European Journal of Chemistry

View Journal Online
View Article Online

Modeling the aromatase inhibitor activity of indole-imidazole derivatives: Quantitative structure activity relationship and molecular docking

Rawan Mustafa Ali Massad * and Ahmed Elsadig Mohammed Saeed

Department of Chemistry, College of Science, Sudan University of Science and Technology, 407, Khartoum, Sudan

** Corresponding author at: Department of Chemistry, College of Science, Sudan University of Science and Technology, 407, Khartoum, Sudan.*
e-mail: rawanmus5afa@gmail.com (R.M.A. Massad).

RESEARCH ARTICLE

10.5155/eurjchem.17.1.26-33.2713

Received: 24 August 2025
Received in revised form: 12 December 2025
Accepted: 3 January 2026
Published online: 31 March 2026
Printed: 31 March 2026

KEYWORDS

Drug design
Heterocycles
Bioinformatics
Molecular modeling
Computational chemistry
Indole and imidazole derivatives

ABSTRACT

In the present study, 19 compounds of indole-imidazole derivatives were studied to obtain the structure requirements to inhibit the active sites of the aromatase enzyme. 2D quantitative structure-activity relationships (QSARs) were analyzed using the partial least squares (PLS) method. To build the QSAR model, the data set was randomly split into a training set (15 compounds) and a test set (4 compounds) for the external validation of the model. As a result, a model with three descriptors (diameter, Petitjean, Q_VSA_FPNEG) was found to be robust enough to predict the aromatase inhibitor activity of the indole-imidazole derivatives, with an R^2 of 0.892 and Q^2 of 0.741. A series of 57 new compounds was modeled and designed; Of these, only 18 compounds were found to have biological activity greater than that of letrozole (the reference compound). These compounds were docked to the active site of aromatase to understand their inhibitory action and their binding energy toward the aromatase enzyme. Analysis of energy of the 18 compound-aromatase complexes revealed that compound 48 has a low binding energy (strong binding affinity) to aromatase as compared to letrozole; the energy of this compound is less by 7 units than that of letrozole. This compound is enhanced by an electron-withdrawing group (COOH) at the *meta* position of the phenyl ring of indole.

Cite this: *Eur. J. Chem.* **2026**, *17(1)*, 26-33 **Journal website**: www.eurjchem.com

1. Introduction

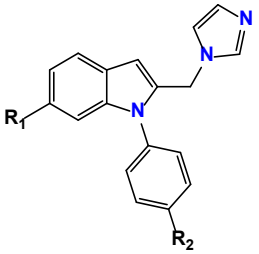
Breast cancer is the most diagnosed form of cancer worldwide [1,2], the second most common cause of cancer death in women [3], and in 2022 about 2.3 million cases were registered [4]. The predominant subtype of breast cancer is hormone dependent, accounting for approximately 70% of all cases [5]. This subtype is characterized by the presence of estrogen receptors (ER) and requires estrogen for tumor growth [6]. A crucial player in this hormonal pathway is aromatase (cytochrome P450), which plays a pivotal role in the conversion of androgens to estrogens [7,8]. Inhibition of aromatase represents one of the most effective therapeutic strategies for the treatment of hormone-dependent breast cancer [9,10]. Aromatase inhibitors (AIs) can be categorized based on their mechanism of action into two main types: steroidal AIs, which irreversibly inhibit the activity of the aromatase enzyme [11], and non-steroidal AIs (NSAIs), which bind to the enzyme's active site through non-covalent interactions, resulting in a reversible binding process [12]. Among nonsteroidal aromatase inhibitors, the most studied compounds have been indole [13]. Compounds containing heterocycles with nitrogen have been identified as having optimal activity, as they can coordinate with the heme iron of the enzyme, resulting in its inactivation [6]. The concept of a single molecule possessing multiple pharmacophores, each with different modes of action, could offer benefits in cancer treatment by potentially reducing the chances of drug resistance [14].

Computational approaches serve as effective facilitation tools in drug design and discovery [15]. Determining the activity of small molecule inhibitors *in vitro* is costly and time-consuming. Computer-aided drug design (CADD) not only predicts inhibitor activity, but also provides guidance for the design of more effective inhibitors by exploring the reaction mechanism at the molecular level, thus saving experimental costs [16].

The quantitative structure-activity relationship (QSAR) model employs statistical analysis and theoretical calculation tools to establish a quantitative relationship between the structural properties of a compound and its biological activity [17], and has been widely used in the development and design of anticancer agents [18]. Molecular docking represents a computational method that is used to identify potential binding modes of a compound with its biological target. It has been used successfully to investigate the binding modes of various classes of aromatase inhibitors [15].

https://dx.doi.org/10.5155/eurjchem.17.1.26-33.2713

Table 1. Structures and *in vitro* CYP19 inhibitory activities of indole-imidazole derivatives and the predicted relative potency from QSAR study [21].



