Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 512–520

Quantitative structure–activity analysis of 5-arylidene-2,4-thiazolidinediones as aldose reductase inhibitors

S. V. Sambasivarao, Love K. Soni,* Arun K. Gupta, P. Hanumantharao and S. G. Kaskhedikar

Molecular Modelling Study Group, CADD Laboratory, Department of Pharmacy, S.G.S.I.T.S., 23 Park Road, Indore-452003, M.P., India

Received 7 August 2005; revised 28 September 2005; accepted 19 October 2005
Available online 15 November 2005

Abstract—Design of aldose reductase (ALR2) inhibitors has received considerable attention. Aldose reductase inhibitors, when administered from the onset of hyperglycemia, prevent the progression of polyol accumulation-linked complications. The feasibility that inhibition of aldose reductase provides a pharmacologically direct treatment for diabetic complications that is independent of the control of blood sugar levels has spurred the development of structurally diverse aldose reductase inhibitors. In this work, we report quantitative structure–activity relationship (QSAR) analysis performed by 3D-QSAR analysis, Hansch analysis, and Fujita-Ban analysis on a series of 5-arylidene-2,4-thiazolidinediones as aldose reductase inhibitors. The 2D & 3D-QSAR models were generated using 18 compounds and Fujita-Ban analysis models were obtained using 23 compounds. The predictive ability of the resulting 2D and 3D models was evaluated against a test set of 5 compounds. Analyses of results from the present QSAR study inferred that 3rd position of the phenyl ring and acetic acid substitution at *N*-position of thiazolidinediones play a key role in the aldose reductase inhibitory activity.

Diabetes mellitus is a chronic disease characterized by hyperglycemia. Diabetic patients are at risk for developing long-term complications including neuropathy, nephropathy, retinopathy, and cardiovascular diseases.[1] The clinical consequences of these complications include lower limb amputation, end-stage renal failure, and loss of vision. The Diabetic Complications and Control Study[2] and the United Kingdom Prospective Diabetes Study[3] demonstrated that strict and sustained control of glucose excursions through interventions including intensive insulin therapy reduces risk of developing these complications in type 1 and type 2 diabetes. However, relatively few diabetics have adopted this strict, physician-monitored regimen, and this type of round-the-clock control is not practical for patients at large. At present, there is no specific therapy available for diabetic complications. A metabolic approach is to control excess glucose flux in diabetic tissue through the first step of polyol pathway (Fig. 1) by aldose reductase inhibitors.[4,5] Aldose reductase (ALR2), the first enzyme of the polyol pathway, catalyzes the reduction of glucose to sorbitol using nicotinamide adenine dinucleotide phosphate as an essential cofactor, which is further processed by sorbitol dehydrogenase to fructose. Aldose reductase has been localized at the sites of tissue damage, and inhibitors of this enzyme prevent the development of neuropathy, nephropathy, retinopathy, and cataract formation in animal models of diabetes.[4,5] Aldose reductase inhibitors have shown promising results in both preclinical models and early clinical trials.[6] In search for potent inhibitors, many compounds of diverse structures have been identified. Kinetic studies suggest that these compounds appear to interact with the enzyme at a site independent of either substrate or nucleotide cofactor fold. Moreover, from the pioneering studies on sorbinil and alrestatin to recent investigation on zopolrestat and zenarestat, several compounds in clinical trials or in market for the treatment of the diabetic complications have been developed but they were withdrawn, suggesting that currently no 'universally potent' inhibitor exists. This demands the development of potent inhibitors.[7,8] To gain insight into the structural and molecular requirements influencing the aldose reductase inhibitor activity, we herein describe the quan-

Keywords: Aldose reductase inhibitors; 2,4-Thiazolidinediones; Diabetes mellitus; Quantitative structure–activity relationship; Hansch analysis; Fujita-Ban analysis.
* Corresponding author. Tel./fax: +91 0731 2546031; e-mail: lovesoni@hotmail.com

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.10.069

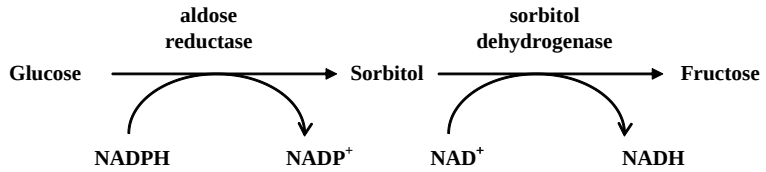


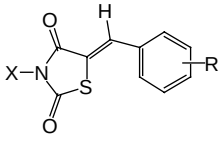
Figure 1. Polyol pathway.

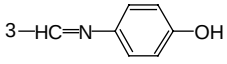
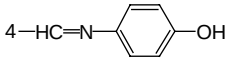
titative structure–activity relationship (QSAR) analysis of 5-arylidene-2,4-thiazolidinediones.[9] The relevance of the best QSAR model obtained for the design of novel derivatives should be assessed not only in terms of predictivity, either internal or external, but also in terms of their ability to provide a chemical and structural explanation of their binding interaction. Here, we propose a model for the aldose reductase inhibitors and present minimal structural requirements for an aldose reductase inhibitor. These results could serve as a guideline in the design of more potent and selective aldose reductase inhibitors.

