

# Quantitative Structure Activity Relationship (QSAR) Investigations and Molecular Docking Analysis of Plasmodium Protein Farnesyltransferase Inhibitors as Potent Antimalarial Agents

Mebarka Ouassaf<sup>1</sup>, Salah Belaidi<sup>1</sup>, Amneh Shtaiwi<sup>2</sup>, Samir Chtita<sup>\*3</sup>

<sup>1</sup> University of Biskra, Group of Computational and Medicinal Chemistry, LMCE Laboratory, Algeria.

<sup>2</sup> Faculty of Pharmacy, Middle East University, Jordan.

<sup>3</sup> Laboratory Physical Chemistry of Materials, Faculty of Sciences Ben M'sik, Hassan II University of Casablanca, Morocco.

## ABSTRACT

The development of *farnesyltransferase* inhibitors based on the benzophenone scaffold directed against *Plasmodium falciparum* is considered a strategy in malaria treatment. In this work, quantitative structure–activity relationship (QSAR) was performed to predict the protein *farnesyltransferase* (PFT) inhibitory activities for a series of 36 benzophenone derivatives. The data set was divided into two subsets of training and test sets, and the best model using **multiple linear regression (MLR)**, with the values of internal and external validity ( $R^2 = 0.884$ ,  $R^2_{adj} = 0.865$ ,  $R^2_{pred} = 0.821$ ,  $Q^2_{cv} = 0.822$  and  $R^2_p = 0.811$ ) was found in agreement with the Tropsha and Golbraikh criteria. The applicability domain (AD) was determined using the Williams plot to describe the chemical space for the model used in this study. The model shows that antimalarial activities of benzophenone depend on logP, bpol, MAXDn, and FMF descriptors. These indications prompted us to design new benzophenones PFT inhibitors and predict the value of their anti-malarial activities based on the MLR equation. Docking results reveal that the newly designed benzophenones bind to the hydrophobic pocket and polar contact with high affinity. The predicted results from this study can help to design novel benzophenone as inhibitors of human PFT with high antimalarial activities.

**Keywords:** QSAR, docking, benzophenone, PFT inhibitory, antimalarial.

## I. INTRODUCTION

Malaria is one of the most important infectious diseases in the world [1]; it affects 400–900 million people each year in the world and is also the cause of death of about one to three million people annually [2]. Malaria can be caused by several species of **Plasmodium** (P) parasites, each of which has a complex life cycle: *Plasmodium vivax*, *P. ovale*, *P. malariae*, and *P. falciparum*. [3].

*Plasmodium falciparum* causes a more severe type of

malaria, with a greater risk of mortality for individuals who get it [3]. It causes severe and/or lethal complications including cerebral malaria defined as coma, altered mental status, or multiple seizures. Recent efforts at the control of *falciparum* malaria have generally been unsuccessful, in large part due to the continuous development of *P. falciparum*'s resistance to conventional antimalarial [4]. Thus, there is a great need for the identification of new agents against multi-drug resistant *Plasmodium* and the evaluation of potential new compounds that act as an inhibitor for the treatment of malaria. Inhibitors of the protein *farnesyltransferase* (PFT) enzyme have emerged as a promising target for treating malaria caused by the *Plasmodium falciparum* parasite. *Farnesyltransferase* is a

---

\*Corresponding author: S. CHTITA

[samirchtita@gmail.com](mailto:samirchtita@gmail.com)

Received: 17/9/2021 Accepted: 12/1/2022.

DOI: <https://doi.org/10.35516/jjps.v15i3.407>

catalytic enzyme that catalyzes the transfer of a farnesyl residue from farnesyl diphosphate to the thiol of a cysteine side chain of proteins carrying the CAAX-tetra peptide sequence (C: cysteine, A: aliphatic amino acid, X: serine or methionine) at their C terminus [5, 6]. PFT inhibitors are promising drugs for the treatment of malaria, and a number of different scaffolds have been shown to inhibit the growth of the malaria parasite *in vitro* and *in vivo* [7]. Previously, a class of *farnesyltransferase* inhibitors based on a benzophenone scaffold has developed by Wiesner and his co-workers, to find new synthetic inhibitors with simple structure and low-cost properties [8]. It was observed that compounds of this type suppress the growth of the multiresistant *Plasmodium falciparum* strain Dd2 in the nanomolar range [8].

In order to find new leads in the process of drug design and discovery, it would be helpful to examine chemical databases and virtual libraries against molecules with known activities or properties by using computational procedures. For this purpose, analysis of the relationship between structure and quantitative activity is widely used [9]. The Two-Dimensional Quantitative Structure-Activity Relationship (2D-QSAR) is a useful tool for describing the relationships between chemical structure and experimental data [10]. The relationship can be used to achieve the modeler's intended purpose, such as increasing the efficiency of an operation, lowering the toxicity of dangerous substances, or improving the pharmacological activity of pharmaceuticals, every phase of the predictive QSAR modeling analysis involves the application of multiple mathematical principles, and a huge quantity of quantitative data is created. The main purpose is to convert chemical information into useful numbers (descriptors), then build a mathematical link with the response [11]. The descriptors for 2D QSAR can be categorized according to their nature as well as calculation method, such as constitutional, topological, geometrical, electrostatic, quantum-chemical, and thermodynamic descriptors [12].

The present work describes the descriptor-based

QSAR studies developed for a series of benzophenone derivatives that were previously reported in the literature and evaluated for their antimalarial activity. In addition, to predict the activity of newly designed compounds based on their molecular properties, we used the 2D-QSAR model by altering molecular descriptors and chemical fragments, which were determined to be important within the model's applicability scope. Furthermore, we explored the binding interactions of the newly designed benzophenone derivatives with protein *farnesyltransferase* receptor, to provide new information that might be useful in the development of new inhibitors with improved antimalarial properties.

## II. DATA SET AND COMPUTATIONAL METHODS

### 1. Data set for analyze

The 2D-QSAR studies performed on 36 benzophenone derivatives were obtained with their anti-antimalarial activities against Dd2 strain of *Plasmodium falciparum* from the work of Wiesner and co-workers [8, 13-16]. After conversion of the IC<sub>50</sub> values to micromole, we generated the pIC<sub>50</sub> values for each of the 36 compounds using the following:  $pIC_{50} = -\log(IC_{50})$  (1)

The 2D structures of the molecules were prepared using Marvin Sketch (<https://www.chemaxon.com>) [17], and converted to 3D and optimized the molecular geometry using HyperChem software [18], as shown in Table 1. The geometries of benzophenone derivatives were first fully optimized by molecular mechanics, with MM<sup>+</sup> force-field (RMS = 0.001 Kcal/Å). Then, geometries were fully re-optimized by using PM3 method.

### 2. Calculation of molecular descriptors

The PaDEL descriptor [19] was used to calculate a pool of descriptors of the optimized molecules of benzophenone derivatives. The constitutional, autocorrelation, Basak, BCUT, Burden, connectivity, E-state, Kappa, extended topochemical atom (ETA), molecular property, and topological descriptors were computed. Among these descriptors, 19 descriptors were

selected and the other descriptors were eliminated after pretreatment on the basis of the correlation coefficient cut-off value of 0.80 (Tables S1 and S2) and the variance cut-off value of less than 0.0001, as presented in Table 2.

### 3. Data Splits and Model Development

The data set was divided into two subsets, the training data set and the test data set, the division was random by the software. Training set (70%) was used in building the model while test set (30%) in validating the model.

In this stage, the multiple linear regression (MLR) analysis of the training set was carried out using XLSTAT software between a response variable Y (pIC<sub>50</sub>) and independent variables X (2D) molecular descriptors to find a linear model of the activity of interest, which takes the form of the multiple linear regression equation [20].

The best model QSAR was chosen based on the statistical validation parameters like the correlation coefficient of determination (**R**<sup>2</sup>), Adjusted correlation coefficient **R**<sup>2</sup> (**R**<sup>2</sup><sub>adj</sub>), predicted residual sum of squares (**PRESS**), total sum of squares (**SSY**) and standard deviation based on predicted residual sum of squares (**S<sub>PRESS</sub>**) [21] all

are represented in Eqs. (2), (3), (4) and (5):

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

$$PRESS = \sum_i^{i=n} (y_i - \hat{y}_i)^2 \quad (3)$$

$$SSY = \sum_i^{i=n} (y_i - \bar{y})^2 \quad (4)$$

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

Where:

$y_i$  is the observed activity of the training set compounds

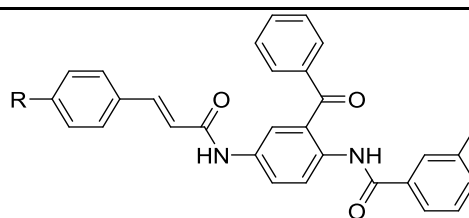
$\hat{y}_i$  is the predicted activity of the training set compounds.

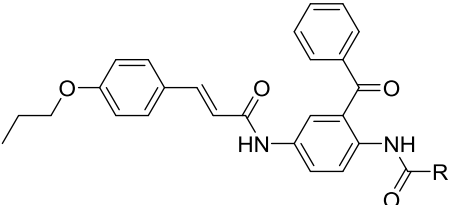
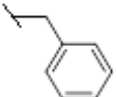
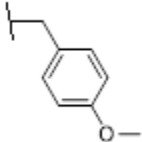
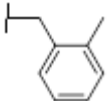
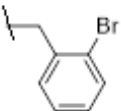
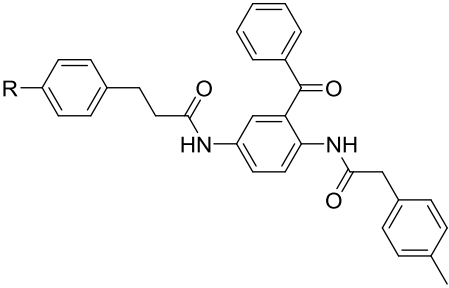
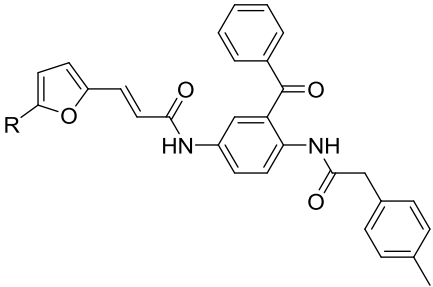
$\bar{y}$  is mean observed activity of the training set compounds and n number of objects.

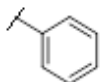
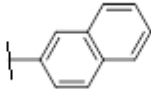
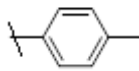
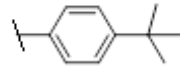
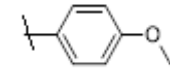
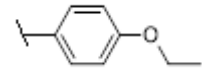
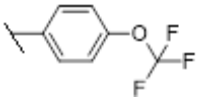
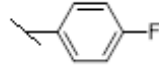
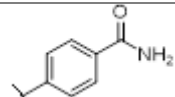
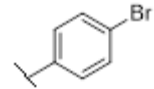
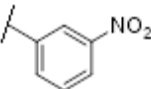
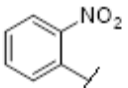
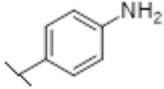
p number of predictor variables.

