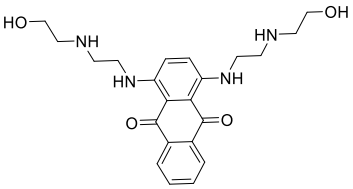
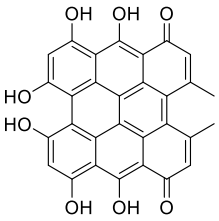
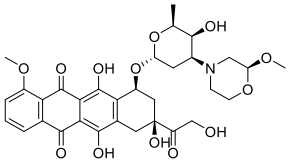
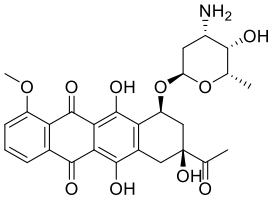
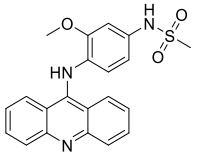
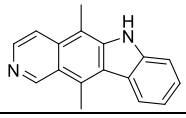
3.2 Descriptors and QSAR model

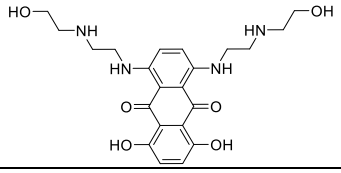
We have used PaDEL software [34] to determine the 1D, 2D, and 3D non-QC descriptors. In accordance with the pharmacophore model that was successfully obtained, a selection of compounds was made based on their demonstrated biological activity in the context of breast cancer, with a particular focus on their efficacy in MCF-7 cells (Table 5). The IC_{50} data for these molecules was subsequently collected for the construction of the QSAR model.

Table 5 The molecular structure of the molecules used by the model QSAR. The compound ID (CID) was displayed with the PubCHEM code.

Molecule	Structure	Reference
4a		[16]
5a		[16]
5b		[16]
CID443939		[16]
CID3885		[16]
CID57186217		[35]
CID2764		[36]

CID135908529		[37]
CID68897		[38]
CID137100894		[39]
CID381319		[40]
CID41867		[41]
CID42890		[42]
CID68063		[43]
CID36462		[44]
CID30623		[45]
CID12567021		[46]
CID135494322		[47]

CID3220		[48]
CID139585971		[49]
CID1382		[50]
CID2948		[51]
CID2134		[52]
CID3663		[53]
CID65907		[54]
CID30323		[55]
CID2179		[56]
CID3213		[57]

CID4212		[58]

The Quantum Chemical DFT reactivity descriptors (hardness, softness, including electrophilicity, ionization potential, and electronic affinity), were calculated using the M06-2X/6-311+G(d,p) (level of theory), and the Gaussian 09 software [59]. The neutral species were subject to fully geometry optimization; the minimum were confirmed by frequency analysis, followed by a single point calculation for cationic and anionic species of each molecule. The QC descriptors were determined using the following equations, Table 6.

Table 6 Quantum descriptors of the conceptual DFT.

Descriptor	Equation	Reference
Ionization Potential	$I = E(N-1) - E(N)$	[60]
Electronic affinity	$A = E(N) - E(N+1)$	[60]
Electronegativity	$\chi = \frac{I+A}{2}$	[60]
Chemical Potential	$\mu = -\left(\frac{I+A}{2}\right)$	[60]
Overall hardness	$\eta = \frac{I-A}{2}$	[60]
Global softness	$S = \frac{2}{I-A}$	[60]
Electrophilicity	$\omega = \frac{(I+A)^2}{4(I-A)}$	[60]
Electrodonating power	$\omega^- = \frac{(3I+A)^2}{16(I-A)}$	[61]
Electro-acceptor power	$\omega^+ = \frac{(I+3A)^2}{16(I-A)}$	[61]
Electrophilicity index	$\Delta\omega^{\pm} = \omega^- + \omega^+$	[62]
GAP HOMO LUMO	$GAP_{HOMO-LUMO} = E(N)_{LUMO} - E(N)_{HOMO}$	[63]

Where E(N), *E(N-1)* and *E(N+1)* are the energies for neutral, cation, and anion, respectively.

