March 2010 ISSN: 0974 - 7516 Volume 6 Issue 1

Trade Science Inc.

Organic CHEMISTRY

An Indian Journal

Full Paper

OCAIJ, 6(1), 2010 [30-38]

Correlation potential of Wiener index vis-à-vis molecular refractivity: Anti-malarial activity of xanthone derivatives

Vijay H.Masand[1]*, Komalsingh N.Patil[1], Ghulam M.Nazerruddin[2], Rahul D.Jawarkar[3], Sumit O.Bajaj[4]

[1]Department of Chemistry, Vidyabharati College, Camp, Amravati, Maharashtra, (INDIA)

[2]Department of Chemistry, Poona College, Pune-1, Maharashtra, (INDIA)

[3]Department of Pharmaceutical Chemistry, Vidyabharati College of Pharmacy, Camp, Amaravati, Maharashtra, (INDIA)

[4]C.Eugene Bennett Department of Chemistry, West Virginia University, Morgantown, West Virginia 26506-6045, (USA)

E-mail : vijaymasand@gmail.com; vijaymasand@rediffmail.com

Received: 15th November, 2009 ; Accepted: 25th November, 2009

ABSTRACT

A quantitative structure-activity relationship (QSAR) analysis of seventeen structurally diverse derivatives of xanthone recently reported as anti-malarial has been performed using Wiener index, MR, logP, suitable indicator variables and various 2D descriptors. These revealed several important physicochemical and structural requirements for anti-malarial activity. Twenty QSAR models reported herein provide interesting insights in understanding the hydrophobic, steric, electronic, and structural requirements of anti-malarial activity among these individual set of compounds. The application of a multiple linear regression analysis indicated that a combination of topological indices with the ad hoc molecular descriptors and the indicator parameters yielded a statistically significant model for the activity. $R^2_{(LOO)}$, R^2pred, R^2adj, PSE are used to validate the models. We have found that, among the various parameters, Wiener index has highest prediction ability. These results may be used to further the design and development of new anti-malarial compounds with better activity.

KEYWORDS

Wiener index;
Molecular refractivity;
Anti-malarial;
Xanthone derivatives.

INTRODUCTION

According to the WHO report 2008[1] Malaria, one of the life threatening diseases, still remains the most significant parasitic disease in the tropics and sub-tropics, where it causes at least 500 million clinical episodes and claims 1.5 million lives each year, mostly young children and pregnant women. Emerging widespread resistance to the available best and less expensive antimalarials like Quinine, Chloroquine, Mefloquine (Figure 1) and S/P *(i.e.*, a combination of sulfadoxine and pyrimethamine), combined with an increasing tolerance to insecticides in the mosquito vector, threaten a global malaria tragedy unless new countermeasures are developed[2]. Multidrug-resistant malaria is a serious problem in Southeast Asia and travelers to this region are recommended to use mefloquine or halofantrine-expensive drugs which have serious side effects such as psychiatric or cardiotoxic complications, respectively.

The xanthone nucleus or *9H*-xanthen-*9*–one (Figure 2) comprises important class of oxygenated heterocycles with diverse pharmaceutical applications in-

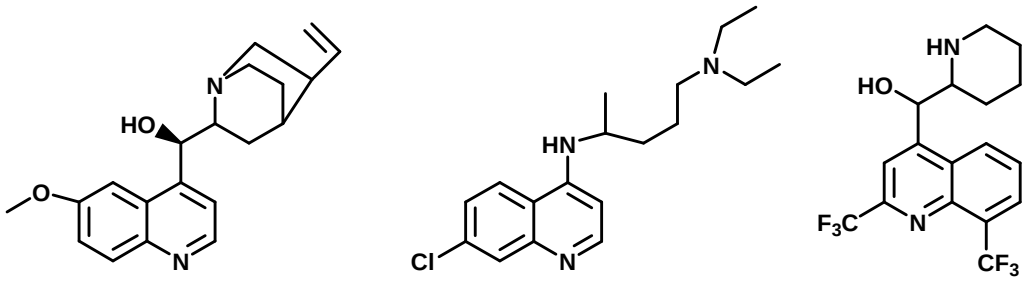


Figure 1 : Chemical structures of commercially available antimalarial drugs

cluding anti-malarial activity[3]. The discovery of xanthones as novel antiparasitic agents with potent activity against Plasmodium parasites has been widely reported in literature but search for a xanthone derivative with very high activity with minimum side effects is still under development.

Quantitative Structure–Activity Relationships (QSARs) are efforts to associate molecular structure with chemical properties or biochemical activities. The basic concept of a QSAR model is that the numerical value of a specified biological activity measured for a set of molecules depends on the structure of these molecules. QSAR has been applied successfully in many disciplines, pertaining to drug design and environmental risk assessments[4]. To get an insight into the structure-activity relationship we need molecular descriptors that can efficiently characterize molecular size, molecular branching or the variations in molecular shapes, and can influence the structure and its activities. In present study, a number of successful multiple regression QSAR models are developed on well established physicochemical parameters like Weiner index, MR, logP, total valence connectivity (TVC), total connectivity (TC), suitable indicator parameters and various 2D descriptors. The work describes QSAR studies on structurally diverse xanthone derivatives synthesized by Riscoe and coworkers[5]. This study may help us to design new analogues with better biological profile.