The aldose reductase inhibitory activity data of 5-arylidene-2,4-thiazolidinediones were taken from the reported work of Bruno et al.[9] (Table 1). The biological activity data (IC_{50} in μm) were converted to a negative logarithmic mole dose (pIC_{50}) for quantitative structure activity relationship analysis. Initially the series was subjected to Fujita-Ban analysis using regression technique in order to estimate the de novo contribution of substituents to the activity of the molecules. Further, Hansch approach was carried out to establish correlations between ALR2 inhibitory activity and various substituent constants at position R and X of the molecule (Fig. 2). Values of substituent constants like hydrophobic (π), steric (Molar refractivity or MR), hydrogen acceptor (HA), hydrogen donor (HD), and electronic descriptor (field effect or $\mathscr{F}$, resonance effect or $\mathscr{R}$, and Hammett's constant or σ) were taken from the reported work of Hansch et al.[10]

The molecular modeling study was performed on a P-III processor using MOE[11] and the regression

Table 1. Structure and activities of 5-arylidene-2,4-thiazolidinediones[8] used in training and test sets



S. No.	Compound	R	X	IC_{50}[a]	pIC_{50}[b]
1	T-1	3-F	H	9.10	5.04
2	T-2	3-CH_3	H	9.23	5.03
3	T-3	3-OCH_3	H	13.28	4.88
4	T-4	3-CF_3	H	12.81	4.89
5	T-5		H	1.86	5.73
6	T-6	4-F	H	8.21	5.09
7	T-7	4-CF_3	H	31.49	4.50
8	T-8		H	22.93	4.64
9	T-9	3-CH_3	CH_2COOCH_3	10.10	5.00
10	T-10	3-OCH_3	CH_2COOCH_3	8.87	5.05
11	T-11	3-CF_3	CH_2COOCH_3	28.67	4.54
12	T-12	4-CF_3	CH_2COOCH_3	3.44	5.46
13	T-13	3-F	CH_2COOH	0.74	6.13
14	T-14	3-CH_3	CH_2COOH	0.65	6.19
15	T-15	3-OC_6H_5	CH_2COOH	0.13	6.89
16	T-16	3-OCH_3	CH_2COOH	0.48	6.32
17	T-17	4-F	CH_2COOH	1.14	5.94
18	T-18	4-CF_3	CH_2COOH	0.46	6.34
19	Test-1	3-OC_6H_5	H	6.14	5.21
20	Test-2	4-OCH_3	H	40.83	4.39
21	Test-3	3-OC_6H_5	CH_2COOCH_3	1.32	5.88
22	Test-4	4-F	CH_2COOCH_3	12.90	4.89
23	Test-5	3-CF_3	CH_2COOH	0.47	6.33

[a] In vitro IC_{50} (50% inhibitory concentration in μM) against bovine lenses aldose reductase.
[b] Negative logarithmic of IC_{50} value in mole.

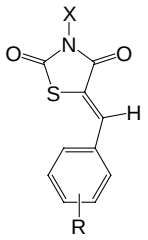


Figure 2. General structure for the present study.

analysis program VALSTAT.[12] The molecular structures of all 23 compounds (except for compounds with % inhibitory activity) were sketched using the builder module software and energy minimized via steepest descent, conjugate gradient, and truncated Newton method in sequence using MMFF94 as force field with an energy tolerance value of root mean square gradient 0.001 kcal/mol and the maximum number of iterations set to 1000. Conformational search of each energy-minimized structure was performed using a stochastic approach. Stochastic conformational search method is similar to the RIPS method,[13] which generates new molecular conformation by randomly perturbing the position of each coordinate of each atom in the molecule followed by energy minimization. All conformers generated for each structure were analyzed in conformational geometry panel with great care and the lowest energy conformation of each structure was selected and added to a molecular database to compute various physicochemical properties from three classes: 2D-descriptors based on atoms and connection information of the molecules, i3D-descriptors used three-dimensional coordinate information about each molecule, which are invariant to rotations and translations of the conformation, and x3D descriptors, which supported by three-dimensional coordinate information, require an absolute frame of reference using a QuaSAR module.[14–19]

Series was divided into a training set of 18 compounds and a test set of 5 compounds on the basis of structural diversity and cover the complete range of variations in inhibitory activity. The data were transferred to the statistical program in order to establish a correlation between physicochemical parameters as independent variables and aldose reductase inhibitory activity as a dependent variable. The sequential multiple linear regression analysis method was employed. In sequential multiple regression, the program searches for all permutations and combinations sequentially for the data set. The ± data within parentheses are the standard deviation, associated with the coefficient of descriptors in regression equations. The best model is selected from the various statistically significant equations on the basis of the observed squared correlation coefficient (r^2), the standard error of the estimate (SE), sequential Fischer test (F), the bootstrapping squared correlation coefficient (r^2_{bs}), the bootstrapping standard deviation (S_{bs}), the cross-validated squared correlation coefficient using leave-one-out (LOO) procedure (q^2), chance statistics (evaluated as the ratio of the equivalent regression equations to the total number of randomized sets; a chance value of 0.001 corresponds to 0.1% chance of fortuitous correlation), outliers (on the basis of Z-score value), and the predictive squared correlation coefficient of test set (r^2_{pred}).

