Journal of Enzyme Inhibition and Medicinal Chemistry, February 2009; 24(1): 77–85

Topological descriptors in modeling malonyl coenzyme A decarboxylase inhibitory activity: *N*-Alkyl-*N*-(1,1,1,3,3,3-hexafluoro-2-hydroxypropylphenyl)amide derivatives

P. SINGH[1], R. KUMAR[1], B. K. SHARMA[1], & Y. S. PRABHAKAR[2]

[1]*Department of Chemistry, S.K. Government College, Sikar 332 001, India, and* [2]*Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow 226 001, India*

(*Received 8 August 2007; in final form 10 December 2007*)

Abstract

The malonyl-CoA decarboxylase (MCD) inhibition activity of derivatives of *N*-alkyl-*N*-(1,1,1,3,3,3-hexafluoro-2-hydroxypropylphenyl)amide has been analyzed through combinatorial protocol in multiple linear regression (CP-MLR) using different topological descriptors obtained from Dragon software for the energy minimized 3D-structures of these molecules. Among the topological descriptor classes considered in the study, the MCD inhibition activity is correlated with simple topological descriptors (TOPO) and 2D-autocorrelation descriptors (2DAUTO). The complementary information contents having neighborhood symmetry of 2-order, CIC2 from the TOPO class, the Geary autocorrelations-lag 8, weighted by atomic Sanderson electronegativities, GATS8e and the Moran autocorrelations-lag 6, weighted by atomic Sanderson electronegativities, MATS6e both from 2DAUTO class have contributed significantly in the development of a statistical significant model.

Keywords: *Quantitative structure-activity relationship (QSAR), malonyl-CoA decarboxylase inhibitors (MCD),* N-*alkyl*-N-*(1,1,1,3,3,3-hexafluoro-2-hydroxypropylphenyl)amides, combinatorial protocol in multiple linear regression (CP-MLR), topological descriptors*

Introduction

The enzyme malonyl-CoA decarboxylase (MCD, EC 4.1.1.9; CoA = coenzyme A) catalyzes the conversion of malonyl-CoA to acetyl-CoA and thus regulates malonyl CoA levels. Malonyl-CoA is a potent endogenous inhibitor of carnitine palmitoyl-transferase-I (CPT-I), an enzyme necessary for the mitochondrial metabolism of long-chain fatty acids. CPT-I, the rate-limiting enzyme in fatty acid oxidation, catalyzes the formation of acylcarnitine, which is transported from the cytosol across the mitochondrial membranes by acylcarnitine translocase. Inside the mitochondria, the long-chain fatty acids are transferred back to their CoA compounds by a complementary enzyme, CPT-II, where acyl-CoA enters the β-oxidation pathway, generating acetyl-CoA.

A number of studies indicate that shifting energy metabolism in the heart towards glucose oxidation is an effective approach to decrease the symptoms associated with myocardial ischemia [1,2]. The antianginal drugs such as ranolazine inhibit fatty acid β-oxidation and stimulate glucose oxidation [3]. Trimetazidine has been shown to specifically inhibit the long-chain 3-ketoacetyl CoA thiolase (3-KAT), an essential step in fatty acid oxidation [4]. Inhibiting CPT-I activity through increasing malonyl-CoA levels with MCD inhibitors would result in a novel and perhaps a much safer method, compared to other known small-molecule CPT-I inhibitors, to the prophylaxis and treatment of ischemic heart disease [5].

Recently, Cheng et al. [6] have argued that increasing the malonyl-CoA level in tissues has potential application in heart disease and diabetes therapy. They have screened in-housed library compounds employing either MCD from rat heart or maltose-binding protein (MBP)-fused human

Correspondence: P. Singh, Department of Chemistry, S.K. Government College, Sikar 332 001, India. E-mail: psingh_sikar@rediffmail.com

ISSN 1475-6366 print/ISSN 1475-6374 online © 2009 Informa UK Ltd.
DOI: 10.1080/14756360801915336

