Turkish Computational and Theoretical Chemistry

Turkish Comp Theo Chem (TC&TC)

Volume(Issue): 4(1) – Year: 2020 – Pages: 1-11

e-ISSN: 2602-3237

https://doi.org/10.33435/tcandtc.545369

Received: 29.03.2019 ***Accepted***: 29.04.2020 *Research Article*

Quantitative Structure-Activity Relationships of 1.2.3 Triazole Derivatives as Aromatase Inhibition Activity

Mebarka OUASSAF [a], Salah BELAIDI[1,a], Imane BENBRAHIM [a], Houmam BELAIDI [a,b], Samir CHTITA [c]

[a] University of Biskra, Group of Computational and Medicinal Chemistry, LMCE Laboratory, BP 145 Biskra 07000, Algeria

[b] Institut des Sciences Chimiques de Rennes, UMR 6226 CNRS-Université de Rennes 1, Campus de Beaulieu, 35042, Rennes Cedex, France

[c] Laboratory of Physical Chemistry of Materials, Faculty of Sciences Ben M'Sik, University Hassan II, Casablanca, Morocco

Abstract: Aromatase is an estrogen biosynthesis enzyme belonging to the cytochrome P450 family that catalyzes the rate-limiting step of converting androgens to estrogens. As it is pertinent toward tumor cell growth promotion aromatase is a lucrative therapeutic target for breast cancer. In the pursuit of robust aromatase inhibitors, a set of thirty 1-substituted mono- and bis-benzonitrile or phenyl analogs of 1.2.3-triazole letrozole were employed in quantitative structure activity relationship (QSAR) study using multiple linear regression (MLR).The results demonstrated good predictive ability for the MLR model. After dividing the dataset into training and test set. The models were statistically robust internally (R^2 = 0.982) and the model predictability was tested by several parameters, including the external criteria (R^2_{pred} = 0.851. CCC= 0.946). Insights gained from the present study are anticipated to provide pertinent information contributing to the origins of aromatase inhibitory activity and therefore aid in our on-going quest for aromatase inhibitors with robust properties.

Keywords: *1.2.3-triazole, Aromatase inhibitors, QSAR, MLR.*

1. Introduction

Breast cancer is a kind of malignant tumor for women [1], which account about 30% incidences for all malignant tumors in different age groups of women. In recent years, the mortality rate of breast cancer also shows an increasing trend and it has become one of the major causes of cancer death in female [2].

A great majority of breast cancers is hormone-dependent [3] and it is widely accepted that estrogen plays an important role in the genesis and evolution of breast tumors [4]. Aromatase (CYP19) a cytochrome P450 enzyme is responsible for the conversion of androgens including androstenedione and testosterone into estrogens [5], therefore it is considered as a particularly attractive target for inhibition in the endocrine treatment of hormone-dependent breast cancer such as Non-steroidal aromatase inhibitors:aminoglutethimide [6]anastrozole (Arimidex™)[7] and letrozole (Femara™)[8].

Competitively inhibit the enzymatic activity of aromatase in a reversible manner and play an

[1] Corresponding Authors

e-mail:prof.belaidi@gmail.com

important role in the endocrine treatment for hormone-dependent breast cancers.

Triazoles are common pharmacophore found in a diverse range of biologically active molecules due to their potential structural features[9].Among the AIs (inhibitory aromatase)letrozole and anastrozole both containing 1.2.4-triazole ring, were approved by the Food and Drug Administration (FDA) and using as the first-line therapy in the treatment of breast cancer in postmenopausal women since they have been shown to be superior to tamoxifen[10].Based on the AIs, the triazole ring plays a pivotal role in chelation with heme iron along the line [11].Touaibia et al has studied on an aromatase inhibitory activity of various substituted-1.2.3-triazole letrozole-based analogs[12].The results revealed that 1.2.3-trizole analog of letrozole showed equipotent activity to the parent compound.

In last decades, quantitative Structure-Activity Relationships (QSAR)[13], have been applied in many areas enabling to prevent time consuming and cost during the analysis of biological activities of interest. The main hypothesis involved in any QSAR is the assumption that the variation of the behavior of chemical compounds, as expressed by any experimentally measured biological or physicochemical property, can be correlated with numerical entities related to some aspect of the chemical structure termed molecular descriptors [14-16].

Descriptors are generally used to describe different characteristics/ attributes of the chemical structure in order to yield information about the activity/property being studied.

Herein, a series of 1-substituted mono- and bis-benzonitrile or phenyl analogs of 1.2.3-triazole letrozole are employed for QSAR modeling of the aromatase inhibitory activity. A diverse set of quantum chemical and molecular descriptors were employed to provide numerical description of the investigated compounds, using multiple linear regressions (MLR).

2. Materials and Methods

2.1. Data Set

A series of thirty molecules belonging to 1.2.3-triazole derivatives have aromatase inhibitory activity, were taken from literature [17]. The studied compounds were randomly divided into training set (twenty-four compounds) and test set (six compounds). Training and test set compounds are represented in (Table1). These compounds in the series were sketched using ChemDraw module, which is available in ChemOffice.

