Chem Biol Drug Des 2009; **73:** 428–441

Research Article

doi: 10.1111/j.1747-0285.2009.00788.x

Quantitative Structure–Activity Relationship Models with Receptor-Dependent Descriptors for Predicting Peroxisome Proliferator-Activated Receptor Activities of Thiazolidinedione and Oxazolidinedione Derivatives

Viney Lather, Visvaldas Kairys and Miguel X. Fernandes*

Centro de Química da Madeira, Departamento de Química, Universidade da Madeira, Campus da Penteada, 9000-390 Funchal, Portugal
**Corresponding author: Miguel X. Fernandes, mxf@uma.pt*

A quantitative structure–activity relationship study has been carried out, in which the relationship between the peroxisome proliferator-activated receptor α and the peroxisome proliferator-activated receptor γ agonistic activities of thiazolidinedione and oxazolidinedione derivatives and quantitative descriptors, V_{site} calculated in a receptor-dependent manner is modeled. These descriptors quantify the volume occupied by the optimized ligands in regions that are either common or specific to the superimposed binding sites of the targets under consideration. The quantitative structure–activity relationship models were built by forward stepwise linear regression modeling for a training set of 27 compounds and validated for a test set of seven compounds, resulting in a squared correlation coefficient value of 0.90 for peroxisome proliferator-activated receptor α and of 0.89 for peroxisome proliferator-activated receptor γ. The leave-one-out cross-validation and test set predictability squared correlation coefficient values for these models were 0.85 and 0.62 for peroxisome proliferator-activated receptor α and 0.89 and 0.50 for peroxisome proliferator-activated receptor γ respectively. A dual peroxisome proliferator-activated receptor model has also been developed, and it indicates the structural features required for the design of ligands with dual peroxisome proliferator-activated receptor activity. These quantitative structure–activity relationship models show the importance of the descriptors here introduced in the prediction and interpretation of the compounds affinity and selectivity.

Key words: multiple linear regression, peroxisome proliferator-activated receptor, quantitative structure–activity relationship

Received 12 August 2008, revised 15 January 2009 and accepted for publication 21 January 2009

The study of relationships between molecular structure and properties, either physicochemical or biological, has attracted considerable attention and forms the basis of quantitative structure property/activity relationship (QSPR/QSAR) studies. During the past decades, thousands of molecular descriptors such as graph-theoretic, geometric, electrostatic, quantum-chemical, etc. have been extensively used for the prediction of physical, chemical, environmental and biological properties of molecules (1). Computation of these molecular descriptors for QSAR studies is generally carried out in a receptor-independent manner except in some cases, where the alignment of molecules is based either on a co-crystallized ligand or on the docked conformations for CoMFA and CoMSIA studies (2,3). A receptor-dependent QSAR approach has been incorporated in 4D-QSAR and 5D-QSAR approaches, and 4D-QSAR analysis has been used in the design of inhibitors of glycogen phosphorylase (4) and HIV-1 protease inhibitors (5). With the available panoply of descriptors, several QSAR equations could describe equally well the experimental data under study. Nonetheless, from a medicinal chemist's standpoint QSAR models should be easily interpreted and their predictions easily translated into synthesis of new compounds with improved properties. In this sense, descriptors and QSAR models with simple and intuitive interpretation are at premium.

An increasing realization in drug discovery is that modulating a multiplicity of targets can be an asset in the treatment of a range of disorders. Examples include: Omapatrilat, a dual angiotensin-converting enzyme and neutral endopeptidase inhibitor (6); SKI-606, a dual inhibitor of Src and Abl kinases (7); Netoglitazone, a peroxisome proliferator-activated receptor; PPARα and PPARγ agonist (8). In this study, we introduce novel molecular descriptors which incorporate receptor information, and are conceived to be used in cases where ligands are being designed to target several receptors. These quantitative descriptors, V_{site}, measure the volume occupied by a ligand within the common or specific regions defined by the superimposed binding sites of the targets under consideration, as receptors targeted by the same ligand should share some common features in their binding sites. Therefore, ligands that are active against several receptors should preferentially remain, and interact,

within the region that is common to the binding sites of all targeted receptors. We want to test if the selectivity of ligands for multiple receptors is correlated with the volume, they occupy within the common region defined by the binding sites. The descriptors that we are introducing are very simple and take into account only the steric nature of the binding sites to define common or specific regions, and sometimes other factors (chemical nature, flexibility of ligands, conformational changes of receptors, etc.) could play a more important role than these simple descriptors, we are introducing. The proposed descriptors were used along with other molecular descriptors, in the development of 2D-QSAR models for the prediction of PPARα, PPARγ and PPARα/γ dual agonistic activity of a dataset of thiazolidinedione and oxazolidinedione derivatives from the lead compound KRP-297 of Merck which showed PPARα/γ dual activity. This set of compounds was studied by Khanna *et al.* and the authors established a 3D-QSAR model based on the additivity of molecular fields, but the study was carried out in a receptor-independent manner (9).

Materials and Methods

Dataset for analysis

A dataset consisting of a series of 5-aryl thiazolidinedione and oxazolidinedione derivatives (10,11) acting as PPARα and PPARγ dual activators has been selected to develop three QSAR models: (i) α model, (ii) γ model and (iii) dual model. The basic structures of these compounds are shown in Table 1 along with their induced activities for PPARα and PPARγ (pIC_{50}). In the dual model development, pIC_{50} (dual) have been calculated as the summation of pIC_{50} values for PPARα and PPARγ. The dataset of 34 molecules was subdivided randomly into a training set and a test set of 27 and seven molecules, respectively (as selected by Khanna *et al.* for the CoMFA studies). Two molecules which were excluded from Khanna's study because of not-fitting into the model were included in this study as part of the test set molecules. Khanna's report includes compounds for which IC_{50} values of PPARα activity could not be specifically ascertained, but a minimum range was defined. For such molecules, the reported minimum value of activity was employed in the development of QSAR models.

Computational details

Active/inactive classification

The new descriptors here introduced (see below) will have a zero value for those compounds that are inactive towards a certain target. We are assuming that inactive compounds will not remain in the target's binding site and therefore they do not occupy any volume in that binding site. This does not pose any problem for compounds with known activities, where we calculate the descriptors only for the active compounds. However, in a context of quantitative prediction of activities of unknown compounds, we will need an *a priori* qualitative classification of these compounds, either as active or inactive, before the calculation of the descriptors. To determine if an unknown compound will be considered as active or inactive for a particular receptor isoform, and therefore if the descriptors here introduced need to be calculated or not, a qualitative approach for classifying the molecules from the dataset has been implemented. The relationship of the *Zagreb index* – an adjacency based topological descriptor (12,13) and the *Balaban-type index from polarizability weighted distance matrix,* J_p – a distance-based topological descriptor (14,15) with the PPARα and the PPARγ agonistic activities was investigated by the development of suitable models. The Zagreb index M_1 proposed by Gutman *et al.* (12,13) is defined as the sum of squares of degree over all vertices and is represented by the following equation:

$$M_1 = \sum_{i=1}^{n} (V_i^2) \tag{1}$$

where V_i is the degree of vertex i in a hydrogen-suppressed molecular structure. The vertex degree V_i for a vertex i is given as the sum of the entries in a row i of the adjacency matrix.

The Balaban index (J) proposed by Balaban (14) is calculated using the following formula:

$$J = \frac{B}{C+1} \sum_{k=1}^{B} \sqrt{(v_i v_j)_k} \tag{2}$$

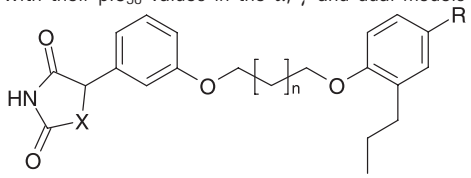
where v_i and v_j are the vertex distance degrees of two atoms connected by bond k, B is the number of bonds of the molecule and C is the cyclomatic number. The Balaban-type index, J_p is obtained by weighting the contributions of atoms and bonds with polarizability (15). The topological indices M_1 and J_p were calculated using E-Dragon (VCCLAB, Virtual Computational Chemistry Laboratory, 2005).

