European Journal of Chemistry 1 (4) (2010) 325-334

European Journal of Chemistry

Journal homepage: www.eurjchem.com

QSAR rationales for the 1,2-diarylcyclopentenes as prostaglandin EP_1 receptor antagonists: Potentially useful in the treatment of inflammatory pain

Brij Kishore Sharma*, Pradeep Pilania and Prithvi Singh

Department of Chemistry, Shri Kalyan Government College, Sikar, IN-332001, India

**Corresponding author at: Department of Chemistry, Shri Kalyan Government College, Sikar, IN-332001, India, Tel.: +91.1572.251039; fax: +91.1572.270039.*
E-mail address: bksharma_sikar@rediffmail.com (B.K. Sharma).

ARTICLE INFORMATION

Received: 06 July 2010
Received in revised form: 29 August 2010
Accepted: 01 September 2010
Online: 31 December 2010

KEYWORDS

1,2-Diarylcyclopentenes
EP_1 receptor antagonists
Dragon descriptors
QSAR
CP-MLR
Partial least square analysis

ABSTRACT

The EP_1 receptor inhibitory activity of 1,2-diarylcyclopentene derivatives have been quantitatively analyzed in terms of Dragon descriptors. The derived QSAR models have provided rationales to explain the EP_1 receptor inhibitory activity of 1,2-diarylcyclopentene derivatives. The 2D-autocorrelation descriptors (MATS4e, MATS5e, MATS7v, GATS5e and GATS7v) have highlighted the role of atomic properties in respective lags of autocorrelations to explain the biological actions of 1,2-diarylcyclopentene analogues. Presence of fluorine atom (nF) and smaller distance between N and O atoms (T(N..O)) in molecular structures, in addition to Kier-Hall electrotopological states (Ss) have also shown prevalence to optimize the EP_1 receptor inhibitory activity. Partial least square analysis has confirmed the dominance of information content of the combinatorial protocol in multiple linear regression identified descriptors. Applicability domain analysis revealed that the suggested model matches the high quality parameters with good fitting power and capability of assessing external data. All the compounds are within the applicability domain of the proposed model and were evaluated correctly.

1. Introduction

The key mediator of pain and inflammation is prostaglandin PGE_2 [1-5]. The first-line treatments for inflammatory pains, including chronic low back pain, pain associated with osteoarthritis and rheumatoid arthritis, are selective COX-2 inhibitors and non-steroidal anti-inflammatory drugs (NSAIDs), which interrupt the biosynthesis of PGE_2 and other prostaglandins (PGs) [6-11]. The gastrointestinal side effects associated with NSAIDs have been successfully reduced with selective COX-2 inhibitors [12-14]. The withdrawal of Vioxx due to increased cardiovascular risk [15] alarmed toward the safety profile of COX-2 inhibitors [16-23]. The PGE_2 exert its biological actions through binding to specific receptors with seven transmembrane domains, the E-prostanoid (EP) receptors. The EP receptors are classified into four subtypes, EP_1, EP_2, EP_3 and EP_4 [24-27], which are located both peripherally and in the CNS [28]. These receptor subtypes are distinguished by their distinct pattern of tissue distribution, signaling pathways, and physiological functions. EP_1 is coupled to intracellular Ca^{2+} mobilization, EP_2 and EP_4 are coupled to stimulation of adenylate cyclase via G_s protein, and EP_3 is coupled to inhibition of adenylate cyclase via G_i protein. Studies which were aimed to identify key structural requirements to synthesize EP selective agonists and/or antagonists and to provide insights to the mechanism of receptor ligand selectivity revealed that sensitive positions for agonist-activity at the EP_1 receptor is the hydroxyl group at the carbon 15 position and C-1 carboxylate [29]. The selective EP_1 antagonists may aid in characterization of the effects mediated by this receptor subtype [30-32]. The reported selective EP_1 receptor antagonists are the acylhydrazide derivative SC51322 (by Searle group) [31], ZD6416 (by AstraZeneca) [33,34], ONO-8713 (by Ono) [35] and thiophene derivative ((by Merck Frosst) [36]. Studies revealed that EP_1 receptor plays a central role in PGE_2-mediated allodynia [37,38] and inflammatory pain [39]. The efficacy of EP_1 receptor antagonists shown in preclinical models [40-42], led to hypothesize that these antagonists may furnish improved safety profile by sparing the synthesis of PGs.

As an attempt to identify novel EP_1 receptor antagonists, recently, a series of 1,2-diarylcyclopentene derivatives has been reported as potential clinically effective analgesics [43]. This study led to the discovery of GW848687X, a potential candidate for the treatment of inflammatory pain. In view of the importance of anti-inflammatory agents in the clinical management of several disorders, a quantitative structure-activity relationship is attempted on the EP_1 inhibitory activity of these 1,2-diarylcyclopentene derivatives. The present study is aimed at rationalizing the substituent variations of these analogues to provide insight for the future endeavors.

2. Experimental

2.1. Data set

The reported thirty three 1,2-diarylcyclopentene derivatives are considered as data set for the present study [43]. These compounds were evaluated for their ability to inhibit CYP450 enzymes in human recombinant CYP450 assays in terms of IC_{50} (molar concentration for 50% inhibition) values. The structural variations and reported EP_1 inhibitory activity (as pIC_{50}, on molar basis) of these analogues are given in Table 1. For the purpose of modeling study all 33 analogues

ISSN 2153-2249 (Print) / ISSN 2153-2257 (Online)
DOI:10.5155/eurjchem.1.4.325-334.195

have been divided into training and test sets. Out of the 33 analogues, one fourth compounds (8) have been placed in the test set for the validation of derived models. The test set was generated in the SYSTAT [44] using the single linkage hierarchical cluster procedure involving the Euclidean distances of the activity. The selection of the test set from the cluster tree was done in such a way to keep the test compounds at a maximum possible distance from each other. The training and test set compounds are also listed in Table 1.

2.2. Theoretical molecular descriptors

The structures of the compounds under study have been drawn in 2D ChemDraw [45]. The drawn structures were then converted into 3D modules using the default conversion procedure implemented in the CS Chem3D Ultra. The energy of these 3D-structures was minimized in the MOPAC module using the AM1 procedure for closed shell systems. This will ensure a well defined conformer relationship among the compounds of the study. All these energy minimized structures of respective compounds have been ported to DRAGON software [46] for the computation of descriptors for the titled compounds (Table 1). This software offers several hundreds of descriptors from different perspectives corresponding to 0D-, 1D-, and 2D-descriptor modules. The outlined modules comprised of ten different classes, namely, the constitutional (CONST), the topological (TOPO), the molecular walk counts (MWC), the BCUT descriptors (BCUT), the Galvez topological charge indices (GALVEZ), the 2D autocorrelations (2D-AUTO), the functional groups (FUNC), the atom-centered fragments (ACF), the empirical descriptors (EMP), and the properties describing descriptors (PROP). For each of these classes the DRAGON software computes a large number of descriptors which are characteristic to the molecules under multi-descriptor environment. The definition and scope of these descriptor's classes is given in Table 2.

