Computer-aided molecular design of 2-anilino 4-amino substituted quinazolines derivatives as malarial inhibitors

Zakari Ya'u Ibrahim¹ · Adamu Uzairu¹ · Gideon Adamu Shallangwa¹ · Stephen Eyije Abechi¹ · Sulaiman Isyaku¹

Received: 7 April 2021 / Accepted: 27 July 2021

Published online: 05 August 2021

© The Author(s) 2021

Abstract

Quantitative structure–activity relationship studies conducted on forty-five (45) derivatives of 2-anilino 4-amino substituted quinazolines as malaria inhibitors to determine the structures responsible for their antimalarial properties and design novel derivatives with improved activities. The molecular descriptors generated were selected to develop the theoretical model using the genetic approximation component of the material studio. The developed model found to be a function of ATSC8c, GATS8i, SpMin1_Bhi, JGI10, and TDB6u descriptors, shows excellent statistical parameters ($R^2 = 0.7913$, $R^2_{adj} = 0.7553$, $Q^2_{cv} = 0.7112$, $LOF = 0.2125$, and $R^2_{pred} = 0.7650$). The mean effect (MF) analysis revealed the descriptor SpMin1_Bhi, as the most influential by its largest percentage contribution (54%) to the developed model. The descriptor decodes the information on the first ionization potentials and was found to have positive MF. Hence, activity increases with increases the descriptor value. Structural modifications of the template (compound 13; $pEC_{50} = 7.387$) using electron-withdrawing groups increases the descriptor value (first ionization potentials) of the template, which by extension increases the antimalarial activity lead to the design of ten (10) novel theoretical derivatives with improved antimalarial activities. Compound 3, N4-(3-bromo-5-fluorobenzyl)-N2-(4-fluorophenyl)-6,7-dimethoxyquinazoline-2,4-diamine was found to have the highest antimalarial activities among all the designed derivatives ($pEC_{50} = 8.0515$).

Keywords QSAR · Design · Antimalarial · 2-Anilino 4-amino substituted quinazolines · GA · *P. falciparum* strain

1 Introduction

Malaria has maintained its status as one of the most devastating diseases in evolving countries, which increases several socioeconomic challenges [3]. Estimated malaria casualty in the year 2017 is said to be in the region of 200 million infections worldwide leading to over 400 thousand death [10], with children in the age bracket of 5 years mostly affected [22]. The protozoan, Plasmodium consisting of five species of the renowned parasite (*P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale*, and *P. knowlesi*) is the genesis

of the contagious disease, with *P. falciparum* as the deadliest of the species [10].

The use of drugs such as chloroquine and artemisinin to treat malaria remains the major means of getting rid of the disease [21], despite several efforts put in at producing a malaria vaccine [15]. The uses of drugs for the treatment are however not without its challenges, in form of the resistance of *P. falciparum* against the available antimalarial drugs. The optimization of novel antimalarial drugs using new target molecules, having the capacity of overcoming this resistance menace has become a most and urgent quest to embark [5]. Several compounds such as

✉ Zakari Ya'u Ibrahim, zakariyyadibrahim@gmail.com | ¹Department of Chemistry, Faculty of Physical Sciences, Ahmadu Bello University, P.M.B 1045, Zaria, Nigeria.

2-anilino 4-amino substituted quinazolines have indicated antimalarial potency in recent times. Gibson et al. claimed that the 2-anilino 4-amino substituted quinazolines target Plasmodium dihydrofolate reductase (DHFR) [8]. Although no pattern or pharmacophores were reported to be responsible for their antimalarial activity, the unlimited substitution at specific positions may be responsible for their observed potency. Hence, designing derivatives of 2-anilino 4-amino substituted quinazoline through substitution of a variety of groups at specific positions will collaborate this claim.

The synthesis of compounds with improved activities in the drug industry has always been a huge task as it cost time and resources to carry out what is practically a trial and error procedure. Hence, an alternative method of predicting the activity of the compounds before their synthesis becomes a necessity. And quantitative structure–activity relationship (QSAR) is one of the methods available for the prediction of the antimalarial activity of designed derivatives of 2-anilino 4-amino substituted quinazolines.