Values of the Zagreb index were computed for each compound. For the selection and evaluation of range-specific features, exclusive activity ranges were discovered from the frequency distribution of response level, and subsequently identifying the active range by analyzing the resultant data by maximization of the moving average with respect to the active compounds (16). Subsequently, each compound in the dataset was classified using this model, and its classification was compared with the experimental values of PPARα and PPARγ activities (IC_{50}). Compounds possessing IC_{50} values ≤ 2 μM were considered as active and analogs possessing IC_{50} values >2 μM were considered as inactive in this study. The percentage of successful prediction of active and inactive range was calculated from the ratio of the number of compounds predicted correctly to the total number of compounds present in that range. The overall percentage of successful prediction was calculated from the ratio of the total number of compounds predicted correctly to that of the total number of compounds present in both the active and inactive ranges. The aforementioned procedure was similarly adopted for the *Balaban type index*. As no compound in the dataset possessed PPARγ IC_{50} value >2 μM, all the compounds were considered as active and the calculation of volumes within the PPARγ binding site was performed.

Optimization of ligands and calculation of V_{site}

The molecular structures of the compounds were built with HYPERCHEM (Hypercube Inc, Gainsville, FL, USA) using the crystal structure

Table 1: Dataset used for QSAR analysis with their pIC_{50} values in the α, γ and dual models



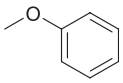
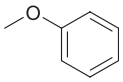
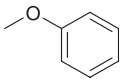
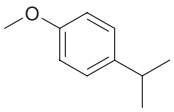
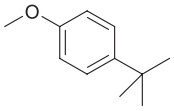
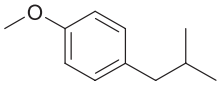
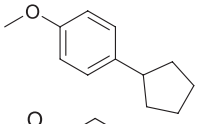
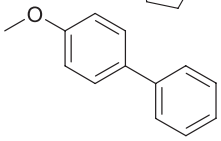
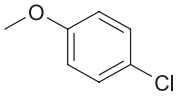
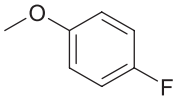
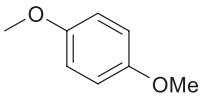
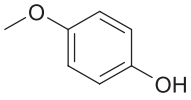
Comp. No.	X	R	M_1	J_p	Exp. PPARα activity	Pred. PPARα activity		Exp. pIC_{50} (α)	Pred. pIC_{50} (α)	Exp. pIC_{50} (γ)	Pred. pIC_{50} (γ)	Exp. pIC_{50} (dual)	Pred. pIC_{50} (dual)
						M_1	J_p						
1	S		172	0.861	+	+	+	7.55	7.25	7.24	7.29	14.80	14.20
2[a,b]	S		158	0.800	+	+	+	7.33	7.04	7.12	7.77	14.45	16.37
3[c]	S		188	0.870	–	–	–	5.30	5.53	6.71	7.01	12.01	13.34
4	S		196	0.877	–	–	–	5.68	5.11	6.77	6.80	12.45	12.22
5	S		192	0.870	–	–	–	5.0	5.09	6.47	6.39	11.47	11.29
6	S		202	0.763	–	–	–	5.0	5.22	6.54	6.64	11.54	10.91
7[a]	S		206	0.774	–	–	–	5.70	5.40	6.48	6.50	12.18	12.99
8	S		212	0.780	–	–	–	5.0	5.04	6.65	6.54	11.65	12.20
9[a]	S		212	0.774	+	–	–	7.0	7.25	7.14	7.10	14.14	14.47
10	S		216	0.775	+	+	–	7.55	7.18	7.11	7.17	14.66	14.14
11	S		212	0.777	–	–	–	5.59	5.78	6.52	6.98	12.11	12.26
12	S		218	0.788	+	+	+	6.02	6.12	7.52	7.59	13.54	13.52

Table 1: (Continued)

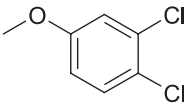
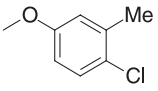
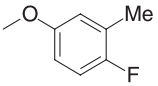
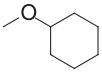
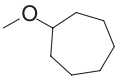
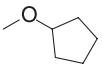
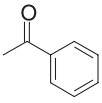
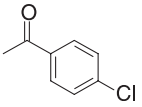
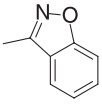
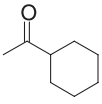
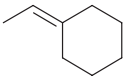
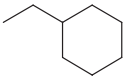
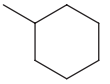
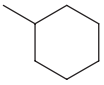
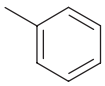
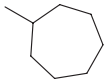
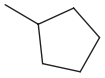
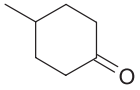
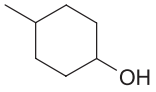
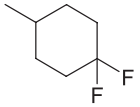
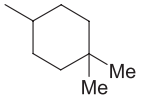
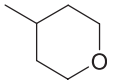
Comp. No.	X	R	M_1	J_p	Exp. PPARα activity	Pred. PPARα activity		Exp. pIC_{50} (α)	Pred. pIC_{50} (α)	Exp. pIC_{50} (γ)	Pred. pIC_{50} (γ)	Exp. pIC_{50} (dual)	Pred. pIC_{50} (dual)
						M_1	J_p						
13[a]	S		218	0.788	+	+	+	7.17	7.09	7.19	6.94	14.36	15.54
14	S		218	0.782	+	+	+	6.79	6.78	7.25	6.94	14.04	14.46
15	S		172	0.849	+	+	+	6.95	6.73	7.11	7.00	14.06	13.80
16	S		176	0.852	+	+	+	5.85	6.66	7.43	7.21	13.28	13.86
17	S		168	0.850	+	+	+	6.68	6.50	7.07	7.02	13.75	12.98
18	S		178	0.949	+	+	+	7.38	7.16	7.20	7.28	14.58	14.45
19	S		184	0.961	+	+	+	7.05	7.30	7.72	7.72	14.77	13.92
20[a]	S		190	0.850	–	–	+	5.0	6.20	6.84	6.91	11.84	14.23
21	S		178	0.935	+	+	+	7.43	7.30	7.52	7.50	14.95	14.89
22	S		172	0.927	+	+	+	6.92	6.80	7.19	7.00	14.11	13.85
23	S		172	0.913	+	+	+	7.21	7.29	6.82	6.75	14.03	14.52
24	S		168	0.917	+	+	+	7.34	7.14	7.23	7.17	14.57	14.73
25[a]	S		168	0.882	–	+	–	5.30	5.87	6.47	7.16	11.77	N.A.
26	O		168	0.896	+	+	+	6.54	6.70	6.31	6.23	12.85	12.55

Table 1: (Continued)

Comp. No.	X	R	M_1	J_p	Exp. PPARα activity	Pred. PPARα activity		Exp. pIC_{50} (α)	Pred. pIC_{50} (α)	Exp. pIC_{50} (γ)	Pred. pIC_{50} (γ)	Exp. pIC_{50} (dual)	Pred. pIC_{50} (dual)
						M_1	J_p						
27	O		172	0.893	–	+	+	5.58	5.65	6.35	6.26	11.92	12.34
28	O		164	0.873	+	+	–	6.21	6.58	6.19	6.13	12.40	11.93
29[a]	O		174	0.890	+	+	+	6.20	7.04	5.97	6.36	12.17	13.30
30	O		174	0.886	–	+	–	4.82	5.24	5.94	6.20	10.76	11.33
31	O		182	0.886	–	+	–	4.82	4.46	6.74	6.52	11.56	11.88
32	O		182	0.901	+	+	+	6.23	6.23	6.27	6.07	12.50	12.68
33	O		168	0.876	–	+	–	5.48	5.34	5.75	6.02	11.23	11.71
34	O		176	0.848	–	+	+	5.54	5.30	6.12	6.25	11.66	11.34

[a]Test set molecules; [b]without propyl group; [c]n = 2.

of rosiglitazone with PPARγ [PDB:2PRG (17)] as the basic skeleton. All compounds were then optimized individually using semi-empirical quantum mechanics (PM3 method with restricted Hartree-Fock) to get the correct bond lengths and angles. The compounds with correct geometries were docked against the PPAR isoforms using VDock (18). If for a given receptor, a particular compound was classified as inactive, then that compound would not be docked against the receptor.

