



Scholars Research Library

Der Pharma Chemica, 2010, 2(5): 22-32
(<http://derpharmachemica.com/archive.html>)



ISSN 0975-413X
CODEN (USA): PCHHAX

3D-QSAR studies on xanthone derivatives to understand pharmacological activities as MAO inhibitors

Vijay H. Masand*¹, Komalsingh N. Patil¹, Devidas T. Mahajan¹, Rahul D. Jawarkar² and Ghulam M. Nazerruddin³

¹Department of Chemistry, VidyaBharati College, Amravati, Maharashtra, India

²Department of Pharmaceutical Chemistry, Vidyabharati College of Pharmacy, Amravati, Maharashtra, India

³Department of Chemistry, Poona College, Pune, Maharashtra, India

ABSTRACT

Monoamine Oxidase (MAO) play vital role in regulation of monoamine neurotransmitters such as serotonin, dopamine and nor-adrenaline. MAO inhibitors are used in the treatment of Parkinson's disease, Alzheimer's disease, depression etc. In present work we have carried out SAR analysis and developed many QSAR equations to understand the pharmacological activities of Xanthone derivatives as MAO inhibitors. The SAR and QSAR analysis provides interesting insights in understanding the hydrophobic, steric, electronic, and structural requirements for MAO inhibitory activity. Quantum mechanical studies support the SAR analysis. Combination of different types of 3-D descriptors like WHIM, GATEWAY and so on provide useful QSAR models. The QSAR models were tested for their statistical significance by using Y-randomization, R^2 , R^2_{adj} and R^2_{LOO} methods. The significant model is with $R^2 > 91\%$. These QSAR studies help us in the design and prediction of novel Xanthone derivatives as MAO inhibitors.

Keywords: QSAR, Xanthone Derivatives, MAO inhibitors, 3D Descriptors.

INTRODUCTION

Monoamine Oxidase (MAO) regulates monoamine neurotransmitters such as serotonin, dopamine and nor-adrenaline by oxidative deamination [1]. Serotonin and norepinephrine, play important role in controlling mood. But other substances in the brain may interfere with mood control by breaking down these neurotransmitters. MAO inhibitors work by blocking the chemicals that break down serotonin and norepinephrine [2]. The inhibition of this enzyme allows these neurotransmitters to remain active in the brain longer, which in turn causes a stimulation effect, thereby correcting a presumed deficit in monoamine function.

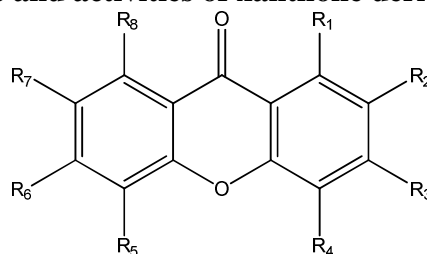
Therefore MAO inhibitors are a type of antidepressant that helps to reduce the extreme sadness, hopelessness, and lack of interest in life that are typical in people with depression. MAO inhibitors are especially useful in treating people whose depression is combined with other problems such as anxiety, panic attacks, phobias, or the desire to sleep too much. Unfortunately the available drugs have serious side effects like ‘Cheese effect’ especially when taken with special foods, beverages and medicines [2]. To ameliorate the situation new MAO inhibitors like xanthone derivatives are under development but the search for potent xanthone derivative as MAO inhibitor is still continue. Previously Clare et al [3] used Nodal descriptors and Nunez applied 2D descriptors to get useful QSAR models to predict the MAO inhibitory activity of xanthone derivatives. But these attempts gave models with large number of descriptors as well as statistically low significant equations. Here we have applied the well-established 3D descriptors to understand the role of various 3D structural features responsible for MAO inhibitory activity of xanthenes.

The objectives of this work are (1) to determine the best variables that afford the most significant linear QSAR models correlating the structure of these compounds with their MAO inhibitory activity. (2) To understand the various characters responsible for MAO inhibitory activity especially for xanthenes. (3) To develop QSAR equation with no problem of “Over Fitting” with optimum number of descriptors. The relevance of the models for the design of novel derivatives should be assessed not only in terms of predictivity, but also in terms of their ability to provide a chemical and structural explanation also. These results should serve as a guideline in designing more potent and selective MAO inhibitor.

