# QSAR, molecular docking and ADMET studies of Quinoline, Isoquinoline and Quinazoline derivatives against Plasmodium falciparum malaria

**EL Rhabori Said** ( said.elrhabori@usmba.ac.ma )
USMBA FSTF: Universite Sidi Mohamed Ben Abdellah Faculte des Sciences et Techniques de Fes
https://orcid.org/0000-0003-1468-950X

**El Aissouq Abdellah**
USMBA FSTF: Universite Sidi Mohamed Ben Abdellah Faculte des Sciences et Techniques de Fes

**Chtita Samir**
Universite Hassan II Casablanca Faculte des Sciences Ben M'Sik

**Khalil Fouad**
USMBA FSTF: Universite Sidi Mohamed Ben Abdellah Faculte des Sciences et Techniques de Fes

**Research Article**

**Keywords:** QSAR, DFT, Plasmodium falciparum, molecular docking, ADMET

**Posted Date:** February 15th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1318934/v1

# QSAR, molecular docking and ADMET studies of Quinoline, Isoquinoline and Quinazoline derivatives against Plasmodium falciparum malaria

El Rhabori Said [1*], El Aissouq Abdellah [1†], Chtita Samir [2†] and Khalil Fouad [1†]

[1*] Laboratory of Processes, Materials and Environment (LPME), Sidi Mohamed Ben Abdellah University, Faculty of Science and Technology - Fez, Morocco
[2] Laboratory of Analytical and Molecular Chemistry, Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, B.P. 7955, Casablanca, Morocco.

[*] Corresponding Author: E-mail: said.elrhabori@usmba.ac.ma
Contributing authors: abdellah.elaissouq@usmba.ac.ma; samirchtita@gmail.com; fouad.khalil@usmba.ac.ma
[†]These authors contributed equally to this work.

## Abstract

With the aim of researching new antimalarial drugs, a series of quinoline, isoquinoline and quinazoline derivatives were studied against the Plasmodium falciparum CQ-sensitive and MQ-resistant strain 3D7 protozoan parasite. DFT with B3LYP functional and 6-311G basis set was used to calculate quantum chemical descriptors for QSAR models. The molecular mechanics (MM2) method was used to calculate constitutional, physicochemical, and topological descriptors. By randomly dividing the dataset into training and test sets, we were able to construct reliable models using linear regression (MLR), nonlinear regression (MNLR) and artificial neural networks (ANN). The determination coefficient values indicate the predictive quality of the established models. The robustness and predictive power of the generated models were also confirmed via internal validation, external validation, the Y-Randomization test and the applicability domain. Furthermore, molecular docking studies were conducted to identify the key interactions between the studied molecules and the PfPMT receptor's active site.

The findings of this contribution study indicate that the antimalarial activity of these compounds against Plasmodium falciparum appears to be largely determined by four descriptors, i.e., Total Connectivity (Tcon), percentage of carbon (C (%)), density (D) and bond length between the two nitrogen atoms (Bond N-N).

On the basis of the reliable QSAR model and molecular docking results, several new antimalarial compounds have been designed. The selection of drug-candidates was performed according to drug-likeness and ADMET parameters.

**Keywords:** QSAR, DFT, Plasmodium falciparum, molecular docking, ADMET.

## 1 Introduction

Despite the high cost assigned to it, malaria remains one of the major diseases affecting the most vulnerable populations, with more than 500.000 deaths per year worldwide [1–3]. Among the various Plasmodium parasite species, Plasmodium falciparum is the most prevalent. The absence of a safe and effective vaccine and the emergence of antimalarial drug resistance are significant public health concerns. To ascertain the primary factors underlying malaria resistance, efforts must be directed toward developing novel inhibitor compounds capable of

specifically targeting and inhibiting plasmodium parasite growth and transmission [4–7]. Phosphatidylcholine (PC) is a key phospholipid component of parasite membranes and is required for parasite growth, proliferation, and survival. As a result, it is believed that phospholipid biosynthesis pathways represent promising targets for antimalarial drug development. In the case of malaria, the Parasites require active PC generation via a different mechanism, the Serine Decarboxylation Pathway (SDPM), in order to rapidly proliferate intraerythrocytic and gametocytes [8]. The unusual synthesis of phosphocholine from ethanolamine is catalyzed by phosphoethanolamine methyl transferase (PMT). Thus, inhibiting the catalytic domain of PfPMT can completely shut down the SDPM pathway. Apart from the fact that there are no human orthologues (mammalian absence), PfPMT may be used as a target template for the development of novel antimalarial drugs [9]. As a result, the crystal structure of PfPMT 3D7 (PDB ID: 3UJ9) can be used to conduct molecular docking experiments [10].

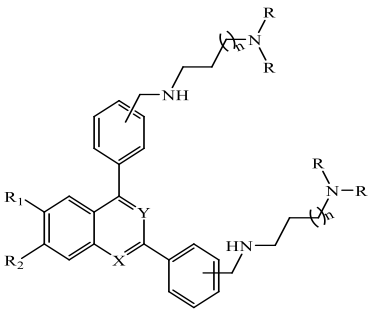
In the context of developing future strategies for plasmodium falciparum resistance prevention, the use of 2D-QSAR, molecular docking and ADMET studies have become critical methods [11–19]. This objective is accomplished by performing molecular modeling on fifty-five quinoline, isoquinoline and quinazoline derivatives synthesized by Jean Guillon et al [20]. To do so, we created MLR, MNLR and ANN models with fourteen descriptors and attempted to interpret the propriety of these molecules using multivariate statistical analyses [21].

To optimize the biological activity of the molecules studied via molecular docking, the most active and least active molecules were chosen as the key to discovering new antimalarial drug candidates. As a result, numerous molecules with potential biological benefits have been proposed. The leverage values effect hi and the h* threshold were also calculated in order to select designed molecules within the applicability domain. Then, based on their pharmacokinetic properties, the molecules of the hits drug candidates were chosen. Finally, we confirmed the interactions of the newly designed compounds with the PfPMT receptor active site using the validated molecular docking method [22].

## 2 Material and methods

### 2.1 Dataset and molecular descriptors

The experimental biological activities of 2,4-bis[(substituted-aminomethyl) phenyl] quinoline,1,3-bis[(substituted-aminomethyl) phenyl] isoquinoline, and 2,4-bis[(substituted-aminomethyl) phenyl] quinazoline derivatives were used as independent variables in order to perform the molecular modeling (refer Figure1 and Table1) [20].




X and Y: N, CH; $R_1$ and $R_2$: -H, -$OCH_3$; n =0, 1, 2; Chain's position (p) =3, 4

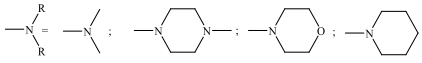
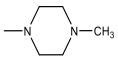
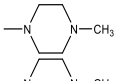
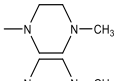
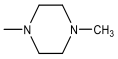
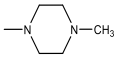
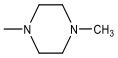
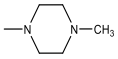
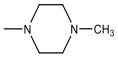
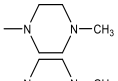
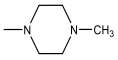
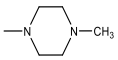
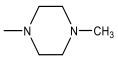
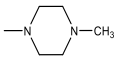
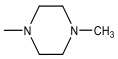
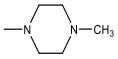


**Fig .1** Structures of quinoline, isoquinoline and quinazoline derivatives

**Table 1** Structures of 54 quinoline, isoquinoline and quinazoline derivatives and their activities

| N° | X | Y | $R_1$ | $R_2$ | $-NR_2$ | p | n | $pIC_{50}$ | N° | X | Y | $R_1$ | $R_2$ | $-NR_2$ | p | n | $pIC_{50}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1a** | N | CH | H | H |  | 4 | 2 | 5.389 | **2h** | CH | N | $CH_3O$ | H |  | 4 | 1 | 5.742 |
| **1b** | N | CH | H | H |  | 4 | 1 | 6.328 | **2i** | CH | N | H | $CH_3O$ |  | 4 | 2 | 5.073 |
| **1c** | N | CH | H | H |  | 4 | 0 | 7.495 | **2j** | CH | N | H | $CH_3O$ |  | 4 | 1 | 5.142 |
| **1d** | N | CH | H | H |  | 4 | 2 | 6.161 | **2k** | CH | N | H | $CH_3O$ |  | 4 | 2 | 5.050 |
| **1e** | N | CH | H | H |  | 4 | 1 | 6.328 | **2l** | CH | N | H | $CH_3O$ |  | 4 | 1 | 5.015 |
| **1f** | N | CH | H | H |  | 4 | 0 | 5.616 | **3a** | N | N | H | H |  | 4 | 2 | 5.668 |
| **1g** | N | CH | H | H |  | 4 | 1 | 7.004 | **3b** | N | N | H | H |  | 4 | 1 | 6.208 |
| **1h** | N | CH | H | H |  | 4 | 0 | 5.873 | **3c** | N | N | H | H |  | 4 | 0 | 5.975 |
| **1i** | N | CH | $CH_3O$ | H |  | 4 | 2 | 5.510 | **3d** | N | N | H | H |  | 4 | 2 | 5.629 |
| **1j** | N | CH | $CH_3O$ | H |  | 4 | 1 | 5.577 | **3e** | N | N | H | H |  | 4 | 1 | 5.818 |
| **1k** | N | CH | $CH_3O$ | H |  | 4 | 2 | 5.587 | **3f** | N | N | H | H |  | 4 | 1 | 6.377 |
| **1l** | N | CH | $CH_3O$ | $CH_3O$ |  | 4 | 1 | 5.745 | **3g** | N | N | H | H |  | 4 | 0 | 6.824 |
| **1m** | N | CH | H | $CH_3O$ |  | 4 | 2 | 5.047 | **3h** | N | N | $CH_3O$ | H |  | 4 | 2 | 5.416 |
| **1n** | N | CH | H | $CH_3O$ |  | 4 | 1 | 5.498 | **3i** | N | N | $CH_3O$ | H |  | 4 | 1 | 5.393 |
| **1o** | N | CH | H | $CH_3O$ |  | 4 | 2 | 5.618 | **3j** | N | N | $CH_3O$ | H |  | 4 | 0 | 6.252 |
| **1p** | N | CH | H | H |  | 4 | 1 | 5.363 | **3k** | N | N | $CH_3O$ | H |  | 4 | 1 | 5.506 |
| **1q** | N | CH | H | H |  | 3 | 1 | 5.842 | **3l** | N | N | $CH_3O$ | H |  | 4 | 0 | 5.742 |
| **1r** | N | CH | H | H |  | 3 | 1 | 5.710 | **3m** | N | N | H | $CH_3O$ |  | 4 | 2 | 5.928 |
| **1s** | N | CH | H | H |  | 3 | 1 | 6.678 | **3n** | N | N | H | $CH_3O$ |  | 4 | 1 | 5.599 |
| **1t** | N | CH | H | H |  | 3 | 0 | 6.638 | **3o** | N | N | H | $CH_3O$ |  | 4 | 0 | 6.215 |
| **2a** | CH | N | H | H |  | 4 | 2 | 5.759 | **3p** | N | N | H | $CH_3O$ |  | 4 | 2 | 5.183 |
| **2b** | CH | N | H | H |  | 4 | 1 | 5.740 | **3q** | N | N | H | $CH_3O$ |  | 4 | 1 | 4.982 |
| **2c** | CH | N | H | H |  | 4 | 2 | 5.708 | **3r** | N | N | H | $CH_3O$ |  | 4 | 0 | 5.363 |
| **2d** | CH | N | H | H |  | 4 | 1 | 5.239 | **3s** | N | N | H | H |  | 3 | 1 | 5.947 |
| **2e** | CH | N | $CH_3O$ | H |  | 4 | 2 | 5.356 | **3t** | N | N | H | H |  | 3 | 1 | 5.650 |
| **2f** | CH | N | $CH_3O$ | H |  | 4 | 1 | 5.833 | **3u** | N | N | H | H |  | 3 | 1 | 6.367 |
| **2g** | CH | N | $CH_3O$ | H |  | 4 | 2 | 5.485 | **3v** | N | N | H | H |  | 3 | 0 | 6.444 |