The volumes of the compounds within a 4 Å threshold from receptor atoms were calculated using a simple in-house grid-based algorithm programmed in FORTRAN[a]. The receptor atom co-ordinates were read from a specially prepared PDB file containing the two superimposed PPAR variants (PPARγ; PDB:2PRG and PPARα; PDB:1I7G). The structural alignment of the two proteins was carried out based upon the co-ordinates of the protein's Cα atoms using the SHEBA program (19). For each docked ligand, a relatively small virtual 3D grid with a 0.2 Å spacing was created, completely enveloping the entire van der Waals volume of the ligand molecule. The grid points that fall outside of the ligand atom's van der Waals radii were ignored. For each grid point inside the ligand's van der Waals radii, the test of the proximity to the receptor atoms was performed. The program classified each grid point as belonging to one of the following categories: (i) within 4 Å from both PPARα and γ receptors; (ii) within 4 Å from PPARα but outside of the 4 Å zone of PPARγ (PPARα only); (iii) within 4 Å from PPARγ but outside of the 4 Å zone of PPARα (PPARγ only); and (iv) outside of the 4 Å zone of both PPARα and PPARγ receptors. Figure 1 illustrates schematically the calculated volumes used in this study. The numbers of grid points falling within each category were further multiplied by the volume of a grid cell, i.e. 0.008 Å^3, and produced the numerical values of V_{common}, V_α, V_γ and V_{out}, corresponding to the four categories mentioned above. Only the heavy (non-hydrogen) atoms of the ligand and of the receptor were taken into account by the program. In cases where compounds presented dual activity, V_α and V_γ were calculated using the compounds' docked conformation in PPARα and PPARγ, respectively. V_{common} and V_{out} of a dual active compound were calculated as averages for both conformations of the compound docked separately against the PPARα- and the

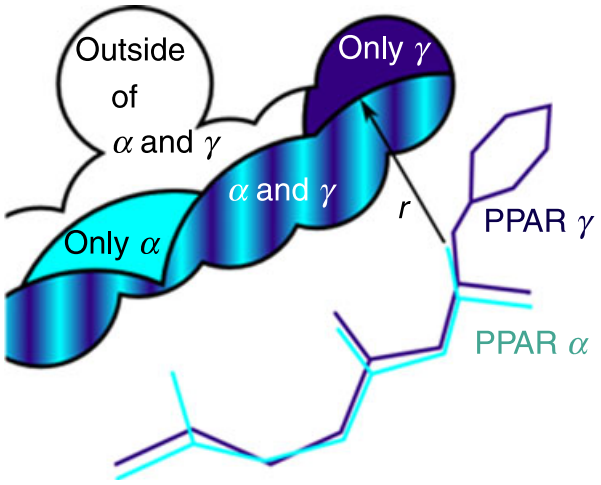


Figure 1: Schematic illustration of volumes used in this paper. The relevant volumes are computed by superposing two PPAR isoforms (thin sticks), creating van der Waals spheres around the ligand atoms (the upper half of the picture), and calculating volumes of the ligand within the $r = 4$ Å threshold from the protein atoms of either isoform. The volumes obtained can be classified into four categories: part of the ligand far from atoms belonging to either PPARα or PPARγ isoforms (uncolored field), close to PPARα but far from PPARγ (light blue), close to PPARγ only (dark blue) and close to both PPARα and PPARγ (stripes).

PPARγ binding sites. If a compound was considered inactive towards PPARα, then it was not docked in the receptor's binding site and its V_α value was set to zero. For a compound inactive towards PPARα, V_{common} and V_{out} were calculated using only the compound's docked conformation against the PPARα binding site, and the same happened obviously for the calculation of V_γ. The values of V_α, V_γ, V_{common}, V_{out} and the molecular volume for some compounds used in this study are shown in Table 2.

Choice of descriptors

The molecular structures were extracted from the proteins after docking and transferred to a database for the calculation of other molecular descriptors including constitutional, topological, geometrical, electrostatic and quantum descriptors using CODESSA™ (Comprehensive Descriptors for Structural and Statistical Analysis) program[b] (20). A total of 527 descriptors were calculated for all the structures. Selection of descriptors was performed in a heuristic manner as employed in the framework of CODESSA which reduces the pool of descriptors by eliminating those that satisfy the following conditions (21): (i) descriptors that were not available for every structure; (ii) the descriptor has a constant value for all the investigated compounds; (iii) the descriptors with a correlation coefficient smaller than 0.3 with the dependent variable (pIC_{50}) were regarded as redundant; (iv) in the monoparametric correlation with (pIC_{50}), the descriptor has a squared correlation coefficient lower than 0.1; (v) in the monoparametric correlation the descriptor has a *t*-test value lower than 0.1; (vi) in the monoparametric correlation the descriptor has an *F*-test value lower than 1 at a probability level of 0.05; (vii) highly correlated descriptors provide approximately identical information, if their pair-wise correlation coefficient exceeded 0.75.

Table 2: Numerical values of V_α, V_γ, V_{common}, V_{out} and molecular volume of representative thiazolidinedione and oxazolidinedione derivatives

Comp. No.	V_α	V_γ	V_{common}	V_{out}	Molecular volume
1	17.94	5.86	333.36	1.40	368.30
4	0	5.66	361.54	2.73	411.30
10	18.71	5.87	342.62	1.33	372.30
13	18.52	6.03	363.54	1.38	399.70
15	27.70	5.87	339.72	3.65	386.60
18	23.43	6.46	332.90	1.34	371.50
23	22.40	5.66	344.72	1.29	387.10
25	0	7.35	330.90	1.46	378.50
28	20.57	6.95	331.80	1.56	381.40
32	21.42	7.34	335.85	1.54	375.10

QSAR models generation

A heuristic approach was used for the initial selection of molecular descriptors followed by forward feature selection with multiple linear regressions (MLRs) to establish the final QSAR models. The final QSAR models were built using STATISTICA 7.0 (Statsoft Inc, Tulsa, OK, USA). Using the Fisher test (*F*-value) for the analysis of variance, squared correlation coefficient (R^2) and standard error of estimate (SEE), of training set as criteria of selection, subsets of descriptors were examined for establishing the best linear QSAR. The size of the descriptors subset used for model establishment was increased until no improvement was seen as well as keeping in view that the number of compounds in the training set should not be smaller than five times the number of descriptors. Variance–covariance matrices were calculated for each of the descriptors in all of the resulting linear models and the descriptors which had multicollinearity, were discarded. Tolerance and variation inflation factor (VIF) were chosen as the parameters for determining the collinearity among the variables. VIF values between 1 and 10 are acceptable for indicating the non-collinearity. Among the remaining models after the elimination process, the one that had the minimum RMSE was chosen as the best. The goodness of the regression fits were estimated using parameters, such as R^2, RMSE and *F*-statistics. To assess the self-consistency, the final model was validated using leave-one-out (LOO) and the predictive ability was checked using cross-validated squared correlation coefficient (q^2). After model development with a randomly selected set of training compounds, the best model was further examined by the test set molecules.

Results and Discussion

Active/inactive classification

The relationship of the *Zagreb index* and the *Balaban-type index* with the PPARα activity of compounds in the dataset was investigated and suitable models were developed to provide the behavior (active or inactive) of an unknown compound (not included in the dataset), through exploitation of the active ranges drawn from the

proposed models. Retrofit analysis of the data in Tables 3 and 4 reveals the following information concerning the different models developed in this study.

