ORIGINAL ARTICLE

In-silico Design of Aryl and Aralkyl Amine-Based Triazolopyrimidine Derivatives with Enhanced Activity Against Resistant *Plasmodium falciparum*

Zakari Ya'u Ibrahim[1] · **Adamu Uzairu[1]** · **Gideon Shallangwa[1]** · **Stephen Abechi[1]**

Abstract

A blend of genetic algorithm with multiple linear regression (GA-MLR) method was utilized in generating a quantitative structure–activity relationship (QSAR) model on the antimalarial activity of aryl and aralkyl amine-based triazolopyrimidine derivatives. The structures of derivatives were optimized using density functional theory (DFT) DFT/B3LYP/6–31+G* basis set to generate their molecular descriptors, where two (2) predictive models were developed with the aid of these descriptors. The model with an excellent statistical parameters; high coefficient of determination (R^2) = 0.8884, cross-validated R^2 (Q^2cv) = 0.8317 and highest external validated R^2 (R^2_{pred}) = 0.7019 was selected as the best model. The model generated was validated through internal (leave-one-out (LOO) cross-validation), external test set, and Y-randomization test. These parameters are indicators of robustness, excellent prediction, and validity of the selected model. The most relevant descriptor to the antimalarial activity in the model was found to be GATS6p (Geary autocorrelation—lag 6/weighted by polarizabilities), in the model due to its highest mean effect. The descriptor (GATS6p) was significant in the *in-silico* design of sixteen (16) derivatives of aryl and aralkyl amine-based triazolopyrimidine adopting compound DSM191 with the highest activity (pEC_{50} = 7.1805) as the design template. The design compound D8 was found to be the most active compound due to its superior hypothetical activity (pEC_{50} = 8.9545).

Keywords In silico design · QSAR · Aryl and aralkyl amine-based triazolopyrimidine · *Plasmodium falciparum* · GA-MLR

1 Introduction

Despite decades of efforts to curb and eradicate the malaria pandemic, it still retains its status as one of the most dangerous and fatal infections in the tropical and subtropical endemic countries, having over 500 million reported cases with approximately one million deaths yearly [1]. The disease is caused by *Plasmodium falciparum* (*P. falciparum*), the most lethal of all the five *Plasmodium* species, others are; Plasmodium malariae (*P. malariae*), Plasmodium knowlesi (*P. knowlesi*), Plasmodium falciparum (*P. falciparum*), Plasmodium vivax (*P. vivax*), and Plasmodium ovale (*P. ovale*) [2–4]. The protozoan parasite *Plasmodium falciparum* produced in the red blood cell (RBCs) requires various hosts for its growth and survival. The host is unable to manufacture purine rings de novo, hence purine is reclaimed from the host by the parasite [5]. Within the digestive vacuole of the parasite, the hemoglobin releases heme molecules that are toxic to the parasite. The heme is crystallized into hemozoin to nullify its toxicity. A large amount of hemoglobin is accumulated in the digestive vacuole [6]. The rate of heme release must equal to that of hemozoin crystallization for the survival of the parasite otherwise toxic heme buildup in the parasite [7].

Several compounds including chloroquine, artemisinin, amodiaquine, and quinine have been reported to have antimalarial activity and are employed in the treatment of malaria [8]. The treatment of malaria is faced with great challenges which include the rapid resistance developed by *Plasmodium* parasites to the available antimalarial drugs as well as the lack of an effective antimalarial vaccine. This resistance ensures that the available antimalarial drugs become ineffective in the treatment of pandemic disease. Hence, the need to develop new and novel antimalarial

✉ Zakari Ya'u Ibrahim
  zakariyyadibrahim@gmail.com

[1] Department of Chemistry, Faculty of Physical Sciences, Ahmadu Bello University, P.M.B 1045, Zaria, Nigeria

agents through chemical modification of existing compounds becomes a necessity [9].

Triazolopyrimidines identified with either lack of potency or poor plasma exposure due to metabolic stability were found to have improved potency as a result of the substitution of arylamine moiety [10]. In light of this, aryl and aralkyl amine-based triazolopyrimidine compounds reported for their antimalarial activity against the multi-drug resistant *Plasmodium falciparum* strain, 3D7 [10] could provide an alternative application to the orthodox antimalarial drugs.

Quantitative structure–activity relationship (QSAR) is an essential tool in the drug design field by reducing the time as well as the cost of drug development [11]. QSAR relates the physicochemical properties of congeneric compounds with the structural features responsible for their activity [12].

Several QSAR studies related to the design of antimalarial drugs were reported. In 2018, Hadanu and co. investigated the antimalarial activity of nitrobenzothiazole derivatives [13]. Beheshti and Co study the modeling of antimalarial activity of urea derivatives [14]. Prediction of anti-malarial activity in a set of new imidazolopiperazines using artificial neural networks was reported [9], but no systematic QSAR studies were conducted on a series of aryl and aralkyl amine-based triazolopyrimidine compounds as an antimalarial. This research is aimed at establishing a virgin QSAR regression model for predicting the antimalarial activity of 32 derivatives of aryl and aralkyl amine-based triazolopyrimidine compounds through the application of a hybrid of GA–MLR approach. Thereafter, we performed in silico design of a template (a compound with the highest activity), into new derivatives with enhanced activity against *P. falciparum* with the aid of the most contributive molecular descriptor determined by the descriptors mean effect, which saves time and cost enquired in the drug development analysis.

2 Material and Methods

2.1 Sources of Data

The dataset of thirty-two aryl and aralkyl amine-based triazolopyrimidine derivatives consisting of their structures and activities (Table 1) were extracted from the literature [15]. Their biological activities were expressed as pEC$_{50}$ (-Log$_{10}$ drug concentration to cause a 50% reduction in cell proliferation) suitable for QSAR analysis (Tables 2 and 3).

2.2 Geometry Optimization

The structures of the compounds were drawn with the help of ChemDraw software, while Spartan 14 software (Spartan 14v114) was utilized in the optimization of the

molecular geometries [16] based on the density functional theory (DFT) level using Becke's three parameters Lee–Yang–Parr hybrid functional (B3LYP) in combination with the 6–31 + G* basis set. The optimized Spartan 14 structures are saved in SDF format.