2.2. Descriptors Generation

Firstly, the thirty investigated molecules were pre-optimized by the Molecular Mechanics Force Field (MM+) included in HyperChem version 8.03 package [18]. After that, the resulted minimized structures were further refined using the semi empirical PM3 Hamiltonian implemented also in HyperChem. We chose a gradient norm limit of 0.01kcal/Å for the geometry optimization.

QSAR properties module from HyperChem 8.03 was used to calculate physical and chemical proprieties of a series of thirty 1.2.3-triazole derivatives: the molar polarizability (Pol), the molar refractivity (REF), logarithm of partition coefficient octanol/water (log P), hydration energy (HE), Surface area grid (S) and molar weight (M); these properties are described in (Table 2).

Then, these 1.2.3-triazoles were re-optimized by using Gaussian 09 program package [19] at the density functional theory level (DFT) using Becke's three-parameter Lee-Yang- Parr (B3LYP) With the 6-311G (d. p) basis set, this theory was used to calculate a number of electronic descriptors (Table 3).such as; LUMO energies and atomic net charges (qN1, qN2, qN3, qC4 and qC5place the atoms shown in Fig1)

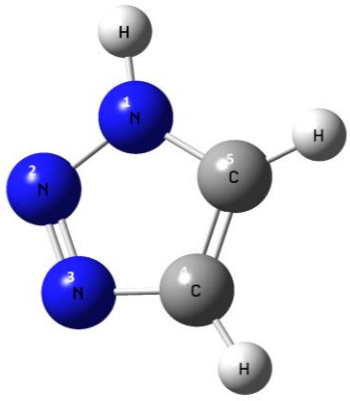
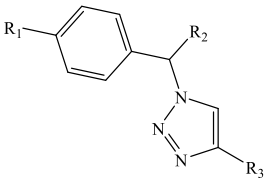


Fig .1.3D conformation of 1H-1,2,3-triazole

Table 1 Chemical structures of the 1.2.3-triazole derivatives



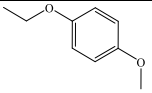
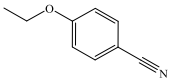
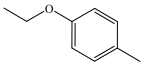
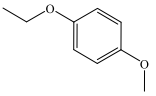
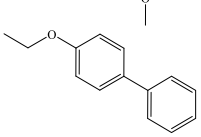
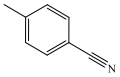
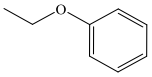
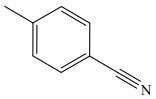
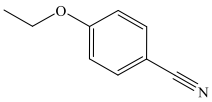
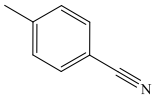
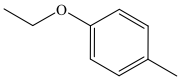
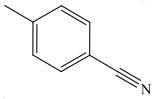
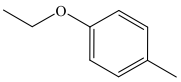
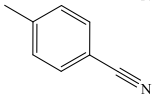
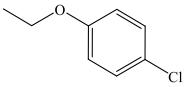
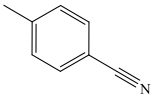
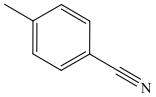
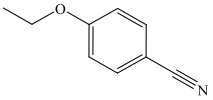
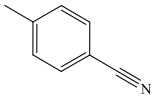
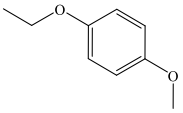
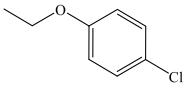
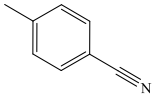
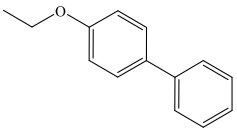
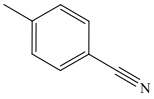
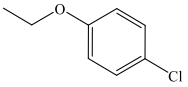
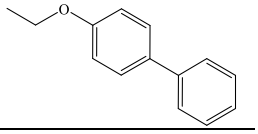
	R1	R2	R3		R1	R2	R3
1	CN	H	-Ph	16	H	-Ph	
2	CH3	H		17	H	-Ph	H
3	CN	H		18	H	-Ph	$-(CH_2)_2CH_3$
4	CN	H		19	H	-Ph	$-(CH_2)_5CH_3$
5	CN	H		20	H	-Ph	
6	CN		$-(CH_2)_2CH_3$	21	H	-Ph	
7	H	H	-Ph	22	H	-Ph	
8	CN		H	23	H	-Ph	
9	CN		$-(CH_2)_5CH_3$	24	H	-Ph	
10	CN			25	CN	H	$-(CH_2)_2CH_3$
11	CN			26	CN		
12	CN			27	CN	H	
13	CN			28	H	H	
14	CN			29	CN		$-(CH_2)_9CH_3$
15	H	-Ph		30	H	-Ph	