Fujita-Ban analysis was carried out in order to find out the de novo contribution of the substituents to the activity of the molecules. The multivariant regression expression Eq. 1 indicated that few substituents have poor contribution to the ALR2 inhibitory activity, which is further supported by a high standard error of the substituent coefficient. Values of the Fujita-Ban matrix and calculated activities of 5-arylidene-2,4-thiazolidinediones are given in Table 2.

$$\begin{aligned} p\mathrm{IC}_{50} &= 0.056(\pm0.198)[3\mathrm{CH_3}] + 0.642(\pm0.198) \\ &\quad \times [3\mathrm{OC_6H_5}] + 0.066(\pm0.198)[3\mathrm{OCH_3}] \\ &\quad - 0.096(\pm0.198)[3\mathrm{CF_3}] + 0.889(\pm0.305) \\ &\quad \times [3-\mathrm{CH{=}N{-}C_6H_4{-}OH}] \\ &\quad - 0.453(\pm0.305)[4\mathrm{OCH_3}] \\ &\quad + 0.084(\pm0.198)[4\mathrm{CF_3}] - 0.202(\pm0.305) \\ &\quad \times [4-\mathrm{CH{=}N{-}C_6H_4{-}OH}] \\ &\quad + 0.170(\pm0.151)[\mathrm{CH_2COOCH_3}] \\ &\quad + 1.355(\pm0.144)[\mathrm{CH_2COOH}] + 4.842 \\ &\quad n = 23,\ r = 0.960,\ r_2 = 0.922, \\ &\quad \mathrm{SE} = 0.270,\ F = 14.1838 \end{aligned} \tag{1}$$

Contributions of substituents to the model are:

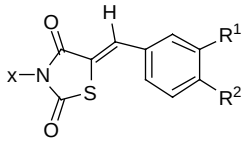
$[3\mathrm{CH_3}]$:$[3\mathrm{OC_6H_5}]$:$[3\mathrm{OCH_3}]$:$[3\mathrm{CF_3}]$:[3-CH=N–C6H4–OH]:$[4\mathrm{OCH_3}]$:$[4\mathrm{CF_3}]$:[4-CH=N–$\mathrm{C_6H_4}$–OH]:$[\mathrm{CH_2COOCH_3}]$:$[\mathrm{CH_2COOH}]$ :: [1]:[11.5586]:[1.18044]:[1.72786]:[5.3317]:[2.7172]:[1.50969]:[1.21373]:[6.12105]:[56.9202]

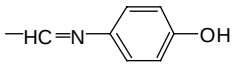
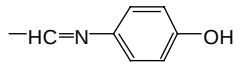
On refinement of the de novo contribution using stepwise multiple linear regression, significant tri-variant regression expression Eq. 2 was obtained. Tri-variant expression accounts for more than 87% variance in activity with an internal statistical significance level better than 99.9% as it exceeded the tabulated $F_{(3,19\alpha0.001)} = 9.42$ against calculated $F_{(3,19)} = 45.271$.

$$\begin{aligned} p\mathrm{IC}_{50} &= 0.672(\pm0.167)[3-\mathrm{OC_6H_5}] \\ &\quad + 0.854(\pm0.279)[3-\mathrm{CH{=}N{-}C_6H_4{-}OH}] \\ &\quad + 1.332(\pm0.123)[\mathrm{CH_2COOH}] + 4.877 \\ &\quad n = 23,\ r = 0.937,\ r^2 = 0.877, \\ &\quad \mathrm{SE} = 0.269,\ F = 45.271 \end{aligned} \tag{2}$$

Fujita-Ban analysis of the ALR2 inhibitory data of 5-arylidene-2,4-thiazolidinediones inferred that 3rd position of phenyl ring is more susceptible as com-

Table 2. Fujita-Ban matrix and calculated activities of 5-arylidene-2,4-thiazolidinediones



S. No.	μ	R^1					R^2			X		Cal. pIC_{50} *
		$-CH_3$	$-OC_6H_5$	$-OCH_3$	$-CF_3$		$-OCH_3$	$-CF_3$		$-CH_2COOCH_3$	$-CH_2COOH$	
1	1	0	0	0	0	0	0	0	0	0	0	4.877
2	1	1	0	0	0	0	0	0	0	0	0	4.877
3	1	0	0	1	0	0	0	0	0	0	0	4.877
4	1	0	0	0	1	0	0	0	0	0	0	4.877
5	1	0	0	0	0	1	0	0	0	0	0	5.730
6	1	0	0	0	0	0	0	0	0	0	0	4.877
7	1	0	0	0	0	0	0	1	0	0	0	4.877
8	1	0	0	0	0	0	0	0	1	0	0	4.877
9	1	1	0	0	0	0	0	0	0	1	0	4.877
10	1	0	0	1	0	0	0	0	0	1	0	4.877
11	1	0	0	0	1	0	0	0	0	1	0	4.877
12	1	0	0	0	0	0	0	1	0	1	0	4.877
13	1	0	0	0	0	0	0	0	0	0	1	6.208
14	1	1	0	0	0	0	0	0	0	0	1	6.208
15	1	0	1	0	0	0	0	0	0	0	1	6.880
16	1	0	0	1	0	0	0	0	0	0	1	6.208
17	1	0	0	0	0	0	0	0	0	0	1	6.208
18	1	0	0	0	0	0	0	1	0	0	1	6.208
19	1	0	1	0	0	0	0	0	0	0	0	5.549
20	1	0	0	0	0	0	1	0	0	0	0	4.877
21	1	0	1	0	0	0	0	0	0	1	0	5.549
22	1	0	0	0	0	0	0	0	0	1	0	4.877
23	1	0	0	0	1	0	0	0	0	0	1	6.208

* Calculated pIC_{50} values using the Fujita-Ban matrix

pared to 4th position of the ring for change in the activity. It is also suggested that substitution at the N-atom of thiazolidinedione with an acetic acid group is favorable to activity as compare to unsubstituted or methyl acetate substituted N-atom of thiazolidinedione. De novo contribution of groups also helped us to understand the binding of ALR2 inhibitors with aldose reductase by means of a possible hydrogen bond interaction between the acetic acid moiety of the side chain of thiazolidinediones and the polar region of the ALR2 active site. The 3rd position of phenyl ring depicted the hydrophobic interaction of the substituted moiety and the lipophilic pocket of ALR2.

