ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES Knowledge to Innovation

Vol 7, Issue 1, 2014

ISSN - 0974-2441

Research Article

RATIONALIZATION OF MOLECULAR DESCRIPTORS OF AURONE ANALOGS TOWARD ANTI-MALARIAL ACTIVITY

ARUN K. GUPTA*[1], PRIYANKA SINGH[2], NEETU SABARWAL[2]

1RKDF Inst. of Pharm. Sci., Indore M.P., India-452016., 2Smriti College of Pharmaceutical Education, Indore M.P., India-452010.
Email: arunkg73@gmail.com

Received: 26November 2013, Revised and Accepted: 5 December 2013

ABSTRACT

Malaria is a vector-borne infectious disease that is widespread in tropical and subtropical regions. Every year approximately 450 million clinical cases are caused by malaria parasites and 3.2 billion people in the world are living under the risk of malaria. The growing resistance of the parasites to anti-malarial agents is responsible for some of the worst cases in the tropical world. In view of above and as a part of our effort to develop newer anti-malarial agents, molecular modeling analysis was performed to develope QSAR models that show substantial predictive promise for aurones. The QSAR model explains 86.9 percent variance in activity with low standard error of estimation (0.221). Model showed statistical significant internal predictivity (Q^2=0.816) and external predictivity (r^2_{pred}=0.751) values. The detailed structural investigation revealed that the anti-malarial activity is predominantly explained by the GETAWAY (*H6m & R2v*) and MoRSE-Code (*Mor30v*) descriptors. The structural insights gleaned from the study are helpful in design of inhibitors with enhanced potency.

Keywords: QSAR; Aurone analogs; Anti-malarial activity; Molecular descriptors

INTRODUCTION

Malaria kills over 1.5 million people a year and it is responsible for human misery in tropical countries. Malaria is caused by a protozoal parasite of the genus *Plasmodium* and remains worldwide problem and there is an urgent need for identifying new class of antimalarials [1].The major hurdle in malaria treatment is the spread of resistance to natural or semi-synthetic antiplasmodial drugs [2]. Among the various mechanisms identified, those based on drug transport proteins of the ATP-binding cassette (ABC) family appear to play an important role by pumping drugs out of their target sites. Currently meta-analysis provides compelling evidence that drug resistant malarial parasites are continuing their inexorable global march. Due to the wide spread of malaria and owing to the resistance of the parasite to major drugs, it is urgent to pursue identification for more active and inexpensive drugs to fight malaria [3-6].

Aurones, (2-benzylidenebenzofuran-3(2H)-ones) are secondary metabolites belonging to the flavonoids family. Aurones are structural isomers of flavones, and compared to other flavonoids subclasses, somehow, aurones still by large less studied. The natural origin of aurones, theirs inhibitory effects on (1) erythrocytic stages of *Plasmodium falciparum* strains in vitro [7-10] and (2) efflux pumps involved in the resistance of the parasites [11,12] has prompted us to investigate them as anti-malarial agents targeting resistant strains.

Quantitative structure activity relationship (QSAR) studies are tools of prediction endpoints of interest on organic compounds acting as drugs, which have not been experimentally determined. Many physiological activities of compounds can be related to their composition and structure. The QSAR analysis of the anti-malarial agents is the current highly concerned area of research. The purpose of this study was to develop some good, statistically significant QSAR models to correlate and predict anti-malarial activity of aurones. These QSAR models allow the prediction of anti-malarial activity with an aim to reduce the number of compounds to be synthesized with respect to cost and time. It also aids in the designing of potent and safer inhibitors. The quantification of responsible physicochemical properties was done with the help of regression techniques.

Experimental

Biological activity

The *in vitro* activity data of aurones, 2-benzylidenebenzofuran-3(2*H*)-ones for antimalarial activity was taken from the reported work of Souard *et al.* [1] (**Table 1**) In attempting QSAR, these inhibitory data (IC_{50}) were converted to negative logarithmic dose in moles (pIC_{50}) because a QSAR is a linear free energy relationship, and from the van't Hoff isotherm, free energy change during a process is proportional to the logarithm of the rate or equilibrium constant of the process.

$$\Delta G = - 2.303\ RT \log K$$

Table 1: Structure and activities of azaaurones derivatives used in training and test sets

Comp. No	X	R_1	R_2	R_3	R_4	R_5	R_6	R_7	IC_{50}	pIC_{50}
1	O	OH	OH	H	H	H	H	H	94.5	4.025
2	O	OCH_3	OCH_3	H	H	H	H	H	60.3	4.220
3	O	OH	OH	H	H	CH_3	H	H	63.4	4.198