Classification based upon the *Zagreb index*:

• A total of 27 of 34 compounds were classified correctly in both the active and inactive ranges. The overall accuracy of prediction was found to be 80% with respect to PPARα agonistic activity.

• The inactive range traversed between the two active ranges and had a M_1 value of 188–212. Eight of nine analogs in the inactive range were predicted correctly resulting in a prediction of 89%. The average IC_{50} value for the inactive range was found to be 5.75 μM, indicating the presence of highly inactive compounds.

• The lower active range had a M_1 value of <188 and the upper active range had a M_1 value of >212. The lower active range had a predictability percentage of 72%, while the upper range resulted in 100% predictability. The average IC_{50} value for the upper active range was found to be 0.30 μM, indicating the presence of highly active compounds.

Classification based upon the *Balaban-type index*:

• A total of 28 of 34 compounds were classified correctly in both the active and inactive ranges. The overall accuracy of prediction was found to be 82.4% with respect to PPARα agonistic activity.

• Two active ranges, i.e. lower and upper active range were identified. The lower active traversed between the two inactive ranges and had a J_p value of 0.782–0.861. Eight of 10 analogs were predicted correctly resulting in a prediction of 80%. The average IC_{50} value for the lower active range was found to be 1.58 μM, indicating the presence of more active compounds.

• The upper active range had J_p values of >0.886–0.961. Nine of 10 analogs in the active range exhibited the PPARα agonistic activity. The average IC_{50} value for the upper active range was found to be 0.45 μM, indicating the presence of highly active compounds.

• The inactive ranges had J_p values of <0.782 and >0.861–0.886. The average IC_{50} values of the inactive ranges were found to be 4.11 and 7.00 μM, respectively, indicating the presence of highly inactive compounds compared with active ranges.

The developed classification models revealed significant correlations between the topological indices and PPARα agonistic activity of thiazolidinedione and oxazolidinedione derivatives. The model based upon the Balaban-type index; a distance-based topological descriptor resulted in a better model as compared with the model based on Zagreb index; an adjacency based topological descriptor. Thus, these models could be used to predict the behavior (active/inactive) of unknown compounds whose numerical activity will be predicted using the QSAR equations presented below.

QSAR models

Three independent QSAR models were developed using the volumes occupied by the ligands in the receptors' binding sites and other molecular descriptors calculated by CODESSA. The generated models include: (i) the PPARα model, developed using the ligands' induced activity of PPARα as the dependent variable, (ii) the PPARγ model, using the ligands' induced activity of PPARγ as the dependent variable, (iii) the PPAR dual model, using the multiplication of the activities induced by the ligands for both receptors. A linear regression with forward feature selection approach was carried out for the development of all three models. The volumes calculated for the molecules docked against the PPARα and the PPARγ binding sites were used in the QSAR models for the prediction of PPARα and PPARγ agonistic activities, respectively. As mentioned earlier, the

Table 3: Proposed model for PPARα activity based on Zagreb index

Model index	Nature of range in the proposed model	Index value	No. compounds falling in the range		Percent accuracy	Average IC_{50} (μM)
			Total	Correct		
M_1	Lower active	<188	21	15	72	2.29
	Inactive	188–212	09	08	89	5.75
	Upper active	>212	04	04	100	0.30

Table 4: Proposed model for PPARα activity based on Balaban-type index

Model index	Nature of range in the proposed model	Index value	Nature of compounds falling in the range		Percent accuracy	Average IC_{50} (μM)
			Total	Correct		
J_p	Inactive	<0.782	06	04	67	4.11
	Lower active	0.782–0.861	10	08	80	1.58
	Inactive	>0.861–0.886	08	07	87	7.00
	Upper active	>0.886–0.961	10	09	90	0.45

volumes, V_α, for molecules which were inactive for PPARα were set to zero.

PPARα model

The initial pool of descriptors calculated by CODESSA was reduced to a minimum of 21 in heuristic manner followed by a stepwise MLR model generation using forward feature selection. The best MLR model resulted in a 5-parameter equation:

pIC_{50} (α) = 0.054 (±0.006) V_α − 0.707 (±0.0165) Topograhic electronic index (all pairs) + 3.433 (±1.160) XY shadow − 0.348 (±0.127) HA-dependent HDCA-1 − 6.805(±3.096) ZX shadow + 12.899 (±2.042)

n = 27, R^2 = 0.903, $F(5,21)$ = 39.007, SEE = 0.322, q^2 = 0.849

The pentavariant model explained 90% of the variance in activity. The statistical details are shown in Table 5. All the *t*-values are significant with low p-values which confirm the significance of each descriptor. The *F*-statistics (on 5 and 21 degrees of freedom) for this model was found to be 39.0 with a p-value lower than 0.000. Further, the variance inflation factors are smaller than 1.6, which indicates the absence of multicollinearities in the model. Thus, the model is considered to be statistically valid. The predictive ability of the model was validated internally by LOO method resulting in a q^2 value of 0.85, which eliminates the possibility of chance correlation in the model and also indicates that the model is highly predictive. The RMSE for the training set (R^2 = 0.90) was 0.28 and the RMSE for the prediction/test set (R^2 = 0.62) was 0.60. Figure 2 shows a fit plot of predicted versus experimental pIC_{50} values for the training and test set molecules.

The main descriptor for the prediction of the PPARα agonistic activity as deduced from the *t*-statistics was found to be the V_α which is the volume occupied by the ligands in the region specific of PPARα's binding site. A positive coefficient value indicates that an increase in the volume or size of ligands in this PPARα-specific binding region increases the biological activity. Probably, this happens because an increase in the bulk leads to increased hydrophobic interactions (the region of the binding site accounting for V_α is lined with hydrophobic residues) thus enhancing biological activity.

The second descriptor of importance is the topographic electronic index (all pairs), which is derived from Zefirov's partial charges (22).

Table 5: Statistics for the best MLR model for prediction of PPARα agonistic activity

Descriptor	beta	SE	*t*	p level	Tolerance	VIF
Constant	12.899	2.042	6.315	0.000		
V_α	0.054	0.006	8.835	0.000	0.840	1.190
Topographic electronic index	−0.707	0.165	−4.263	0.000	0.845	1.183
XY shadow	3.433	1.160	2.958	0.007	0.886	1.128
HA dependent HDCA-1	−0.348	0.127	−2.740	0.010	0.836	1.196
ZX shadow	−6.805	3.096	−2.198	0.010	0.838	1.193

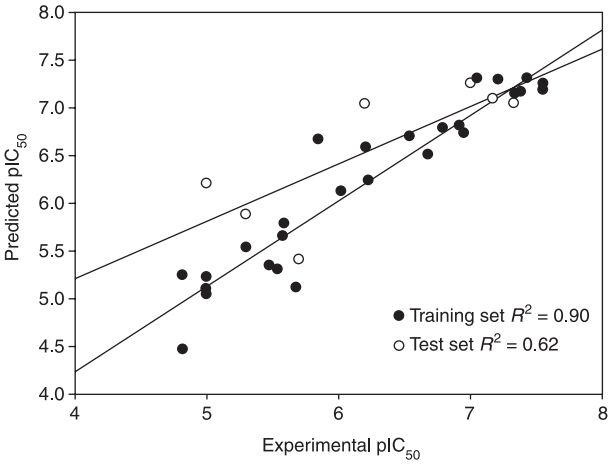


Figure 2: Plots of predicted versus experimental PPARα activities of the training and test set molecules.

This descriptor reflects the electrostatic interactions between the ligands and receptor molecules, and can characterize the charge distribution in the ligands. Negative contribution of this descriptor indicates the importance of other forces like hydrophobic interactions between ligands and PPARα.