To obtain a statistically significant model, we have followed maximum R^2 method; we note that the maximum R^2 method actually includes a combination of standard error, adjusted R^2 value[6], R, standard error of estimation, *F*- ratio and Q-test. In order to avoid the probability of "Over fitting" the maximum number of descriptors in each multivariate equation is restricted to four by using variable selection method and to validate the models, we calculated $R^2_{pred.}$, in addition, for better validation; we have considered the PSE and $R^2_{(LOO)}$ values also.

EXPERIMENTAL

Database

The data set of IC_{50} related to anti-malarial activity was collected from literature[2] and converted into – $logIC_{50}$ (pIC_{50}) for convenience. The compounds include structurally diverse xanthone derivatives with substituents like -Cl, -OH etc. (Figure 2). A complete list of compounds names, corresponding -$logIC_{50}$ pIC_{50}) along with indicator parameter values are listed in TABLE 1.

Molecular descriptors and computer programs

Chem Sketch software (ACD labs 12. freeware)[7] was used to draw 2D and 3D structures of the molecules and the structures were optimized to obtain descriptors like MR, logP, index of refraction (IR) Surface tension (ST). The topological indices like Wiener Index (W), Total Connectivity Index (TC), Total Valence Connectivity (TVC), energies of HOMO, LUMO and dipole were calculated by Chem Draw 3D version 11.0[8]. The various molecular descriptors are listed in TABLE 2.

Hyperchem 8.0 was used to obtain the electrostatic potential 2Dcontour diagrams of hematin, compound **(11)** and compound **(2)**[8]. MS-Excel and Minitab 14.0 were used to perform various statistical functions. Since

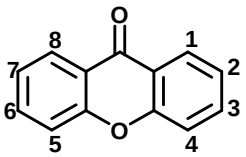


Figure 2 : Xanthone nucleus with numbering

Full Paper

TABLE 1 : Name, pIC_{50} and indicator parameters used in the present study

Sr. No.	Compound Name	IC_{50}	pIC_{50}	I2	I4	I5	I7
1	3,6-bis-(diethylamino)xanthone	20	-1.3010	0	0	0	0
2	4,5-diamidinoxanthone	50	-1.6989	0	0	0	0
3	3,6-bis-(ω-hydroxypentyloxy)xanthone	31.4	-1.4969	0	0	0	0
4	3,6-bis-(ω-diethylaminopentyloxy)xanthone	0.12	0.9208	0	0	0	0
5	4,4'-bis-(ω-diethylaminopentyloxy)benzophenone	0.55	0.2596	0	0	0	0
6	3-ω-diethylaminopentyloxyxanthone	4.60	-0.6627	0	0	0	0
7	3,5-bis-(ω-diethylaminopentyloxy)xanthone	0.93	0.0315	0	0	0	0
8	4,5-bis-(ω-diethylaminopentyloxy)xanthone	0.72	0.1426	0	0	0	0
9	3,6-bis-(5-morpholinopentyloxy)xanthone	0.96	0.0177	0	0	0	0
10	3,6-bis-[5-(4-methyl-piperazinyl)pentyloxy]xanthone	0.21	0.6778	0	0	0	0
11	3,6-bis-(5-piperidinylpentyloxy)xanthone	0.04	1.3979	0	0	0	0
12	3,6-bis-(5-pyrrolidinylpentyloxy)xanthone	0.04	1.3979	0	0	0	0
13	3,6-bis-ω-diethylaminohexyloxyxanthone	0.1	1.0000	0	0	0	0
14	4-chloro-3,6-bis-ω-diethylaminohexyloxyxanthone	0.05	1.3010	0	1	0	0
15	4,5-dichloro-3,6-bis-ω-diethylaminohexyloxyxanthone	0.15	0.8239	0	1	1	0
16	2,4,5,7-tetrachloro-3,6-bis-ω-diethylaminohexyloxyxanthone	0.27	0.5686	1	1	1	1
17	Mangostin	5.1	-0.7076	0	0	0	0

I2, I4, I5, I7 = 1 if –Cl is as substitution on xanthone nucleus at respective position, 0 otherwise

the calculations of these topological indices are well documented in the literature, it is not necessary to duplicate the same here.

Technique

Before the quantum chemical computations the energy of conformation was minimized by using MM2 force field with following used parameters:

Minimum RMS gradient: 0.100 step interval: 2.0 fs frame interval: 10 fs termination after: 2000 steps

Target temperature: 300K

The energy minimization was followed by calculations of energies of HOMO, LUMO and dipole for the compounds considered here by using Chem Draw 11.0 3D version. The dipole was calculated with following parameters:

Method: AM1 wave function, close shell (restricted)

Charges: Mulliken.

The Electrostatic potential 2D contour diagrams were obtained by using Semi-empirical ZINDO/S method (with default settings except Convergence limit: 0.001 and Iteration limit: 50) followed by Single point calculations in HyperChem 8.0. In multiple regression analysis, the independent variables must be orthogonal[9] and consequently the autocorrelation among the descriptors was checked and is given in the correlation matrix in TABLE 3.