Fourteen molecular descriptors were calculated to create QSAR models, some of which were linear and some of which were non-linear, and to identify spatial regions associated with the activity of the compounds under study.

After optimizing the energy of each compound using the MM2 method (force field method with gradient for root mean square (RMS) of 0.01 kcal/mol) [23], topological, physicochemical and geometrical descriptors were calculated using the Chem3D V16 [24], ChemSketch12 [25], and Marvin Sketch [26] software packages. Then, using the Gaussian 09 program [27], we calculated quantum chemical descriptors by employing the DFT approach with the B3LYP functional and the 6-311G basis (refer table 2 and table S1).

**Table 2** Software used and descriptors calculated

| **Software** | **Descriptors calculated** |
|---|---|
| **Chemoffice3D** | Number of HBond Acceptors (**NHA**), Number of HBond Donors (**NHD**), the octanol–water partition coefficient (**LogP**), Number of Rotatable Bonds (**NRB**). |
| **ChemSketch** | Percent ratio of atoms (**C %, H %, N %** and **O %**) |
| **Gaussian 09** | Dipolar moment (**DM**), energy of highest occupied molecular orbital ($\mathbf{E_{HOMO}}$), energy of the lowest unoccupied molecular orbital ($\mathbf{E_{LUMO}}$), electrophilicity index (**ω**), bond length between two atoms of nitrogen inside chain (**Bond** $_{\mathbf{NN}}$), Angle between benzene carbon and nitrogen (**Angle Bn-C-N**), dihedral angle between benzene ring and carbon before nitrogen and nitrogen and carbon after nitrogen (DA) |

These descriptors served as input data for establishing a quantitative relationship between them and antimalarial activity using MLR, MNLR and ANN methods as well as statistical analysis.

## 2.2 Principal component analysis and data splits

Principal component analysis (PCA) is a statistical technique that uses descriptive statistics. In this study, the PCA was used to extract as much information from the database as possible and to identify the various chemical descriptors that will help build the QSAR models [28]. Following this, the database was split into training and test sets comprising 80% and 20% of the total data, respectively [29]. This approach is carried out using the k-means classification technique (implemented in the XLSTAT software) [30]. Following this division, a randomly chosen compound from each cluster is included in the test set.

## 2.3 Models' development and validation

We employed multiple linear regression (MLR) [31], multiple nonlinear regression (MNLR) and artificial neural network (ANN) approaches to construct QSAR models [32]. The MNLR and MLR models were created using the XLSTAT V. 2019 software [33], whereas the ANN model was created using the MatlabV.2015 software [34]. The principal parameters utilized in statistical analysis of QSAR models are the determination coefficient ($R^2$) (Eq. 1), adjusted coefficient ($R^2_{adj}$) (Eq. 2), mean squared error (MSE) (Eq. 3), Fisher's statistical parameter (F-value) and level of significance (p-value) [35–37].

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_{iobs} - Y_{ical})^2}{\sum_{i=1}^{n} (Y_{iobs} - \bar{Y})^2} \qquad \text{(Eq. 1)}$$

$$R^2_{adj} = \frac{(n-1)\times R^2 - P}{n-1-p} \quad \text{(Eq. 2)}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_{iobs} - Y_{cal})^2 \quad \text{(Eq. 3)}$$

With; $Y_{iobs}$ is the value of the observed response i, $Y_{cal}$ is the value of the predicted response i, $\bar{Y}_{cal}$ is the average value of predicted responses, p is the number of explicative variables in the model and n is the number of molecules in the training set.

Internal validation was conducted to ensure the developed models' validity. In this step, the validation procedure named leave-one-out cross-validation (LOO-CV) and based on the calculation of the $R^2_{cv}$coefficient value was used to validate the robustness of developed models.

According to following equation (Eq.4), the value of $R^2_{cv}$ should be more than 0.5 [38-41].

$$R^2_{cv} = 1 - \frac{\sum_{j=1}^{N}(Y_{jobs}(train) - Y_{jcal}(train))^2}{\sum_{j=1}^{N}(Y_{jobs}(train) - \bar{Y}_{cal}(train))^2} \quad \text{(Eq. 4)}$$

Where;

$Y^2_{jobs}$(train) is the value of the observed response;

$Y_{jcal}$(train) is the value of the response predicted by Loo-cv;

$\bar{Y}_{cal}$ (train) is the mean value of the predicted responses.

The predictive power of generated models is determined by calculating the $R^2_{test}$coefficient between the observed and projected pIC50 values for the test set [41].

The Y-randomization test was used to exclude the possibility of random association between selected descriptors and their associated activities in the original model [42]. To ensure that the model was not obtained by chance via the Y-randomization test, the average random correlation coefficient ($R^2_r$) of the randomly constructed models must be less than the correlation coefficient ($R^2$) of the original non-random model [43].

## 2.4 Applicability domain (AD)

The applicability domain is defined as a space including molecules with accurately predicted activities; because the model is based on a limited number of compounds, it does not encompass the complete chemical space [44,45].

Among the several ways for defining AD models, the most widely utilized is determining the leverage values effect hi for each compound (Eq. 5) [46]. If the leverage effect hi of a compound exceeds the alert leverage h* (Eq. 6), the compound is regarded to be outside the application domain [47].

$$h_i = x_i^T (X^T X)^{-1} x_i \quad \text{(Eq. 5)}$$

$$h^* = 3\times\frac{(K+1)}{n} \quad \text{(Eq. 6)}$$

Where i=1, 2,..., n, xi is a vector descriptive of the compound to be discovered and X is a matrix containing the model's k descriptor values for the n compounds in the training set.

## 2.5 Drug likeness and pharmacokinetics ADMET prediction

To find possible drug candidates, the created compounds' drug similarity is evaluated using Lipinski rule-based filters, synthetic accessibility and an in-silico analysis of adsorption, distribution, metabolism and excretion. The toxicity levels of the compounds under investigation were determined mostly during the drug research phase [48–50].

In this work, the drug-likeness and ADMET of the selected compounds are evaluated using the online SwissADME [51] and pkCSM [52-53] servers, respectively.

### 2.6 Molecular docking

The PfPMT receptor is required for the catalytic production of PC. The amino acid His132 acts as a general base in the reaction mechanism of this receptor (Fig. S1), abstracting a proton from the hydroxyl group of Tyr-19 and activating the residue [10]. As a result, the crystal structure of PfPMT3D7 can be employed in molecular docking to deduce the critical structural requirements for anti-plasmodium falciparum activity.

To conduct molecular docking analysis in this study, the AutoDock Vina tool was used. The Protein Data Bank has revealed the X-ray crystal structure of PfPMT3D7 (code PDB: 9UJ9). The Discovery Studio software [54] was used to visualize existing interactions in order to gain insight into the activity's critical structural requirements.

After removing the complex's original ligand and water molecules, the 3D grid maps were constructed using the AUTOGRID technique. The AutoDock software MGLTools 1.5.6 packages [55] was used to prepare the receptor and the original ligand (Phosphocholine) for re-docking with the PfPMT receptor, as well as for active site identification using the same 3D grid [56]. To guarantee that the docking technique is acceptable and valid, the root mean square deviation (RMSD) range must not exceed 2 Å [57].

## 3 Results and discussion

### 3.1 Dataset collection and selected descriptors

PCA method was applied using fourteen descriptors calculated for each quinoline, isoquinoline and quinazoline derivatives (refer table S2). The descriptors with the lowest correlation coefficients are chosen and assigned to each of the 54 molecules analyzed in the form of a 13-column, 43-row matrix (refer table S3). The acquired database is then partitioned into training and test sets using the K-means technique as follows: The test set contains eleven molecules (1h-1i-1i-1n-1p-1q-1p-3r-3v), while the training set contains forty-three molecules (refer table S4).

### 3.2 Models' development

Several models may be created utilizing specified chemical descriptors and experimental values of antimalarial activity. Any model that did not meet the Organization for Economic Cooperation and Development's principles, as well as Golbraikh and Tropsha's criteria, was discarded [58–60]. The equation for the resultant model produced using the MLR technique is provided below (Eq.7):

$$pIC50 = -39.269 + 598449\ \mathbf{Tcon} + 0.249\ \mathbf{C\ (\%)} + 24{,}603\ \mathbf{D} - 0.172\mathbf{Bond\ NN} \qquad \text{(Eq.7)}$$

Because the model's p-value is less than 0.0001, we would take a risk of less than 0.01 percent assuming that H0 is incorrect and the model equation is statistically significant at a level larger than 95 percent.

The variance inflation factor (VIF), defined as VIF= 1/ (1-$r^2$), was applied to selected model with descriptors' VIF values less than 5 (refer table 3) [61].

**Table 3** VIF values of the models' descriptors

| Model | T Con | C (%) | D | Bond NN |
|---|---|---|---|---|
| **VIF** | 1.235 | 2.287 | 2.249 | 1.136 |

According to this table, all descriptors chosen for the created model have VIF less than 5. This shows that there is no multicollinearity among the descriptors chosen and that the resulting model is stable.