4	O	OCH_3	OCH_3	C_2H_5	H	H	H	H	21	4.678
5	O	OH	OH	C_2H_5	H	H	H	H	113.5	3.945
6	O	OH	H	H	H	C_2H_5	H	H	28	4.553
7	O	OCH_3	OCH_3	H	H	$C(CH_3)_3$	H	H	13.3	4.876
8	O	OCH_3	OCH_3	H	H	C_4H_{10}	H	H	11.8	4.928
9	O	OCH_3	OCH_3	H	H	Br	H	H	49.8	4.303
10	O	OCH_3	OCH_3	H	H	F	H	H	86.7	4.062
11	O	OH	H	H	H	OH	H	H	130	3.886
12	O	OCH_3	OCH_3	H	H	OCH_3	H	H	11	4.959
13	O	OCH_3	OCH_3	H	H	C_6H_5	H	H	234	3.631
14	O	OCH_3	OCH_3	H	H	C_6H_5N	H	H	85	4.071
15	N	OCH_3	OCH_3	H	H	Br	H	H	49.8	4.303
16	N	OCH_3	OCH_3	H	H	Cl	H	H	17	4.770
17	N	OCH_3	OCH_3	Cl	H	H	H	H	9.9	5.004
18	N	OCH_3	OCH_3	Cl	H	H	Cl	H	8.4	5.076
19	N	OCH_3	OCH_3	Cl	H	H	H	F	9	5.046
20	N	OCH_3	OCH_3	H	H	C_2H_5	H	H	1	6.000
21	N	OCH_3	OCH_3	C_2H_5	H	H	H	H	12.8	4.893
22	N	OCH_3	OCH_3	CH_3	H	H	H	CH_3	9.1	5.041
23	N	OCH_3	OCH_3	CH_3	H	CH_3	H	H	3.6	5.444
24	N	OCH_3	OCH_3	CH_3	H	CH_3	CH_3	H	5.6	5.252
25	N	OCH_3	OCH_3	CH_3	CH_3	H	CH_3	CH_3	8.9	5.051
26	N	OCH_3	OCH_3	H	H	$CH(CH_3)_2$	H	H	4.4	5.357
27	N	OCH_3	OCH_3	H	H	$C(CH_3)_3$	H	H	7.2	5.143
28	N	OCH_3	OCH_3	H	H	C_4H_{10}	H	H	4.1	5.387
29	N	OCH_3	OCH_3	H	H	CCH	H	H	13.4	4.873
30	N	OCH_3	OCH_3	OCH_3	H	OCH_3	H	H	5	5.301
31	N	OCH_3	OCH_3	OCH_3	H	OCH_3	H	OCH_3	1.9	5.721
32	N	OCH_3	OCH_3	H	OCH_3	OCH_3	OCH_3	H	1.9	5.721
33	N	OCH_3	OCH_3	H	H	SCH_3	H	H	6.7	5.174
34	N	OCH_3	OCH_3	H	H	Morpholino	H	H	8.9	5.051
35	N	OCH_3	OCH_3	H	H	$N(CH_3)_2$	H	H	3.7	5.432

Descriptors computation

In pursuit of potent anti-malarial drugs, quantitative structure activity relationship has been investigated through Hansch substituent constant [13]. Computer aided drug design study was performed using *CS ChemBioOffice* [14]. Structure of all the compounds was sketched using builder module of the program. The sketched structures were subjected to energy minimization using molecular mechanics (MM2) until the root mean square (RMS) gradient value became smaller than 0.1 kcal/mol Å. The energy minimized molecules were subjected to re-optimization via Austin model-1 (AM1) method until the RMS gradient attained a value smaller than 0.0001 kcal/mol Å using MOPAC. The geometry optimization of the lowest energy structure was carried out using EF routine. The minimized molecule was saved as "MOL file". These files were used for calculation of various molecular descriptors with the help of *DRAGON* [15].

Regression Analysis

In order to gain an insight to the essential structural and physiochemical requirements for the anti-malarial activity in this class of molecules, 35 compounds were selected. These were divided into training set of 25 compounds and test set of 10 compounds. The data was transferred to the statistical program *VALSTAT* [16] in order to establish a correlation between molecular descriptors as independent variables and pIC_{50} as dependent variable employing sequential multiple linear regression (SEQ-MLR) analysis method. The statistical quality of the SEQ-MLR equations were assessed by parameters like correlation coefficient (r), standard error of estimate (*SEE*), sequential Fischer test (F) at specified degree of freedom (df) and explained variance (r^2_{adj}). The internal predictive powers of the equations were validated by "leave n out" method using predicted residual sum of squares (*PRESS*), cross validation squared correlation coefficient (Q^2), standard deviation based on *PRESS* (S_{PRESS}), total sum of squares (*SSY*) and standard deviation of error of prediction (S_{DEP}). Chances of fortuitous correlation were tested with the help of Y-scrambled test. Robustness of the equations was assessed through bootstrapping method. Finally, selected equations have been validated using test set considering predictive squared correlation coefficient (r^2_{pred}). The ±data within the parenthesis is the standard deviation associated with the coefficient of descriptor in regression equation.