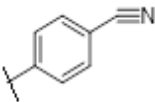
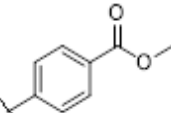
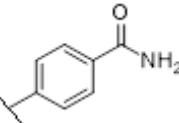
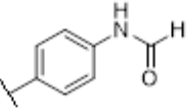
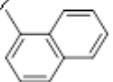
**Table 1: Chemical structures and PFT inhibitors activities of benzophenone derivatives**

Comp. No	R	pIC <sub>50</sub> exp μM	pIC <sub>50</sub> pred μM	Residue
1	-H	5.24	5.234	0.006
2	-Cl	5.26	5.396	-0.136
3t	-NO <sub>2</sub>	5.19	5.318	-0.128
4	-Br	5.49	5.598	-0.108
5	-NH <sub>2</sub>	5.26	5.314	-0.054
6t	-CH <sub>3</sub>	5.85	5.568	0.282
7	-CF <sub>3</sub>	5.24	5.560	-0.320
8	-O-CH <sub>3</sub>	5.89	5.730	0.160
9	-CH <sub>2</sub> -CH <sub>3</sub>	5.92	5.641	0.279
10	-CH (CH <sub>3</sub> ) <sub>2</sub>	5.92	5.590	0.330
11	-C (CH <sub>3</sub> ) <sub>3</sub>	5.52	5.641	-0.121



<b>12</b>	-O-CH <sub>2</sub> -CH <sub>3</sub>	6.07	5.780	0.290
<b>13</b>	-O-(CH <sub>2</sub> ) <sub>2</sub> -(CH <sub>3</sub> ) <sub>3</sub>	5.96	5.847	0.113
				
<b>14</b>		5.60	5.552	0.048
<b>15</b>		5.89	5.848	0.042
<b>16t</b>		5.62	5.719	-0.099
<b>17</b>		5.55	5.671	-0.121
				
<b>18</b>	H	5.57	5.658	-0.088
				

19t		6.38	6.148	0.232
20t		6.70	6.664	0.036
21t		6.92	6.534	0.386
22t		6.52	6.833	-0.313
23		6.89	6.754	0.136
24		6.52	6.664	-0.144
25		6.68	6.736	-0.056
26		6.49	6.505	-0.015
27		6.25	6.589	-0.339
28		6.90	6.492	0.408
29		6.23	6.248	-0.018
30		6.17	6.248	-0.078
31		6.59	6.568	0.022

32		6.66	6.619	0.041
33		6.77	6.642	0.128
34t		6.25	6.289	-0.039
35		6.55	6.502	0.048
36		6.00	5.672	0.328

t: Test set compounds.

Table 2: Values of parameters calculated for the studied compounds

Comp.	ALogP	ALogp2	AMR	apol	bpol	Lipa	MAX DP	DEL S2	FMF	MAX DN	VAB C	VAdj Mat	MW	AM W	WPA TH	WPOL	XLogP	Zagreb	TPSA
1	-0.49	0.245	40.5	76.5	32.6	12.3	6.315	36.41	0.51	1.992	457.8	6.169	474	7.64	4550	52	11.43	182	75.3
2	0.35	0.123	46.2	76.4	29.0	11.7	6.350	43.49	0.54	1.977	471.0	6.209	528	8.95	4956	54	9.19	188	75.3
3	0.16	0.027	44.4	78.5	32.7	10.6	6.342	51.61	0.50	2.496	483.8	6.285	519	8.11	5844	58	10.26	198	118.4
4	0.39	0.154	49.1	78.8	32.8	12.2	6.349	37.88	0.51	1.986	477.1	6.209	552	8.90	4956	54	10.81	188	75.3
5	-1.24	1.551	44.7	78.2	33.0	11.4	6.328	39.92	0.50	2.001	468.8	6.200	489	7.64	4956	54	9.35	188	101.2
6	0.15	0.021	46.0	79.5	34.8	12.7	6.342	36.68	0.49	1.990	475.1	6.209	488	7.51	4956	54	11.75	188	75.3
7	0.60	0.363	46.9	79.2	35.1	13.9	6.333	64.67	0.49	5.693	493.3	6.321	542	8.34	6291	60	11.91	206	75.3
8	-0.54	0.299	47.8	80.3	36.7	12.2	6.357	40.04	0.48	2.018	483.9	6.247	504	7.63	5399	56	9.93	192	84.5
9	-0.18	0.035	50.8	82.6	36.9	13.2	6.367	37.22	0.47	1.987	492.4	6.247	502	7.38	5399	56	12.32	192	75.3
10	0.08	0.005	54.7	85.7	39.1	13.9	6.390	37.86	0.45	1.988	509.7	6.285	516	7.27	5844	58	12.80	198	75.3
11	1.06	1.125	59.6	88.8	1.06	14.7	6.413	38.49	0.43	1.991	527.0	6.321	530	7.10	6291	60	13.43	206	75.3
12	-0.29	0.088	52.1	83.4	38.9	12.6	6.378	40.82	0.46	2.015	501.2	6.285	518	7.51	5880	57	10.35	196	84.5
13	-1.16	1.351	58.5	89.6	43.2	13.8	6.415	41.85	0.42	2.011	535.8	6.357	546	7.28	6960	59	11.28	204	84.5
14	-1.24	1.549	55.9	86.5	41.0	13.3	6.462	42.30	0.43	2.015	518.5	6.321	532	7.39	6210	61	10.86	200	84.5
15	-1.29	1.680	63.2	90.4	45.1	13.3	6.542	46.59	0.40	2.028	544.6	6.392	562	7.39	7039	65	9.79	210	93.7
16	-0.60	0.363	61.4	89.6	43.2	13.8	6.516	42.87	0.41	2.015	535.8	6.357	546	7.28	6600	64	11.19	206	84.5
17	-0.35	0.126	64.5	88.9	41.2	13.2	6.525	44.26	0.43	2.008	537.8	6.357	610	8.47	6600	64	10.45	206	84.5
18	-0.75	0.576	39.6	77.8	34.8	11.8	6.329	36.83	0.50	1.901	460.5	6.169	476	7.44	4550	52	11.17	182	75.2
19	-0.06	5.775	47.6	88.8	38.9	14.0	6.440	42.24	0.51	2.057	524.6	6.392	554	7.69	7278	62	12.08	218	88.4
20	-0.07	5.775	47.6	97.1	41.1	16.1	6.518	44.10	0.52	2.061	565.2	6.523	604	7.74	9358	72	13.95	244	88.4
21	0.63	0.402	53.1	91.8	41.0	14.5	6.466	43.03	0.49	2.056	541.9	6.426	568	7.57	7705	65	12.40	224	88.4
22	0.19	0.036	59.2	95.7	45.1	14.5	6.495	47.06	0.46	2.072	568.0	6.491	598	7.57	8829	68	11.42	232	97.6
23	-0.05	0.003	54.9	92.6	43.0	14.0	6.478	46.27	0.48	2.074	550.7	6.459	584	7.68	8245	67	11.00	228	97.6
24	1.55	2.398	66.7	101.0	47.6	16.6	6.521	45.17	0.44	2.057	593.8	6.523	610	7.26	9331	71	14.07	242	88.4
25	0.08	0.006	54.6	92.5	43.1	14.1	6.477	54.53	0.48	2.433	556.7	6.491	602	7.92	8829	68	11.27	232	97.6
26	0.47	0.222	48.5	88.6	39.0	14.2	6.443	53.34	0.51	2.116	530.6	6.426	572	7.94	7705	65	11.40	224	88.4
27	-0.63	0.400	60.5	96.1	43.8	13.7	6.502	52.68	0.46	2.109	576.3	6.523	611	7.73	9310	69	10.90	238	117.0
28	0.88	0.776	56.2	91.1	39.1	14.0	6.471	44.14	0.51	2.044	543.9	6.426	632	8.77	7705	65	12.09	224	88.4
29	-0.02	5.904	64.1	93.4	43.4	11.6	6.477	60.91	0.47	2.402	566.7	6.491	603	7.73	8749	69	10.87	234	12.0
30	-0.02	5.904	64.1	93.4	43.4	11.8	6.474	60.07	0.47	2.402	566.7	6.491	603	7.73	8851	69	10.45	234	127.0
31	-1.43	2.057	64.4	93.2	43.7	12.8	6.471	47.54	0.47	2.017	551.7	6.426	573	7.34	7705	65	9.96	224	110
32	-0.64	0.422	66.4	93.6	42.8	12.8	6.481	50.08	0.48	2.042	563.8	6.459	583	7.57	8210	67	10.54	228	108
33	0.00	1.296	59.6	95.2	43.9	13.4	6.499	55.89	0.47	2.148	574.1	6.523	612	7.84	9269	71	11.42	238	114
34	-0.64	0.412	55.0	93.1	40.2	12.8	6.480	55.15	0.48	2.236	559.0	6.491	597	7.85	8717	69	10.37	234	131
35	-0.71	0.505	56.4	93.1	41.6	13.2	6.484	50.88	0.48	2.110	559.0	6.491	597	7.85	8759	68	10.66	232	117
36	-0.48	0.234	52.5	82.9	37.6	11.7	6.379	48.37	0.47	2.107	507.4	6.321	532	7.82	6328	60	10.35	202	101

#### 4. Model Validation

A QSAR study's major aim is to create a model with the best predictive and generalization abilities [22]. This is done to test the internal stability and predictive ability of the **QSAR** models, in this paper two principal type of validation (internal validation and external validation) were performed [23].

Common method for internally validating **QSAR** models is cross-validation (**CV**,  $Q^2$ ,  $q^2$ , or jack-knifing). The CV method repeats the regression on subsets of data multiple times. Usually each molecule is left out once (only), in turn, and the  $Q^2$  is computed using the predicted values of the missing molecule. CV is frequently used to calculate the maximum size of a model that may be utilized for a particular data collection. A cross-validated  $Q^2$  is usually smaller than the overall  $R^2$  for a **QSAR** equation. It's a diagnostic tool for determining an equation's predictive capability. The cross-validation regression coefficient ( $Q^2_{cv}$ ) was calculated with the following equation [18]:

$$Q^2_{cv} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{PRESS}{SSY} \quad (6)$$

It has been reported that high estimation of statistical attributes is not enough to justify the ability of a model, and so to assess the predictive capacity of the new **QSAR** model, the method was used by Golbraikh and his co-workers [24], Roy and his co-workers [25].