The MLR model was tested for robustness by doing 100 random trials with randomized training set compound activity [62,63]. This technique generated 100 new models with new $R_r$ ,$R_r^2$ and $Q_r^2$ values. Then, a comparison was done between the findings of the created model and the acceptable limit of Golbraikh and Tropsha's threshold values in order to corroborate the validation model. Table 4 explains how the optimal model's reliability and acceptability are determined [50].

**Table 4** Comparison of the MLR model's statistical parameters with Golbraikh and Tropsha's criteria

| Model's Parameter | Scor | Golbraikh / Tropsha's Threshold | Validation |
|---|---|---|---|
| $R^2$ | 0.734 | Greater than 0.600 | Passed |
| $R_{adj}^2$ | 0.706 | Greater than 0.600 | Passed |
| $R_{test}^2$ | 0.664 | Greater than 0.600 | Passed |
| $Q_{CV}^2$ | 0.676 | Greater than 0.500 | Passed |
| MSE | 0.096 | A low value | Passed |
| $F_{test}$ | 26.271 | A high value | Passed |
| $R_r^2$ | 0.099 | Less than $R^2$ | Passed |
| $Q_{cv\ LOO}^2$ (Rand) | -0.173 | Less than $Q^2$ cv | Passed |
| $cR_p^2$ | 0.688 | Greater than 0.500 | Passed |

The MLR model's coefficient of determination ($R^2$= 0.734) indicates strong activity-descriptor relationship efficiency (73%). Because the constructed model has excellent descriptive capacity to descriptors, the high adjusted coefficient of determination ($R_{adj}^2$) shows the genuine impact of employed descriptors on the analyzed pIC50. The cross-validated square correlation coefficient ($Q_{cv}^2$= 0.676) indicates that this model has a high degree of internal predictive capacity. The large value of $R_{test}^2$ ($R_{test}^2$= 0.664) indicates that the created model has a high degree of predictive capacity for the novel compounds. Additionally, the VIF, $R_r^2, Q_r^2$ and MSE values demonstrate that this ideal model may be used to forecast pIC50 values for novel quinoline, isoquinoline and quinazoline compounds. The correlation diagram with calculated versus experimental pIC50 values of the MLR model of training and test sets is shown in the figure 2.

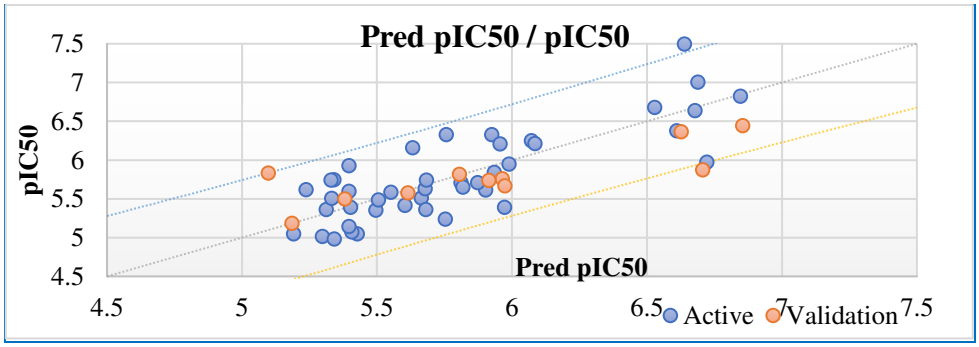



**Fig. 2** The correlation diagram of the calculated pIC50 versus experimental pIC50 of the generated model of training and test sets.

Observing the distribution of observed and predicted pIC50 values, we can see that the two values are significantly correlated (refer figure 2).

With the four molecular descriptors (TCon, C percent, D and BondNN) as input parameters, MNLR and ANN models are developed to improve the relationship between predicted activities and these pertinent descriptors. The equation below (Eq.8) illustrates the nonlinear model obtained via MNLR:

pIC50 = -68.898 + 304539 Tcon - 91.665 D -1.435 $\boldsymbol{Bond_{N-N}}$ – 0.018 $\boldsymbol{C\%^2}$+ 52.004 $\boldsymbol{D^2}$ + 0.169 $\boldsymbol{Bond_{N-N}^2}$ (Eq.8)

Where: $R^2$ =0.701, MSE=0.121 and $R_{test}^2$=0.690

When comparing experimental and predicted values of pIC50, the MNLR technique has a more consistent distribution of values (refer figure 3).

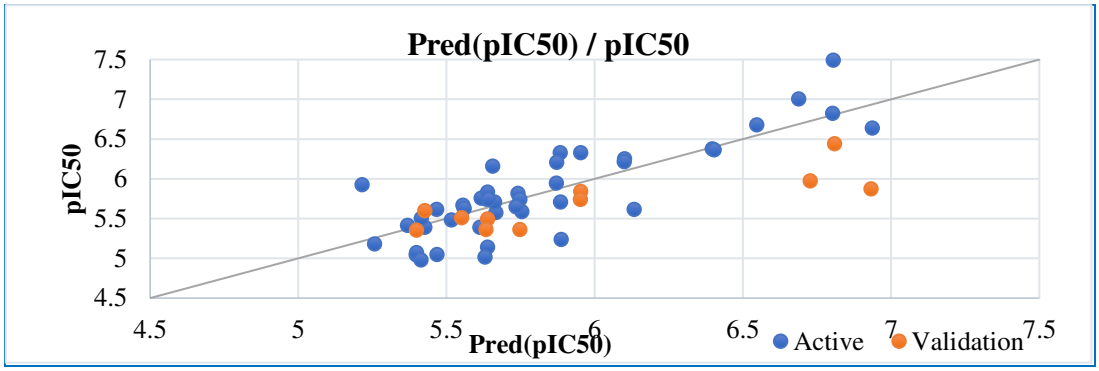



**Fig.3** Experimental pIC50 versus calculated pIC50 with the MNLR model.

To determine whether the descriptors chosen were effective at predicting pIC50 values, an ANN technique with the 4-4-1 architecture was used. The value 1.520 of the parameter ρ with $1 < \rho < 3$ indicate that the number 4 in the hidden layer is proportional to the number of descriptors in the input layer in order to predict the pIC50 values expressed as 1 in the output layer.

With a high value of the determination coefficient ($R^2$ = 0.740), a low value of the mean square error (MSE = 0.087) and a high value of the test-validation coefficient ($R_{test}^2$ = 0.865), the ANN model was found to be highly effective in predicting the antimalarial activity of the investigated molecules. A similar even distribution of candidate pIC50 values was observed across the training and test sets, as illustrated in Figure 4. With this distribution, it is ensured that the predicted values of pIC50 are extremely close to the values observed experimentally.

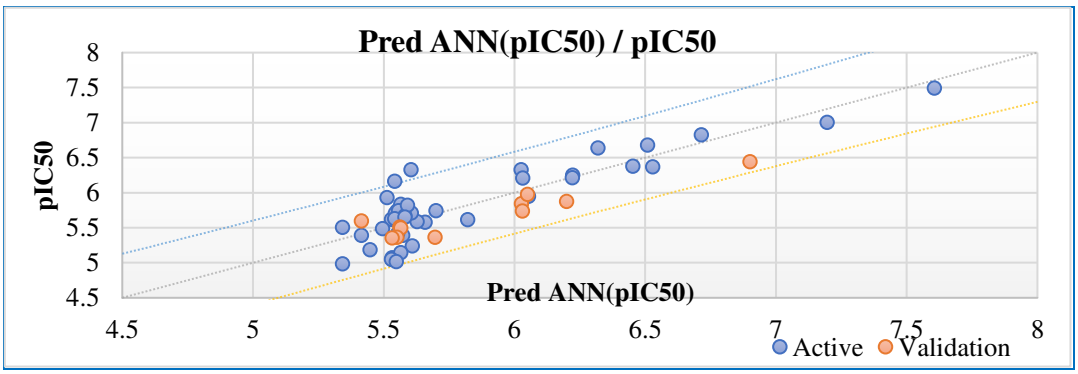



**Fig.4** Correlations between observed and the predicted activity using the ANN model.

MLR, MNLR and ANN models all exhibit statistical significance (refer table S6). As a result, these models can be used to forecast the biological activity of previously unknown drug candidates.

## 3.3 Applicability domain (AD)

To determine the appropriate application of the MLR model, researchers examined the relationship between residual value and leverage effect. The leverage effect threshold value h*, with h* = 3*(k+1) /n); k = 4; n = 43 and the distribution of normalized residual values and leverage level values were calculated (refer figure 5) [49]. From this diagram, compound which has hi>h*, with h* =0.348, or with standardized residual greater than y = ± 3 is considered outside of the AD.