RESULTS AND DISCUSSION

Table 2: Quantification of substituent constant of azaaurone analogs

C. No	MR	π	HA	R HD	$\mathcal{F}$	$\mathcal{R}$	σ	MR_1	π_1	HA_1	R_1 HD	$\mathcal{F}_1$	$\mathcal{R}_1$	σ_1	MR_2	π_2	HA_2	R_2 HD	$\mathcal{F}_2$	$\mathcal{R}_2$	σ_2	IV
1	1.03	0.0	0	0	0.00	0.00	0.00	2.85	-0.67	1	1	0.29	-0.6	-0.4	2.85	-0.67	1	1	0.29	-0.64	0.12	0
2	1.03	0.0	0	0	0.00	0.00	0.00	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	0
3	5.65	0.6	0	0	-0.04	-0.1	-0.2	2.85	-0.67	1	1	0.29	-0.6	-0.4	2.85	-0.67	1	1	0.29	-0.64	0.12	0

4	10.3	1.0	0	0	-0.05	-0.1	-0.1	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	0
5	10.3	1.0	0	0	-0.05	-0.1	-0.1	2.85	-0.67	1	1	0.29	-0.6	-0.4	2.85	-0.67	1	1	0.29	-0.64	0.12	0
6	10.3	1.0	0	0	-0.05	-0.1	-0.2	2.85	-0.67	1	1	0.29	-0.6	-0.4	1.03	0	0	0	0.00	0.00	0.00	0
7	19.6	2.0	0	0	-0.07	-0.1	-0.2	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	0
8	19.6	2.1	0	0	-0.06	-0.1	-0.2	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	0
9	8.88	0.9	0	0	0.44	-0.2	0.23	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	0
10	0.92	0.1	0	1	0.43	-0.3	0.06	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	0
11	2.85	-0.7	1	0	0.29	-0.6	-0.4	2.85	-0.67	1	1	0.29	-0.6	-0.4	1.03	0	0	0	0.00	0.00	0.00	0
12	7.87	-0.7	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	0
13	25.4	2.0	0	0	0.08	-0.1	-0	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	0
14	8.88	0.9	0	0	0.44	-0.2	0.23	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
15	6.03	0.7	0	0	0.41	-0.2	0.23	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
16	6.03	0.7	0	0	0.41	-0.2	0.37	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
17	6.03	0.7	0	0	0.41	-0.2	0.37	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
18	6.95	0.9	0	0	0.84	0.49	0.71	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
19	10.3	1.0	0	0	-0.05	-0.1	-0.2	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
20	10.3	1.0	0	0	-0.05	-0.1	-0.1	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
21	5.65	0.6	0	0	-0.04	-0.1	-0.1	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
22	5.65	0.6	0	0	-0.04	-0.1	0.24	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
23	5.65	0.6	0	0	-0.04	-0.1	0.24	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
24	5.65	0.6	0	0	-0.04	-0.1	0.24	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
25	15	1.5	0	0	-0.05	-0.1	-0.2	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
26	19.6	2	0	0	-0.07	-0.1	-0.2	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
27	19.6	2.1	0	0	-0.06	-0.1	-0.2	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
28	9.55	0.4	0	0	0.19	0.05	0.23	7.87	-0.0	1	0	0.26	-0.5	-0.3	7.87	-0.0	1	0	0.26	-0.5	0.12	1

29	7.87	-0	1	0	0.26	-0.5	-0.2	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
30	7.87	-0	1	0	0.26	-0.5	-0.2	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
31	7.87	-0	1	0	0.26	-0.5	-0.2	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
32	13.8	0.6	0	0	0.2	-0.2	0	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1
33	15.6	0.2	0	0	0.1	-0.9	-0.8	7.87	-0.02	1	0	0.26	-0.5	-0.3	7.87	-0.02	1	0	0.26	-0.51	0.12	1

In the present study, efforts have been made to find the structural requirements for the inhibitory activity of aurones analogs against *P. falciparum*. Quantitative models were developed by means of substituent constant contribution and structural contribution considering regression methodology. Quantification of substituent constant and indicator variable (**Table 2**) with anti-malarial activity in mixed approach offers various mathematical equations. Substantial tri-parametric expression is presented as equation 1.

pIC_{50} = 0.614(± 0.315) π -0.636(± 0.240) σ + 0.801(± 0.151) *IV* + 4.436

n=33, *r*=0.813, *r2adj*=0.626, *SSE*=0.357, *F*=18.86 (Eqⁿ.1)

This equation shows a correlation coefficient (*r*=0.813), which accounted for 62.6% of the explained variance in the activity calculated as $r^2_{adj} = r^2(1–1/F)$ that accounts in percentage when multiplied by 100. The data showed overall internal statistical significance level better than 99.9% as it exceeded the tabulated $F_{(3,29\alpha0.001)}$ = 7.977. Mixed approach analysis revealed that Hansch substituent constant (π) and indicator variable (*IV*) contributed positively while Hammett's substituent constant (σ) contributed negatively to the activity. Positively contribution of Hansch substituent constant at R position of the nucleus indicates that hydrophobic nature of the group favor for the activity. Hydrophocity could be helpful in cell wall in filtration of the *P. falciparum*. Positive contribution of indicator variable (*IV*) indicates that the indolinone ring have higher activity in comparison to benzofuranone system. Negative contribution of Hammett's substituent constant indicates that electron withdrawing groups are unfavorable for the activity. Hammett's constant at R position suggested that the localized electron density on scaffold is favorable. Thus, the electronic interactions seem to be dominating for the activity of the compounds.