Both the 2D- and 3D-descripors may be used to obtain significant QSARs. However, a QSAR study involving 0D to 2D-descriptors is quite simple to interpret the biological data in terms of different descriptors obtained from the two dimensional structures of the compounds. In a congeneric series, where a relative study is being carried out, the 2D-descriptors may play important role in deriving the significant correlations with biological activities of the compounds. Thus the novelty and importance of a 2D-QSAR study is mainly due to its simplicity for the calculations of different descriptors and their interpretation (in physical sense) to explain the biological actions of compounds in a congeneric series.

The combinatorial protocol in multiple linear regression (CP-MLR) [47] and partial least-squares (PLS) [48-50] procedures have been used in the present work for developing QSAR models. Before the application of CP-MLR procedure, all those descriptors which are intercorrelated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor vs. activity, $r < 0.1$) were excluded. This has reduced the total dataset of the compounds from 457 to 88 descriptors as relevant ones for the EP_1 inhibitory activity. A brief description of the computational procedure is given below.

2.3. Model development

The CP-MLR is a 'filter' based variable selection procedure for model development in QSAR studies [47]. Its procedural aspects and implementation are discussed in some of our recent publications [51-55]. It involves selected subset regressions. In this procedure a combinatorial strategy with appropriately placed 'filters' has been interfaced with MLR to result in the extraction of diverse structure-activity models, each having unique combination of descriptors from the dataset under study. In this, the contents and the number of variables to be evaluated are mixed according to the predefined confines. Here the 'filters' are significance evaluators of the variables in regression at different stages of model development. Of these, filter-1 is set in terms of inter-parameter correlation cutoff criteria for variables to stay as a subset (filter-1, default value 0.3 and upper limit ≤ 0.79). In this, if two variables are correlated higher than a predefined cutoff value the respective variable combination is forbidden and will be rejected. The second filter is in terms of t-values of regression coefficients of variables associated with a subset (filter-2, default value 2.0). Here, if the ratio of regression coefficient and associated standard error of any variable is less than a predefined cutoff value then the variable combination will be rejected. Since successive additions of variables to multiple regression equation will increase successive multiple correlation coefficient (r) values, square-root of adjusted multiple correlation coefficient of regression equation, r-bar, has been used to compare the internal explanatory power of models with different number of variables. Accordingly, a filter has been set in terms of predefined threshold level of r-bar (filter-3, default value 0.71) to decide the variables' 'merit' in the model formation. Finally, to exclude false or artificial correlations, the external consistency of the variables of the model have been addressed in terms of cross-validated R^2 or Q^2 criteria from the leave-one-out (LOO) cross-validation procedure as default option (filter-4, default threshold value $0.3 \le Q^2 \le 1.0$). All these filters make the variable selection process efficient and lead to unique solution. In order to collect the descriptors with higher information content and explanatory power, the threshold of filter-3 was successively incremented with increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for next generation.

2.4. Model validation

In this study, the data set is divided into training set for model development and test set for external prediction. Goodness of fit of the models was assessed by examining the multiple correlation coefficient (r), the standard deviation (s), the F-ratio between the variances of calculated and observed activities (F). A number of additional statistical parameters such as the Akaike's information criterion, AIC [56,57], the Kubinyi function, FIT [58,59], and the Friedman's lack of fit, LOF [60], (Eqs. 1-3) have also been derived to evaluate the best model.

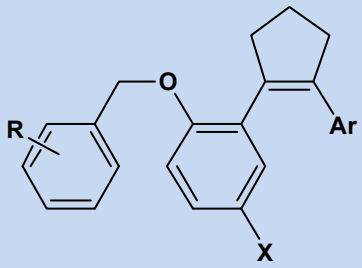
$$AIC = \frac{RSS \times (n + p')}{(n - p')^2} \quad (1)$$

$$FIT = \frac{r^2 \times (n - k - 1)}{(n + k^2) \times (1 - r^2)} \quad (2)$$

$$LOF = \frac{RSS/n}{\left[1 - \frac{k(d+1)}{n}\right]^2} \quad (3)$$

where, RSS is the sum of the squared differences between the observed and the estimated activity values, k is the number of variables in the model, p' is the number of adjustable parameters in the model, and d is the smoothing parameter. The AIC takes into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. The FIT, closely related to the F-value (Fisher ratio), was proved to be a useful parameter for assessing the quality of the models.

Table 1. Structures and observed EP_1 receptor inhibitory activity of 1,2-diarylcyclopentene derivatives.



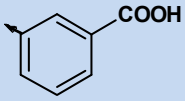
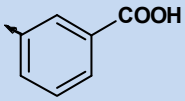
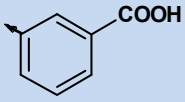
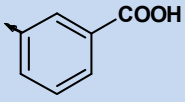
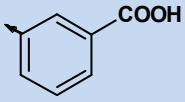
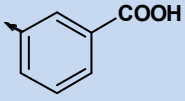
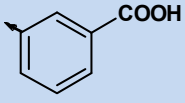
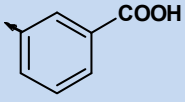
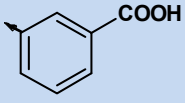
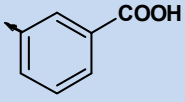
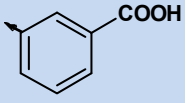
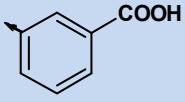
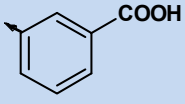
Compound	X	R	Ar	pIC_{50}[a]				
				Obsd.[b]	Calcd.			
					Eq. 8	Eq. 9	Eq. 10	PLS
1	H	H		8.2	7.7	7.8	7.8	7.8
2	H	2,4-F_2		7.6	7.9	7.9	7.9	7.9
3	H	3,4-Cl_2		6.4	6.2	6.3	6.4	6.1
4[c]	H	2-F,4-Cl		7.3	7.4	7.7	7.6	7.7
5[c]	H	4-OMe		7.0	6.6	7.0	6.8	7.0
6[c]	Cl	H		8.3	7.8	7.4	7.5	7.7
7	Br	H		7.9	8.2	8.0	8.0	8.1
8	Br	4-Cl		7.9	7.8	7.9	7.8	7.8
9	Br	4-F		7.7	8.1	8.3	8.1	7.9
10	Br	3,4-Cl_2		7.0	7.2	6.9	7.0	7.0
11	Br	2,4-F_2		8.3	8.7	8.5	8.5	8.7
12	Br	2-F, 4-Cl		8.6	8.3	8.3	8.2	8.5
13	Br	4-OMe		7.4	7.1	7.2	7.1	7.3