MCD protein and reported a number of small molecules as MCD inhibitors. These compounds were derived from the phenylhexafluoroisopropanol moiety, which was found to be essential for the high *in vitro* MCD inhibitory activity having favorable pharmacokinetic properties with efficacy in regulating energy metabolism in isolated working rat hearts. However, their structure-activity relationship (SAR) study was targeted at the alterations of substituents at different positions of the said moiety and provided no rationale to reduce the trial-and error factors. Hence, in the present communication, a 2D-quantitative SAR (2D-QSAR) study on these analogues (Table I, along with the generalized structure) has been conducted to provide the rationale for drug-design and to explore the possible mechanism of their action at molecular level. Recently a 3D-QSAR study on similar type of inhibitors has been reported [7], using the comparative molecular field analysis (CoMFA) and the comparative molecular similarity indices analysis (CoMSIA). This study could reveal the importance of steric, electrostatic, hydrophobic and H-bond acceptor fields. Additionally the 3D-contour maps, obtained through these fields, were proposed to be used in the design of more potent MCD inhibitors. Both the 2D- and 3D-QSAR studies are equally important if the developed statistical models under them have high predictive power for the new potential congeners. The emerged models from such studies may assist to identify the type of interactions involved between a drug molecule and the receptor sites. The possible mechanism of action of the compounds anticipated by one study may, therefore, be corroborated through the other one. However, a 2D-QSAR study is quite simple to interpret the biological data in terms of different descriptors obtained from the two dimensional structures of the compounds without involving energy minimization procedures. In a congeneric series, where a relative study is being carried out, the 2D- descriptors may play important role in deriving the significant correlations with biological activities of the compounds. Thus the novelty and importance of a 2D-QSAR study is mainly due to its simplicity for the calculations of different descriptors and their interpretation (in physical sense) to explain the inhibition actions of compounds in a congeneric series.

Material and methods

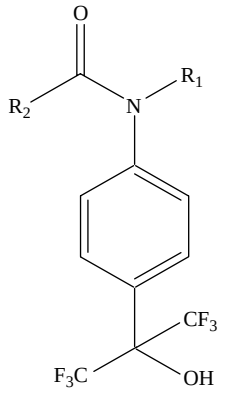
Data set

In the present study, the compounds derived from phenylhexafluoro- isopropanol moiety (Table I) have been considered from the literature [6] along with their *in vitro* malonyl coenzyme A decarboxylase (MCD) inhibition activity. The activity, IC_{50}, was measured as the concentration of an inhibitor to achieve 50% inhibition of the MCD enzyme. For present work, the same is expressed as $-\log IC_{50}$ on a molar basis. DRAGON software [8] has been used for the parameterization of the compounds of this study (Table I). This software offers several hundreds of descriptors from different perspectives relating to empirical, constitutional and topological indices characteristic to the molecules under multi-descriptor class environment. The structures of the compounds under study have been drawn in 2D ChemDraw [9] using the standard procedure. These structures were converted into 3D objects using the default conversion procedure implemented in the CS Chem3D Ultra. The generated 3D-structures of the compounds were subjected to energy minimization in the MOPAC module, using the AM1 procedure for closed shell systems, implemented in the CS Chem3D Ultra. This will ensure a well defined conformer relationship across the compounds of the study. All these energy minimized structures for the varying R_1 and R_2 groups of respective compounds have been ported to DRAGON software for computing the parameters corresponding to 0D-, 1D-, and 2D-descriptor classes. The descriptor classes considered in the study along with their definitions and scope in addressing the structural features have been presented in Table II. As the total number of descriptors involved in this study is very large, only the names of descriptor classes and the actual descriptor involved in the models have been listed. The combinatorial protocol-multiple linear regression (CP-MLR) computational procedure used for present work in developing QSAR models is briefly described below.

Model development

The CP-MLR is a 'filter' based variable selection procedure for model development in QSAR studies [10–13]. The procedure employs a combinatorial strategy with MLR to result in selected subset regressions for the extraction of diverse structure-activity models, each having unique combination of descriptors from the generated data set of the compounds under study. The 'filters' set in CP-MLR are intended at (i) having inter-parameter correlation to a predefined cutoff value (filter-1; default acceptable value ≤ 0.3, (ii) optimize the variable entry to a model through *t*-value of regression coefficients (filter-2; default acceptable value ≥ 2.0, (iii) comparability of models (regression equations) with different number of descriptor in terms of square root of adjusted multiple correlation coefficient (filter-3; *r*-bar, default acceptable value ≥ 0.74), and (iv) addressing the external consistency of the model with leave-one-out (LOO) cross-validation as default option (filter-4; cross-validated Q^2 criteria, default acceptable limits are $0.3 \leq Q^2 \leq 1.0$). All these filters make the variable selection process efficient and lead

Table I. Observed and modeled malonyl coenzyme A decarboxylase inhibition activity of *N*-alkyl-*N*-(1,1,1,3,3,3-hexafluoro-2-hydroxypropylphenyl)amide derivatives.