Several applications of QSAR studies have been reported for lots of biological compounds. Da Silva and his co-researchers conducted a QSAR analysis on some series of arylsulfonamide derivatives on 5-hydroxytryptamine subtype 6 (5-HT₆) receptor to design novel anti-Alzheimer disease using PLS and CoMFA. The results of the study show that 23 compounds were designed to have better activity than the original compounds [6]. Thirty inhibitors of the acetylcholinesterase enzyme were modeled using the QSAR method, results in a robust and predictive model for the prediction of anti-Alzheimer disease [11].

This research targeted the development a descriptive model to predict the activity of antimalarial compounds through the use of experimentally determined compounds with activity values. The descriptors used for this analysis were obtained from Padel software and the contribution of each descriptor was determined which played a role in the design of novel antimalarial derivatives.

2 Materials and methods

2.1 Data source and preparation

The forty-five (45) derivatives of 2-Anilino 4-Amino Substituted Quinazolines deployed for this research (Table 1), were extracted from literature [8]. The 2-D structures were drawn with the aid of ChemDraw Ultra-version 12.0 [19] and thereafter opened in the Spartan software in 3-D for full geometry optimization. The Padel software was used to calculate 1875 molecular descriptors for each of the optimized compounds. The activities of the derivatives were converted to $-\log_{10}^{EC_{50}}$ (pEC₅₀) for better QSAR studies.

Furthermore, the data set was slashed into 35 for model construction (training set) and 10 for the model validation (test sets).

2.2 Feature selection

2.2.1 Data pre-treatment and selection

The molecular descriptors were subjected to some series of treatments such as discarding all constant value descriptors as well as those that are highly correlated with each other. Furthermore, descriptors containing empty cells are also discarded [2]. This kind of descriptors are normally referred to as noisy descriptors and could make the selection of the informative descriptors more difficult. The whole of data pre-treatment was carried out with the aid of the "Data Pre-Treatment GUI 1.2" tool that employs the V-WSP algorithm [16].

2.2.2 Data division

The forty-five (45) derivatives of 2-Anilino 4-Amino Substituted Quinazolines data set were slashed into 35 training set (77.7% of the data set) for model construction and 10 test set (22.2% of the data set) for model validation. In slashing the data into a test and training set, the Kennard Stone algorithm technique of "Dataset Division GUI 1.2" software was deployed [4].

2.3 QSAR model development and validation

The predictive models were initiated from the training set where the activities serve as the response variable (pEC₅₀), and the descriptor values as the explanatory variable by subjecting them to the material studio's genetic approximation (GA) component. The genetic algorithm (GA) is a heuristic search algorithm that capitalized on the evolutionary concept of natural selection and genetics [9]. The method has the advantage of addressing the challenges of constrained and unconstrained encountered during optimization. The GAs makes use of historical information through a definite search for excellent performance [17]. A genetic algorithm may find tremendous use in searching for a large pool of descriptors. The crossovers, smoothness values, are fixed at 800,000 and 1.00 respectively during the algorithm, as well as allowing other default settings. To construct the model, between five (5) to seven (7) descriptors were set as possible numbers in a model. The GA accessed the model fitness through evaluation of the Friedman lack of-fit (LOF) expressed by Eq. 1 below;

$$LOF = SSE / (1 - z + \frac{dn}{M})^2 \quad (1)$$

Table 1 Molecular structures of 2-Anilino 4-Amino Substituted Quinazolines derivatives, their PubChem ID, along side their activities against *P. falciparum* strain, 3D7

S/N	PubChem ID	STRUCTURE	EC ₅₀	pEC ₅₀
1 [^]	137653280		0.283	6.548
2	137640710		0.134	6.873
3	137642131		0.051	7.292
4	137642342		0.105	6.979
5	137642713		0.539	6.268
6	137645594		0.104	6.983
7	137648236		1.300	5.886

Table 1 (continued)

8	137648458		0.102	6.991
9	137649875		0.329	6.483
10	137650600		0.057	7.244
11	50946356		0.136	6.866
12	137651712		0.077	7.114
13	137652254		0.041	7.387
14	137653812		0.057	7.244

Table 1 (continued)

15	52110629		0.056	7.252
16	137655317		1.200	5.921
17	137656431		0.121	6.917
18	137656788		1.800	5.745
19	137657146		0.118	6.928
20	137658134		0.866	6.062

Table 1 (continued)

21	50947795		0.169	6.772
22	137660574		0.042	7.377
23	137661630		0.110	6.959
24^	11689305		0.574	6.241
25	1316842		0.125	6.903
26	1330474		0.264	6.578

where SSE = sum of squares of errors, z = number of terms other than the constant in the model, d , represents the

smoothing parameter, n , represents the model descriptor count, and M , representing the training set counts.