Table 1 *(Continued).*

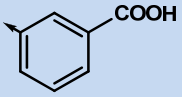
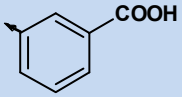
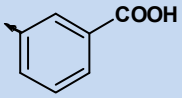
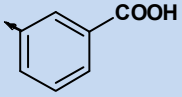
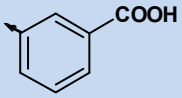
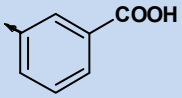
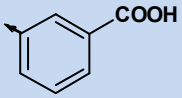
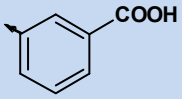
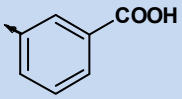
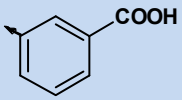
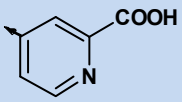
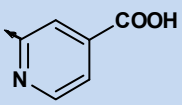
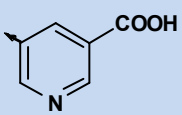
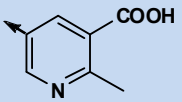
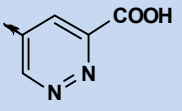
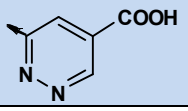
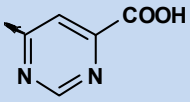
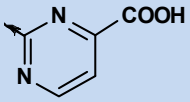
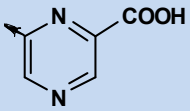
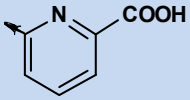
Compound	X	R	Ar	pIC$_{50}$[a]				
				Obsd.[b]	Calcd.			
					Eq. 8	Eq. 9	Eq. 10	PLS
14	SMe	H		7.2	7.2	7.4	7.5	7.5
15[c]	SMe	4-F		7.5	7.2	7.7	7.6	7.4
16	SMe	2,4-F$_2$		7.8	7.8	7.9	8.0	8.1
17	SO$_2$Me	H		7.1	7.0	7.2	7.1	7.1
18	SO$_2$Me	4-F		7.1	7.1	7.3	7.3	7.2
19	SO$_2$Me	2,4-F$_2$		7.8	7.4	7.4	7.6	7.6
20	CN	H		7.0	7.2	6.8	6.9	7.0
21	CN	4-F		6.4	7.1	7.0	6.9	6.8
22	CN	2,4-F$_2$		7.4	7.5	7.1	7.2	7.5
23	CN	4-Cl		6.7	6.6	6.7	6.6	6.6
24[c]	CF$_3$	H		7.4	7.0	7.3	7.9	7.6
25	Br	4-F		8.5	8.2	7.9	7.9	7.6
26[c]	CF$_3$	H		8.0	7.8	8.2	8.7	8.1
27	Cl	4-F		8.8	8.4	8.8	8.7	8.5
28	Cl	H		7.4	7.4	7.3	7.1	7.4
29	Cl	2,4-F$_2$		8.2	8.4	8.1	8.1	8.2

Table 1 *(Continued).*

Compound	X	R	Ar	Obsd.[b]	pIC_{50}[a] Calcd.			
					Eq. 8	Eq. 9	Eq. 10	PLS
30[c]	Cl	H		6.7	6.4	6.4	6.6	6.7
31	Cl	H		6.8	6.9	7.4	7.6	7.2
32[c]	Cl	2,4-F_2		8.7	8.1	8.6	8.5	8.4
33	Cl	2,4-F_2		8.6	8.2	8.4	8.4	8.4

[a] On molar basis.
[b] Taken from reference [43].
[c] Compounds included in test set.

Table 2. Descriptor classes [a] used along with their definition and scope for modeling the EP_1 receptor inhibitory activity of 1,2-diarylcyclopentene derivatives.

Descriptor class (acronyms)	Definition and scope
Constitutional (CONST)	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations
Topological (TOPO)	2D-descriptor from molecular graphs and independent conformations
Molecular walk counts (MWC)	2D-descriptors representing self-returning walks counts of different lengths
Modified Burden eigenvalues (BCUT)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms
Galvez topological charge indices (GALVEZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix
2D-autocorrelations (2D-AUTO)	Molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag)
Functional groups (FUNC)	Molecular descriptors based on the counting of the chemical functional groups
Atom centered fragments (ACF)	Molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen
Empirical (EMP)	1D-descriptors represent the counts of non-single bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule
Properties (PROP)	1D-descriptors representing molecular properties of a molecule

[a] Reference [46].

The main disadvantage of the F-value is its sensitivity to changes in k (the number of variables in the equation, which describe the model), if k is small, and its lower sensitivity if k is large. The FIT criterion has a low sensitivity toward changes in k-values, as long as they are small numbers, and a substantially increasing sensitivity for large k-values. The model that produces the minimum value of AIC and the highest value of FIT is considered potentially the most useful and the best. The LOF takes into account the number of terms used in the equation and is not biased, as are other indicators, toward large numbers of parameters. A minimum LOF value infers that the derived model is statistically sound.

The internal validation of derived model was ascertained through the cross-validated index, Q^2, from leave-one-out and leave-five-out procedures. The LOO method creates a number of modified data sets by taking away one compound from the parent data set in such a way that each observation has been removed once only. Then one model is developed for each reduced data set, and the response values of the deleted observations are predicted from these models. The squared differences between predicted and actual values are added to give the predictive residual sum of squares, PRESS. In this way, PRESS will contain one contribution from each observation. The cross-validated Q^2_{LOO} value may further be calculated as

$$Q^2_{LOO} = 1 - PRESS/SSY \quad (4)$$

where, SSY represents the variance of the observed activities of molecules around the mean value. In leave-five-out procedure, a group of five compounds is randomly kept outside the analysis each time in such a way that all the compounds, for once, become the part of the predictive groups. A value greater than 0.5 of Q^2-index hints toward a reasonable robust model.

The external validation or predictive power of derived model is based on test set compounds. The squared correlation coefficient between the observed and predicted values of compounds from test set, r^2_{Test}, has been calculated as

$$r^2_{Test} = 1 - \frac{\sum (Y_{Pred(Test)} - Y_{(Test)})^2}{\sum (Y_{(Test)} - \overline{Y}_{(Training)})^2} \quad (5)$$

where, $Y_{Pred(Test)}$ and $Y_{(Test)}$ indicate predicted and observed activity values, respectively of the test-set compounds, and $\overline{Y}_{(Training)}$) indicate mean activity value of the training set. r^2_{Test} is the squared correlation coefficient between the observed and predicted data of the test-set. It suggests the fraction of explained variance in the test set which is not part of regression/model derivation. It is a measure of goodness of the derived model equation. A high r^2_{Test} value is always good. But

considering the stringency of test set procedures, often r^2_{Test} values in the range of 0.500–0.600 are regarded as indicative predictive models.

2.5. Y-randomization

Chance correlations, if any, associated with the CP-MLR models were recognized in randomization test [61,62] by repeated scrambling of the biological response. The data sets with scrambled response vector have been reassessed by multiple regression analysis. The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to the unscrambled response data were counted. Every model has been subjected to 100 such simulation runs. This has been used as a measure to express the percent chance correlation of the model under scrutiny.