The training set of 18 compounds was subjected to a stepwise multiple linear regression analysis, in order to develop a 2D-QSAR model between the inhibitory activity of aldose reductase inhibitors as dependent variables and substituent constants as independent variables, the several statistically significant equations were obtained Eqs. 3 and 4.

$$\begin{aligned} p\mathrm{IC}_{50} &= 1.300(\pm 0.168)\mathrm{HD}_2 + 0.015(\pm 0.007)\mathrm{MR}_1 \\ &\quad + 0.110(\pm 0.120)\pi + 4.815 \\ n &= 18,\ \ r = 0.913,\ \ r^2 = 0.833, \\ \mathrm{SE} &= 0.323,\ \ F = 23.303 \end{aligned} \tag{3}$$

$$\begin{aligned} p\mathrm{IC}_{50} &= 1.338(\pm 0.161)\mathrm{HD}_2 + 0.012(\pm 0.007)\mathrm{MR}_1 \\ &\quad + 4.862 \\ n &= 18,\ \ r = 0.907,\ \ r^2 = 0.823, \\ \mathrm{SE} &= 0.322,\ \ F = 34.905 \end{aligned} \tag{4}$$

Eqs. 3 and 4 account for more than 82% of the variance in activity. The fitness of Eq. 4 is more as compared to the Eq. 3 as shown by the value of the Fischer sequential test. The Eq. 4 was considered as the best 2D-model for the data set. The model is used for the internal and external predictivity of the training and the test sets (Figs. 3, 4 and Tables 3, 4). The value of the LOO cross-validation squared correlation coefficient ($q^2 = 0.659$) and the predictive squared correlation coefficient ($r^2_{\mathrm{pred}} = 0.673$) suggested good predictability. Equation reveals that the hydrogen donor groups at the nitrogen atom of thiazolidinediones contributed positively and linearly to the activity, while molar refractivity at the phenyl ring of the nucleus plays a significant role in the hydrophobic interaction with the enzyme.

To explore the physicochemical requirements for aldose reductase inhibition in terms of three-dimensional (3D) characteristics, the series was subjected to QSAR analysis. Training set was used to explore the conformation and geometry related physicochemical properties that could be helpful in understanding the probable binding site of the drug with the enzyme. A correlation was established between physicochemical parameters and ALR2 inhibitory activity using the sequential multiple linear regression technique. Several statistically significant equations with a coefficient of correlation (r) $\geqslant 0.925$ were obtained, which accounts for more than 85% of the variance in activity.

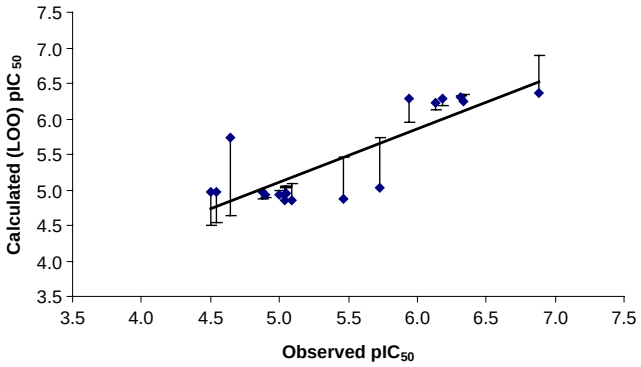


Figure 3. Plot between observed $p\mathrm{IC}_{50}$ and predicted (LOO) $p\mathrm{IC}_{50}$ with residual presentation for training set of 2D-QSAR analysis.

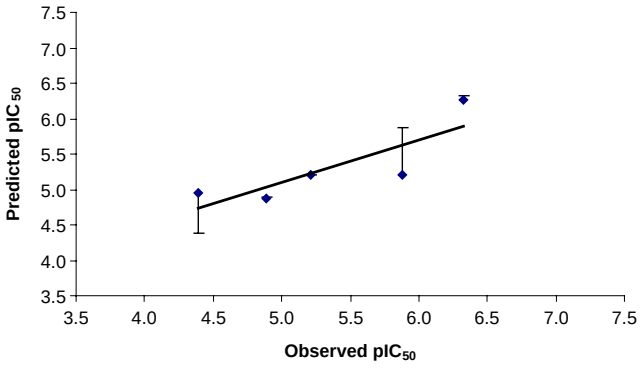


Figure 4. Plot between observed $p\mathrm{IC}_{50}$ and predicted $p\mathrm{IC}_{50}$ with residual presentation for test set of 2D-QSAR analysis.