A correlation was established between molecular descriptors and biological activity using the sequential multiple linear regression technique considering adjusted square correlation coefficient (r^2_{adj}) [17]. Initially uni-variant expressions were explored and ARR showed significant correlation coefficient value (Eqn. 2) with 68.7% explained variance in the activity.

pIC_{50} = - 8.616(±1.175) *ARR* +10.436

n=25, *r*=0.836, r^2_{adj} =0.687, *SEE*=0.341, *F*=53.79 (Eqⁿ.2)

SEQ-MLR revealed that the r^2_{adj} value is increasing significantly from the uni to the bivariant expressions i.e. 0.687, and 0.826 respectively. Boosting of r^2_{adj} value from uni to bivariant revealed that incorporation of second physicochemical descriptor improve the quality of mathematical expression in a comprehensible manner.

pIC_{50} = -3.038 (±0.691) *Mor30v* -6.870 (±0.962) *ARR* +9.604

n=25, *r*=0.917, r^2_{adj} =0.826, *SEE*=0.255, *F*=57.98 (Eqⁿ.3)

Significant improvement in r^2_{adj} value emphasizes to explore the higher variant expressions. Therefore several tri-variant expressions were developed through SEQ-MLR method. Several statistically significant equations with a coefficient of correlation (*r*) ≥ 0.930, which accounts for more than 85% of the explain variance in the activity were considered for further study (Eqns. 4-8 and **Table 3**).

Table 3: Regression parameters and quality of correlation of tri-parametric equations

Eqn. No.	*n*	r^2	r^2_{adj}	*F*	*QF*	*PE*	*Outlier*
4	25	0.90	0.89	62.89	4.59	0.013	NIL
5	25	0.89	0.87	54.11	4.26	0.015	NIL
6	25	0.89	0.87	53.89	4.25	0.015	NIL
7	25	0.88	0.86	50.11	4.10	0.016	NIL
8	25	0.88	0.86	50.11	4.10	0.016	NIL

pIC_{50} = 4.965(±0.668) *H6m* -10.566(±0.881) *R2v* -0.301(±0.058) *O-059* + 12.311

n=25, *r*=0.949, r^2_{adj} =0.885, *SSE*=0.207, *F*=62.89 (Eqⁿ.4)

pIC_{50} = 3.123(±0.668) *H6m* -2.789(±0.608) *Mor30v* -5.920(±0.750) *R2v* + 8.863

n=25, *r*=0.941, r^2_{adj} =0.869, *SEE*=0.221, *F*=54.11 (Eqⁿ.5)

pIC_{50} = 0.842(±0.162) *Mor03m* -134.488(±12.370) *G2v* - 1.388(±0.146) *Mor07v* + 33.051

n=25, *r*=0.941, r^2_{adj} =0.869, *SSE*=0.221, *F*=53.88 (Eqⁿ.6)

pIC_{50} = 0.604(± 0.119) *nNHR* -1.125(±0.320) *E1m* -5.377(±0.987) *ARR* + 8.476

n=25, *r*=0.937, r^2_{adj} =0.860, *SSE*=0.228, *F*=50.11 (Eqⁿ.7)

pIC_{50} = -1.125(±0.320) *E1m* -0.604(± 0.119) *nROR* -5.377(±0.987) *ARR* + 9.080

n=25, *r*=0.937, r^2_{adj} =0.860, *SSE*=0.228, *F*=50.11 (Eqⁿ.8)

A high correlation coefficient alone is not enough to select the equation as a model and hence various statistical approaches were employed to confirm the robustness and the practical applicability of the equations. The orthogonality of the descriptors in the equations was established through variance inflation factor (*VIF*) [18,19] values and pair-wise correlation among the descriptors (**Table 4**). The *VIF* is defined as $1/(1–r^2_i)$, where r_i is the multiple correlation coefficient for the i^{th} variable regressed on D-1 others, D is being the number of descriptors contributed to the model. *VIF* value larger than 10 indicates that the information of the descriptors may be hidden by the correlation of the other descriptors. In models *VIF* values of these descriptors positioned in the range of 2.05 to 2.33 (**Table 4**). Therefore, from *VIF* analysis it is clear that the

descriptors used in equations are considerably self-governing. The low value of pair-wise correlation (*PWC*) among the descriptors (< 0.430) also supported comparatively independent contribution. We have also made efforts to investigate predictive power of the equations by using quality factor (*QF*) [20,21] considering Pogliani's method. *QF* is defined as the ratio of correlation coefficient to standard error of estimation (*SEE*) that is *QF = r/SEE*. Obviously, the larger value of r, the smaller *SEE*, and higher will be *QF*, as well as better will be the predictive power of the model. *QF* value for Eqn. 4 to 8 falls in between 4.10–4.59.

