VERSITA

Chemical Papers 67 (11) 1462–1473 (2013)
DOI: 10.2478/s11696-013-0398-5

ORIGINAL PAPER

QSAR study of amidino bis-benzimidazole derivatives as potent anti-malarial agents against *Plasmodium falciparum*

[a,b]Apilak Worachartcheewan, [a,b]Chanin Nantasenamat*, [b]Chartchalerm Isarankura-Na-Ayudhya, [b]Virapong Prachayasittikul*

[a]*Center of Data Mining and Biomedical Informatics,* [b]*Department of Clinical Microbiology and Applied Technology, Faculty of Medical Technology, Mahidol University, 10700 Bangkok, Thailand*

Received 8 January 2013; Revised 17 February 2013; Accepted 21 February 2013

A data set of amidino bis-benzimidazoles, in particular 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidine derivatives with anti-malarial activity against *Plasmodium falciparum* was employed in investigating the quantitative structure-activity relationship (QSAR). Quantum chemical and molecular descriptors were obtained from B3LYP/6-31g(d) calculations and Dragon software, respectively. Significant variables, which included total energy (E_T), highest occupied molecular orbital (HOMO), Moran autocorrelation-lag3/weighted by atomic masses (MATS3m), Geary autocorrelation-lag8/weighted by atomic masses (GATS8m), and 3D-MoRSE-signal 11/weighted by atomic Sanderson electronegativities (Mor11e), were used in the construction of QSAR models using multiple linear regression (MLR) and artificial neural network (ANN). The results indicated that the predictive models for both the MLR and ANN approaches using leave-one-out cross-validation afforded a good performance in modelling the anti-malarial activity against *P. falciparum* as observed by correlation coefficients of leave-one-out cross-validation ($R_{\text{LOO-CV}}$) of 0.9760 and 0.9821, respectively, root mean squared error of leave-one-out cross-validation ($\text{RMSE}_{\text{LOO-CV}}$) of 0.1301 and 0.1102, respectively, and predictivity of leave-one-out cross-validation ($Q^2_{\text{LOO-CV}}$) of 0.9526 and 0.9645, respectively. Model validation was performed using an external testing set and the results suggested that the model provided good predictivity for both MLR and ANN models with correlation coefficient of the external set (R_{Ext}) values of 0.9978 and 0.9844, respectively, root mean squared error of the external set (RMSE_{Ext}) of 0.0764 and 0.1302 respectively, and predictivity of the external set (Q^2_{Ext}) of 0.9956 and 0.9690, respectively. Furthermore, the robustness of the QSAR models is corroborated by a number of statistical parameters, comprising adjusted correlation coefficient (R^2_{Adj}), standard deviation (s), predicted residual sum of squares (PRESS), standard error of prediction (SDEP), total sum of squares deviation (SSY), and quality factor (Q). The QSAR models so constructed provide pertinent insights for the future design of anti-malarial agents.

Keywords: quantitative structure-activity relationship, malaria, *Plasmodium falciparum*, multiple linear regression, artificial neural network, anti-malarial agent, data mining

Introduction

Malarial infection caused by the *Plasmodium* genus is a public health concern as it gives rise to morbidity and mortality in tropical and subtropical areas. The disease is transmitted by the parasitic female anopheline mosquito (Farooq & Mahajan, 2004; Greenwood et al., 2005; White, 2004). Two of the five species commonly found in malarial infection are *Plasmodium falciparum* and *Plasmodium vivax* (Greenwood et al., 2005; White, 2004). In particular, *P. falciparum* has been documented as causing severe cerebral malaria

*Corresponding author, e-mail: chanin.nan@mahidol.ac.th, virapong.pra@mahidol.ac.th

Table 1. Molecular descriptors and anti-malarial activity of the amidino bis-benzimidazoles dataset

Compound	Descriptors					$IC_{50}/(\mu g\ mL^{-1})$	pIC_{50}
	E_T	HOMO	MATS3m	GATS8m	Mor11e		
I	–1237.65	–0.2003	–0.205	0.872	0.975	0.007	2.155
II	–1237.66	–0.2003	–0.165	1.261	0.982	0.010	2.000
III	–1336.88	–0.2030	–0.086	1.236	1.060	0.015	1.824
IV	–1295.68	–0.1971	–0.246	1.163	0.236	0.031	1.509
V	–1370.90	–0.2013	–0.259	0.940	0.204	0.211	0.676
VI	–1935.05	–0.2057	–0.096	0.674	1.646	0.064	1.194
VII	–1812.50	–0.2079	–0.129	1.152	1.295	0.630	0.201
VIII	–1367.47	–0.1973	–0.211	1.218	0.327	0.017	1.770
IX	–1367.46	–0.1961	–0.051	1.218	0.870	0.008	2.097
X	–1484.00	–0.1941	–0.275	1.272	1.088	0.017	1.770
XI	–1598.52	–0.1971	–0.074	1.260	0.493	0.023	1.638
XII	–1829.57	–0.1970	–0.089	1.219	1.096	0.061	1.215
XIII	–1904.28	–0.2009	–0.054	1.309	1.072	0.193	0.714
XIV	–2003.51	–0.2036	–0.061	1.288	1.231	0.634	0.198
XV	–1904.28	-0.1985	-0.051	1.007	0.942	0.028	1.553
XVI	–1673.73	–0.1939	–0.100	1.117	1.039	0.015	1.824
XVII	–1292.06	–0.1983	–0.228	1.200	0.661	0.016	1.796
XVIII	–1292.07	–0.1970	–0.237	1.163	1.195	0.016	1.796
XIX	–1213.64	–0.1951	–0.237	1.346	0.934	0.013	1.886
XX	–1252.95	–0.1939	–0.241	1.283	1.072	0.015	1.824

culminating in the death of patients (Idro et al., 2005, 2010). Although the use of drugs has effectively resulted in successful treatment of *Plasmodium* infections, drug resistances in *Plasmodium* have been reported and the drugs are known to have failed against malarial infection (Bloland, 2001; Gogtay et al., 2006; Kshirsagar, 2006). Owing to growing resistance, the development of novel compounds is recommended in order to address this challenge. In the literature, aromatic diamidines have been reported as anti-parasitic agents (Nwaka & Ridley, 2003; Nwaka & Hudson, 2006) while bisbenzimidazole amidine has been used as antifungal as well as DNA binding recognition (Del Poeta et al., 1998; Tanious et al., 2004). Accordingly, twenty analogues of a series of 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidines were synthesised and evaluated for their anti-fungal and anti-parasitic activities by Alp et al. 2009. These compounds exhibited good activity against *Trypanozoma brucei rhodesiense*, *P. falciparum*, and *Candida albicans*. In particular, the compounds displayed superior activity against *P. falciparum*. In the present study, the quantitative structure-activity relationship of the physiochemical properties of these compounds was correlated with their biological activity against *P. falciparum*. Important descriptors were used in constructing quantitative structure-activity relationship (QSAR) models for predicting the biological activity, using multiple linear regression and artificial neural network as regression methods. As the QSAR investigation performed here employs a small dataset (N = 20), caution is enjoined in analysing the data. Notably, several successful examples on the use of small data sets (i.e. N = 12 (Podunavac-Kuzmanović & Cvetković, 2011; Verma, 2006), N = 13 (Vahdani & Bayat, 2011), N = 19 (Saghaie et al., 2013), N = 20 (Sawant et al., 2013), N = 21 (Rastija & Medić-Šarić, 2009), and N = 23 (Jain et al., 2012)) have previously been demonstrated. In this respect, sampling of the data set was performed by means of leave-one-out cross-validation (LOO-CV) in order to make the most economical use of the available data. LOO-CV was demonstrated as useful in previous QSAR studies (Jalali-Heravi & Parastar, 2000; Nantasenamat et al., 2005, 2007; Suvannang et al., 2011; Thippakorn et al., 2009; Worachartcheewan et al., 2009, 2011, 2012; Zou & Zhou, 2007).