Table 1 (continued)

27 [^]	797650		0.077	7.114
28 [^]	886392		0.109	6.963
29 [^]	2952933		0.144	6.842
30 [^]	2929638		0.317	6.499
31	1333198		0.159	6.799
32	940433		0.124	6.907

Table 1 (continued)

33^	797083		0.112	6.951
34	2929218		0.526	6.279
35	1285085		0.134	6.873
36	2951910		0.230	6.638
37^	1333189		0.113	6.947
38^	2952061		1.100	5.959

Table 1 (continued)

39	50947631		0.155	6.810
40	137633004		0.114	6.943
41	137633575		0.636	6.197
42	137634078		0.137	6.863
43	137634534		0.064	7.194
44^	8546968		7.600	5.119
45	137635910		0.702	6.154
^ Test Set				

2.3.1 Model validation-internal

The leave-one-out (LOO) cross-validation was applied in model validation. As the name implies, the method involves leaving a molecule out of the data set before model development, which is then used to calculate the activity of the compound left out. The procedure is repeated again and again until all the data were left out once and were predicted with the constructed model. Equation 2, ensured the determination of the cross-validated correlation coefficient squared, R^2_{cv} (Q^2).

$$R^2_{cv} = 1 - \left\{ \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - \hat{Y})^2} \right\} \quad (2)$$

where Y_{obs} stands for the training set activity, Y_{pred} stands for the training set predicted activity while $\hat{Y}$ stands for the training set observed mean activity of the.

2.3.2 Model validation-external

The predictive strength of the developed model is determined through the validation of the model constructed externally. The method involves splitting the data into training and test sets, where the training set is used to develop a model. The developed model is used to predict the activity of the test set, thereby estimating the value of the predictive R^2 (R^2_{pred}) of the test set expressed by the formula presented in Eq. 3.

$$R^2_{pred} = 1 - \left\{ \frac{\sum (Y_{Pred(Test)} - Y_{Test})^2}{\sum (Y_{Test} - \hat{Y}_{Training})^2} \right\} \quad (3)$$

where $Y_{Pred(Test)}$ and Y_{Test} respectively represent predicted and observed activity of the test compounds. $\hat{Y}_{Training}$ represents the training set mean activity value.

2.4 Y-Randomization

The application of the Y-Randomization technique is to test the robustness of the model developed. The techniques, the values of the activity (Y) are randomized while the descriptors remained unchanged [2] and used in developing a model. The value of R^2_m parameter, expressed in Eq. 4 [1] measures the disparity in the squared mean correlation coefficient values of the randomized model, R^2_{rand} from the values of the squared correlation coefficient of the non-random model (R^2).

$$R^2_m = R^2 X \sqrt{(R^2 - R^2_r)} \quad (4)$$

where R^2_m represents the squared coefficient of regression of the randomized activity, R^2 , the squared coefficient of

regression in non-randomized activity, and R^2_r , the average value of R^2 .

2.5 Model applicability domain

The ability of the developed QSAR model to make an excellent prediction of the activity of compounds is greatly a function of the applicability domain. The model predicts the activity of the compounds under its domain more correctly than those outside the domain [20]. Building the model applicability domain involves plotting the leverages of each compound against their respective standardized residuals. The diagonals of the hat matrix, $H_i = X_i(X_i^T X_i)^{-1} X_i^T$ produces the leverages for each of the compounds [12], with H_i , the training/test hat matrix, X_i as the initial matrix of test/training set, and X_i^T as the training/test set transpose matrix. The domain has warning leverage, $h^* = 3(t + 1)/z$, where z and t stand for the training set and model descriptors counts respectively. Beyond the warning leverage, a compound is regarded as an outlier and not reliably predicted by the model.