MATERIALS AND METHODS

Data Set: The MAO inhibitory activity data of xanthone derivatives were taken from the reported work of Clare et al [3] (Table 1). The compounds include structurally diverse xanthone derivatives mostly positional isomers with substituents like -OMe, -OH etc. Chem Sketch software (ACD labs 12.0 freeware) was used to draw 2D and 3D structures of the molecules. The 3D optimized structures were used to calculate the descriptors. We tested all 3D descriptors like MoRSE, RDF, WHIM and GATEAWAY etc. available in Dragon 3.0. Minitab 14.0 and MolGro Data Modeller were used to perform various regression analyses. Since the calculations of these descriptors are well documented in the literature, it is not necessary to duplicate the same here. We performed Y- Randomization, “Leave One Out”, “Leave Many Out” and calculated R^2 , PSE, PRESS as well as adjustable R^2 to determine robustness of the models.

Table 1. Structure and activities of xanthone derivatives used in QSAR.



No.	R1	R2	R3	R4	R5	R6	R7	R8	IC ₅₀ MAO-A (μ M)	Log IC ₅₀ (Obs)
1	H	H	H	H	H	H	H	H	0.84	-0.07572

2	OH	H	H	H	H	□	H	H	0.31	-0.50864
3	MeO	H	H	H	H	H	H	H	0.90	-0.04576
4	H	OH	H	H	H	H	H	H	3.80	0.579784
5	H	MeO	H	H	H	H	H	H	5.30	0.724276
6	H	H	OH	H	H	H	H	H	1.10	0.041393
7	H	H	MeO	H	H	H	H	H	0.18	-0.74473
8	H	H	H	OH	H	H	H	H	1.30	0.113943
9	H	H	H	MeO	H	H	H	H	30.00	1.477121
10	OH	H	H	H	OH	H	H	H	0.73	-0.13668
11	H	H	OH	H	OH	H	H	H	4.50	0.653213
12	H	H	OH	H	MeO	H	H	H	23.00	1.361728
13	OH	H	MeO	H	H	H	H	H	0.11	-0.95861
14	MeO	H	MeO	H	H	H	H	H	20.20	1.305351
15	H	H	MeO	H	MeO	H	H	H	36.00	1.556303
16	MeO	H	H	H	OH	H	H	H	51.00	1.707570
17	H	H	MeO	OH	H	H	H	H	18.00	1.255273
18	H	H	OH	MeO	H	H	H	H	65.00	1.812913
19	H	H	MeO	MeO	H	H	H	H	31.00	1.491362
20	OH	H	OH	H	OH	H	H	H	3.80	0.579784
21	OH	H	MeO	H	OH	H	H	H	0.04	-1.39794
22	OH	H	MeO	H	MeO	H	H	H	29.00	1.462398
23	MeO	H	MeO	H	MeO	H	H	H	58.00	1.763428
24	OH	H	OH	Me	H	H	H	H	4.30	0.633468
25	OH	Me	OH	H	H	H	H	H	3.70	0.568202
26	OH	Me	OH	Cl	H	H	H	H	27.00	1.431364
27	OH	Me	OH	Br	H	H	H	H	14.90	1.173186
28 ^a	OH	H	OH	C ₁₀ H ₁₇	OH	H	H	H	37.00	1.568202
29 ^b	OH	C ₅ H ₉	H	OH	OH	H	H	H	3.30	0.518514
30 ^c	OH	H	C ₅ H ₉	OH	OH	H	H	H	40.00	1.60206
31	OH	MeO	OH	H	OH	H	H	H	2.70	0.431364
32	OH	MeO	OH	H	MeO	H	H	H	51.00	1.70757
33	MeO	MeO	MeO	H	MeO	H	H	H	37.00	1.568202
34	OH	H	OH	H	H	H	OH	H	8.00	0.90309
35	OH	H	OH	H	OH	H	H	OH	13.00	1.113943
36	OH	H	MeO	H	OH	H	H	OH	0.66	-0.18046
37	OH	H	OH	H	H	H	OH	OH	24.00	1.380211
38	OH	H	MeO	H	H	H	OH	OH	8.50	0.929419
39	OH	H	MeO	H	H	H	MeO	MeO	19.00	1.278754
40	OH	H	OH	H	H	OH	OH	H	25.00	1.39794
41	MeO	H	H	Me	OH	H	MeO	H	24.00	1.380211
42	OH	MeO	OH	H	MeO	OH	H	H	32.00	1.50515

^a C₁₀H₁₇ is Me₂C = CH-CH₂-CH₂-C(Me) = CH-CH₂; ^b C₅H₉ is CH₂ = CH-CMe₂; ^c C₅H₉ is Me₂ = CH-CH₂.