The coefficient of determination for the test set  $R^2$  test and other statistical characteristics of the test set are represented in Eqs. (7-11).

$$R^2_{pred} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \quad (7)$$

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

$$R'^{\circ 2} = 1 - \frac{\sum_{i=1}^{ntest} (y_i - \hat{y}_i^{r\circ})^2}{\sum_{i=1}^{ntest} (y_i - \bar{y})^2} \cdot \hat{y}_i^{r\circ} = K' y_i \quad (9)$$

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

$$K' = \frac{\sum_{i=1}^{ntest} y_i \hat{y}_i}{\sum_{i=1}^{ntest} y_i^2} \quad (11)$$

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

In addition, Roy and his co-workers [26] proposed a new simple external validation metric, as shown in the following equations:

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

$$r_m'^2 = R'^{\circ 2} (1 - \sqrt{|R^2 - R'^{\circ 2}|}) \quad (13)$$

$$\overline{r_{m(test)}^2} = \frac{r_m^2 + r_m'^2}{2} \quad (14)$$

This formula can be applied for both external and internal validation and the present study focuses on the external validation form.

In addition, we choose the concordance correlation coefficient (**CCC**) proposed by Lin [27] to measure the agreement between experimental and predicted data, which should be the real aim of any predictive **QSAR** models:

$$CCC = \frac{2 \sum_{i=1}^{ntest} (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sum_{i=1}^{ntest} (y_i - \bar{y})^2 + \sum_{i=1}^{ntest} (\hat{y}_i - \bar{y})^2 + n_{test} (\bar{y} - \bar{y})^2} \quad (15)$$

#### 5. Y-Randomization test

To establish model robustness, Y-randomization, randomization of the response variable, test was used this test consists of redoing all of the computations from the training set with scrambled activities. Calculations were repeated at least five times, to ensure reproducibility in the results, and after each iteration, a new **QSAR** model is developed [28].

New **QSAR** models had lower  $Q^2$  and  $R^2$  than those of the original models. This technique was performed to eliminate the possibility of random correlation. If higher values of  $Q^2$  and  $R^2$  are obtained, it means that an

acceptable **QSAR** cannot be generated for this dataset due to structural redundancy and random correlation.

Coefficient of determination,  $cR^2_p$  value has been reported to be greater than **0.5** for passing this test, and it is also calculated in the Y-randomization test and is expressed as:

$$cR_p^2 = R_x(R^2 - R_p^2)^2 \quad (16)$$

Where **R** is the correlation coefficient for Y-randomization and  $R^2_r$  is the average 'R' of the random models [22].

## 6. Assessment of the applicability domain of the model

The reliability of a **QSAR** classification model depends on its capacity to achieve confident predictions of new compounds not considered in the building of the model [29]. The ability of a **QSAR** classification model to make credible predictions of novel compounds not addressed in the model's construction determines its dependability. The domain of applicability (**AD**) is an important concept in **QSAR** which makes it possible to estimate the uncertainty in the prediction of a new compound according to its similarity with the compounds used to build the model [30].

The Williams plot, the plot of standardized residuals versus leverage values (**h**), was used in the present study to visualize the **AD** of the **QSAR** model.

Leverage value of a given chemical compound **i** is defined as:

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (17)$$

Where  $x_i$  is the descriptor row-vector of the query compounds **i**.

The warning leverage ( $h^*$ ) is the limit of normal values for **X** outliers and  $h^*$  is generally fixed at  $3(k + 1)/n$  (**k** is the number of model parameters and **n** is the number of training set compounds), whereas **x** = **2** or **3**. Prediction was considered unreliable for compounds with a high leverage value ( $h > h^*$ ). When a compound's leverage value is less than the threshold value, on the other hand, the agreement probability between observed and predicted values is as high as it is for the training set compounds [31-33].

## 7. Preparation of farnesyltransferase Protein and docking studies

Molecular docking calculations were carried out using MOE-Dock software [34]. The crystal structure of the protein farnesyltransferase (**PFT**) was obtained from the Protein Data Bank (PDB ID: 3E37) [35]. Crystal structure was edited to remove water molecules and was imported into MOE, and then all hydrogen atoms were added to the structure followed by their optimization using Amber10: EHT force field. Active site was identified, as presented in Figure 1. The 3D structures of the ligands were optimized using MOE software with MMFF94x force field and Root Mean Square (**RMS**) gradient value of 0.001 kcal/mol Å. The ligand database that was developed from the total set of 11 newly designed compounds were used for docking with the known **PFT** receptor active site. Thirty ligand-receptor complex conformations were generated for each test compound, and the conformation with the least docking score was considered for further analysis.



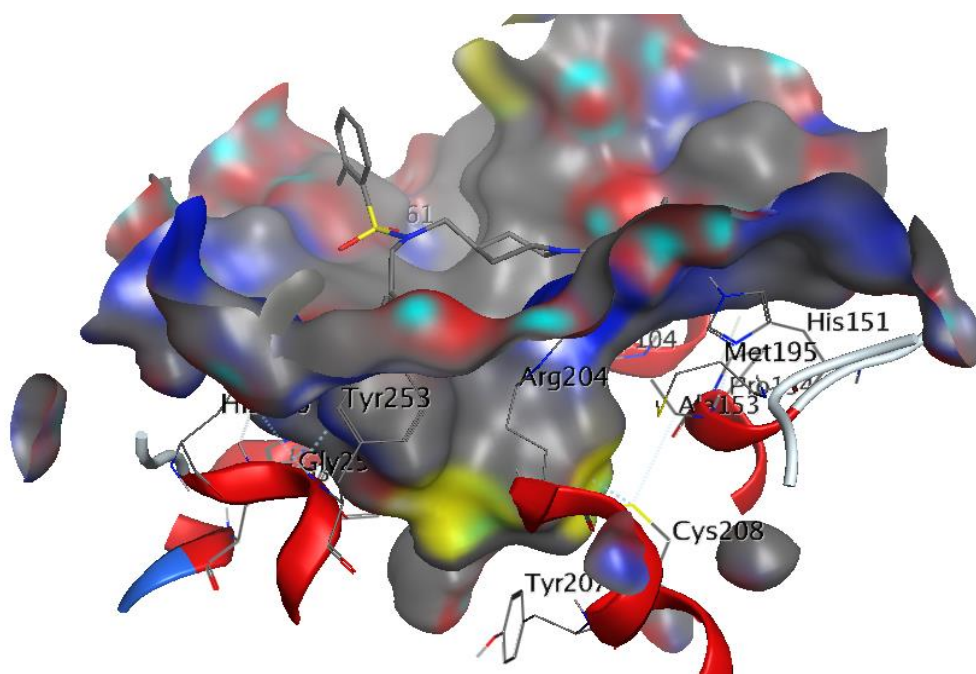


Figure 1: Binding site of PFT receptor (PDB ID: 3E37) in the complex with the substrate ethylenediamine ED5.

### III. RESULTS AND DISCUSSION

#### 1. Developed QSAR model and validation

The QSAR analysis was performed using calculated molecular descriptors and the experimental values of the antimalarial activities for the thirty-six benzophenone derivatives, thirteen multiple linear regression (MLR) models based on the same size of training sets and using the 19 descriptors already selected after preprocessing on the basis of the correlation coefficient. It is interesting to observe that bpol, FMF and Alogp2 are the most important descriptors among the other models. In addition, it can be observed that MLR has the best model as compared to the other statistical parameters. Indeed, model 13 shows the best one between the 13 MLR models as expressed in the following QSAR equation 19 with 4 variables. The statistical parameters of all the generated models with threshold values were presented in Table S3.

$$\begin{aligned} \text{pIC50} = & -6.818 - 6.234 \cdot 10^{-2} \text{Alogp2} + 0.156 \text{bpol} \\ & - 0.098 \text{MAXD} \\ & + 14.528 \text{FMF} \end{aligned} \quad (18)$$

$$\begin{aligned} N = 27 \quad R^2 = 0.884; \quad R^2 \text{ adj} = 0.865; \quad S = 0.038; \quad N_{\text{test}} = 9; \quad p < 0.0001; \quad F = 42; \quad Q^2 \text{ LOO} = 0.822 \end{aligned}$$

The obtained coefficient of correlation in equation (18) exhibits high value of correlation coefficient, 0.884, and low value of mean squared error, 0.038, which indicates that the model is more reliable. In addition, the coefficient p shows a lower value than 0.05 and the F-test has the value of 42, and these results demonstrate that the regression equation is statistically significant.

It can be observed the high adjusted value of the regression coefficient ( $R^2 \text{ adj} = 0.865$ ) from the QSAR model, as shown in Table 3; In addition, it has the same value as that of the regression coefficient ( $R^2 = 0.884$ ), this indicates that the developed model has a perfect descriptive capacity for the descriptors and illustrates the real impact of descriptors used on pIC50.

**Table 3: The validation parameters of the model which passed the three holds required for a QSAR model to be accepted**

Validation Tools	Interpretation	Acceptable Value	model Value
<b>R<sup>2</sup></b>	Co-efficient of determination	>0.6	0.884
<b>Q<sup>2</sup><sub>cv</sub></b>	Cross-Validation Coefficient	>0.5	0.822
<b>R<sup>2</sup><sub>adj</sub></b>	Adjusted R-squared	>0.6	0.865
<b>press</b>	predicted residual sum of squares	Press/ssy<	1.835
<b>SSY</b>	total sum of squares	0.4	0.209
<b>VIF</b>	Variance Inflation Factor	<5	1-2
<b>N<sub>Ext testset</sub></b>	Minimum number of external and test sets	5	8.0
<b>R<sup>2</sup><sub>Test set</sub></b>	Co-efficient of determination of external and test set	>0.5	0.821
<b>cR<sup>2</sup><sub>p</sub></b>	Coefficient of determination for Y-randomization	>0.5	0.811

Cross-validation is important way to explore the stability of a predictive model by using the analysis of the influence of each one of the individual objects that configure the final model.

The QSAR model expressed by equation (18) is cross-validated by its appreciable Q<sup>2</sup><sub>cv</sub> values (Q<sup>2</sup><sub>cv</sub> = 0.821) obtained using the leave-one-out method. The value of Q<sup>2</sup><sub>cv</sub> is higher than 0.5, which is important criterion for

qualifying a **QSAR** model as valid [36].

In addition, the low PRESS/SSY ratio, 0.114, indicates the accuracy of the developed **QSAR** model used in this study and this is in agreement with the previous study, which states that the PRESS/SSY ratio should be lower than 0.4 [37]. The four model descriptors MAXDn, bpol, logp2 and FMF used in this study are shown in Table 4.