$$\begin{aligned} p\mathrm{IC}_{50} &= 0.029(\pm 0.006)\mathrm{PEOE\_VSA} - 0 \\ &\quad + 0.454(\pm 0.083)\mathrm{a\_don} \\ &\quad - 0.412(\pm 0.047)\mathrm{E\_ang} + 14.511 \\ n &= 18,\ \ r = 0.955,\ \ r^2 = 0.912, \\ \mathrm{SE} &= 0.234,\ \ F = 48.639 \end{aligned} \tag{5}$$

$$\begin{aligned} p\mathrm{IC}_{50} &= 0.048(\pm 0.006)\mathrm{PEOE\_VSA} - 0 \\ &\quad - 0.502(\pm 0.050)\mathrm{E} - \mathrm{ang} \\ &\quad - 0.044(\pm 0.009)\mathrm{vsa\_acc} + 17.827 \\ n &= 18,\ \ r = 0.948,\ \ r^2 = 0.898, \\ \mathrm{SE} &= 0.253,\ \ F = 41.179 \end{aligned} \tag{6}$$

$$\begin{aligned} p\mathrm{IC}_{50} &= 0.086(\pm 0.008)\mathrm{vsa\_other} \\ &\quad - 0.146(\pm 0.034)\mathrm{E\_tor} \\ &\quad - 0.070(\pm 0.010)\mathrm{vsa\_acc} + 5.003 \\ n &= 18,\ \ r = 0.946,\ \ r^2 = 0.895, \\ \mathrm{SE} &= 0.256,\ \ F = 39.809 \end{aligned} \tag{7}$$

Table 3. Calculated and predicted $p\mathrm{IC}_{50}$ (by the LOO method) of training set, predicted value of the test set, and value of the substituent constant used in the 2D-QSAR model

Compound	MR_1	HD_2	Calculated $p\mathrm{IC}_{50}$	Residual	Z value	Predicted $p\mathrm{IC}_{50}$ (LOO)	Residual (LOO)
T-1	0.92	0	4.874	0.167	0.553	4.851	0.190
T-2	5.65	0	4.933	0.102	0.337	4.923	0.112
T-3	7.87	0	4.961	−0.084	−0.278	4.969	−0.092
T-4	5.02	0	4.925	−0.033	−0.108	4.929	−0.036
T-5	35.86	0	5.310	0.420	1.391	5.038	0.692
T-6	0.92	0	4.874	0.212	0.700	4.845	0.241
T-7	5.02	0	4.925	−0.423	−1.401	4.970	−0.468
T-8	35.86	0	5.310	−0.671	−2.219	5.744	−1.105
T-9	5.65	0	4.933	0.063	0.207	4.927	0.069
T-10	7.87	0	4.961	0.091	0.302	4.952	0.100
T-11	5.02	0	4.925	−0.383	−1.266	4.965	−0.423
T-12	5.02	0	4.925	0.538	1.781	4.869	0.595
T-13	0.92	1	6.212	−0.081	−0.269	6.231	−0.100
T-14	5.65	1	6.271	−0.084	−0.278	6.288	−0.101
T-15	27.68	1	6.546	0.340	1.125	6.366	0.520
T-16	7.87	1	6.299	0.020	0.066	6.295	0.024
T-17	0.92	1	6.212	−0.269	−0.890	6.275	−0.332
T-18	5.02	1	6.263	0.074	0.245	6.248	0.089
Test-1	27.68	0	—	—	—	5.208	0.004
Test-2	7.87	0	—	—	—	4.961	−0.572
Test-3	27.68	0	—	—	—	5.208	0.671
Test-4	0.92	0	—	—	—	4.874	0.015
Test-5	5.02	1	—	—	—	6.263	0.065

Table 4. QSAR statistics of significant equations

Equation No.	r^2	SE	F	ICAP[a] (Up to)	r^2_{bs}	s_{bs}	Chance	q^2	S_{PRESS}	S_{DEP}	r^2_{pred}	Outlier
3	0.913	0.323	23.303	0.315	FPE[b]	FPE[b]	0.001	0.572	0.517	0.457	0.776	Nil
4	0.823	0.322	34.905	0.090	0.831	0.103	0.001	0.659	0.447	0.408	0.673	Nil
5	0.912	0.234	48.639	0.590	0.917	0.055	0.001	0.843	0.314	0.277	0.728	Nil
6	0.898	0.253	41.179	0.590	0.902	0.070	0.001	0.839	0.317	0.280	0.243	Nil
7	0.895	0.256	39.809	0.501	0.912	0.068	0.001	0.818	0.338	0.298	0.709	1
8	0.860	0.296	28.731	0.433	0.869	0.091	0.001	0.710	0.426	0.376	0.870	Nil
9	0.857	0.300	27.899	0.221	0.877	0.064	0.001	0.688	0.442	0.390	0.659	Nil

[a] The maximum limit of inter-correlation among the parameters used in the generation of equations.
[b] FPE, floating point error.

$$\begin{aligned} p\mathrm{IC}_{50} &= 0.243(\pm 0.082)\mathrm{chi1\_C} \\ &\quad - 0.343(\pm 0.052)\mathrm{E\_ang} \\ &\quad + 0.544(\pm 0.097)\mathrm{a\_don} + 13.138 \\ n &= 18,\ \ r = 0.928,\ \ r^2 = 0.860, \\ \mathrm{SE} &= 0.296,\ \ F = 28.731 \end{aligned} \tag{8}$$

$$\begin{aligned} p\mathrm{IC}_{50} &= 0.546(\pm 0.094)\mathrm{a\_don} \\ &\quad + 0.045(\pm 0.008)\mathrm{vsa\_other} \\ &\quad - 0.064(\pm 0.015)\mathrm{E} + 5.613 \\ n &= 18,\ \ r = 0.926,\ \ r^2 = 0.857, \\ \mathrm{SE} &= 0.300,\ \ F = 27.899 \end{aligned} \tag{9}$$

A high correlation coefficient alone is not enough to select the equation as a model and hence various statistical approaches were employed to confirm the robustness and the practical applicability of the equations. The Eqs. 5–9 were tested for the presence of outliers. In case of Eq. 7, one outlier was present which suggested that although the equation has a good correlation coefficient it is unable to explain the deviation of the prediction of activity of a compound which was involved in the generation of expression. Eqs. 5, 6, 8, and 9 on randomized biological activity test show that the probability of chance correlation was less than 0.1%. Bootstrapping technique was employed to confirm the contribution of physicochemical properties of the molecules to the activity whether equi-intense or of different rank. The value of the bootstrapping squared correlation coefficient and the bootstrapping standard deviation implies that the equations were proper representative of the group of analogs.