Additionally, we have included in the QC descriptors the traces of the quadrupole and the dipole moments. The dipole moment vector is a first-order moment $\mu = \sum e_i \bar{r}_i$ where e_i is the charge ith and $\bar{r}_i$ is vector position.

On the other hand, the quadrupole moment is a moment tensor (Q), is the second-order moment of the charge density:

$$Q = \frac{1}{2}\sum_i e_i \,(3r_i r_i - r_i^2 U)$$

Where U is the unit tensor, e_i is the charge and r_i is the position of the ith particle [2].

Since the quadrupole moment is a tensor that, in general, has nine non-zero components, if the trace is zero, it means that the charge distribution is symmetrical with respect to any axis [2, 64].

$$Q = \begin{bmatrix} Q_{xx} & Q_{xy} & Q_{xz} \\ Q_{yx} & Q_{yy} & Q_{yz} \\ Q_{zx} & Q_{yz} & Q_{zz} \end{bmatrix}$$

$$Q_{tz} = Q_{xx} + Q_{yy} + Q_{zz} = 0$$

The QSAR model was calculated with the QSARINS software [23]. The training set was used for external validation, and the validation criteria included the following: $R^2 > 0.6$, $Q^2_{LOO} > 0.5$, $R^2_{adj} > 0.6$, $Q^2 - F3 > 0.6$, CCC > 0.8 [65-67]. QSAR model validation graphs were generated using the QSARINS software.

3.3 ADMET properties

The determination of the pharmacokinetic properties of the studied molecules was performed by means of the free server SwissADME [68], ProTox 3.0 [69] and Toxicity Estimation Software Tool (T.E.S.T) [70].

3.4 Molecular dynamics simulation

The molecular dynamics of ligand-protein systems were performed at 500 ns, maintaining a temperature of 300 K and a constant pressure of 1 atm. The force field used was AMBER03 for protein, DNA, and ligand. During the molecular dynamic simulation, water molecules were added with the TIP3P model [71] (17432 water molecules), as well as sodium ions (Na^+), which were considered to neutralize the protein charge and the representation of the correct biological physiological medium (0.15 μM), requiring 20 Na+ ions. Long-range non-bonded Van der Waals interactions were computed using a cutoff of 10, and the electrostatic interactions were computed using the Particle Mesh Ewald method [72] with a 1.0 nm cutoff.

The simulation of system minimization was calculated with 10.0 kJ/mol convergence criterion, and 10 ns, followed by NVT and NPT ensemble calculated at 20 ns, using the V-rescale thermostat and Perrinello-Rahman [73] barostate and GROMACS software.

4 Discussion

In the present study, we have obtained a QSAR model for Topoisomerase IIα inhibitors and DNA intercalators. Such a model showed robust statistical results, considering QC descriptors (Q_{tz}) and classical descriptors (ATSC5s, GATS2p, GATS3s, SpMax2Bhi). The QC descriptors provide relevant electronic information about the molecules to explain changes in intermolecular interactions [74]. This has enabled our QSAR model to be oriented towards a mechanistic interpretation, as suggested by Wang et al [7].

It is known that the quadrupole moment is defined as the measure of the charge distribution of a molecule relative to a specific axis [75]. Furthermore, it is widely accepted that intermolecular interactions involving aromatic electron density are governed by the aromatic molecular quadrupole moment [76]. Additionally, the π-cation interactions had been relatively with the negative of the quadrupole moment, while π-anion interactions had been associated with the positive of quadruple moment. That means, the negative value of the quadrupole moment is related to the distribution of electronic density as it moves from one π-cloud to another through the hydrocarbon´s structure.

In this regard, our QSAR model demonstrates that the negative of quadrupole trace (Q_{tz}) is pertinent to account for the binding affinity of ligands with the biological target. Therefore, an increase in the negative charge of Q_{tz} corresponds to a corresponding increase in the negative quadrupole moment. This phenomenon has the potential to promote favorable interactions between the π-cation bonds of molecules and the biological target. This observation was made during the examination of molecular docking interactions between molecules and Topoisomerase IIα-DNA. The analysis revealed that the predominant interactions were of the following types: π-π stacking, bond hydrogen, and π-sigma. It has been established that the final interactions pertain to the functional groups -CH and π-systems [77]. Consequently, the present study observed the π-sigma interaction between the aromatic rings' π-cloud and the -CH group of T9, A12, and Met766.