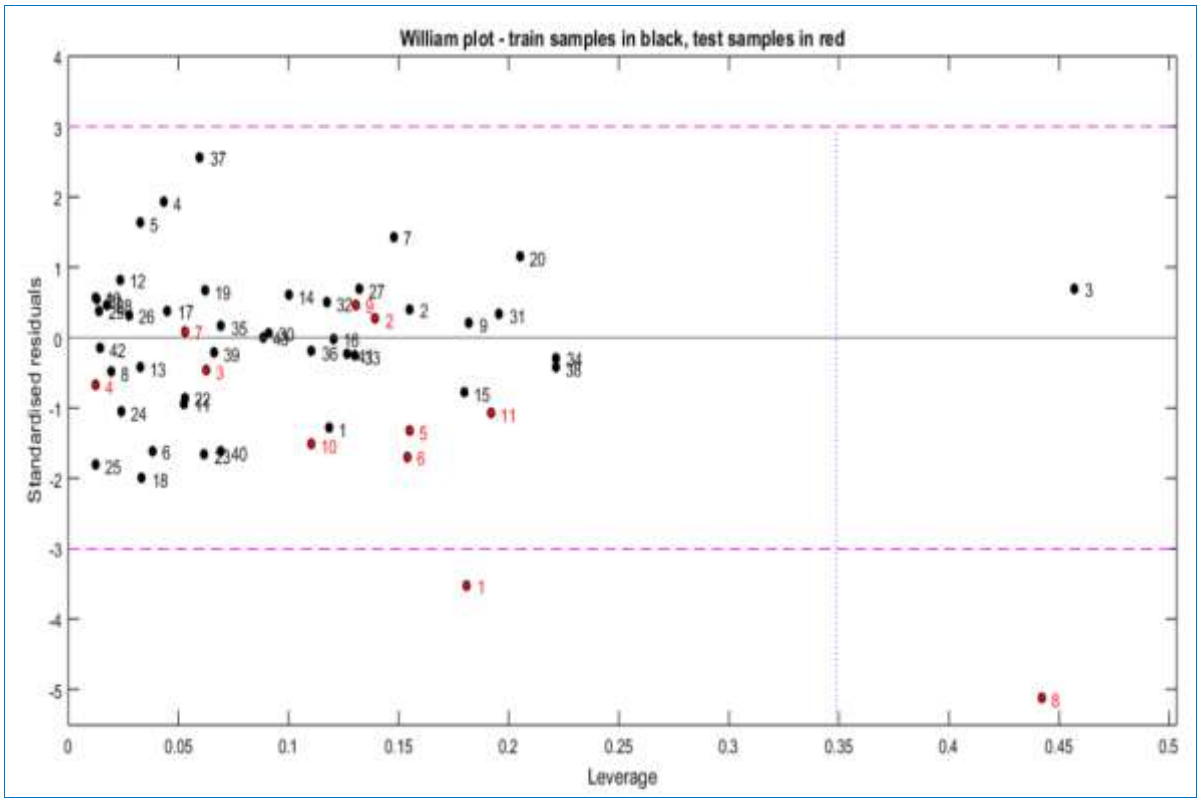



**Fig.5** Applicability domain obtained by the William's plot of best MLN model (h*=0.348)

The MLR model predicts correct and valid pIC50 values for compounds that fall within the applicability domain, which is located to the left of the leverage threshold h*=0.348, but not for molecules that fall outside of the applicability domain. As seen in this figure, one molecule 1 (1h) belonging to the test set has a standard deviation outside the y range (y = -3), while two molecules, one belonging to the training set's molecule 3 (1c) and the other to the test set's molecule 8 (3c), are outside of the domain of applicability. It is possible that the activity of these compounds was underestimated due to the lack of experimental data for these molecules. As a result, these molecules should be removed from the list of molecules and excluded from molecular modeling in the future study, which will focus exclusively on molecules within the applicability domain.

## 3.4 Molecular docking of most and less active molecules

In order to optimize the biological activity, molecular docking was used to compare the key interactions between the most active molecule 1c and the least active molecule 3q with the PfPMT protein receptor.

For this, the first step is to figure out where the co-crystallized ligand phosphocholine interacted with the PfPMT receptor pocket. Figure 6 shows the active site (ASP10, PHE13, LEU14, ASN17, GLN18, TYR19, THR20, GLU22, GLY23, VAL26, TYR27, PHE31, ASN34, TYR35,ILE62, GLY63, SER64, GLY67, GLY68, GLY69, GLY70, ILE84, ILE85, ILE86, ILE87, ILE90, PHE106, ASN109, ASP110, ILE111, LEU112, TYR125, SER126, ARG127, ASP127, ASP128, ASP129, ILE130, LEU131, HIS132,LEU133, SER134, ASN137, TYR160, GLU171, GLU174, TYR175, GLN178, ARG179, LYS180, TYR181, THR182, PHE228, LYS236, SER239, LEU240, ASP242, GLY243, TRP244, ARG246, LYS247 and GLN256) by visualizing the structure of the 3UJ9 crystal complex using MOE software.

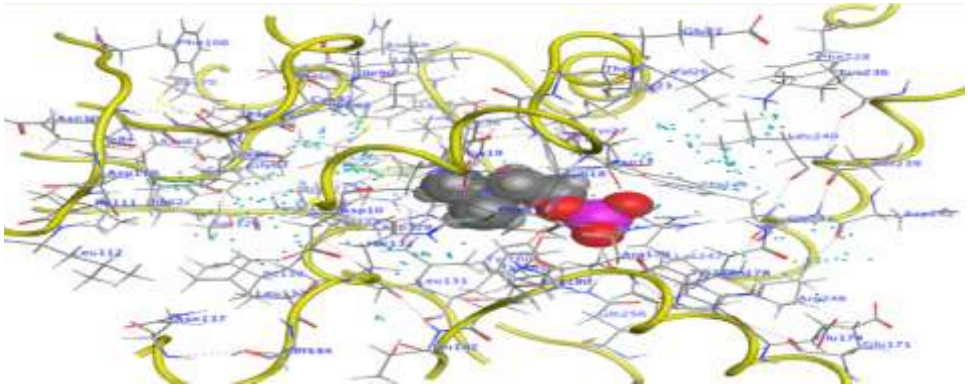


**Fig.6** Ligand phosphocholine and the active site in the 3UJ9 crystal complex.

This active site is critical for inhibiting PfPMT activity. Figure 7 shows the interactions based on 2D and 3D visualizations of the 3UJ9 complex using Discovery Studio 2016 software.

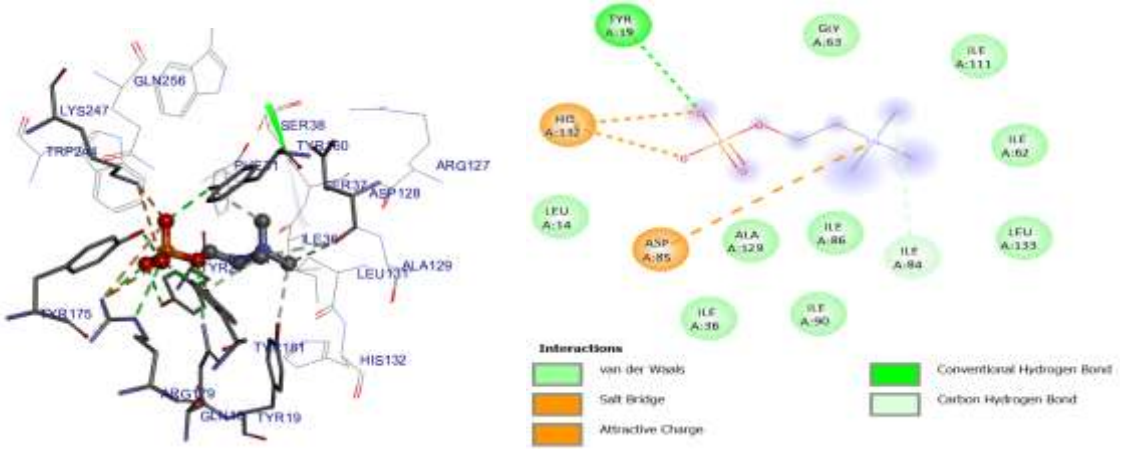



**Fig.7** Interactions (3D and 2D) between the co-crystallized ligand phosphocholine and the active sites in the 3UJ9 crystal complex.

This visualization clearly shows that the catalytic dyad between Tyr19 and His132 is formed by phosphocholine binding sites.

The next phase will be to re-dock the phosphocholine with the PfPMT receptor for confirming the identity of its active site. This technique corroborates the validity of molecular docking as demonstrated in the following study. For this purpose, the grid maps were constructed using 68, 68 and 68 pointing in x, y and z directions with grid point spacing of 0.375 A˚ and the center grid box is of 19.437 A˚, 11.576 A˚ and 17.437 A˚. Figure 8 depicts the conformational relationship between a docked ligand and its native crystallized form in the PfPMT receptor pocket, which is superimposed on the docked ligand.

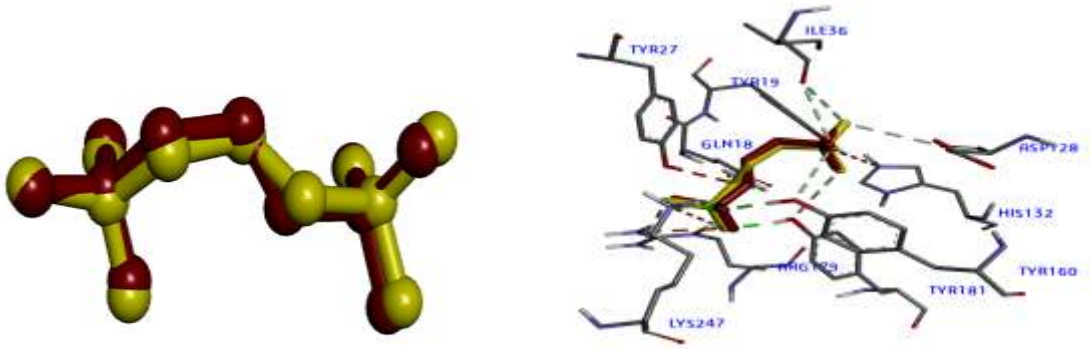
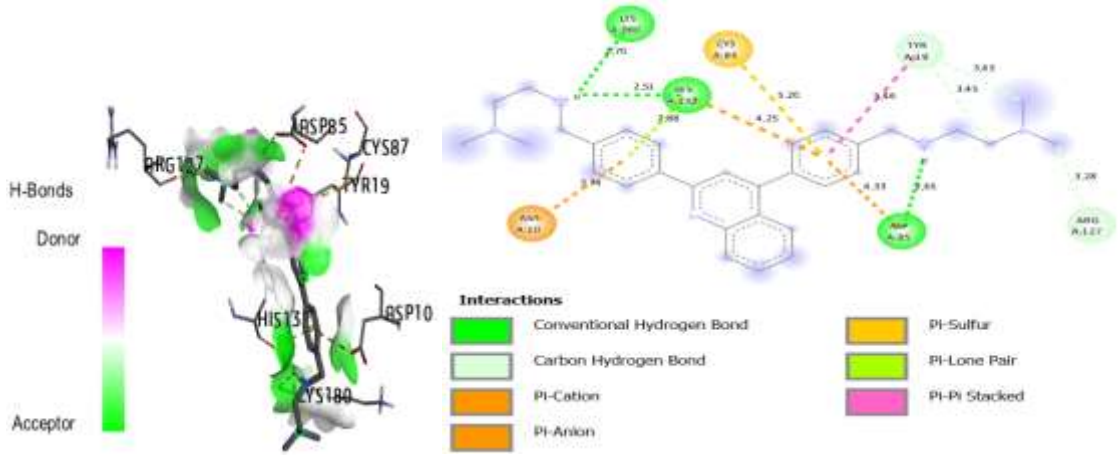


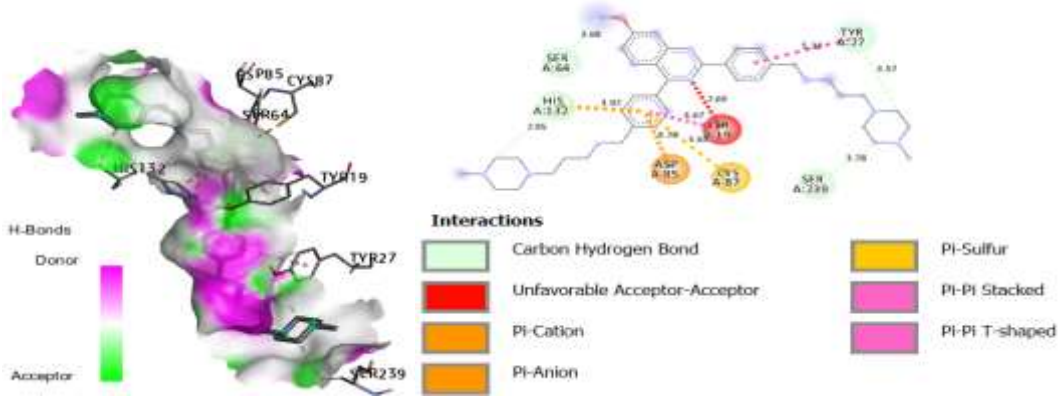

**Fig.8** Re-docking pose with RMSD= 0.256 Å (Brown = Original, Yellow= Docked) and interactions with the active site in the PfPMT receptor predicted by re-docking.