Table 2. Physicochemical descriptors

	S (A°2)	M(uma)	HE [Kcal.mol-1]	Pol. [Å3]	Ref (A°3)	TPSA	ABS	LE	LogP
1	458.7	246.27	-14.22	28.06	85.85	87.5	90.19	0.380	0.89
2	510.0	287.28	-21.48	30.55	91.44	96.7	78.79	0.320	1.11
3	507.0	276.30	-16.11	30.53	90.74	87.5	87.00	0.320	1.54
4	520.1	292.30	-18.99	31.17	92.83	49.1	83.82	0.350	0.39
5	516.9	338.37	-18.15	38.36	114.9	30.7	87.00	0.300	1.99
6	621.7	258.43	-5.25	34.90	98.20	30.7	81.98	0.410	4.14
7	450.9	227.31	-6.36	26.78	77.86	30.7	98.400	0.280	2.06
8	469.2	251.29	-11.96	28.68	85.76	63.7	95.210	0.390	1.82
9	520.5	285.31	-17.81	93.62	31.75	78.3	81.987	0.515	1.52
10	609.4	327.39	-14.54	108.5	37.25	78.3	81.987	0.265	2.42
11	822.9	425.52	-11.87	140.7	50.10	87.5	78.799	0.231	5.19
12	632.8	367.45	-13.87	120.4	41.98	87.5	78.799	0.216	3.11
13	633.7	361.40	-17.14	123.4	41.41	111.	70.591	0.233	2.22
14	660.4	377.40	-18.88	124.0	42.05	96.7	75.614	0.215	2.71
15	683.0	411.85	-18.50	128.7	43.97	87.5	78.799	0.189	2.49
16	702.0	407.430	-20.50	130.4	44.52	49.1	92.029	0.241	1.72
17	747.0	453.500	-20.86	152.5	51.74	30.7	98.402	0.415	3.32
18	454.7	235.29	83.66	28.04	0.331	30.7	98.402	0.331	2.08
19	544.1	277.370	98.57	33.55	0.282	30.7	98.402	0.282	2.97
20	629.8	319.450	112.3	39.05	0.285	30.7	98.402	0.285	4.16
21	570.5	311.390	113.4	37.70	0.284	30.7	98.402	0.284	2.78
22	588.6	317.430	110.4	38.28	0.243	39.9	95.217	0.243	3.66
23	614.8	361.830	118.7	40.27	0.239	63.7	87.010	0.239	3.05
24	627.5	352.400	119.0	40.19	0.237	39.9	95.217	0.237	2.99
25t	442.2	212.25	23.91	70.97	0.44	78.3	81.98	0.44	1.09
26t	653.1	310.44	36.78	103.1	0.24	39.9	95.21	0.24	3.86
27t	701.0	369.47	42.76	122.3	0.21	78.3	81.98	0.21	3.6
26t	642.4	375.86	42.1	124.7	0.25	39.9	95.21	0.25	1.97
29t	549.6	290.32	32.37	96.62	0.373	63.4	87.00	0.373	0.83
30t	627.8	341.41	40.18	120.1	0.28	39.9	95.217	0.28	1.71

Molinspiration[21]software was used to obtain TPSA parameter (topological polar surface area) which was used to calculate the percentage of absorption (%ABS) according to the equation[22]:

$$\%ABS = 109 \pm 0.345 \times TPSA \quad (a)$$

Also, we calculated the Ligand Efficiency (LE) according to the equation:

$$LE = 1.4 * \frac{pIC_{50}}{NH} \quad (b)$$

Where, N*H* is the number of heavy atoms [23].

2.3. Regression Analysis

Multiple linear regression analysis of molecular descriptors was carried out using the stepwise strategy in SPSS version 19 for Windows [24].

Table 3. Quantum descriptors

	E_{LUMO} [au]	qN_1	qN_2	qC_4	qC_5
1	-0.06	-0.49	-0.001	0.01	0.21
2	-0.09	-0.49	-0.012	0.05	0.22
3	-0.06	-0.49	-0.017	0.06	0.22
4	-0.06	-0.49	-0.017	0.28	0.22
5	-0.08	-0.70	0.019	0.27	0.25
6	-0.07	-0.48	-0.023	0.10	0.18
7	-0.01	-0.49	-0.016	0.11	0.17
8	-0.01	-0.49	0.023	0.08	0.17
9	-0.07	-0.68	0.007	0.28	0.22
10	-0.07	-0.49	-0.023	0.10	0.18
11	-0.07	-0.48	-0.019	0.36	0.18
12	-0.07	-0.49	-0.010	0.09	0.20
13	-0.07	-0.49	-0.006	0.09	0.21
14	-0.07	-0.49	-0.015	0.09	0.20
15	-0.07	-0.49	-0.015	0.09	0.21
16	-0.02	-0.48	-0.007	0.09	0.28
17	-0.01	-0.48	-0.019	0.09	0.18
18	-0.02	-0.49	-0.003	0.018	0.21

19	-0.01	-0.49	-0.013	0.11	0.17
20	-0.02	-0.47	-0.010	0.36	0.18
21	-0.02	-0.47	-0.005	0.36	0.10
22	-0.03	-0.52	0.034	0.04	0.22
23	-0.03	-0.48	-0.002	0.09	0.20
24	-0.02	-0.47	-0.010	0.36	0.18
25t	-0.25	-0.5	-0.005	-0.29	0.01
26t	-0.26	-0.49	0.017	-0.06	0225
27t	-0.07	-0.49	-0.02	0.101	0.18
26t	-0.02	-0.48	-0.001	0.098	0.20
29t	-0.04	-0.49	-0.011	0.05	0.27
30t	-0.06	-0.02	-0.558	-0.00	0.20

3. Validation of the QSAR Model

Testing the stability predictive power and generalization ability of the models is a very important step in QSAR study, as for the validation of predictive power of a QSAR model. Two basic principles (internal validation and external validation) are used.