Theoretical

Data set

The IC_{50} values of the anti-malarial activity against *P. falciparum* were obtained from the study by Alp et al. (2009) and converted to pIC_{50} by taking the negative logarithm to base 10 of the IC_{50} values ($-\log IC_{50}$), as shown in Table 1. The twenty molecular structures of 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidines derivatives are presented in Fig. 1 while their activities are shown in Table 1.

Calculation of descriptors

The molecular structures of the compounds were drawn using the GaussView software package (Dennington et al., 2003). The optimal geometrical structures and their quantum chemical descriptors were

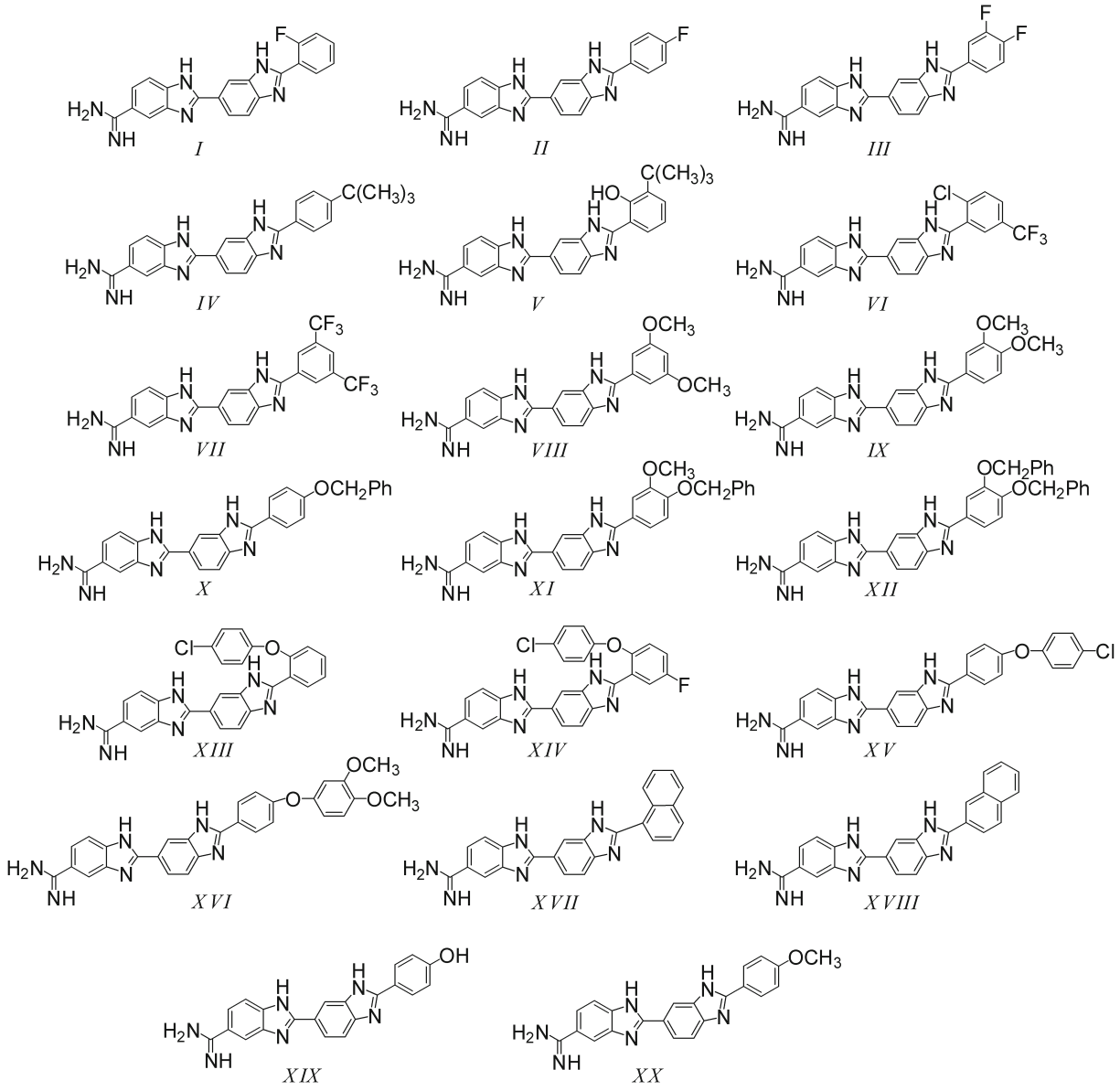

Fig. 1. Molecular structures of 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidines.

derived from Gaussian 03W software (Frisch et al., 2004).

Quantum chemical descriptors

The initial geometry of the molecular structures was optimised using the semi-empirical Austin Model 1 (AM1) method. The structures were then subjected to further geometrical optimisation at the density functional theory (DFT) using Becke's three parameter hybrid Lee-Yang-Parr (B3LYP) functional together with the 6-31g(d) basis set. All quantum chemical descriptors were calculated by Gaussian 03W. The descriptors thus generated include total energy (E_T), energy of the highest occupied molecular orbital (HOMO), energy of the lowest unoccupied molecular orbital (LUMO), dipole moment (μ), difference between the energy values of HOMO and LUMO ($GAP_{HOMO-LUMO}$), hardness (η), electrophilicity (ω), softness (S), Mulliken electronegativity (χ), electrophilic index (ω_i), and the mean absolute atomic charge (Q_m) (Karelson et al., 1996; Parr et al., 1978, 1999; Parr & Pearson, 1983; Thanikaivelan et al., 2000).

Molecular descriptors

The optimised structure of compounds calculated by Gaussian 03W was saved in MOL2 format and further used as input data for calculating 3224 molecular descriptors in Dragon software, version 5.5 (Talete, 2007). The molecular descriptors from Dragon are classified into the following 22 categories: 48 constitutional descriptors, 119 topological descriptors, 47

walk and path counts, 33 connectivity indices, 47 information indices, 96 2D autocorrelation, 107 edge adjacency indices, 64 burden eigenvalues, 21 topological charge indices, 44 eigenvalue-based indices, 41 randic molecular profiles, 74 geometrical descriptors, 150 RDF descriptors, 160 3D-MoRSE descriptors, 99 WHIM descriptors, 197 GETAWAY descriptors, 154 functional group counts, 120 atom-centred fragments, 14 charge descriptors, 29 molecular properties, 780 2D binary fingerprints and 780 2D frequency fingerprints.

Data pre-processing

The quantum chemical and molecular descriptors used as independent variables were obtained from Gaussian 03W and Dragon, respectively. These descriptors were standardised to comparable scale with a mean of zero and standard derivation of one using Waikato Environment for Knowledge Analysis (Weka) (Witten et al., 2011), version 3.4.5, according to the following equation:

$$x_{ij}^{\mathrm{stn}} = \frac{x_{ij} - \bar{x}_j}{\sum_{i=1}^{N} (x_{ij} - \bar{x}_j)^2 / N} \qquad (1)$$

where x_{ij}^{stn} is the standardised value, x_{ij} is the value of each sample, $\bar{x}_j$ is the mean of each descriptor and N is the sample size of the data set.

Once the full set of 3224 molecular descriptors from Dragon was standardised, redundant and constant descriptors were removed using the unsupervised forward selection (UFS) algorithm of the UFS software package, version 1.8. The UFS algorithm was described by Whitley et al. (2000) and summarised in one of our previous works (Nantasenamat et al., 2005). In addition, significant descriptors were selected by stepwise regression using the statistical data analysis software, SPSS Statistics 18.0 (IBM, USA).

In order to detect outliers in the data set, the cut-off value was set to the absolute standardised residual of 2. This method has been widely used for the identification of outliers in QSAR investigations (Afantitis et al., 2006; García et al., 2011; Verma & Hansch, 2010; Vlaia et al., 2009). Several other techniques are also available for outlier detection, such as Williams plot (Gramatica, 2007; Khosrokhavar et al., 2010; Papa et al., 2005), Monte Carlo cross-validation (Cao et al., 2010), principal component analysis (Shahlaei et al., 2009; Sun et al., 2009), self-organising map and adaptive non-linear map (Yan, 2011). Collinearity of descriptors was identified by constructing an inter-correlation matrix that can be used to reveal non-dependence of descriptors with each other (Gupta et al., 2010; Rastija & Medić-Šarić, 2009; Veerasamy et al., 2009). This inter-correlation matrix was created by calculating the Pearson's correlation coefficient in a pairwise manner.