Compounds	R_1	R_2	IC_{50} (nM) [a]	RP_{exp} [b]	RP_{pre}	Residual
10a [T]	H	F	9.01±12.61	1.8346	0.2348	1.5998
10b	H	CH_3	148.93±12.61	0.1110	0.2013	-0.0903
10c	H	OCH_3	77.36±6.31	0.2137	0.1557	0.0580
10d [T]	H	CF_3	4.93±0.23	3.3529	0.9769	2.2376
10e	H	H	16.58±1.39	0.9970	0.9077	0.0893
10f	CH_3	H	21.39±1.76	0.7728	0.8621	-0.0893
10g	CH_3	CH_3	164.01±14.63	0.1008	0.1607	-0.0599
10h	CH_3	OCH_3	138.72±11.46	0.1192	0.1155	0.0037
10i	CH_3	F	56.83±5.18	0.2909	0.1905	0.1004
10j	OCH_3	CN	27.01±1.92	0.6120	0.4973	0.1147
10k [T]	OCH_3	H	6.23±0.51	2.6576	0.2269	2.4307
10l	OCH_3	F	48.93±4.36	0.3378	0.2162	0.1216
10m [T]	OCH_3	CF_3	25.56±2.14	0.6467	0.8837	-0.2370
10n	OCH_3	CN	46.92±4.13	0.3523	0.5048	-0.1525
10o	OCH_3	CH_3	57.43±4.96	0.2878	0.1867	0.1011
10p	OCH_3	OCH_3	111.09±10.47	0.1488	0.1393	0.0095
10q	Cl	OCH_3	203.34±18.91	0.0813	0.1147	-0.0334
10r	Cl	CH_3	235.33±22.49	0.0702	0.1599	-0.0897
10s	Cl	Cl	217.43±20.67	0.0760	0.1591	-0.0831
Letrozole	-	-	16.53±1.24	1.00	-	-

[a] Values are the mean of at least three experiments performed in triplicate.
[b] Relative potency RP = IC_{50} (letrozole)/IC_{50} (tested compound).
[T] Test set compounds.

The aim of this study is to develop a robust and predictive quantitative structure-activity relationship (QSAR) model for indole-imidazole derivatives and to elucidate their inhibitory interactions with the aromatase enzyme through molecular docking. The scientific novelty of this work lies in the integration of a rigorously validated 2D-QSAR model with virtual design and structure-based docking to identify new indole-imidazole derivatives with enhanced aromatase inhibitory potential. Unlike previously reported QSAR studies that primarily focus on activity prediction, the present study establishes a direct link between statistically significant molecular descriptors and binding interactions within the aromatase active site. These findings provide important molecular-level insights into the structure-activity relationships of indole-imidazole derivatives and offer a rational computational framework for guiding the future design and optimization of potent aromatase inhibitors.

2. Experimental

2.1 Molecular modeling software

In this study, the Molecular Operating Environment (MOE) software developed by the Chemical Computing Group [19] was used to construct the QSAR model and perform molecular docking studies. MOE is a versatile computational platform widely used in bioinformatics, drug discovery, and integrated computational chemistry [20].

2.2. QSAR study

2.2.1. Data set

In this study, a set of 19 indole-imidazole derivatives that have aromatase inhibitor activity were obtained from the literature and used as a data set to build the QSAR model [21]. Their biological data and chemical structure are shown in Table 1. The data was split into a training set (15 compounds) to build the QSAR regression model and a test set (4 compounds) to estimate the model predictive capability.

2.2.2. Molecular descriptor

Molecular descriptor of the training set and test set compounds were chosen in order to find the best model, all the descriptors that calculated are 2D descriptors. The calculated molecular descriptors include Diameter, Petitjean index, Q_VSA_FPNEG, VSA_HYD, BCUT_SlogP_0, LogS, LogP (o/w), BCUT SMR0, and VDistEQ. The values of these descriptors are summarized in Table 2.

2.2.3. Partial least-square analysis

The PLS regression method [22] is a technique that merges the features of multiple regression and principal component analysis. It is used to build the linear model equation and find the relation between the dependent and independent variables [23].

2.2.4. Model development

The PLS method is used to create the model equation. A number of good QSAR equations were obtained; the most robust equations (Equations 1-4) were listed.

In principle, Q^2 and R^2 should have higher values. High Q^2 and R^2 (in general, $Q^2 > 0.5$ and $R^2 > 0.6$) indicate that the model is highly predictive [24]. Equation 1 was the best QSAR model equation, with a high square of the correlation coefficient (R^2 = 0.892) and a high cross- validation coefficient (Q^2 = 0.740).

Table 2. The value of chemical descriptors used in the QSAR models.

No	Compounds	Q_VSA_FPNEG	Diameter	Petitjean index	LogS	BCUT_SlogP_0	VSA_HYD	LogP(o/w)	BCUT SMR0	VDistEQ
1	10a	0.1015	10.0000	0.5000	-4.1067	-2.2412	246.4117	4.5795	-2.0324	3.0924
2	10b	0.0972	10.0000	0.5000	-4.2856	-2.2441	257.0992	4.7245	-2.0395	3.0924
3	10c	0.1005	11.0000	0.4545	-3.8621	-2.2570	269.5991	4.3825	-2.0352	3.2049
4	10d	0.1015	11.0000	0.4545	-4.8682	-2.2424	272.4621	5.3613	-2.0386	3.3119
5	10e	0.1031	9.0000	0.4440	-3.8117	-2.2408	241.3390	4.4265	-2.0324	2.9933
6	10f	0.0972	9.0000	0.4444	-4.2856	-2.2417	257.0992	4.7615	-2.0332	3.0313
7	10g	0.0919	10.0000	0.5000	-4.7595	-2.2450	272.9077	5.0595	-2.0404	3.1194
8	10h	0.0953	11.0000	0.4545	-4.3360	-2.2575	285.4076	4.7175	-2.0361	3.2261
9	10i	0.0958	10.0000	0.5000	-4.5806	-2.2421	262.2202	4.9145	-2.0332	3.1194
10	10j	0.1449	11.0000	0.4545	-4.6365	-2.2420	243.4992	4.4215	-2.0331	3.2261
11	10k	0.1005	10.0000	0.5000	-3.8621	-2.2495	269.5991	4.4195	-2.0327	3.1099
12	10l	0.0991	10.0000	0.5000	-4.1570	-2.2497	274.7201	4.5725	-2.0327	3.1738
13	10m	0.1952	11.0000	0.4545	-4.9186	-2.2504	300.7706	5.3543	-2.0389	3.3563
14	10n	0.1459	11.0000	0.4545	-4.2130	-2.2496	255.9920	4.0795	-2.0325	3.2683
15	10o	0.0953	10.0000	0.5000	-4.3360	-2.2513	285.4076	4.7175	-2.0399	3.1738
16	10p	0.0984	11.0000	0.4545	-3.9125	-2.2593	287.8923	4.3755	-2.0356	3.2683
17	10q	0.0952	11.0000	0.4545	-4.5964	-2.2569	257.0992	5.0115	-2.0352	3.2261
18	10r	0.0918	10.0000	0.5000	-5.0199	-2.2440	269.5991	5.3535	-2.0394	3.1194
19	10s	0.0917	10.0000	0.5000	-5.2803	-2.2403	241.3390	5.6475	-2.0320	3.1194