Figure 8 shows that the original and redocked ligands are almost perfectly superimposed in the PfPMT receptor pocket. A further point to note is the low value of the root mean squared deviation (0.256 A˚), which indicates that the grid maps and AutoDock software were effective in achieving excellent molecular docking results.

After identifying the active site involved in the inhibition of the PfPMT receptor, we perform validated docking with the PfPMT receptor for the most active molecule 1c and the least active molecule 3q. Figure 9 depicts the three-dimensional and two-dimensional visualizations of the interactions of the ligands 1c and 3q with the PfPMT receptor, as well as the obtained binding energies.




**a.** Most active molecule (Binding energy = -9.6 kcal/mol)




**b.** Least active molecule 3q (Binding energy = -7.7 kcal/mol)

**Fig.9** 2D and 3D docking poses interactions between ligand 1c (a) and ligand 3q (b) with active site of PfPMT

As can be seen in figure 9.a, the most active molecule 1c has conventional H-bonds between the nitrogen hydrogen chain and His132 (2.180 A°) and Lys180 (2.930 A°) and Asp 85 (2.750 A°). Carbon Hydrogen Bonds were formed with Tyr 19 as crucial amino acid and Arg127. Electrostatic interactions were formed with His132, Asp10, and Asp85 through Pi-ion bonding and formed Pi-sulfur interactions with Cys87. Both the benzene ring and heterocyclic ring packed performing hydrophobic interaction through pi-pi bonding with important residues like Tyr19 that is essential for biological function of PfPMT. While, the least active molecule 3q (11.500 uM) is docked into this pocket by interactions with the following residues; Tyr19 through unfavorable acceptor-acceptor and pi-pi interactions, His132 and Asp85 through Pi-ion interactions, Cys87 by pi-sulfur interactions and through Carbon Hydrogen Bond with His132, Tyr27, Ser64 and Ser239 (refer fig.9.b).

We have observed that the drug is most potent when the binding energy is lowest and there are the greatest number of interactions with the receptor. By comparing the binding energies of the ligand 1c (-9.6 Kcal/mol) and the ligand 3q (-7.7 Kcal/mol) with the PfPMT receptor, we can interpret the lowest experimental IC50 value for the active molecule 1c (0.032 µM) in comparison to the less active molecule 3q (11.500 µM).

Based on the results of the molecular docking predictions, it is clear that the structure of ligand 1c can be used to improve the inhibition of the enzymatic activity of the PfPMT protein, which has been demonstrated in this study. We can also modify the structure of molecule 1c and assess the impact of these modifications on the pIC50 values in order to develop new antimalarial molecules.

When looking for structural clues for new antimalarial activity, we can deduce that determining how long a carbon chain is between two amino functions on a side chain and its shape and atom types at the terminal amine will be key. Therefore, increasing antimalarial activity can be achieved by decreasing the shape or surface area of interacting molecules ( for 3q PSA = 72 ; VWSA = 1082.31 and for 1c PSA = 43.43; VWSA = 806,08 ) as well as by decreasing the volume (increasing density D), increasing the percent of atoms carbon (C% = 77.3 for 1c and for 3q C% = 72), increasing total connectivity (Tcon) by increasing the number of Hydrogen-Bond ( NHB = 7 for 1c and NHB =4 for 3q) and decreasing the bond length between two nitrogen atoms by decreasing the number of rotatable bond (NRB) or number of active torsion (12 for 1c and 15 for 3q).

## 3.5 Proposition of new compounds anti-plasmodium candidates

As shown in the MLR model equation, steric characteristics of the substituents affect antimalarial activity; bond NN decreases activity, while total Connectivity (Tcon), carbon percent ratio (C (%)) and density (D) increase activity.

The significance of each descriptor was determined by comparing its absolute value to the t-test (Student-test) or standardized coefficient; A larger t-test value indicates that the model's descriptor has a higher impact. The best model descriptors have T-values of 5.021, 5.757, 8.856 and -1.396 for total Connectivity, C (%), density (D), and bond length NN, respectively. The length of the bond between two nitrogen atoms (Bond NN) has a negative sign, implying that shortening it can increase activity (increase value of pIC50). The carbon atoms' percent ratio (C (%)), density (D), and total Connectivity (Tcon) are all positive, indicating that increasing these values results in increased activity. As a consequence, the models may be used to predict the biological activity of additional molecules formed by modifying the structure of studied derivatives. By substituting R1, R2, and R7 as well as other modifications to active molecule 1c, twelve derivatives were designed using the zinc-database as a free library of commercially available compounds for virtual screening for substitutions (refer figure 10 and table S7) [**64**].

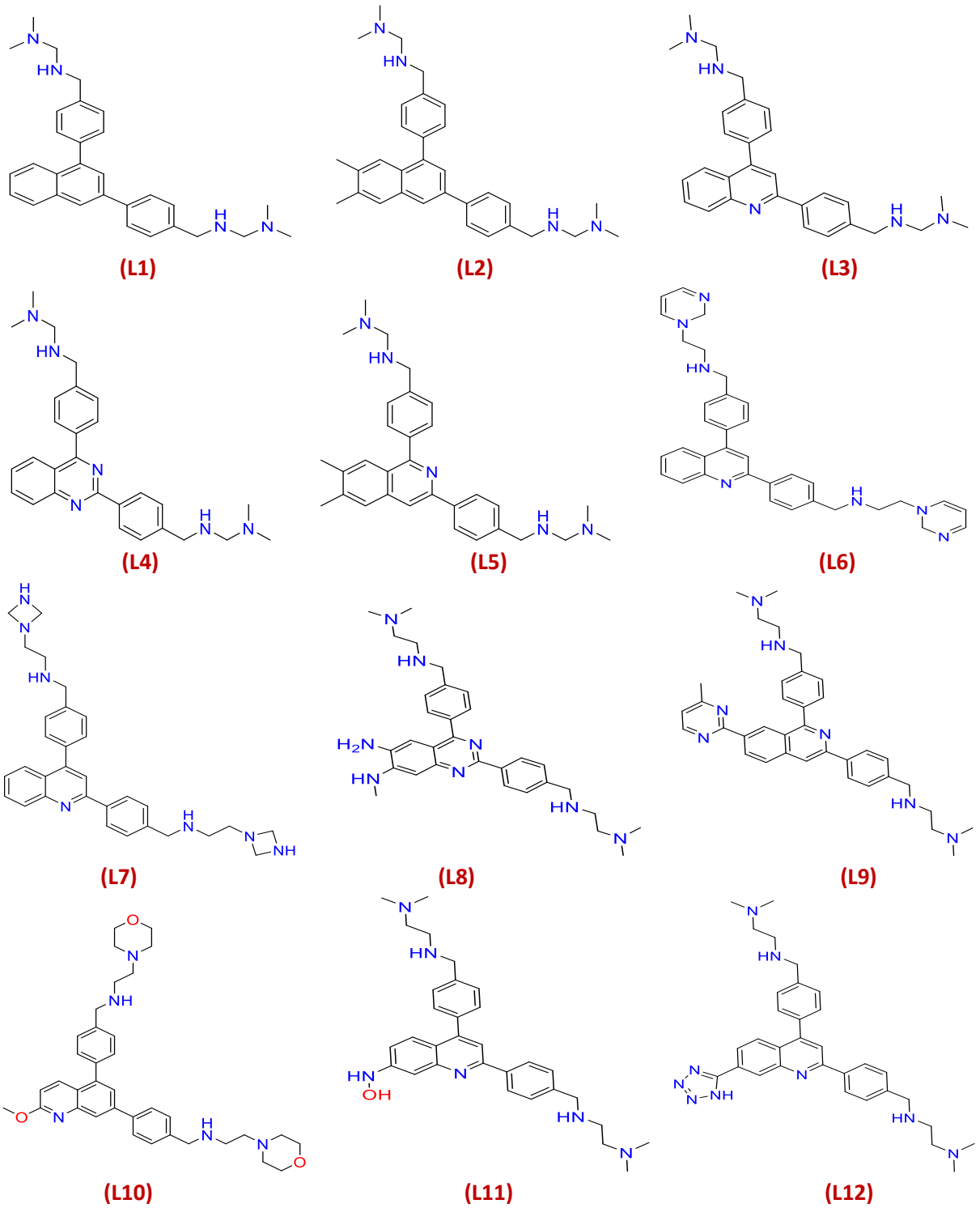



**Fig.10** Structure of designed molecules

We have seen that molecules that lie beyond the application area of the MLR model are not considered drug-candidates. Drug-likeness was also assessed by utilizing the SwissADME online tool, which includes the Lipinski rule of five and synthetic accessibility (SA), which is a crucial feature to examine in this selection process since profiteering may occur in the pharmaceutical industry. Table 5 shows the predicting biological activities, drug-likeness property and hi values of proposed molecules.