Data preparation and feature selection

A data set of twenty analogues of 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidines possessing anti-malarial activity against *P. falciparum* (Fig. 1, Table 1) was employed in the QSAR investigation presented herein. Quantum chemical and molecular descriptors were calculated using Gaussian 03W and Dragon software, respectively. From the masses of calculated descriptors as obtained from Dragon, constant or near-constant variables were excluded. This led to the elimination of 1834 descriptors from an initial set of 3224 descriptors, yielding a reduced set of 1390 descriptors. The descriptors were then standardised according to Eq. (1). This set of molecular descriptors was subjected to further reduction of redundant descriptors using the UFS algorithm. Descriptors where the pairs of variables displayed a squared correlation coefficient (R^2) greater than 0.99 were excluded. This yielded a final set of 19 molecular descriptors consisting of: (i) constitutional descriptors, nCIC; (ii) 2D binary fingerprints, B05[N-F], B01[C-O], and B04[O-O]; (iii) WHIM descriptors, G2u, P2u, G3e, and G2m; (iv) GETAWAY descriptors, R6e, R7u+, R2p+, and HATS5p; (v) 2D autocorrelations, GATS8v, MATS7e, Moran autocorrelation-lag3/weighted by atomic masses (MATS3m), and Geary autocorrelation-lag 8/weighted by atomic masses (GATS8m); (vi) information indices: BIC3; (vii) molecular properties: Depressant-80 and; (viii) 3D-MoRSE descriptors, 3D-MoRSE-signal 11/weighted by atomic Sanderson electronegativities (Mor11e). DRAGON descriptors have previously been shown to be useful for QSAR studies (Castillo-Garit et al., 2008; Durand et al., 2007; Fernández et al., 2005; Zhang et al., 2008). Thus, 19 DRAGON descriptors along with 11 quantum chemical descriptors were subjected to further descriptor reduction by means of stepwise multiple linear regression. This process led to the selection of five significant descriptors from a set of thirty descriptors that was then used in QSAR study as shown in Table 1. The final set of descriptors consists of two quantum chemical descriptors: E_{T} and HOMO; and three molecular descriptors: MATS3m, GATS8m, and Mor11e. The inter-correlation matrix of the final set of descriptors as presented in Table 2 revealed that the descriptors were independent from one another.

Multivariate regression analysis

QSAR models of 2′-arylsubstituted-1*H*,1′*H*-[2,5′] bisbenzimidazolyl-5-carboximidines as anti-malarial agents against *P. falciparum* were constructed using multiple linear regression (MLR) and artificial neural network (ANN).

MLR was used for the construction of QSAR mod-

Table 2. Inter-correlation matrix of five significant descriptors

Descriptor	Parameters				
	E_T	HOMO	MATS3m	GATS8m	Mor11e
E_T	1.0000	–	–	–	–
HOMO	0.4448	1.0000	–	–	–
MATS3m	–0.5457	–0.3964	1.0000	–	–
GATS8m	0.1958	0.4104	0.0092	1.0000	–
Mor11e	–0.4374	–0.2664	0.3647	–0.0771	1.0000

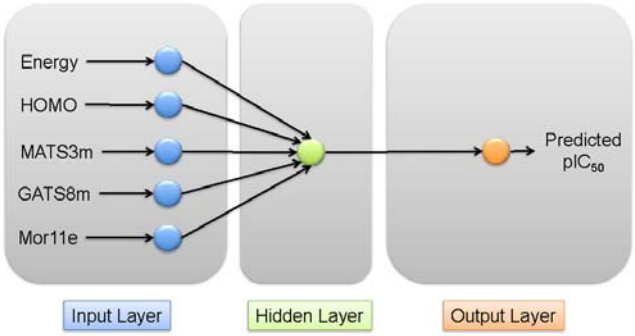

Fig. 2. Schematic representation of QSAR model using ANN.

els using the following equation:

$$Y = B_0 + \sum B_n X_n \quad (2)$$

where Y is the pIC_{50} of the compound, B_0 is the intercept, and B_n are the regression coefficients of descriptors X_n. MLR was employed for generating linear models that correlate the anti-malarial activity with their respective molecular descriptors. Stepwise MLR employing both forward and backward algorithms were used to select important descriptors from the initial set consisting of a large number of variables. Several efforts that successfully used MLR to construct QSAR models have been documented (Alves et al., 2001; Caballero & Fernández, 2006; Nantasenamat et al., 2007; Prasad et al., 2008; Zahouily et al., 2006, Worachartcheewan et al., 2011, 2012; Saghaie et al., 2013).

ANN is a data mining technique whose architectural design is similar to the learning process of neurons in the human brain. ANN implementing the back-propagation algorithm is a popular approach for constructing non-linear models that correlate physicochemical properties with their observed biological activity (Jalali-Heravi & Parastar, 2000; Nantasenamat et al., 2005; Thippakorn et al., 2009; Worachartcheewan et al., 2009; Zahouily et al., 2006; Zou & Zhou, 2007). The layout of ANN used in this study is composed of 3 layers: input layer, hidden layer, and output layer (Fig. 2). The multivariate analysis was performed using WEKA, version 3.4.5. (Witten et al., 2011).

Model validation procedure

The data set of compounds exhibiting anti-malarial activity against *P. falciparum* was divided into two sets as training and testing sets, using LOO-CV. This procedure was performed by leaving one sample of the data set out as the testing set while the remaining data samples were used as the training set. The process was repeated until all data samples were used as the testing set. Additional data sampling approaches were performed to evaluate the predictive performance of the QSAR model by using an external test set to serve as the unknown data.

Statistical assessment of QSAR models

A range of statistical parameters was calculated to assess the predictive ability of the QSAR models via cross-validation. These parameters included R, R^2, root mean squared error (RMSE), predictivity (Q^2), variance ratio (F ratio), critical F value, adjusted correlation coefficient (R^2_{Adj}), standard deviation (s), predicted residual sum of squares (PRESS), standard error of prediction (SDEP), total sum of squares deviation (SSY), and quality factor (Q), as successfully deployed in previous QSAR studies (Alves et al., 2001; Jain et al., 2012; Nantasenamat et al., 2008; Podunavac-Kuzmanović & Cvetković, 2011; Rastija & Medić-Šarić 2009).

Results and discussion

QSAR modelling of anti-malarial activities

The set of five descriptors was subsequently used in construction of the QSAR model in order to discern the relationship between its physicochemical properties and biological activity. QSAR has previously been demonstrated as successful in studying the structure-activity relationship of biological and chemical systems (Nantasenamat et al., 2005, 2009; Prachayasittikul et al., 2010; Suksrichavalit et al., 2009; Suvannung et al., 2011; Worachartcheewan et al., 2009, 2011, 2012).