						$-\log IC_{50}$ (M)	
S.No	R_1	R_2	CIC2	MATS6e	GATS8e	Obsd*	Calcd Eq (3)
1	Me	CH_2OPh	1.704	−0.223	0.758	6.03	5.87
2	H	CH_2OPh	1.470	−0.310	0.982	5.18	5.19
3	H	CH_2CH_2Ph	1.638	−0.227	0.913	5.36	5.54
4	Me	CH_2CH_2Ph	1.851	−0.127	0.731	6.11	6.19
5	Et	CH_2CH_2Ph	1.574	−0.066	0.682	6.76	6.96
6	*i*-Pr	CH_2CH_2Ph	1.692	−0.029	0.642	6.70	7.08
7	*n*-Bu	CH_2CH_2Ph	1.732	−0.089	0.519	7.10	7.07
8	CH_2COOMe	CH_2CH_2Ph	1.617	−0.194	0.608	6.98	6.51
9	CH_2COOH	CH_2CH_2Ph	1.517	−0.206	0.716	6.43	6.30
10	benzyl	CH_2CH_2Ph	2.070	−.114	0.527	6.24	6.50
11	cyclohexyl	CH_2CH_2Ph	2.084	−0.073	0.491	6.82	6.77
12	H	*i*-Pr	1.343	−0.250	1.009	NA†	5.56
13	Me	*i*-Pr	1.544	−0.040	0.913	6.75	6.51
14	Me	3-Py	1.369	−0.166	0.869	5.99	6.28
15	Me	4-Py	1.475	−0.114	0.891	6.41	6.32
16	Et	*i*-Pr	1.513	−0.002	0.897	7.39	6.77
17	*n*-Pr	*i*-Pr	1.455	−0.036	0.626	7.55	7.39
18	*n*-Bu	*i*-Pr	1.490	−0.038	0.530	7.70	7.59
19	$CH_2CH_2NMe_2$	*i*-Pr	1.719	−0.047	0.498	5.94‡	7.34
20	CH_2CH_2OH	*i*-Pr	1.328	−0.054	0.773	6.98	7.09
21	CH_2COOMe	*i*-Pr	1.222	−0.123	0.612	7.47	7.33
22	CH_2COOH	*i*-Pr	1.244	−0.128	0.774	6.20	6.86
23	$(CH_2)_4COOMe$	*i*-Pr	1.463	−0.015	0.554	7.59	7.67
24	$(CH_2)_4COOH$	*i*-Pr	1.331	−0.022	0.555	8.10	7.80
25	$(CH_2)_4$-(5-1*H*-tetrazole)	*i*-Pr	1.218	−0.028	0.542	8.10	7.95
26	$(CH_2)_4CN$	*i*-Pr	1.344	−0.035	0.536	7.70	7.77
27	Et	*p*-CNPh	1.402	−0.067	0.701	7.43	7.12
28	Et	4-Py	1.214	−0.055	0.791	7.02	7.19
29	H	NHPh	1.758	−0.345	0.862	NA†	4.97
30	Me	NHPh	1.836	−0.225	0.705	5.81	5.83
31	H	morpholinyl	1.269	−0.264	0.836	NA†	6.04
32	Et	NMe_2	1.587	−0.025	0.890	7.23	6.59
33	Et	$N(Et)(CH_2)_4OH$	1.444	−0.035	0.542	7.22	7.63
34	Me	morpholinyl	1.456	−0.072	0.790	6.49	6.80
35	Et	morpholinyl	1.373	−0.036	0.726	7.13	7.24
36	*n*-Pr	morpholinyl	1.329	−0.079	0.578	7.47	7.48
37	*n*-Bu	morpholinyl	1.368	−0.076	0.527	7.52	7.58
38	*n*-pentyl	morpholinyl	1.431	−0.076	0.574	7.72	7.38
39	*n*-hexyl	morpholinyl	1.540	−0.063	0.568	7.51	7.31
40	*n*-heptyl	morpholinyl	1.662	−0.055	0.573	7.14	7.18
41	octyl	morpholinyl	1.790	−0.048	0.562	6.57	7.07
42	CH_2CH_2Ph	morpholinyl	1.642	−0.055	0.588	7.37	7.16
43	CH_2CH_2OH	morpholinyl	1.506	−0.084	0.729	6.78	6.84
44	CH_2COOMe	morpholinyl	1.329	−0.135	0.707	7.10	6.89
45	CH_2CH_2COOMe	morpholinyl	1.385	0.012	0.893	7.00	7.01
46	CH_2CH_2COOH	morpholinyl	1.313	0.012	0.874	6.61	7.15

Table I – *continued*

S.No	R_1	R_2	CIC2	MATS6e	GATS8e	$-\log IC_{50}$ (M) Obsd*	Calcd Eq (3)
47	$(CH_2)_3COOMe$	morpholinyl	1.457	−0.048	0.677	7.13	7.20
48	$(CH_2)_3CN$	morpholinyl	1.327	−0.072	0.589	7.40	7.49
49	$(CH_2)_4COOMe$	morpholinyl	1.536	−0.050	0.530	7.38	7.47
50	$(CH_2)_4COOH$	morpholinyl	1.417	−0.061	0.528	7.85	7.58

* IC_{50} represents the concentration of an inhibitor to bring out 50% inhibition of the MCD enzyme; taken from ref. (6); † inactive at 10 μM.; ‡ 'outlier' compound of present study.

to unique solution. Further, to find out any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been subjected to randomization test [14,15] by repeated randomization of the biological responses. The datasets with randomized response vector have been reassessed by multiple regression analysis. The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted. Every model has been subjected to 100 such simulation runs. This has been used as a measure to express the percent chance correlation of the model under scrutiny. The CP-MLR protocol has been applied with default filter thresholds to identify the all possible models that could emerge from the descriptors of compounds.