2.3 Descriptors Calculation

The lowest energy-optimized geometries were imported into the PaDel-Descriptor version 2.20 for descriptors calculation. About 1500 varieties of molecular descriptors were calculated for each molecular geometry present in Table 1. The descriptors range from Atom Count, Autocorrelation, BaryszMatrix, BCUT, Burden Modified Eigenvalues, Carbon Types, Electrotopological State Atom Type, Extended Topochemical Atom, Autocorrelation3D, and RDF descriptors were calculated [17].

2.4 Dataset Division

The dataset was split in the ratio 78:22 with 25 training compounds (78% of the data set) and 7 test compounds (covering the remaining 22%) with the aid of the DatasetDivision1.2 program [18] by employing the Kennard-Stone's algorithm method.

2.5 QSAR Model Development

The training set was utilized in the development of the QSAR model, where the activities (dependent variable) and the molecular descriptors (independent variables) were exposed to model development by the Genetic Function Approximation (GFA) method contained in the material studio software. The Genetic algorithms (GAs), a copycat of biological evolution, are currently the ultimately used variable selection method that addresses both the constrained and unconstrained optimization challenges based on an essential selection process [19, 20]. GA is a prying search method that looks for the correct or approximate solutions to any optimization challenges [21]. Its ability to produce more than one blend of descriptors useable in model construction is an added advantage. With GAs, users have the liberty of imposing the equation length and employs a lack of fit (LOF) function to check over-fitting in the models produced. The lack-of-fit (LOF) was calculated in the material studio to estimate the model fitness using Eq. 1.

$$LOF = SSE / \left(1 - {}^{c+dp/M}\right)^2 \tag{1}$$

where SSE stands for the sum of squares of errors, c, number of terms other than the constant in the model, d stands for the user-defined smoothing parameter, p is the number

Table 1 Molecular structures of substituted aryl amine-based triazolopyrimidine derivatives and their biological activities against *P. falciparum* strain 3D7
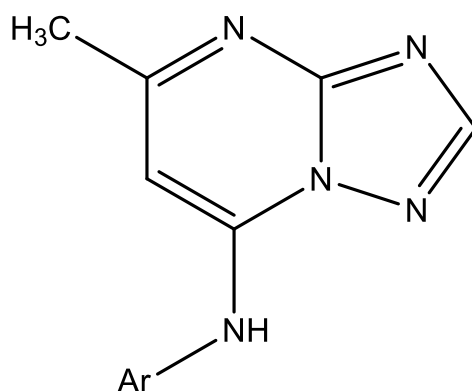
Compd	Name	Ar	EC$_{50}$ (µM)	pEC$_{50}$	Predicted pEC$_{50}$	Residuals
1	DSM101	4-ethyl-Ph	3.3	5.481	5.673	−0.192
2	DSM102	4- n-propyl-Ph	6.8	5.167	5.264	−0.097
3*	DSM104	4- n-butyl-Ph	4.8	5.319	4.466	0.853
4	DSM103	4-iso-propyl-Ph	0.78	6.108	5.776	0.332
5	DSM316	4-cyclo-propyl-Ph	1.4	5.854	5.812	0.042
6*	DSM115	4-venyl-Ph	0.34	6.469	5.802	0.667
7*	DSM157	4-ethynyl-Ph	5.1	5.292	5.761	−0.468
8	DSM217	4-S-CF$_3$-Ph	0.27	6.569	6.599	−0.030
9	DSM161	4-SF$_5$-Ph	1.3	5.886	5.861	0.025
10	DSM130	4-N(CH$_3$)$_2$-Ph	2.7	5.569	5.353	0.215
11	DSM251	4-C(CH$_3$)$_2$CN-Ph	3.7	5.432	5.808	−0.376
12*	DSM228	3-F-4-C(CH$_3$)$_3$-Ph	0.097	7.013	8.339	−1.325
13	DSM215	3-F-4-Cl-Ph	3.5	5.456	5.435	0.021
14	DSM238	3,4-di-CF$_3$-Ph	2.3	5.638	5.786	−0.147
15	DSM190	3,5-di-F-4-CF$_3$-Ph	1.1	5.959	5.655	0.303
16	DSM159	5,6,7,8-tetrahydro-2-naphthyl	0.16	6.796	6.429	0.367
17	DSM160	5-indanyl	0.8	6.097	6.466	−0.369
18	DSM127	5-benzodioxolyl	21	4.678	4.791	−0.113
19	DSM191	5-benzothiophenyl	0.066	7.180	6.962	0.218
20	DSM86	2-CF3-5-benzo- imidazolyl	25	4.602	4.486	0.116

NB: *Test Sets

of descriptors in the model, and M, represent the samples counts in the training set.

2.6 Multi-co-linearity analysis

The possibility of a huge degree of correlation among the model descriptors was evaluated using a variance inflation factor (VIF). For a descriptor, i, the VIF$_i$ was calculated from Eq. 2.

$$VIF_i = \left(1 - R_{ij}^2\right)^{-1} \qquad (2)$$

where R$_{ij}^2$ represents the multiple regression's correlation coefficients between the descriptor i and the remaining j model descriptors.

2.7 Model Validation

2.7.1 Internal Validation

The leave- one- out (LOO) cross-validation technique was deployed to internally validate the model. It involves the elimination of one compound from the data set at random in each cycle and the model is built with the remaining compounds. The constructed model is used for predicting the activity of

Table 2 Molecular structures of alicyclic/aralkyl and alkyl amine-based triazolopyrimidine derivatives and their biological activities against *P. falciparum* strain 3D7
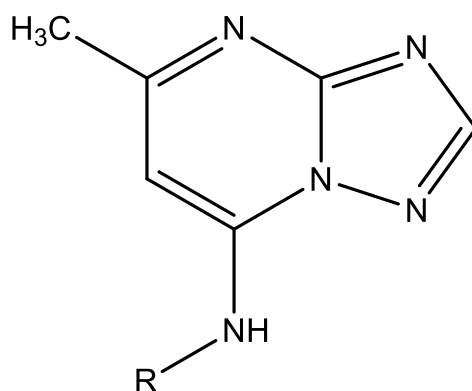
Compd	Name	R	EC$_{50}$ (µM)	pEC$_{50}$	Predicted pEC$_{50}$	Residuals
21	DSM162	2-indanyl	0.4	6.398	6.353	0.045
22	DSM218	5-F-2-indanyl	1.3	5.886	5.645	0.241
23	DSM220	5-Br-2-indanyl	6.6	5.180	5.280	−0.100
24	DSM219	5,6-di-OCH$_3$-2-indanyl	7.7	5.114	4.954	0.160
25	DSM192	1,2,3,4-tetrahydro-2-naphthyl	1.8	5.745	6.126	−0.382
26	DSM236	6-F-1,2,3,4-tetra-hydro-2-naphthyl	2.5	5.602	5.646	−0.044
27	DSM237	6-Cl-1,2,3,4-tetra-hydro-2-naphthyl	10	5.000	4.999	0.001
28*	DSM77	3-phenylpropyl	25	4.602	4.880	−0.278
29	DSM76	n-heptyl	13	4.886	5.000	−0.114