**Table 4: Names of the model descriptors and their respective degree of contribution.**

Java class Descriptor	Descriptor	Description	Class	ME%
ALOGP Descriptor	Alogp2	Square of ALogP	2D	0.06
BPolDescriptor	bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens)	2D	45.38
Electrotopological State Atom Type Descriptor	MAXDn	Maximum negative intrinsic state difference in the molecule (related to the nucleophilicity of the molecule)	2D	1.60
FMF Descriptor	FMF	Complexity of a molecule	2D	52.37

The simplest method of investigating occurrence of inter correlation is to calculate Pearson correlation coefficients for the descriptors in the model [38], reported in Table 5. The model's low correlation coefficients suggest that there is no substantial inter correlation between the descriptors. Furthermore, the multicollinearity between the four descriptors for the model was detected by calculating their variation inflation factors (**VIF**), as shown in Table 5, using the following equation:

$$VIF_i = \frac{1}{1-R_i^2} \quad (19)$$

Where **R<sup>2</sup>** is the value obtained by regressing and *i* is predictor on the other predictors. Surprisingly, the calculated **VIF** values with less than 2 were observed, and the results are tabulated in Table 5. This result confirms that there is no significant intercorrelation among the descriptors used in building the model, and this is in

agreement with a previous study, which states that the high **VIF** value more than 5.0 indicates that the model is

unstable, while the value between 1.0 and 4.0 means that the model is acceptable [39].

**Table 5: Correlation matrix (Pearson (n)) between the different descriptors and VIF values.**

Descriptors	pbol	Logp2	FMF	MAXDn	VIF
pbol	1				2.018
Logp2	0.298	1			1.195
FMF	-0.181	-0.589	1		1.789
MAXDN	-0.103	0.028	0.140	1	1.034

Recently, the maximum negative intrinsic state difference (**MAXDn**), which relates with the molecular nucleophilicity have been studied as Kier-Hall intrinsic state atom type descriptor [40-41]. The ratio of Kier-Hall atomic electronegativity to the vertex degree is used to calculate atom. Thus, the number of bonds of the atom and encoding information were related to both partial charges of atoms and their topological negative relative to the whole molecule.

Since **MAXDn** has negative sign for the linear Eq. (18), increasing the value of descriptors via electrophilicity behavior of compounds has been shown to decrease the **pIC50** values.

FMF is a word that refers to the idea of molecular topology and the percentage of a molecular framework made up of terminal rings and a molecular bridge [42]. It has been shown that **FMF** correlates to the **ADMET** properties, such as solubility, permeability and Cytochrome P450 isoform 3A4 inhibition, as well [43]. As for the third descriptor **bpol** which is the sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule [44]. The results correlate with the antimalarial activities of the benzophenone derivatives. This implies that an overall increasing in the polarizability of the compound improves the antimalarial activity of benzophenone.

Molecular lipophilicity, usually quantified in log P, is an important molecular characteristic in medicinal chemistry and in rationalized drug design; the log P

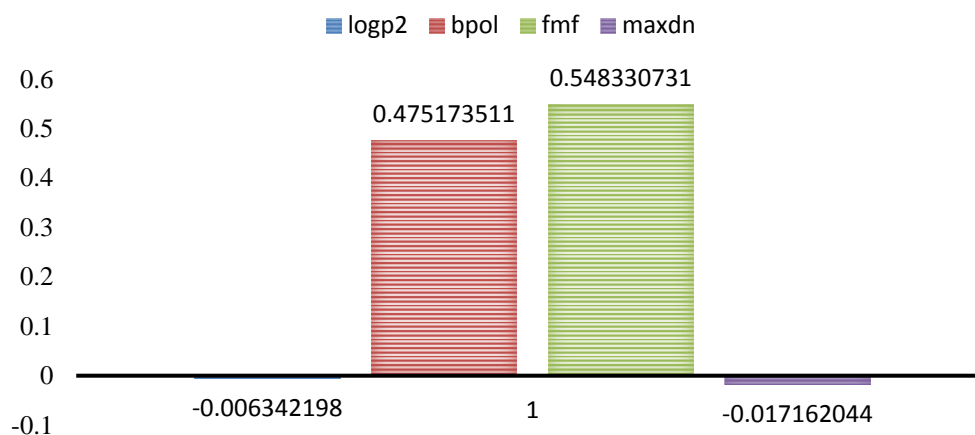
coefficient is well known as one of the main parameters for the estimation of lipophilicity of chemical compounds and determines their pharmacokinetic properties [45].

Log P has been linked to a wide range of biological activities, including pharmacological activity, toxicity, pesticidal action, genotoxic activity, and more. The lipophilicity is a main physico-chemical determinant influencing the bioavailability, and refers to the octanol/water partition coefficient descriptor **logP(o/w)** [46]. To study the lipophilicity of the benzophenone derivatives, the descriptor **logp2** was analyzed, and the results are presented in Table1. The negative coefficient of the **logp2** in model **MLR** suggests the increasing in the overall lipophilicity of the molecule that leads to decrease the **PFT** inhibitory activity of benzophenone derivatives.

On the other hand, the contribution of every descriptor in the built model was evaluated by the computation of the mean effect (**ME**) value [47] by using Eq. (20). The values for the **ME** are shown in both Figure 2 and Table 4:

$$ME_i = \frac{\beta_j \sum_{i=1}^n d_{ij}}{\sum_j^n \beta_j \sum_i^n d_{ij}} \quad (20)$$

Where **Mei** is the mean effect of the descriptor **j**,  **$\beta_j$**  represents the coefficient of the descriptor **j**, and  **$d_{ij}$**  represents the value of the selected descriptors of each compound and the total number of descriptors in the produced model is given by **m**.



**Figure 2: Contribution of descriptors in models.**

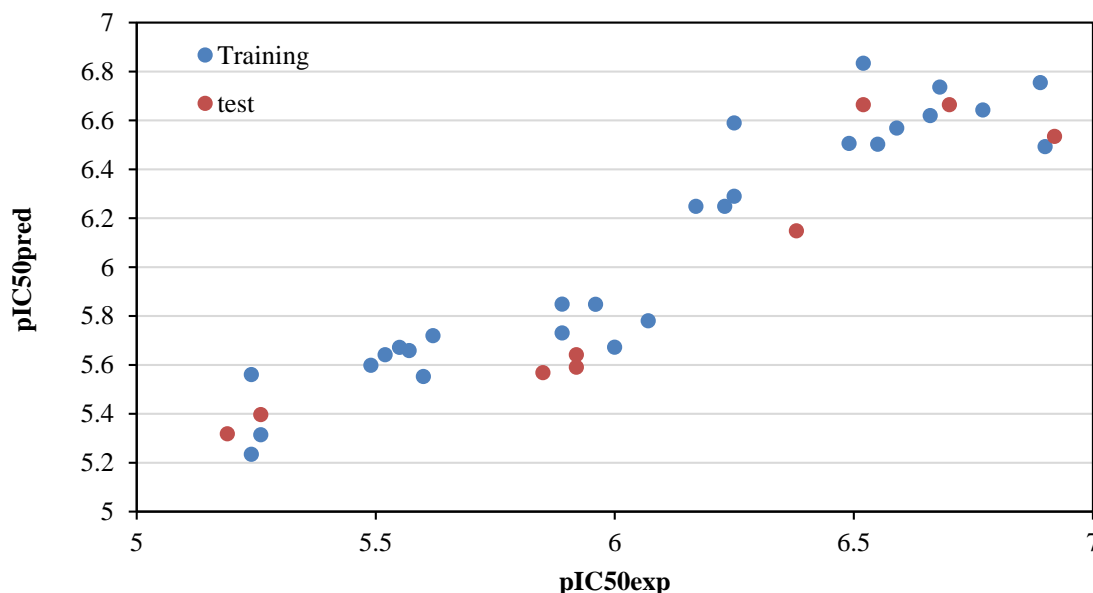
Based on the calculated **ME** values, the greatest influence on the antimalarial activities among the four descriptors was **FMF**, and the trend is in the order of **FMF** > **bpol** > **logp2** > **MAXDn**. This suggests that the **FMF** should be highly considered when designing high potent benzophenone derivatives.

Previously, Roy suggested the best way to estimate the true predictive power of a **QSAR** model is by comparing between the predicted and observed activities of an external test set of compounds in the developed **QSAR** model [48].

The values  $R^2_{pred} = 0.821$  indicate absolute quality of fitness the predicted model. The predicted activities of the test data compounds were studied for the **PFT** inhibitors

by the developed **QSAR** model, and the results are tabulated in Table 1. Generally, Low residue values found in both the training set and the test, this indicated that the model has an ability to correlate activity and structure. The statistical analysis results showed that the correlation between experimental activity and predicted activity according to the model was highly significant.

The correlation between experimental and predicted values of **pIC50** of both the training and test sets obtained by **MLR** model is presented in Figure 3. The correlation showed clearly that the obtained **MLR** model from Eq. (18) represents reasonably well over the entire range of the **pIC50** values.



**Figure 3: A correlation plot between predicted pIC50 (PRED) values on y-axis and pIC50 values on x-axis for both the training and test sets.**

All the external validation results were above the threshold values for the various parameters presented in Table 3.

The squared correlation coefficient values between the observed and predicted values of the test set compounds ( $r^2$ ) and ( $r^2_0$ ), respectively were observed, and the model had satisfied the requirement of the term  $(r^2 - r^2_0)/r^2$ . This was in agreement with a previous study reported by Golbraikh and co-workers, which states that the value  $(r^2 - r^2_0)/r^2$  exhibits less than 0.1.

In case of good external prediction, predicted values will be very close to observed activity values. Therefore,  $R^2$  value will be very near to  $R^2_0$  value. In the best case,  $r^2_m$  will be equal to  $r^2$  whereas in the worst-case  $r^2_m$  value will be zero; including values of  $r^2_m < 0.6$  indicate these models are useless for external predictivity [49].

In the present study  $r^2_m$  value of the model [Equation 18] are acceptable (Table 6). This developed model passed all the Golbraikh and Tropsha criteria for the acceptability of the model (Table 7).

**Table 6: Validation characteristics of developed model according to  $r^2_m$  metrics and Concordance correlation coefficient**

$r^2_m$ parameter			Concordance correlation coefficient
$r^2_m$	$r'^2_m$	$\Delta r^2_{m(test)}$	CCC
0.553	0.533	0.02	0.946
>0.5	>0.5	<0.2	>0.85

**Table 7: Golbraikh and Tropsha's criteria for the model.**

$R^2_{pred}$	K	K'	$R^{\circ 2}$	$R'^{\circ 2}$	$\frac{R^2 - R^{\circ 2}}{R^2}$	$\frac{R^2 - R'^{\circ 2}}{R^2}$	$ R_0^2 - R^{\circ 2} $
<b>0.821</b>	1,012	0.978	0.945	0.944	0.149	-0.148	0.001
>0.6	>0.85	<1.15	close to 1	close to 1	<0.1	<0.1	< 0.3

$R$ : Correlation coefficient between the predicted and observed activities

$R^{\circ 2}$ : Coefficients of determination predicted versus observed activities

$R'^{\circ 2}$ : Coefficients of determination observed versus predicted activities.