$$RP = 5.73683 - 0.33652 \times Diameter - 5.83615 \times Petitjeam\ index + 7.69314 \times Q\_VSA\_FPNEG$$
$$N = 15 \quad R^2 = 0.892 \quad F = 30.305 \quad Q^2 = 0.740 \quad (1)$$

$$RP = 81.79346 + 0.36196 \times logS - 5.58242 \times Petitjean\ index + 34.36055 \times BCUT\_SlogP\_0$$
$$N = 15 \quad R^2 = 0.862 \quad F = 22.930 \quad Q^2 = 0.754 \quad (2)$$

$$RP = 7.16269 - 0.19911 \times Diameter - 0.00703\ VSA\_HYD - 6.15301 \times Petitjean\ index$$
$$N = 15 \quad R^2 = 0.849 \quad F = 20.600 \quad Q^2 = 0.712 \quad (3)$$

$$RP = 79.56019 + 34.23290 \times BCUT\ SMR\ 0 - 2.40683 \times VDistEQ - 0.41722 \times logP(o/w)$$
$$N = 15 \quad R^2 = 0.811 \quad F = 15.808 \quad Q^2 = 0.664 \quad (4)$$

2.2.5. Validation of the QSAR models

The statistical parameters such as the square of the correlation coefficient (R^2), the cross-validation coefficient (Q^2) [25], the Fischer's value (F), the significance of the model(P) [26] and the standard error of estimate (SEE) [27], were used to verify the predictive capability of the QSAR model.

2.3. Modeling of new indole-imidazole derivatives

All of the new set of 57 compounds (indole-imidazole derivatives) were designed using ACD/ChemSketch [28]. The structure of all compounds was adjusted, saved as mol file format, opened by MOE software [19], and their energy was minimized. Equation 1 was selected to predict the biological activity of the designed compounds. The predicted bioactivity of the new compounds and their structure are shown in Table 3.

2.4. Crystal structure from PDB

The structure of aromatase is available in the Protein Data Bank (PDB: 3EQM Classification: Oxidoreductase, Organism(s): Homo sapiens) [29]. The structure was prepared using MOE software. Protein was initially prepared by adding hydrogen bond (3D protonation), and then water molecule and iron porphyrin were removed from the protein structure, the remaining amino acid and androstenedione were reserved to the docking study.

2.5. Database generation

All ligands (57 compounds) were sketched using ACD/ChemSketch. The molecular structures were then imported into the Molecular Operating Environment (MOE) software [19] in Mol file format. Each compound was subjected to Protonate 3D to assign appropriate protonation states, followed by hydrogen atom hiding. Energy minimization was performed using the MMFF94X force field with a gradient convergence criterion of 0.05. The optimized structures were subsequently added to a database file, which was saved and used for the molecular docking studies.

2.6. Procedure for performing molecular modeling

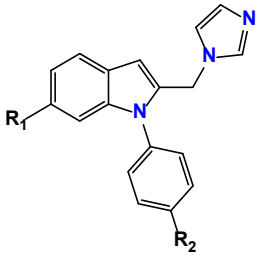
The molecular docking of the newly designed indole-imidazole derivatives at the active site of the aromatase enzyme (PDB ID: 3EQM) was performed using the MOE-Dock module implemented in the Molecular Operating Environment (MOE) software [19]. The docking parameters were established as follows: placement method, Triangle Matcher; scoring function, London dG; number of retained poses, 10; refinement, force field-based optimization; and rescoring function, none, with 10 poses retained [30]. The binding energies and root-mean-square deviation (RMSD) values obtained from the docking studies are summarized in Tables S1 and S2.

3. Results and discussion

3.1. QSAR study

In the present study, a dataset consisting of 19 indole-imidazole derivatives was collected from the literature [21]. Of these, 15 compounds were used as a training set to construct the QSAR model, while the remaining four compounds were used as an external test set for model validation. The QSAR model was developed using the partial least squares (PLS) method [22] implemented in the MOE software package [19]. Descriptor selection was carried out according to the following criteria [31]: (i) descriptors with a standard deviation lower than 0.0001 were eliminated; (ii) descriptors containing at least one missing value were removed; (iii) pairs of descriptors showing a correlation coefficient ≥ 0.8 were excluded to avoid

Table 3. Designed indole and imidazole derivatives and their predicted relative potency.