The third descriptor is the *XY* shadow/*XY* rectangle which is a geometrical descriptor characterizing the shape and extent of the molecule in terms of its 3D co-ordinates (23). This descriptor represents a 2-dimensional projection on the *X-Y* plane of a 3-dimensional molecule. The orientation of a molecule along the axes of inertia (*X*-co-ordinate) casts a shadow of the molecule projected on the *X-Y* plane. Normalized shadows are calculated by *XY* shadow/*XY* rectangle. A positive coefficient shows that the activity increases with increase in the value of *XY* shadow, which means that a larger area of molecular shadow in the enclosing rectangle will benefit the activity.

The fourth descriptor, HA-dependent HDCA-1 is related to the hydrogen-acceptor(s) charged surface area of molecules (22). This descriptor is related to the hydrogen-bonding ability of the ligands and to the ability of forming polar interactions between the ligand and the receptor.

The fifth descriptor in the equation is a geometrical descriptor representing the 2-dimensional projection on the *Z-X* plane of a 3-dimensional molecule. A negative contribution to the linear equation indicates the decrease in activity with increase in the value of *ZX* shadow.

Overall, this reflects that the size/volume of molecules positively govern the PPARα activity, while the biological activity is influenced negatively by the hydrogen bond forming ability of thiazolidinedione and oxazolidinedione molecules. The positive contribution of size to the activity shows that hydrophobic interactions are playing a major role on the PPARα activity.

The importance of the V_α descriptor (which had a low beta coefficient value but a high *t*-statistics value) in the modeling of PPARα

activity was observed by the development of a QSAR model without V_α (while keeping everything else unchanged relatively to the model described at the beginning of this subsection). The QSAR model without V_α resulted in an R^2 value of 0.54 with SEE of 0.68 for the training set of compounds.

pIC_{50} (α) = −1.019 (±0.344) Topograhic electronic index (all pairs) + 6.435 (±2.732) *XY* shadow − 0.481 (±0.268) HA-dependent HDCA-1 − 9.649 (±6.534) *ZX* shadow + 15.881 (±4.275)

n = 27, R^2 = 0.541, F(4,22) = 6.49, SEE = 0.683, q^2 = 0.307

The model resulted in a low LOO cross-validation variance and *F*-test value of 6.49 (on 4 and 22 degrees of freedom). The predictability of this model was further tested on the test set molecules resulting in an R^2_{test} value of 0.10.

PPARγ model

The best MLR model for the prediction of PPARγ agonistic activity resulted in a 5-parameter equation explaining up to 89% variance. The initial pool of descriptors calculated by CODESSA was reduced to a minimum of 20 in a heuristic manner followed by a forward stepwise MLR model generation in combination with V_γ for PPARγ. The best model is shown as follows:

pIC_{50} (γ) = 63.312 (±8.397) HASA-2/TMSA + 63.068 (±13.110) Relative No. of *N* atoms − 0.481(±0.121) No. of double bonds − 0.001(±0.0002) Gravitation index (all bonds) − 0.115 (±0.0510) V_γ + 8.719 (±0.951)

n = 27, R^2 = 0.890, F (5,21) = 33.306, SEE = 0.195, q^2 = 0.80

The statistical details are shown in Table 6. All the *t*-values are significant with low p-values which confirm the significance of each descriptor. The *F*-statistic (on 5 and 21 degrees of freedom) for this model was found to be 33.3 with a p-value less than 0.000. The variance inflation factors were less than 2.5 indicating the absence of multicollinearities in the model. VIF values upto 10 are considered as statistically valid, but independent variables with values between 1 and 4 are considered as highly significant. Hence, the model is considered to be statistically valid. The predictive ability of the model was validated internally by the LOO method giving a q^2 value of 0.80, which indicates the high predictability of the model. The RMSE for the training set (R^2 = 0.89) was 0.17. The model was further validated externally by the prediction of test set compounds resulting in an R^2_{test} value of 0.50. Figure 3 shows a fit plot of predicted versus experimental pIC_{50} values for the training and test set molecules.

Table 6: Statistics for the best MLR model for prediction of PPARγ agonistic activity

Descriptor	beta	SE	*t*	p level	Tolerance	VIF
Constant	8.719	0.951	9.166	0.000		
HASA-2/TMSA	63.312	8.397	7.540	0.000	0.405	2.469
Relative No. *n* atoms	63.068	13.110	4.810	0.000	0.733	1.364
No. double bonds	−0.481	0.121	−3.973	0.000	0.638	1.567
Gravitation index (all bonds)	−0.001	0.0002	−3.127	0.005	0.738	1.355
V_γ	−0.115	0.051	−2.260	0.030	0.456	2.193

The main descriptor for the prediction based on the *t*-statistics was found to be HASA-2/TMSA (24), a Zefirov's partial charge-based descriptor, which is equal to the square root of hydrogen bonding acceptor surface area normalized by the total molecular solvent accessible surface area, which is related to the hydrogen bonding ability of the ligands with the receptor.

The second descriptor of importance for the PPARγ model is the relative number of *N* atoms and the positive coefficient indicates the increase in the activity of ligands with an increase in the relative no. of *N* atoms, which could reflect the importance of having more polar compounds for an increased activity.

The gravitation index (all bonds) (25) is defined as

$$G_b = \sum_{i<j}^{N_b} \frac{m_i m_j}{r_{ij}^2} \quad (3)$$

where m_i and m_j are the atomic masses of bonded atoms *i* and *j*, r_{ij} denotes the respective bond lengths and N_b is the number of chemical bonds in the molecule. G_b accounts for both the atomic masses (volumes) and for their distribution within the molecular space. It quantifies effectively the bulk cohesiveness of compounds arising from the dispersion and hydrophobic interactions. A negative coefficient for this descriptor implies that ligands with higher molecular mass have smaller PPARγ agonisitc activity.

The number of double bonds is a constitutional descriptor and negative coefficient in the PPARγ model shows that a decrease in the activity was observed with an increase in the number of double

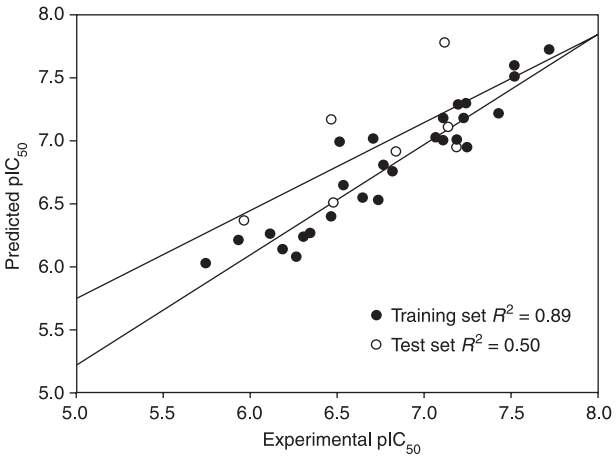


Figure 3: Plots of predicted versus experimental PPARγ activities of the training and test set molecules.

bonds. It could be due to the fact that the introduction of more double bonds reduced the flexibility of the molecules and thus decrease the activity.

The last descriptor for the prediction of the PPARγ agonistic activity was V_γ, which is the volume occupied by the ligands in the binding region specific to PPARγ. A negative coefficient value indicates that a decrease in the size or volume of ligands in this region increases the biological activity. Although the coefficient for V_γ in the equation is low, it had an individual correlation coefficient of 0.65 with PPARγ activity for the molecules in the training set, which shows that it influences the PPARγ activity significantly. The negative contribution to the PPARγ activity can be explained on the basis that as the size of the substituents at the terminal phenyl ring increases, the biological activity decreases. As it can be observed in Figure 4, the terminal phenyl ring fits in a small hydrophilic pocket (which is specific of the PPARγ isoform and accounts for the V_γ values), and molecules substituted with hydrophilic moieties (e.g. molecule 12 having hydroxyl substituted terminal phenyl ring) are more active. Molecules with larger and hydrophobic substituents (e.g. molecule 6 having isobutyl substituted terminal phenyl ring) make less favorable interactions with the receptor atoms and tend to be less active.