2.6 Mean effect (MF)

The relative contributions of each molecular descriptor to the development of a model are analyzed through the mean effect (MF) estimation. The size and sign of the mean effect respectively show the relevance and the direction of descriptors effect in the developed model. The mean effect can be evaluated using Eq. 5.

$$\text{Mean Effect} = \beta_j \sum_i^n D_j / \sum_j^m \left(\beta_j \sum_i^n D_j \right) \quad (5)$$

where β_j stands for the coefficient of j , D_j stands for the training set value of the matrix descriptors and m , the descriptors sum in the model, and n is the training set count [12].

2.7 Molecular design

The careful analyzes of the molecular features play a significant role in designing compounds with improved anti-malarial activities. Structural modifications of the molecule with the highest activity (template) were carried out by substitutions of groups with others using the information retrieved from the mean effect analysis, thereby designing several derivatives. The theoretical activities of the designed derivatives were estimated after optimizing the derivatives and their molecular descriptors calculated.

Table 2 Regression equations with their statistical validation parameters

S/N	Model	R ²	R ² _{adj}	Q ² _{cv}	LOF	R ² _{pred}
A	$pEC_{50} = -0.000119092 * ATSC0m - 2.56091 * GATS6m - 3.36157 * GATS8i + 7.47634 * SpMin1_Bhi + 0.12704 * maxHBint9 - 0.0926707 * TDB8s - 0.125462 * RDF45m + 0.676907$	0.8881	0.8591	0.8014	0.1303	0.7531
B	$pEC_{50} = 2.88412 * MATS8p - 9.70191 * SpMax2_Bhv + 6.75189 * SpMin1_Bhi - 0.332489 * C1SP2 + 0.104861 * maxHBint9 + 0.331642 * TDB6e + 15.8527$	0.8315	0.7954	0.7331	0.1832	0.5030
C ⁺⁺	$pEC_{50} = -2.87386 * ATSC8c - 1.98497 * GATS8i + 9.0232 * SpMin1_Bhi + 80.7891 * JGI10 + 2.1469 * TDB6u - 22.096$	0.7913	0.7553	0.7112	0.2125	0.7650
D	$pEC_{50} = 3.11204 * AATSC1s - 2.45823 * GATS8i + 9.64979 * SpMin1_Bhi + 0.0817275 * SHBint5 + 2.76508 * TDB6u - 25.9623$	0.7909	0.7549	0.6853	0.2128	0.7197

+ + Best model

Table 3 Contributive descriptors in the selected model, their meaning as well as their classes

Descriptor	Description	Descriptor class	Class
ATSC8c	Centered Broto-Moreau autocorrelation—lag 8 / weighted by charges	Autocorrelation Descriptor	2D
GATS8i	Geary autocorrelation—lag 8 / weighted by first ionization potential	Autocorrelation Descriptor	2D
SpMin1_Bhi	The smallest absolute eigenvalue of Burden modified matrix—n 1 / weighted by relative first ionization potential	Burden Modified Eigenvalues descriptor	2D
JGI10	Mean topological charge index of order 10	Topological Charge Descriptor	2D
TDB6u	3D topological distance-based autocorrelation—lag 6/unweighted	Autocorrelation3D Descriptor	3D

Fig. 1 The predicted against the experimental pEC_{50} values for the training as well as the test sets

3 Results and discussions

3.1 Models developed

The results of the QSAR analysis of 2-anilino 4-amino substituted quinazolines derivatives were presented in Table 2.

Table 4 Correlation matrix of the model descriptors with their respective VIF values

	ATSC8c	GATS8i	SpMin1_Bhi	JGI10	TDB6u	VIF
ATSC8c	1					1.4856
GATS8i	−0.46513	1				1.5381
SpMin1_Bhi	0.279373	−0.01966	1			1.1637
JGI10	0.371194	−0.45268	0.141247	1		1.4169
TDB6u	0.121106	−0.17746	−0.19441	−0.14351	1	1.1670

3.2 Model selection

The statistical parameters of the generated models (Table 2) show the significant and dependable nature of the models. From these, model C was selected as the best model due to its highest external validation, the predictive R^2 ($R^2_{pred} = 0.765$), despite not having the highest coefficient of determination ($R^2 = 0.7913$), and internal validation coefficient ($Q^2 = 0.7112$). The selected model is a five parametric equation with ATSC8c, GATS8i, SpMin1_Bhi, JGI10, and TDB6u as the contributive descriptors, whose definitions are provided in Table 3.