Compound	R_1	R_2	RP_{pre}*
1	CN	CF_3	1.260
2	NO_2	CF_3	1.610
3	NH_2	CF_3	1.064
4	CH_3	CF_3	0.893
5	Br	CF_3	0.840
6	COH	CF_3	1.183
7	CF_3	CF_3	1.635
8	OH	CF_3	1.112
9	CH_2CH_3	CF_3	0.818
10	H	NO_2	0.944
11	H	COOH	0.616
12	H	Cl	0.201
13	H	CN	0.560
14	H	SO_3H	1.025
15	H	CCl_3	1.923
16	H	$COOCH_3$	-0.186
17	OCH_2CH_3	CH_3	0.079
18	CH_2Ph	CH_3	-1.287
19	Br	CH_3	0.134
20	F	CH_3	0.191
21	OH	CH_3	0.370
22	NH_2	CH_3	0.331
23	NO_2	CH_3	0.934
24	$OCOCH_3$	CH_3	0.365
25	Cl	NO_2	0.861
26	Cl	SO_3H	0.943
27	Cl	N_2	0.595
28	Cl	$OCOCH_3$	-0.238
29	Cl	CH_2CH_3	0.052
30	Cl	OH	0.369
31	Cl	F	0.189
32	Cl	Br	0.134
33	Cl	CCl_3	1.804
34	Cl	COOH	0.551
35	Cl	OCH_2CH_3	-0.523
36	Cl	COH	0.402
37	Cl	$COCH_3$	0.350
38	Cl	CH_2OCH_3	-0.523
39	CCl_3	CN	2.123
40	COCl	CN	1.385
41	CF_3	CN	1.269
42	CH_2OCH_3	CN	-0.151
43	CN	CN	0.889
44	CH_3	NO_2	0.862
45	CH_3	SO_3H	0.944
46	CH_3	$COOCH_2CH_3$	-0.394
47	CH_3	COCl	1.021
48	COOH	CF_3	1.460
49	OSO_2NH_2	CF_3	1.006
50	SO_3H	CF_3	1.793
51	SO_3H	CH_3	1.164
52	SO_3H	CN	1.448
53	H	CH_2CH_3	0.089
54	H	Br	0.172
55	CH_3	COOH	0.553
56	CH_2CH_3	CH_3	0.124
57	NH_2	CN	0.668

* RP_{pre} = Predicted relative potency by Model 1.

multicollinearity; and (iv) descriptors exhibiting a Pearson's correlation coefficient (|r|) lower than 0.3 with the relative potency of indole-imidazole derivatives were discarded. The remaining descriptors were retained for subsequent QSAR modeling. The Pearson correlation matrix for the selected descriptors is presented in Table 4 and was generated using MOE software to ensure the appropriateness of the chosen descriptors.

The QSAR model with the best performance was selected based on its statistical robustness, where the values of $Q^2 > 0.5$ and $R^2 > 0.6$ indicate good predictive capability. Accordingly, Model 1 (Equation 1) was identified as the optimal model, exhibiting a high coefficient of determination (R^2 = 0.892; Figure 1). Additionally, the model showed a high Fisher ratio (F = 30.02) and a strong cross-validated correlation coefficient (Q^2 = 0.741; Figure 2), confirming its reliability and predictive

Table 4. Pearson correlation matrix between molecular descriptors and relative potency (RP).

	RP	BCUT_SlogP_0	Diameter	LogP (o/w)	LogS	Petitjean index	VSA_HYD	Q_VSA_FPNEG
RP	100	44	-53	-45	44	-52	-73	36
BCUT_SlogP_0	44	100	-70	36	-44	32	-70	10
Diameter	-53	-70	100	-29	4	-24	39	40
LogP (o/w)	-45	36	-29	100	-85	59	32	-62
LogS	44	-44	4	-85	100	-55	-14	15
Petitjean index	-52	32	-24	59	-55	100	22	-42
VSA_HYD	-73	-70	39	32	-14	22	100	-55
Q_VSA_FPNEG	36	10	40	-62	15	-42	-55	100

Table 5. Statistical parameters for the best QSAR model equations.

Equation	No of training set	No of test set	R^2	Q^2	R^2_{pre}	RMSE	SEE	F value	P value
1	15	3	0.892	0.741	0.840	0.0907	0.1390	30.305	< 0.005
2	15	3	0.862	0.754	0.950	0.1005	0.1174	22.934	< 0.005
3	15	3	0.849	0.712	0.997	0.1051	0.1220	20.600	< 0.005
4	15	3	0.823	0.715	0.916	0.1137	0.1328	17.090	<0.005

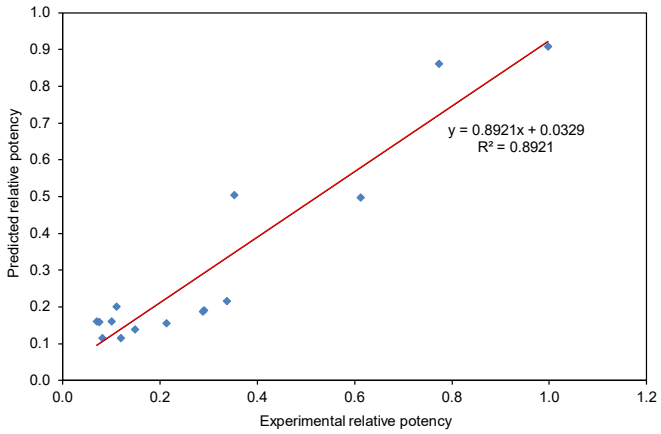


Figure 1. Correlation plot of experimental versus predicted relative potency (RP) for the training set used in QSAR model development.