Table 4: Pair wise correlation and VIF values of the descriptors used in QSAR models

Eqn.	*Descriptors*	*VIF*	*Pair wise correlation*		
			H6m	R2V	O-059
	H6m	1.18	1		
	R2v	1.93	0.07	1	
4	*O-059*	2.05	0.25	0.65	1
			Mor 30V	*H6m*	*R2V*
	Mor 30v	1.24	1		
	H6m	1.03	0.12	1	
5	*R2v*	1.23	0.41	0.06	1
			G2v	*Mor03m*	*Mor07v*
	G2v	2.31	1		
	Mor03m	2.33	0.59	1	
6	*Mor07v*	1.66	0.28	0.29	1
			E1m	*nNHR*	*ARR*
	E1m	1.13	1		
	nNHR	1.56	0.15	1	
7	*ARR*	1.57	0.17	0.54	1
			E1m	*nROR*	*ARR*
	E1m	1.13	1		
	nROR	1.56	0.15	1	
8	*ARR*	1.57	0.17	0.54	1

Table 5: Internal and external statistics of tri-parametric equations

Eqn. No.	*n*	*Bootstrapping*		*Randomized*				*Leave One Out*			*Leave Three Out*			*Test set (n=10)*	
		r^2_{bs}	SE_{bs}	*Chance*	R^2_{max}	R^2_{mean}	R^2_{std}	$_1Q^2$	S_{PRESS}	S_{DEP}	$_7Q^2$	S_{PRESS}	S_{DEP}	r^2_{pred}	SEP
4	25	0.902	0.057	0.001	0.536	0.124	0.09	0.865	0.239	0.219	0.817	0.26	0.256	0.645	0.274
5	25	0.907	0.056	0.001	0.496	0.121	0.089	0.816	0.28	0.257	0.774	0.289	0.284	0.751	0.253
6	25	0.899	0.062	0.001	0.49	0.122	0.089	0.800	0.292	0.268	0.806	0.263	0.263	0.461	0.374
7	25	0.886	0.088	0.001	0.536	0.127	0.089	0.826	0.272	0.25	0.846	0.173	0.158	0.466	0.349
8	25	0.891	0.033	0.001	0.518	0.122	0.085	0.826	0.272	0.25	0.846	0.22	0.217	0.613	0.317

Reliability of the model, we have calculated regression associated statistical parameter called probable error of correlation (*PE*). Goodness of fit is calculated as $PE = 2\ (1- r^2)/3\sqrt{n}$, if the value of correlation coefficient (*r*) is more than six times of *PE* than the expression is good and reliable. The robust QSAR models should have to satisfy both statistical quality and predictive power. Therefore, all the expressions were tested for internal and external corroboration (**Table 5**). Both the validations put forward decision-making input for selection of QSAR models. Internal corroboration was carried out using leave-n-out cross-validation method and Y-scrambling test. Although equations showed good internal consistency ($_1Q^2$ = 0.800-0.865 & $_7Q^2$ = 0.816-0.846), they may not be applicable for the analogs, which were never used in the generation of the correlation. Therefore, predictive power of Eqns. (4-8) was further confirmed by a test set of ten compounds. On the basis of statistical studies, Eqn. 5 is considered as best model for further study.

The best QSAR model having coefficient of correlation (r=0.941) which explain 86.9% variance in the activity (**Table 6 & Fig. 1**). The model showed overall internal statistical significance level more than 99.9% as it exceeded the tabulated $F_{(3,21\ \alpha\ 0.001)}$ =8.995. Sequential Fischer test recommended that equations are applicable for more than 999 times out of 1000. The QSAR model has correlation coefficient significantly higher than 6*PE* supporting reliability and goodness. The model was further analyzed for the outlier by the Z score method (Z_{value}), the outlier test helps in the identification of unexplainable structurally diverse analogs. The persuasive QSAR model should not have any outlier. The Z_{value} for individual compounds lies within the specific range <|2.5|, which indicated the absence of outliers (**Table 6**). Test revealed that the model is able to explain the structurally diverse analogs and is helpful in the designing of more potent compounds using physiochemical descriptors.

Bootstrapping technique was employed to confirm the contribution of physicochemical properties of the molecules to the activity whether equi-intense or of different rank. The value of the bootstrapping squared correlation coefficient (r^2_{bs} = 0.907) and the bootstrapping standard deviation (SE_{bs} = 0.056) implies that the equations were proper representatives of the group of analogs. The chance of fortuitous correlation was checked with the help of Y-scrambling data test. *Chance* value of 0.001 corresponds to 0.1 % chance of fortuitous correlation. *Chance* value (less than 0.001) of model revealed that the result was not based on prospective correlation. Similarly mean randomized squared correlation coefficient (R^2_{mean} = 0.121) and randomized standard deviation (SE_{rand} = 0.089) are also supporting that the results are not based on chance correlation (**Table 5**).