2.6. Applicability domain

The utility of a QSAR model is based on its accurate prediction ability for new compounds. A model is valid only within its training domain, and new compounds must be assessed as belonging to the domain before the model is applied. The applicability domain is assessed by the leverage values for each compound [63]. A Williams plot (the plot of standardized residuals versus leverage values (h) can then be used for an immediate and simple graphical detection of both the response outliers (Y outliers) and structurally influential chemicals (X outliers) in the model. In this plot, the applicability domain is established inside a squared area within $\pm x$ (standard deviations) and a leverage threshold h^*. The threshold h^* is generally fixed at $3(k+1)/n$ (n is the number of training-set compounds, and k is the number of model parameters) whereas $x = 2$ or 3. Prediction must be considered unreliable for compounds with a high leverage value ($h > h^*$). On the other hand, when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training set compounds.

3. Results and discussion

3.1. QSAR results

To rationalize the substituent variations of the 1,2-diarylcyclopentene derivatives to provide insight for the future endeavors, 88 descriptors accounting 0D-, 1D- and 2D-molecular features of the analogues have been subjected to CP-MLR analysis with default 'filters' set in it. Statistical models in two, three and four descriptor(s) have been derived successively to achieve the best relationship correlating EP_1 inhibitory activity. These models (with 88 descriptors) were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this the optimum r-bar value of the preceding level model has been used as the new threshold of filter-3 for the next generation. The highest significant models in three and four descriptors are given below.

pIC_{50} = −3.514 − 0.404(0.099)X0sol + 0.059(0.015)T(F..Cl) + 5.345(1.061)BEHm4

n = 25, r = 0.821, s = 0.423, F = 14.501, r^2_{randY}(sd) = 0.335(0.126), Q^2_{LOO} = 0.521, Q^2_{L50} = 0.556, FIT = 1.279, LOF = 0.261, AIC = 0.247, r^2_{Test} = 0.543, SE_{Test} = 0.491, R_0^2 = 0.984, R'^2_0 = 0.999, k = 1.007, k' = 0.989 (6)

pIC_{50} = −5.280 + 18.704(5.169)BELm3 − 10.160(2.418)BEHe6 + 5.700(1.058)MATS4e + 9.002(1.630)GATS7v

n = 25, r = 0.875, s = 0.368, F = 16.308, r^2_{randY}(sd) = 0.376(0.133), Q^2_{LOO} = 0.614, Q^2_{L50} = 0.653, FIT = 1.591, LOF = 0.235, AIC = 0.203, r^2_{Test} = 0.517, SE_{Test} = 0.505, R_0^2 = 0.533, R'^2_0 = 0.997, k = 1.046, k' = 0.953 (7)

In above and all follow up regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The r^2_{randY}(sd) is the mean random squared multiple correlation coefficient of the regressions in the activity (Y) randomization study with its standard deviation from 100 simulations. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The SE_{Test} is the standard error of estimation for the test set activity values. R_0^2 and R'^2_0 are the coefficients of determination, predicted versus observed activity and observed versus predicted activity, respectively and k and k' are the slopes of regression lines, predicted versus observed activity and observed versus predicted activity, respectively with the intercept set to 0) through the origin [64]. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models.

The participated descriptors X0sol and T(F..Cl), and BEHm4 in model (6) are from TOPO and BCUT class of Dragon descriptors, respectively. The descriptor X0sol, solvation connectivity index chi-0, has shown negative influence on the activity. Thus, suggesting that a lower value of 0^{th} order solvation connectivity index would be favorable to the activity. From the positive sign of regression coefficient of descriptor T(F..Cl), the sum of topological distance between F and Cl atoms, it is evident that a bigger value of such distance would augment the activity. The descriptor BEHm4, the highest eigenvalue n.4 of Burden matrix weighted by atomic masses, contributed positively to the activity recommending a higher value of this descriptor for improved inhibitory activity of 1,2-diarylcyclopentene derivatives.

It is apparent from the participating descriptors in model (7) that atomic properties such as atomic masses, Sanderson electronegativities and van der Waals volumes played a pivotal role in explaining the inhibitory actions of the titled compounds. The descriptors, BELm3 and BEHe6 are BCUT descriptors. From the signs of regression coefficients of it may be ascertained that a higher value of descriptor BELm3, the lowest eigenvalue n.3 of Burden matrix weighted by atomic masses, and a lower value of descriptor BEHe6, the highest eigenvalue n.6 of Burden matrix wighted by atomic Sanderson electronegativities would be beneficiary to the activity. Descriptors MATS4e and GATS7v are the 2D-autocorrelation descriptors. Both these descriptors have shown positive influence on activity recommending higher values of Moran autocorrelation of lag 4 weighted by atomic Sanderson electronegativities (MATS4e) and Geary autocorrelation of lag 7 weighted by atomic van der Waals' volumes (GATS7v) for elevated activity.

The four descriptor models could estimate 76.56 percent variance in observed activity of the compounds. Considering the number of observation in the dataset, models with up to five descriptors were explored. A total number of 9 models, sharing 14 descriptors among them, were obtained through CP-MLR. All these 14 descriptors along with their brief meaning, average regression coefficients and total incidence are listed in Table 3, which will serve as a measure of their estimate across these models. The given below are some five-descriptor models for the activity. These models have accounted for up to 81 percent variance in the observed activities.

pIC_{50} = 1.434 + 0.035(0.016)Ss + 23.074(5.143)BELm3 − 15.631(3.343)BEHe6 + 3.992(1.245)MATS4e + 8.767(1.498)GATS7v

n = 25, r = 0.901, s = 0.338, F = 16.472, r^2_{randY}(sd) = 0.460(0.117), Q^2_{LOO} = 0.648, Q^2_{L50} = 0.680, FIT = 1.647, LOF =

Table 3. Physical meaning, average regression coefficients and the total incidences, and MLR-like coefficients from PLS model of descriptors identified from five parameter CP-MLR models for the EP_1 receptor inhibitory activity of 1,2-diarylcyclopentene derivatives.

No.	Descriptors' Class	Descriptors' symbol, and meaning	Avg. reg. coef. (incidence)[a]	MLR-like coef. (fc) Order[b]
1	CONST	Ss, sum of Kier-Hall electrotopological states	0.038 (2)	-0.048 (-0.018) 13
2	CONST	nF, number of fluorine atoms	0.321 (1)	0.175 (0.066) 11
3	CONST	nX, number of halogen atoms	0.314 (2)	0.179 (0.068) 8
4	TOPO	T(N..O), sum of topological distance between N and O atoms	-0.020 (4)	-0.276 (-0.104) 3
5	TOPO	T(Cl..Cl), sum of topological distance between Cl and Cl atoms	-0.480 (1)	-0.287 (-0.108) 2
6	TOPO	MWC10, 10th order molecular walk count	4.663 (2)	0.178 (0.067) 10
7	BCUT	BELm3, lowest eigenvalue n.3 of Burden matrix weighted by atomic masses	20.677 (5)	0.148 (0.056) 12
8	BCUT	BEHe6, highest eigenvalue n.6 of Burden matrix weighted by atomic Sanderson electronegativities	-13.661 (9)	-0.340 (-0.128) 1
9	2D-AUTO	MATS7v, Moran autocorrelation of lag 7 weighted by atomic van der Waals volumes	-12.505 (6)	-0.184 (-0.069) 6
10	2D-AUTO	MATS4e, Moran autocorrelation of lag 4 weighted by atomic Sanderson electronegativities	4.213 (5)	0.209 (0.079) 5
11	2D-AUTO	MATS5e, Moran autocorrelation of lag 5 weighted by atomic Sanderson electronegativities	-11.805 (4)	-0.183 (-0.069) 7
12	2D-AUTO	MATS6e, Moran autocorrelation of lag 6 weighted by atomic Sanderson electronegativities	4.682 (1)	-0.179 (-0.067) 9
13	2D-AUTO	GATS7v, Geary autocorrelation of lag 7 weighted by atomic van der Waals volumes	8.540 (2)	0.245 (0.092) 4
14	2D-AUTO	GATS5e, Geary autocorrelation of lag 5 weighted by atomic Sanderson electronegativities	-2.596 (1)	0.025 (0.009) 14