Moreover, the incorporation of molecular descriptors associated with electronegativity, such as the ATC5, GATS3, and the Broto-Moreau and Geary autocorrelation descriptors, has been shown to enhance the prediction of plogIC_{50}. Tuning of molecular descriptors in molecules has the potential to yield enhanced antiproliferative activity in MCF-7. These findings suggest that a decrease in molecular electronegativity could improve the IC_{50} value in estrogen-receptor-positive breast cancer. From a mechanistic perspective, highly electronegative substituents (electron-withdrawing groups) can influence intermolecular interactions, thereby modifying the quadrupole moment of the aromatic ring [75]. It has been demonstrated that this behavior impacts the nature of stacking interactions; for instance, by reducing the electron density in the π system, it favors π-anion interactions

[75]. This finding is consistent with our docking observations, as no intermolecular interactions of the π-anion type were observed during docking process between the Topoisomerase IIα-DNA complex and the molecules under study.

Regarding Geary's autocorrelation coefficient, weighted by a polarization term based on the topological distance between two bonds [78], the QSAR model shows that a higher polarization value increases the response activity, as measured by the plogIC_{50} value. The increase in polarizability is significant, as it has been suggested that a molecule's polarization may preferentially lead to scattering and Van der Waals interactions [79].

Finally, the Burden matrix trace of the Level 2 maximum electro-topological index for the first ionization energy is a descriptor that estimates first ionization energy based on the molecular topology [78]. The increase in ionization potential *(I)* is indicative of a molecule's propensity to resist electron donation. This assertion can be substantiated by the outcomes of the conceptual DFT calculations, in where *I* is greater than *A* for all molecules. From a metabolic perspective, ligands with a low *I* value can be rapidly oxidized by CYPP450, which may lead to an increased risk of toxic metabolites [74]. This finding suggests that the high, *I* value of potential anti-cancer agents for estrogen receptor-positive breast cancer, as indicated by the present QSAR model, not only increases plogIC_{50} but also influences the metabolic and toxicological characteristics of the molecules.

The molecular dynamics simulation results indicated that the **CID443939** interactions occurred predominantly via hydrogen bonds with Topoisomerase IIα. The **CID2948** ligand exhibited a predominant interaction with DNA in the topoisomerase IIα-DNA complex, a finding that aligns with the outcomes of the docking study.

5 Conclusion

The integration of QC and classical molecular descriptors within the QSAR model facilitates a multifaceted approach to correlating chemical structures with activities. This integration enables mechanistic interpretations based on trends observed in the training set of molecules. This phenomenon can be attributed to the ability of QC descriptors to elucidate the nature of intermolecular interactions between the molecules and biological targets. The present work examines a series of molecules that demonstrate inhibitory activity in estrogen receptor-positive breast cancer cells. These molecules comply with a novel pharmacophore model proposed for potential Topoisomerase IIα inhibitors and DNA intercalators. Consequently, the QSAR model obtained in this study revealed the significant role played by π-stacking interactions, as determined by docking studies between the molecules under investigation and the Topoisomerase IIα-DNA complex. The obtained QSAR model demonstrated the correlation between the influence of the quadrupole moment and the proposed molecular mechanism. It has been posited that modifying the quadrupole moment by substituting with more electronegative groups could result in a decrease of the observed π-stacking molecular interactions in docking.

Acknowledgements

I.M-S. acknowledges Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Mexico PhD. grant # CVU: 673473. DAIP-UGto (CIIC/277) is acknowledged by the research grant to M.A.G-R.

Supplementary information.

The additional supplementary information contains followed:
Table S1 The values of binding energies result from docking between the molecules studied and the DNA sequence.
Table S2 Intermolecular interactions between molecules and the enzyme Topoisomerase IIα.
Table S3 Values of the QC descriptors of the molecules
Fig. S1 Structural stability analysis of Topoisomerase IIα from molecular dynamics simulations over a period of 500ns.