**Table 5** Biological activities, drug-likness and leverage hi values of new designed molecules

| N° | pIC50 pred | Lipinsky | Synthetic Accessibility | $h_i$ | AD (h*=0.34) Comment |
|---|---|---|---|---|---|
| L1 | 9.170 | Yes | 3.240 | 1.810 | Outside (hi more than h*) |
| L2 | 8.080 | Yes | 3.480 | 0.940 | Outside (hi more than h*) |
| L3 | 8.970 | Yes | 3.380 | 1.910 | Outside (hi more than h*) |
| L4 | 8.790 | Yes | 8.790 | 2.110 | Outside (hi more than h*) |
| L5 | 7.860 | Yes | 3.720 | 0.880 | Outside (hi more than h*) |
| L6 | 7.990 | Yes | 4.820 | 0.080 | Inside (hi less than h*) |
| L7 | 7.400 | Yes | 3.730 | 0.140 | Inside (hi less than h*) |
| L8 | 6.470 | Yes | 4.040 | 0.320 | Inside (hi less than h*) |
| L9 | 6.480 | Yes | 4.190 | 0.030 | Inside (hi less than h*) |
| L10 | 6.520 | Yes | 4.210 | 0.100 | Inside (hi less than h*) |
| L11 | 7.510 | Yes | 3.920 | 0.300 | Inside (hi less than h*) |
| L12 | 6.570 | Yes | 4.050 | 0.310 | Inside (hi less than h*) |

Each of the twelve proposed compounds has a synthetic accessibility value between 3.24 and 8.79, which is greater than 1 but less than 10. Thus, all designed compounds were chosen as drug candidates based on their ease of synthesis.

According to values of $h_i$ that are less than h*, seven molecules (L6, L7, L8, L9, L10, L11 and L12) are selected as drug-candidates with the highest biological activities. Moreover, the designed ligands L6 and L11 generally exhibited better antimalarial activities on comparison with the molecule 1c as the most active molecule in the studied series.

Based solely on the best anticipated pIC50 values of the seven compounds, it is difficult to favor one molecule over another and pick it as the greatest inhibitor of PfPMT activity. The pharmacokinetic parameters ADMET were employed through the pkCSM online tool to ensure that the identified compounds were viable drugs. As a result, we'll only select compounds with drug-like characteristics. The results of the ADMET properties prediction are presented in Table 6.

**Table 6** ADMET properties of selected molecules and comparison with active molecule (1c)

| **Ligands** | **Proprieties** | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Absorption** | | **Distribution** | | | **Metabolism** | | | | | | | **Excretion** | **Toxicity** |
| | | | | | | CYP | | | | | | | | |
| | | | | | | Substrate | | Inhibitor | | | | | | |
| | pIC50 | Intestinal absorption | VDss | BBB | CNS | 2D6 | 3A4 | 1A2 | 2C19 | 2C9 | 2D6 | 3A4 | Total clearance | AMES toxicity |
| | - | Numeric (%absorb.) | Numeric (Log L kg1) | Numeric (Log BB) | Numeric (Log PS) | Categorical (yes/no) | | | | | | | Numeric (log mL min1 kg1) | Categoical (yes/no) |
| **L.6** | 7.99 | 91.813 | 1.237 | -0.891 | -2.324 | Yes | Yes | Yes | No | No | Yes | No | 0.792 | No |
| **L.7** | 7.40 | 94.854 | 1.476 | -0.23 | -2.998 | Yes | Yes | Yes | No | Yes | Yes | No | 1.471 | No |
| **L.8** | 6.47 | 71.7 | 1.318 | -1.341 | -2.922 | Yes | Yes | Yes | No | Yes | Yes | No | 0.974 | No |
| **L.9** | 6.48 | 97.393 | 0.481 | -1.375 | -2.509 | Yes | Yes | Yes | No | No | No | Yes | 0.576 | No |
| **L.10** | 6.52 | 95.803 | 0.924 | -1.263 | -2.998 | Yes | Yes | No | No | No | Yes | Yes | 1.257 | No |
| **L.11** | 7.51 | 89.778 | 1.02 | -1.368 | -2.739 | Yes | Yes | Yes | Yes | Yes | Yes | Yes | 0.789 | No |
| **L.12** | 6.57 | 63.682 | -0.326 | -1.761 | -3.256 | Yes | Yes | No | No | Yes | No | Yes | 0.436 | No |
| **1C** | 7.49 | 92.962 | 1.19 | -0.001 | -1.414 | Yes | Yes | Yes | No | No | Yes | No | 0.752 | Yes |

If the absorbance is less than 30%, the absorption is low. However, the seven compounds with a value more than 90% have a good absorbance in the human intestine [**65**], indicating that they are well-absorbed. Several studies have shown that the values for large volume of distribution (VDss) exceed 0.45 [**52**]. As a result, all of the proposed compounds have the most significant potential in terms of volume of distribution.

In terms of blood-brain barrier (BBB) and central nervous system (CNS) permeability standard values, when LogBB< -1, compounds are poorly distributed to the brain and when LogBB >0.3, compounds have the potential to cross the BBB. Furthermore, if LogPS > -2 compounds are considered to penetrate the CNS, while LogPS < -3 is difficult to move in the CNS [66-67]. Thus, all proposed compounds have the best significant potential to cross the barriers.

The enzymatic metabolism of a medication in the body indicates its chemical biotransformation. CYP enzymes are found in all bodily tissues and oxidize foreign germs to aid in their elimination. When inhibitors of this enzyme impair its metabolism, the drug may have the opposite effect [68]. CYP3A4 is the most important inhibitor in this study among the CYP families (CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4), which are responsible for the biotransformation of more than 90% of drugs undergoing phase metabolism. This implies that all newly synthesized compounds must function as both a substrate and an inhibitor of CYP3A4 [69]. As a result, only the compounds L9, L10, L11 and L12 were chosen as CYP3A4 substrates and inhibitors.

Clearance measures how quickly drugs are excreted from the body in comparison to their concentrations inside. Drug persistence is not a problem for any of the new compounds in the table 6.

During the earliest stages of drug development, the toxicity study of the predicted compounds plays a significant role. Almost every drug in this research is being tested for toxicity using the AMES test [70]. Table 6 shows that all of the molecular designs are non-toxic. We have found that almost all of the compounds in this series are poisonous, including molecule 1c, which was previously identified as the most effective antimalarial candidate.

We conclude that compound L11 fits all of the pharmacokinetic parameters tested in this contribution research based on the findings of the ADMET characteristics. Thus, compound L11 may be employed as an antimalarial medication in the future by blocking the PfPMT protein's enzymatic activity. Additionally, this molecule may be exploited to develop novel drugs with enhanced anti-plasmodium falciparum activity.

To acquire insight into the critical interactions between ligands and PfPMT, validated molecular docking was performed on all hit compounds. The interactions of the most active molecule in the studied series 1c with PfPMT were then compared to those of the best designed ligand L11.

## 3.7 Molecular docking of selected molecules as drug-candidates

To facilitate docking analysis, all ligands were docked to the active site of the protein using a validated 3D grid.

Table 7 summarizes the binding energies and interactions of the ligands L6, L7, L8, L9, L10, L11, L12 and 1c (as a reference molecule) with the PfPMT receptor.

**Table 7** Docking results of the selected ligands and the molecule 1c as reference.

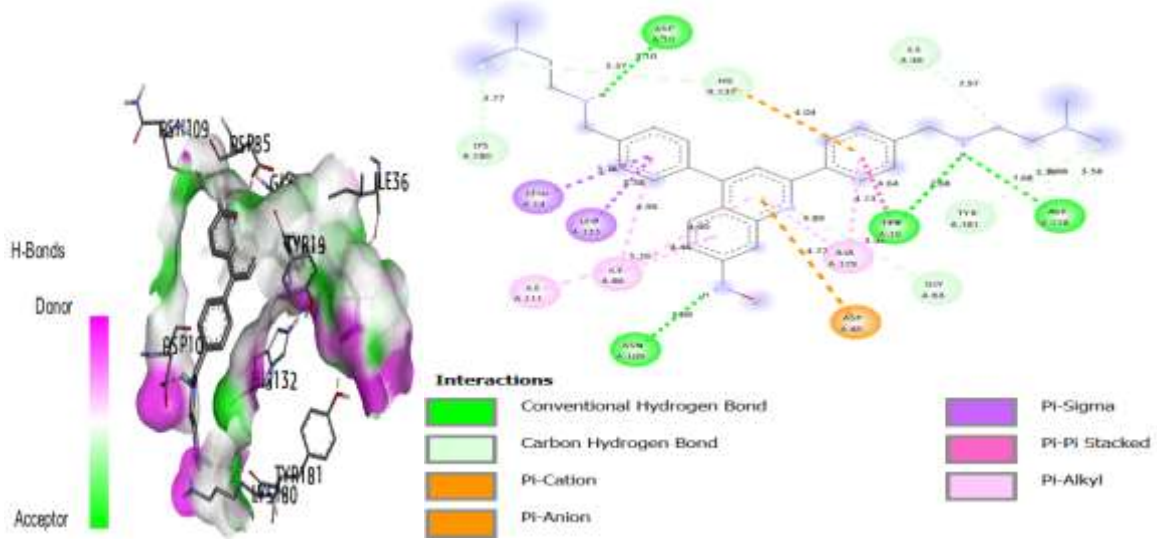
| Ligands | Complex | Binding energy (Kcal /mol) | Hydrogen-Binding interactions (HB) | | Hydrophobic interactions | Electrostatic Interactions |
|---|---|---|---|---|---|---|
| | | | Conventional - HB | C-HB | | |
| **L6** | L6-PfPMT | -10.400 | HIS132-ASN137-TYR160-TYR181 | ASP10-ASP36-ASP128 | TYR19 | ASP85 – ASP85 |
| **L7** | L7-PfPMT | -9.600 | HIS132-GLY63 | LYS180-ASP128-ARG127 | TYR19 | CYS87-ASP61-ASP85-HIS132 |
| **L8** | L8-PfPMT | -4.100 | - | ASP85-ALA129 | TYR19-TYR27-ALA129-ILE36 | HIS132-CYS87-ASP85 |
| **L9** | L9-PfPMT | -6.500 | - | TYR19-ASP128-ARG127-GLY63-LYS236 | TRP244-SER239 | ARG179 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **L10** | L10-PfPMT | -7.900 | HIS132-ILE111-TYR160 | TYR19-ASP128-LYS180 | TYR19-ILE111 | HIS132-CYS87-ASP85 |
| **L11** | L11-PfPMT | -9.910 | TYR19-ASP10-ASP128-ASN109 | HIS132-ILE36-TYR181-GLY63-SER134-LYS180-ASP128 | TYR19 | HIS132-ASP85 |
| **L12** | L12-PfPMT | -6,800 | GLY63-GLY69-LYS247 | TYR27-TYR175-ASN109 | TYR27 | HIS132-ASP85 |
| **1c** | 1C-PfPMT | -9.500 | HIS132-LYS180-ASP85 | TYR19-ARG127 | TYR19 | ASP10 |

From table 7, the best ligands L6 and L11 have the lowest binding energy (-10.400 Kcal/mol and -9.910 Kcal/mol respectively) compared to others ligands in the order (1c, L7, L8, L9, L10 and L12). Thus, by comparing the values of the binding energies of the best ligands L6 and L11 with the receptor PfPMT, we can interpret the highest pIC50 value of the designed molecules L6 (7.990) and L11 (7.510) that reflect the high activity of these molecules compared to the experimental pIC50 value of the molecule 1c (7.490). In addition, the stronger biological activity of L6 and L11 can be also explained by the number and type of interactions with the active site of receptor (refer figure S2).