Experimental protocol / Computational approach

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 involves [4]:

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 0.95 ($R > 0.95$) and has a large correlation coefficient with the other descriptors in each class was eliminated.

Optimum Number of Descriptors to be used:

A major decision in developing successive QSAR model is when to stop adding descriptors to the model. A simple technique to control the model expansion is the so-called “breaking point” in the improvement of the statistical quality of the model [4, 5], 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 1. The reason behind using adjusted R^2 is that it depends upon number of descriptors used and its values decreases as the number of unnecessary descriptors increases. The figure indicates that the optimum number of descriptors is to be used is ten. Therefore QSAR models with descriptors more than ten may give rise to “highly fit” results due to “Over fitting”. Statistically over fitting is a demerit and should be avoided and hence equation with more than ten descriptors were not considered.

In multiple regression analysis, the independent variables must be orthogonal. Consequently the autocorrelation among the descriptors was checked and is given in the correlation matrix.

Table 2 . LogIC₅₀ values and various descriptors for all the molecules

Cpd.	logIC ₅₀	I _{1-OH}	I _{2-OH}	I _{3-OH}	I _{4-OH}	I _{5-OH}	I _{6-OH}	I _{7-OH}	I _{8-OH}
1	-0.07572	0	0	0	0	0	0	0	0
2	-0.50864	1	0	0	0	0	0	0	0
3	-0.04576	0	0	0	0	0	0	0	0
4	0.579784	0	1	0	0	0	0	0	0
5	0.724276	0	0	0	0	0	0	0	0
6	0.041393	0	0	1	0	0	0	0	0
7	-0.74473	0	0	0	0	0	0	0	0
8	0.113943	0	0	0	1	0	0	0	0
9	1.477121	0	0	0	0	0	0	0	0
10	-0.13668	1	0	0	0	1	0	0	0
11	0.653213	0	0	1	0	1	0	0	0
12	1.361728	0	0	1	0	0	0	0	0
13	-0.95861	1	0	0	0	0	0	0	0
14	1.305351	0	0	0	0	0	0	0	0
15	1.556303	0	0	0	0	0	0	0	0
16	1.707570	0	0	0	0	1	0	0	0
17	1.255273	0	0	0	1	0	0	0	0
18	1.812913	0	0	1	0	0	0	0	0
19	1.491362	0	0	0	0	0	0	0	0
20	0.579784	1	0	1	0	1	0	0	0

21	-1.39794	1	0	0	0	1	0	0	0
22	1.462398	1	0	0	0	0	0	0	0
23	1.763428	0	0	0	0	0	0	0	0
24	0.633468	1	0	1	0	0	0	0	0
25	0.568202	1	0	1	0	0	0	0	0
26	1.431364	1	0	1	0	0	0	0	0
27	1.173186	1	0	1	0	0	0	0	0
28	1.568202	1	0	1	0	1	0	0	0
29	0.518514	1	0	0	1	1	0	0	0
30	1.60206	1	0	0	1	1	0	0	0
31	0.431364	1	0	1	0	1	0	0	0
32	1.70757	1	0	1	0	0	0	0	0
33	1.568202	0	0	0	0	0	0	0	0
34	0.90309	1	0	1	0	0	0	1	0
35	1.113943	1	0	1	0	1	0	0	1
36	-0.18046	1	0	0	0	1	0	0	1
37	1.380211	1	0	1	0	0	0	1	1
38	0.929419	1	0	0	0	0	0	1	1
39	1.278754	1	0	0	0	0	0	0	0
40	1.39794	1	0	1	0	0	1	1	0
41	1.380211	0	0	0	0	1	0	0	0
42	1.50515	1	0	1	0	0	1	0	0

^a $I_{n-OH} = 1$ if $-OH$ at that position otherwise 0.