NB: *Test Sets

Table 3 Molecular structures of pyrimidine ring-modified triazolopyrimidine derivatives and their biological activities against *P. falciparum* strain 3D7
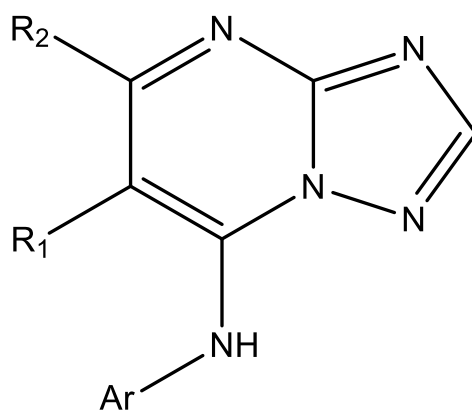
Compd	Name	R$_1$/R$_2$	Ar	EC$_{50}$ (µM)	pEC$_{50}$	Predicted pEC$_{50}$	Residuals
30*	DSM247	C$_2$H$_5$/H	4-CF$_3$-Ph	2.3	5.638	5.636	0.002
31	DSM252	Cl/H	4-CF$_3$-Ph	3.4	5.469	5.593	−0.124
32*	DSM174	-C$_3$H$_6$-	3-F-4-CH$_3$-Ph	8.3	5.081	4.978	0.103

NB: *Test Sets

the compound eliminated. This process is carried out again and again until the entire compounds are eliminated once. The cross-validated squared coefficient of correlation, R^2_{cv} (Q^2) was calculated from the equation below.

$$Q^2 = 1 - \frac{\sum \left(Y_{obs} - Y_{pred}\right)^2}{\sum \left(Y_{obs} - \hat{Y}\right)^2} \tag{3}$$

where Yobs is the observed activity of the training set, Ypred represents the predicted activity of the training set and Ŷ is the mean observed activity of the training set.

2.7.2 External Validation

The developed model was externally validated to determine its predictive strength, through predicting the activity of some dataset (test set) separated from the training set. This is achieved using Eq. 4 below.

$$R^2_{pred} = 1 - \frac{\sum \left(Y_{obs(test)} - Y_{pred(test)}\right)^2}{\sum \left(Y_{obs(test)} - \hat{Y}_{training}\right)^2} \tag{4}$$

where $Y_{obs\ (test)}$ and $Y_{(test)}$ indicates the test set predicted and observed activity values respectively, $\hat{Y}_{training}$, indicates training set means activity. R^2_{pred} is the predicted coefficient of correlation estimated from the predicted test set activity.

2.8 Y-Randomization

The possibility of the model occurring by chance was determined through the Y-randomization test, which also ascertains the robustness of the developed model. The Y-Randomization test involves constructing a model with a randomizing the activities of the training set. A detour in the square of the mean correlation coefficient of the randomized activity model (R^2r) from that of the non-randomized activity model (R^2) is reflected in the value of R^2m parameter calculated with help of the Eq. 5 [22].

$$R^2_m = R^2 X \sqrt{\left(R^2 - R^2_r\right)} \tag{5}$$

Normally, the average value of R^2r for the randomized activity model is expected to be zero, implying equal values of R^2m and R^2r for the developed model, necessitate the use of cRp^2 in place of R^2m. A model is deemed not due to a chance when the coefficient of determination (cRp^2) is found to be greater than 0.5 [23]. The value of cRp^2 was calculated for each randomized activity model with the aid of Eq. 6 given below.

$$cR^2_p = RX \sqrt{\left(R^2 - R^2_r\right)} \tag{6}$$

2.9 Model Applicability Domain

The applicability domain (AD) provides an assessment of the model in terms of its ability to predict the activity of unknown compounds excellently in form of chemical space. Every model is only able to predict the activity of a certain number of compounds; such prediction is conditioned on the satisfaction of the assumptions applied in the construction of the model [24]. The chemical space of the AD is designed by plotting the values of leverage against the standardized residuals (Williams plot) of the entire data set. The leverages are obtained from the diagonals of a hat matrix obtainable from Eq. 7 [17].

$$H_i = X_i \left(X^T_i X_i\right)^{-1} X^T_i \tag{7}$$

where, H_i, is the hat matrix of the training/test set, X_i is the original matrix of training/test set, and X^T_i is the transpose matrix of the training/test set.

Being a prediction tool, the AD has warning leverage (h*) defining the limit of normal values for X outliers and it's defined as $h^* = 3(z + 1)/n$, where n is the number of the training set, and z is the number of descriptors in the model. Only compounds with leverages less than or equal to the warning (h*) are considered to be within the chemical space and their activities are reliably predicted by the model. In addition to the AD, a plot of the normalized mean distance of the model descriptors against the standardized residuals of the predicted activities of the data set (UZAIRU'S plot) was employed to detect outliers present in the data set [25].

2.10 Mean Effect (MF)

The model descriptor's relevance was analyzed by evaluating the descriptors mean effect (MF). The contribution by individual descriptor to the model are reflected by the size as well as the signs of their mean effect. The most contributive descriptor has the largest mean effect, while the small mean effect size is for the least contributive descriptor. The mean effect is evaluated from Eq. 7.

$$Mean\ Effect = \frac{\beta_j \sum_i^n D_j}{\sum_j^m \left(\beta_j \sum_i^n D_j\right)} \tag{8}$$

where β_j is the coefficient of j, D_j is the value of each training set in the descriptor matrix and m is the tally of model descriptors and n is the tally of molecules used as the training set [26].