Declarations

Competing interests the authors report no conflict of interest.

Author contribution responsible for computational analysis, Marco A. García-Revilla and Itzel Mercado-Sánchez. Computational analysis, Marco A. García-Revilla, Itzel Mercado-Sánchez. Original Draft, Itzel Mercado-Sánchez. Computational Software and Hardware, Adan Bazan-Jimenez. Methodology Itzel Mercado-Sánchez, Ismael Várgas-Rodríguez and Rogelio Chávez-Rocha. Data curation and Visualization Itzel Mercado-Sánchez, Cristina Foseca-Yepez, Esthela P. García-Tejada, Miguel A. Vazquez.

References

1. Schnitt SJ (2010) Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. Mod Pathol 23:S60–S64. https://doi.org/10.1038/modpathol.2010.33
2. Prakash V (1966) Evaluation of molecular quadrupole moments. Rev Mod Phys 38:690. https://doi.org/10.1103/RevModPhys.38.690
3. Levenson AS, Jordan VC (1997) MCF-7: the first hormone-responsive breast cancer cell line. Cancer Res 57:3071–3078.
4. Gurung AB, Ali MA, Lee J, Farah MA, Al-Anazi KM (2021) An updated review of computer-aided drug design and its application to COVID-19. Biomed Res Int 2021:8853056. https://doi.org/10.1155/2021/8853056
5. Bajorath J (ed) (2011) Chemoinformatics and Computational Chemical Biology. Humana Press, New York. https://doi.org/10.1007/978-1-60761-839-3
6. Horvath D (2011) Pharmacophore-based virtual screening. In: Bajorath J (ed) Chemoinformatics and Computational Chemical Biology. Humana Press, New York, pp 261–298. https://doi.org/10.1007/978-1-60761-839-3_11
7. Wang L, Ding J, Pan L, Cao D, Jiang H, Ding X (2021) Quantum chemical descriptors in quantitative structure–activity relationship models and their applications. Chemom Intell Lab Syst 217:104384. https://doi.org/10.1016/j.chemolab.2021.104384
8. Cheng Z, Chen Q, Cervantes S, Tang Q, Gao X, Tan Y, Liu S, Ma Y, Shen Z (2020) Two-dimensional and three-dimensional quantitative structure-activity relationship models for the degradation of organophosphate flame retardants during supercritical water oxidation. J Hazard Mater 394:121811. https://doi.org/10.1016/j.jhazmat.2019.121811
9. Karelson M, Lobanov VS, Katritzky AR (1996) Quantum-chemical descriptors in QSAR/QSPR studies. Chem Rev 96:1027–1044. https://doi.org/10.1021/cr950202r
10. Parr RG (1980) Density functional theory of atoms and molecules. In: Fukui K, Pullman B (eds) Horizons of Quantum Chemistry. Springer, Dordrecht, pp 5–15. https://doi.org/10.1007/978-94-009-9027-2_2
11. Pearson RG (1968) Hard and soft acids and bases, HSAB, part 1: fundamental principles. J Chem Educ 45:581–587. https://doi.org/10.1021/ed045p581
12. Guégan F, Merzoud L, Chermette H, Morell C (2022) Conceptual DFT and excited states. In: Liu S (ed) Conceptual Density Functional Theory: Towards a New Chemical Reactivity Theory, vol 1. Wiley-VCH, Weinheim, pp 75-91. https://doi.org/10.1002/9783527829941.ch5
13. Chikamori K, Grozav AG, Kozuki T, Grabowski D, Ganapathi R, Ganapathi MK (2010) DNA topoisomerase II enzymes as molecular targets for cancer chemotherapy. Curr Cancer Drug Targets 10:758–771. https://doi.org/10.2174/156800910793605785
14. Gilad Y, Senderowitz H (2014) Docking studies on DNA intercalators. J Chem Inf Model 54:96–107. https://doi.org/10.1021/ci400352t
15. Tomczyk MD, Walczak KZ (2018) 1,8-Naphthalimide based DNA intercalators and anticancer agents. A systematic review from 2007 to 2017. Eur J Med Chem 159:393–422. https://doi.org/10.1016/j.ejmech.2018.09.055