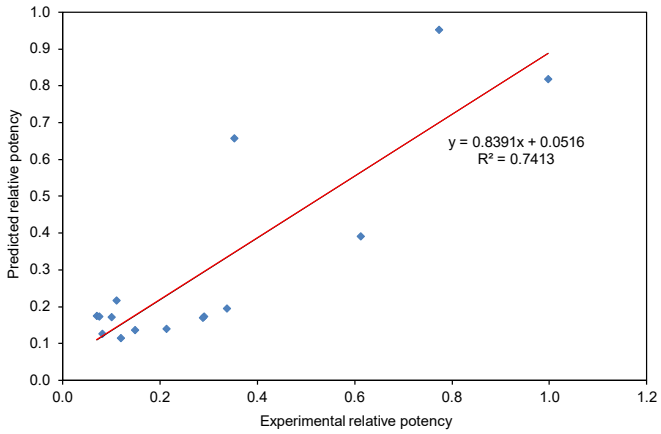


Figure 2. Correlation plot of experimental versus predicted relative potency (RP) obtained from cross-validation of the training set.

power. The small difference between the R^2 and Q^2 values (0.1507) indicates that Model 1 is not affected by overfitting [32]. Furthermore, external validation using the test set produced a satisfactory predictive correlation (R^2 = 0.84; Figure 3), further supporting the robustness of the developed QSAR model.

Equation 1 is based on three molecular descriptors. The relative potency (RP) increases with decreasing values of the Petitjean index and molecular diameter, while it increases with an increase in the fractional negative polar van der Waals surface area (Q_VSA_FPNEG). The statistical parameters of the best-performing models are summarized in Table 5.

Wang *et al*. observed that any proton or a small electron withdrawing group (R_2) on the phenyl ring, that connected to indole nitrogen, might enhance aromatase inhibitor activities, and any bulky group should be avoided to keep a relatively small value for these kinds of molecules [21]. Therefore, in this study, the design of the new indole-imidazole derivatives depended on this observation, but also some molecules that possess an electron-donating group at the R_2 position were designed to make a comparison between the two types of substituents.

A group of 57 compounds was designed and modeled, and their biological activity was predicted using Model 1. All compounds that exhibit high inhibitor activity with relative potency (RP) range of 2.12 to 1.02 have an electron withdrawing group in the *meta* position of the indole ring R_1. Compounds 17 and 30, which possess an electron-donating group in R_2, show a low inhibitor activity of 0.0793 and 0.3690, respectively.

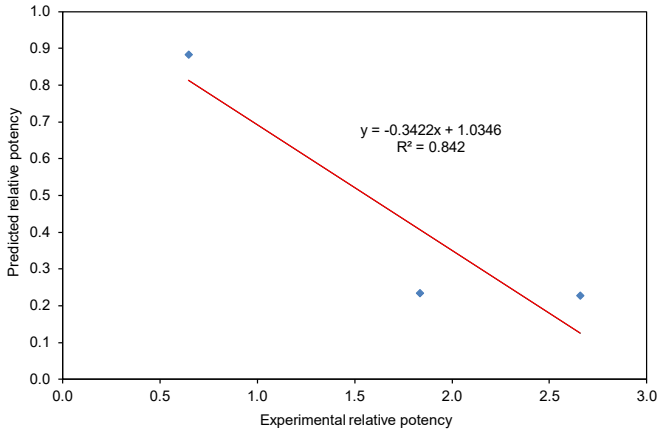


Figure 3. Correlation between experimental and predicted relative potency (RP) values for the test set.

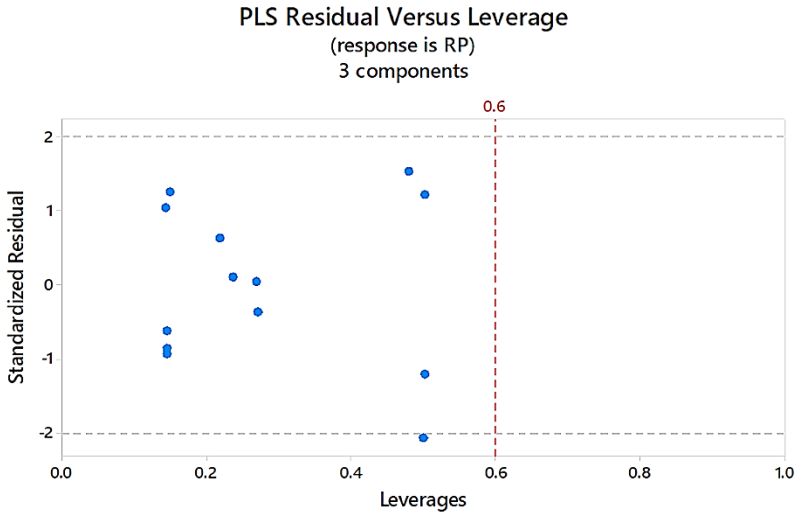


Figure 4. Williams plot of the developed QSAR model (Model 1).

All compounds that show high inhibitory activity towards aromatase (more than letrozole) have an electron-withdrawing group at the *para* position of the phenyl ring R_2. So, the R_2 favorite electron withdrawing group and these support the observation of Wang *et al.* [21]. The most potent compounds in the newly designed compounds are 39, 15, 33, 7, 50, 2, and 48 (RP: 2.12, 1.92, 1.80, 1.63, 1.79, 1.61, 1.46), which are more powerful in inhibiting aromatase as compared to letrozole.

3.1.1. Application domain of the QSAR model

The applicability domain (AD) of the developed QSAR model was defined using the leverage approach [32], which is commonly employed to evaluate the reliability of model predictions. The leverage value (h_i) for each compound was calculated according to Equation (5):

$$h_i = x_i^T (X^T X)^{-1} x_i \ (i = 1, \dots, n) \quad (5)$$

where x_i is the descriptor row vector of the query compound, X represents the $n \times k$ matrix of the k selected descriptor values for the n n compounds in the training set, and the superscript T denotes matrix transposition [31].