3.1. Internal validation (LOO validation technique)

Predictive power of the mentioned models was tested by leave-one-out *cross*-validation method the calculation of the following parameters: *cross*-validated coefficient of determination (r^2_{CV}) adjusted coefficient of determination (r^2adj) predicted residual sum of squares (*PRESS*) total sum of squares (*TSS*) and standard deviation based on predicted residual sum of squares (S_{PRESS}) [25, 26].

These parameters are defined as below:

$$R^2_{cv} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)} = 1 - \frac{PRESS}{TSS} \quad (1)$$

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-p-1} = \frac{(n-1)R^2 - p}{n-p+1} \quad (2)$$

$$PRESS = \sum_{i}^{i=n} (y_i - \hat{y}_i) \quad (3)$$

$$TSS = \sum_{i}^{i=n} (y_i - \bar{y}) \quad (4)$$

$$S_{PRESS} = \sqrt{\frac{\int_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p-1}} = \sqrt{\frac{PRESS}{n-p-1}} \quad (5)$$

Where:

y_i is the observed activity of the training set compounds

$\hat{y}_i$is the predicted activity of the training set compounds.

$\bar{y}$ is mean observed activity of the training set compounds and n number of objects.

p number of predictor variables.

3.2. External Validation

Several authors have suggested that the only way to estimate the true predictive power of a QSAR model is to compare the predicted and observed activities of an external test set of compounds that were not used in the model development [27-31].

To estimate the predictive power of a QSAR model, Golbraikh and Tropsha recommended the use of the following statistical parameters using the test set [32, 33]:

That reflects the degree of correlation between the observed and predicted activity data of the test set.

$$R^2_{pred} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \quad (6)$$

Here,$y_i$and $\bar{y}_i$are the observed and predicted activity data for the test set compoundswhile $\bar{y}$ indicates the mean observed activity of the training set molecules. Thus, model with values of R^2_{pred} above the stipulated value of 0.5 are considered predictive.

$$R^{\circ 2} = 1 - \frac{\sum_{i=1}^{ntest}(\hat{y}_i - y_i^{r^\circ})^{k2}}{\sum_{i=1}^{ntest}(\hat{y}_i - \bar{\hat{y}})^2} \cdot y_i^{r^\circ} = K\hat{y}_i \quad (7)$$

$$R^{\circ 2} = 1 - \frac{\sum_{i=1}^{ntest}(y_i - \hat{y}_i^{r^\circ})^2}{\sum_{i=1}^{ntest}(y_i - \bar{y})^2} \cdot \hat{y}_i^{r^\circ} = K' y_i \quad (8)$$

Where,$R^{\circ 2}$or $R^{\circ 2}$are the squared correlation coefficient obtained using predicted versus observed activities and observed versus predicted activities respectively.

$$K = \frac{\sum_{i=1}^{ntest} y_i \hat{y}_i}{\sum_{i=1}^{ntest} y^2_i} \quad (9)$$

$$K' = \frac{\sum_{i=1}^{ntest} y_i \hat{y}_i}{\sum_{i=1}^{ntest} y_i{}^2} \quad (10)$$

K and K' are the slopes of regression lines through the origin for fits to experimental and predicted data respectively.

3.2.1. The $r^2_{m(test)}$matrice for external validation

The external predictability of the selected model was also checked by r_m as proposed by Roy Paul

(2008) [34] and the different r_m^2 values were calculated using Equations.

$$r_m^2 = R^2(1 - \sqrt{|R^2 - R^{*2}|}) \quad (11)$$

$$r_m'^2 = R^2(1 - \sqrt{|R^2 - R'^{*2}|}) \quad (12)$$

$$\overline{r_{m(test)}^2} = \frac{r_m^2 + r_m'^2}{2} (13)$$

Where R^2 is squared correlation coefficient between observed and predicted values and R^{*2} is squared correlation coefficient between observed and predicted values with intercept value set to zero.

A value of r_m^2 is greater than 0.5 may be taken as an indicator of good external predictability [35]. For the prediction the value of $\Delta r_{m(test)}^2$ should preferably be lower than 0.2 provided that the value $\overline{r_{m(test)}^2}$ is more than 0.5 [36].

3.2.2. Concordance correlation coefficient

The external predictability of the selected model was also checked by concordance correlation coefficient (CCC), as proposed by Gramaticaand al. [37] and calculated using Equation (14):

$$ccc = \frac{2\sum_{i=1}^{ntest}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_i^{ntest}(y_i - \bar{y})^2 + \sum_i^{ntest}(\hat{y}_i - \bar{\hat{y}})^2 + n_{test}(\bar{y} - \bar{\hat{y}})^2} (14)$$

4. Results and Discussion

In the present study, we tried to develop the best QSAR model to explain the correlations between the physicochemical parameters and the biological activities IC_{50} values of 1.2.3-triazole derivatives with aromatase inhibitory effects.