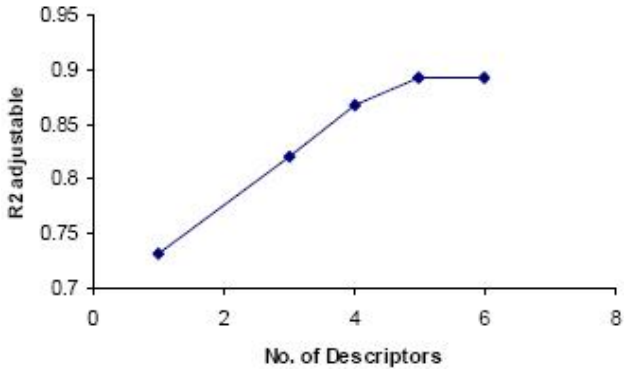


Figure 3 : Correlation of the adjusted R^2 value with number of descriptors

Descriptor selection

Once the descriptors had been generated, variable selection was performed to reduce the number of descriptors per compound. Objective feature selection was carried out to choose a subset of descriptors that are best in encoding the activity of interest, since many of the calculated descriptors carry redundant and highly correlated information or very little useful information. Objective feature selection uses the independent variables alone to filter out non-useful descriptors without using the dependent variables. This procedure[10] involves:

1. All descriptors with same values for all molecules were omitted.
2. The input variables in Multiple Linear Regression (MLR) must not be highly correlated. Therefore, one of the two descriptors that has the pair wise correlation coefficient above 0.9 (R>0.9) and has a large correlation coefficient with the other descriptors in each class was eliminated.

Interestingly, we obtained MR (a physico-chemical parameter) and W (a topological index) with high correlation coefficient (0.973) with each other (see TABLE 3), so according to the procedure of objective feature selection the best thing is to avoid any one of them, instead of this we chose to check the prediction ability of W as well as MR.

Optimum number of descriptors to be used

A major decision in developing successive QSAR model is when to stop adding descriptors to the model.

TABLE 2 : Molecular descriptors of the compounds

Sr. no.	pIC50	W	logP	MR	TC	TVC	HOMO	LUMO	IR	ST	Dipole
1	-1.3010	1496	6.05	102.69	0.0005	2.4112	-8.161	-4.187	1.611	48.3	4.864
2	-1.6989	820	-0.11	73.92	0.0006	1.0046	-10.726	-5.382	1.761	71.9	3.473
3	-1.4969	2720	3.46	109.02	2.9531	5.0234	-11.392	-4.615	1.577	51.8	2.937
4	0.9208	5856	7.15	150.45	2.461	1.2558	-9.335	-4.625	1.539	41.7	1.482
5	0.2596	5488	6.87	146.33	7.383	5.0086	-9.382	-4.812	1.524	38.3	5.123
6	-0.6627	2046	5.27	102.99	8.3528	8.9862	-9.298	-4.974	1.563	43.8	2.628
7	0.0315	5636	6.8	150.45	2.4609	1.2558	-9.352	-4.800	1.539	41.7	0.636
8	0.1426	5416	6.45	150.45	2.461	1.2558	-9.387	-4.921	1.539	41.7	2.348
9	0.0177	6858	4.24	149.21	3.0762	5.2328	-7.963	-4.616	1.557	45.6	1.079
10	0.6778	7996	3.68	162.55	2.0508	6.2793	-9.133	-4.577	1.559	43.9	1.849
11	1.3979	6858	7.75	154.78	3.0762	1.3082	-10.435	-4.587	1.543	41	1.969
12	1.3979	5846	6.62	145.56	6.1524	2.6164	-10.242	-4.584	1.552	42.4	2.123
13	1.0000	6998	8.21	159.72	1.2305	6.2793	-9.209	-4.576	1.534	41.3	2.652
14	1.3010	7302	8.72	164.61	1.0046	6.1662	-9.081	-4.484	1.539	42.1	1.741
15	0.8239	7612	9.23	169.51	8.2032	6.0551	-9.079	-4.386	1.543	42.9	2.196
16	0.5686	8316	10.34	179.3	5.4688	5.8388	-9.09	-4.342	1.552	44.5	1.197
17	-0.7076	2106	5.44	114.55	6.0623	4.7656	-10.883	-3.804	1.624	53.9	2.953

TABLE 3 : Correlation matrix of molecular descriptors

	pIC50	W	logP	MR	TC	TVC	HOMO	LUMO	IR	ST	Dipole
W	0.864	1									
logP	0.700	0.643	1								
MR	0.865	0.973	0.768	1							
TC	0.165	0.097	0.318	0.164	1						
TVC	0.059	0.181	0.182	0.167	0.47	1					
HOMO	0.248	0.397	0.352	0.406	-0.087	0.227	1				
LUMO	0.169	0.156	0.464	0.276	0.164	0.218	0.095	1			
IR	-0.705	-0.707	-0.734	-0.79	-0.327	-0.27	-0.383	-0.246	1		
ST	-0.729	-0.675	-0.748	-0.763	-0.319	-0.219	-0.461	-0.236	0.977	1	
Dipole	-0.484	-0.589	-0.266	-0.554	0.0328	0.007	-0.126	-0.012	0.345	0.263	1

A simple technique to control the model expansion is the so-called "breaking point" in the improvement of the statistical quality of the model, by analyzing the plot of the number of descriptors involved in the models obtained versus the adjusted R^2 value. Consequently, the model corresponding to the breaking point is considered the optimum model.

The graph between the numbers of parameters used in the models against the adjusted R^2 value is as shown in figure 3. The figure indicates that the optimum number of descriptors is to be used is four. Therefore QSAR models with descriptors more than four are not considered.