The following relevant studies employing computational modelling of anti-malarial activity have previously been documented. QSAR studies of anti-

Table 3. Generated equations from MLR model

Model	N	Equation
1	20	$pIC_{50} = 0.0017E_T + 104.6713HOMO + 2.6032MATS3m - 1.2850GATS8m + 0.3978Mor11e + 26.4391$
2	19	$pIC_{50} = 0.0016E_T + 91.8124HOMO + 2.2474MATS3m - 1.5615GATS8m + 0.0659Mor11e + 24.3676$
3	18	$pIC_{50} = 0.0015E_T + 86.4931HOMO + 1.9721MATS3m - 1.6700GATS8m - 0.1463Mor11e + 23.5660$
4	17	$pIC_{50} = 0.0015E_T + 93.2926HOMO + 1.9811MATS3m - 1.7734GATS8m - 0.1746Mor11e + 24.8559$
5	16	$pIC_{50} = 0.0012E_T + 98.3822HOMO + 1.2197MATS3m - 1.8671GATS8m - 0.2462Mor11e + 25.5277$

Table 4. Summary of predictive performance of MLR model

Model	N	R_{Tr}	$RMSE_{Tr}$	$R_{LOO\text{-}CV}$	$RMSE_{LOO\text{-}CV}$	R^2_{Tr}	$Q^2_{LOO\text{-}CV}$	F ratio	Critical F value
1	20	0.9457	0.1879	0.8557	0.3065	0.8943	0.7322	7.6556	2.9582[a]
2	19	0.9678	0.1415	0.9323	0.2050	0.9366	0.8692	17.2777	3.0254[b]
3	18	0.9795	0.1163	0.9581	0.1660	0.9594	0.9180	26.8683	3.1059[c]
4	17	0.9895	0.0840	0.9666	0.1505	0.9791	0.9343	31.2855	3.2039[d]
5	16	0.9922	0.0741	0.9760	0.1301	0.9845	0.9526	40.1941	3.3258[e]

Critical F values at 95 % confidence level with m and $n - m - 1$ degrees of freedom ($F_{(m,n-m-1)}$) as follows: *a*) $F_{(5,14)}$, *b*) $F_{(5,13)}$, *c*) $F_{(5,12)}$, *d*) $F_{(5,11)}$, *e*) $F_{(5,10)}$; F ratio – calculated F ratio of leave-one-out cross-validation testing set.

malarial agents based on 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulphonate and acylate derivatives ($N = 63$) were reported by Zahouily et al. (2006). The physicochemical and geometrical properties of the molecules were calculated to obtain the following descriptors: molecular weight, hydrogen-bond acceptors and log$P$. QSAR models were subsequently derived using the MLR and ANN methods, which displayed good statistical performance as deduced from Q^2, Q, PRESS, and SSY. Another example is the QSAR study of anti-malarial 3-hydroxypyridinone agents ($N = 19$) by Saghaie et al. (2013). Molecular and quantum chemical descriptors consisting of molecular volume, HOMO, number of ring secondary carbon, C (sp^3), atoms and number of hydrogen bond donors were used in the construction of predictive models using MLR and principle component regression (PCR). The statistical parameters exhibited robust performance as observed by the following statistical parameters: R^2, RMSE, PRESS, correlation coefficients of leave-one-out cross-validation ($R^2_{LOO\text{-}CV}$), and root mean squared error of leave-one-out cross-validation ($RMSE_{LOO\text{-}CV}$).

In this study, QSAR models were constructed using the MLR and ANN approaches which were calculated using the Weka, version 3.4.5. To obtain the best model, correlation coefficient (R) was used as a relative measure of predictive performance, which corresponded to the degree of correlation existing between the predicted and experimental values. RMSE was used to measure the predictive error of the model.

QSAR model developed by MLR

The QSAR model of 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidines was used with the MLR approach according to Eq. (2) and led to the formulation of several QSAR equations (Table 3) along with the corresponding statistical analysis (Table 4). Compounds in the data set were identified as outliers if the absolute value of their standardised residual was greater than 2. This process was performed until the model had no outliers. The results indicated that compounds *II*–*V* were outliers in the MLR model and they were excluded from the data set (Tables 3 and 4). This led to a performance improvement as indicated by the statistical parameters presented in Table 4. The best model for the MLR method was model 5, possessing correlation coefficients $R_{LOO\text{-}CV}$, $RMSE_{LOO\text{-}CV}$, and $Q^2_{LOO\text{-}CV}$ of 0.9760, 0.1301, and 0.9526, respectively, as obtained from leave-one-out cross-validation. The experimental and predicted activities along with their respective residuals are presented in Table 5.

The best MLR equation (Table 4) for modelling the pIC_{50} of amidino bis-benzimidazole derivatives was obtained from Model 5 as a function of the set of five important descriptors (i.e. E_T, HOMO, MATS3m, GATS8m and Mor11e):

$$pIC_{50} = 0.0012E_T + 98.3822HOMO + 1.2197MATS3m - 1.8671GATS8m - 0.2462Mor11e + 25.5277 \quad (3)$$

$N = 16$, $R_{LOO\text{-}CV} = 0.9760$, $RMSE_{LOO\text{-}CV} = 0.1301$, correlation coefficient of training set (R^2_{Tr}) = 0.9845, $Q^2_{LOO\text{-}CV} = 0.9526$, F ratio = 40.1941, critical F value = 3.3258. Furthermore, additional statistical parameters corroborated the robustness of the MLR model: $R^2_{Adj} = 0.9354$, $s = 0.1339$, PRESS = 0.2706, SDEP = 0.1300, SSY = 5.4890, and $Q = 7.2890$.

QSAR model developed by ANN

In addition, a QSAR study of 2′-arylsubstituted-

Table 5. Experimental and predicted pIC_{50} values as obtained using MLR and ANN methods

Compound	Experimental pIC_{50}	MLR		ANN	
		Predicted pIC_{50}	Residual	Predicted pIC_{50}	Residual
I	2.155	2.218	–0.063	2.030	0.125
II	2.000	–	–	–	–
III	1.824	–	–	1.967	–0.143
IV	1.509	–	–	–	–
V	0.676	–	–	–	–
VI	1.194	1.061	0.133	1.206	–0.012
VII	0.201	0.270	–0.069	0.277	–0.076
VIII	1.770	1.530	0.240	1.575	0.195
IX	2.097	1.850	0.247	1.984	0.113
X	1.770	1.902	–0.132	1.810	–0.040
XI	1.638	1.613	0.025	1.700	–0.062
XII	1.215	1.263	–0.048	1.317	–0.102
XIII	0.714	0.639	0.075	0.572	0.142
XIV	0.198	0.299	–0.101	0.249	–0.051
XV	1.553	1.476	0.077	1.486	0.067
XVI[a]	1.824	1.990	–0.166	1.967	–0.143
XVII[a]	1.796	1.745	0.051	1.792	0.004
XVIII	1.796	1.820	–0.024	1.870	–0.074
XIX	1.886	1.768	0.118	1.855	0.031
XX	1.824	2.021	–0.197	1.967	–0.143

a) Present in data sets of MLR and ANN models, respectively.

Table 6. Summary of predictive performance of ANN model

Model	N	P_{ANN}	R_{Tr}	$RMSE_{Tr}$	$R_{LOO\text{-}CV}$	$RMSE_{LOO\text{-}CV}$	R^2_{Tr}	$Q^2_{LOO\text{-}CV}$	F ratio	Critical F value
1	20	1 100 0.3 0.1	0.9449	0.1639	0.9064	0.2525	0.8928	0.8216	12.8951	2.9582[a]
2	19	1 70 0.1 0.6	0.9779	0.1314	0.9436	0.1877	0.9563	0.8906	21.1660	3.0254[b]
3	18	1 60 0.2 0.2	0.9869	0.1076	0.9639	0.1549	0.9740	0.9291	31.4505	3.1059[c]
4	17	1 300 0.1 0.2	0.9929	0.0708	0.9821	0.1102	0.9859	0.9645	59.7718	3.2039[d]

Critical F values at 95 % confidence level with m and $n - m - 1$ degrees of freedom ($F_{(m,n-m-1)}$) for the various ANN models as follows: *a*) $F_{(5,14)}$, *b*) $F_{(5,13)}$, *c*) $F_{(5,12)}$, and *d*) $F_{(5,11)}$; P_{ANN} – optimal ANN parameters consisting of hidden node, learning epoch, learning rate and momentum, respectively.