The full linear equation for the prediction of the inhibitory IC_{50} activity is the following equation (15):

$$pIC_{50} = 2.726 + 0.005\,S - 0.366\,logP - 0.026\,ref + 6.894\,LE - 6.604\,qN1 + 0.01\,TPSA + 0.057\,EH + 9.431\,E_{LUMO} - 5.333\,qN2 + 0.722\,qC4 - 0.026\,ABS + 0.006\,M - 0.02\,POL - 2.661\,qC5 (15)$$

n = 24; R = 0.991; R^2 = 0.982; S = 0.149;
F = 34.569; Q = 6.651

The significant equation consists of 14 descriptors: Polarizability (Pol),Molar refractivity (ref), Partition coefficient octanol/water (log P), Hydration energy (HE),Surface area grid (S), Molar weight (M), Topological polar surface area (TPSA),Percentage of absorption (%ABS),Ligand efficiency (LE), Energy LUMO (E_{LUMO}) and Atomic net charges (qN1, qN2,qC4,qC5).

The F-value has found to be statistically significant at 95 % level since the calculated F value is higher as compared to tabulated value.

The positive value of quality factor (Q) for this QSAR's model suggests its high predictive power and lack of over fitting.

The positive coefficient of hydration energy and negative Log P indicates that the hydrophilic derivatives give a good biological activity.

From the equation, we can see any increase in the molecular surface causes an increase of the biological activity, which results in increased surface of contact between the ligand and the receptor the same for the molecular weight. The positive coefficients of MW explain that any decrease molecular weight of the compounds causes a decrease in the biological activity.

It can be observed that high coefficients of Ligand efficiency LE**,** Thus, high LE lead to increasing aromatase inhibitory activity high LE prefers compounds that gain to escape the affinity-biased selection and optimization towards larger ligands. The focus should be directed towards the generation of compounds that use their atoms most efficiently.

In the model, positive coefficient of TPSA indicates that the substituants that increase molecular polar surface area will lead to increased activity. This relates to the molecular transport through membranes Suggested that a decrease in the permeability might decrease the activity.

It can be observed that high coefficients of atomic charges on atoms N2, C4 and C5 (qN2, qC4 and qC5 respectively). Thus, high negative charges lead to increasing aromatase inhibitory activity.

The charges allowed a physical explanation and electronic molecular properties contributing to aromatase inhibitory potency as the electronic character related directly to the electron distribution of interacting molecule at the site active.

Once the equation is obtained, it is important to determine its reliability and significance. The validation of the equation is done by cross-validation "leave-one out" method. The results are shown below.

Table 4. Cross-validation parameters

PRESS	SSY	PRESS /SSY	SPRESS	R^2cv	R^2adj
0.202	11.072	0.018	0.149	0.982	0.953

Also, for reasonable QSAR model the PRESS/SSY ratio should be lower than 0.4 [38].

The data presented in (Table.3) indicate that for the developed model this ratio is 0.018.

Our result of $R^2_{cv}$and R^2_{adj} for this QSAR model has been to be 0.982 and 0.953 respectively. The high value of R^2_{cv} and R^2_{adj} are essential criteria for the best qualification of the QSAR model.

One can also use the S_{press} parameter that reflects the error changes predictions. Developed QSAR models have low values of S_{press} (<0.200) indicating that the model has small residual value between observed and predicted biological activities. We can see from the Table 5, that all residual values less than twice of standard error of estimate (0.149) therefore, are not any outliers.

In order to confirm our results, we have estimated the aromatase inhibitory activity pIC_{50} of training sets using the model expressed by equation (15) and compared them with the observed values. The data presented in (Table 4) show that the observed and predicted activities are very close each other.

The plots for this model show to be more convenient with R^2= 0.991. It indicates that the model can be successfully applied to predict the aromatase inhibitory activity of these compounds.

To investigate the presence of a systematic error in developing the QSAR models. The residuals of predicted values of the biological activity pIC_{50} were plotted against the experimental values as shown in (Fig.3.).

The propagation of the residuals on both sides of zero indicates that no systemic error exists. As suggested by Jalali-Heravi and Kyani [39]. It indicates that this model can be successfully applied to predict the aromatase inhibitory activity of this class of molecules.

The simplest method of investigating occurrence of multicolinearity is to obtain the correlation matrix which indicates that most of the descriptors used are not highly correlated (Table 6). In this study, no descriptor strongly correlated with the others. $(\rho < 0,9)$

The proposed model (Eq.15) passed all the tests for the predictive ability (Eqs.7 – 14).

The results obtained show that the predicted values (Table 8) are very close to the observed values (Fig.2). The value of R^2 is equal to 0.851 which confirms that model adequately describes the relationship between pIC_{50} predicted and observed model. Further, the above QSAR model is confirmed its external predictability by predicting.

The ortep diagram and the optimized geometries at the optiumum conformation of the title compound is shown in Fig. 1. The asymmetric unit of the title compound, $C_{16}H_{14}OS_2$, has one-half-molecule and it is completed with a twofold symmetry axis [symmetry code:x, y, -z]. The molecular structure of the compound, $C_{16}H_{14}OS_2$, has an E-configuration so that the substituents at the vinyl group of the compound [(C5=C6, C10=C12)] indicate a trans conformation, and the two thiophene rings adopt a syn orientation and are located on both side of the cyclohexanone.