2.11 Molecular Design

The relevance of the model's descriptors cannot be overemphasized as such the interpretations of the most contributive descriptor could provide a guide to design several hypothetical derivatives of aryl and aralkyl amine-based triazolopyrimidine having enhanced antimalarial activities. This is made possible by selecting the compound with the highest activity as a design template. The structural feature of the most informative descriptor forms the basis of manipulating the design template by substituting group(s) based on the information from the size of the descriptors mean effect, to design some theoretical compound possessing higher antimalarial activity than the design template through replacing atom/group of atoms at different positions on the template. The design derivatives were then optimized and their descriptor calculated from which the antimalarial activity was obtained.

3 Results and Discussion

3.1 QSAR Studies

Two QSAR models developed through the genetic function algorithm analysis of the datasets were as presented in Table 4:

3.2 QSAR Model Selection

The second tetra parametric equation (model II) was announced as the best QSAR model based on its statistical data obtained from multiple linear regression (MLR) of the material studio. The model was used in evaluating the theoretical activity of the aryl and aralkyl amine-based

triazolopyrimidine derivatives used in the analysis. The model may not have the highest R^2 (0.8884) compared to that of the model I (0.8931), but its higher square coefficient of correlation for the test set ($R^2_{ext} = 0.7019$) which is in line with the prerequisite for a dependable, foretelling and robust QSAR model [27] ensures model II emerges as the best model. The definition of the descriptors in the model as well as their class were reported in Table 5.

The meager residual values of the data set (Tables 1, 2, and 3) calculated by the best model, is an indication of a good agreement between the observe and predicted activity as well as the reliability of the model, this was supported by the distribution of the data set around the legend, displayed in the plot of experimental pEC_{50} vs predicted pEC_{50} presented as Fig. 1.

3.3 Model Validation

A significant QSAR model is one that can predict the activities of an external data set, (R^2_{ext}) not just in its ability to clone the training set (Q^2_{LOO}). Hence, the training set prediction was affirmed by the leave-one-out (LOO) cross-validation method. The large Q^2_{LOO} value, 0.8317 calculated is a good indication of an excellent internal validation. The results of the external validation, $R^2_{ext} = 0.7019$, show the predictive strength of the model and hence its robustness. Analysis of the correlation matrix shown in Table 6, shows a large value of correlation coefficients between the descriptors and the activity, an indication of a great interrelation between the two. Whereas the low correlation coefficients between the descriptors themselves indicate noncollinearity between the descriptors. The descriptor's multicollinearity is explained with variation inflation factors (VIF). The VIF values of the model's descriptors as reflected in Table 6, all fall within, $1 < VIF \leq 5$ range, which is a VIF criterion for

Table 4 Generated models with their statistical parameters

S/N	MODEL	R^2	R^2_{adj}	Q^2cv	LOF	R^2_{pred}
I	$pEC_{50} = 2.5963(MATS4e) - 1.87976(GATS5m) - 8.01883(GATS6p) + 0.142531(RDF50p) + 14.0942$	0.8931	0.8717	0.7911	0.1500	0.5106
II*	$pEC_{50} = -2.3661(MATS6i) + 2.1115(GATS5e) - 7.93447(GATS6p) + 0.0926193(RDF30m) + 10.0917$	0.8884	0.8661	0.8317	0.1566	0.7019

*Best model

Table 5 List of the physiochemical descriptors that constitute the model

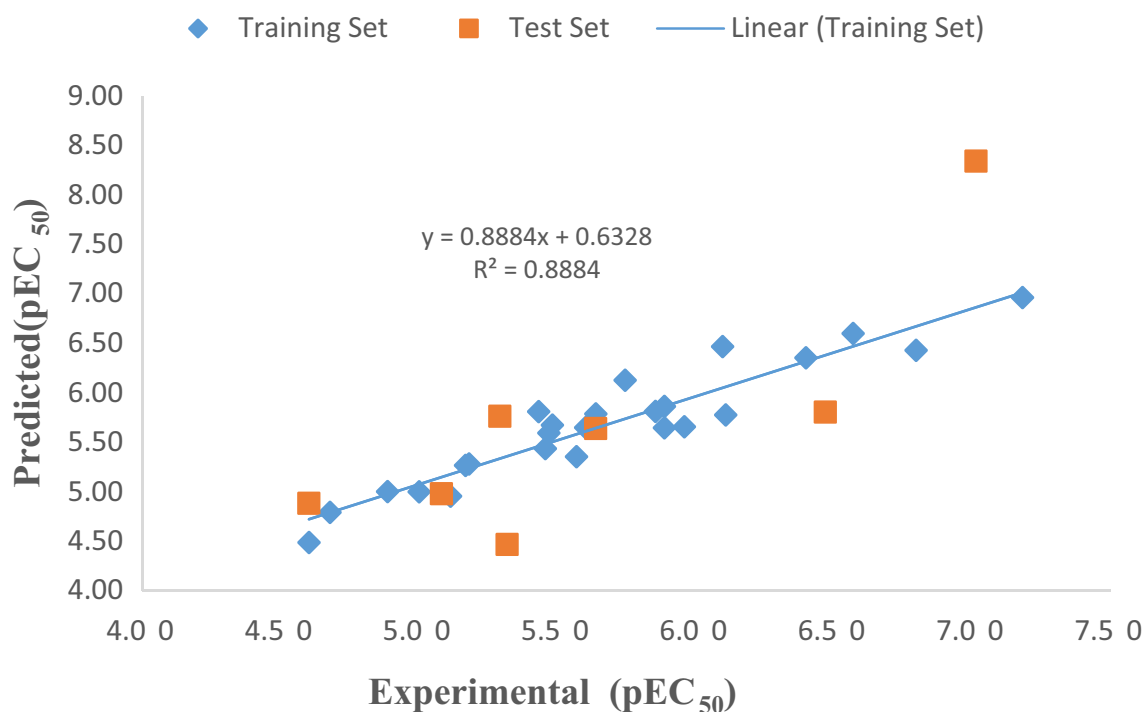| S/N	Symbol	Names of descriptors	Class
1	MATS6i	Moran autocorrelation—lag 6/weighted by first ionization potential	2D
2	GATS5e	Geary autocorrelation—lag 5/weighted by Sanderson electronegativities	2D
3	GATS6p	Geary autocorrelation—lag 6/weighted by polarizabilities	2D
4	RDF30m	Radial distribution function—030/weighted by relative mass	3D