[a] The average regression coefficient of the descriptor corresponding to all models and the total number of its incidences; the arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.
[b]MLR like regression coefficient of three-component PLS model; (fc) is fraction contribution of the regression coefficient to the activity; order indicates the order of their significance in the PLS model; the constant term of PLS model is 7.592; number of compounds are 28. PLS regression and validation statistics: r = 0.906, s = 0.314, F = 32.099, Q^2_{LOO} = 0.739, Q^2_{L50} = 0.730, r^2_{Test} = 0.794.

0.241, AIC = 0.186, r^2_{Test} = 0.599, SE_{Test} = 0.460, R_0^2 = 0.657, $R'_0{}^2$ = 0.997, k = 1.046, k' = 0.954 (8)

pIC_{50} = 54.539 – 0.018(0.006)T(N..O) –14.550(2.225)BEHe6 – 14.116(2.174)MATS7v –15.617(3.126)MATS5e – 2.595(0.641)GATS5e

n = 25, r = 0.898, s = 0.343, F = 15.858, r^2_{randY}(sd) = 0.416(0.119), Q^2_{LOO} = 0.666, Q^2_{L50} = 0.728, FIT = 1.586, LOF = 0.248, AIC = 0.192, r^2_{Test} = 0.600, SE_{Test} = 0.459, R_0^2 = 0.991, $R'_0{}^2$ = 0.999, k = 1.007, k' = 0.989 (9)

pIC_{50} = 43.765 + 0.321(0.085)nF – 0.020(0.006)T(N..O) – 11.828(2.168)BEHe6 –12.694(2.276)MATS7v – 10.140(3.058)MATS5e

n = 25, r = 0.891, s = 0.353, F = 14.704, r^2_{randY}(sd) = 0.441(0.134), Q^2_{LOO} = 0.660, Q^2_{L50} = 0.607, FIT = 1.470, LOF = 0.264, AIC = 0.204, r^2_{Test} = 0.500, SE_{Test} = 0.513, R_0^2 = 0.990, $R'_0{}^2$ = 0.999, k = 0.991, k' = 1.005 (10)

In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. For above EP_1 receptor inhibition models, ten high r_{randY} values from hundred Y-randomizations are included in Table 4.

Table 4. Ten random correlation coefficients [a] (r_{randY}) from the activity (Y) randomization study of the models.

No.	Eq. (8)	Eq. (9)	Eq. (10)
1	0.709	0.693	0.775
2	0.686	0.641	0.773
3	0.655	0.640	0.737
4	0.630	0.631	0.688
5	0.519	0.613	0.673
6	0.363	0.605	0.647
7	0.389	0.605	0.624
8	0.501	0.591	0.563
9	0.433	0.566	0.562
10	0.572	0.521	0.542

[a] Values of 10 high r_{randY} recorded from 100 Y-randomizations.

The values greater than 0.5 of Q^2-indices is in accordance to a reasonable robust QSAR model. The pIC_{50} values of the training set compounds calculated using Equations (8) to (10) have been mentioned in Table 1. The models (8) to (10) are validated with an external test set of eight compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values and the predicted activity values are also reported in Table 1. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

The descriptors BELm3, BEHe6, MATS4e and GATS7v, which were emerged in models (6 and 7), have once again shown their importance in five parameter models and convey same inferences to the activity. The descriptor Ss, the sum of Kier-Hall electrotopological states, and nF, number of fluorine atoms in molecular structure are from CONST class. The positive influence of these descriptors, in models, to the activity demanded a higher value of Kier-Hall electrotopological states and more number of fluorine atoms in a molecular structure for improved activity. The TOPO class descriptor, T(N..O), the sum of topological distance between N and O atoms, advocates smaller distances between N and O atoms in a molecule for the better inhibition activity. The remaining descriptors MATS7v, MATS5e and GATS5e are 2D-AUTO class descriptors. All these descriptors have contributed negatively to the activity. A higher values of Moran autocorrelation of lag 7 weighted by atomic van der Waals volumes (MATS7v), Moran and Geary autocorrelations of lag 5 weighted by atomic Sanderson electronegativities (MATS5e and GATS5e, respectively) would be detrimental to the activity.

According to the test set R^2 values and the corresponding R_0^2 and $R'_0{}^2$ suggest that equations (7) and (8) have higher predictive power when compared to other equations listed. However, the k and k' values are within acceptable range for all the models. In a comprehensive manner, the statistics emerged from the test sets have validated the models and ranked equations (7) and (8) as best bets.

A PLS (partial least squares) analysis has been carried out on the 14 CP-MLR identified descriptors (Table 4) to facilitate the development of a 'single window' structure-activity model and to identify their (descriptors) potential in explaining the EP_1 receptor inhibition actions of 1,2-diarylcyclopentene derivatives. It also gives an opportunity to make a comparison