Table 5. Experimental and predicted aromatase inhibitory activities (pIC_{50}) of aromatase inhibitory activity (1-24) obtained from the model

Compound	pIC_{50} exp.	pIC_{50} pred.	Resid.	Compound	pIC_{50} exp.	pIC_{50} pred.	Resid.
1	5.470	5.599	-0.129	13	5.330	5.276	0.054
2	5.040	5.129	-0.089	14	4.920	4.903	0.017
3	5.380	5.227	0.153	15	4.870	4.782	0.088
4	5.630	5.652	-0.022	16	4.820	4.817	0.003
5	4.940	5.002	-0.062	17	5.340	5.298	0.042
6	5.330	5.309	0.021	18	4.960	4.993	-0.033
7	5.100	4.928	0.172	19	4.830	5.025	-0.195
8	5.650	5.655	-0.005	20	4.890	4.871	0.019
9	8.100	8.042	0.058	21	4.860	4.770	0.090
10	5.290	5.342	-0.052	22	4.510	4.416	0.094
11	4.960	5.039	-0.079	23	4.780	4.958	-0.177
12	4.790	4.713	0.077	24	4.570	4.605	-0.034

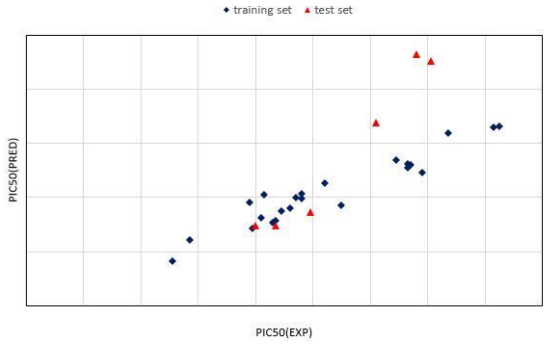


Fig. 2. Scatter Plot between the Observed and Predicted Activities of Model of training set and the test set.

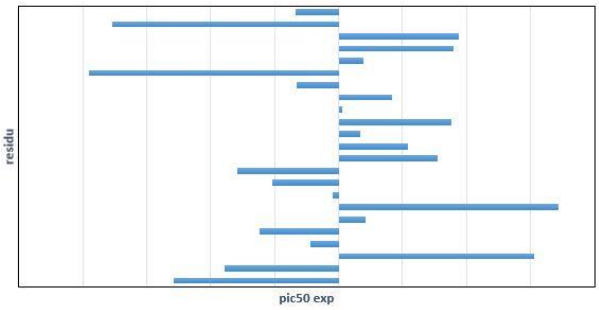


Fig. 3. Plots of the residual values against the experimentally observed

Table 6. Correlation of the fourteen selected descriptors

	pIC50	*S*	*M*	*logP*	*EH*	*Pol*	*Ref*	*LE*	E_{LUMO}	*qN1*	*qN2*	*qC4*	*qC5*	*TPSA*	*ABS*
pIC50	*1*														
S	*0.26*	*1*													
M	*0.26*	*0.90*	*1*												
logp	*0374*	*0.676*	*0.41*	*1*											
EH	*0.258*	*0.180*	*0.45*	*0.529*	*1*										
Pol	*0.224*	*0.800*	*0.80*	*0.570*	*0.07*	*1*									
Ref	*0.064*	*0.404*	*0.36*	*0.387*	*0.180*	*-0.83*	*1*								
LE	*0.776*	*0.413*	*0.45*	*0.314*	*0.063*	*-0.46*	*0.34*	*1*							
E_{LUMO}	*0.288*	*0.095*	*0.12*	*0.168*	*0.474*	*0.151*	*0.302*	*0.003*	*1*						
qN1	*0.555*	*0.267*	*0.09*	*0.230*	*0.22*	*0.202*	*0.21*	*-0.40*	*0.348*	*1*					
qN2	*0.094*	*0.356*	*0.20*	*0.125*	*0.14*	*-0.14*	*0.04*	*0.128*	*0.166*	*0.500*	*1*				
qC4	*0.1*	*0.27*	*0.2*	*0.297*	*0.07*	*0.199*	*0.07*	*-0.01*	*0.010*	*0.22*	*0.064*	*1*			
qC5	*0.071*	*0.044*	*0.16*	*0.472*	*0.55*	*-0.08*	*0.18*	*-0.06*	*0.335*	*0.397*	*0.267*	*-0.1*	*1*		
TPSA	*0.218*	*0.148*	*0.21*	*0.228*	*0.51*	*0.071*	*0.02*	*-0.19*	*0.700*	*0.06*	*0.143*	*-0.2*	*0.205*	*1*	
ABS	*0.20*	*0.2*	*0.3*	*0.072*	*0.48*	*-0.06*	*0.10*	*0.193*	*0.89*	*0.12*	*0.249*	*0.08*	*-0.27*	*0.812*	*1*

From the Table 7 it is obvious that the predicted responses of all the test compounds are in good agreement with their corresponding observed responses as well as ideal fit is attained produced by plotting a graph (Fig. 3) by correlating observed activity versus predicted activity of the test set compounds, the squared correlation coefficient is calculated as 0.923.

The predictive abilities of the best MLR was tested (Table 8) using the Golbraikh–Tropsha criteria and the R^2_{pred} test (see Model Validity section). All the calculated parameters indicated the model showed a good predictive power.