A QSAR model was also developed without V_γ (while keeping everything else unchanged relatively to the model described at the beginning of this subsection) for modeling the PPARγ activity. The QSAR model without V_γ resulted in an R^2 value of 0.86 with SEE of 0.21 for the training set of compounds.

pIC_{50} (γ) = 63.312 (±8.397) HASA-2/TMSA + 63.068 (±13.110) Relative No. of N atoms − 0.481(±0.121) No. of double bonds − 0.001(±0.0002) Gravitation index (all bonds)

n = 27, R^2 = 0.860, F(5,21) = 34.000, SEE = 0.21, q^2 = 0.50

This model resulted in a LOO cross-validation variance of 50%, which was lower than that of 5-parametric model obtained using V_γ (q^2 = 0.80). Further, predictability of this model was tested on the test set compounds resulting in an R^2_{test} value of 0.35, which was also lower than that of model with V_γ (R^2_{test} = 0.50). Thus, the comparative results of the two models validate the role of V_γ in the prediction of PPARγ activity.

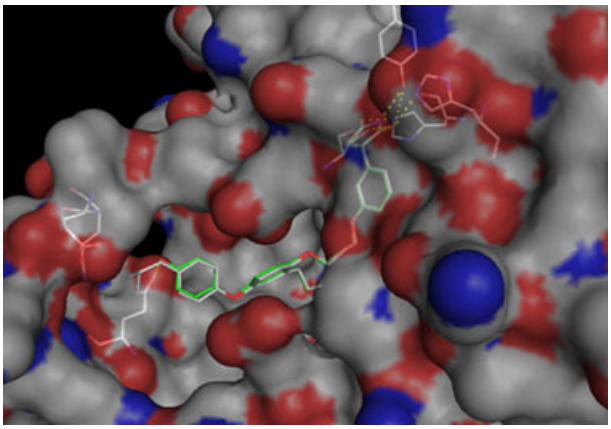


Figure 4: Overlay of molecules 6 (Grey) and 12 (Green) docked in binding site of PPARγ.

PPAR dual model

The MLR model for the prediction of PPAR dual activity resulted in a pentavariant equation explaining the 85% variance in the activity. The initial pool of descriptors calculated by CODESSA was reduced to a minimum of 21 in a heuristic manner followed by a forward stepwise MLR model generation. The product (sum of pIC_{50}) of the activities induced by the ligands for PPARα and for PPARγ was used as the dependent variable. The best model is shown as:

pIC_{50} (dual) = −0.544 (±0.081) No. of atoms + 2.405 (±0.374) Kier & Hall index (order 3) − 1.272(±0.352) No. of O atoms + 3.858 (±1.627) Topographic electronic index + 80.682 (±51.480) PNSA-3/TMSA + 32.390 (±3.508)

n = 27, R^2 = 0.850, F(5,21) = 24.08, SEE = 0.566, q^2 = 0.773

The statistical details of this model are shown in Table 7. All the t-values are significant with low p-values which confirm the significance of each descriptor. The F-statistics (on 5 and 21 degrees of freedom) for this model was found to be 24.08 with a p-value lower than 0.000. The variance inflation factors were smaller than 8.0 indicating the absence of multicollinearities in the model. Hence, the model is considered to be statistically valid. The predictive ability of the model was validated internally using the LOO method resulting in a q^2 value of 0.77, which indicates the high predictability as well as eliminating the chance correlation of the model. The RMSE for the training set (R^2 = 0.85) was 0.49. The model was further validated externally by prediction of the test set compounds activities resulting in an R^2 value of 0.3. However, upon removal of a single outlier (compound 25) from the test set using the Z-score method (2.5 times the SD), the R^2 value increased to 0.69. Figure 5 shows a fit plot of predicted versus experimental pIC_{50} values for the training and test set molecules.

The most important descriptor based on the t-statistics was the number of atoms, a constitutional descriptor that describes the size of the molecule. A negative coefficient value for the descriptor

Table 7: Statistics for the best MLR model for prediction of PPARdual agonistic activity

Descriptor	beta	SE	t	p level	Tolerance	VIF
Constant	32.390	3.508	9.232	0.000		
No. atoms	−0.544	0.081	−6.715	0.001	0.131	7.633
Kier & Hall index (order 3)	2.405	0.374	6.427	0.000	0.322	3.105
No. O atoms	−1.272	0.352	−3.610	0.000	0.387	2.583
Topographic electronic index	3.858	1.627	2.370	0.020	0.237	4.219
PNSA-3/TMSA	80.682	51.480	1.567	0.010	0.399	2.506

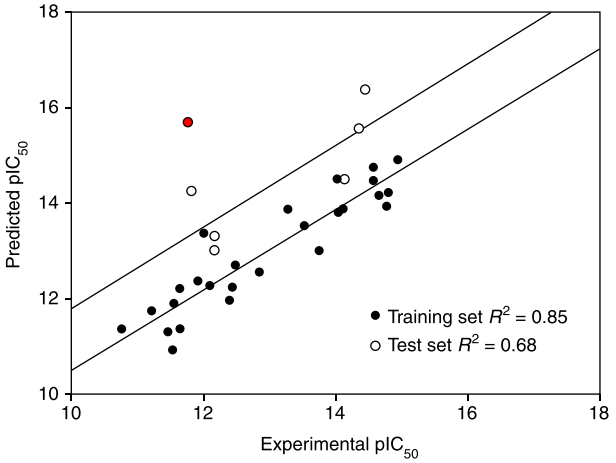


Figure 5: Plots of predicted versus experimental PPAR dual activities of the training and test set molecules. Outlier is shown as red circle.

indicates that an increase in the number of atoms results in a decrease of the dual activity.

The second descriptor of importance is Kier & Hall valence connectivity index (order 3) (26). The Kier–Hall indices of molecular similarity measurements employ the Kier–Hall κ and χ connectivity indices. The χ indices, for example, are computed according to following formula:

$$mx^{v} = \sum_{i=1}^{N_s} \prod_{k=1}^{m+1} \left(\frac{1}{\delta_k^v}\right)^{1/2} \quad (4)$$

where m represents the atomic valence connectivity indices, e.g. one bond path valence connectivity indices, two bond fragment valence connectivity indices, and three contiguous bond fragment valence connectivity indices, etc. δ_k^v is the valence connectivity for the kth atom in the molecular graph, which is defined as following:

$$\delta_k^v = \frac{(Z_k^v - H_k)}{(Z_k - Z_k^v - 1)}$$

where Z_k is the total number of electrons in the kth atom, Z_k^v is the number of valence electrons in the kth atom, H_k is the number of hydrogen atoms directly attached to the kth non-hydrogen atom. This descriptor depicts different aspects of atom connectivity within a molecule, such as branching or flexibility. Further, this can be considered to describe the steric crowding around an atom and/or bond, which in turns impacts the accessibility of an atom/bond to interact with the receptor. Simply, this descriptor depicts the shape of the molecules and their role in binding to the target proteins.

The third descriptor represents the number of O atoms in the ligand molecules. A negative coefficient indicates the increase in the activity with a decrease in the number of O atoms.

Table 8: Proposed model for PPARdual activity based on V_{common}

Model index	Nature of range in the proposed model	Index value	No. compounds falling in the range		Percent accuracy
			Total	Correct	
V_{common}	Inactive	296.88–330.38	09	07	78
	Active	332.89–363.54	20	16	80
	Inactive	366.26–393.57	05	05	100

The fourth descriptor of importance is the topographic electronic index (all pairs) (22). Positive contribution of this descriptor indicates the importance of electrostatic forces for the dual activity.