Number of interactions with important amino acids (His132 and Tyr19) involved in PfPMT inhibition shows that these interactions are important for binding energy and biological activities. The ligands (L6, 1C, L11, L7, L10 and L12) with pIC50 values above 6.500 have the most interactions in PfPMT inhibition compared to L8 and L9. Therefore, the formation of a hydrogen bond with the amino acid His132 is also important in the inhibition of drug candidates against P. falciparum.

The primary hit ligand L11 also exhibited strong interaction within the binding pocket by occupying critical amino acids Tyr19 and His132 that functions as a general base to extract a proton(H+) from the hydroxyl group (OH) of Tyr19 to activate the residue (refer figure 11).




(a) (b)

**Fig.11** The 3D interactions (a) and 2D interactions (b) of the best designed ligand (L11)

The hydrogen atom linked to nitrogen atom formed conventional hydrogen bonds with crucial amino acid Tyr19 (4.64A°), Asn10 (3.10A°), Asn109 (3.09) and Asp128 (2.82A°). The carbon that is linked to nitrogen atom plays a main role of formation of carbon-hydrogen bonds with crucial amino acid His132 (3.37A°), Tyr181 (3.19A°), Lys180 (3.77A°), ILE36 (2.97A°), Asp128 (3.58A°) and Gly63 (4.89 A°). The benzene and heterocyclic rings play a crucial role of hydrophobic and electrostatic interactions.

According to the docking results for selected ligand L11, we can conclude that the drug-like properties against plasmodium falciparum may be improved by increasing the number of hydrogen bonds in proposed ligands in order to avoid the liberation of toxic ions due to the stronger interactions between these ligands and the PfPMT receptor. Thus, it is possible that hydrogen bonds suppressed in the chemical graph structure for molecular connectivity calculations could be considered a key parameter in this research. As a result, they should be included in any discussion of molecular connectivity indicators for new antimalarial drug candidates [71].

We notice that the Tcon descriptor establishes a direct relationship between the number of interacting hydrogen bonds (C-H and X-H) and the molecular connectivity index for each molecule. These results corroborate the QSAR and molecular docking analyses, as well as the experimental results.

Based on the results that are achieved by combining the study of 2D-QSAR with ADMET and molecular docking studies, it is clear that the structure of ligand L11 can be used to improve the inhibition of the enzymatic activity of PfPMT protein. In addition, we can modify the structure of this ligand in order to design new antimalarial drugs that can be added to the quinoline, isoquinoline and quinazoline derivatives.

## 4 Conclusion

In this contribution study, molecular modeling was performed to improve the antimalarial activity against the Plasmodium falciparum CQ-sensitive and MQ-resistant strain 3D7 protozoan parasite of a set of fifty-four quinoline, isoquinoline and quinazoline derivatives.

The statistical results obtained through internal and external validations corroborate the generated models' performance. According to the best MLR model, it is clear that steric parameters have the greatest effect on the investigated activities. As a result, the proposed models can be used to predict the antimalarial activity of novel quinoline, isoquinoline and quinazoline derivatives.

To optimize the studied activity, the most and least active molecules were chosen based on the developed models' predictions. Then, docking study was conducted on these molecules with the PfPMT receptor in order to use them as a starting point for the development of new molecules with the highest activity.

Based on the obtained results, twelve molecules (L1-L12) with the highest antimalarial activity were designed. The results of applicability domain, drug-likeness and ADMET properties showed that only the compound L11 had acceptable pharmacokinetics properties. Then, using molecular docking, the inhibitory activity of the best drug candidate L11 was confirmed. Due to its ability to interfere with the enzymatic activity of the PfPMT protein, this compound may be proposed as a novel antimalarial agent for use in the treatment of malaria. Moreover, it may also provide a wealth of opportunities for medicinal chemists to develop new plasmodium falciparum drug candidates.

## Acknowledgment

We are grateful to the “Association Marocaine des Chimistes Théoriciens” (AMCT) for its pertinent help concerning the programs and softwares.

## Declarations

### *Funding

No external funding was obtained for this study.

### *Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### *Availability of data and materials

The data availability is included in the manuscript

### *Code availability

Not applicable

### *Authors' contributions

These authors contributed equally to this work

## References

1. Health Organization W GLOBAL TECHNICAL STRATEGY FOR MALARIA 2016-2030 Global Technical Strategy for Malaria 2016-2030 Global Malaria Programme World Health Organization
2. Health Organization W World malaria report 2020: 20 years of global progress and challenges
3. Mwenesi H, Mbogo C, Casamitjana N, et al Confidential-Not for Distribution "Rethinking Human Resources and Capacity Building Needs for Malaria Control and Elimination in Africa," by
4. Hocart SJ, Liu H, Deng H, et al (2011) 4-aminoquinolines active against chloroquine-resistant Plasmodium falciparum: basis of antiparasite activity and quantitative structure-activity relationship analyses. Antimicrob Agents Chemother 55:2233–2244. https://doi.org/10.1128/AAC.00675-10
5. Hadni H, Elhallaoui M (2020) 3D-QSAR, docking and ADMET properties of aurone analogues as antimalarial agents. Heliyon 6:e03580. https://doi.org/10.1016/J.HELIYON.2020.E03580
6. Hadni H, Mazigh M, Charif E, et al (2018) Molecular Modeling of Antimalarial Agents by 3D-QSAR Study and Molecular Docking of Two Hybrids 4-Aminoquinoline-1,3,5-triazine and 4-Aminoquinoline-oxalamide Derivatives with the Receptor Protein in Its Both Wild and Mutant Types. Biochem Res Int 2018:. https://doi.org/10.1155/2018/8639173
7. Hadni H, Elhallaoui M (2020) 2D and 3D-QSAR, molecular docking and ADMET properties in silico studies of azaaurones as antimalarial agents. New J Chem 44:6553–6565. https://doi.org/10.1039/C9NJ05767F
8. Witola WH, Pessi G, El Bissati K, et al (2006) Localization of the Phosphoethanolamine Methyltransferase of the Human Malaria Parasite Plasmodium falciparum to the Golgi Apparatus *. J Biol Chem 281:21305–21311. https://doi.org/10.1074/JBC.M603260200
9. Witola WH, El Bissati K, Pessi G, et al (2008) Disruption of the Plasmodium falciparum PfPMT Gene Results in a Complete Loss of Phosphatidylcholine Biosynthesis via the Serine-Decarboxylase-Phosphoethanolamine-Methyltransferase Pathway and Severe Growth and Survival Defects *. J Biol Chem 283:27636–27643. https://doi.org/10.1074/JBC.M804360200
10. Lee SG, Kim Y, Alpert TD, et al (2012) Structure and reaction mechanism of phosphoethanolamine methyltransferase from the malaria parasite Plasmodium falciparum: An antiparasitic drug target. J Biol Chem 287:1426–1434. https://doi.org/10.1074/JBC.M111.315267/ATTACHMENT/F9BD9DDE-084F-4060-81A9-8E07149522FB/MMC2.PDF
11. El Aissouq A, Toufik H, Stitou M, et al (2019) In Silico Design of Novel Tetra-Substituted Pyridinylimidazoles Derivatives as c-Jun N-Terminal Kinase-3 Inhibitors, Using 2D/3D-QSAR Studies, Molecular Docking and ADMET Prediction. Int J Pept Res Ther 2019 263 26:1335–1351. https://doi.org/10.1007/S10989-019-09939-8
12. Hadni H, Bakhouch M, Elhallaoui M (2021) 3D-QSAR, molecular docking, DFT and ADMET studies on quinazoline derivatives to explore novel DHFR inhibitors. https://doi.org/101080/0739110220212004233. https://doi.org/10.1080/07391102.2021.2004233
13. El Aissouq A, Chedadi O, Bouachrine M, Ouammou A (2021) Identification of Novel SARS-CoV-2 Inhibitors: A Structure-Based Virtual Screening Approach. J Chem 2021:. https://doi.org/10.1155/2021/1901484
14. El-Mernissi R, El Khatabi K, Khaldan A, et al (2021) Design of new 3, 5-disubstituted indole as hematological anticancer agents using 3D-QSAR, molecular docking and drug-likeness studies. Mater Today Proc 45:7608–7614. https://doi.org/10.1016/J.MATPR.2021.03.080