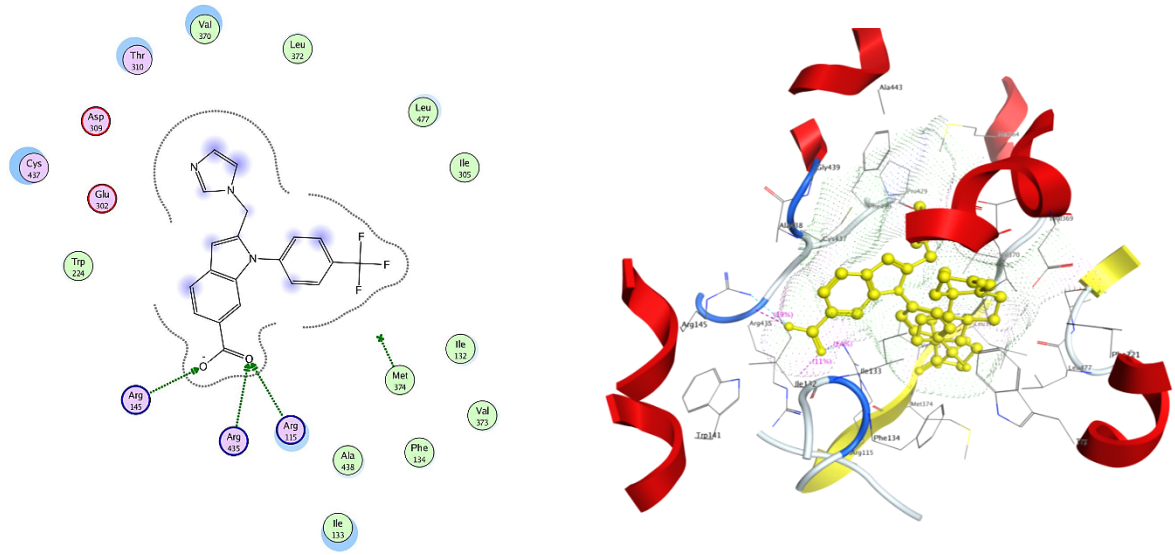
The warning leverage value (h*) for Model 1 was set to 0.6. As shown in Figure 4, no compounds fall outside the applicability domain of Model 1, and no response outliers were observed. This indicates that the developed QSAR model is reliable and sufficiently robust to predict the aromatase inhibitory activity of the compounds studied.

3.2. Molecular docking study

Molecular docking is a widely used computational approach to investigate the interactions and preferred binding modes of small-molecule ligands with protein receptors during complex formation [33]. In the present study, docking was employed to gain insight into the binding behavior of the designed indole-imidazole derivatives within the active site of the aromatase enzyme.

The ligand-binding pocket of aromatase, defined within a radius of 4 Å, comprises both nonpolar residues (Ala306, Ala307, Ile133, Ile305, Leu477, Met374, Phe134, Phe221, Trp224, Val369, Val370, and Val373) and polar residues (Arg115, Arg375, Asp309, Asp371, Ser478, Thr310, and Glu302) [34]. Previous computational studies on nonsteroidal aromatase inhibitors, such as letrozole, have shown that interactions with these residues play a crucial role in stabilizing ligand binding and determining inhibitory activity [35].

The 57 newly designed compounds were docked into the active site of the aromatase enzyme to investigate their binding energies and modes of interaction (Supplementary Table S2).



(a) (b)

Figure 5. (a) 2D and (b) 3D interaction patterns of compound 48 with the aromatase enzyme.

The accuracy of the predicted docking poses was evaluated using the root mean square deviation (RMSD) values [36], where RMSD values lower than 2.0 Å are generally considered indicative of reliable docking results. The calculated RMSD values are summarized in Supplementary Tables S1 and S2. Among the compounds studied, compound 48 exhibited an RMSD value of 1.5 Å, confirming the reliability of its predicted binding pose. Kang *et al.* [12] previously investigated the binding energies of phenyl indole derivatives with aromatase and compared them with those of letrozole, reporting that the most active compound displayed a binding energy lower than that of letrozole by approximately 3 units.

In the present study, compound 48 demonstrated a significantly lower binding energy, indicating higher complex stability, with a value of 7 units lower than that of letrozole. Furthermore, compound 48 forms three hydrogen bond interactions with key active site residues Arg145, Arg115, and Arg435, with hydrogen-bond distances of 2.19, 1.84, and 2.31 Å, respectively. The two- and three-dimensional interaction patterns of compound 48 within the aromatase active site are illustrated in Figure 5.

4. Conclusions

In this study, 57 indole-imidazole derivatives were designed and their potential aromatase inhibitory activity was evaluated using a QSAR model developed with three descriptors (Q_VSA_FPNEG, diameter, and Petitjean index). The model was built using the partial least squares (PLS) method and was validated both internally and externally. The application of the QSAR model identified 18 compounds with predicted values of relative potency (RP) higher than that of letrozole. These compounds were further subjected to molecular docking to investigate their binding modes and interaction energies with aromatase. Among them, compound 48 exhibited the lowest binding energy (-27.68 kcal/mol), which is 7 kcal/mol lower than that of letrozole (-20.18 kcal/mol), forming three hydrogen bond interactions with key residues at the enzyme active site. These findings indicate that compound 48 is a promising aromatase inhibitor. Further experimental studies, including in vitro evaluation of aromatase inhibitory activity, are warranted to validate these computational predictions.

Disclosure statement

Conflict of interest: The authors declare that they have no conflict of interest.
Ethical approval: All ethical guidelines have been adhered to.

CRediT authorship contribution statement

Conceptualization: Ahmed Elsadig Mohammed Saeed, Rawan Mustafa Ali Massad; Methodology: Rawan Mustafa Ali Massad; Software: Rawan Mustafa Ali Massad; Validation: Rawan Mustafa Ali Massad, Ahmed Elsadig Mohammed Saeed; Formal analysis: Rawan Mustafa Ali Massad; Investigation: Rawan Mustafa Ali Massad; Resource: Rawan Mustafa Ali Massad; Data curation: Rawan Mustafa Ali Massad; Writing- original draft: Rawan Mustafa Ali Massad; Writing - Review & Editing: Ahmed Elsadig Mohammed Saeed; Supervision: Ahmed Elsadig Mohammed Saeed; Project administration: Rawan Mustafa Ali Massad, Ahmed Elsadig Mohammed Saeed; Funding acquisition: Rawan Mustafa Ali Massad.