$K$  and  $K'$ : Slopes of the regression lines

## 2. Randomization test

The Y-Randomization method was carried out to validate the **MLR** and to detect and quantify chance correlations between the dependent variable and descriptors. By comparing the resulting scores between randomization test and the original **QSAR** equation which

generated with non-randomized data, we have found that the new **QSAR** models, after several iterations, have low  $R^2$  and  $Q^2_{Loo}$  values, as shown in Table 8. This result indicates that the good **MLR** models obtained in this study cannot be attributed to the chance correlation of the training set.

**Table 8: Y-randomization table for QSAR Analysis.**

Model	R	$R^2$	$Q^2$
Original	0.940	0.884	0.822
Random 1	0.459	0.211	-2.191
Random 2	0.458	0.210	-0.233
Random 3	0.213	0.045	-0.437
Random 4	0.283	0.080	-5.857
Random 5	0.332	0.110	-0.779
Random 6	0.280	0.078	-0.426
Random 7	0.362	0.131	-1.084
Random 8	0.531	0.282	-0.440
Random 9	0.406	0.164	-3.358
Random 10	0.403	0.162	-0.573
<b>Random Models Parameters</b>			
Average r	0.373		
Average $r^2$	0.147		
Average $Q^2$	-1.538		
cRp <sup>2</sup>	0.811		

### 3. Applicability Domain

The applicability domain (AD) approaches were used to estimate the prediction reliability for each modeled compound individually using many prediction methods. The plot of standardized residuals in prediction vs leverage values is a typical technique for visualizing the AD of a QSAR model. This plot, called as the Williams plot, allows for quick and easy graphical detection of response outliers and structurally influential chemicals in a model ( $h_i > h^*$ ). Where  $h^*$  is a threshold value, in fact, when a compound's leverage value is less than the crucial value  $h^*$ , the chance of predicted and actual values agreeing is as great as it is for the training set chemicals. A high leverage chemical,

on the other hand, is structurally distinct from the other chemicals and may thus be regarded outside the AD of the model. [50]. In this study, the results of the leverage values, standardized residuals of the observables and the test of the model which used to develop the applicability domain, are tabulated in Table 9. It has been observed that the standardized residual values for all compounds are in the range of  $-2.5$  to  $2.5$ . The lifts obtained are all below the critical value  $h^* = 0.55$ . Domain As shown in the developed Williams plot on the selected descriptors for predicting the PFT inhibitory activity all the compounds from the data set are in this area (Figure 4)

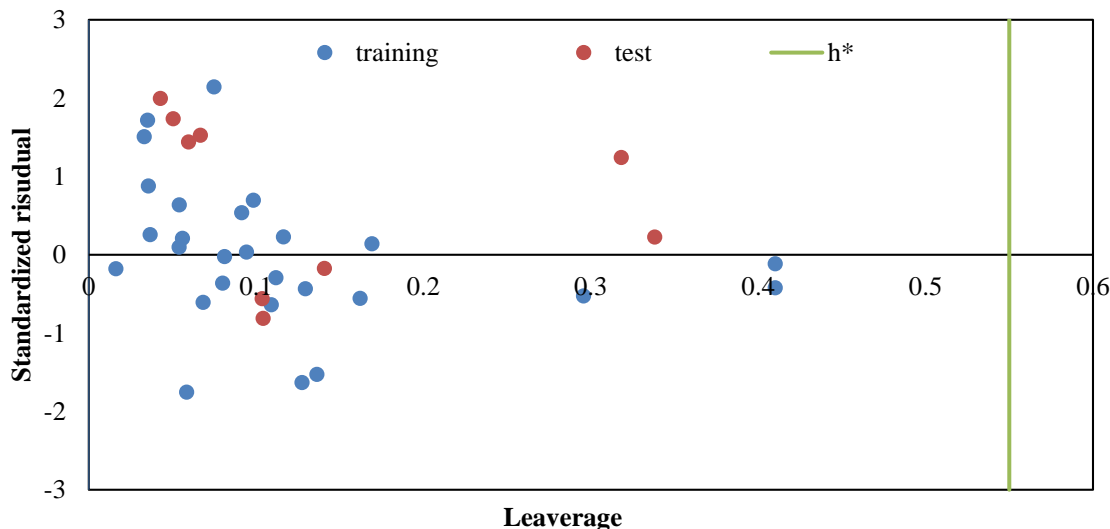


Figure 4: The Williams plot of MLR model for the training and test sets.

Table 9: The values of leverage and standardized residuals of the observables and test set

N. Compound	S. Residual	Leverage	N. Compound	S. Residual	Leverage
1	-1.528	0.136	25	0.033	0.094
2	0.877	0.035	26	0.097	0.054
4	-0.607	0.068	27	0.208	0.056
5	1.506	0.033	28	0.636	0.054
6	0.536	0.091	29	-0.180	0.016
7	0.227	0.116	30	0.256	0.036
8	0.139	0.169	31	-1.755	0.058
9	-0.557	0.162	32	-0.526	0.295

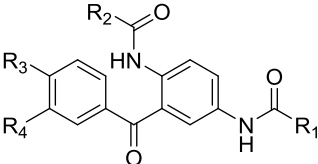
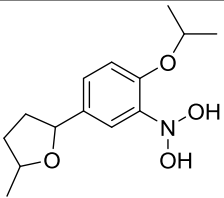
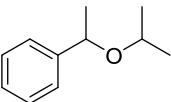

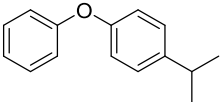
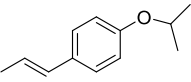

N. Compound	S. Residual	Leverage	N. Compound	S. Residual	Leverage
<b>10</b>	-0.638	0.109	<b>33</b>	-0.435	0.129
<b>11</b>	-0.362	0.079	<b>35</b>	-0.175	0.140
<b>12</b>	1.717	0.035	<b>36</b>	1.525	0.066
<b>13</b>	0.695	0.098	<b>3t</b>	-0.813	0.104
<b>14</b>	-1.631	0.127	<b>6t</b>	1.240	0.318
<b>15</b>	-0.293	0.111	<b>16t</b>	0.225	0.338
<b>17</b>	-0.024	0.081	<b>19t</b>	1.995	0.042
<b>18</b>	2.141	0.074	<b>20t</b>	-0.561	0.103
<b>23</b>	-0.116	0.410	<b>21t</b>	1.7364	0.050
<b>24</b>	-0.423	0.410	<b>22t</b>	1.4385	0.059

#### 4. Design of novel derivatives

QSAR method has been playing important role for synthesis lead molecules and detect their biological activities. Therefore, the development of new benzophenones derivatives with strong affinity for **PFT** receptor could be achieved using core scaffolds that mimicked the best effective lead molecule. In this study, the equation of **MLR** model was applied for the 15 of newly designed benzophenone hybrids. A detailed profile of **bpol**

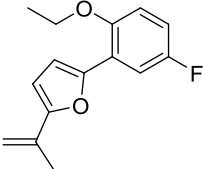
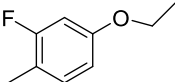

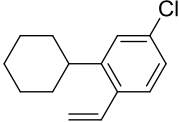
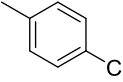

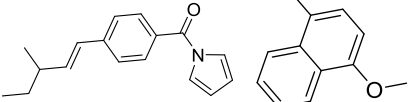
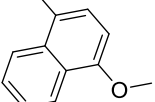
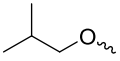
and **FMF** descriptors was analyzed using the equation of **MLR** model, and the descriptors have been shown to increase the activity by increasing their values. Thus, the benzophenone backbones were developed by adding polar OH and NH<sub>2</sub> functionality and incorporation ring into the benzophenones scaffold. The predicted values and the calculated leverages (*h*) values were calculated for the different derivatives, and they are tabulated in Table 10.

**Table 10: Structures of the newly designed inhibitors, predicted values and calculated *h* of pIC<sub>50</sub> (in μM) used in this study.**

COMP. N					pIC <sub>50</sub> pred μM	<i>h</i>
	R1	R2	R3	R4		
<b>P1</b>				H	8.067	0.541
<b>P2</b>				H	7.783	0.391



<b>P3</b>				H	9.259	0.799
<b>P4</b>				H	10.374	0.866
<b>P5</b>				H	8.360	0.439
<b>P6</b>				H	7.647	0.279
<b>P7</b>				H	7.976	0.673
<b>P8</b>				H	7.842	0.360
<b>P9</b>				H	8.614	0.713
<b>P10</b>			-OCH <sub>3</sub>	CH <sub>3</sub> -	7.776	0.441
<b>P11</b>			-CH <sub>2</sub> CH <sub>3</sub>	-OCH <sub>3</sub>	7.923	0.486
<b>P12</b>					8.804	0.509

<b>P13</b>				-F	7.588	0.399
<b>P14</b>				-F	8.250	0.434
<b>P15</b>				H	8.251	0.471

### 5. Molecular docking studies of the newly designed inhibitors

The Protein-Ligand interaction plays a vital role in structural based drug design [51-54]. In this present study, Molecular docking of benzophenone derivatives (**P1-P15**) was studied with **PFT** receptor, PDB ID: 3E37, to compare the binding interactions and the binding free energies ( $\Delta G$ ). The newly designed ligands with the most binding affinities and the synthetic substrate ethylenediamine inhibitor (**ED5**) used in this study are shown in Table 11. It can be observed that all the benzophenone derivatives occupy the same binding pocket, cavity 1, as that of synthetic substrate **ED5** (Figure S1).

The docking results of the complexes show that the binding free energies of benzophenone derivatives are in the same range as that of **ED5**,  $-10.28 \text{ kcal mol}^{-1}$ , whereas **P1** has the lowest binding free energy of  $-12.38 \text{ kcal mol}^{-1}$ . The complex **P7-PFT** show a binding energy of  $-8.79$

$\text{kcal mol}^{-1}$ , this indicates that **P7-PFT** have the lowest binding affinities.

In addition, the highest binding affinity, **P1-PFT** complex is engaged in hydrogen bonding interaction with the Asp354, Lys358, Cys301 and His250 amino acid residues, the hydrogen bonds were examined based on the acceptor-donor atom distance, as shown in Table 12. Upon comparing the occupancy of hydrogen bonds for the higher binding affinity **P1**, **22** and the ligand reference **ED5**, the results suggest that the newly designed complex **P1-PFT** has higher hydrogen bonds occupancy and provides stability to the PFT receptor. Furthermore, it was observed that benzophenone derivatives along with **P1** formed obvious hydrophobic interactions with Trp305 as shown in Figure 5.