Analyzing the results of the external test set listed in Table 7, it could be observed that all the Golbraikh–Tropsha criteria were fulfilled.

r_m^2 value of 0.556 whereas values of average r_m^2 of 0.581 and $\Delta r^2_{m(test)}$ of 0.042 extend more efficient evidence of external predictability of the generated QSAR.(Table9)

Once the QSAR model formulated and validated properly. Its utility is to predict the biological responses of the compounds which are generated by combinatorial deign and experimentally non-investigated.

Table 7. Observed and predicted activity of test compounds

Compounds	pIC_{50} exp.	pIC_{50} pred.
25t	4.99	4.87
26t	5.64	5.87
27t	4.80	4.74
28t	4.87	4.96
29t	5.87	6.10
30t	5.00	5.11

Table 8. Predictive power results for the external test set; Golbraikh and Tropsha criteria

Golbraikh and Tropsha's criteria							
R^2_{pred}	K	K'	$R^{'2}$	$R^{''2}$	$\frac{R^2 - R^{'2}}{R^2}$	$\frac{R^2 - R^{''2}}{R^2}$	$\lvert R_0^2 - R_c^{'2} \rvert$
0.851	0.982	1.017	0.972	0.952	-1.04	-0.11	0.02
>0.6	>0.85	<1.15	close to R^2	close to R^2	<0.1	<0.1	< 0.3

Table 9. Validation characteristics of developed model according to r2m metrics and Concordance correlation coefficient

r_m^2 parameter				Concordance correlation coefficient
r_m^2	$r_m^{'2}$	$\Delta r^2_{m(test)}$	$\overline{r^2_{m\,(test)}}$	CCC
0.556	0.581	0.042	0.568	0.946
>0.5		<0.2	>0.5	>0.85

5. Conclusion

In this study, SW-MLR was used to develop linear QSAR model for prediction of aromatase inhibitory activity of Triazole derivatives. The built model displayed good correlations between the structure and activity of the studied compounds. The model was validated using LOO cross-validation and external test set. The built model has a good self-and external-predictive power. Based on QSAR model results coefficients of Ligand efficiency LE atomic net charges (qN1) and partition coefficient octanol/water (log P), were found to be important factors controlling aromatase inhibitor activity.

References

[1] M. P. Leze, M. Le Borgne, P.Pinson, A. Palusczak, M. Duflos, G. LeBaut,and R. W. Hartmann,2- and 3-[(aryl)(azolyl)methyl]indoles as potential non-steroidal aromatase inhibitors, Bioorg. Med. Chem. Lett. 1134(2004)549-557.

[2] L. A. Torre, F. Islami, R. L. Siegel, E. M. Ward, et A. Jemal,Global Cancer in Women: Burden and Trends, Cancer Epidemiology and Prevention Biomarkers 26 (2017) 444-57.

[3] P. P. Koonings, K.Campbell, D. R. J. Mishell,and D. A.Grimes,Global Cancer in Women: Burden and Trends , Obstet. Gynecol.74,(1989) 921-926.

[4] M. J. Reed, The role of aromatase in breast tumors, Breast Cancer Res. Treat. 30, (1994) 7-17.

[5] E. R. Simpson, M. S. Mahendroo, G. D.Means, M. W.Kilgore, M. M.Hinshelwood, S. Graham-Lorence, B.Amarneh,Y.Ito, C. R. Fisher, andM. D. Michael, Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis, Endocr. Rev.15 (1994) 342-355.

[6] M. A. C. Neves,T. C. P. Dinis, G.Colombo, M. L. S Melo,An efficient steroid pharmacophore-based strategy to identify new aromatase inhibitors, Eur. J. Med. Chem. 44,(2009) 4121-4127.

[7] A.Howell, J. F. R.Robertson, I. Vergote, A review of the efficacy of anastrozole in postmenopausal women with advanced breast cancer with visceral metastases., Breast Cancer Res. Treat. 82(2003) 215-222.

[8] D. Simpson, M. P. Curran, andC. M. Perry, Letrozole: a review of its use in postmenopausal women with breast cancer. Drugs. 64(2004) 1213-1230.

[9] T. Fornander, A.C. Hellstrom, B.Moberger, Descriptive clinicopathologic study of 17 patients with endometrial cancer during or after adjuvant tamoxifen in early breast cancer. J. Natl.Cancer Inst. 85,(1993)1850-1855.

[10] Jeong, H.-J., Shin, Y. G., Kim, I.-H., & Pezzuto, J. M. Inhibition of aromatase activity by flavonoids. Archives of Pharmacal Research, 22(1999), 309. 312

[11] S. G. Agalave, S. R.Maujan,and V. S. Pore, Click Chemistry: 1,2,3-Triazoles as Pharmacophores.Chem. Asian J. 6(2011) 2696-2718.

[12] A. D.Favia,O. Nicolotti, A.Stefanachi,F.Leonetti,and A. Carotti, Computational methods for the design of potent aromatase inhibitors, Expert Opin. Drug Discov. 8 (2013) 395-409.

[13] C. Hansch and A. Leo, Exploring QSAR, Fundamentals and Applications in Chemistry and Biology.American Chemical Society. Washington. D. C.(1995).