15. Daoui O, Elkhattabi S, Chtita S, et al (2021) QSAR, molecular docking and ADMET properties in silico studies of novel 4,5,6,7-tetrahydrobenzo[D]-thiazol-2-Yl derivatives derived from dimedone as potent anti-tumor agents through inhibition of C-Met receptor tyrosine kinase. Heliyon 7:. https://doi.org/10.1016/J.HELIYON.2021.E07463
16. Chedadi O, El Aissouq A, El Ouardi Y, et al (2021) 3D-QSAR and molecular docking studies of 4-methyl quinazoline derivatives as PI3Kα inhibitors. J Indian Chem Soc 98:100183. https://doi.org/10.1016/J.JICS.2021.100183
17. El Aissouq A, Chedadi O, Kasmi R, et al (2021) Molecular Modeling Studies of C-Glycosylfavone Derivatives as GSK-3β Inhibitors Based on QSAR and Docking Analysis. J Solut Chem 2021 505 50:808–822. https://doi.org/10.1007/S10953-021-01083-6
18. Belete TM (2020) Recent Progress in the Development of New Antimalarial Drugs with Novel Targets . Drug Des Devel Ther 14:3875–3889. https://doi.org/10.2147/DDDT.S265602
19. Vilar S, Cozza G, Moro S (2008) Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. Curr Top Med Chem 8:1555–1572. https://doi.org/10.2174/156802608786786624
20. Guillon J, Cohen A, Boudot C, et al (2020) Design, synthesis, and antiprotozoal evaluation of new 2,4-bis[(substituted-aminomethyl)phenyl]quinoline, 1,3-bis[(substituted-aminomethyl)phenyl]isoquinoline and 2,4-bis[(substituted-aminomethyl)phenyl]quinazoline derivatives. J Enzyme Inhib Med Chem 35:432–459. https://doi.org/10.1080/14756366.2019.1706502
21. Chtita S, Ghamali M, Ousaa A, et al (2019) QSAR study of anti-Human African Trypanosomiasis activity for 2-phenylimidazopyridines derivatives using DFT and Lipinski's descriptors. Heliyon 5:e01304. https://doi.org/10.1016/J.HELIYON.2019.E01304
22. Singh J, Vijay S, Mansuri R, et al (2019) Computational and experimental elucidation of Plasmodium falciparum phosphoethanolamine methyltransferase inhibitors: Pivotal drug target. PLoS One 14:e0221032. https://doi.org/10.1371/JOURNAL.PONE.0221032
23. Allinger NL (2002) Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. J Am Chem Soc 99:8127–8134. https://doi.org/10.1021/JA00467A001
24. ChemDraw and ChemOffice | Computing. https://www.ch.cam.ac.uk/computing/software/chemdraw-and-chemoffice. Accessed 31 Dec 2021
25. Structure Drawing Software for Academic and Personal Use | ACD/ChemSketch. https://www.acdlabs.com/resources/freeware/chemsketch/index.php. Accessed 31 Dec 2021
26. Marvin | ChemAxon. https://chemaxon.com/products/marvin. Accessed 31 Dec 2021
27. Gaussian 09 Citation | Gaussian.com. https://gaussian.com/g09citation/. Accessed 31 Dec 2021
28. David CC, Jacobs DJ (2014) Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol Biol 1084:193–226. https://doi.org/10.1007/978-1-62703-658-0_11
29. Lagunin A, Zakharov A, Filimonov D, Poroikov V (2011) QSAR Modelling of Rat Acute Toxicity on the Basis of PASS Prediction. Mol Inform 30:241–250. https://doi.org/10.1002/MINF.201000151
30. Leonard JT, Roy K (2006) On selection of training and test sets for the development of predictive QSAR models. QSAR Comb Sci 25:235–251. https://doi.org/10.1002/QSAR.200510161
31. Papa E, Dearden JC, Gramatica P (2007) Linear QSAR regression models for the prediction of bioconcentration factors by physicochemical properties and structural theoretical molecular descriptors. Chemosphere 67:351–358. https://doi.org/10.1016/J.CHEMOSPHERE.2006.09.079
32. Salt DW, Yildiz N, Livingstone DJ, Tinsley CJ (1992) The Use of Artificial Neural Networks in QSAR. Pestic Sci 36:161–170. https://doi.org/10.1002/PS.2780360212
33. XLSTAT version 2019.1 | Logiciel statistique pour Excel. https://www.xlstat.com/fr/articles/xlstat-version-2019-1. Accessed 31 Dec 2021
34. MATLAB - Télécharger. https://matlab.fr.softonic.com/. Accessed 31 Dec 2021
35. Kiralj R, Ferreira MMC (2009) Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. J Braz Chem Soc 20:770–787. https://doi.org/10.1590/S0103-50532009000400021
36. Gramatica P, Sangion A (2016) A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology. J Chem Inf Model 56:1127–1131. https://doi.org/10.1021/ACS.JCIM.6B00088
37. Javidfar M, Ahmadi S (2020) QSAR modelling of larvicidal phytocompounds against Aedes aegypti using index of ideality of correlation. https://doi.org/101080/1062936X20201806922 31:717–739. https://doi.org/10.1080/1062936X.2020.1806922
38. Rácz A, Bajusz D, Héberger K (2018) Modelling methods and cross-validation variants in QSAR: a multi-level analysis$. https://doi.org/101080/1062936X20181505778 29:661–674. https://doi.org/10.1080/1062936X.2018.1505778
39. Kůrková V (1992) Kolmogorov's theorem and multilayer neural networks. Neural Networks 5:501–506.

https://doi.org/10.1016/0893-6080(92)90012-8
40. Andrea TA, Kalayeh H (1991) Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. J Med Chem 34:2824–2836. https://doi.org/10.1021/JM00113A022
41. Golbraikh A, Tropsha A (2002) Beware of q2! J Mol Graph Model 20:269–276. https://doi.org/10.1016/S1093-3263(01)00123-1
42. Rücker C, Rücker G, Meringer M (2007) y-Randomization and its variants in QSPR/QSAR. J Chem Inf Model 47:2345–2357. https://doi.org/10.1021/CI700157B
43. Roy K, Mitra I (2011) On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. Comb Chem High Throughput Screen 14:450–474. https://doi.org/10.2174/138620711795767893
44. Netzeva TI, Worth AP, Aldenberg T, et al (2005) Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. Altern Lab Anim 33:155–173. https://doi.org/10.1177/026119290503300209
45. Gramatica P (2007) Principles of QSAR models validation: internal and external. QSAR Comb Sci 26:694–701. https://doi.org/10.1002/QSAR.200610151
46. Eriksson L, Jaworska J, Worth AP, et al (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect 111:1361. https://doi.org/10.1289/EHP.5758
47. Frey A, Di Canzio J, Zurakowski D (1998) A statistically defined endpoint titer determination method for immunoassays. J Immunol Methods 221:35–41. https://doi.org/10.1016/S0022-1759(98)00170-7
48. Hansch C, Leo A, Mekapati SB, Kurup A (2004) QSAR and ADME. Bioorg Med Chem 12:3391–3400. https://doi.org/10.1016/J.BMC.2003.11.037
49. Jin Z, Wang Y, Yu XF, et al (2020) Structure-based virtual screening of influenza virus RNA polymerase inhibitors from natural compounds: Molecular dynamics simulation and MM-GBSA calculation. Comput Biol Chem 85:. https://doi.org/10.1016/J.COMPBIOLCHEM.2020.107241
50. Chtita S, Aouidate A, Belhassan A, et al (2020) QSAR study of N-substituted oseltamivir derivatives as potent avian influenza virus H5N1 inhibitors using quantum chemical descriptors and statistical methods. New J Chem 44:1747–1760. https://doi.org/10.1039/C9NJ04909F
51. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Reports 2017 71 7:1–13. https://doi.org/10.1038/srep42717
52. Pires DEV, Blundell TL, Ascher DB (2015) pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J Med Chem 58:4066–4072. https://doi.org/10.1021/ACS.JMEDCHEM.5B00104/SUPPL_FILE/JM5B00104_SI_001.PDF
53. Chtita S, Belhassan A, Aouidate A, et al (2021) Discovery of Potent SARS-CoV-2 Inhibitors from Approved Antiviral Drugs via Docking and Virtual Screening. Comb Chem High Throughput Screen 24:441–454. https://doi.org/10.2174/1386207323999200730205447
54. Free Download: BIOVIA Discovery Studio Visualizer - Dassault Systèmes. https://discover.3ds.com/discovery-studio-visualizer-download. Accessed 31 Dec 2021
55. MGLTools including AutoDockTools | WestGrid. https://www.westgrid.ca/support/software/mgltools. Accessed 31 Dec 2021
56. Verdonk ML, Cole JC, Hartshorn MJ, et al (2003) Improved protein-ligand docking using GOLD. Proteins Struct Funct Genet 52:609–623. https://doi.org/10.1002/PROT.10465
57. Westermaier Y, Barril X, Scapozza L (2015) Virtual screening: an in silico tool for interlacing the chemical universe with the proteome. Methods 71:44–57. https://doi.org/10.1016/J.YMETH.2014.08.001
58. Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. Chemom Intell Lab Syst 145:22–29. https://doi.org/10.1016/J.CHEMOLAB.2015.04.013
59. OECD PRINCIPLES FOR THE VALIDATION, FOR REGULATORY PURPOSES, OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP MODELS
60. Tropsha A (2010) Best Practices for QSAR Model Development, Validation, and Exploitation. Mol Inform 29:476–488. https://doi.org/10.1002/MINF.201000061
61. Belhassan A, Chtita S, Lakhlifi T, Bouachrine M (2017) QSPR study of the retention/release property of odorant molecules in water using statistical methods. Orbital 9:234–247. https://doi.org/10.17807/ORBITAL.V9I4.978
62. Roy K, Mitra I (2011) On Various Metrics Used for Validation of Predictive QSAR Models with Applications in Virtual Screening and Focused Library Design. Comb Chem High Throughput Screen 14:450–474. https://doi.org/10.2174/138620711795767893
63. Garg R, Smith CJ (2014) Predicting the bioconcentration factor of highly hydrophobic organic

chemicals. Food Chem Toxicol 69:252–259. https://doi.org/10.1016/J.FCT.2014.03.035
64. Irwin JJ, Tang KG, Young J, et al (2020) ZINC20 - A Free Ultralarge-Scale Chemical Database for Ligand Discovery. J Chem Inf Model 60:6065–6073. https://doi.org/10.1021/ACS.JCIM.0C00675/SUPPL_FILE/CI0C00675_SI_001.PDF
65. Kalantzi L, Goumas K, Kalioras V, et al (2006) Characterization of the human upper gastrointestinal contents under conditions simulating bioavailability/bioequivalence studies. Pharm Res 23:165–176. https://doi.org/10.1007/S11095-005-8476-1
66. Speciale A, Muscarà C, Molonia MS, et al (2021) Silibinin as potential tool against SARS-Cov-2: In silico spike receptor-binding domain and main protease molecular docking analysis, and in vitro endothelial protective effects. Phytother Res 35:4616–4625. https://doi.org/10.1002/PTR.7107
67. Han Y, Zhang J, Hu CQ, et al (2019) In silico ADME and toxicity prediction of ceftazidime and its impurities. Front Pharmacol 10:434. https://doi.org/10.3389/FPHAR.2019.00434/BIBTEX
68. Domínguez-Villa FX, Durán-Iturbide NA, Ávila-Zárraga JG (2021) Synthesis, molecular docking, and in silico ADME/Tox profiling studies of new 1-aryl-5-(3-azidopropyl)indol-4-ones: Potential inhibitors of SARS CoV-2 main protease. Bioorg Chem 106:104497. https://doi.org/10.1016/J.BIOORG.2020.104497
69. Zanger UM, Schwab M (2013) Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. Pharmacol Ther 138:103–141. https://doi.org/10.1016/J.PHARMTHERA.2012.12.007
70. Ferraz ERA, Umbuzeiro GA, de-Almeida G, et al (2011) Differential toxicity of Disperse Red 1 and Disperse Red 13 in the Ames test, HepG2 cytotoxicity assay, and Daphnia acute toxicity test. Environ Toxicol 26:489–497. https://doi.org/10.1002/TOX.20576
71. Pogliani L (2000) From Molecular Connectivity Indices to Semiempirical Connectivity Terms: Recent Trends in Graph Theoretical Descriptors. Chem Rev 100:3827–3858. https://doi.org/10.1021/CR0004456

# Supplementary Files

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

- SupportingmaterialTables.docx