Table 6: Calculated, calculated (loo), residual and Z-score of azaaurone derivatives obtained from model

C. No.	[a]cal	[b]cal_{res}	[c]Z_{value}	[d]loo	[e]loo_{res}
1	4.034	-0.009	-0.046	4.036	-0.011
2	4.646	-0.426	-2.063	4.695	-0.475
3	4.130	0.068	0.331	4.116	0.082
4	4.671	0.007	0.034	4.670	0.007
7	4.601	0.275	1.332	4.533	0.343
8	5.215	-0.287	-1.387	5.235	-0.307
10	3.896	0.166	0.802	3.794	0.268
11	3.854	0.032	0.153	3.846	0.040
13	3.987	-0.356	-1.721	4.099	-0.468
15	4.333	-0.030	-0.146	4.336	-0.034
16	4.843	-0.073	-0.355	4.850	-0.081
17	4.925	0.079	0.384	4.916	0.089
18	4.923	0.152	0.737	4.456	0.620
20	5.542	0.458	2.215	5.466	0.534
21	4.843	0.050	0.240	4.839	0.053
22	5.079	-0.038	-0.184	5.082	-0.041
24	5.122	0.130	0.630	5.113	0.139
25	5.304	-0.253	-1.224	5.330	-0.280

26	5.423	-0.066	-0.320	5.431	-0.074
29	4.956	-0.083	-0.402	4.960	-0.087
30	5.321	-0.020	-0.098	5.325	-0.024
31	5.883	-0.162	-0.782	5.923	-0.201
32	5.693	0.028	0.135	5.684	0.037
33	4.779	0.394	1.908	4.760	0.414
35	5.468	-0.036	-0.176	5.472	-0.041

[a]Observed pIC_{50} value of compound, [b]Calculated pIC_{50} value of compound using model-2 [c]Residual pIC_{50} value of calculated data, [d]Z-score value obtained from model-2,[e]Calculated (loo) data of the compounds, [f]Residual value of calculated (loo) data of the compounds.

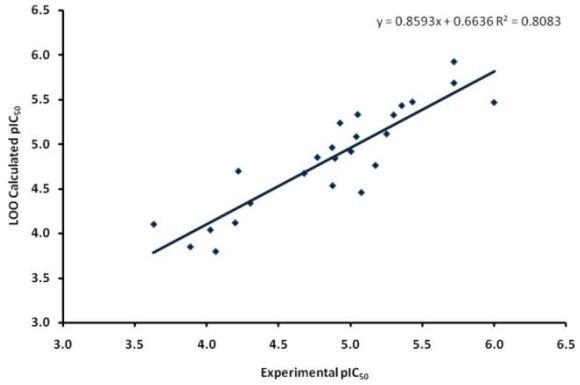

Fig. 1: Graphical representation of experimental versus calculated loo pIC_{50} of training set

The selected mathematical expression (model) are able to predict the activity of test set compound, which supported by r^2_{pred} (0.751) and low standard error of prediction (0.253) values. The predicted activity of test set compounds are very close to their actual activity, which indicate the robustness of model (**Table 7 & Fig. 2**). Reproducibility of predictiveness of the inhibitory activity of azurones by contributing descriptors was confirmed by random fifty run of data set. Random training and test set run depicted the mean r^2_{pred} value (0.768) equivalent to model r^2_{pred} value (0.751) with low standard error (0.162).

QSAR model assessment revealed that molecular descriptors *H6m* is contributing positively while *Mor30v* and *R2v* are contributing negatively to inhibitory activity. *H6m* is belong to the GETAWAY [22,23] class of descriptors, which represents [GEometry, Topology and Atom-Weights AssemblY] group of descriptors. These molecular descriptors match the three dimensional molecular geometry provided by the molecular influence matrix and atom relatedness by molecular topology, with chemical information by using various atomic weight schemes like atomic mass, polarizability, van der Waals volume, and electro- negativity. Therefore, this class of descriptors is highly sensitive to the 3-dimensional molecular structure. GETAWAY descriptors are used to compare molecules or even conformers taking into account their molecular shape, size, symmetry and atom distributions. The *H6m* is an H index autocorrelation of lag 6 weighted by atomic mass, and all H values provided information on the interaction degree between atom pairs. *Mor30v* is 3D MoRSE code descriptor; Morse code is a 3D molecular representation of structure based on electron diffraction. MoRSE code [24-27] was calculated by summing atom weights viewed by a different angular scattering function. The values of these code functions were calculated at 32 evenly distributed values of scattering angle(s) in the range of 0-31 Å-1 from the three dimensional atomic co-ordinates of a molecule. The 3D-MoRSE code was calculated using following expression;

$$I(s) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} A_i A_j \left(\sin(sr_{ij})/sr_{ij}\right)$$

Where, s is scattering angle, r_{ij} is interatomic distance of i^{th} and j^{th} atom, A_i and A_j are atomic properties of i^{th} and j^{th} atom respectively including van der Waals volume, atomic mass, Sanderson atomic electronegativity and atomic polarizability. The negatively contribution of *Mor30v* revealed that after 30.0 Å interatomic distance non-bonded clashes might be appears as enthalpy penalty for the interactions.