The internal consistency of the training set was confirmed by using LOO cross-validation method to ensure the robustness of the equations. Equations showed good internal consistency ($q^2 = 0.688$–0.843), which reduces the probability of coincidental correlation of the expression. Expressions that have significant internal consistency may not be applicable for the analogs, which were never used in the generation of correlation. Therefore, the predictive power of Eqs. 5–9 was further confirmed by a test set of five compounds. Eq. 6 shows

poor predictivity of the test set with a low r^2_{pred} value (0.243). Eqs. 5, 8, and 9 fulfill all the statistical criteria (Table 4) to consider these equations as a QSAR model. Eq. 5 has inter-correlation among the parameters used in the generation of Equations (ICAP) up to 0.590 suggesting an interdependency of the contributed physicochemical properties. Therefore, Eqs. 8 and 9 were considered as models 1 and 2, respectively.

Model-1 has a better correlation coefficient ($r = 0.928$), which accounts for more than 86.0% of the variance in the activity, also the inter-correlation among the parameters (Table 5) being less (<0.433). The multivariant model shows that the dependent variable can be predicted from a linear combination of the independent variables. The P value is less than 0.02 for each physicochemical parameter involved in model generation. The data showed an overall internal statistical significance level to be better than 99.9% as it exceeded the tabulated $F_{(3,14\alpha0.001)} = 11.3$. The model was further tested for outlier by the Z-score method and no compound was found to be an outliers (Table 6), which suggested that the model is able to explain the structurally diverse analogs of the series and is helpful in designing more potent compounds using physicochemical parameters. The LOO cross-validation method was employed for the prediction of the activity (Fig. 5 and Table 6), and a q^2 value (in the biological activity data of leave one compound) of 0.3 corresponds to a confidence limit greater than 95%, which minimizes the risk of finding a significant explanatory equation for the biological activity just by mere chance. The cross-validated squared correlation coefficient ($q^2 = 0.710$), predictive residual sum of square ($S_{PRESS} = 0.426$), and standard error of prediction (S_{DEP}=0.376) suggested a good internal consistency as well as predictive ability of the biological activity with low S_{DEP}. The r^2_{bs} is at par with the conventional squared correlation coefficient (r^2). Randomized biological activity test (Chance < 0.001) revealed that the results were not based on chance correlation. The robustness and wide applicability of the model were further explained by a significant r^2_{pred} value (0.870) of the test set data (Fig. 6 and Table 7). In general, the model fulfills the statistical validation criteria to a significant extent and to be a useful theoretical base for proposing more active compounds. In model-1, chi1_C and a_don contributed positively, while E_ang contributed negatively, to the observed variance in activity. a_don is a pharmacophoric

Table 5. Correlation matrix of descriptors used in the model

	chi1_C	a_don	E_ang	E	vsa_other
Chi1_C	1.000				
A_don	0.402	1.000			
E_ang	0.433	0.145	1.000		
E	0.746	0.202	0.819	1.000	
Vsa_other	0.531	0.220	0.371	0.185	1.000

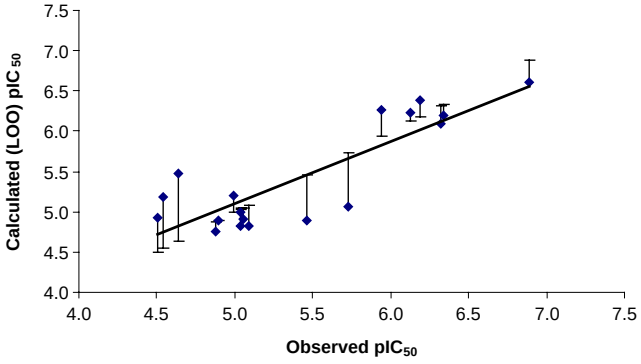


Figure 5. Plot between observed pIC_{50} and predicted (LOO) pIC_{50} with residual presentation for training set using model-1.