1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidines against *P. falciparum* was performed using ANN as the non-linear method. To obtain the best model of ANN, a systematic parameter search was performed to obtain an optimal set of parameters, which included the number of hidden layer, learning epochs, learning rate, and momentum. ANN implementing the back-propagation algorithm (Worachartcheewan et al., 2009) of Weka was employed in this study. The RMS was used as a measure of predictive error in order to obtain the most optimal ANN model. As each ANN run randomly initialises the interconnection weights of the neurons, ten runs of ANN were performed for each parameter and their average values were used. The outliers inherent in the data set were identified as compounds *II*, *IV*, and *V*. This process of outlier iden-

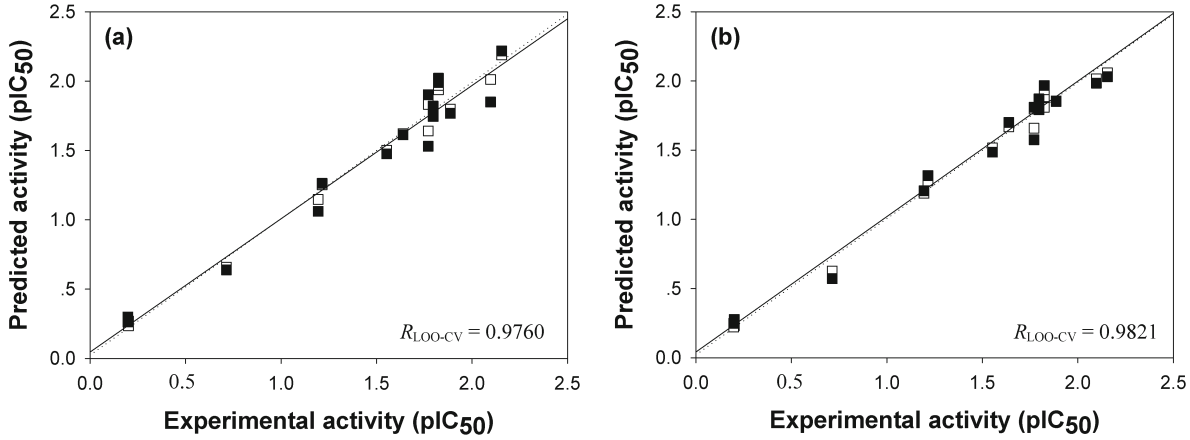

Fig. 3. Plot of experimental and predicted activity (pIC_{50}) for training set (□) (regression line is represented by dotted line) and leave-one-out cross-validated testing set (■) (regression line is represented by solid line) using MLR (a) and ANN (b) methods for 16 and 17 compounds, respectively.

tification was iteratively performed until no outliers could be detected in the data set, shown as Models 1 through 4 (Table 6).

The statistical results from the QSAR study of 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidines employing ANN are presented in Table 5. The ANN model using optimal parameters of ANN models was subjected to a series of iterative outlier exclusion as presented in Table 5. Model 4 was shown to be the best ANN model exhibiting an optimal set of parameters consisting of hidden nodes, learning epochs, learning rate, and momentum of 1, 300, 0.1, and 0.2, respectively. The R_{CV}, RMS_{CV}, and calculated using leave-one-out cross-validation were 0.9821, 0.1102, and 0.9645, respectively. The experimental and predicted activities along with their respective residuals are presented in Table 5. Moreover, the best ANN model exhibited the following statistical parameters: $N = 17$, $R_{LOO\text{-}CV} = 0.9821$, $RMSE_{LOO\text{-}CV} = 0.1102$, $R^2_{Tr} = 0.9859$, $Q^2_{LOO\text{-}CV} = 0.9645$, F ratio = 59.7718, critical F value = 3.2069, $R^2_{Adj} = 0.9527$, $s = 0.1065$, PRESS = 0.1832, SDEP = 0.1038, SSY = 7.9442, and $Q = 9.2194$.

Interpreting QSAR models

The five descriptors obtained through feature selection were shown to be useful in the construction of robust QSAR models. Compounds *I*, *II*, and *IX* possessed potent inhibitory activity against *P. falciparum* with IC_{50} (pIC_{50}) values of 0.007 (2.155) μg mL^{-1}, 0.010 (2.000) μg mL^{-1}, and 0.008 (2.097) μg mL^{-1}, respectively, as compared with the reference chloroquine (Alp et al., 2009). We observed that the total energy of compounds *I* and *II* of –1237.65 and –1237.66 , respectively, was higher than that of compound *IX*. The structures of compounds *I* and *II*, possessed a fluoride atom in the structure whereas compound *IX* had methoxy groups in the structure. Moreover, we found that HOMO, MATS3m, and More11e displayed lower values in compounds *I* and *II* than in compound *IX*, which correspondingly referred to the energy of the highest occupied molecular orbital in the ground state, atomic masses and electronegativities.

Models 5 of MLR and 4 of ANN were used in predicting the anti-malarial activity (pIC_{50}) against *P. falciparum*. The experimental and predicted activities are presented in Table 5. Comparisons of the experimental values in relation to their predicted values are given for MLR in Table 4 and for ANN in Table 6 while MLR (Fig. 3a) and ANN (Fig. 3b) display the plots of experimental values against those of their predicted values. The plots indicated that both QSAR models possessed a high correlation coefficient, implying that the experimental and predicted values were well correlated. Assessment of the statistical parameters from QSAR models developed by the MLR (model 5) and ANN (model 4) approaches indicated a satisfactory performance. PRESS is an essential statistical parameter used in approximating the real predictive error of models and its value was shown to be less than the SSY value, thereby suggesting a statistical significance. Q is the ratio of R to the standard error of estimation (S_e) in which S_e ($Q = r/s$) describes the quality of the models. It was observed that the Q value of ANN was higher than that of the MLR model.

Model validation using external testing set

The external predictivity of the QSAR models was assessed using an external testing set constructed by random selection. The data sets from model 5 and model 4 of MLR and ANN, respectively, were used and divided into two sets of data: (i) 13 compounds and 14 compounds of the MLR and ANN models, respectively, were used as the training set and LOO-CV

Table 7. Results of MLR and ANN models for predicting anti-*Plasmodium falciparum* activity

Method	N	Training set		LOO-CV			External test set[c]	
		R_{Tr}	$\mathrm{RMSE_{Tr}}$	$R_{\mathrm{LOO\text{-}CV}}$	$\mathrm{RMSE_{LOO\text{-}CV}}$	F ratio	R_{Ext}	$\mathrm{RMSE_{Ext}}$
MLR[a]	16	0.9922	0.0741	0.9760	0.1301	40.1941	–	–
	13	0.9899	0.0757	0.9537	0.1909	14.0771	0.9978	0.0764
ANN[b]	17	0.9929	0.0708	0.9821	0.1102	59.7718	–	–
	14	0.9959	0.0564	0.9626	0.1422	20.1980	0.9844	0.1302

a) Critical F value at 95 % confidence level with 5 and 10 degrees of freedom is 3.3258 for data set with $N = 16$ and critical F value at 95 % confidence level with 5 and 7 degrees of freedom is 3.9715 for data set with $N = 13$; *b*) critical F value at 95 % confidence level with 5 and 11 degrees of freedom is 3.2039 for data set with $N = 17$ and critical F value at 95 % confidence level with 5 and 8 degrees of freedom is 3.6875 for data set with $N = 14$; *c*) compounds *VII*, *XII*, *XIX* were used as external validation for MLR ($N = 13$) and ANN ($N = 14$) model.

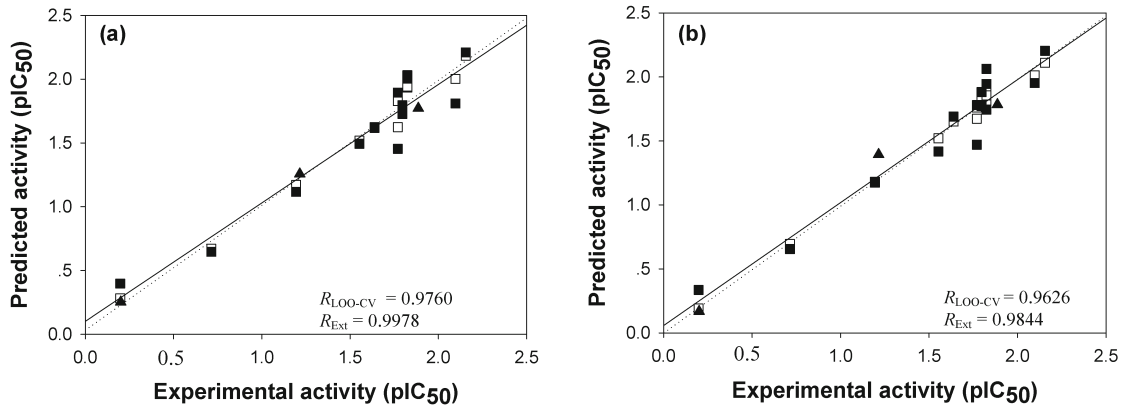

Fig. 4. Plot of predicted and experimental activity ($\mathrm{pIC_{50}}$) for training set (□) (regression line is represented by dotted line), leave-one-out cross-validated test set (■) (regression line is represented by solid line) and external testing set (▲) using MLR (a) and ANN (b) methods comprising of 13, 14, and 3 compounds, respectively.

for QSAR model development; and (ii) an external testing set (compounds *VII*, *XII*, and *XIX*) was used for external validation of the QSAR model.