Fig. 1 The predicted pEC$_{50}$ against the experimental values for the data set

Table 6 The correlation matrix of the descriptors and their VIF

	pEC50	MATS6i	GATS5e	GATS6p	RDF30m	VIF
pEC50	1					
MATS6i	0.2912	1				**1.5299**
GATS5e	0.4527	0.0292	1			**1.1060**
GATS6p	−0.6946	−0.5528	0.1164	1		**1.8359**
RDF30m	−0.1496	−0.0638	−0.1797	0.3686	1	**1.2886**

Table 7 Y-randomization model parameters

Average r	0.3974
Average r^2	0.1655
Average Q^2	−0.3099
cRp^2	0.8056

model acceptance [28, 29]. The results of the VIF confers orthogonality and acceptability on the model. The outcome of the Y-randomization inquiry after scrambling the activity ten folds shows that the R^2 and Q^2 values for the scrambling activity model are significantly low, confirming the robustness of the model. Table 7, shows the Y-randomization parameters with cR_p^2 greater than 0.5 reaffirming the model's quality. The standardized residuals plotted against the calculated leverage values (Williams plot) gives a graphical identification of both the outliers and the influential substances within a model. The applicability domain (Fig. 2), is entrenched within ± 3 standardized residual range and

a leverage threshold (h* = 0.6). Figure 2, clearly shows no compound of the data set beyond the standard deviation, ± 3d, and a leverage value, h > 0.6. Hence, the entire data set contain neither an outlier nor an influential substance. Furthermore, the Uzairu's plot (the normalized mean distance of the model descriptors versus the standardized residuals of the predicted activities of the data set) presented in Fig. 3, also confirms the absence of an outlier.

3.4 Interpretation of Descriptors

In this research, the four molecular descriptors chosen by GA play a significant role in the discovery of new drugs for the treatment of malaria. The contributions of each descriptor to the model were carried out in the MF analysis displaced in Fig. 4. The first two descriptors are MATS6i and GATS5e, which are both 2D-autocorrelated descriptors obtained from a molecular graph through the addition of the atomic weight of the terminal atoms of the
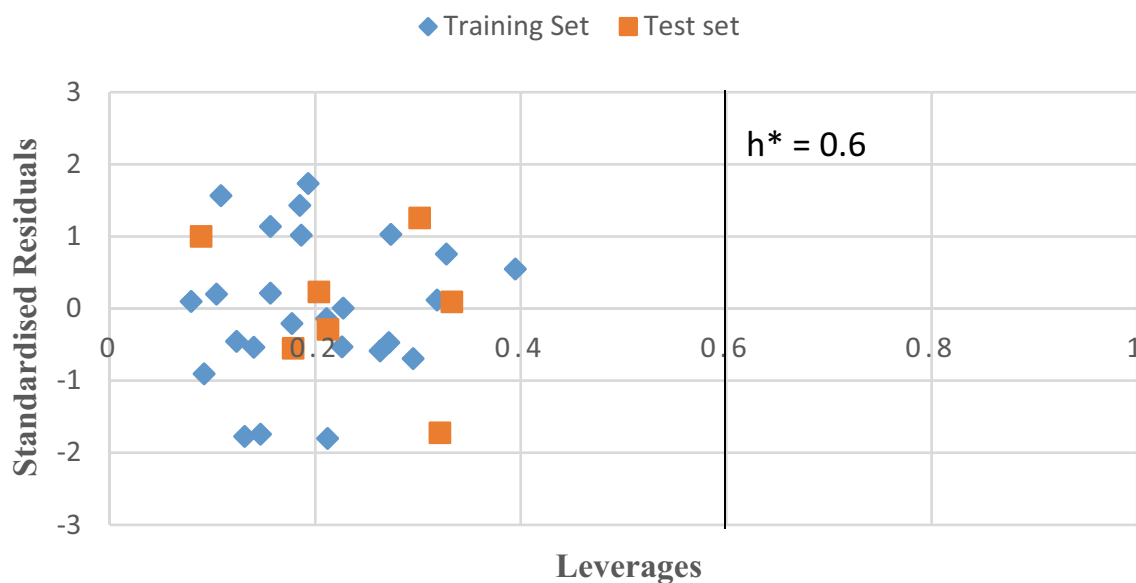
WILLIAMS PLOT

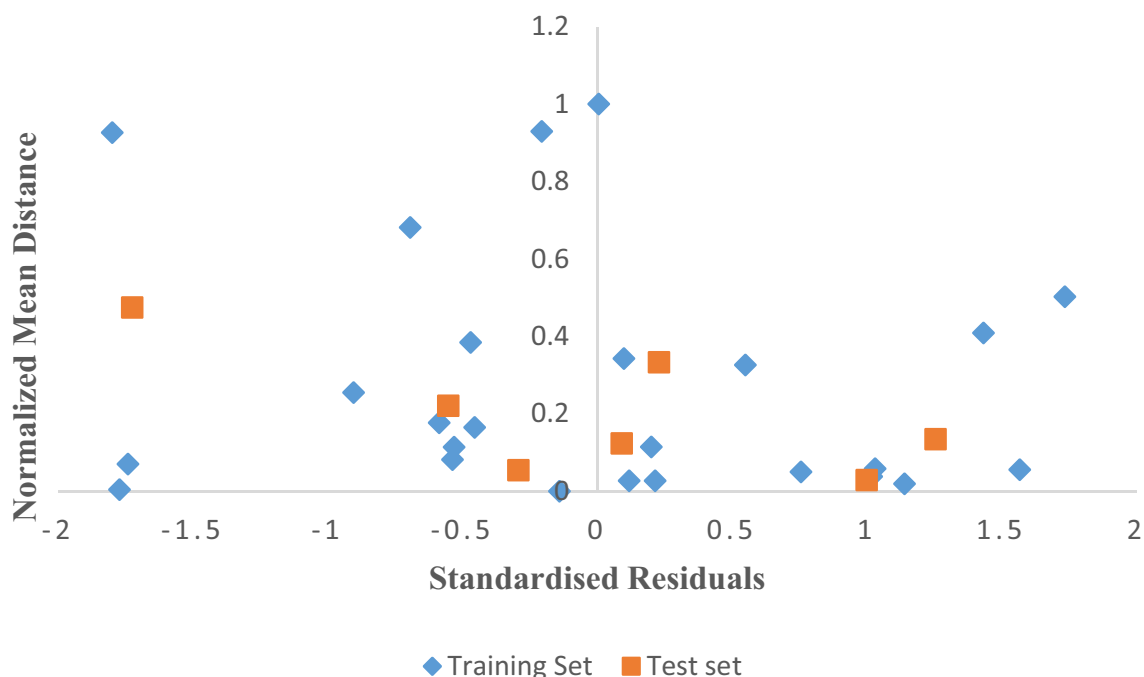**Fig. 2** The plot of the standardized residuals versus the leverage values