[14] A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation. Molecular Informatics 29, (2010). 476–488.

[15] S. Belaidi, A. Dibi and M. Omari, Contribution à l'étude du contrôle stéréochimique dans les macrolides à 16 chaînons par la mécanique moléculaire, J. Soc. Alg. Chim., 10(2000), 221-232,

[16] S. Belaidi, H. Belaidi and D. Bouzidi, Computational Methods Applied in Physical-Chemistry property Relationships of Thiophene Derivatives, J. Comput. Theor. Nanosci., 12 (2015)1737-1745

[17] C. Nantasenamat, A. Worachartcheewan, S.Prachayasittikul, C. Isarankura-Na-Ayudhya, and V. Prachayasittikul, QSAR modeling of aromatase inhibitory activity of 1-substituted 1,2,3-triazole analogs of letrozole. Eur. J. Med. Chem.69, (2013) 99-114

[18] HyperChem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA (2007).

[19] Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian Inc., Wallingford, CT (2010).

[20] Ghose, A. K., & Crippen, G. M. Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. Journal of Chemical Information and Computer Sciences, 27(1987) 21-35.

[21] Database,(http://www.molinspiration.com).

[22] M. Y. Zhao, M. H. Abraham, J. Le, A. Hersey, C. N. Luscombe, G.Beck,B.Sherborne, Rate-limited steps of human oral absorption and QSAR studies. Pharm. Res. 19,(2002) 1446-1457.

[23] P. R. Andrews,D. J.Craik,J. L. Martin, Functional group contributions to drug-receptor interactions J. Med. Chem. 27, (1984) 1648-1657.

[24] SPSS software packages, SPSS Inc., 444 North Michigan Avenue, Suite 3000, Chicago, Illinoi, 60611, USA.

[25] Z. Almi, S. Belaidi, L. Segueni, Structural Exploration and Quantitative Structure-Activity Relationships Properties for 1.2. 5-Oxadiazole Derivatives, Rev. Theor. Sci. 3, (2015) 264-272.

[26] A.E Ivanescu, LiP., B. George, A.W. Brown, S.W.Keith, D. Raju, and D.B. Allison, The Importance of Prediction Model Validation and Assessment in Obesity and Nutrition Research. Int J Obes (Lond) 40 (2016) 887–894.

[27] K. Roy, A. S.Mandal, Development of linear and nonlinear predictive QSAR models and their external validation using molecular similarity principle for anti-HIV indolyl aryl sulfones. Journal of Enzyme Inhibition and Medicinal Chemistry. 23 (2008) 980–995

[28] R. Guha,P.C.Jurs,Determining the validity of a QSAR model A Classification Approach.J. Chem. Inf. Model. 45, (2005) 65-73

[29] E.Novellino,C.Fattorusso,G. Greco, Use of comparative molecular field and cluster

analysis in series design.Pharm. Acta Helv. 70,(1995) 149-154.

[30] N. S. Zefirov, V. A. Palyulin, QSAR for Boiling Points of "Small" Sulfides. Are the "High-Quality Structure-Property-Activity Regressions" the Real High Quality QSAR Models? J. Chem. Inf. Comput. Sci. 41 (2001) 1022–1027.

[31] N. Frimayanti, M. L. Yam, H. B. Lee, R. Othman, S. M.Zain., N. A. Rahman, Design of new competitive dengue Ns2b/Ns3 protease inhibitors-a computational approach. International Journal of Molecular Sciences 12 (2011) 8626–8644.

[32] A. Golbraikh, A. Tropsha, Predictive QSAR Modeling Based on Diversity Sampling of Experimental Datasets for the Training and Test Set Selection. Mol. Divers. 5 (2000) 231–243.

[33] L. Sachs ,Applied Statistics: A Handbook of Techniques, Springer-Verlag ,BerlirdNew York,(1984).

[34] P. P. Roy, S. Paul, I. Mitra, K. Roy, On Two Novel Parameters for Validation of Predictive QSAR Models. Molecules. 14 (2009) 1660–1701.

[35] R. Veerasamy, H.Rajak, A. Jain, S. Sivadasan,C.P. Varghese, and R. K. Agrawal,Validation of QSAR models-strategies and importance .Int J Drug &Discov. 2(2011) 511-519.

[36] D. L. J. Alexander. A. Tropsha, Winkler, D. A. Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. J Chem Inf Model, 55 (2015) 1316–1322.

[37] N.Chirico and P.Gramatica, Real External Predictivity of QSAR Models. Part 2. New Intercomparable Thresholds for Different Validation Criteria and the Need for Scatter Plot Inspection. J. Chem. Inf. Model. 52, (2012) 2044-2058.

[38] S. O. P. Kuzmanovic, D. D. Cvetkovic, D. J. Barna, QSAR Analysis of 2-Amino or 2-Methyl-1-Substituted Benzimidazoles Against Pseudomonas aeruginosa. Int. J. Mol. Sci. 10, (2009) 1670-1682.

[39] M. J. Heravi , A. Kyani, Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds: a PCA-MLR-ANN approach. J. Chem. Inf. Comput. Sci. 44, (2004) 1328-1335.