The last descriptor PNSA-3/TMSA (22), which is the fractional atomic charge weighted partial negative surface area, is based on the Zefirov's partial charge calculations. PNSA-3 is calculated as:

$$\text{PNSA} - 3 = \sum_{A} q_A S_A A \in \{\delta_A < 0\} \quad (5)$$

where q_A is the atomic partial charge and S_A is the negatively charged solvent-accessible atomic surface area, and TMSA is the total molecular surface area. This descriptor is related to the induced charge asymmetry in the molecule and to the total molecular surface area. The positive contribution of this descriptor indicates the positive role of electrostatic interactions for the PPAR dual activity.

The descriptor V_{common} calculated as the volume occupied by the ligands in the region common to both PPARα and PPARγ binding sites does not appear in the final QSAR equation for the dual model. However, a model was established based on the relationship of V_{common} with the PPAR dual activity. The results are shown in Table 8. Retrofit analysis of the data in Table 8 reveals the following information regarding the model based on V_{common}:.

- A total of 28 of 34 compounds were classified correctly in both the active and inactive ranges. The overall prediction accuracy was found to be 82.4% with respect to PPAR dual agonistic activity.

- The active range traversed between the two inactive ranges and had a V_{common} value of 332.90–363.54 Å^3. Sixteen of 20 analogs in the active range were predicted correctly resulting in a success rate of prediction of 80%.

The results show that an increase in the PPAR dual activity is observed as the value of V_{common} increases. The occurrence of second inactive range can be explained on the basis that this descriptor depicts only the steric interactions between the ligand and receptor, and possibly other factors (see dual QSAR model at the beginning of this subsection) may also be playing a major role in the present case.

Overall, the QSAR results shows that size and hydrophobicity are the factors which play a major role in defining the PPARα activity as shown by the positive contribution of V_α and *XY* shadow. On the other hand, PPARγ activity is influenced more by the H-bond forming ability of the molecules with the receptor or in other words it is more controlled by the polarity of these molecules. The PPAR dual model reflects that dual PPAR agonistic activity for the thiazolidinedione and oxazolidinedione derivatives is affected by both the size and hydrogen bond forming ability as shown by the positive influence of size and polarity parameters. This model thus appears to be a combination of PPARα and PPARγ model and could be successfully used for the design of compounds with dual activity.

The main focus of the research work carried out in this case was the development of QSAR models which are simple, versatile, highly predictive and simultaneously taking into account the receptor information for the design of PPAR agonists. Quantitative descriptors, V_{site} accounting for the volume occupied by a ligand within the common or specific regions of the binding sites, in combination with other descriptors, have been used for the development of QSAR models. The geometrical descriptors accounting for 3D information of the molecules have been computed for the same conformations as those used for the calculation of V_{site} descriptors. The models developed were highly predictive as shown by the resulting statistics. This study differs from the CoMFA analysis on the same dataset carried out by Khanna *et al.* (9) as mentioned earlier, because it was carried out in a receptor-dependent manner. The success of a CoMFA model depends on a number of factors such as the accuracy with which one can describe the compound's bioactive conformer(s), and the ability to find superimposition rules of the bioactive conformer(s) consistent with ligand–protein interactions. Two outliers in the study by Khanna *et al.* were probably the outliers because of inappropriate superposition on the core molecule. This study also differs from Khanna *et al.* because we use different docked conformations for the development of PPARα and PPARγ models, effectively taking receptor-dependent information under consideration in the analysis of biological activities.

Further evaluation of the QSAR models and the physical significance of the descriptors involved, lead to identify the following key features for PPAR activities of thiazolidinedione and oxazolidinedione derivatives:

- PPARα activity is positively influenced more by the size and shape, whereas negatively by the electrostatic interactions. This shows that hydrophobic interactions are playing major role in governing the activity. Substitution with the hydrophobic groups at R1 position will improve the PPARα activity. However, the more bulky groups at R1 make the molecule inactive for PPARα.

- On the other hand, PPARγ activity is positively influenced by the electrostatic parameters rather than the size. Substituents with hydrogen-bond forming ability at R1 will be helpful in improving the PPARγ activity as shown by the positive contribution of HASA-2/TMSA descriptor in describing the activity.

- The dual model represented the influence of size and electrostatics on the PPAR dual activity of the thiazolidinediones and oxazolidinediones. Substituents such as heteroaryl groups at R1 will improve the PPAR dual activity.

Design of New Molecules

Based on the chemical features extracted from the developed QSAR models, new molecules were designed having selectivity for either of the PPAR isoforms and also dual activators. The docking studies of these molecules were carried out using VDock after optimization by PM3 semi-empirical method using HYPERCHEM. The numerical values for the V_α, V_γ were calculated using the methodology described earlier. The chemical structures for the three designed molecules are shown in Figure 6. The predicted activities along with the calculated V_{site} numerical values are given in Table 9. Analysis of these results shows clearly that the designed molecule T1 is a dual agonist, whereas T2 and T3 are more specific for the PPARα and PPARγ, respectively. Figure 7 shows the docked poses of T1 in the PPARα and PPARγ binding sites. A PPAR dual agonist T1 having the features of PPAR dual model was designed initially by substituting the terminal phenyl ring with a heterocyclic ring substituted with a trifluoromethyl group. T2 was designed to increase the specificity for the PPARα by replacing the trifluoromethyl group with a phenyl ring, as the QSAR model showed us that PPARα activity is influenced mainly by size and thus hydrophobic interactions. T3 molecule has been designed with specificity for PPARγ by incorporating a phenyl-substituted heterocyclic ring.

Conclusions

Structure–activity relationship studies based on the information derived from the receptor can provide a useful insight in the design of multiple activating drug molecules. Design of PPAR ligands with dual activity is one such example of targeting multiple receptors. The thiazoidinedione and oxazolidinedione derivatives used in this study are the dual-activating molecules of PPARα and PPARγ. Novel descriptors have been conceptualized which account for the volume occupied by the ligands in regions that are either common or specific to the superimposed binding sites of the targets, PPARα and PPARγ, under consideration. Statistically valid QSAR models were developed with these descriptors in combination with other geometrical, electrostatics constitutional and topological descriptors for the prediction of PPARα and PPARγ agonistic activity. A dual PPAR model has also been developed which incorporated the structural features required for the dual PPAR activity. Correlation with these novel quantitative descriptors proves the importance of adding receptor dependent descriptors to describe the activity of ligands when designing compounds intended to target multiple receptors, and additionally they can provide an easy and intuitive way to interpret the QSAR equations facilitating the synthesis of new compounds with improved properties based on the predictions of QSAR models.

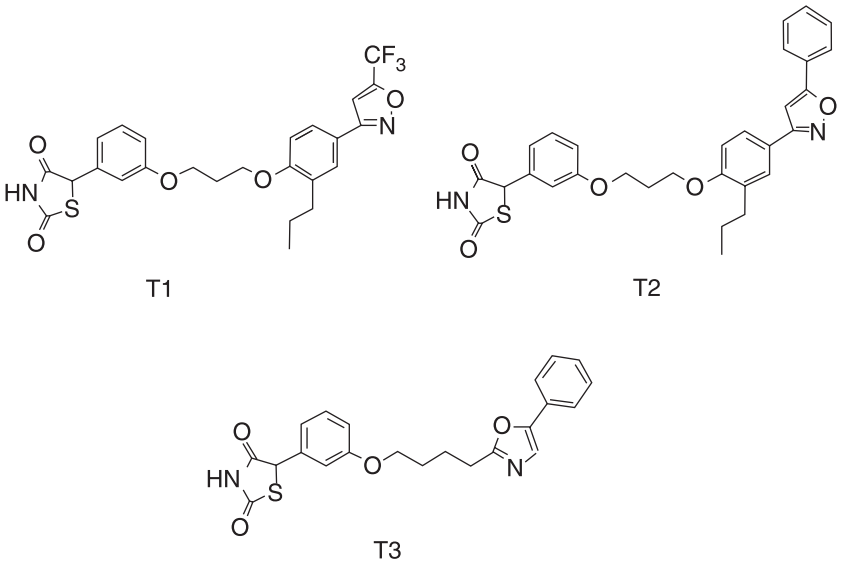


Figure 6: Chemical structures of designed molecules.