Table 3 . Correlation matrix for LogIC₅₀ values and various descriptors

	logIC50	RDF085m	R5v	R4m+	Mor14e	G2u	Mor20u	G2e
RDF085m	0.362							
R5v	0.543	0.155						
R4m+	-0.275	-0.296	0.310					
Mor14e	-0.184	0.032	-0.062	-0.233				
G2u	-0.384	-0.573	-0.046	0.192	-0.207			
Mor20u	0.440	0.652	0.323	-0.343	-0.286	-0.436		
G2e	-0.216	-0.565	-0.053	0.165	-0.365	0.577	-0.427	
RDF095m	0.283	0.825	0.283	-0.170	0.198	-0.571	0.455	-0.564
Mor16u	-0.264	-0.464	0.081	0.324	0.097	0.287	-0.521	0.324
RDF030v	0.260	0.598	0.359	-0.238	0.520	-0.596	0.534	-0.743
DP01	0.450	0.881	0.259	-0.275	0.046	-0.663	0.784	-0.735
E2u	-0.085	-0.604	-0.110	0.140	-0.228	0.504	-0.483	0.623
Mor02m	0.373	0.882	0.140	-0.379	-0.100	-0.529	0.784	-0.497
E2e	-0.221	-0.661	-0.100	0.404	-0.208	0.420	-0.589	0.528
Mor08u	0.289	0.602	0.200	-0.433	0.387	-0.541	0.607	-0.636
I7-OH	-0.127	0.086	-0.272	-0.324	0.098	-0.009	-0.051	0.225
RDF020e	0.408	0.509	0.129	-0.420	-0.065	-0.450	0.715	-0.418
RDF050m	0.352	0.647	0.211	-0.168	-0.101	-0.403	0.722	-0.541
H8m	0.276	0.742	0.194	-0.306	0.063	-0.474	0.716	-0.456
Mor16m	-0.019	-0.337	-0.036	-0.013	0.028	0.014	-0.235	0.179
RDF055u	0.524	0.663	0.382	-0.390	0.320	-0.555	0.665	-0.743

	RDF095m	Mor16u	RDF030v	DP01	E2u	Mor02m	E2e	Mor08u
Mor16u	-0.341							
RDF030v	0.624	-0.339						
DP01	0.740	-0.498	0.749					
E2u	-0.531	0.278	-0.704	-0.668				
Mor02m	0.727	-0.440	0.491	0.867	-0.391			
E2e	-0.495	0.349	-0.643	-0.638	0.879	-0.496		
Mor08u	0.522	-0.372	0.772	0.759	-0.413	0.686	-0.500	
I7-OH	0.072	-0.165	-0.141	-0.103	0.055	-0.005	-0.075	-0.155
RDF020e	0.405	-0.510	0.419	0.690	-0.136	0.757	-0.262	0.790
RDF050m	0.347	-0.470	0.555	0.759	-0.717	0.535	-0.661	0.405
H8m	0.558	-0.568	0.632	0.836	-0.588	0.746	-0.593	0.697
Mor16m	-0.138	0.348	-0.261	-0.249	0.542	-0.091	0.508	0.125
RDF055u	0.607	-0.357	0.810	0.830	-0.590	0.674	-0.670	0.819
		RDF020e	RDF050m	H8m	Mor16m			
RDF020e	-0.076							
RDF050m	-0.079	0.311						
H8m	0.007	0.643	0.706					
Mor16m	-0.042	0.312	-0.650	-0.333				
RDF055u	-0.128	0.651	0.574	0.724	-0.140			