16. Mercado-Sánchez I, López J, Chávez-Rocha R, Vargas-Rodríguez I, Bazán-Jiménez A, Segovia-Mendoza M, Prado-Garcia H, Vázquez MA, García-Becerra R, Garcia-Revilla MA (2024) Evaluation of doxorubicin and β-lapachone analogs as anticancer agents, a biological and computational study. Chem Biol Drug Des 104:e14596. https://doi.org/10.1111/cbdd.14596
17. Wang Y-R, Chen S-F, Wu C-C, Liao Y-W, Lin T-S, Liu K-T, Chen Y-S, Li T-K, Chien T-C, Chan N-L (2017) Producing irreversible topoisomerase II-mediated DNA breaks by site-specific Pt(II)-methionine coordination chemistry. Nucleic Acids Res 45:10861–10871. https://doi.org/10.1093/nar/gkx742
18. Li Z, Kim H-Y, Tamura PJ, Harris CM, Harris TM, Stone MP (1999) Role of a polycyclic aromatic hydrocarbon bay region ring in modulating DNA adduct structure: the non-bay region (8S,9R,10S,11R)-N6-[11-(8,9,10,11-tetrahydro-8,9,10-trihydroxybenz[a]anthracenyl)]-2‘-deoxyadenosyl adduct in codon 61 of the human N-ras protooncogene. Biochemistry 38:14820–14832. https://doi.org/10.1021/bi991607z
19. Parr RG, Chattaraj PK (1991) Principle of maximum hardness. J Am Chem Soc 113:1854–1855. https://doi.org/10.1021/ja00005a072
20. Pearson RG (1993) The principle of maximum hardness. Acc Chem Res 26:250–255. https://doi.org/10.1021/ar00029a004
21. Forrest J, Bazylewski P, Bauer R, Hong S, Kim CY, Giesy JP, Khim JS, Chang GS (2014) A comprehensive model for chemical bioavailability and toxicity of organic chemicals based on first principles. Front Mar Sci 1:31. https://doi.org/10.3389/fmars.2014.00031
22. Boelsterli UA (2002) Mechanistic Toxicology: The Molecular Basis of How Chemicals Disrupt Biological Targets. CRC Press, Boca Raton. https://doi.org/10.3109/9780203361764
23. Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S (2013) QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. J Comput Chem 34:2121–2132. https://doi.org/10.1002/jcc.23361
24. Gramatica P, Cassani S, Roy PP, Kovarich S, Yap CW, Papa E (2012) QSAR modeling is not "push a button and find a correlation": a case study of toxicity of (benzo-)triazoles on algae. Mol Inform 31:817–835. https://doi.org/10.1002/minf.201200075
25. Gramatica P (2020) Principles of QSAR modeling: comments and suggestions from personal experience. Int J Quant Struct-Prop Relatsh 5:61–97. https://doi.org/10.4018/IJQSPR.20200701.oa1
26. Banerjee P, Kemmler E, Dunkel M, Preissner R (2024) ProTox 3.0: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res 52:W513–W520. https://doi.org/10.1093/nar/gkae303
27. Ames BN, Lee FD, Durston WE (1973) An improved bacterial test system for the detection and classification of mutagens and carcinogens. Proc Natl Acad Sci 70:782–786. https://doi.org/10.1073/pnas.70.3.782
28. Pavlin M, Herlah B, Valjavec K, Perdih A (2023) Unveiling the interdomain dynamics of type II DNA topoisomerase through all-atom simulations: implications for understanding its catalytic cycle. Comput Struct Biotechnol J 21:3746–3759. https://doi.org/10.1016/j.csbj.2023.07.019
29. Koes DR, Camacho CJ (2012) ZINCPharmer: pharmacophore search of the ZINC database. Nucleic Acids Res 40:W409–W414. https://doi.org/10.1093/nar/gks378
30. Sunseri J, Koes DR (2016) Pharmit: interactive exploration of chemical space. Nucleic Acids Res 44:W442–W448. https://doi.org/10.1093/nar/gkw287
31. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA (2016) PubChem substance and compound databases. Nucleic Acids Res 44:D1202–D1213. https://doi.org/10.1093/nar/gkv951

32. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E (2017) The ChEMBL database in 2017. Nucleic Acids Res 45:D945–D954. https://doi.org/10.1093/nar/gkw1074
33. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461. https://doi.org/10.1002/jcc.21334
34. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32:1466–1474. https://doi.org/10.1002/jcc.21707
35. Wang W, Li T, Milburn R, Yates J, Hinnant E, Luzzio MJ, Noble SA, Attardo G (1998) Novel 1,3-disubstituted-5,10-dihydro-5,10-dioxo-1H-benzo[g]isochromene-3-carboxamides as potent antitumor agents. Bioorg Med Chem Lett 8:1579–1584. https://doi.org/10.1016/S0960-894X(98)00274-1
36. Eisa MA, Fathy M, Abu-Rahma GE-DA, Abdel-Aziz M, Nazmy MH (2021) Anti-proliferative and pro-apoptotic activities of synthesized 3,4,5-trimethoxy ciprofloxacin chalcone hybrid, through p53 up-regulation in HepG2 and MCF7 cell lines. Asian Pac J Cancer Prev 22:3393. https://doi.org/10.31557/APJCP.2021.22.10.3393
37. Huang H-S (2013) Method for inhibiting growth of cancer cells and cell telomere and diseases of cell proliferation by using heteroannelated anthraquinone derivative compounds. US Patent 8,445,492, 21 May 2013
38. Bennis S, Garcia C, Robert J (1993) Aspects of the cellular pharmacology of N1-leucyldoxorubicin in human tumor cell lines. Biochem Pharmacol 45:1929–1931. https://doi.org/10.1016/0006-2952(93)90453-4
39. Shchekotikhin AE, Glazunova VA, Luzikov YN, Buyanov VN, Susova OY, Shtil AA, Preobrazhenskaya MN (2006) Synthesis and structure–activity relationship studies of 4,11-diaminonaphtho[2,3-f]indole-5,10-diones. Bioorg Med Chem 14:5241–5251. https://doi.org/10.1016/j.bmc.2006.03.052
40. Lansiaux A, Laine W, Baldeyrou B, Mahieu C, Wattez N, Vezin H, Martinez FJ, Piñeyro A, Bailly C (2001) DNA topoisomerase II inhibition by peroxisomicine A1 and its radical metabolite induces apoptotic cell death of HL-60 and HL-60/MX2 human leukemia cells. Chem Res Toxicol 14:16–24. https://doi.org/10.1021/tx000145j
41. El-Sherbeni S, Al-Ashmawy GM (2021) Myricitrin from Physalis pubescens L. leaves and frankincense decrease resistance of MCF-7 cells and ameliorate efficacy of epirubicin. J Pharm Pharmacogn Res 9:497–508. https://doi.org/10.56499/jppres21.1031_9.4.497
42. Żwierełło W, Maruszewska A, Nowak R, Kostrzewa-Nowak D, Tarasiuk J (2017) DNA damage induced by NADPH cytochrome P450 reductase-activated idarubicin in sensitive and multidrug resistant MCF7 breast cancer cells. Pharmacol Rep 69:185–195. https://doi.org/10.1016/j.pharep.2016.10.002
43. Valla V, Bakola-Christianopoulou M, Kojic V, Bogdanovic G (2007) Synthesis, spectroscopy and in vitro cytotoxicity of new hydroxyanthraquinonato triorganotin compounds. Synth React Inorg Met-Org Nano-Met Chem 37:41–51. https://doi.org/10.1080/15533170601172450
44. Yang X-H, Sladek TL, Liu X, Butler BR, Froelich CJ, Thor AD (2001) Reconstitution of caspase 3 sensitizes MCF-7 breast cancer cells to doxorubicin- and etoposide-induced apoptosis. Cancer Res 61:348–354
45. Budman DR, Calabro A (2002) In vitro search for synergy and antagonism: evaluation of docetaxel combinations in breast cancer cell lines. Breast Cancer Res Treat 74:41–46. https://doi.org/10.1023/A:1016070230538
46. Huang H-S, Chen T-C, Chen R-H, Huang K-F, Huang F-C, Jhan J-R, Chen C-L, Lee C-C, Lo Y, Lin J-J (2009) Synthesis, cytotoxicity and human telomerase inhibition activities of a series