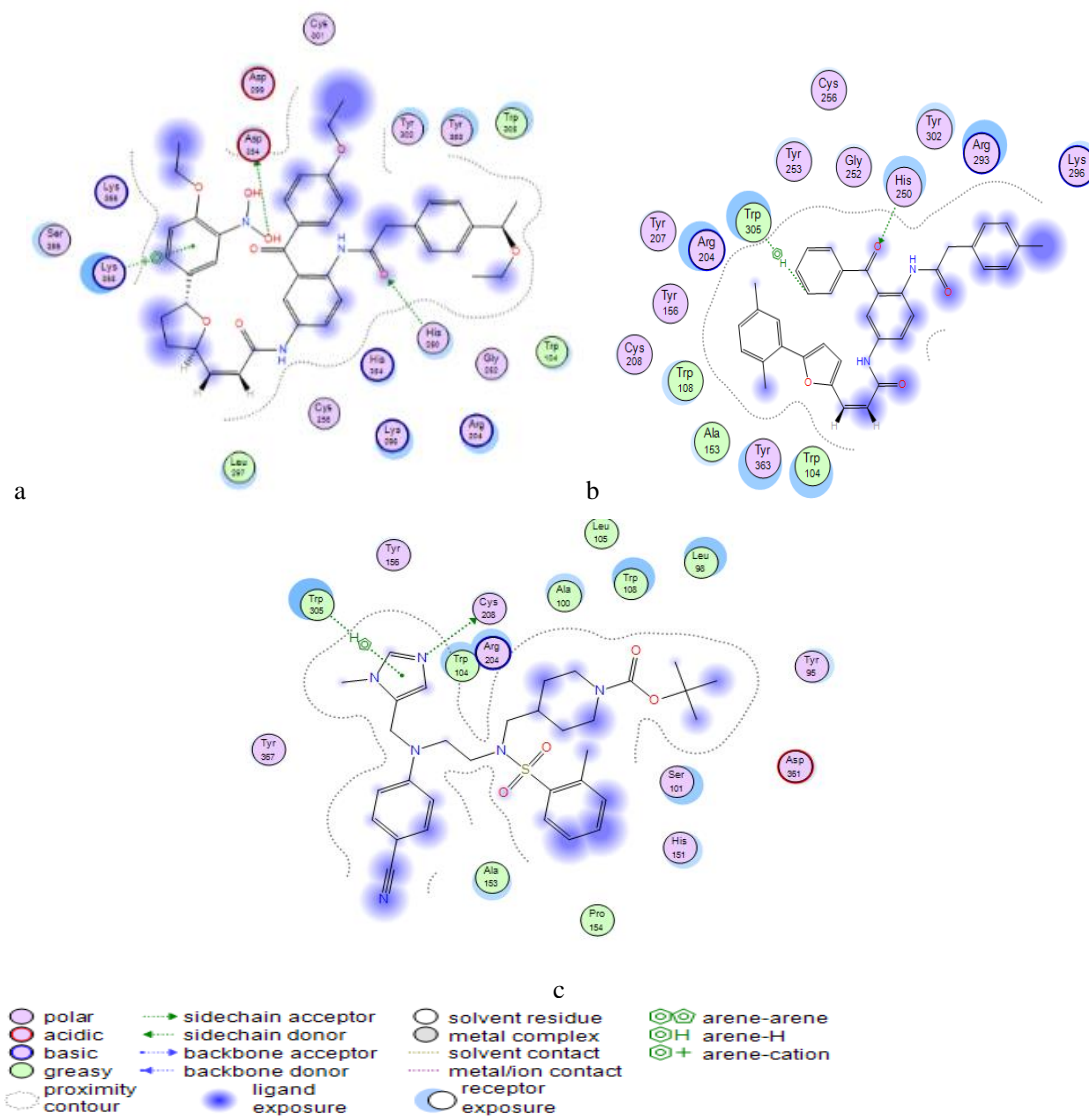
Overall, the obtained results show that benzophenone derivatives can form stable complexes with PFT, and the best candidate **P1** binds to the receptor with higher affinity and stability.

**Table 11: Binding free energies (in  $\text{kcal mol}^{-1}$ ) of the PFT receptor interacting with the most active benzophenone derivatives and the design substrate.**

Compound	Binding energy, $\Delta G$	Compound	Binding energy, $\Delta G$
<b>ED5</b>	-10.28	<b>P10</b>	<b>-8.796</b>
<b>22</b>	-9.06	<b>P11</b>	<b>-9.226</b>
<b>P1</b>	<b>-12.38</b>	<b>P12</b>	<b>-9.702</b>
<b>P2</b>	-9.372	<b>P13</b>	<b>-9.140</b>
<b>P5</b>	<b>-11.18</b>	<b>P14</b>	<b>-10.04</b>
<b>P6</b>	-9.490	<b>P15</b>	<b>-9.991</b>
<b>P8</b>	-9.129	/	/

**Table 12: Molecular interactions between PFT receptor and P1, 22 and ED5**

Compound	Ligand	Receptor	Interaction	Distance	E (kcal/mol)
P1	O	ASP 354	H-donor	2.82	-6.6
	N	LYS 358	H-acceptor	3.95	-0.7
	O	CYS 301	H-acceptor	3.98	-0.7
	O	HIS 250	H-acceptor	3.49	-1.4
	6-ring	LYS 358	pi-cation	3.64	-3.0
22	N	HIS 250	H-acceptor	3.02	-2.8
	5- ring	TRP 305	H-pi	4.39	-1.1
ED5	N	CYS 208	H-donor	3.55	-1.1
	5-ring	TRP 305	pi-H	4.60	-0.7



**Figure 5: 2D representations of (a) Ligand ED5 (b) Ligand 22 (c) Ligand P1 docked in the binding pocket of the PFT. The important residues involved in the H-bond formation and hydrophobic interactions are highlighted.**

**CONCLUSION**

Benzophenone derivatives as protein *farnesyltransferase* (PFT) inhibitors were modified using quantitative structure–activity relationship (QSAR) screening and molecular docking calculations. Thirty-six proposed ligands were studied by QSAR method, which used to propose and compare thirty MLR model results, the best model was found in agreement with the Tropsha and Golbraikh criteria. The results indicate that the descriptors such as pbol and FMF values modulate the activity of the molecules. In addition, the applicability domain (AD) results demonstrate that all the proposed compounds were within the defined domain. The pIC50 activity of fifteen proposed molecules was calculated

using the proposed MLR model. Molecular docking results of the best predicted active compounds ( $h < h^*$ ) with the PFT receptor demonstrate that **P1** binds well to the hydrophobic contacts and polar amino acid residues, leading to the increasing the binding affinity and thus enhancing the complex stability. Finally, the reasonable strategy for QSAR method and the high binding affinity for the newly designed ligands with PFT suggest that this benzophenone scaffolds may be worth exploring in the development of new inhibitors to treat malaria disease.

**Acknowledgements**

The authors are grateful to the University of Biskra, Algeria, and Middle East University; Amman, Jordan for the academic license of the software is used in this research article.

**Supplementary Material****Table S1: Pearson's correlation matrix for descriptors used in QSAR model**

	ALogP	ALogP2	AMR	Apol	Bpol	Lipo	MAXDP	DELS2	FMF	MAXDN	VABC	VAdjM	MW	AMW	WPATH	WPOL	XLogP	Zagreb	Tpsa	
ALogP	1																			
ALogP2	0.002	1																		
AMR	-0.090	0.124	1																	
Apol	0.120	0.367	0.724	1																
Bpol	-0.091	0.279	0.771	0.894	1															
Lipo	0.387	0.083	0.280	0.628	0.586	1														
MAXDP	0.000	0.252	0.785	0.882	0.836	0.552	1													
DELS2	0.133	0.182	0.380	0.380	0.360	0.105	0.380	1												
FMF	0.372	0.155	-0.68	-0.28	-0.58	-0.21	-0.460	-0.120	1											
MAXDN	0.177	-0.060	0.017	-0.08	0.013	0.193	-0.042	0.723	-0.12	1										
VABC	0.122	0.328	0.792	0.979	0.875	0.558	0.894	0.520	-0.33	0.033	1									
VAdj	0.187	0.359	0.647	0.944	0.794	0.536	0.828	0.617	-0.15	0.116	0.970	1								
MW	0.279	0.266	0.679	0.845	0.683	0.479	0.847	0.574	-0.13	0.121	0.890	0.895	1							
AMW	0.393	-0.158	-0.11	-0.21	-0.42	-0.18	-0.041	0.228	0.411	0.214	-0.124	-0.049	0.315	1						
WPATH	0.200	0.382	0.607	0.937	0.768	-0.519	0.785	0.595	-0.09	0.086	0.957	0.994	0.874	-0.06	1					
WPOL	0.182	0.363	0.661	0.933	0.804	0.550	0.875	0.634	-0.20	0.152	0.959	0.982	0.905	-0.02	0.967	1				
XLogP	0.581	0.193	-0.02	0.335	0.232	0.743	0.132	-0.097	0.034	0.101	0.246	0.246	0.147	-0.24	0.263	0.243	1			
Zagreb	0.260	0.407	0.551	0.924	0.737	0.542	0.769	0.597	-0.01	0.107	0.932	0.985	0.876	-0.02	0.989	0.968	0.297	1		
Tpsa	-0.17	0.319	0.423	0.491	0.334	-0.27	0.368	0.537	0.014	-0.065	0.565	0.620	0.482	-0.02	0.641	0.579	-0.311	0.607	1	

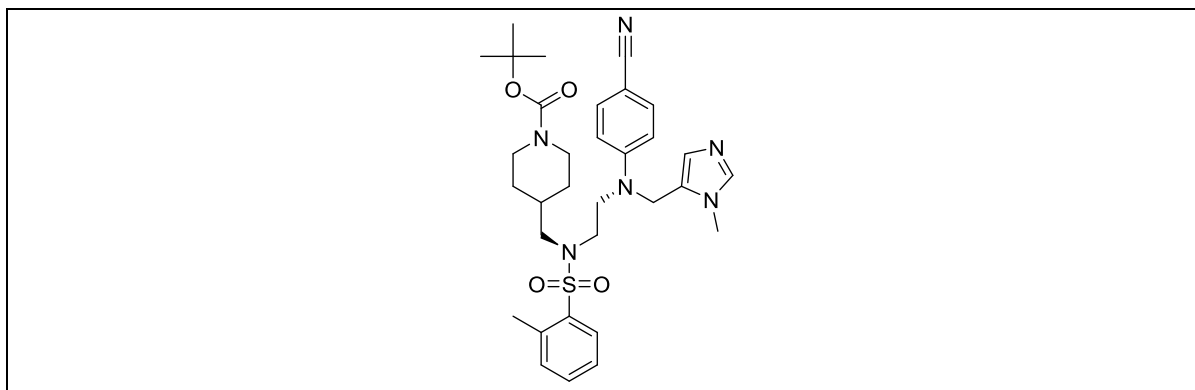
**Table S2: List of physiochemical descriptors used for the best model**

Descripteurs	Description
ALogP	Ghose-CrippenLogKow
ALogP2	Square of ALogP
Bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens)
TPSA	Topological polar surface area
MAXDN	Maximum negative intrinsic state difference in the molecule (related to the nucleophilicity of the molecule). Using $\Delta V = (Z_v - \max \text{BondedHydrogens}) / (\text{atomicNumber} - Z_v - 1)$ . Gramatica, P., Corradi, M., and Consonni, V. (2000). Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. Chemosphere 41, 763-777.
MAXDP	Maximum positive intrinsic state difference in the molecule (related to the electrophilicity of the molecule). Using $\Delta V = (Z_v - \max \text{BondedHydrogens}) / (\text{atomicNumber} - Z_v - 1)$ . Gramatica, P., Corradi, M., and Consonni, V. (2000). Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. Chemosphere 41, 763-777.