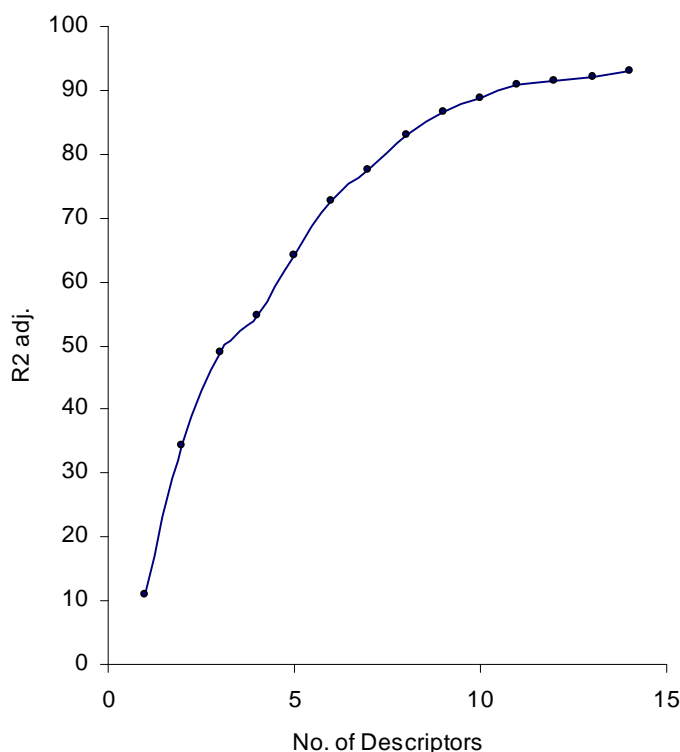


Fig. 1. Number of Descriptors Vs. R² adj.

In order to develop 3D QSAR model, the data set was subjected to a stepwise multiple linear regression analysis. This resulted into several correlation equations between the logIC₅₀ values as a dependent variable and several quantifying parameters as an independent variable. Equation 1 involves use of ten variables only, a low number of variables compared to previously reported

work which involved more than ten variables. There it can be considered as statistically significant model for MAO inhibitory activity. This model shows a better correlation coefficient ($R\text{-Sq} = 91.548\%$) with low standard error of estimation. Eqn. 1 accounts for more than 91% variance in the biological activity.

Cross-validation technique:

Deriving 10-parametric equations from 42 molecules may be done by *chance*. Therefore, in order to prove that the models are not *chancy* we have calculated R^2_{pred} and $R^2_{(\text{LOO})}$ also.

Predictive correlation coefficient (R^2_{pred}): The predictive correlation [25] (R^2_{pred}), based on the test set molecules, is computed using: $R^2_{\text{pred}} = (\text{SD-PRESS})/\text{SD}$ 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.

Since a high-correlation coefficient R^2 merely points out how well the equations fit the data, hence to check the reliability of the proposed models we used cross-validation procedure. The well-known “leave-one-out” (LOO) approach is very useful for this purpose, in which a number of models were developed with one sample ignored each time, then the ignored data were predicted by each model and the differences between predicted and observed activity values were evaluated. The cross validation was performed by using Molgro Data Modeler 2009. In general, LOO cross-validated coefficient R^2 being higher than 0.5 can be considered as a statistical proof of the high-predictive ability [5].

For a more thorough testing of the predictive power of the models, except for the classical LOO cross validation technique, the validation of the model was carried out by “Leave Many Out” cross-(LMO) validation procedure also. The results are $R\text{-Sq}(\text{LOO}) = 85.307\%$ and $R\text{-Sq}(\text{L50}) = 83.6953\%$ $R\text{-Sq}(\text{L100}) = 84.4691\%$. It is important that the model is quite stable to the inclusion – exclusion of compounds as measured by values of LOO, L50 and L100 correlation coefficients. The results of predictions on the cross-validation test illustrated the quality of the obtained model (see table 4).

Table 4 . Actual and Predicted LogIC_{50} values

Actual LogIC_{50}	Predicted $-\text{LOGIC}_{50}$ MLR-1	Residual	Predicted $-\text{LOGIC}_{50}$ MLR-2	Residual	Predicted $-\text{LOGIC}_{50}$ MLR-3	Residual
-0.07572	-0.2515	0.1758	-0.4282	0.353	-0.3317	0.256
-0.13668	0.1110	-0.247227	0.1871	-0.324	0.1751	-0.312
0.65321	0.8130	-0.1598	0.7428	-0.090	0.8501	-0.197
1.36173	1.3348	0.0269	1.4458	-0.084	1.3479	0.014
-0.95861	-0.5366	-0.4220	-0.6160	-0.343	-0.4163	-0.542
1.30535	1.5704	-0.2650	1.1310	0.174	1.4047	-0.099
1.55630	1.2553	0.3010	1.7605	-0.204	1.4096	0.147
1.70757	1.4461	0.2615	1.7714	-0.064	1.7191	-0.012
1.25527	0.9581	0.2972	0.9021	0.353	1.0788	0.177
1.81291	1.7780	0.0349	1.8745	-0.062	1.3027	0.510
1.49136	1.4394	0.0519	1.5492	-0.058	1.4542	0.037
-0.50864	-0.5266	0.0179	-0.6291	0.121	-0.3940	-0.115
0.57978	0.4408	0.1390	0.4219	0.158	0.1525	0.427