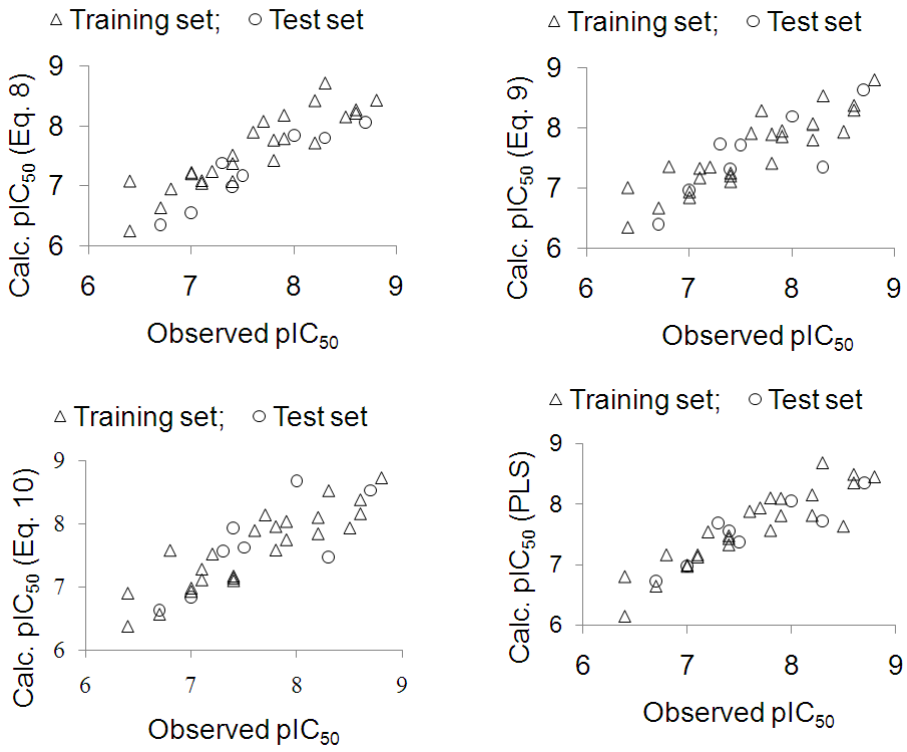


Figure 1. Plot of observed versus calculated pIC_{50} values for training and test set compounds.

of the relative significance among the descriptors. The fraction contributions obtainable from the normalized regression coefficients of the descriptors allow this comparison within the modeled activity. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit s.d.) to give each one of them equal weight in the analysis. In the PLS cross-validation, three components are found to be the optimum for these 14 descriptors and they explained 82.08% variance in the activity ($r = 0.906$, $Q^2_{LOO} = 0.739$, $s = 0.314$, $F = 32.099$, $r^2_{Test} = 0.794$). The MLR-like PLS coefficients of these 14 descriptors are given in Table 3. The calculated activity values of training and test set compounds are in close agreement to that of the observed ones and are listed in Table 1. For the sake of comparison, the plot showing goodness of fit between observed and calculated activities (through PLS analysis) for the training and test set compounds is given in Figure 1. Figure 2 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity (Table 3).

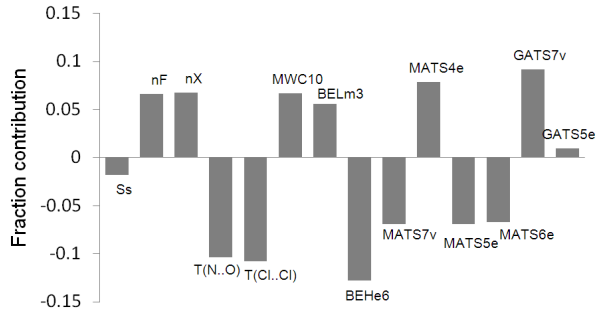


Figure 2. Plot of fraction contribution of MLR-like PLS coefficients (normalized) of the 14 descriptors (Table 3) to EP_1 receptor inhibitory activity of 1,2-diarylcyclopentene derivatives.

The PLS analysis has also suggested BEHe6 (a 2D-AUTO class descriptor) as the most determining descriptor for modeling the activity of the compounds (descriptor S. No. 8 in Table 3; Figure 2). The other nine significant descriptors in decreasing order of significance are T(Cl..Cl), T(N..O), GATS7v, MATS4e, MATS7v, MATS5e, nX, MATS6e and MWC10 (descriptors S. No. 5, 4, 13, 10, 9, 11, 3, 12 and 6 in Table 3; Figure 2). Except, T(Cl..Cl), nX, MATS6e and MWC10, all these descriptors are part of Equations 8-10 and convey same inference in the PLS model as well. The TOPO class descriptor T(Cl..Cl), the sum of topological distance between Cl and Cl atom, advocates that the shorter distance between chlorine atoms in a molecular structure would be beneficiary to the activity. The positive influence of descriptor, nX (number of halogen atoms) to the activity recommended the presence of halogen atoms in a compound for improved activity. The negative regression coefficient of the Moran autocorrelation of lag 6 weighted by atomic Sanderson electronegativities (descriptor MATS6e) advocates that a higher positive value of it is detrimental to the activity. Descriptor MWC10 (a TOPO class descriptor) representing 10^{th} order molecular walk count suggests a higher value of 10^{th} order walk count to enhance the activity. In comparison to these ten descriptors, the remaining ones appear in lower order of significance to influence the activity of the compounds (Table 3; Figure 2). It is also observed that PLS model from the dataset devoid of 14 descriptors (Table 3) is inferior in explaining the activity of the analogues.

3.2. Applicability domain

On analyzing the model applicability domain (AD) in the Williams plot (Figure 3) of the model based on the whole data set (Table 5), it has appeared that none of the compounds were identified as an obvious outlier for the EP_1 receptor inhibitory activity if the limit of normal values for the Y outliers (response outliers) was set as 3 (standard deviation) units. None of the compounds was found to have leverage (h) values greater than the threshold leverages (h^*). For both the training set and test set, the suggested model matches the high quality parameters

Table 5. The derived models for the whole data set (n=33) in identified descriptors of Eqs. 8-10 for the EP_1 receptor inhibitory activity of 1,2-diarylcyclopentene derivatives.

Model	r	s	F	Q^2_{LOO}	Q^2_{L50}	Eq.
pIC_{50} = 2.407 + 0.033(0.012)Ss + 20.562(3.865)BELm3 – 14.289(2.801)BEHe6 + 3.825(1.079)MATS4e + 8.556(1.352)GATS7v	0.886	0.342	19.814	0.683	0.678	8a
pIC_{50} = 52.768 – 0.018(0.005)T(N..O) – 14.092(1.967)BEHe6 – 13.952(1.885)MATS7v – 12.840(2.306)MATS5e – 2.119(0.585)GATS5e	0.884	0.345	19.281	0.674	0.679	9a
pIC_{50} = 45.743 + 0.235(0.072)nF – 0.021(0.005)T(N..O) – 12.415(1.975)BEHe6 –12.681(2.043)MATS7v –8.425(2.527)MATS5e	0.875	0.357	17.685	0.665	0.662	10a

with good fitting power and the capability of assessing external data. Furthermore, almost all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