Descripteurs	Description
FMF	Complexity of a molecule
MW	Molecular weight
AMW	Average molecular weight (Molecular weight / Total number of atoms)
WPOL	Weiner polarity number
WPATH	Weiner path number
XLogP	Partition coefficient of Wang's octanol water
Lipo	Lipoaffinity index
DELS	Sum of all atoms intrinsic state differences (measure of total charge transfer in the molecule). Using $\Delta V = (Z_v - \max \text{BondedHydrogens}) / (\text{atomicNumber} - Z_v - 1)$ . Gramatica, P., Corradi, M., and Consonni, V. (2000). Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. Chemosphere 41, 763-777.
AMR	Molar refractivity
Apol	Sum of the atomic polarizabilities (including implicit hydrogens)
vabc	Van der Waals volume calculated using the method proposed in [Zhao, Yuan H. and Abraham, Michael H. and Zissimos, Andreas M., Fast Calculation of van der Waals Volume as a Sum of Atomic and Bond Contributions and Its Application to Drug Compounds, The Journal of Organic Chemistry, 2003, 68:7368-7373]
V adjamat	Vertex adjacency information (magnitude)
Z agreb	Sum of the squares of atom degree over all heavy atoms i

**Table S3: Statistical results of different QSAR model.**

Descriptor	$R^2_{adj}$	S	$R^2_{test}$
<b>1</b> <i>WPATH-TopoPSA</i>	0.686	0.099	0.358
<b>2</b> <i>bpol FMF</i>	0.640	0.175	0.645
<b>3</b> <i>bpolFMF</i>	0.679	0.133	0.530
<b>4</b> <i>WPATH-XLogP-TopoPSA</i>	0.649	0.119	0.434
<b>5</b> <i>bpol-FMF-WPATH</i>	0.797	0.082	0.408
<b>6</b> <i>ALogp2bpolFMF</i>	0.669	0.117	0.664
<b>7</b> <i>ALogp2-bpol-FMF-MW</i>	0.707	0.103	0.578
<b>8</b> <i>ALogp<sup>2</sup>bpol-FMF-TopoPSA</i>	0.755	0.135	0.508
<b>9</b> <i>bpol- MW-WPATH</i>	0.823	0.081	0.395
<b>10</b> <i>bpol-MAXDN-FMF-XLogP-TopoPSA</i>	0.798	0.092	0.477
<b>11</b> <i>Apol-MAXDN-FMF-MW-AMW</i>	0.708	0.082	0.026
<b>12</b> <i>ALogp2-bpol-MAXDN-FMF-TopoPSA</i>	0.758	0.082	0.555
<b>13</b> <i>ALogp2-bpol-MAXDN-FMF</i>	0.865	0.038	0.821



**Figure S1: structure of ligand of reference tert-butyl 4-((2-((4-cyanophenyl) (1-methyl-1H-imidazol-5-yl)methyl)amino}ethyl)[(2-methylphenyl)sulfonyl]amino)methyl)piperidine-1-carboxylate C<sub>32</sub> H<sub>42</sub> N<sub>6</sub> O<sub>4</sub> S**

## REFERENCES

- [1] World malaria report 2020, World Health Organization, Geneva, Switzerland, 2020, p 300
- [2] Prakash N., Patel S., Faldu N.J., Ranjan R. and Sudheer D.V.N. Molecular Docking Studies of Antimalarial Drugs for Malaria. *J. Comput. Sci.Syst. Biol.* 2010; 3: 70-73. doi:10.4172/jcsb.1000059
- [3] Hameed A., Masood S., Hameed A., Ahmed E., Sharif A. and Abdullah M.I. *J. Comput. Aided. Mol. Des.* 2019; 33:677-688.
- [4] Sharma K. A. Review on Plasmodium Falciparum-Protein Farnesyltransferase Inhibitors as Antimalarial Drug Targets. *Curr. Drug. Targets.* 2017; 18:1676–1686. <https://doi.org/10.2174/1389450117666160823165004>
- [5] Singh J., Mansuri R., Vijay S., Sahoo G.C., Sharma A. and Kumar M. Docking predictions-based Plasmodium falciparum phosphoethanolamine methyl transferase inhibitor identification and in-vitro antimalarial activity analysis. *BMC.Chem.* 2019; 13:43. <https://doi.org/10.1186/s13065-019-0551-5>
- [6] Subramanian T., Liu S., Troutman J.M., Andres D.A. and Spielmann H.P. Protein Farnesyltransferase-Catalyzed Isoprenoid Transfer to Peptide Depends on Lipid Size and Shape, not Hydrophobicity. *Chem.Bio.Chem.* 2008; 9:2872-2882. <https://doi.org/10.1002/cbic.200800248>
- [7] Kumar S., Bhardwaj T.R., Prasad D.N. And Singh R.K. Drug targets for resistant malaria: Historic to future perspectives. *Biomed. Pharmacother.* 2018;104: 8–27. <https://doi.org/10.1016/j.biopha.2018.05.009>
- [8] Wiesner J., Kettler K., Sakowski J., Ortmann R., Jomaa H. and Schlitzer M. Structure–Activity relationships of novel anti-Malarial agents: Part 5. N-(4-acylamino-3-benzoylphenyl)-[5-(4-nitrophenyl)-2-furyl] acrylic acid amides. *Bioorg. Med. Chem. Lett.* 2003; 13:361–363. [https://doi.org/10.1016/S0960-894X\(02\)01003-X](https://doi.org/10.1016/S0960-894X(02)01003-X)
- [9] Choudhari P. B., Bhatia M. S., Bhatia N. M. Application of pocket modeling and k-nearest neighbor molecular field analysis (kNN-MFA) for designing of some anticoagulants: potential factor IXa inhibitors. *Med. Chem. Res.* 2013; 22:976-985.
- [10] Roy K., Kar S., Das R.N. Background of QSAR and Historical Developments. Editors. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment . Boston: Academic Press, 2015; Chapter 1, pp 1–46. <https://doi.org/10.1016/B978-0-12-801505-6.00001-6>
- [11] Neves B.J., Braga R.C., Melo-Filho C.C., Moreira-Filho J.T., Muratov E.N. and Andrade C.H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front Pharmacol.* 2018; 9:1275. <https://doi.org/10.3389/fphar.2018.01275>

- [12] Moukhliiss Y., ElKhatabi K., Koubi Y., Maghat H., Sbai A, Bouachrine. and M.Lakhlifi T. 2D-QSAR modeling of novel pleconaril derivatives (isoxazole-based molecules) as antiviral inhibitors against Cocksackievirus B3 (CVB3). *Jordan Journal of Pharmaceutical Sciences*. 2021; 14:137- 156
- [13] Wiesner J., Kettler K., Sakowski J., Ortmann R., JomaaH.and Schlitzer M. Structure–Activity relationships of novel anti-Malarial agents: Part 5. N-(4-acylamino-3-benzoylphenyl)-[5-(4-nitrophenyl)-2-furyl] acrylic acid amides. *Bioorg. Med. Chem. Lett.* 2003;13: 361-363. [doi:10.1016/S0960-894X\(02\)01003-X](https://doi.org/10.1016/S0960-894X(02)01003-X)
- [14] Wiesner J., Fucik K., Kettler K., Sakowski J., Ortmann R. and Jomaa H. Structure–Activity relationships of novel anti-malarial agents. Part 6: N-(4-Arylpropionylamino-3 benzoylphenyl)-[5-(4-nitrophenyl)-2-furyl]acrylic acid amides; *Bioorg. Med. Chem. Lett.* 2003; 13:1539-1541. [https://doi.org/10.1016/S0960-894X\(03\)00179-3](https://doi.org/10.1016/S0960-894X(03)00179-3)
- [15] Wiesner J., Mitsch A., Wißner P., Krämer O., JomaaH.and Schlitzer M. Structure–Activity relationships of novel anti-Malarial agents. Part 4: N-(3-Benzoyl-4-tolylacetylaminophenyl)-3-(5-aryl-2-furyl)acrylic acid amides. *Bioorg. Med. Chem.Lett.* 2002;12: 2681-2683. [doi:10.1016/S0960-894X\(02\)00555-3](https://doi.org/10.1016/S0960-894X(02)00555-3)
- [16] Wiesner J., Mitsch A., Jomaa H. and Schlitzer M. Structure–activity relationships of novel anti-malarial agents. Part 7: N-(3-Benzoyl-4-tolylacetylaminophenyl)-3-(5-aryl-2-furyl)acrylic acid amides with polar moieties. *Bioorg. Med. Chem. Lett.* 2003; 13:2159-2161. [doi:10.1016/S0960-894X\(03\)00353-6](https://doi.org/10.1016/S0960-894X(03)00353-6)
- [17] Calculator Plugins, Marvin 6.3.0, 2014, ChemAxon (<http://www.chemaxon.com>).
- [18] HyperChem (Molecular Modeling System) (2007) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA
- [19] Ruiz I.L. and Gómez-Nieto M.Á. Study of the Applicability Domain of the QSAR Classification Models by Means of the Rivality and Modelability Indexes. *Molecules*. 2018; 23:2756. <https://doi.org/10.3390/molecules23112756>
- [20] Saxena A.K. and Prathipati P. Comparison of MLR, PLS and GA-MLR in QSAR analysis. *SAR. QSAR. Environ. Res.* 2003; 14:433–45. <https://doi.org/10.1080/10629360310001624015>
- [21] Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Info.*2010; 29: 476–88. <https://doi.org/10.1002/minf.201000061>
- [22] TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Info.* 2010; 29: 476 – 488.
- [23] Gramatica P. Principles of QSAR models validation: internal and external. *QSAR. Comb. Sci.* 2007;26 :694–701. <https://doi.org/10.1002/qsar.200610151>
- [24] Golbraikh A. and Tropsha A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J. Comput. Aided. Mol. Des.* 2002 ;16: 357–369. <https://doi.org/10.1023/A:1020869118689>
- [25] Roy K., Kar S. and Ambure P. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Syst.* 2015; 145:22–29. <https://doi.org/10.1016/j.chemolab.2015.04.013>
- [26] Lin L.I. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989; 45:255–68.
- [27] Golbraikh A., Wang X.S., Zhu H., Tropsha A., *Predictive QSAR Modeling: Methods and Applications in Drug Discovery and Chemical Risk Assessment*: J. Leszczynski (Ed.), Handbook of Computational Chemistry, Springer Netherlands, Dordrecht, 2016
- [28] Rücker C., Rücker G.and Meringer M. y-Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.* 2007; 47:2345–2357. <https://doi.org/10.1021/ci700157b>
- [29] Ouassaf M., Belaidi S., Benbrahim I., Belaidi H. and Chtita S. Quantitative Structure-Activity Relationships of 1.2.3 Triazole Derivatives as Aromatase Inhibition Activity. *Turkish Comp. Theo. Chem.* 2020; 4:1–11. <https://doi.org/10.33435/tcandtc.545369>