Table 9: Numerical values of V_{site} descriptors and predicted activities for designed molecules

Comp. No.	V_α	V_γ	pIC_{50} (α)	pIC_{50} (γ)	pIC_{50} (dual)[a]
T1	30.95	7.68	7.10	7.70	15.48
T2	35.50	33.14	6.97	4.35	13.95
T3	28.07	5.14	8.09	9.54	18.67

[a]Predicted pIC_{50} (dual) based on PPARdual QSAR model.

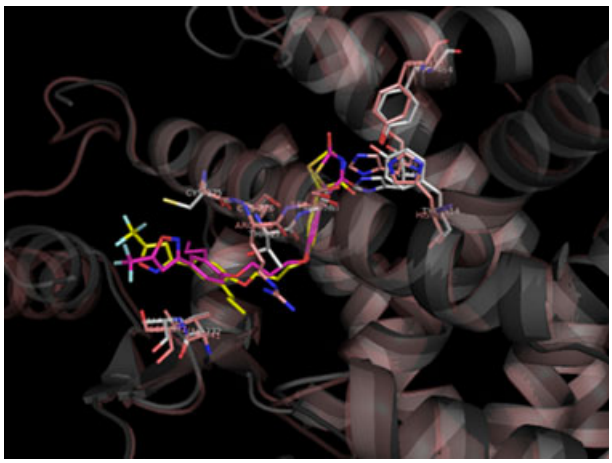


Figure 7: Docked poses of T1 in PPARα (Grey) and PPARγ (Orange). T1 docked pose is shown as yellow for PPARα and magenta for PPARγ.

Acknowledgments

We thank Fundação para a Ciência e Tecnologia (Portugal) for Grant SFRH/BPD/30954/2006 attributed to Viney Lather and SFRH/BPD/20716/2004 attributed to Visvaldas Kairys.

References

1. Baumann K. (2002) An alignment-indpendent versatile structure descriptor for QSAR and QSPR based on the distribution of molecular features. J Chem Inf Comput Sci;42:26–35.
2. Datar P.A., Coutinho E.C. (2004) A CoMFA study of COX-2 inhibitors with receptor based alignment. J Mol Graph Model;23:239–251.
3. Huang H., Pan X., Tan N., Zeng G., Ji C. (2007) 3D-QSAR study of sulfonamide inhibitors of human carbonic anhydrase II. Eur J Med Chem;42:365–372.
4. Pan D., Tseng Y., Hopfinger A.J. (2003) Quantitative structure-based design: formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen phosphorylase. J Chem Inf Model;43:1591–1607.
5. Santos-Filho O.A., Hopfinger A.J. (2006) Structure-based QSAR analysis of a set of 4-hydroxy-5,6-dihydropyrones as inhibitors of HIV-1 protease: An application of the receptor-dependent (RD) 4D-QSAR formalism. J Chem Inf Model;46:345–354.
6. Graul A., Lesson P., Castaner J. (1999) Omapatrilat. Drugs Future;24:269–277.
7. Golase J.M., Arndt K., Etienne C., Lucas J., Nardin D., Gibbons J., Frost P., Ye P., Boschelli D.H., Boschelli F. (2003) SKI-606, a 4-Anilino-3-quinolinecarbonitrile dual inhibitor of Src and Abl kinases, is a potent antiproliferative agent against chronic myelogenous leukemia cells in culture and causes regression of K562 xenografts in nude mice. Cancer Res;63:375–381.
8. Sorbera L.A., Castaner J., del Fresno M., Silvestre J. (2002) Netoglitazone. Drugs Future;27:132–139.
9. Khanna S., Sobhia M.E., Bharatam P.V. (2005) Additivity of molecular fields; CoMFA study on dual activators of PPARα and PPARγ. J Med Chem;48:3015–3025.
10. Desai R.C., Han W., Metzger E.J., Bergman J.P., Gratale D.F., MacNaul K.L., Berger J.P., Doebber T.W., Leung K., Moller D.E., Heck J.V., P S.S. (2003) 5-aryl thiazolidine-2,4-diones: discovery of PPAR dual α/γ agonists as antidiabetic agents. Bioorg Med Chem Lett;13:2795–2798.

11. Desai R.C., Gratale D.F., Han W., Koyama H., Metzger E.J., Lombardo V.K., MacNaul K.L., Doebber T.W., Berger J.P., Leung K., Franklin R., Moller D.E., Heck J.V., P S.S. (2003) Aryloxazolidinediones: identification of potent orally active PPAR dual α/γ agonists. Bioorg Med Chem Lett;13:3541–3544.
12. Gutman I., Russic B., Trinajstic N., Wicox C.X. Jr (1975) Graph theory and molecular orbitals. XII. Acyclic polyenes. J Chem Phys;62:3399–3405.
13. Gutman I., Trinajstic N. (1972) Graph theory and molecular orbitals III. Chem Phys Lett;17:535–538.
14. Balaban A.T. (1982) Highly discriminating distance-based topological index. Chem Phys Lett;80:399–404.
15. Devillers J., Balaban A.T. (editors) (1999) Topological Indices and Related Descriptors in QSAR and QSPR. The Netherlands: Gordon and Breach.
16. Gupta S., Singh M., Madan A.K. (2001) Predicting anti-HIV activity: computational approach using novel topological descriptor. J Comput Aid Mol Des;15:671–678.
17. Nolte R.T., Wisley G.B., Westin S., Cobb J.E., Lambert M.H., Kurokawa R., Rosenfeld M.G., Willson T.M., Glass C.K., Milburn M.V. (1998) Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor gamma. Nature;395:137–143.
18. Kairys V., Gilson M.K. (2002) Enhanced docking with the mining minima optimizer: acceleration and side-chain flexibility. J Comput Chem;23:1656–1670.
19. Jung J., Lee B. (2000) Protein structure alignment using environmental profiles. Protein Eng;13:535–543.
20. Katritzky A.R., Oliferenko A., Lomaka A., Karelson M. (2002) Six-membered cyclic ureas as HIV-1 protease inhibitors: A QSAR study based on CODESSA PRO approach. Bioorg Med Chem Lett;12:3453–3457.
21. Ivanciuc T., Ivanciuc O. (2002) Quantitative structure–retention relationship study of gas chromatographic retention indices for halogenated compounds. Int Elec J Mol Des;1:94–107.
22. Katritzky A.R., Kuanar M., Fara D.C., Karelson M., Acree W.E., Solovev V.P., Varnek A. (2005) QSAR modeling of blood: air and tissue: air partition coefficients using theoretical descriptors. Bioorg Med Chem;13:6450–6463.
23. Code J.E., Perko K.E., Yourtee D.M., Holder A.J., Kostoryz E. (2007) A semi-empirical computational method for the inhibition of porcine cholesterol esterase. J Biomater Sci Polym Ed;18:1457–1474.
24. Spreafico M., Boriani E., Benfenati E., Novic M. (2007) Structural features of diverse ligands influencing binding affinities to estrogen α and estrogen β receptors. Part II. Molecular descriptors calculated from conformation of the ligands in the complex resulting from previous docking study. Mol Divers;11:171–181.
25. Katritzky A.R., Dobchev D.A., Tulp I., Karelson M., Carlson D.A. (2006) QSAR study of mosquito repellents using Codessa Pro. Bioorg Med Chem Lett;16:2306–2311.
26. Kier L.B., Murray W.J., Randic M., Hall L.H. (1975) Molecular connectivity I. Relationship to nonspecific local anesthesia. J Pharm Sci;64:1971–1974.

Notes

[a]The FORTRAN source code, along with example files, needed to calculate the V_{site} descriptors are available, free of charge, for download from the authors' website http://www.uma.pt/molmodel/m3l/software.html.

[b]CODESSA 2.7, © 1992–2004, Shawnee, KS, USA: Semichem Inc.