Results and discussion

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. With this, attempts have been made to develop one-, and two-descriptor models for the malonyl coenzyme A decarboxylase inhibition activity of *N*-alkyl-*N*-(1,1,1,3,3,3-hexafluoro-2-hydroxypropylphenyl)-amide derivatives. The study has led to forty-one models in the topological (TOPO) class descriptors, one model each in modified Burden eigenvalues (BCUT), and Galvez topological charge indices (GVZ) and twenty-six models in 2D-autocorrelations (2DAUTO) class descriptors (Table II). In these individual classes, the models derived in two descriptors, compared to the models in one- or three-descriptors, are only statistically sound. Thus, the collective descriptors of these classes may further result into the improved models. However, no derived model in collective descriptors could explain the variance in activity beyond 67%. This prompted us to search out for the models in three-descriptors from the collective class, resetting filter-3 at 0.84 without altering the other filters. The extract of all such models obtained from the study has been provided in Table III in the form of identified descriptor's average of regression coefficients along with standard deviation across the models and the total incidence corresponding to all the models. This, while providing the averages of the estimated regression coefficients of all the identified

Table II. Descriptor classes used for the analysis of malonyl coenzyme A decarboxylase inhibitiory activity of *N*-alkyl-*N*-(1,1,1,3,3,3-hexafluoro-2-hydroxypropyl phenyl)amide derivatives and identified categories in modeling the activity.

Descriptor class (acronyms)	Definition and scope	Descriptors per model (Number of models)*
Topological (TOPO)	2D-descriptor from molecular graphs & independent of conformations	2 (41)
Modified Burden eigenvalues (BCUT)	2D-descriptors representing positive & negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms	2 (1)
Galvez topological charge indices (GVZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix	1 (1); 2 (1)
2D-autocorrelations (2DAUTO)	Molecular descsriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag)	1 (1); 2 (26)

* Models emerged from CP-MLR protocol with filter-1 as 0.3; filter-2 as 2.0; filter-3 as 0.74; filter-4 as $0.3 \leq Q^2 \leq 1.0$, number of compounds in each dataset was 47.

descriptors, shows their variance across the models emerging from the study as well. To be concise, the complete regression equations have been shown for selected models only. The following regression equations represent collective class structure-activity model of the compounds.

$$-\log IC_{50} = 9.686(1.540)\text{BIC2} + 4.292(0.850)\text{MATS8e} - 1.693(0.445)\text{GATS7e} + 0.483 \quad (1)$$

$$n = 47,\ r = 0.859,\ Q^2 = 0.672,$$

$$s = 0.384,\ F(3,43) = 40.328$$

$$-\log IC_{50} = 1.427(0.272)\text{CIC2} + 4.382(0.806)\text{MATS6e} - 2.395(0.398)\text{GATS8e} + 11.085 \quad (2)$$

$$n = 47,\ r = 0.868,\ Q^2 = 0.753,$$

$$\text{hskip11pt}s = 0.367,\ F(3,43) = 43.798$$

In above and all the follow up regression equations, n is the number of compounds, r is the correlation coefficient, Q^2 is cross-validated index from leave-one-out (LOO) procedure, s is the standard error of the estimate and F is the F-ratio between the variances of calculated and observed activities. The values given in the parentheses are the standard errors of the regression coefficients. Also, in the randomization study, where 100 simulations per model were carried out, none of the identified models has shown any chance correlation. The six models, including the above-shown two, emerged from the collective class of descriptors belonging to TOPO, BCUT, GVZ and 2DAUTO classes. However the above two models, being more significant compared to remaining four (not given here), shared the descriptors belonging to TOPO and 2DAUTO classes only. Among these descriptors, the bond information contents, BIC2 and the complementary information contents, CIC2, both being neighborhood symmetry of 2-order, belong to TOPO class. The 2DAUTO descriptors, MATSke and GATSke have their origin in autocorrelation of topological structure of Moran and of Geary [16,17], respectively. The computation of these descriptors involves the summation of different autocorrelation functions corresponding to the different fragment lengths and lead to different autocorrelation vectors corresponding to the lengths of the structural fragments [18]. Also a weighting component in terms of a physicochemical property has been embedded in these descriptors. As a result, these descriptors address the topology of the structure or parts thereof in association with a selected physicochemical property. In these descriptors' nomenclature, the penultimate character, a number, indicates the number of consecutively connected edges considered in its computation and is called as the autocorrelation vector of lag k (corresponding to the number of edges in the unit fragment). The very last character of the descriptor's nomenclature indicates the physicochemical property considered in the weighting component— e for atomic Sanderson electronegativity— for its computation. In above two equations, Geary autocorrelation lag 7 and 8 weighted by electronegativity (GATS7e and GATS8e) and Moran autocorrelation lag 6 and 8 weighted by electronegativity (MATS6e and MATS8e) are correlated with the activity. From Table III, it appears that though the descriptors of Equation (1) have participated in more number of models compared to the descriptors of Equation (2) but the later one is statistically more significant. This implies that the descriptors of Equation (2) are more enriched in information content to influence the malonyl coenzyme A decarboxylase inhibition activities of the compounds under study. This equation was, therefore, considered further to draw some meaningful conclusions. The lone compound **19**, bearing a $CH_2CH_2NMe_2$ substituent at R_1, does not fit into the trend followed by other congeners of the series. A basic group such as the N,N-dimethylamino group at the terminal position, separated through 4-lags, is undesirable and may significantly decrease the potency [6]. Possibly, an acidic or neutral moiety present at lag-4 is essential for desired interaction of R_1 substituents. This compound was, therefore, eliminated from the data-set and the derived correlation is shown in Equation (3)