Supporting information

The online version of this article contains supplementary material, which is available to authorized users.

ORCID and Email

Rawan Mustafa Ali Massad
rawanmus5afa@gmail.com
cgs@sustech.edu
https://orcid.org/0009-0008-0645-8502
Ahmed Elsadig Mohammed Saeed
aemsaeed@gmail.com
https://orcid.org/0000-0002-7317-8040

References

[1]. Spinello, A.; Ritacco, I.; Magistrato, A. Recent advances in computational design of potent aromatase inhibitors: open-eye on endocrine-resistant breast cancers. *Expert Opin. Drug Discov.* **2019**, *14* (10), 1065–1076.
[2]. Bassatne, A.; Bou Khalil, A.; Chakhtoura, M.; Arabi, A.; Van Poznak, C.; El-Hajj Fuleihan, G. Effect of antiresorptive therapy on aromatase inhibitor induced bone loss in postmenopausal women with early-stage breast cancer: A systematic review and meta-analysis of randomized controlled trials. *Metabolism* **2022**, *128*, 154962.
[3]. Kaur, K.; Jaitak, V. Recent Development in Indole Derivatives as Anticancer Agents for Breast Cancer. *ACAMC.* **2019**, *19* (8), 962–983.
[4]. De Filippis, B.; Agamennone, M.; Ammazzalorso, A.; Amoroso, R.; Giampietro, L.; Maccallini, C.; Sağlık, B. N.; De Simone, C.; Zuccarini, M.; Kaplancıklı, Z. A.; Fantacuzzi, M. Discovery and Evaluation of Novel