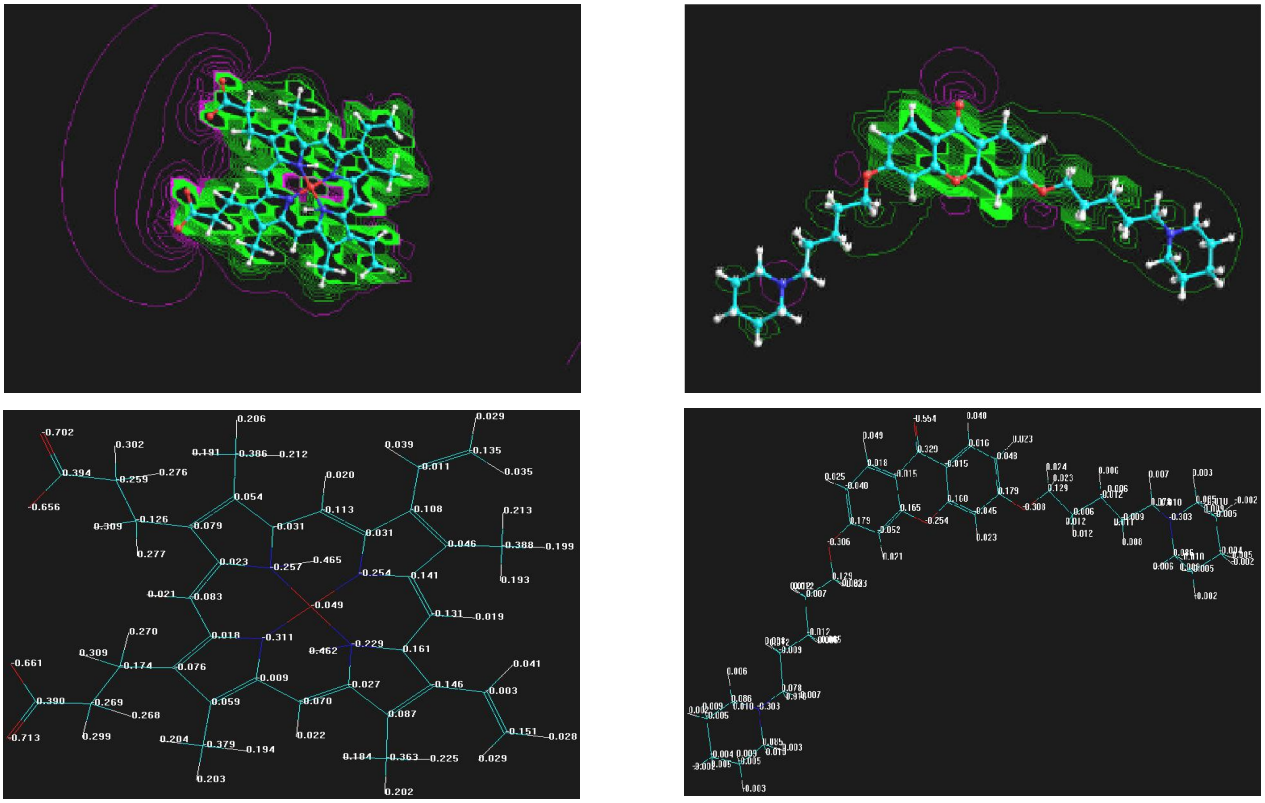
Defining model applicability domain

For a QSAR model to be more useful for screening new compounds, its domain of application[11,12] must be defined and predictions which fall into this domain may be considered reliable for only those compounds. Extent of extrapolation[13] is one simple approach to define the applicability of the domain. It is based on the calculation of the leverage Hi[14] for each chemical, where the QSAR model is used to predict its activity. Leverages are obtained from the hat matrix (H), which is a n x n projection matrix specified as:

$$\mathbf{H = X (X'X)^{-1} X'}$$

where X is the matrix of x-values.

Full Paper



(a) (b)

Figure 4 : Electrostatic potential 2D contour diagrams and atomic charges calculated by HyperChem 8.0 (ZINDO/S) of (a) Hematin and (b) most active compound (11). The red color lines are for negative while green lines indicate positive electrostatic potential.

Leverages values fall between 0 and 1. A leverage value greater than 3p/n where p is the number of predictors plus the constant and n is the number of observations, is considered as large. It means that the predicted response is the result of a substantial extrapolation of the model and may be not reliable.

MLR equations

The following significant mono to tetra variate models were developed. These are as follows along with the interpretation of QSAR model in terms of the specific contribution of substituents and other molecular features to the modeled activity.

pIC50 = - 1.7154 + 0.0003562 W **(1)**

n = 17, S = 0.525, R = 0.864, R^2 = 0.747, R^2(adj) = 0.730, PRESS = 5.156, Q^2= 0.684, $R^2_{(LOO)}$ = 0.686

The small but positive coefficient of W indicates that bulkiness play small but crucial role in deciding activity and hence groups enhancing the bulkiness of molecules should be retained.

pIC50 = - 4.0904 + 0.030262 MR **(2)**

n = 17, S = 0.523397, R = 0.865, R^2= 0.749, R^2(adj)= 0.732, PRESS = 5.10681, Q^2= 0.687, $R^2_{(LOO)}$ = 0.690

The small but positive coefficient of MR indicates that bulkier and highly polar groups are more favorable.