Fig. 3 The Uzairu's plot of the normalized mean distance versus the standardized residuals

considered path length. While MATS6i descriptor encodes the ionization potential of the substituted aryl amine-based triazolopyrimidine analogs, it expresses the tendency of an electron cloud or molecule to interact with a target [30], from the MF analysis, the descriptor has the least contribution (MF = 0.00298) to the antimalarial activity and its positive sign is an indication of a direct relationship with the activity. The GATS5e relates to the electronegativity of the compound which has an inverse relation with the activity as a result of the MF sign.
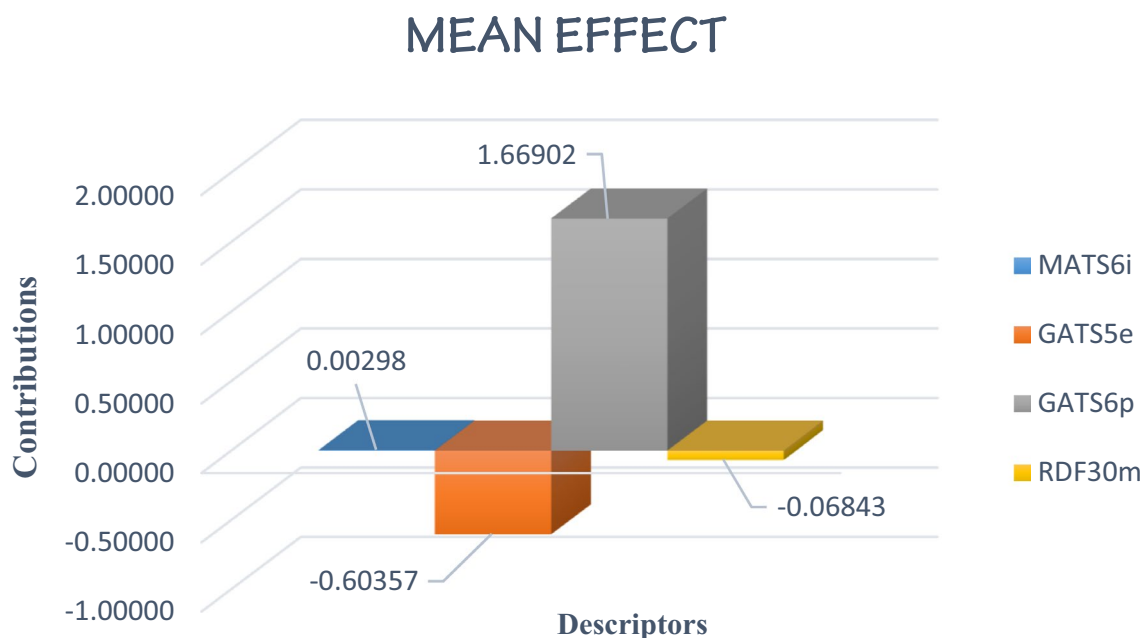
MEAN EFFECT

Fig. 4 3-D clustered column descriptors mean effect

The third descriptor, GATS6p is also a 2D descriptor [31] that reflects the topological structure of the molecules. The difference in the atomic polarizabilities of the molecule is explained by the descriptor [32]. This descriptor is the most relevant with MF = 1.66902, and its positive MF sign indicates increasing the value of the descriptor, increases the antimalarial activity.

The last descriptor, RDF30m indicates the existence of some linear dependence between the inhibitory activities of aryl amine-based triazolopyrimidine derivatives and the 3-dimensional molecular distribution of mass determined at a radius of 4.0 Å from the centers of all molecule. The descriptor's negative MF is a reflection of its indirect relation with the activity. A decrease in the value of the descriptor increases the antimalarial activity.

3.5 Molecular Design

The compound with the highest antimalarial activity, DSM191 (pEC_{50} = 7.1805) was exploited as the design template. The most contributive descriptor, GATS6p (MF = 1.6690) influences the antimalarial activity of the derivatives of aryl and aralkyl amine-based triazolopyrimidine greatly, and relate to the atomic polarizabilities of the molecule. With the positive mean effect, antimalarial activity increases with an increase in the descriptor value (atomic polarizabilities) of the molecule. Atomic polarizability increases with an increase in the number of electrons (leading to a reduction in the effective nuclear charge). Therefore, various electron-donating groups such as; -CH_3,

-CH_2CH_3, -$COCH_3$, -OCH_3, -OH, -NH_2, and -$N(CH_3)_2$ were added to the design template at ortho/para positions to release electron density into the ring thereby creating derivatives of the template which increase the polarizability and by extension, increase the antimalarial activity of the designed derivatives. Sixteen (16) derivatives of the template were designed, out of which nine of the designed compounds (2D, 4D, 8D, 11D, 12D, 13D, 14D, 15D, and 16D) have improved hypothetical antimalarial activity than the design template as well as Chloroquine (pEC_{50} = 4.6352) and Pyrimethamine (pEC_{50} = 6.2593) standards as shown in Table 8. Compound D8, N^7-(benzo[b]thiophen-5-yl)-N^6,N^6,5-trimethyl-[1,2,4] triazolo[1,5-a]pyrimidine-6,7-diamine was identified with the highest antimalarial activity (pEC_{50} = 8.9545) as reflected (Table 8), hence is the most active designed derivative (Fig. 5).

4 Conclusions

This research is focused on designing antimalarial derivatives through the developed QSAR model used for predicting the activities of aryl and aralkyl amine-based triazolopyrimidine derivatives. The PaDel-Descriptor version 2.20 software was employed to calculate several descriptors used for the regression analysis carried conducted by the GA–MLR process and the developed model validated comprehensively. The selected model revealed descriptors; MATS6i, GATS5e, GATS6p, and RDF30m played a significant role in the antimalarial activity of the compounds.