The MLR equation for modelling the $\mathrm{pIC_{50}}$ of amidino bis-benzimidazole derivatives was obtained as shown in the following equation:

$$\mathrm{pIC_{50}} = 0.0012E_{\mathrm{T}} + 96.774\mathrm{HOMO} + 1.2385\mathrm{MATS3m} - 1.88857\mathrm{GATS8m} - 0.2594\mathrm{Mor11e} + 25.1863 \quad (4)$$

Number of compound in the training set (N_{Tr}) = 13, number of compound of the external set (N_{Ext}) = 3, $R_{\mathrm{LOO\text{-}CV}} = 0.9537$, $\mathrm{RMSE_{LOO\text{-}CV}} = 0.1909$, $R_{\mathrm{Ext}} = 0.9978$, $\mathrm{RMSE_{Ext}} = 0.0764$, $R^2_{\mathrm{Tr}} = 0.9799$, $Q^2_{\mathrm{LOO\text{-}CV}} = 0.9095$, $Q^2_{\mathrm{Ext}} = 0.9095$, standard deviation of the external set $s_{\mathrm{Ext}} = 0.1024$, F ratio = 14.0771, Critical F value = 3.9715, $R^2_{\mathrm{Adj}} = 0.8643$, $s = 0.1343$, PRESS = 0.2487, SDEP = 0.1383, SSY = 3.8221, and Q = 7.1004.

The statistical parameters used in assessing the predictive performance were computed using an external test set as presented in Table 7. It was found that the predicted anti-malarial activity against *P. falciparum* for the external test set demonstrated a high correlation coefficient value of 0.9978 and root mean square error of 0.0764 (Table 7). The plot of the experimental against predicted ($\mathrm{pIC_{50}}$) values for the 13 compounds data set of the training set and LOO-CV testing set, as well as the 3 compounds data set of external testing set, was shown in Fig. 4a. The experimental and predicted activities along with their respective residuals are presented in Table 8.

For the ANN model, optimisation of the network parameters was performed to obtain the optimal parameters as follows: 8 nodes in the hidden layer, 500 learning epochs, learning rate of 0.1, and momentum of 0.6, which were also used for the external test set. The results (Table 7) indicated that the QSAR model gave accurately predicted values of the anti-malarial activity against *P. falciparum* as observed from the correlation coefficient of 0.983 and root mean square error of 0.130 (Table 7). Furthermore, additional statistical parameters corroborated the robustness of the MLR model: number of compounds in the training set $N_{\mathrm{Tr}} = 14$, number of compounds in the external set (N_{Ext}) = 3, $R_{\mathrm{LOO\text{-}CV}} = 0.9626$, $\mathrm{RMSE_{LOO\text{-}CV}} = 0.1422$, $R_{\mathrm{Ext}} = 0.9844$, $RMSE_{\mathrm{Ext}} = 0.1302$, $R^2_{\mathrm{Tr}} = 0.9959$, $Q^2_{\mathrm{LOO\text{-}CV}} = 0.9266$, $Q^2_{\mathrm{Ext}} = 0.9690$, $s_{\mathrm{Ext}} = 0.1463$, F ratio = 20.1980, critical F value = 3.6875, $R^2_{\mathrm{Adj}} = 0.8940$, $s = 0.1550$, PRESS = 0.3058, SDEP

Table 8. Experimental and predicted pIC_{50} by MLR and ANN methods[a]

Compound	pIC_{50}	MLR		ANN	
		Predicted	Residual	Predicted	Residual
I	2.155	2.210	–0.055	2.204	–0.049
II	2.000	–	–	–	–
III	1.824	–	–	2.063	–0.239
IV	1.509	–	–	–	–
V	0.676	–	–	–	–
VI	1.194	1.117	0.077	1.180	0.014
VII[b]	0.201	0.254	–0.053	0.169	0.032
VIII	1.770	1.895	–0.125	1.470	0.300
IX	2.097	1.810	0.287	1.952	0.145
X	1.770	1.454	0.316	1.780	–0.010
XI	1.638	1.618	0.020	1.691	–0.052
XII[b]	1.215	1.258	–0.430	1.394	–0.179
XIII	0.714	0.647	0.067	0.655	0.059
XIV	0.198	0.398	–0.200	0.336	–0.138
XV	1.553	1.494	0.059	1.418	0.136
XVI	1.824	2.006	–0.182	1.945	–0.121
XVI	1.796	1.796	0.000	1.882	–0.086
XVIII	1.796	1.728	0.068	1.776	0.020
XIX[b]	1.886	1.773	0.113	1.784	0.102
XX	1.824	2.032	–0.208	1.744	0.080

a) 13 and 14 compounds were presented in data sets modeled by MLR and ANN methods, respectively; *b*) external test set.

= 0.1478, SSY = 3.8716 and $Q = 6.2088$.

Plots of the experimental versus the predicted (pIC_{50}) values of the data set (e.g. comprising 14 compounds) and external testing set (e.g. containing 3 compounds) are shown in Fig. 4b. The experimental and predicted activities along with their respective residuals are presented in Table 8. The results suggested that the QSAR models constructed using MLR and ANN afforded a good performance in predicting the anti-malarial activity against *P. falciparum* of 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidines.

Conclusions

The MLR and ANN methods were successful in predicting the activity of 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboximidines against *P. falciparum*. The quantum chemical and molecular descriptors consisting of E_T, HOMO, MATS3m, GATS8m, and Mor11e were significant variables used in constructing robust QSAR models. The descriptors selected could account for the important roles in the interaction of compounds with the parasites. The results indicated that the MLR and ANN methods exhibited good predictive performances for modelling the anti-malarial activity against *P. falciparum*. The computational methodology employed herein has a potential application as a tool to complement high-throughput screening in the design of novel and robust anti-malarial agents.

Acknowledgements. This research project is financially supported by the Office of the Higher Education Commission and Mahidol University under the National Research Universities Initiative. A.W. is supported by the Royal Golden Jubilee (PhD) scholarship of the Thailand Research Fund under the supervision of V.P.

References

Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J., & Igglessi-Markopoulou, O. (2006). A novel QSAR model for predicting induction of apoptosis by 4-aryl-4*H*-chromenes. *Bioorganic & Medicinal Chemistry*, *14*, 6686–6694. DOI: 10.1016/j.bmc.2006.05.061.

Alp, M., Göker, H., Brun, R., & Yıldız, S. (2009). Synthesis and antiparasitic and antifungal evaluation of 2′-arylsubstituted-1*H*,1′*H*-[2,5′]bisbenzimidazolyl-5-carboxamidines. *European Journal of Medicinal Chemistry*, *44*, 2002–2008. DOI: 10.1016/j.ejmech.2008.10.003.

Alves, C. N., Pinheiro, J. C., Camargo, A. J., Ferreira, M. M. C., Romero, R. A. F., & da Silva, A. B. F. (2001). A multiple linear regression and partial least squares study of flavonoid compounds with anti-HIV activity. *Journal of Molecular Structure: THEOCHEM*, *541*, 81–88. DOI: 10.1016/s0166-1280(00)00755-7.

Bloland, P. B. (2001). *Drug resistance in malaria*. Retrieved November 25, 2011, from http://www.who.int/csr/resources/publications/drugresist/malaria.pdf

Caballero, J., & Fernández, M. (2006). Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks. *Journal of Molecular Modeling*, *12*, 168–181. DOI: 10.1007/s00894-005-0014-x.