$$-\log IC_{50} = -1.285(0.224)\text{CIC2} + 4.579(0.661)\text{MATS6e} - 2.603(0.328)\text{GATS8e} + 11.057 \quad (3)$$

$$n = 46,\ r = 0.911,\ Q^2 = 0.791,$$

$$s = 0.285,\ F(3,42) = 68.620$$

All the statistical parameters of this equation are improved over to that of Equation (2). The r^2 now accounts for 83% of variance in observed activity values while F-value remained significant at 99% level [$F_{3,43}(0.01) = 4.30$]. Also, the value obtained for cross-validated Q^2 index is in favor of a robust QSAR model. Above model equation was further subjected to randomization process, where 100 simulations per model were carried out but none of the identified models has shown any chance correlation.

Table III. Descriptors identified for modeling the malonyl coenzyme A decarboxylase inhibitory activity of *N*-alkyl-*N*-(1,1,1,3,3,3-hexafluoro-2-hydroxypropylphenyl)amide derivatives along with the average regression coefficients, standard deviation and the total incidence.

Descriptor*	Avg reg coeff (sd) total incidence†	Descriptor*	Avg reg coeff (sd) total incidence†
TOPO			
X2A	−27.459 (–)1	X0Av	13.395 (–) 1
IVDE	1.959 (–) 1	BIC2	10.675 (1.247) 4
IC1	1.247 (–) 1	IC3	0.794 (–) 1
IC2	2.037 (0.221) 5	SIC3	10.696 (–) 1
TIC2	0.013 (–) 1	BIC3	10.913 (3.004) 3
SIC2	10.694 (1.654) 3	CIC5	1.826 (–) 1
CIC2	−1.462 (0.535) 3		
BCUT			
BELm3	3.327 (0.224) 2	BELm6	3.406 (–) 1
GVZ			
GGI2	1.246 (–) 1	JGI8	152.478 (9.716) 4
GGI4	1.240 (–) 1	JGI9	122.281 (–) 1
GGI8	3.453 (0.083) 2		
2DAUTO			
ATS5p	0.045 (0.011) 2	MATS8e	5.769 (0.505) 4
MATS4m	−61.449 (2.748) 2	GATS4e	3.261 (–) 1
MATS5m	−47.798 (–) 1	GATS6e	−3.611 (0.227) 8
MATS7m	21.965 (5.034) 4	GATS7e	−2.256 (0.729) 3
MATS6e	6.475 (–) 1	GATS8e	−2.466 (0.174) 2
MATS7e	2.599 (0.713) 2		

* The descriptors are identified from the one and two parameter models emerged from CP-MLR protocol with filter-1 as 0.3; filter-2 as 2.0; filter-3 as 0.76; filter-4 as $0.3 \leq Q^2 \leq 1.0$; number of compounds in the study was 47; X2A and X0Av are the average connectivity indices (chi-2 and chi-0, respectively); IVDE is the mean information content on the vertex degree equality; BIC2 and BIC3 are the bond information contents (neighborhood symmetry of 2- and 3-order, respectively); IC1, IC2 and IC3 are the information content indices (neighborhood symmetry of 1-, 2- and 3-order, respectively); SIC2 and SIC3 are the structural information contents (neighborhood symmetry of 2- and 3-order, respectively); TIC2 is the total information content index (neighborhood symmetry of 2-order); CIC2 and CIC5 are the complementary information contents (neighborhood symmetry of 2- and 5-order, respectively); BELm3 and BELm6 are the lowest eigenvalue n = 3 and 6 of Burden matrices, respectively / weighted by atomic masses; GGI2, GGI4 and GGI8 are the topological charge indices of order 2, 4 and 8, respectively; JGI8, JGI9 are the mean topological charge indices of order 8 and 9, respectively; ATS5p is the Broto-Moreau autocorrelation of a topological structure–lag 5/weighted by atomic polariazabilities; MATS4m, MATS5m, MATS7m are the Moran autocorrelations-lag 4, 5 and 7, respectively / weighted by atomic masses; MATS6e, MATS7e, MATS8e are the Moran autocorrelations-lag 6, 7 and 8, respectively / weighted by atomic Sanderson electronegativities; GATS4e, GATS6e, GATS7e, GATS8e are the Geary autocorrelations-lag 4, 6, 7 and 8, respectively / weighted by atomic Sanderson electronegativities; also see ref. (7); † the average regression coefficient of the descriptor corresponding to all models and the total number of its indices; the arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.