- [30] Ouassaf M., Belaidi S., Lotfy K., Daoud I. and Belaidi H. Molecular Docking Studies and ADMET Properties of New 1,2,3-Triazole Derivatives for Anti-Breast Cancer Activity. *J. Bionanosci.* 2018; 12: 26-36. DOI: <https://doi.org/10.1166/jbns.2018.1505>
- [31] Dermeche K., Tchouar N., Belaidi S., Salah T. Qualitative Structure-Activity Relationships and 2D-QSAR Modeling of TNF- $\alpha$  Inhibition by Thalidomide Derivatives. *J. Bionanosci.* 2015; 9: 395-400. DOI: <https://doi.org/10.1166/jbns.2015.1320>
- [32] Almi Z., Belaidi S., Segueni L., [Structural Exploration and Quantitative Structure-Activity Relationships Properties for 1,2,5-Oxadiazole Derivatives](#), *Rev. Theo. Sci.* 2015; 3: 264-272
- [33] Weaver S. and Gleeson M.P. The importance of the domain of applicability in QSAR modeling. *J.Mol. Graph. Model.* 2008; 26:1315–1326. <https://doi.org/10.1016/j.jmgm.2008.01.002>
- [34] Chtita S., Belhassan A., Bakhouch M., Taourati A.I., Aouidate A., Belaidi S., Moutaabbid M., Belaouad S., Bouachrine M. and Lakhlifi T. QSAR study of unsymmetrical aromatic Disulfides as potent avian SARS-CoV main protease inhibitors using quantum chemical descriptors and statistical methods. *Chemometr. Intell. Lab. Syst.* 2021; 210:104266. <https://doi.org/10.1016/j.chemolab.2021.104266>
- [35] Chtita S., Ghamali M., Ousaa A., Aouidate A., Belhassan A., Taourati A. I., Masand V. H., Bouachrine M. and Lakhlifi T. QSAR study of anti-Human African Trypanosomiasis activity for 2-phenylimidazopyridines derivatives using DFT and Lipinski's descriptors. *Heliyon*, 2019; 5:01304.
- [36] Al-Shar'i N.A., Hassan M.A., Al-Barqi H.M., Al-Balas Q.A. and El-Elimat T. Discovery of Novel Glyoxalase-I Inhibitors Using Computational Fragment-Based Drug Design Approach. *Jordan Journal of Pharmaceutical Sciences.* 2020; 13:225-245
- [37] Ouassaf M., Belaidi S., AlMogren M.M., Chtita S., UllahKhan S. and ThetHtar T. Combined docking methods and molecular dynamics to identify effective antiviral 2, 5-diaminobenzophenonederivatives against SARS-CoV-2. *J. King Saud Univ. Sci.* 2021; 33:101352. <https://doi.org/10.1016/j.jksus.2021.101352>
- [38] Hast M.A., Fletcher S., Cummings C.G., Pusateri E. E., Blaskovich., M. A, Rivas K., Gelb M.H., Van Voorhis W. C., Sebti S. M., Hamilton A D. and Beese L. S. *Chem. Biol.* 2009;16:181-192
- [39] Ouassaf M., Belaidi S., Khamouli S., Belaidi H. and Chtita S. Combined 3D-QSAR and Molecular Docking Analysis of Thienopyrimidine Derivatives as Staphylococcus aureus Inhibitors. *Acta Chim. Slov.* 2021; 68:289–303. <https://doi.org/10.17344/acsi.2020.5985>
- [40] Cherkasov A., Muratov E.N., Fourches D., Varnek A., Baskin I.I. and Cronin M. QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* 2014; 57:4977–5010. <https://doi.org/10.1021/jm4004285>
- [41] Timmerman H., Mannhold R., Krogsgaard LP, *Chemometric methods in molecular design*, John Wiley & Sons, Hoboken, 2008
- [42] Bakdash J.Z. and Marusich L.R. Repeated Measures Correlation. *Front Psychol.* 2017; 8:456. <https://doi.org/10.3389/fpsyg.2017.00456>
- [43] Akinwande M.O., Dikko H.G. and Samson A. Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open J. Stat.* 2015; 5:754–67. <https://doi.org/10.4236/ojs.2015.57075>
- [44] Kier L.B. and Hall L.H. An Electrotopological-State Index for Atoms in Molecules. *Pharm. Res.* 1990; 7:801–807. <https://doi.org/10.1023/A:1015952613760>
- [45] Galvez J., Garcia-Domenech R., De Julian-Ortiz J.V. and Soler R. Topological Approach to Drug Design. *J. Chem. Inf. Comput. Sci.* 1995; 35:272-284 <https://doi.org/10.1021/ci00024a017>
- [46] Yang Y., Engkvist O., Llinàs A. and Chen H. Beyond Size, Ionization State, and Lipophilicity: Influence of Molecular Topology on Absorption, Distribution, Metabolism, Excretion, and Toxicity for Druglike



- Compounds. *J. Med. Chem.* 2012; 26; 55:3667–77.  
<https://doi.org/10.1021/jm201548z>
- [47] Mitroy J., Safronova M.S. and Clark C.W. Theory and applications of atomic and ionic polarizabilities. *J. Phys; B: At. Mol. Opt. Phys.* 2010; 43: 202001.  
<https://doi.org/10.1088/0953-4075/43/20/202001>
- [48] Martin Y.C., *Quantitative Drug Design: A Critical Introduction*, Second Edition, CRC Press, 2010, [Boca Raton, Florida, USA](#)
- [49] Arnott J.A. and Planey S.L. The influence of lipophilicity in drug discovery and design. *Expert. Opin. Drug Discov.* 2012;7: 863–75.  
<https://doi.org/10.1517/17460441.2012.714363>
- [50] Jalali-Heravi M. and Konuze E. Use of quantitative structure property relationships in predicting the Kraft point of anionic surfactants, *Elec. J. Mol. Des.* 2002; 1:410–417.
- [51] Roy K., Mitra I., Kar S., Ojha P. K., Das R. N., and Kabir H. *J. Chem. Info. and Mod.* 2012 ;52: 396-408. DOI: 10.1021/ci200520g
- [52] Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Info.* 2010; 29:476–88.  
<https://doi.org/10.1002/minf.201000061>
- [53] Roy K., Kar S. and Ambure P. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Syst.* 2015; 145:22–9.  
<https://doi.org/10.1016/J.CHEMOLAB.2015.04.013>
- [54] Kalirajan R., Gowramma B., Gomathi S. and Vadivelan R. Activity of Isoxazole substituted 9-aminoacridines against SARS CoV-2 main protease for COVID19: A computational approach. *Jordan Journal of Pharmaceutical Sciences.* 2021; 14:403-416.

## تحقيقات التركيب الكمي للنشاط (QSAR) وتحليل الالتحام الجزيئي لمثبطات بروتين البلازموديوم فرنازيل ترانسفيراز كعوامل قوية مضادة للملاريا

مباركة وصاف<sup>1</sup>، صلاح بلعدي<sup>1</sup>، آمنة شتيوي<sup>2</sup>، سمير اشطيطة<sup>3\*</sup>

<sup>1</sup>جامعة بسكرة، مجموعة الكيمياء الحاسوبية والطبية، مختبر LMCE، الجزائر

<sup>2</sup>كلية الصيدلة، جامعة الشرق الأوسط، الأردن.

<sup>3</sup>مختبر الكيمياء التحليلية والجزيئية، كلية العلوم بن مسيك، جامعة الحسن الثاني بالدار البيضاء، المغرب.

### ملخص

يعتبر تطوير مثبطات farnesyltransferase على أساس سقالة بنزوفينون الموجهة ضد *falciparum Plasmodium* استراتيجية في علاج الملاريا. في هذا العمل، تم إجراء علاقة التركيب الكمي بالنشاط (QSAR) للتنبؤ بالأنشطة المثبطة للبروتين (PFT) farnesyltransferase لسلسلة من 36 مشتقاً من مشتقات بنزوفينون. تم تقسيم مجموعة البيانات إلى مجموعتين فرعيتين من مجموعات التدريب والاختبار، وأفضل نموذج باستخدام الانحدار الخطي المتعدد (MLR)، مع قيم الصلاحية الداخلية والخارجية  $R^2 = 0.884$ ،  $R^2_{adj} = 0.865$ ،  $R^2_{pred} = 0.821$ ،  $R^2_{cv} = 0.822$  و 0.811 بالاتفاق مع معايير Tropsha و Golbraikh. تم تحديد مجال التطبيق (AD) باستخدام مخطط ويليامز لوصف الفضاء الكيميائي للنموذج المستخدم في هذه الدراسة. يوضح النموذج أن الأنشطة المضادة للملاريا للبنزوفينون تعتمد على واصفات  $\log P$  و  $\text{bpol}$  و  $\text{MAXDn}$  و  $\text{FMF}$ . دفعنا هذه المؤشرات إلى تصميم مثبطات جديدة للبنزوفينونات PFT والتنبؤ بقيمة أنشطتها المضادة للملاريا بناءً على معادلة MLR. تكشف نتائج الإرساء أن البنزوفينونات المصممة حديثاً ترتبط بالجيب الكارهة للماء والتلامس القطبي مع التقارب العالي. يمكن أن تساعد النتائج المتوقعة من هذه الدراسة في تصميم بنزوفينونات جديدة كمثبطات لـ PFT البشري مع أنشطة مضادة للملاريا عالية.

الكلمات الدالة: QSAR، الالتحام، بنزوفينون، مثبط PFT، مضاد للملاريا.

\* المؤلف المراسل: سمير اشطيطة

[samirchtita@gmail.com](mailto:samirchtita@gmail.com)

تاريخ استلام البحث 2021/9/17 وتاريخ قبوله للنشر 2022/1/12.