The plots of the predicted antimalarial activities of the 2-anilino 4-amino substituted quinazolines derivatives against their experimental activities (Fig. 1), shows the linearity of the data set around the training set legendary line. The correlation matrix of the selected model descriptors, Table 4, shows low values of the correlation coefficient between descriptors which may indicate a lack of collinearity among the descriptors. Further analysis of the selected model ensured that the determined variation

inflation factors (VIF), which study the collinearity existing between the descriptors revealed that the VIF values of the model descriptors all fall below 2 (Table 4). These low VIF values which ensure the orthogonality among the descriptors were found to be within the allowed VIF range of $1 < \text{VIF} \leq 5$ [13], [18], hence indicates the model acceptability.

3.3 Model validation

Model C, selected to be the best model was internally as well as externally validated. The internal validation involves the determination of both the leave-one-out (LOO) and leave-5-out (L5O) cross-validations, as well as the Y-randomization. The LOO and L5O cross-validation produced correlation coefficients, $R^2 = 0.7144$ and 0.6754 respectively. The high values of the results are an indication that the model does not occur by chance and can predict reasonably. The results of the Y-randomization after 10 different randomization trials, show the R^2 and Q^2 to have low values compared to those of the original

model as reflected in Table 5, which shows the robustness of the model. In the external validation, the square of the regression coefficient (coefficient of determination) was calculated to be $R^2_{\text{pred}} = 0.765$, which is again greater than the minimum value for model acceptance.

3.4 Model applicability domain (AD)

The end-use of a developed QSAR model is for property prediction purposes and will predict effectively only compounds found within the model applicability domain. The applicability domain, Fig. 2, shows that all the data set were found to fall the domain with no outliers except compound 10 having the standardized residual greater than 3σ . The threshold (h^*) of the model was calculated to be $h^* = 0.514$, and no compound was found beyond the threshold, which points to the model strength of prediction.

3.5 Descriptors mean effect

The percentage contribution of the model descriptors towards the antimalarial activity was determined and the results were displaced in a chart presented in Fig. 3. The first descriptor in the model is ATSC8c, a centered Broto-Moreau autocorrelation—lag 8/weighted by charges contributes 39% to the activity. The mean effect shows the descriptor to have a positive charge, indicating an increase in activity with an increase in the descriptor. The second descriptor, GATS8i, Geary autocorrelation—lag 8/weighted by first ionization potential has about 6% contribution to the activity, and since it is negatively charged, decreasing the descriptor value, increases the activity of the compound. Descriptor SpMin1_Bhi (smallest absolute eigenvalue of Burden modified matrix – n 1/weighted by relative first ionization potential) was found to contribute more (54%) to the antimalarial activity and is positively charged, hence, increasing the descriptor values increases the activity. The next descriptor determines charge transference between two atoms separated by 10 bonds [7], JGI10, with 1% contributions, increases the activity by increasing its values. The last descriptor is the positively charged TDB6u that belongs to the TDB descriptors. These descriptors sum up the products of bonds found between two atoms i and j [14]. Having a positive mean effect, the activity increases with increasing the value of the descriptor.

3.6 Molecular design

The structural modification of compound 13, N4-benzyl-N2-(4-fluorophenyl)-6,7-dimethoxyquinazoline-2,4-diamine, Fig. 4 the most active compound ($\text{pEC}_{50} = 7.387$) as the template could be exploited in the design of several

Table 5 Results of Y-randomization performed

Model	R	R^2	Q^2
Initial	0.8895	0.7913	0.7112
Random 1	0.4437	0.1969	-0.2904
Random 2	0.3341	0.1117	-0.1803
Random 3	0.4313	0.1860	-0.0931
Random 4	0.3489	0.1217	-0.3059
Random 5	0.3028	0.0917	-0.6304
Random 6	0.5071	0.2572	-0.0397
Random 7	0.5470	0.2992	-0.0428
Random 8	0.2997	0.0898	-0.3968
Random 9	0.3462	0.1199	-0.1895
Random 10	0.2525	0.0638	-0.2356