pIC50 = - 2.3288 + 0.15768 logP + 0.0003071 W - 0.9638 I5 **(3)**

n = 17, S = 0.429034, R = 0.924, R^2= 0.854, R^2(adj) = 0.82, PRESS = 4.39438, Q^2= 0.731, $R^2_{(LOO)}$ = 0.756

pIC50 = - 4.328 + 0.08787 logP + 0.02876 MR - 0.851 I5 **(4)**

n = 17, S = 0.49204, R = 0.899, R^2= 0.807, R^2(adj) = 0.763, PRESS = 3.1474, Q^2= 0.643, $R^2_{(LOO)}$ = 0.680

The negative coefficient of I5 indicates that Cl at this position has negative effect on biological activity therefore for better activity it should be removed.

pIC50 = - 2.2343 + 0.0003006 W + 0.14072 logP - 1.1522 I2 **(5)**

n = 17, S = 0.441184, R = 0.916, R^2= 0.84, R^2(adj) = 0.80, PRESS = 2.53035, Q^2= 0.694, $R^2_{(LOO)}$ = 0.730

pIC50 = - 4.234 + 0.02845 MR + 0.07384 logP - 1.1061 I2 **(6)**

n = 17, S = 0.4954, R = 0.897, R^2 = 0.805, R^2(adj) = 0.760, PRESS = 3.1899, Q^2 = 0.636, $R^2_{(LOO)}$ = 0.681 The negative coefficient of I2 indicates that Cl at position 2 should be avoided to enhance the activity.

pIC50 = - 2.2343 + 0.0003006 W + 0.14072 logP - 1.1522 I7 (7)

n = 17, S = 0.441184, R= 0.919, R^2= 0.845, R^2(adj) = 0.809, PRESS = 2.5303, Q^2 = 0.694, $R^2_{(LOO)}$ = 0.730

pIC50 = - 4.234 + 0.02845 MR + 0.07384 logP - 1.1061 I7 (8)

n = 17, S = 0.4954, R = 0.897, R^2= 0.805, R^2(adj) = 0.760, PRESS = 3.1899, Q^2= 0.636, $R^2_{(LOO)}$ = 0.681 The negative coefficient of I7 indicates that presence of Cl is unfavorable at this position.

pIC50 = 0.711 + 0.0755 I4 + 0.00027684 W - 0.03896 ST - 0.05829 TVC (9)

n = 17, S = 0.518294, R = 0.896, R^2= 0.803, R^2 (adj) = 0.737, PRESS = 6.8248, Q^2= 0.582, $R^2_{(LOO)}$ = 0.654

pIC50 = -2.460 - 0.0854 I4 + 0.002694 MR - 0.0216 ST - 0.0392 TVC (10)

n = 17, S = 0.5583, R = 0.878, R^2= 0.771, R^2 (adj) = 0.695, PRESS = 3.7402, Q^2= 0.417, $R^2_{(LOO)}$ = 0.579 The positive coefficient of I4 in eq. 9 and negative in eq. 10 indicates that groups like Cl which has positive resonance but negative inductive effect should be retained at position 4 for good activity.

pIC50 = - 21.89 + 0.0468 LUMO + 0.0003172 W - 0.14943 ST + 17.49 IR (11)

n = 17, S = 0.488, R = 0.908, R^2 = 0.825, R^2(adj) = 0.767, PRESS = 26.662, Q^2= -0.632, $R^2_{(LOO)}$ = 0.503

pIC50 = - 26.66 - 0.2207 LUMO + 0.030193 MR - 0.13573 ST + 17.71 IR (12)

n = 17, S = 0.5144, R = 0.898, R^2= 0.806, R^2 (adj) = 0.741, PRESS = 32.227, Q^2= -0.972, $R^2_{(LOO)}$ = 0.477

The positive coefficient of LUMO in eq. 11 and negative in eq. 12 indicates that the ligand donate as well as accept electron density from receptor. This is consistent with finding of Riscoe that the ligand probably involves co-ordination between Fe of receptor with keto oxygen of xanthone nucleus in ligand and ionic interaction between protonatable nitrogen atoms with heme propionate groups. This proposition could be further justified by the quantum chemistry calculations studies when Electrostatic potential 2D contour diagrams (Figure 4) of hematin and most active compound **(11)** are evaluated.The presence of noticeable amount of negative charge on Fe in hematin supports the idea of possible back bonding between Fe and xanthone derivative.

pIC50 = - 2.2662 + 0.00030114 W + 0.15943 logP - 0.0172 Dipole - 0.9667 I5 (13)

n = 17, S = 0.4461, R = 0.924, R^2= 0.854, R^2 (adj) = 0.805, PRESS = 4.84265, Q^2= 0.704, $R^2_{(LOO)}$ = 0.729

pIC50 = - -4.050+ 0.02716 MR + 0.9717 logP -0.0451Dipole - 0.864 I5 (14)

n =17, S = 0.5097, R = 0.900, R^2= 0.809, R^2 (adj) = 0.746, PRESS = 3.1173, Q^2 = 0.613, $R^2_{(LOO)}$ = 0.658 The dipole plays a negative role towards activity.

pIC50 = - 2.233 + 0.0003093 W + 0.15771 logP - 0.02714 TVC - 0.9157 I5 (15)

n =17, S = 0.4405, R = 0.926, R^2 = 0.857, R^2 (adj) = 0.810, PRESS = 4.9632, Q^2 = 0.696, $R^2_{(LOO)}$= 0.731

pIC50 = - 4.267 + 0.08814 MR + 0.02883 logP - 0.0183 TVC - 0.818 I5 (16)

n = 17, S = 0.5097, R = 0.900, R^2= 0.809, R^2(adj) = 0.746, PRESS = 3.118, Q^2 = 0.576, $R^2_{(LOO)}$ = 0.648

pIC50 = - 2.478 + 0.05147 TC + 0.0003204 W - 1.1316 I5 + 0.1433 logP (17)

n = 17, S = 0.4232, R = 0.932, R^2= 0.868, R^2 (adj) =

TABLE 4 : Comparison between observed and calculated value (by eq. 20) of pIC_{50} and leverages