of 1,2-heteroannelated anthraquinones and anthra[1,2-d]imidazole-6,11-dione homologues. Bioorg Med Chem 17:7418–7428. https://doi.org/10.1016/j.bmc.2009.09.033
47. Stefanska B, Dzieduszycka M, Martelli S, Tarasiuk J, Bontemps-Gracz M, Borowski E (1993) 6-[(Aminoalkyl)amino]-substituted 7H-benzo[e]perimidin-7-ones as novel antineoplastic agents. Synthesis and biological evaluation. J Med Chem 36:38–41. https://doi.org/10.1021/jm00053a005
48. Li W-Y, Chan RY-K, Yu PH-F, Chan S-W (2013) Emodin induces cytotoxic effect in human breast carcinoma MCF-7 cell through modulating the expression of apoptosis-related genes. Pharm Biol 51:1175–1181. https://doi.org/10.3109/13880209.2013.782322
49. Pejin B, Jovanovic KK, Mojovic M, Savic AG (2013) New and highly potent antitumor natural products from marine-derived fungi: covering the period from 2003 to 2012. Curr Top Med Chem 13:2745–2766. https://doi.org/10.2174/15680266113136660197
50. Aziz MYA, Omar AR, Subramani T, Yeap SK, Ho WY, Ismail NH, Ahmad S, Alitheen NB (2014) Damnacanthal is a potent inducer of apoptosis with anticancer activity by stimulating p53 and p21 genes in MCF-7 breast cancer cells. Oncol Lett 7:1479–1484. https://doi.org/10.3892/ol.2014.1898
51. Sharma K, Pachauri S, Khandelwal K, Ahmad H, Arya A, Biala P, Agrawal S, Pandey R, Srivastava A, Srivastav A (2016) Anticancer effects of extracts from the fruit of Morinda citrifolia (Noni) in breast cancer cell lines. Drug Res 66:141–147. https://doi.org/10.1055/s-0035-1555804
52. Fisher GR, Patterson LH (1992) Lack of involvement of reactive oxygen in the cytotoxicity of mitoxantrone, CI941 and ametantrone in MCF-7 cells: comparison with doxorubicin. Cancer Chemother Pharmacol 30:451–458. https://doi.org/10.1007/BF00685596
53. Mirmalek SA, Azizi MA, Jangholi E, Yadollah-Damavandi S, Javidi MA, Parsa Y, Parsa T, Salimi-Tabatabaee SA, Ghasemzadeh Kolagar H, Alizadeh-Navaei R (2015) Cytotoxic and apoptogenic effect of hypericin, the bioactive component of Hypericum perforatum on the MCF-7 human breast cancer cell line. Cancer Cell Int 16:1–9. https://doi.org/10.1186/s12935-016-0279-4
54. Broggini M (2008) Nemorubicin. In: Anthracycline Chemistry and Biology II: Mode of Action, Clinical Aspects and New Drugs. Springer, Berlin, pp 191–206. https://doi.org/10.1007/978-3-540-75813-6
55. Sadava D, Coleman A, Kane SE (2002) Liposomal daunorubicin overcomes drug resistance in human breast, ovarian and lung carcinoma cells. J Liposome Res 12:301–309. https://doi.org/10.1081/LPR-120016196
56. Zhou Z, Zwelling L, Ganapathi R, Kleinerman E (2001) Enhanced etoposide sensitivity following adenovirus-mediated human topoisomerase IIα gene transfer is independent of topoisomerase IIβ. Br J Cancer 85:747–751. https://doi.org/10.1054/bjoc.2001.1966
57. Stiborová M, Poljaková J, Martínková E, Bořek-Dohalská L, Eckschlager T, Kizek R, Frei E (2011) Ellipticine cytotoxicity to cancer cell lines—a comparative study. Interdiscip Toxicol 4:98. https://doi.org/10.2478/v10102-011-0017-7
58. Zou Y, Cao Z, Wang J, Chen X, Chen Y-q, Li Y, Liu J, Zhao Y, Wang A, He B (2020) A series of novel HDAC inhibitors with anthraquinone as a cap group. Chem Pharm Bull 68:613–617. https://doi.org/10.1248/cpb.c20-00206
59. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, et al (2009) Gaussian 09 (Revision A.02). Gaussian Inc, Wallingford, CT
60. Parr RG, Szentpály Lv, Liu S (1999) Electrophilicity index. J Am Chem Soc 121:1922–1924. https://doi.org/10.1021/ja983494x
61. Gázquez JL, Cedillo A, Vela A (2007) Electrodonating and electroaccepting powers. J Phys Chem A 111:1966–1970. https://doi.org/10.1021/jp065459f