-1.39794	-1.4873	0.0894	-1.3746	-0.023	-1.3889	-0.009
1.46240	1.3430	0.1194	1.2279	0.235	1.2206	0.242
1.76343	1.7028	0.0606	1.6374	0.126	1.7790	-0.016
0.63347	0.6732	-0.0397	0.6947	-0.061	0.9921	-0.359
0.56820	0.2371	0.3311	0.3755	0.193	0.3592	0.209
1.43136	1.6499	-0.2185	1.2157	0.216	1.5106	-0.079
1.17319	1.1517	0.0215	1.1851	-0.012	1.2924	-0.119
1.56820	1.6553	-0.0870	1.3127	0.256	1.6724	-0.104
0.51851	0.3164	0.2022	0.6265	-0.108	0.6777	-0.159
-0.04576	0.0865	-0.1322	0.4847	-0.530	-0.0207	-0.025
1.60206	1.3601	0.2420	1.6799	-0.078	1.5111	0.091
0.43136	0.3955	0.0358	0.4039	0.028	0.5255	-0.094
1.70757	1.4887	0.2189	1.7413	-0.034	1.9194	-0.212
1.56820	1.6701	-0.1019	1.3631	0.205	1.6107	-0.043
0.90309	0.4767	0.4264	0.6108	0.292	0.3861	0.517
1.11394	1.5301	-0.4162	1.2385	-0.125	1.1699	-0.056
-0.18046	0.0854	-0.2659	0.2861	-0.467	-0.1019	-0.079
1.38021	1.4564	-0.0762	1.3785	0.002	1.3312	0.049
0.92942	0.6692	0.2602	0.7301	0.199	0.6812	0.248
1.27875	1.5437	-0.2649	1.2639	0.015	1.2613	0.018
0.57978	0.7869	-0.2072	0.5040	0.076	0.6099	-0.030
1.39794	1.4095	-0.0116	1.6478	-0.250	1.8631	-0.465
1.38021	1.5591	-0.1789	1.6456	-0.265	1.2976	0.083
1.50515	1.4608	0.0444	1.7353	-0.230	1.3482	0.157
0.72428	0.4119	0.3124	0.1683	0.556	0.4290	0.295
0.04139	-0.2486	0.2900	0.2554	-0.214	-0.3168	0.358
-0.74473	-0.1005	-0.6442	-0.3430	-0.402	-0.3099	-0.435
0.11394	0.3587	-0.2447	0.0229	0.091	0.4981	-0.384
1.47712	1.4538	0.0233	1.0974	0.380	1.3673	0.110

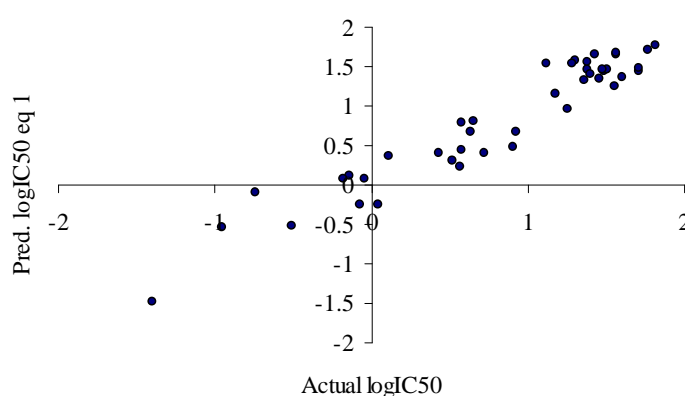


Fig. 2. Graph between actual Vs. predicted logIC₅₀

Y-Randomization test:

Y-Randomization is a widely used approach to establish the robustness of a given QSAR model. In this approach, dependent variable vector (inhibitory activity in this study) is randomly shuffled and a new QSAR model is built using the original independent variables. If the new

QSAR models have lower R^2_{LOO} values for several trials, then the given QSAR model is thought to be robust. Thus Y-randomization is useful to avoid any chancy correlation between dependent variable vector and independent variables. None of the model showed high R^2_{LOO} even after many Y-randomizations.