Table 6. Calculated and predicted pIC_{50} (by the LOO method) of training set with residual and Z-score value using models 1 and 2

Compound	Model-1					Model-2				
	Calculated pIC_{50}	Residual	Z value	Predicted pIC_{50} (LOO)	Residual (LOO)	Calculated pIC_{50}	Residual	Z value	Predicted pIC_{50}(LOO)	Residual (LOO)
T-1	4.861	0.180	0.670	4.831	0.210	4.909	0.132	0.487	4.883	0.158
T-2	4.999	0.036	0.134	4.996	0.039	4.896	0.139	0.510	4.870	0.165
T-3	4.777	0.099	0.370	4.758	0.119	5.012	−0.135	−0.498	5.024	−0.148
T-4	4.898	−0.006	−0.021	4.899	−0.006	4.731	0.161	0.593	4.699	0.194
T-5	5.325	0.406	1.510	5.067	0.663	5.410	0.320	1.178	5.296	0.435
T-6	4.861	0.225	0.838	4.823	0.263	4.908	0.177	0.652	4.874	0.212
T-7	4.868	−0.366	−1.363	4.918	−0.417	4.730	−0.228	−0.839	4.775	−0.274
T-8	5.071	−0.431	−1.606	5.473	−0.833	5.025	−0.385	−1.416	5.563	−0.924
T-9	5.147	−0.151	−0.562	5.202	−0.206	5.097	−0.102	−0.374	5.127	−0.132
T-10	4.939	0.113	0.422	4.910	0.142	5.210	−0.158	−0.581	5.290	−0.238
T-11	5.039	−0.496	−1.847	5.185	−0.642	4.936	−0.394	−1.447	5.043	−0.500
T-12	5.014	0.449	1.672	4.885	0.579	4.933	0.531	1.951	4.790	0.674
T-13	6.207	−0.076	−0.283	6.225	−0.094	6.290	−0.159	−0.584	6.330	−0.199
T-14	6.345	−0.158	−0.587	6.378	−0.191	6.275	−0.088	−0.324	6.297	−0.110
T-15	6.751	0.135	0.503	6.602	0.284	6.483	0.403	1.481	6.278	0.608
T-16	6.136	0.182	0.679	6.096	0.223	6.399	−0.081	−0.296	6.418	−0.099
T-17	6.204	−0.261	−0.971	6.266	−0.323	6.292	−0.349	−1.283	6.381	−0.438
T-18	6.219	0.118	0.441	6.195	0.142	6.121	0.216	0.793	6.086	0.251

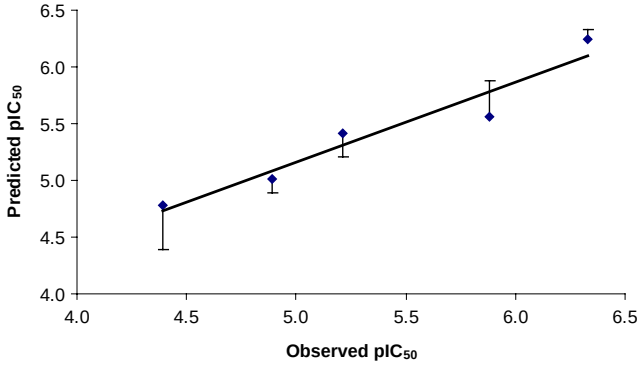


Figure 6. Plot between observed $p\mathrm{IC}_{50}$ and predicted $p\mathrm{IC}_{50}$ with residual presentation for test set using model-1.

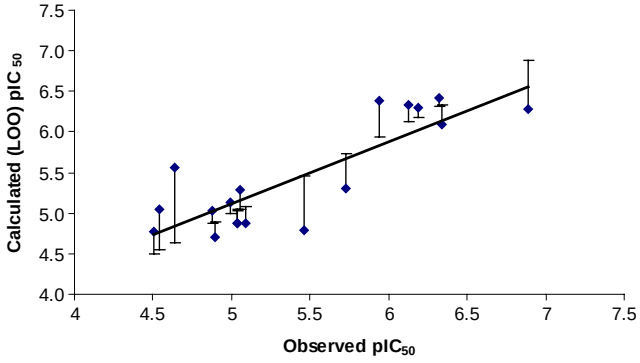


Figure 7. Plot between observed $p\mathrm{IC}_{50}$ and predicted (LOO) $p\mathrm{IC}_{50}$ with residual presentation for training set using model-2.

Table 7. Observed and predicted $p\mathrm{IC}_{50}$ of test set with residual using models 1 and 2

Compound	Observed $p\mathrm{IC}_{50}$	Model-1		Model-2	
		Predicted $p\mathrm{IC}_{50}$	Residual	Predicted $p\mathrm{IC}_{50}$	Residual
Test-1	5.212	5.411	−0.199	5.104	0.108
Test-2	4.389	4.778	−0.389	5.004	−0.615
Test-3	5.879	5.563	0.316	5.305	0.574
Test-4	4.889	5.009	−0.120	5.113	−0.224
Test-5	6.328	6.249	0.079	6.115	0.213

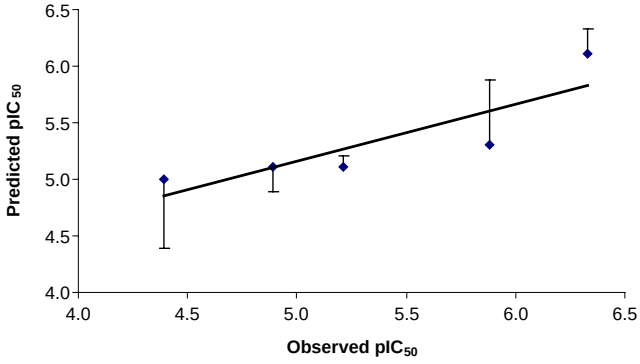


Figure 8. Plot between observed $p\mathrm{IC}_{50}$ and predicted $p\mathrm{IC}_{50}$ with residual presentation for test set using model-2.

feature descriptor which is representative of a number of hydrogen bond donor atoms, except for basic atoms, but including atoms that are both hydrogen bond donors and acceptors present in the molecules, which is helpful for rationalizing the interaction between molecule and enzyme. chi1_C is Kier and Hall connectivity index and related with the geometry of the molecule suggested that specific arrangement is required for the molecule-ALR2 interaction. E_ang, an energy descriptor and related with the conformation of the molecule suggested that the conformer, which deviated from the bend angle, plays a significant role in interaction. The study revealed that distal end of hydrogen donor groups at the nitrogen atom of thiazolidinediones may interact through hydrogen bond formation with the enzyme pocket.