Cao, D. S., Liang, Y. Z., Xu, Q. S., Li, H. D., & Chen, X. (2010). A new strategy of outlier detection for QSAR/QSPR. *Journal of Computational Chemistry*, *31*, 592–602. DOI: 10.1002/jcc.21351.

Castillo-Garit, J. A., Marrero-Ponce, Y., Escobar, J., Torrens, F., & Rotondo, R. (2008). A novel approach to predict aquatic toxicity from molecular structure. *Chemosphere*, *73*, 415–427. DOI: 10.1016/j.chemosphere.2008.05.024.

Del Poeta, M., Schell, W. A., Dykstra, C. C., Jones, S. K., Tidwell, R. R., Kumar, A., Boykin, D. W., & Perfect, J. R. (1998). In vitro antifungal activities of a series of dication-substituted carbazoles, furans, and benzimidazoles. *Antimicrobial Agents and Chemotherapy*, *42*, 2503–2510.

Dennington, R., II, Keith, T., Millam, J., Eppinnett, K., Hovell, W. L., & Gilliland, R. (2003). GaussView, Version 3.09 [computer software]. Shawnee Mission, KS, USA: Semichem.

Durand, A. C., Farce, A., Carato, P., Dilly, S., Yous, S., Berthelot, P., & Chavatte, P. (2007). Quantitative structure-activity relationships studies of antioxidant hexahydropyridoindoles and flavonoid derivatives. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *22*, 556–562. DOI: 10.1080/14756360701425238.

Farooq, U., & Mahajan, R. C. (2004). Drug resistance in malaria. *Journal of Vector Borne Diseases*, *41*, 45–53.

Fernández, M., Caballero, J., Helguera, A. M., Castro, E. A., & González, M. P. (2005). Quantitative structure-activity relationship to predict differential inhibition of aldose reductase by flavonoid compounds. *Bioorganic & Medicinal Chemistry*, *13*, 3269–3277. DOI: 10.1016/j.bmc.2005.02.038.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Jr., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., & Pople, J. A. (2004). Gaussian 03, Revision C.02 [computer software]. Wallingford, CT, USA: Gaussian.

García, J., Duchowicz, P. R., Rozas, M. F., Caram, J. A., Mirífico, M. V., Fernández, F. M., & Castro, E. A. (2011). A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases. *Journal of Molecular Graphics and Modelling*, *31*, 10–19. DOI: 10.1016/j.jmgm.2011.07.007.

Gogtay, N. J., Kshirsagar, N. A., & Vaidya, A. B. (2006). Current challenges in drug-resistant malaria. *Journal of Postgraduate Medicine*, *52*, 241–242.

Gramatica, P. (2007). Principles of QSAR models validation: internal and external. *QSAR & Combinatorial Science*, *26*, 694–701. DOI: 10.1002/qsar.200610151.

Greenwood, B. M., Bojang, K., Whitty, C. J., & Targett, G. A. (2005). Malaria. *Lancet*, *365*, 1487–1498. DOI: 10.1016/s0140-6736(05)66420-3.

Gupta, L., Patel, A., Karthikeyan, C., & Trivedi, P. (2010). QSAR studies on dihydro-alkoxy-benzyl-oxopyrimidines (DABOs) derivatives, a new series of potent, broad-spectrum non-nucleoside reverse transcriptase inhibitors. *Journal of Current Pharmaceutical Research*, *1*, 19–25.

Idro, R., Jenkins, N. E., & Newton, C. R. J. C. (2005). Pathogenesis, clinical features, and neurological outcome of cerebral malaria. *Lancet Neurology*, *4*, 827–840. DOI: 10.1016/s1474-4422(05)70247-7.

Idro, R., Marsh, K., John, C. C., & Newton, C. R. J. (2010). Cerebral malaria: Mechanisms of brain injury and strategies for improved neurocognitive outcome. *Pediatric Research*, *68*, 267–274. DOI: 10.1203/pdr.0b013e3181eee738.

Jain, A. K., Veerasamy, R., Vaidya, A., Kashaw, S., Mourya, V. K., & Agrawal, R. K. (2012). QSAR analysis of B-ring-modified diaryl ether derivatives as a InhA inhibitors. *Medicinal Chemistry Research*, *21*, 145–151. DOI: 10.1007/s00044-010-9518-8.

Jalali-Heravi, M., & Parastar, F. (2000). Use of artificial neural networks in a QSAR study of anti-HIV activity for a large group of HEPT derivatives. *Journal of Chemical Information and Modeling*, *40*, 147–154. DOI: 10.1021/ci990314+.

Karelson, M., Lobanov, V. S., & Katritzky, A. R. (1996). Quantum-chemical descriptors in QSAR/QSPR studies. *Chemical Reviews*, *96*, 1027–1044. DOI: 10.1021/cr950202r.

Khosrokhavar, R., Ghasemi, J. B., & Shiri, F. (2010). 2D quantitative structure-property relationship study of mycotoxins by multiple linear regression and support vector machine. *International Journal of Molecular Sciences*, *11*, 3052–3068. DOI: 10.3390/ijms11093052.

Kshirsagar, N. A. (2006). Malaria: Antimalarial resistance and policy ramifications and challenges. *Journal of Postgraduate Medicine*, *52*, 291–293.

Nantasenamat, C., Naenna, T., Isarankura Na Ayudhya, C., & Prachayasittikul, V. (2005). Quantitative prediction of imprinting factor of molecularly imprinted polymers by artificial neural network. *Journal of Computer-Aided Molecular Design*, *19*, 509–524. DOI: 10.1007/s10822-005-9004-4.

Nantasenamat, C., Isarankura-Na-Ayudhya, C., Tansila, N., Naenna, T., & Prachayasittikul, V. (2007). Prediction of GFP spectral properties using artificial neural network. *Journal of Computational Chemistry*, *28*, 1275–1289. DOI: 10.1002/jcc.20656.

Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., & Prachayasittikul, V. (2008). Prediction of bond dissociation enthalpy of antioxidant phenols by support vector machine. *Journal of Molecular Graphics and Modelling*, *27*, 188–196. DOI: 10.1016/j.jmgm.2008.04.005.

Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., & Prachayasittikul, V. (2009). A practical overview of quantitative structure-activity relationship. *EXCLI Journal*, *8*, 74–88.

Nwaka, S., & Ridley, R. G. (2003). Virtual drug discovery and development for neglected diseases through public-private partnerships. *Nature Reviews Drug Discovery*, *2*, 919–928. DOI: 10.1038/nrd1230.

Nwaka, S., & Hudson, A. (2006). Innovative lead discovery strategies for tropical diseases. *Nature Reviews Drug Discovery*, *5*, 941–955. DOI: 10.1038/nrd2144.

Papa, E., Villa, F., & Gramatica, P. (2005). Statistically validated QSARs, beased on theoretical descriptors, for modelling aquatic toxicity of organic chemicals in *Pimephales promelas* (fathead minnow). *Journal of Chemical Information and Modeling*, *45*, 1256–1266. DOI: 10.1021/ci050212l.

Parr, R. G., Donnelly, R. A., Levy, M., & Palke, W. E. (1978). Electronegativity: The density functional viewpoint. *Journal of Chemical Physics*, *68*, 3801. DOI: 10.1063/1.436185.

Parr, R. G., & Pearson, R. G. (1983). Absolute hardness: companion parameter to absolute electronegativity. *Journal of the American Chemical Society*, *105*, 7512–7516. DOI: 10.1021/ja00364a005.

Parr, R. G., Szentpály, L., & Liu, S. (1999). Electrophilicity index. *Journal of the American Chemical Society*, *121*, 1922–1924. DOI: 10.1021/ja983494x.

Podunavac-Kuzmanović, S. O., & Cvetković, D. D. (2011). QSAR modeling of antibacterial activity of some benzimidazole derivatives. *Chemical Industry & Chemical Engineering Quarterly*, *17*, 33–38. DOI: 10.2298/ciceq100405050p.