Fig. 2 The graph of standardized residuals of the data set against their leverage values

Fig. 3 Descriptors Mean Effect 3-D Pie

derivatives with improved antimalarial activities. The modification was done using the most contributive descriptor, SpMin1_Bhi by increasing the first ionization potentials of the template. The first ionization potentials of a substituted system are increased by electron attracting groups such as $-\text{CN}$, $-\text{CF}_3$, $-\text{COCH}_3$, $-\text{Cl}$, $-\text{F}$, $-\text{I}$, $-\text{COOH}$, $-\text{COCl}_2$, $-\text{CN}$, etc. substituting these electron attracting groups at the meta positions of the substituted compounds produce derivatives of the template with improved malarial activities. Ten (10) theoretical derivatives of the template with improved theoretical activities were designed as shown in Table 6. The activities of five (5) of such compounds (3, 4, 5, 6, and 8) were found to be better than that of chloroquine standard with compound 3, N4-(3-bromo-5-fluorobenzyl)-N2-(4-fluorophenyl)-6,7-dimethoxyquinazoline-2,4-diamine as the most active of the theoretical compounds.

Fig. 4 Design template, N4-benzyl-N2-(4-fluorophenyl)-6, 7-dimethoxyquinazoline-2,4-diamine ($\text{pEC}_{50} = 7.3870$)

3.7 Conclusion

This research aims to design enhanced antimalarial derivatives of 2-Anilino 4-Amino Substituted Quinazolines from a developed activity model. Molecular descriptors ATSC8c, GATS8i, SpMin1_Bhi, JGI10, and TDB6u were discovered to be significant to the antimalarial property of the compounds. The mean effect revealed SpMin1_Bhi, as the most influential descriptor and was essential in designing the ten (10) by-products through structural modifications of the template N4-benzyl-N2-(4-fluorophenyl)-6,7-dimethoxyquinazoline-2,4-diamine. The derivatives show better activities than that of the design template, with compounds 3, 4, 5, 6, and 8 showing even better activities than the chloroquine standard drug. When clinically validated, these compounds may pave way for more potent malarial inhibitors.

Table 6 Structures of the template, its designed derivatives, and their theoretical activities

S/N	R_1	R_2	R_3	Theoretical Activity	
				EC_{50} (μM)	pEC_{50}
1	H	CF_3	H	0.0312	7.5063
2	H	NO_2	CHO	0.0375	7.4259
3	Br	F	H	0.0089	8.0515
4	Cl	F	H	0.0090	8.0466
5	Br	NO_2	H	0.0190	7.7202
6	F	NO_2	H	0.0174	7.7592
7	H	CN	H	0.0357	7.4470
8	H	F	H	0.0185	7.7327
9	H	Br	H	0.0253	7.5961
10	H	I	H	0.0254	7.5957
Template	H	H	H	0.0410	7.3870
Chloroquine				0.0230	7.6383