Sr. no.	Obs. Value	Calc.Value by MLR 20	Calc. Value byMLR20 (LOO)	Calc. Value byMLR20 (L5O)	Leverage (limit 0.99)
1	-1.301	-1.1208	-0.8716	-1.5917	0.580293
2	-1.699	-1.7409	-1.9673	-2.5303	0.843588
3	-1.497	-1.2559	-0.4334	-0.2334	0.773402
4	0.921	0.4220	0.3801	0.3249	0.077430
5	0.260	0.6892	0.7792	0.8285	0.173143
6	-0.663	-0.7866	-0.8459	-0.9149	0.323866
7	0.032	0.3489	0.3769	0.4246	0.081214
8	0.143	0.2825	0.2958	0.2632	0.086828
9	0.018	-0.1519	-0.2979	-0.2228	0.462852
10	0.678	1.1653	1.2890	1.3670	0.202443
11	1.398	1.4822	1.5329	1.7272	0.376053
12	1.398	0.9659	0.8378	0.9153	0.228582
13	1.000	0.7401	0.7134	0.7762	0.093067
14	1.301	0.7535	0.6858	0.6434	0.109905
15	0.824	0.7882	0.7826	0.7758	0.135513
16	0.569	0.9223	1.0228	1.0352	0.221256
17	-0.708	-0.8319	-0.8692	-0.8754	0.230564

Full Paper

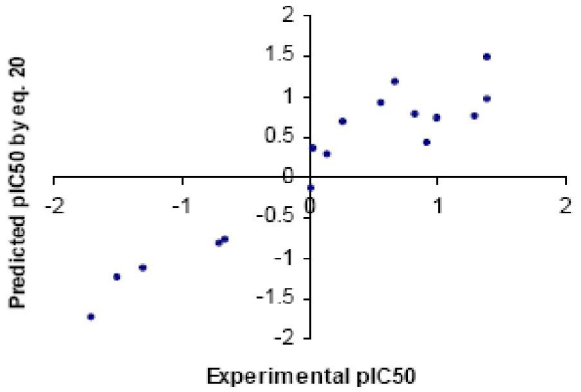


Figure 5 : Comparison between observed and calculated value (by eq. 20) of pIC50

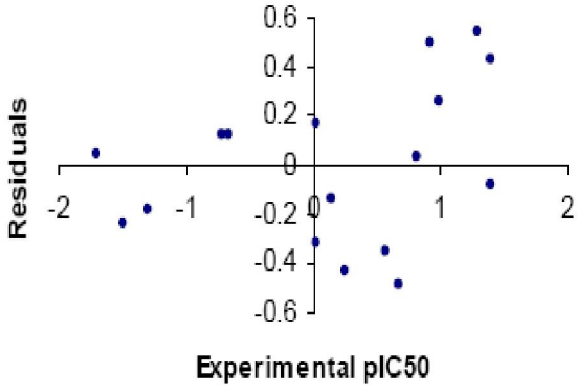


Figure 6 : Plot of residual (equation 20) vs experimental values

0.825, PRESS = 5.3105, Q^2= 0.621, $R^2_{(LOO)}$ = 0.679

pIC50 = -4.501 +0.02962 MR + 0.03969TC - 0.9765 I5 + 0.07533 logP (18)

n = 17, S = 0.50006, R = 0.903, R^2= 0.816, R^2(adj) = 0.755, PRESS = 7.5084, Q^2= 0.622, $R^2_{(LOO)}$ = 0.635

The coefficient of logP is small but positive suggesting either lack of hydrophobic interaction between the ligand and receptor or very little hydrophobic interactions.

pIC50 = - 41.64 + 0.03314 MR - 0.3425 HOMO + 27.87 IR - 0.21503 ST (19)

n = 17, S = 0.431872, R = 0.929, R^2= 0.863, R^2(adj) = 0.817, PRESS = 6.1572, Q^2= 0.623, $R^2_{(LOO)}$ = 0.666

pIC50 = - 39.1563 + 0.0003617 W + 28.4963 IR - 0.23824 ST - 0.3791 HOMO (20)

n = 17, S = 0.36864, R = 0.948, R^2= 0.900, R^2(adj) = 0.867, PRESS = 3.6259, Q^2 = 0.778, $R^2_{(LOO)}$ = 0.782

The large positive coefficient of IR indicates that presence of aliphatic carbon chain and cyclisation has positive impact on activity. Replacing MR with W changed the statistic significantly this shows that W predicts the activity better than MR. The positive coefficient of IR indicates that cyclisation play positive role in determining activity. In all the equations ST, which is considered as inverse steric factor has negative coefficient this indicates that it, play a negative role in determining the biological activity. The negative coefficient of HOMO indicates that the nucleophilicity of ligand has negative effect on biological activity and hence groups which increase nucleophilicity should not be incorporated.