Table 8 The structures of the design template, designed derivatives of aryl and aralkyl amine-based triazolopyrimidine along with their respective theoretical activities
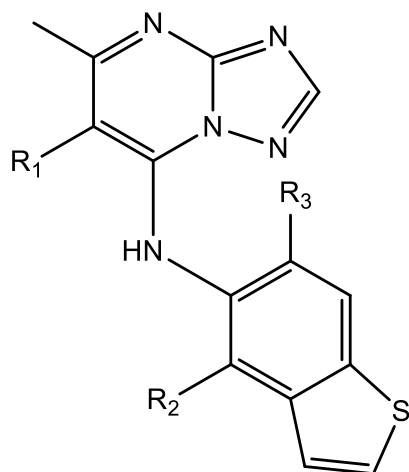
Compd	R_1	R_2	R_3	Theoretical activities	
				EC_{50} (µM)	pEC_{50}
Design template	H	H	H	0.066	7.1805
D1	H	CH_3	H	0.1068	6.9714
D2	CH_2CH_3	H	H	0.0290	7.5373
D3	$OCOCH_3$	H	H	0.4937	6.3065
D4	OCH_3	H	H	0.0231	7.6358
D5	OH	H	H	0.4446	6.3520
D6	H	OH	H	0.2542	6.5948
D7	H	NH_2	H	0.1219	6.9139
D8	$N(CH_3)_2$	H	H	0.0011	8.9545
D9	NH_2	CH_3	H	0.2663	6.5746
D10	OCH_3	H	OH	0.0960	7.0178
D11	$N(CH_3)_2$	CH_3	H	0.0012	8.9186
D12	OCH_3	CH_3	H	0.0231	7.6359
D13	OCH_3	NH_2	H	0.0279	7.5548
D14	$N(CH_3)_2$	NH_2	H	0.0022	8.6498
D15	$N(CH_3)_2$	OH	H	0.0063	8.2020
D16	OCH_3	OH	H	0.0387	7.4125
Pyrimethamine				0.5504	6.2593
Chloroquine				23.1608	4.6352

The mean effect analysis indicated GATS6p as the most contributive descriptor (MF = 1.6690) which estimates the polarizability of the molecule whose positive values increasing the activity of the compounds. Sixteen compounds were designed using DSM191 (a compound with the highest antimalarial activity) as the design template and various electron-donating groups used to increase the atomic polarizability were directed at the ortho/para positions of the aromatic rings of the template. Compound D8, N^7-(benzo[b]thiophen-5-yl)-N^6,N^6,5-trimethyl-[1,2,4] triazolo[1,5-a]pyrimidine-6,7-diamine, was found to have
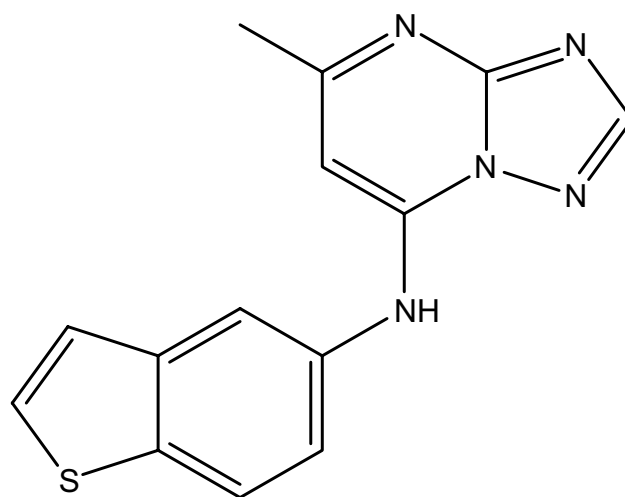

Fig. 5 Design Template, DSM191, N-(benzo[b]thiophen-5-yl)-5-methyl-[1,2,4]triazolo[1,5-a]pyrimidin-7-amine ($pEC_{50} = 7.1805$)

the highest hypothetical activity ($pEC_{50} = 8.9545$) than the template ($pEC_{50} = 7.1805$) and the fellow designed compounds. These designed compounds when synthesized could be clinically validated for antimalarial treatment. Hence, the studies will reduce trial and error as well as minimize the time and cost involved in drug design. Furthermore, in the future, we could conduct the molecular docking studies of the design compounds against their target to observe the nature of interactions responsible for the antimalarial activity.

Data Availability The datasets used for analysis during these studies were included in this published study.

Compliance with Ethical Standards

References

1. Krungkrai SR, Krungkrai J (2016) Insights into the pyrimidine biosynthetic pathway of human malaria parasite *Plasmodium falciparum* as chemotherapeutic target. Asian Pac J Trop Med 9(6):525–534. https://doi.org/10.1016/j.apjtm.2016.04.012

2. Larson B (2019) Origin of two most virulent agents of human malaria: *Plasmodium falciparum* and *Plasmodium vivax*. Malaria. https://doi.org/10.5772/intechopen.84481

3. Yman V, Wandell G, Mutemi DD, Miglar A, Asghar M, Hammar U, Färnert A et al (2019) Persistent transmission of *Plasmodium malariae* and *Plasmodium ovale* species in an area of declining *Plasmodium falciparum* transmission in eastern Tanzania. PLoS Negl Trop Dis 13(5):e0007414. https://doi.org/10.1371/journal.pntd.0007414

4. Haredi Abdelmonsef A, Eldeeb Mohamed M, El-Naggar M, Temairk H, Mohamed Mosallam A (2020) Novel quinazolin-2,4-dione hybrid molecules as possible inhibitors against malaria: synthesis and in silico molecular docking studies. Front Mol Biosci. https://doi.org/10.3389/fmolb.2020.00105

5. Tougan T, Edula JR, Morita M, Takashima E, Honma H, Tsuboi T, Horii T (2020) The malaria parasite *Plasmodium falciparum* in red blood cells selectively takes up serum proteins that affect host pathogenicity. Malar J. https://doi.org/10.1186/s12936-020-03229-1

6. Kapishnikov S, Grolimund D, Schneider G, Pereiro E, McNally JG, Als-Nielsen J, Leiserowitz L (2017) Unraveling heme detoxification in the malaria parasite by in situ correlative X-ray fluorescence microscopy and soft X-ray tomography. Sci Rep 7(1):7610. https://doi.org/10.1038/s41598-017-06650-w