62. Chattaraj PK, Chakraborty A, Giri S (2009) Net electrophilicity. J Phys Chem A 113:10068–10074. https://doi.org/10.1021/jp904674x
63. Levine IN (2009) Quantum Chemistry, 7th edn. Pearson Prentice Hall, Upper Saddle River
64. Villanueva-García M, Robles J, Martínez-Richa A (2001) Quadrupolar moment calculations and mesomorphic character of model dimeric liquid crystals. Comput Mater Sci 22:300–308. https://doi.org/10.1016/S0927-0256(01)00244-0
65. Gramatica P (2007) Principles of QSAR models validation: internal and external. QSAR Comb Sci 26:694–701. https://doi.org/10.1002/qsar.200610151
66. Gramatica P, Sangion A (2016) A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology. J Chem Inf Model 56:1127–1131. https://doi.org/10.1021/acs.jcim.6b00088
67. Consonni V, Ballabio D, Todeschini R (2009) Comments on the definition of the Q2 parameter for QSAR validation. J Chem Inf Model 49:1669–1678. https://doi.org/10.1021/ci900115y
68. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717. https://doi.org/10.1038/srep42717
69. Banerjee P, Eckert AO, Schrey AK, Preissner R (2018) ProTox-II: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res 46:W257–W263. https://doi.org/10.1093/nar/gky318
70. United States Environmental Protection Agency (2020) User's Guide for T.E.S.T. (Version 5.1) (Toxicity Estimation Software Tool): A Program to Estimate Toxicity from Molecular Structure. Chemical Characterization and Exposure Division, Cincinnati, OH
71. Mark P, Nilsson L (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J Phys Chem A 105:9954–9960. https://doi.org/10.1021/jp003020w
72. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593. https://doi.org/10.1063/1.470117
73. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 52:7182–7190. https://doi.org/10.1063/1.328693
74. Rohit H, Tandon H (2025) Conceptual density functional theory in drug discovery: an overview. J Mol Model 31:291. https://doi.org/10.1007/s00894-025-06487-5
75. Schottel BL, Chifotides HT, Dunbar KR (2008) Anion-π interactions. Chem Soc Rev 37:68–83. https://doi.org/10.1039/B614208G
76. Clements A, Lewis M (2006) Arene–cation interactions of positive quadrupole moment aromatics and arene–anion interactions of negative quadrupole moment aromatics. J Phys Chem A 110:12705–12710. https://doi.org/10.1021/jp065175v
77. Nishio M, Hirota M, Umezawa Y (1998) The CH/π Interaction: Evidence, Nature, and Consequences. Wiley, New York
78. Todeschini R, Consonni V (2009) Molecular Descriptors for Chemoinformatics, Volume I: Alphabetical Listing / Volume II: Appendices, References. Wiley-VCH, Weinheim. https://doi.org/10.1002/9783527628766
79. Goel H, Yu W, Ustach VD, Aytenfisu AH, Sun D, MacKerell AD (2020) Impact of electronic polarizability on protein-functional group interactions. Phys Chem Chem Phys 22:6848–6860. https://doi.org/10.1039/D0CP00088D

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SIQSARMODEL250526.docx