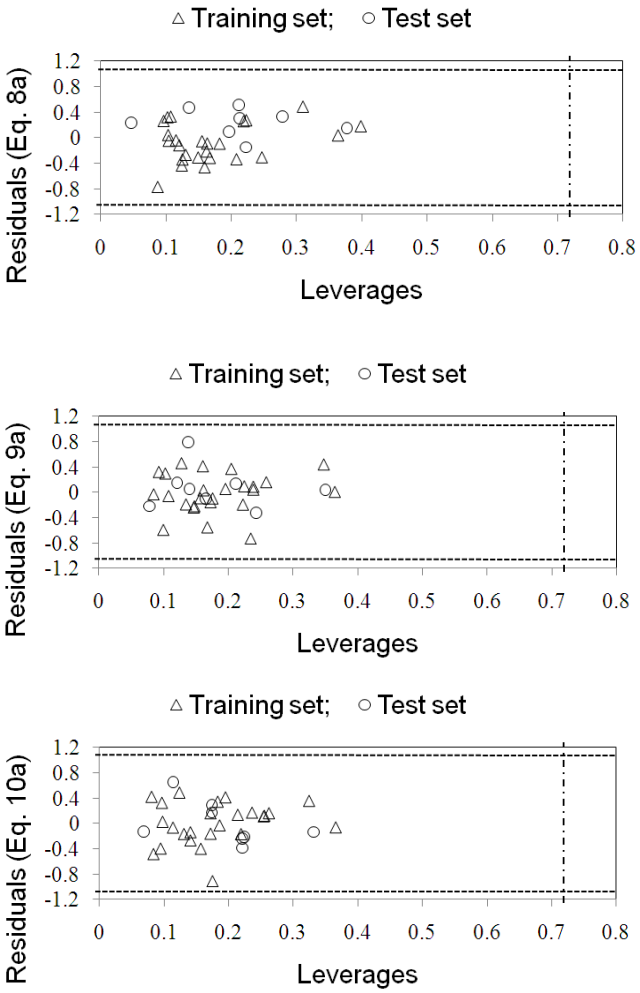


Figure 3. Williams plot for the training set and external prediction set for EP_1 receptor inhibitory activity of 1,2-diarylcyclopentene derivatives. The residuals for training and test set compounds are shown by Δ and O, respectively. The horizontal dotted line refers to the residual limit (±3× standard deviation) and the vertical dotted line represents threshold leverage h* (= 0.720).

4. Conclusion

The derived QSAR models have provided rationales to explain the EP_1 receptor inhibitory activity of 1,2-Diarylcyclopentene derivatives. The 2D-autocorrelation descriptors (MATS4e, MATS5e, MATS7v, GATS5e and GATS7v) and BCUT descriptors (BELm3 and BEHe6) have highlighted the role of atomic properties in respective lags of autocorrelations and eigen-values to explain the biological actions of 1,2-diarylcyclopentene analogues. The presence of fluorine atom (nF) and the smaller distance between N and O atoms (T(N..O)) in molecular structures, in addition to Kier-Hall electrotopological states (Ss) have also shown prevalence to optimize the EP_1 receptor inhibitory activity. The statistics emerged from the test sets have validated the models and suggested that the descriptors Ss, BELm3, BEHe6, MATS4e and GATS7v are more important in comparison to other descriptors to explain the inhibitory activity of the titled compounds. PLS analysis has also confirmed the dominance of information content of the CP-MLR identified descriptors. Applicability domain analysis revealed that the suggested models have acceptable predictability. All the compounds are within the applicability domain of the proposed models and were evaluated correctly.

Acknowledgement

Authors are thankful to their Institution for providing the necessary facilities to complete this study. Pradeep Pilania thanks CSIR, New Delhi (India) for the financial assistance in the form of SRF.

Supplementary material

The supplementary data for this paper can be obtained free of charge via the publisher.

References

[1]. Murakami, M.; Nakatani, Y.; Tanioka, T.; Kudo, I. *Prostaglandins Other Lipid Mediat.* **2002**, *68-69*, 383-399.
[2]. Claveau, D.; Sirinyan, M.; Guay, J.; Gordon, R.; Chan, C-C.; Bureau, Y.; Riendeau, D.; Mancini, J. A. *J. Immunol.* **2003**, *170*, 4738-4744.
[3]. Trebino, C. E.; Stock, J. L.; Gibbons, C. P.; Naiman, B. M.; Wachtmann, T. S.; Umland, J. P.; Pandher, K.; Lapointe, J-M.; Saha, S.; Roach, M. L.; Carter, D.; Thomas, N. A.; Durtschi, B. A.; McNeish, J. D.; Hambor, J. E.; Jakobsson, P-J.; Carty, T. J.; Perez J. R.; Audoly, L. P. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9044-9049.
[4]. Kamei, D.; Yamakawa, K.; Takegoshi, Y.; Mikami-Nakanishi, M.; Nakatani, Y.; Oh-ishi, S.; Yasui, H.; Azuma, Y.; Hirasawa, N.; Ohuchi, K.; Kawaguchi, H.; Ishikawa, Y.; Ishii, T.; Uematsu, S.; Akira, S.; Murakami, M.; Kudo, I. *J. Biol. Chem.* **2004**, *279*, 33684-33695.
[5]. Mabuchi, T.; Kojima, H.; Abe, T.; Takagi, K.; Sakurai, M.; Ohmiya, Y.; Uematsu, S.; Akira, S.; Watanabe, K.; Ito, S. *Neuroreport* **2004**, *15*, 1395-1398.
[6]. Camu, F.; Shi, L.; Vanlersberghe, C. *Drugs* **2003**, *63(suppl. 1)*, 1-7.
[7]. Bianchi, M.; Broggini, M. *Drugs* **2003**, *63(suppl. 1)*, 37-46.
[8]. Jeger, R. V.; Greenberg, J. D.; Ramanathan, K.; Farkouh, M. E. *Exp. Rev. Clin. Immunol.* **2005**, *1*, 37-45.
[9]. Lee, Y.; Rodriguez, C.; Dionne, R. A. *Curr. Pharm. Des.* **2005**, *11*, 1737-1755.
[10]. Akarca, U. S. *Curr. Pharm. Des.* **2005**, *11*, 1779-1793.
[11]. Sheen, C. L.; MacDonald, T. M. *Exp. Opin. Pharmacother.* **2002**, *3*, 265-269.
[12]. Hawkey, C. J.; Jackson, L.; Harper, S. E.; Simon, T. J.; Mortensen, E.; Lines, C. R. *Aliment. Pharmacol. Ther.* **2001**, *15*, 1-9.