R2v is the GETAWAY [22,23] class of descriptors represents [GEometry, Topology and Atom-Weights AssemblY] group of descriptors, which are based on a leverage matrix. Negative contribution of *R2v* descriptor encoding both geometrical information given by the influence molecular matrix and the topological information given by the molecular graph, weighted by van der Waals volumes is significant for the activity. In general the model fulfills all the statistical validation criteria to a significant extent. On the basis of the findings, probable interactions of aurones with macromolecule are illustrated in **Fig. 3**.

Table 7: Predicted and residual value of test set of azaaurone derivatives obtained from model

C. No.	[a]Pred	[b]$Pred_{res}$
5	3.888	0.057
6	4.096	0.457
9	4.086	0.217
12	4.684	0.275
14	4.064	0.007
19	5.035	0.010
23	4.958	0.486
27	5.127	0.016
28	5.110	0.277
34	5.016	0.034

[a]Observed pIC_{50} value of compound, [b]Predicted pIC_{50} value of test compounds using model-2 Residual pIC_{50} value of predicted data.

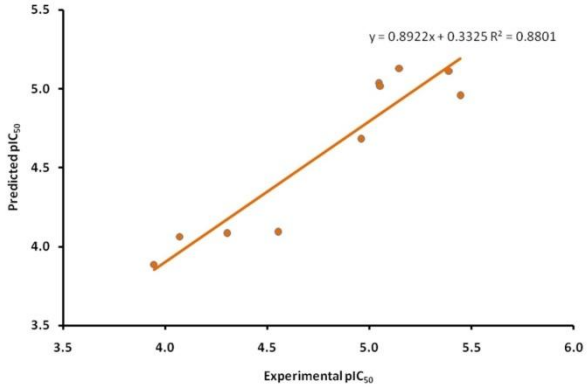

Fig. 2: Graphical representation of experimental versus predicted pIC_{50} of test set

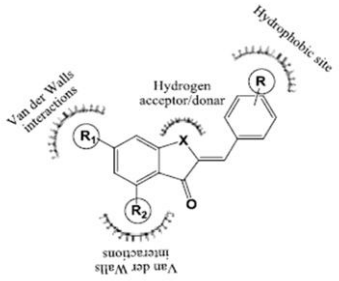

Fig. 3: Illustrative representation of aurones interactions with receptor

CONCLUSION

In this study, molecular feature based quantification of inhibitory activity and binding key interaction with macromolecule have been explored. QSAR results elucidate that the hydrophobic, steric effect (van der Waals volume) and electrostatic interactions affect activities of aurone analogs as anti-malarial. The values of coefficient of determination and cross validated coefficient of prediction obtained from model are 0.885 and 0.816, respectively. Moreover,

test set data showed coefficient of prediction 0.751 with low value of standard error (0.253). The results show that the QSAR model is robust and has good predictive ability. These models are not only able to predict the activity of test compounds but also explained the important structural features of the molecules in a quantitative manner. The study provided useful clues about the structural requirement for effective anti-malarial target interaction chemistry and hence for the improvement of the biological activity. In conclusion, the results derived in present study can provide a preliminary valuable guidance for continuing search for potential anti-malarial prior to synthesis.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