Additionally, the above model equation was also subjected to external validation. For this purpose three test sets, each containing 14 compounds out of the 46 active ones listed in Table I, were considered. Of the three test sets, two were generated in the SYSTAT [19] using the single linkage hierarchical cluster procedure involving the Euclidean distances of the respective descriptors or the activity as the case may be. The selection of the test set from the cluster tree was done in such a way to keep the test compounds at a maximum possible distance from each other. The third test set of the compounds corresponds to the random selection procedure. With this, these test sets represent different cross-sections of compounds. The predictions of the test sets have been done with the models developed using the 46 compounds remaining in the training sets.

The activity values of all three inactive compounds (**12**, **29** and **31**) have also been predicted from these sets. The residuals of the predictions and the corresponding predictive R^2-values of each test set have been given in Table IV. The predictions corresponding to all compounds in three test sets are within the reasonable limits of their actual values. Also, all the models from the training sets have predicted the two out of the three inactive compounds as low active ones. The exception is with compound **31** which is bearing a morpholinyl moiety (Table I) as the varying group in this congener and predicted to have activity values, from three test sets, in the range of slight active compounds. The calculated activity value of outlier compound **19**, using Equation (3), has significantly deviated from the observed ones (Table I). Also, the predicted activity of this congener, from three test sets

Table IV. Predicted residual activity of three test-sets (14 compounds in each) derived from the compounds in Table I.

	Residual*		
Compound number	Des Clus†	Act Clus‡	Rand Clus¶
1			0.25
2	0.00	−0.20	
3	−0.16		
4			−0.05
6		−0.23	
7			0.09
8	0.49		
11	0.10	0.28	0.12
12§	5.55	5.73	5.53
15			0.06
16		0.77	
18	0.13		
19	7.31	7.22	7.32
20			−0.15
22		−0.77	
23			−0.08
24		0.30	
26		−0.09	
27			0.27
28	−0.16	−0.21	
29§	4.95	5.11	4.83
30			0.09
31§	6.04	6.28	5.96
32	0.67	0.79	
34	−0.30		
36		−0.07	0.02
38	0.36	0.33	
40			−0.01
41	−0.47		
42	0.23	0.30	
44			0.24
45	0.00		
46		−0.46	−0.65
47	−0.06		
48	−0.08		
50		0.25	0.31
Test set R^2	0.841	0.678	0.850

* The difference between observed and predicted $-\log IC_{50}$ values; the training models have 32 compounds each; † test-set from the cluster analysis of all descriptors used in deriving Equation (3); ‡ test-set from the cluster analysis of the activity of the compounds; ¶ test-set from random selection of the compounds; § predicted activity for the compounds which are reported inactive at 10 μM.

(Table IV), is comparatively much higher. This favors our consideration that compound **19** is an obvious outlier. The overall satisfactory performance of all the test sets provided us the confidence about the procedure adopted in the study and the variables selected therein. Additionally, the plot in Figure 1 is given to show the goodness of fit and to identify the systematic variation between observed and calculated—logIC_{50} values (Equation 3). Similarly, the plot in Figure 2 is given to demonstrate the trend set by the compounds of three training sets, derived through clustering of descriptors, activity values and random selection. The similar trend is followed by the compounds present in corresponding test set.