Table 5 . Results of Y-randomization

Sr. No.	For eq. 1 value of R^2 After Y-randomization
s	0.174
2	0.219
3	0.344
4	0.295
5	0.268
6	0.343
7	0.374
8	0.272
9	0.217
10	0.217

RESULTS AND DISCUSSION

Previous QSAR analysis:

Clare et al derived MLR equation using Nodal orientations for same set of molecules which gave very good statistics for $n=40$, $R^2=0.879$, $Q^2=0.7178$ with $S=0.364$. The equation is based on fifteen descriptors. It is a well known fact that use of large number of descriptors (in analysis of Clare fifteen are used) may lead to "Over Fitting" which is certainly a demerit [5]. Moreover, the work by Clare et al lacks extensive cross validation like calculation of R^2_{adj} . (to check the optimum number of descriptors) and Leave Many Out (for cross validation). To add further, a high value of S and a significant difference between the value of Q^2 and R^2 indicates that the equation most probably suffers from Over fitting. Hence it will be very useful to derive new equations with new descriptors.

MLR equations:

The following significant multi-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.

$$\text{LogIC}_{50} = 17.2936 + 0.1729 \text{ RDF085m} + 26.1223 \text{ R5v} - 20.5278 \text{ R4m+} - 1.3058 \text{ Mor14e} - 55.0821 \text{ G2u} - 2.6115 \text{ Mor20u} - 61.1559 \text{ G2e} - 0.5335 \text{ RDF095m} - 1.3455 \text{ Mor16u} - 0.2298 \text{ RDF030v} \quad (1)$$

$$n = 42 \quad S = 0.273019 \quad R\text{-Sq} = 91.548\% \quad R\text{-Sq(adj)} = 88.8\% \quad \text{PRESS} = 4.11962 \quad R\text{-Sq(pred)} = 84.93\% \quad R\text{-Sq(LOO)} = 85.307\% \quad q\text{-Sq(LOO)} = 84.932\% \quad R\text{-Sq(L5O)} = 83.6953\% \quad q\text{-Sq} = 81.9695\% \quad R\text{-Sq(L10O)} = 84.4691\% \quad q\text{-Sq(L10O)} = 83.2554\%$$

$$\text{LogIC}_{50} = -18.5376 + 6.1402 \text{ DP01} + 14.2734 \text{ R5v} - 14.3253 \text{ R4m+} + 23.9729 \text{ E2u} - 0.243 \text{ RDF030v} - 48.7015 \text{ G2u} - 0.3577 \text{ Mor02m} - 15.0735 \text{ E2e} - 0.8821 \text{ Mor08u} - 0.6498 \text{ I7-OH} \quad (2)$$

$$n = 42 \quad S = 0.277962 \quad R\text{-Sq} = 91.2\% \quad R\text{-Sq(adj)} = 88.4\% \quad \text{PRESS} = 4.52421 \quad R\text{-Sq(pred)} = 83.45\% \quad R\text{-Sq(LOO)} = 83.828\% \quad q\text{-Sq(LOO)} = 83.452\% \quad R\text{-Sq(L5O)} = 83.222\% \quad q\text{-Sq} = 82.831\% \quad R\text{-Sq(L10O)} = 83.721\% \quad q\text{-Sq(L10O)} = 83.503\%$$

$\text{LogIC}_{50} = 6.324 + 0.1998 \text{ RDF020e} + 19.081 \text{ R5v} - 17.949 \text{ R4m+} - 1.9537 \text{ Mor14e} - 45.758 \text{ G2u} - 4.2394 \text{ Mor20u} + 0.1105 \text{ RDF050m} - 55.77 \text{ H8m} - 1.5404 \text{ Mor16m} + 0.08592 \text{ RDF055u}$
(3)
 $n = 42$ $S = 0.285231$ $R\text{-Sq} = 90.8\%$ $R\text{-Sq}(\text{adj}) = 87.8\%$ $\text{PRESS} = 3.76138$ $R\text{-Sq}(\text{pred}) = 86.24\%$ $R\text{-Sq}(\text{LOO}) = 86.361\%$ $q\text{-Sq}(\text{LOO}) = 86.242\%$ $R\text{-Sq}(\text{L5O}) = 86.097\%$ $q\text{-Sq} = 85.987\%$
 $R\text{-Sq}(\text{L10O}) = 86.277\%$ $q\text{-Sq}(\text{L10O}) = 86.055\%$

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, S is the standard error of estimate. All those equations resulting in low value of R (< 0.85) were not considered being statistically significant. The high values of R , R^2 , $R^2(\text{LOO})$ and low value of S indicates that models have excellent statistical significance. Moreover the values of R^2_{ad} , 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 [4]. We obtained improved models with better correlation statistics when multivariate descriptors are used.