1. Souard F, Okombi S, Beney C, Chevalley S, Valentin A, Boumendjel A 1-Azaaurones derived from the naturally occurring aurones as potential antimalarial drugs. Bioorg Med Chem 2010; 18: 5724-5731.
2. Turschner S, Efferth T Drug resistance in plasmodium: natural products in the fight against malaria. Mini-Rev Med Chem 2009; 9: 206-214.
3. Guglielmo S, Bertinaria M, Rolando B, Crosetti M, Fruttero R, Yardley V, Croft SL, Gasco A A new series of amodiaquine analogues modified in the basic side chain with in vitro antileishmanial and antiplasmodial activity. Eur J Med Chem 2009; 44: 5071-5079.
4. Micale N, Ettari R, Schirmeister T, Evers A, Gelhaus C, Leippe M, Zappala M, Grasso S Novel 2*H*-isoquinolin-3-ones as antiplasmodial falcipain-2 inhibitors. Bioorg Med Chem 2009; 17: 6505-6511.
5. el Sayed I, Van der Veken P, Steert K, Dhooghe L, Hostyn S, Van Baelen G, Lemiere G, Maes BUW, Cos P, Maes L, Joossens J, Haemers A, Pieters L, Augustyns K Synthesis and antiplasmodial activity of alkylamino-substituted neocryptolepine derivatives. J Med Chem 2009; 52: 2979-2988.
6. Van Baelen G, Hostyn S, Dhooghe L, Tapolcsanyi P, Matyus P, Lemiere G, Dommisse R, Kaiser M, Brun R, Cos P, Maes L, Hajos G, Riedl Z, Nagy I, Maes BUW Pieters L Structure-activity relationship of antiparasitic and cytotoxic indoloquinoline alkaloids, and their tricyclic and bicyclic analogues. Bioorg Med Chem 2009; 17: 7209-7217.
7. Kayser O, Kiderlen AF, Folkens U, Kokodziej H In vitro leishmanicidal activity of aurones. Planta Med 1999; 65: 316-319.
8. Kayser O, Chen M, Kharazmi A, Kiderlen AF Aurones interfere with leishmania major mitochondrial fumarate reductase. Z Naturforsch C. 2002; 57: 717-720.
9. Sairafianpour M, Kayser O, Christensen J, Asfa M, Witt M, Staerk D, Jaroszewski JW Leishmanicidal and antiplasmodial activity of constituents of Smirnowia iranica. J Nat Prod 2002; 65: 1754-1758.
10. Boumendjel A Aurones: a subclass of flavones with promising biological potential. Curr Med Chem 2003; 10: 2621-2630.
11. Boumendjel A, Beney C, Deka N, Mariotte AM, Lawson MA, Trompier D, Baubichon-Cortay H, Di Pietro A 4-Hydroxy-6-methoxyaurones with high-affinity binding to cytosolic domain of P-glycoprotein. Chem Pharm Bull 2002; 50: 854-856.
12. Sim HM, Lee CY, Ee PLR, Go ML Dimethoxyaurones: potent inhibitors of ABCG2 (breast cancer resistance protein). Eur J Pharm Sci 2008; 35: 293-306.
13. Hansch C, Leo A Substituent Constants for Correlation Analysis in Chemistry and Biology. New York, USA; John Wiley; 1979.
14. CS ChemOffice, Version 11.0, Cambridge Soft Corporation, Software Publishers Association, 1730 M Street, Suite 700, Washington D.C. 20036 (202) 452-1600, USA.
15. Todeschini R, Consonni V DRAGON-Software for the Calculation of Molecular Descriptors, rel. 1.12 for Windows, 2001.
16. Gupta AK, Babu MA, Kaskhedikar SG VALSTAT : validation program for quantitative structure activity relationship studies. Indian J Pharm Sci 2004; 66: 396-402.
17. Gupta RA, Gupta AK, Soni LK, Kaskhedikar SG 2-(pyrazin-2-yloxy)acetohydrazide analogs QSAR study: an insight into the structural basis of antimycobacterial activity. Chem Biol Drug Des 2010; 76: 441-450.
18. Gupta AK, Sabarwal N, Agrawal YP, Prachand S, Jain S Insights through AM1 calculations into the structural requirement of 3,4,6-substituted-2-quinolone analogs towards FMS kinase inhibitory activity. Eur J Med Chem 2010; 45: 3472-3479
19. Shapiro S, Guggenhei B Inhibition of oral bacteria by phenolic compounds. part 1. QSAR analysis using molecular connectivity. Quant Struct Act Relat 1998; 17: 327-337.
20. Pogliani L Modeling with special descriptors derived from a medium-sized set of connectivity indices. J Phys Chem 1996; 100: 18065-18077.
21. Mandloi D, Joshi S, Khadikar PV, Khosla N QSAR study on the antibacterial activity of some sulfa drugs: building blockers of Mannich bases. Bioorg Med Chem Lett 2005; 15: 405-411.
22. Fedorowicz A, Zheng L, Singh H, Demchuk E QSAR study of skin sensitization using local lymph node assay data. Int J Mol Sci 2004; 5: 56-66.
23. Gonzalez MP, Teran C, Teijeira M, Gonzalez-Moa MJ GETAWAY descriptors to predicting A (2A) adenosine receptors agonists. Eur J Med Chem 2005; 40: 1080-1086.
24. Gasteiger J, Sadowski J, Schuur J, Selzer P, Steinhauer L, Steinhauer V Chemical information in 3D space. J Chem Inf Comput Sci 1996; 36: 1030-1037.
25. Schuur JH, Selzer P, Gasteiger J The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure spectra correlations and studies of biological activity. J Chem Inf Comput Sci 1996; 36: 334-344.
26. Schuur J, Gasteiger J Infrared spectra simulation of substituted benzene derivatives on the basis of a novel 3D structure representation. J Anal Chem 1997; 69: 2398-2405.
27. Todeschini R, Consonni V Handbook of Molecular Descriptors, vol.11, Weinheim: Wiley-VCH; 2000.