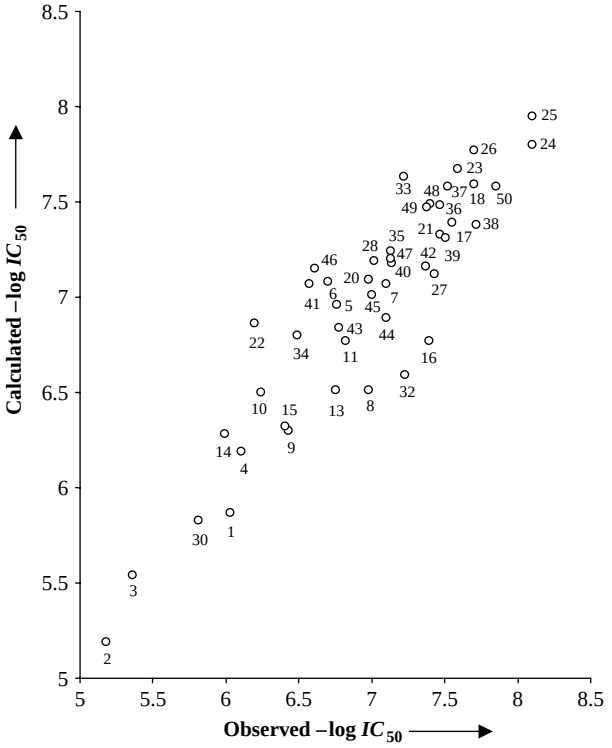


Figure 1. Plot of observed versus calculated $-\log IC_{50}$ values.

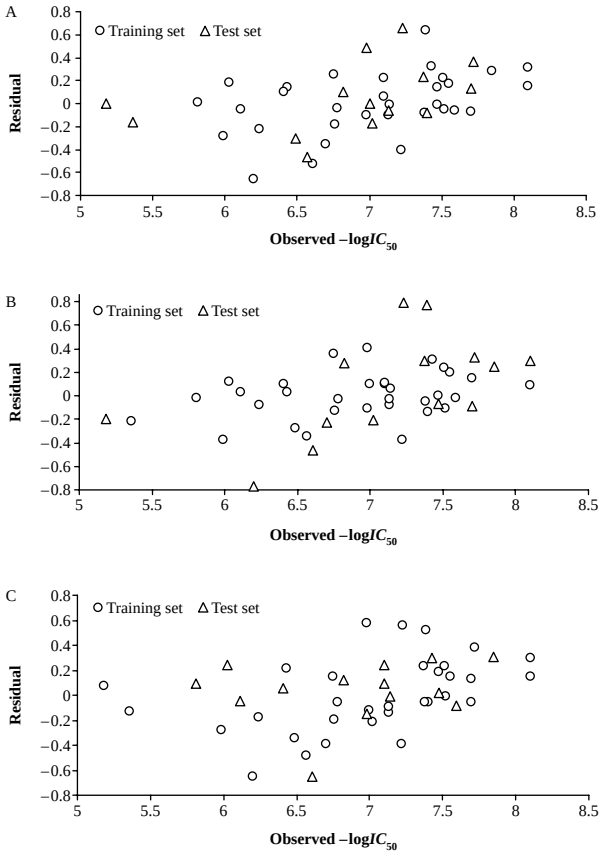


Figure 2. Plot of observed $-\log IC_{50}$ versus residual values of training and test sets developed from (A) Descriptors, (B) Activity values and (C) Random selection of the compounds.

The residuals (observed— calculated activity values) obtained for the aforesaid test sets are found to be evenly distributed about the zero residual line. The predictions of three test sets are, therefore, in parity of what an excellent model can yield.

From Equation (3), it appeared that the complementary information contents having neighborhood symmetry of order-2, CIC2 and the Geary autocorrelations-lag 8, weighted by atomic Sanderson electronegativities, GATS8e have both added negatively to inhibition activity while the Moran autocorrelations-lag 6, weighted by atomic Sanderson electronegativities, MATS6e contributed positively to it. This implies that the smaller positive values of the parameters CIC2 and GATS8e and the larger positive (or smaller negative) value of MATS6e are preferred to improve the activity of a compound. The descriptors, MATS6e and GATS8e have emphasized the importance of electronic enriched centers at the path length-6 and -8, respectively. These lags, in particular, are important for the substituents appeared at R_1 and are apparent in the most potent compounds **24**, **25**, and **50** ($-\log IC_{50} > 7.80$; Table I). It appears that there are two important sites of the receptor which may be 7 and/or 8 bonds away from the benzene ring. These sites being opposite in nature (acidic or basic but not neutral) may engage themselves in electronic type of interaction. The functionality, such as carboxyl group in compounds **24** and **50** and tetrazolyl ring in compound **25,** appears to be electronically enriched through lag-6, while the hydrogen present at lag-8 is unable to do so. The doubly bonded oxygen (of carboxyl group) or nitrogen (of tetrazolyl ring) atom may, therefore, engage itself with proton-donor (acidic) site while hydrogen bonded oxygen or nitrogen perhaps interact with proton-acceptor (basic) site of the receptor. This implies that a descriptor weighted by electronegativity at lag-6, contributes to enhance the activity and the descriptor weighted similarly at lag-8 reduces the same. The positive and negative regression coefficients associated, respectively to MATS6e and GATS8e, in Equation (3), have substantiated the same.