Prachayasittikul, S., Wongsawatkul, O., Worachartcheewan, A., Nantasenamat, C., Ruchirawat, S., & Prachayasittikul, V. (2010). Elucidating the structure-activity relationships of the vasorelaxation and antioxidation properties of thionicotinic acid derivatives. *Molecules*, *15*, 198–214. DOI: 10.3390/molecules15010198.

Prasad, Y. R., Kumar, P. R., Smiles, D. J., & Babu, P. A. (2008). QSAR studies on chalcone derivatives as antibacterial agents against *Bacillus pumilis*. *ARKIVOC*, *11*, 266–276.

Rastija, V., & Medić-Šarić, M. (2009). QSAR modeling of anthocyanins, anthocyanidins and catechins as inhibitors of lipid peroxidation using three-dimensional descriptors. *Medicinal Chemistry Research*, *18*, 579–588. DOI: 10.1007/s00044-008-9151-y.

Saghaie, L., Sakhi, H., Sabzyan, H., Shahlaei, M., & Shamshirian, D. (2013). Stepwise MLR and PCR QSAR study of the pharmaceutical activities of antimalarial 3-hydroxypyridinone agents using B3LYP/6-311++G** descriptors. *Medicinal Chemistry Research*, *22*, 1679–1688. DOI: 10.1007/s00044-012-0152-5.

Sawant, R. L., Bansode, C. A., & Wadekar, J. B. (2013). In vitro anti-inflammatory potential and QSAR analysis of oxazolo/thiazolo pyrimidine derivatives. *Medicinal Chemistry Research*, *22*, 1884–1892. DOI: 10.1007/s00044-012-0189-5.

Shahlaei, M., Fassihi, A., & Nezami, A. (2009). QSAR study of some 5-methyl/trifluoromethoxy-1*H*-indole-2,3-dione-3-thiosemicarbazone derivatives as anti-tubercular agents. *Research in Pharmaceutical Sciences*, *4*, 123–131.

Suksrichavalit, T., Prachayasittikul, S., Nantasenamat, C., Isarankura-Na-Ayudhya, C., & Prachayasittikul, V. (2009). Copper complexes of pyridine derivatives with superoxide scavenging and antimicrobial activities. *European Journal of Medicinal Chemistry*, *44*, 3259–3265. DOI: 10.1016/j.ejmech.2009.03.033.

Sun, M., Zheng, Y. G., Wei, H. T., Chen, J. Q., & Ji, M. (2009). QSAR studies on 4-anilino-3-quinolinecarbonitriles as Src kinase inhibitors using robust PCA and both linear and nonlinear models. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *24*, 1109–1116. DOI: 10.1080/14756360802632906.

Suvannang, N., Nantasenamat, C., Isarankura-Na-Ayudhya, C., & Prachayasittikul, V. (2011). Molecular docking of aromatase inhibitors. *Molecules*, *16*, 3597–3617. DOI: 10.3390/molecules16053597.

Talete (2007). Dragon for windows (software for molecular descriptor calculations), version 5.5 [computer software]. Milano, Italy: Talete.

Tanious, F. A., Hamelberg, D., Bailly, C., Czarny, A., Boykin, D. W., & Wilson, W. D. (2004). DNA sequence dependent monomer-dimer binding modulation of asymmetric benzimidazole derivatives. *Journal of the American Chemical Society*, *126*, 143–153. DOI: 10.1021/ja030403+.

Thanikaivelan, P., Subramanian, V., Raghava Rao, J., & Unni Nair, B. (2000). Application of quantum chemical descriptor in quantitative structure activity and structure property relationship. *Chemical Physics Letters*, *323*, 59–70. DOI: 10.1016/s0009-2614(00)00488-7.

Thippakorn, C., Suksrichavalit, T., Nantasenamat, C., Tantimongcolwat, T., Isarankura-Na-Ayudhya, C., Naenna, T., & Prachayasittikul, V. (2009). Modeling the LPS neutralization activity of anti-endotoxins. *Molecules*, *14*, 1869–1888. DOI: 10.3390/molecules14051869.

Vahdani, S., & Bayat, Z. (2011). A quantitative structure-activity relationship (QSAR) study of anti-cancer drugs. *Der Chemica Sinica*, *2*, 235–243.

Veerasamy, R., Ravichandran, S., Jain, A., Rajak, H., & Agrawal, R. K. (2009). QSAR studies on novel anti-HIV agents using FA-MLR, FA-PLS and PCRA techniques. *Digest Journal of Nanomaterials and Biostructures*, *4*, 823–834.

Verma, R. P. (2006). Anti-cancer activities of 1,4-naphthoquinones: a QSAR study. *Anti-Cancer Agents in Medicinal Chemistry*, *6*, 489–499. DOI: 10.2174/187152006778226512.

Verma, R. P., & Hansch, C. (2010). QSAR modeling of taxane analogues against colon cancer. *European Journal of Medicinal Chemistry*, *45*, 1470–1477. DOI: 10.1016/j.ejmech.2009.12.054.

Vlaia, V., Olariu, T., Vlaia, L., Butur, M., Ciubotariu, C., Medeleanu, M., & Ciubotariu, D. (2009). Quantitative structure-activity relationship (QSAR). V. Analysis of the toxicity of aliphatic esters by means of molecular compressibility descriptors. *Farmacia*, *57*, 549–561.

Whitley, D. C., Ford, M. G., & Livingstone, D. J. (2000). Unsupervised forward selection: A method for eliminating redundant variables. *Journal of Chemical Information and Modeling*, *40*, 1160–1168. DOI: 10.1021/ci000384c.

White, N. J. (2004). Antimalarial drug resistance. *Journal of Clinical Investigation*, *113*, 1084–1092. DOI: 10.1172/jci200421682.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques* (3rd ed.). San Francisco, CA, USA: Morgan Kaufmann.

Worachartcheewan, A., Nantasenamat, C., Naenna, T., Isarankura-Na-Ayudhya, C., & Prachayasittikul, V. (2009). Modeling the activity of furin inhibitors using artificial neural network. *European Journal of Medicinal Chemistry*, *44*, 1664–1673. DOI: 10.1016/j.ejmech.2008.09.028.

Worachartcheewan, A., Nantasenamat, C., Isarankura-Na-Ayudhya, C., Prachayasittikul, S., & Prachayasittikul, V. (2011). Predicting the free radical scavenging activity of curcumin derivatives. *Chemometrics and Intelligent Laboratory Systems*, *109*, 207–216. DOI: 10.1016/j.chemolab.2011.09.010.

Worachartcheewan, A., Prachayasittikul, S., Pingaew, R., Nantasenamat, C., Tantimongcolwat, T., Ruchirawat, S., & Prachayasittikul, V. (2012). Antioxidant, cytotoxicity, and QSAR study of 1-adamantylthio derivatives of 3-picoline and phenylpyridines. *Medicinal Chemistry Research*, *21*, 3514–3522. DOI: 10.1007/s00044-011-9903-y.

Yan, X. F. (2011). Multivariate outlier detection based on self-organizing map and adaptive nonlinear map and its application. *Chemometrics and Intelligent Laboratory Systems*, *107*, 251–257. DOI: 10.1016/j.chemolab.2011.04.007.

Zahouily, M., Lazar, M., Elmakssoudi, A., Rakik, J., Elaychi, S., & Rayadh, A. (2006). QSAR for *anti*-malarial activity of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives. *Journal of Molecular Modeling*, *12*, 398–405. DOI: 10.1007/s00894-005-0059-x.

Zhang, L., Zhu, H., Oprea, T. I., Golbraikh, A., & Tropsha, A. (2008). QSAR modeling of the blood-brain barrier permeability for diverse organic compounds. *Pharmaceutical Research*, *25*, 1902–1914. DOI: 10.1007/s11095-008-9609-0.

Zou, C., & Zhou, L. (2007). QSAR study of oxazolidinone antibacterial agents using artificial neural networks. *Molecular Simulation*, *33*, 517–530. DOI: 10.1080/08927020601188528.