In all the above models, n is number of compounds in data set, R is the correlation coefficient, R^2 is the coefficient of determination, R^2_{ad} is adjusted coefficient of determination, SE is the standard error of estimate, F is the variance ratio, Q is Pogliani test, All those equations resulting in low value of R (< 0.50) were not considered being statistically significant. The high values of R, R^2, F, Q and low value of SE indicates that models have excellent statistical significance. Moreover the values of $R^2_{adj,}$ which is considered as better parameter to judge the predictive power compared to R^2, are close to the values of R^2 thereby confirming high predictive power of models. The values of $R^2_{(LOO)}$ again support our findings. Comparison of values predicted by equation 20 with experimental -logIC_{50} and residual are listed in TABLE 4. In order to investigate the possible existence of outliers, the extent of the extrapolation method was applied to the 17 compounds that constitute the entire dataset. The leverages for all the 17 compounds were computed and no compound lies outside the domain (TABLE 4).

A graph between observed and calculated (by equation 20) values is as shown in figure 5 which clearly indicate a good linear relationship. Figure 6 is a plot of residual (equation 20) vs experimental values. The extensive analysis clearly indicates that the quality of correlation increases as we pass from univariate to tetravariate correlations. Also, the results indicate that multiple correlations give better estimates than the univariate correlations and that the multivariate correlations wherein W is involved are better than those correlations where MR is involved.

Validation

Deriving 4-parametric equations from 17 molecules may be done by chance. Therefore, in order to prove that the models are not *chancy* we have calculated R^2_{pred}, $R^2_{(LOO)}$, and PSE also.

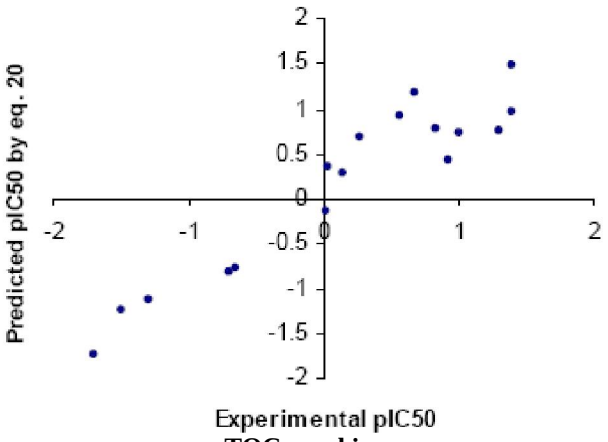
TABLE 5 : Statistical data of various models derived along with predictive square error (PSE)

Model no.	N	R	1-R^2	PRESS	PSE	R^2_{pred}
1	17	0.864	0.253	5.15600	0.303	0.684
2	17	0.865	0.252	5.10681	0.300	0.687
3	17	0.924	0.146	4.39438	0.258	0.731
4	17	0.899	0.192	3.14740	0.185	0.643
5	17	0.916	0.161	2.53035	0.149	--
6	17	0.897	0.195	3.1899	0.187	--
7	17	0.919	0.155	2.53035	0.149	--
8	17	0.897	0.195	3.1899	0.187	--
9	17	0.896	0.197	6.82481	0.401	0.583
10	17	0.878	0.229	3.7402	0.220	0.417
11	17	0.908	0.175	26.6620	1.568	0.000
12	17	0.898	0.194	32.2273	1.896	0.000
13	17	0.924	0.146	4.84265	0.285	0.704
14	17	0.900	0.190	3.1173	0.183	0.613
15	17	0.926	0.142	4.96318	0.292	0.696
16	17	0.900	0.190	3.1183	0.183	0.576
17	17	0.932	0.131	5.31056	0.312	0.675
18	17	0.903	0.185	7.50842	0.441	0.541
19	17	0.929	0.137	6.15719	0.362	0.623
20	17	0.948	0.101	3.62587	0.213	0.778

Predictive correlation coefficient (R^2_{pred})

The predictive capability of the 2D-QSAR models was determined by cross validation. The predictive correlation (R^2_{pred}), based on the test set molecules, is computed using: R^2_{pred} = (SD-PRESS)/SD Where SD is the sum of squared deviations between biological activities of the test set and mean activities of the training set molecules and the predictive residual sum of squares (PRESS) is the sum of squared deviations between calculated and experimental activity values for every molecule. The R^2_{pred}, PRESS, PSE and 1-R^2 values for all the models are listed in TABLE 5.

To have better cross validation we further used "Leave One Out (LOO)" method and obtained $R^2_{(LOO)}$ for all the models. To be a reasonable QSAR model, the proposed equations must be useful to end-users. For practical purposes of end-users the use of square root of PRESS / N, which is called predictive square error (PSE), is more directly related to the uncertainty of the predictions[15]. The PSE values also support our results (TABLE 5).



TOC graphics

For a more exhaustive testing of the predictive power of the model number 20, except for the classical LOO cross validation technique, the validation of the model was carried out by a leave five out cross-(L5O) validation procedure. The results are $R^2_{(LOO)}$ = 0.782 and $R^2_{(L5O)}$ = 0.753 . It is important that the model is quite stable to the inclusion – exclusion of compounds as measured by values of LOO and L5O correlation coefficients. The results of predictions on the cross-validation test illustrated the quality of the obtained model.

Considering the fact that the 2D-QSAR models were able to reproduce the experimental values and that are validated by appropriate statistical procedures, they could be useful in designing more potent anti-malarial xanthone derivatives.