Funding The authors for this research did not receive any funding concerning this research.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Human and animal rights No human or animal subjects were involved in this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ambure P, Roy K (2017) CADD Modeling of multi-target drugs against Alzheimer's disease. *Curr Drug Targets* 18(5):522–533. <https://doi.org/10.2174/1389450116666150907104855>
2. Arthur DE, Uzairu A, Mamza P, Stephen AE, Shallangwa G (2016) Quantum modeling of the structure-activity and toxicity relationship studies of some potent compounds on SR leukemia cell line. *Chem Data Collect* 5–6:46–61. <https://doi.org/10.1016/j.cdc.2016.10.004>
3. Azmi HF, Lhaksmana KM, Kurniawan I (2020) QSAR study of fusidic acid derivative as anti-malaria agents by using artificial neural network-genetic algorithm. 2020 8th international conference on information and communication technology (ICICT). doi:<https://doi.org/10.1109/icoict49345.2020.9166158>
4. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, Garraway LA (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391):603–607. <https://doi.org/10.1038/nature11003>
5. Chakraborty A (2016) Emerging drug resistance in *Plasmodium falciparum*: a review of well-characterized drug targets for novel antimalarial chemotherapy. *Asian Pac J Trop Dis* 6(7):581–588. [https://doi.org/10.1016/s2222-1808\(16\)61090-3](https://doi.org/10.1016/s2222-1808(16)61090-3)
6. Da Silva AP, de Angelo RM, de Paula H, Honório KM, da Silva ABF (2020) Drug design of new 5-HT₆ antagonists: a QSAR study of arylsulfonamide derivatives. *Struct Chem*. <https://doi.org/10.1007/s11224-020-01513-z>
7. Freitas HF, Barros TF, Castilho MS (2013) 2D Chemometric studies of a series of Azole derivatives active against Fluconazole-Resistant *Cryptococcus gattii*. *J Braz Chem Soc*. <https://doi.org/10.5935/0103-5053.20130122>
8. Gilson PR, Tan C, Jarman KE, Lowes KN, Curtis JM, Nguyen W, Sleebs BE (2017) Optimization of 2-Anilino 4-Amino substituted Quinazolines into potent antimalarial agents with oral in vivo activity. *J Med Chem* 60(3):1171–1188. <https://doi.org/10.1021/acs.jmedchem.6b01673>
9. Habibi-Yangjeh A, Pourbasheer E, Danandeh-Jenagharad M (2008) Application of principal component-genetic algorithm-artificial neural network for prediction acidity constant of various nitrogen-containing compounds in water. *Monatshefte Für Chemie Chem Month* 140(1):15–27. <https://doi.org/10.1007/s00706-008-0049-7>
10. 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*. <https://doi.org/10.1039/c9nj05767f>
11. Hammoudi NEH, Benguerba Y, Sobhi W (2019) QSAR Modeling of thirty active compounds for the inhibition of the Acetylcholinesterase enzyme. *Curr Res Bioinform* 8(1):62–65. <https://doi.org/10.3844/ajbsp.2019.62.65>
12. Ibrahim ZY, Uzairu A, Shallangwa G, Abechi S (2020) In-silico design of Aryl and Aralkyl Amine-Based Triazolopyrimidine derivatives with enhanced activity against resistant *Plasmodium falciparum*. *Chem Africa*. <https://doi.org/10.1007/s42250-020-00199-4>
13. Kim JH (2019) Multicollinearity and misleading statistical results. *Korean J Anesthesiol* 72(6):558–569. <https://doi.org/10.4097/kja.19087>
14. Klein CT, Kaiser D, Ecker G (2004) Topological distance based 3D descriptors for use in QSAR and diversity analysis. *J Chem Inf Comput Sci* 44(1):200–209. <https://doi.org/10.1021/ci0256236>
15. Ouattara A, Laurens MB (2015) Vaccines against malaria. *Clin Infect Dis An Off Publ Infect Dis Soc Am* 60(6):930–936. <https://doi.org/10.1093/cid/ciu954>
16. Panagos P, Meusburger K, Ballabio C, Borrelli P, Alewell C (2014) Soil erodibility in Europe: a high-resolution dataset based on LUCAS. *Sci Total Environ* 479–480:189–200. <https://doi.org/10.1016/j.scitotenv.2014.02.010>
17. Pourbasheer E, Aalizadeh R, Ardabili JS, Ganjali MR (2015) QSPR study on solubility of some fullerenes derivatives using the genetic algorithms — Multiple linear regression. *J Mol Liq* 204:162–169. <https://doi.org/10.1016/j.molliq.2015.01.028>
18. Senaviratna NAMR, Cooray TMJA (2019) Diagnosing multicollinearity of logistic regression model. *Asian J Prob Stat* 5(2):1–9
19. Srivastav VK, Tiwari M (2017) QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors. *Arab J Chem* 10:S1081–S1094. <https://doi.org/10.1016/j.arabjc.2013.01.015>
20. Tropsha A, Gramatica P, Gombar V (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb Sci* 22(1):69–77. <https://doi.org/10.1002/qsar.200390007>
21. Welles TE, Plowe CV (2001) Chloroquine-resistant malaria. *J Infect Dis* 184(6):770–776. <https://doi.org/10.1086/322858>
22. WHO (2017) World malaria report, Geneva Switzerland

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.