7. Kapishnikov S, Staalsø T, Yang Y, Lee J, Pérez-Berná AJ, Pereiro E, Als-Nielsen J et al (2019) Mode of action of quinoline antimalarial drugs in red blood cells infected by *Plasmodium falciparum* revealed in vivo. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.1910123116

8. Uzor PF (2020) Alkaloids from plants with antimalarial activity: a review of recent studies. Evid Based Complement Altern Med 2020:1–17. https://doi.org/10.1155/2020/8749083

9. Yousefinejad S, Mahboubifar M, Eskandari R (2019) Quantitative structure–activity relationship to predict the anti-malarial activity in a set of new imidazolopiperazines based on artificial neural networks. Malar J. https://doi.org/10.1186/s12936-019-2941-5

10. Pinheiro LCS, Feitosa LM, Gandi MO, Silveira FF, Boechat N (2019) The development of novel compounds against malaria: quinolines, triazolpyridines, pyrazolopyridines and pyrazolopyrimidines. Molecules 24(22):4095. https://doi.org/10.3390/molecules24224095

11. Lin X, Li X, Lin X (2020) A review on applications of computational methods in drug screening and design. Molecules 25(6):1375. https://doi.org/10.3390/molecules25061375

12. Adeniji SE, Uba S, Uzairu A, Arthur DE (2019) A derived QSAR model for predicting some compounds as potent antagonist against *Mycobacterium tuberculosis*: a theoretical approach. Adv Prev Med 2019:1–18. https://doi.org/10.1155/2019/5173786

13. Hadanu R, Adelin L, Sutapa IW (2018) QSAR studies of nitrobenzothiazole derivatives as antimalarial agents. Makara J Sci. https://doi.org/10.7454/mss.v22i1.7620

14. Beheshti A, Pourbasheer E, Nekoei M, Vahdani S (2016) QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions. J Saudi Chem Soc 20(3):282–290. https://doi.org/10.1016/j.jscs.2012.07.019

15. Gujjar R, El Mazouni F, White KL, White J, Creason S, Shackleford DM, Rathod PK et al (2011) Lead optimization of aryl and aralkyl amine-based triazolopyrimidine inhibitors of *Plasmodium* falciparum dihydroorotate dehydrogenase with antimalarial activity in mice. J Med Chem 54(11):3935–3949. https://doi.org/10.1021/jm200265b

16. Ismail SY, Uzairu A (2019) In silico QSAR and molecular docking studies of sulfur containing shikonin oxime derivatives as anti-cancer agent for colon cancer. Radiol Infect Dis. https://doi.org/10.1016/j.jrid.2019.10.00

17. Ibrahim ZY, Uzairu A, Shallangwa G, Abechi S (2020) Theoretical design of novel antimalarial agents against *P. falciparum* strain, Dd2 through the QSAR modeling of synthesized 2′-substituted triclosan derivatives. Heliyon 6(9):e05032. https://doi.org/10.1016/j.heliyon.2020.e05032

18. Adedirin O, Uzairu A, Shallangwa GA, Abechi SE (2018) Optimization of the anticonvulsant activity of 2-acetamido- N -benzyl-2-(5-methylfuran-2-yl) acetamide using QSAR modeling and molecular docking techniques. Beni-Suef Univ J Basic Appl Sci. https://doi.org/10.1016/j.bjbas.2018.03.010

19. Haq E, Ahmad I, Hussain A, Almanjahie IM (2019) A novel selection approach for genetic algorithms for global optimization of multimodal continuous functions. Comput Intell Neurosci 2019:1–14. https://doi.org/10.1155/2019/8640218

20. Canali C, Lancellotti R (2019) GASP: genetic algorithms for service placement in fog computing systems. Algorithms 12(10):201. https://doi.org/10.3390/a12100201

21. Khezri V, Yasari E, Panahi M, Khosravi A (2020) A hybrid ANN−GA-based technique to optimize a steady-state GTL plant. Ind Eng Chem Res. https://doi.org/10.1021/acs.iecr.9b06477

22. Olasupo SB, Uzairu A, Shallangwa G, Uba S (2020) QSAR modeling, molecular docking and ADMET/pharmacokinetic studies: a chemometrics approach to search for novel inhibitors of norepinephrine transporter as potent antipsychotic drugs. J Iran Chem Soc. https://doi.org/10.1007/s13738-020-01902-5

23. Roy K, Ambure P, Kar S (2018) Prediction reliability indicator: a new tool to judge the quality of predictions from QSAR models for new query compounds. In: 24 May 2018 in MOL2NET 2018, international conference on multidisciplinary sciences

24. Ancuceanu R, Tamba B, Stoicescu CS, Dinu M (2019) Use of QSAR global models and molecular docking for developing new inhibitors of c-src tyrosine kinase. Int J Mol Sci 21(1):19. https://doi.org/10.3390/ijms21010019

25. Abdullahi M, Adeniji SE, Arthur DE et al (2020) Quantitative structure-activity relationship (QSAR) modelling study of some novel carboxamide series as new anti-tubercular agents. Bull Natl Res Cent 44:136. https://doi.org/10.1186/s42269-020-00389-7

26. Arthur DE, Uzairu A, Mamza P, Eyije Abechi S, Shallangwa G (2018) Insilico modelling of quantitative structure-activity

relationship of Pgi50 anticancer compounds on k-562 cell line. Cogent Chem. https://doi.org/10.1080/23312009.2018.1432520

27. 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

28. Kim JH (2019) Multicollinearity and misleading statistical results. Korean J Anesthesiol 72(6):558–569. https://doi.org/10.4097/kja.19087

29. Senaviratna NAMR, Cooray TMJA (2019) Diagnosing multicollinearity of logistic regression model. Asian J Probab Stat 5(2):1–9

30. Sharma A, Teotia D, Nandi S (2018) QSAR and structure based modeling of marine derived anticancer hymenialdisine compounds. J Dev Drugs. https://doi.org/10.4172/2329-6631.1000185

31. Adhikari N, Jana D, Halder KA, Mondal C, Maiti KM, Jha T (2012) Chemometric modeling of 5-phenylthiophenecarboxylic acid derivatives as antirheumatic agents. Curr Comput Aided Drug Des 8:182–195

32. Alisi IO, Uzairu A, Abechi SE, Idris SO (2018) Quantitative structure-activity relationship analysis of coumarins as free radical scavengers by genetic function algorithm. Phys Chem Res 6:208–222