Since the biological activity of a compound is by virtue of the mere presence of its constituent atoms and bonds, therefore, depending upon the number and type of atoms and the arrangement of different bonds, a molecule may conceal unique information content in the form of electron probability fields distributed in space. This information content of different symmetry orders is of considerable importance as its interpretation by the receptor site determines the efficacy of the drug molecule. The descriptor CIC2, being the function of molecular structure is, therefore, embodies the complementary information content of the 2-order symmetry and emerged as one of the promising parameter for the inhibition of MCD. Its negative but significant contribution to inhibition activity has required for those substituents in parent structure which could yield the lower values of CIC2.

Thus the criteria mentioned above may be used to explore new potential compounds of the series.

Acknowledgements

We express sincere thanks to our Institutions for providing necessary facilities to complete this work.

References

[1] Hearse D. Metabolic approaches to ischemic heart disease and its management. New York: Science Press Ltd; 1998.

[2] Kennedy JA, Unger SA, Horowitz JD. Inhibition of carnitine palmitoyltransferase-1 in rat heart and liver by perhexiline and amiodarone. Biochem Pharmacol 1996;52:273–280.

[3] (a) McCormack JG, Barr RL, Wolff AA, Lopaschuk GD. Ranolazine stimulates glucose oxidation in normoxic, ischemic, and reperfused ischemic rat hearts. Circulation 1996;93:135–142, (b) McCormack JG, Stanley WC, Wolff AA. Ranolazine: A novel metabolic modulator for the treatment of angina. Genet Pharmacol 1998;30:639-645. (c) Pepine CJ, Wolff AA. A controlled trial with a novel anti-ischemic agent, ranolazine in chronic stable angina pectoris that is responsive to conventional antianginal agents. Ranolazine study group. Am J Cardiol 1999;84:46-50.

[4] Kantor PF, Lucien A, Kozak R, Lopaschuk GD. The antianginal drug trimetazidine shifts cardiac energy metabolism from fatty acid oxidation to glucose oxidation by inhibiting mitochondrial long-chain 3-ketoacyl coenzyme A thiolase. Circ Res 2000;86:580–588.

[5] Dyck JR, Cheng J-F, Stanley W, Barr R, Chandler MP, Brown S, Wallace D, Arrhenius T, Harmon C, Yang G, Nadzan A, Lopaschuk GD. Malonyl-CoA decarboxylase inhibition protects the ischemic heart by inhibiting fatty acid oxidation and stimulating glucose oxidation. Circ Res 2004;94:e78–e84.

[6] Cheng J-F, Chen M, Wallace D, Tith S, Haramura M, Liu B, Mak CC, Arrhenius T, Reily S, Brown S, Thorn V, Harmon C, Barr R, Dyck JRB, Lopaschuk GD, Nadzan AM. Synthesis and structure-activity relationship of small-molecule malonyl coenzyme A decarboxylase inhibitors. J Med Chem 2006;49:1517–1525.

[7] Patel MR, Talele TT. 3D-QSAR studies on malonyl coenzyme A decarboxylase inhibitors. Bioorg Med Chem 2007;15: 4470–4481.

[8] Dragon software (version 1.11-2001) by Todeschini R, Consonni V. Milano, Italy.

[9] Chemdraw ultra 6.0 and Chem3D ultra. Cambridge Soft Corporation, Cambridge, USA.

[10] Prabhakar YS, A combinatorial approach to the variable selection in multiple linear regression: Analysis of Selwood. data set – A case study. QSAR Comb Sci 2003;22:583–595.

[11] Gupta MK, Sagar R, Shaw AK, Prabhakar YS. CP-MLR directed QSAR studies on the antimycobacterial activity of functionalized alkenols-topological descriptors in modeling the activity. Bioorg Med Chem 2005;13:343–351.

[12] Prabhakar YS, Rawal RK, Gupta MK, Solomon VR, Katti SB. Topological descriptors in modeling the HIV inhibitory activity of 2-aryl-3-pyridyl-thiazolidin-4-ones. Comb Chem High Thro Screen 2005;8:431–437.

[13] Gupta MK, Prabhakar YS. Topological descriptors in modeling the antimalarial activity of 4-(3',5'-disubstituted aniline)quinolines. J Chem Inf Model 2006;46:93–102.

[14] So S-S, Karplus M. Three-dimensional quantitative structure-activity relationship from molecular similarity matrices and genetic neural networks 1. Method and validation. J Med Chem 1997;40:4347–4359.

[15] Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. CP-MLR/PLS directed structure-activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. QSAR Comb Sci 2004;23:234–244.

[16] Moran PAP. Notes on continuous stochastic phenomena. Biometrika 1950;37:17–23.

[17] Geary RC. The contiguity ratio and statistical mapping. Incorp Statist 1954;5:115–145.

[18] Broto P, Moreau G, Vandycke C. 2D autocorrelations. Eur J Med Chem 1984;19:66–70.

[19] SYSTAT, Version 7.0; SPSS Inc., 444 North Michigan Avenue, Chicago, IL, 60611.