Sulfonamide Derivatives Targeting Aromatase in ER+ Breast Cancer. *Pharmaceuticals* **2025**, *18* (8), 1206.
[5]. Boszkiewicz, K.; Piwowar, A.; Petryszyn, P. Aromatase Inhibitors and Risk of Metabolic and Cardiovascular Adverse Effects in Breast Cancer Patients—A Systematic Review and Meta-Analysis. *JCM.* **2022**, *11* (11), 3133.
[6]. Gobbi, S.; Rampa, A.; Belluti, F.; Bisi, A. Nonsteroidal Aromatase Inhibitors for the Treatment of Breast Cancer: An Update. *ACAMC.* **2014**, *14* (1), 54–65.
[7]. Sabet, R.; Khabnadideh, S.; Mohseni, S.; Mojaddami, A.; Rezaei, Z. QSAR analysis of some azole derivatives as potent aromatase inhibitors. *Trends in Pharmaceutical Sciences and Technologies* **2018**, *4*, 205–218.
[8]. Numazawa, M.; Komatsu, S.; Tominaga, T.; Yamashita, K. Structure-Activity Relationships of Estrogen Derivatives as Aromatase Inhibitors. Effects of Heterocyclic Substituents. *Chem. Pharm. Bull.* **2008**, *56* (9), 1304–1309.
[9]. Ghosh, D.; Lo, J.; Egbuta, C. Recent Progress in the Discovery of Next Generation Inhibitors of Aromatase from the Structure–Function Perspective. *J. Med. Chem.* **2016**, *59* (11), 5131–5148.
[10]. Edris, A.; Abdelrahman, M.; Osman, W.; Sherif, A. E.; Ashour, A.; Garelnabi, E. A.; Ibrahim, S. R.; Bafail, R.; Samman, W. A.; Ghazawi, K. F.; Mohamed, G. A.; Alzain, A. A. Design of Novel Letrozole Analogues Targeting Aromatase for Breast Cancer: Molecular Docking, Molecular Dynamics, and Theoretical Studies on Gold Nanoparticles. *Metabolites* **2023**, *13* (5), 583.
[11]. Chamduang, C.; Pingaew, R.; Prachayasittikul, V.; Prachayasittikul, S.; Ruchirawat, S.; Prachayasittikul, V. Novel triazole-tetrahydroisoquinoline hybrids as human aromatase inhibitors. *Bioorg. Chem.* **2019**, *93*, 103327.
[12]. Kang, H.; Xiao, X.; Huang, C.; Yuan, Y.; Tang, D.; Dai, X.; Zeng, X. Potent aromatase inhibitors and molecular mechanism of inhibitory action. *Eur. J. Med. Chem.* **2018**, *143*, 426–437.
[13]. Ozcan-Sezer, S.; Ince, E.; Akdemir, A.; Ceylan, O. O.; Suzen, S.; Gurer-Orhan, H. Aromatase inhibition by 2-methyl indole hydrazone derivatives evaluated via molecular docking and *in vitro* activity studies. *Xenobiotica* **2018**, *49* (5), 549–556.
[14]. Singla, R.; Gupta, K. B.; Upadhyay, S.; Dhiman, M.; Jaitak, V. Design, synthesis and biological evaluation of novel indole-benzimidazole hybrids targeting estrogen receptor alpha (ER-α). *Eur. J. Med. Chem.* **2018**, *146*, 206–219.
[15]. Pingaew, R.; Mandi, P.; Prachayasittikul, V.; Prachayasittikul, S.; Ruchirawat, S.; Prachayasittikul, V. Synthesis, molecular docking, and QSAR study of sulfonamide-based indoles as aromatase inhibitors. *Eur. J. Med. Chem.* **2018**, *143*, 1604–1615.
[16]. Xu, Y.; He, Z.; Liu, H.; Chen, Y.; Gao, Y.; Zhang, S.; Wang, M.; Lu, X.; Wang, C.; Zhao, Z.; Liu, Y.; Zhao, J.; Yu, Y.; Yang, M. 3D-QSAR, molecular docking, and molecular dynamics simulation study of thieno[3,2-*b*]pyrrole-5-carboxamide derivatives as LSD1 inhibitors. *RSC. Adv.* **2020**, *10* (12), 6927–6943.
[17]. Du, M.; Zhang, D.; Hou, Y.; Zhao, X.; Li, Y. Combined 2D-QSAR, Principal Component Analysis and Sensitivity Analysis Studies on Fluoroquinolones' Genotoxicity. *Intern. J. Environ. Res. Public Health* **2019**, *16* (21), 4156.
[18]. Begum, S.; Jaswanthi, P.; Venkata Lakshmi, B.; Bharathi, K. QSAR studies on indole-azole Analogues using DTC tools; imidazole ring is more favorable for aromatase inhibition. *J. Indian Chem. Soc.* **2021**, *98* (1), 100016.
[19]. Molecular Operating Environment (MOE), 2024.0601 Chemical Computing Group ULC, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, 2026.
[20]. Rustagi, V.; Gupta, S. R.; Singh, A.; Singh, I. K. Beyond trial and error: Leveraging advanced software for Therapeutic discovery. *Chem Biol Lett* **2025**, *12* (1), 1251.
[21]. Wang, R.; Shi, H.; Zhao, J.; He, Y.; Zhang, H.; Liu, J. Design, synthesis and aromatase inhibitory activities of novel indole-imidazole derivatives. *Bioorg. amp; Med. Chem. Lett.* **2013**, *23* (6), 1760–1762.
[22]. Mateos-Aparicio, G. Partial Least Squares (PLS) Methods: Origins, Evolution, and Application to Social Sciences. *Commun. Stat. - Theory Methods* **2011**, *40* (13), 2305–2317.
[23]. Park, D. S.; Kim, J. M.; Lee, Y. B.; Ahn, C. H. QSID Tool: a new three-dimensional QSAR environmental tool. *J. Comput Aided Mol Des* **2008**, *22* (12), 873–883.
[24]. Li, R.; Du, Y.; Shen, J. Molecular modelling studies on cinnoline-based BTK inhibitors using docking and structure-based 3D-QSAR. *SAR QSAR Environ. Res.* **2018**, *29* (11), 847–873.
[25]. Tong, W.; Lowis, D. R.; Perkins, R.; Chen, Y.; Welsh, W. J.; Goddette, D. W.; Heritage, T. W.; Sheehan, D. M. Evaluation of Quantitative Structure−Activity Relationship Methods for Large-Scale Prediction of Chemicals Binding to the Estrogen Receptor. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (4), 669–677.
[26]. Grabowski, T.; Jaroszewski, J. J.; Feder, M. Correlations between the selected parameters of the chemical structure of drugs and between-subject variability in area under the curve. *Med Chem Res* **2012**, *22* (4), 1812–1824.
[27]. García-Domenech, R.; Alarcón-Elbal, P.; Bolas, G.; Bueno-Marí, R.; Chordá-Olmos, F. A.; Delacour, S. A.; Mouriño, M. C.; Vidal, A.; Gálvez, J. Prediction of acute toxicity of organophosphorus pesticides using topological indices. *SAR QSAR Environ. Res.* **2007**, *18* (7-8), 745–755.
[28]. ACD/ChemSketch, Version 2012; Advanced Chemistry Development, Inc.: Toronto, ON, Canada, 2012; https://www.acdlabs.com
[29]. Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature* **2009**, *457* (7226), 219–223.
[30]. Haidar, S.; Marminon, C.; Aichele, D.; Nacereddine, A.; Zeinyeh, W.; Bouzina, A.; Berredjem, M.; Ettouati, L.; Bouaziz, Z.; Le Borgne, M.; Jose, J. QSAR Model of Indeno[1,2-b]indole Derivatives and Identification of N-isopentyl-2-methyl-4,9-dioxo-4,9-Dihydronaphtho[2,3-b]furan-3-carboxamide as a Potent CK2 Inhibitor. *Molecules* **2019**, *25* (1), 97.
[31]. Qin, L.; Zhang, X.; Chen, Y.; Mo, L.; Zeng, H.; Liang, Y. Predictive QSAR Models for the Toxicity of Disinfection Byproducts. *Molecules* **2017**, *22* (10), 1671.
[32]. Qin, L.; Liu, S.; Chen, F.; Xiao, Q.; Wu, Q. Chemometric model for predicting retention indices of constituents of essential oils. *Chemosphere* **2013**, *90* (2), 300–305.
[33]. Lin, X.; Li, X.; Lin, X. A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules* **2020**, *25* (6), 1375.
[34]. Roy, P. P.; Roy, K. Docking and 3D-QSAR studies of diverse classes of human aromatase (CYP19) inhibitors. *J. Mol Model* **2010**, *16* (10), 1597–1616.
[35]. Prior, A. M.; Yu, X.; Park, E.; Kondratyuk, T. P.; Lin, Y.; Pezzuto, J. M.; Sun, D. Structure-activity relationships and docking studies of synthetic 2-arylindole derivatives determined with aromatase and quinone reductase 1. *Bioorg. amp; Med. Chem. Lett.* **2017**, *27* (24), 5393–5399.
[36]. Yusuf, D.; Davis, A. M.; Kleywegt, G. J.; Schmitt, S. An Alternative Method for the Evaluation of Docking Performance: RSR vs RMSD. *J. Chem. Inf. Model.* **2008**, *48* (7), 1411–1422.