RESULTS AND DISCUSSIONS

We have used a good number of well established[16-18] QSAR descriptors. An important step for model building was to define the number of independent variables in the main QSAR equation. On MLR method, we built consecutively several equations with different number of descriptors upto six variables and tested objective selection method as well as we plot the graph between R^2 adj. and number of descriptors to get optimum number of parameters essential for best fit model. Since the value of the adjusted R^2 value will decrease if the added variable does not reduce the unexplained variation enough to affect the loss of degrees of freedom, this means if a variable is added that does not contribute its fair share, then the adjusted R^2 value will actually decline and hence the procedure for selection

of optimum number of descriptors is based on the break point rule (the change in the slope) showing the critical improvement of the adjusted R^2 value over the number of the descriptors in each model. This step ensures the over parameterization of the model and prevents to some extent the chance correlations between the descriptors.

From various QSAR models developed here it is clear that even though W and MR have strong correlation with each other and with pIC_{50} but W has better predictive ability than MR Thus Objective feature selection alone is not sufficient to select the best descriptors for model generation.

Replacing logP with HOMO increased the predictive ability; this suggests that logP has little influence on activity. A combination of a set of parameters serves as a better QSAR model predictor than any single parameter. Moreover, the equation 20 indicates that a combination of topological and physico-chemical parameters is best suitable for QSAR model creation.

It should be noted that we have utilized only the MLR method for constructing the QSAR models. For such a relatively low number of molecules, the use of nonlinear models such as artificial neural network (ANN) may produce better predictability for the suggested QSAR. The chemical interpretation of ANN models is difficult compared to MLR analysis and requires high level computer skills. Work is under progress to obtain QSAR models by ANN also. Meanwhile, the predictivity of the proposed QSAR models by MLR analysis was not low. MLR equations with R^2adj greater than 0.75 can be considered as good predictive models. Therefore, in this work we did not try to obtain extra predictive models by nonlinear methods such as ANN. To add further, together with the LOO validation, R^2_{adj} and R^2_{pred} has unambiguously demonstrated the robustness of the QSAR models and their power of predicting external data with accuracy.

In summary, an extensive QSAR analysis has been performed on a wide variety of a structurally diverse set of xanthone derivatives and present investigations revealed several important physicochemical and structural requirements for anti-malarial activity. The results derived for these completely new set of ligands could be beneficial in the hands of medicinal chemists to further the design and development of more and better anti-malarial agents in the future.

ACKNOWLEDGMENT

We are thankful to Mr. MeghShyam Patil, IICT, Hyderabad, India, Mr. Mangesh Tihile, National Chemical laboratory, Pune, India and Naseer Baig, Indian Institute of Science, Banglore, India for providing data, substantial computational work and helpful discussions.

REFERENCES

[1] World Malaria Report 2008, WHO, 141 **(2008)**.
[2] M.Riscoe, J.X.Kelly, R.Winter; Curr.Med.Chem., **12**, 2539 **(2005)**.
[3] M.M.M.Pinto, M.E.Sousa, M.S.J.Nascimento; Curr.Med.Chem., **12**, 2517 **(2005)**.
[4] A.Thakur, S.Baghel, K.Tiwari, M.Thakur, K.S.Chandel, G.L.Badole, A.Thakkar; Res.J.Chem. Environ., **12(3)**,68 **(2008)**.
[5] M.Riscoe, J.X.Kelly, R.Winter; Curr.Med.Chem., **12**, 2539 **(2005)**.
[6] V.K.Agrawal, V.K.Dubey, B.Shaik, J.Singh, K.Singh, P.V.Khadikar; J.Indian Chem.Soc., **86**, 337 **(2009)**.
[7] Advanced Chemistry Development, Inc., 110 Yonge Street 14th Floor, Toronto, Ontario, M5C 1T4, Canada (www.acdlabs.com).
[8] **(a)** Cambridge Soft Corporation, 100 Cambridge Park Drive, Cambridge, MA, 02140 USA; **(b)** Hypercube Inc. www.Hyper.com.
[9] A.K.Srivastava, A.Srivastava, A.Jaiswal, M.Jaiswal; J.Indian Chem.Soc., **85**, 721 **(2008)**.
[10] K.Kraim, D.Khatmi, Y.Saihi, F.Ferkous, M.Brahimi; Chemometrics Intelli.Lab.Syst., **97**, 118 **(2009)**.
[11] A.Tropsha, P.Gramatica, V.K.Gombar; QSAR Comb.Sci., **22**, 1 **(2003)**.
[12] A.Golbraikh, A.Tropsha; J.Mol.Graph.Model., **20**, 269 **(2002)**.
[13] M.Shen, C.Beguin, A.Golbraikh, J.Stables, H.Kohn, A.Tropsha; J.Med.Chem., **47**, 2356 **(2004)**.
[14] A.C.Atkinson; 'Plots; Transformations and Regression', Clarendon Press, Oxford (UK), **(1985)**.
[15] A.K.Srivastava, A.Srivastava, A.Jaiswal, M.Jaiswal; J.Indian Chem.Soc., **85**, 721 **(2008)**.
[16] Qian-Nan Hu, Yi-Zeng Liang, Kai-Tai Fang; J.Data Sci., **1**, 361 **(2003)**.
[17] J.Devillers, A.T.Balaban; 'Topological indices and related descriptos in QSAR and QSPR', Gordon and Breach Science Publishers, The Netherlands, **(1999)**.
[18] O.Ivanciuc, J.Gasteiger; Ed., 'Handbook of Chemoinformatics', Wiley–VCH, **(2003)**.