Model-2 with correlation coefficient at par with model-1 ($r = 0.926$), and low inter-correlation among the parameters (<0.221) revealed that the physicochemical properties are quite independent as compared to those of model-1. Model-2 has somewhat lower internal consistency (Fig. 7 and Table 6) and predictive ability of test set ($r^2_{\mathrm{pred}} = 0.659$) as compared to those of model-1 (Fig. 8 and Table 7). The linear contribution of each physicochemical parameter to the model was significant by more than 99.9% ($P < 0.001$). Model-2 showed that a_don and vsa_other contributed positively, while E contributed negatively. a_don played significant role similar to model-1. vsa_other is a pharmacophoric feature descriptor explained as approximations to the sum of VDW surface areas of atoms typed as 'others.' E is the potential energy of the 3D conformal of the molecules dependent upon the internal coordinates of the molecule and depicted that the lowest energy conformal play a key role in the interaction of drug and enzymes.

Both models gave insight into some common important structural features. Distal end of hydrogen donor groups at the nitrogen atom of thiazolidinediones and the conformers, which deviated from bend angle energy, is favorable for the ALR2 inhibitory activity. The results indicate that the substitution on the nitrogen atom by a polar group could be able bind to the polar recognition region of ALR2. The presence of a carboxylic acid group is essential for the ALR2 inhibitory activity as well as the 3-position of phenylidene ring may be used for modifications to improve the lipophilic character which might result in a more potent ALR2 inhibitory activity.

Acknowledgments

Authors are grateful to the Director, Shri G. S. Institute of Technology and Science, Indore, for providing facilities for this work and Tata Elxsi, India, for providing the MOE software. The author AKG is grateful to CSIR, New Delhi, for providing the senior research fellowship.

References and notes

1. (a) Clarck, C. M., Jr.; Lee, D. A. N. *Engl. J. Med.* **1995**, *332*, 1210; (b) Nathan, D. M. N. *Engl. J. Med.* **1993**, *328*, 1676.
2. The Diabetic Complications and Control Trial Research Group. *N. Engl. J. Med.*, **1993**, *329*, 977.
3. United Kingdom Prospective Diabetes Group. *N. Engl. J. Med.*, **1998**, *352*, 837.;
4. (a) Kador, P. F.; Kinoshita, J. H.; Sharpless, N. J. *J. Med. Chem.* **1985**, *28*(7), 841; (b) Kador, P. F.; Sharpless, N. J. *Mol. Pharmacol.* **1983**, *24*, 521; (c) Kador, P. F. *Med. Res. Rev.* **1988**, *8*, 325.
5. (a) Sarges, R.; Oates, P. J. *Prog. Drug Res.* **1993**, *40*, 99; (b) Eggler, J. F.; Larson, E. R.; Lipinski, C. A.; Mylari, B. L.; Urban, F. *J. Adv. Med. Chem.* **1993**, *2*, 197.
6. (a) Kinoshita, J. H.; Hishimura, C. *Diabetes/Metab. Rev.* **1998**, *4*, 323; (b) Nicolucci, A.; Carinci, F.; Cavaliere, D.; Scorpiglione, N.; Belfiglio, M.; Labrrozi, A. *Diabetic Med.* **1996**, *13*, 1017.
7. (a) Costantino, L.; Rastelli, G.; Vianello, P.; Cignarella, G.; Barlocco, D. *Med. Res. Rev.* **1999**, *19*, 3; (b) Miyamoto, S. *Chem-Bio Inform. J.* **2002**, *2*(3), 74.
8. Maccari, R.; Ottana, R.; Curinga, C.; Vigorita, M. G.; Rakowitz, D.; Steindl, T.; Langer, T. *Bioorg. Med. Chem.* **2005**, *13*, 2809.
9. Bruno, G.; Curinga, C. C.; Maccari, R.; Monforte, F.; Nicolo, F.; Ottana, R.; Vigorita, M. G. *Bioorg. Med. Chem.* **2002**, *10*, 1077.
10. Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley: New York, 1979.
11. MOE User's Manual, Tata Elxsi Ltd, Bangalore.
12. Gupta, A. K.; Arockia, B. M.; Kaskhedikar, S. G. *Indian J. Pharm. Sci.* **2004**, *66*(4), 396.
13. Ferguson, D. M.; Raber, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 4371.
14. Hall, L. H.; Kier, L. B. *Rev. Comput. Chem.* **1991**, 2.
15. Hall, L. H.; Kier, L. B. *Eur. J. Med. Chem. 12—Chim. Ther.* **1977**, *4*, 307.
16. Wildman, S. A.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868.
17. Lin, A. QuaSAR-descriptor. *J. Chem. Compute*. Group <http://www.chemcomp.com/Journal_of_CCG/Fetures/descr.htm>.
18. Labute, P. A. *J. Mol. Graph. Model* **2000**, *18*, 464.
19. Baurin, N.; Mozziconzcci, J.-C.; Arnoult, E.; Chavette, P.; Marot, C.; Morin-Allory, L. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 276.