QSAR equations are frequently derived during drug designing. These equations involve mathematical correlation between dependent variable and independent variables. Usually statistical methods like MLR, ANN etc are used to find out the correlations. After deriving the equation the primary goal and important step is to validate the model. This is done in order to prove that the equation is statistically sound [6, 7]. Most of the time it is assumed that being statistically very stable is more than sufficient for the QSAR equation. With the same assumption we derived many QSAR equations for MAO inhibitory activity of Xanthone derivatives. The equations were subjected to thorough statistical validation and it was observed that all the derived equations are statistically very significant. The equations will be useful in designing new xanthone derivatives with better activity profile.

SAR analysis: Till the date no SAR analysis is available on MAO inhibitory activity of xanthone derivatives. With this logical assumption that similar molecules undergo same metabolism and show similar action because they tend to make similar interactions with the protein and also bind to a common active site in the protein, we carried out SAR analysis. The SAR analysis indicates that (1) Presence of $-\text{OH}$ at position 4 is highly beneficial and blocking the H-bonding capacity of $-\text{OH}$ group decreases the activity therefore it is clear that H-bonding is involved between drug and receptor. (2) $-\text{OH}$ at position 2 is a positive factor for activity. (3) Presence of $-\text{Cl}$ at position 4 is bad for activity. (4) Presence of $-\text{OH}$ at position 2 and 5 is advantage. (5) Simultaneous presence of $-\text{OH}$ at position 2 and 5 is good.

CONCLUSION

From the result and discussion it is clear that only (1) ten 3D descriptors are sufficient for predicting the activity. (2) Presence of $-\text{OH}$ at position 4 is highly beneficial and blocking the H-bonding capacity of $-\text{OH}$ group decreases the activity therefore it is clear that H-bonding is involved between drug and receptor. (3) $-\text{OH}$ at position 2 is a positive factor for activity. (4) Presence of $-\text{Cl}$ at position 4 is bad for activity. (5) Presence of $-\text{OH}$ at position 2 and 5 is advantage. (6) Simultaneous presence of $-\text{OH}$ at position 2 and 5 is good.

Acknowledgement

We are thankful to Dr. F. C. Raghuwanshi for their support and Sumit O. Bajaj, C. Eugene Bennett Department of Chemistry, West Virginia University, Morgantown, West Virginia 26506-6045, USA for providing useful data and helpful discussions.

Supporting Information

The values of all descriptors are available free of charge from corresponding author on request.

REFERENCES

- [1] M. Toprakci, K.Yelekci, *Bioorg. Med. Chem. Lett.*, **2005**, 15, 4438
- [2] R.R. Nadendla, Principles of Organic Medicinal Chemistry, New Age International (P) ltd., New Delhi, **2005**, 99
- [3] O. Deeb, S. Alfalah, B. W. Clare, *J. Enz. Inhib. Med. Chem.*, **2007**, 22(3), 277 and references cited therein.
- [4] V. H. Masand, K. N. Patil, G. M. Nazerruddin, R. D. Jawarkar, S. O Bajaj., *Org. Chem.: An Ind. J.*, **2010**, 6(1), 30
- [5] T. Scior, J. L. M. Franco, Q. T. Do, K. M. Mayorga, J. A. Yunes Rojas, P. Bernard, *Curr. Med. Chem.*, **2009**, 16, 4297
- [6] K. Kraim, D. Khatmi, Y. Saihi, F. Ferkous, M. Brahim, *Chemometrics Intelli. Lab. Syst.*, **2009**, 97, 118
- [7] S. Chatterjee, A. S. Hadi, Regression Analysis by Example Fourth edn, Wiley Interscience, **2006**, 85