[13]. Bombardier, C.; Laine, L.; Reicin, A.; Shapiro, D.; Burgos-Vargas, R.; Davis, B.; Day, R.; Bosi Ferraz, M.; Hawkey, C. J.; Hochberg, M. C.; Kvien T. K.; Schnitzer, T. J.; VIGOR study group. N. *Eng. J. Med.* **2000**, *343*, 1520-1528.
[14]. Silverstein, F. E.; Faich, G.; Goldstein, J. L.; Simon, L. S.; Pincus, T.; Whelton, A.; Makuch, R.; Eisen, G.; Agrawal, N. M.; Stenson, W. F.; Burr, A. M.; Zhao, W. W.; Kent, J. D.; Lefkowith, J. B.; Verburg, K. M.; Geis, G. S. *J. Am. Med. Assoc.* **2000**, *284*, 1247-1255.
[15]. Merck & Co. Press Release; 30 September 2004. http://www.merck.com.
[16]. Mukherjee, D.; Nissen, S. E.; Topol, E. J. *J. Am. Med. Assoc.* **2001**, *286*, 954-959.
[17]. Mamdani, M.; Juurlink, D. N.; Lee, D. S.; Rochon, P. A.; Kopp, A.; Naglie, G.; Austin, P. C.; Laupacis, A.; Stukel, T. A. *Lancet* **2004**, *363*, 1751-1756.
[18]. Scheen, A. J. *Rev. Med. Liege* **2004**, *59*, 565-569.
[19]. Bresalier, R. S.; Sandier, R. S.; Quan, H.; Bolognese, J. A.; Oxenius, B.; Horgan, K.; Lines, C.; Riddell, R.; Morton, D.; Lanas, A.; Konstam, M. A.; Baron, J. A. *N. Engl. J. Med.* **2005**, *352*, 1092-1102.
[20]. Solomon, S. D.; McMurray, J. J. V.; Pfeffer, M. A.; Wittes, J.; Fowler, R.; Finn, P.; Anderson, W. F.; Zauber, A.; Hawk E.; Bertagnolli, M. *N. Engl. J. Med.* **2005**, *352*, 1071-1080.
[21]. Fitzgerald, G. A. *N. Engl. J. Med.* **2004**, *351*, 1709-1711.
[22]. Lagakos, S. W. *N. Engl. J. Med*. **2006**, *355*, 113-117.
[23]. Meagher, E. A. *Curr. Pharm. Des*. **2004**, *10*, 603-611.
[24]. Coleman, R. A.; Smith, W. L.; Narumiya, S. *Pharmacol. Rev.* **1994**, *46*, 205-229.
[25]. Breyer, R. M.; Bagdassarian, C. K.; Myers, S. A.; Breyer, M. D. *Annu. Rev. Pharmacol. Toxicol.* **2001**, *41*, 661-690.
[26]. Narumiya, S.; Sugimoto, Y.; Ushikubi, F. *Physiol. Rev.* **1999**, *79*, 1193-1226.
[27]. Coleman, R. A.; Kennedy, I.; Humphrey, P. P. A.; Bunce, K.; Lumley, P. Comprehensive Medicinal Chemistry, 3rd volume, Pergamon: Oxford, UK, 1990.
[28]. Hata, A. N.; Breyer, R. M. *Pharmacol. Ther.* **2004**, *103*, 147-166.
[29]. Ungrin, M. D.; Carrière, M-C.; Denis, D.; Lamontagne, S.; Sawyer, N.; Stocco, R.; Tremblay, N.; Metters, K. M.; Abramovitz, M. *Mol. Pharmacol*. **2001**, *59*, 1446–1455.
[30]. Hallinan, E.; Hagen, T.; Jusa, R.; Tsymbalov, S.; Rao, S.; van Hoeck, J-P.; Rafferty, M.; Stapelfeld, A.; Savage, M.; Reichman, M. *J. Med. Chem.* **1993**, *36*, 3293-3299.
[31]. Hallinan, E.; Stapelfeld, A.; Savage, M.; Reichman, M. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 509-514.
[32]. Lanthorn, T.; Bianchi, R.; Perkins, W. *Drug Dev. Res.* **1995**, *34*, 35–38.
[33]. Novel EP1 receptor antagonists. Expert Opin. Ther. Patents **2004**, *14*, 435-438.
[34]. Breault, G.A. WO9700863A1, **1997**.
[35]. Watanabe, K.; Kawamori, T.; Nakatsugi, S.; Ohta, T.; Ohuchida, S.; Yamamoto, H.; Maruyama, T.; Kondo, K.; Narumiya, S.; Sugimura, T.; Wakabayashi, K. *Cancer Lett.* **2000**, *156*, 57-61.
[36]. Ducharme, Y.; Blouin, M.; Carrière, M. C.; Chateauneuf, A.; Côte, B.; Denis, D.; Frenette, R.; Greig, G.; Kargman, S.; Lamontagne, S.; Martins, E.; Nantel, F.; O'Neill, G.; Sawyer, N.; Metters, K. M.; Friesen, R. W. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1155-1160.
[37]. Minami, T.; Nakano, H.; Kobayashi, T.; Sugimoto, Y.; Ushikubi, F.; Ichikawa, A.; Narumiya, S.; Ito, S. Br. *J. Pharmacol.* **2001**, *133*, 438-444.
[38]. Mebane, H.; Turnbach, M. E.; Randich, A. *J. Pain* **2003**, *4*, 392-399.
[39]. Stock, J. L.; Shinjo, K.; Burkhardt, J.; Roach, M.; Taniguchi, K.; Ishikawa, T.; Kim, H-S.; Flannery, P. J.; Coffman, T. M.; McNeish, J. D.; Audoly, L. P. *J. Clin. Invest.* **2001**, *107*, 325-331.
[40]. Omote, K.; Kawamata, T.; Nakayama, Y.; Kawamata, M.; Hazama, K.; Namiki, A. *Anesth. Analg.* **2001**, *92*, 233-238.
[41]. Omote, K.; Yamamoto, H.; Kawamata, T.; Nakayama, Y.; Namiki, A. *Anesth. Analg.* **2002**, *95*, 1708-1712.
[42]. Kawahara, H.; Sakamoto, T.; Takeda, S.; Onodera, H.; Imaki, J.; Ogawa, R. *Anesth. Analg.* **2001**, *93*, 1012-1017.
[43]. Giblin, G. M. P.; Bit, R. A.; Brown, S. H.; Chaignot, H. M.; Chowdhury, A.; Chessell, I. P.; Clayton, N. M.; Coleman, T.; Hall, A.; Hammond, B.; Hurst, D. N.; Michel, A. D.; Naylor, A.; Novelli, R.; Scoccitti, T.; Spalding, D.; Tang, S. P.; Wilson, A. W.; Wilson, R. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 385-389.
[44]. SYSTAT, Version 7.0; SPSS Inc., 444 North Michigan Avenue, Chicago, IL, 60611.
[45]. ChemDraw Ultra 6.0 and Chem3D Ultra, Cambridge Soft Corporation, Cambridge, USA. http://www.cambridgesoft.com
[46]. Dragon software (version 1.11-2001) by Todeschini R.; Consonni V. Milano, Italy. http://www.talete.mi.it/dragon.htm
[47]. Prabhakar, Y. S. *QSAR Comb. Sci.* **2003**, *22*, 583-595.
[48]. Wold, S. *Technometrics* **1978**, *20*, 397-405.
[49]. Kettaneh, N.; Berglund, A.; Wold, S. *Comput. Stat. Data Anal.* **2005**, *48*, 69-85.
[50]. Stahle, L.; Wold, S. Biomedical Research, Progress in Medicinal Chemistry, vol. 25, Elsevier Science Publishers, BV, 1998.
[51]. Sharma, B. K.; Pilania, P.; Singh, P.; Sharma, S.; Prabhakar, Y. S. *Cent. Eur. J. Chem.* **2009**, *7*, 909-922.
[52]. Sharma, B. K.; Pilania, P.; Singh, P. *J. Enz. Inhib. Med. Chem.* **2009**, *24*, 607-615.
[53]. Sharma, S.; Sharma, B. K.; Pilania, P.; Singh, P.; Prabhakar, Y. S. *J. Enz. Inhib. Med. Chem.* **2009**, *24*, 1024-1033.
[54]. Sharma, B. K.; Pilania, P.; Singh, P.; Prabhakar, Y. S. *SAR QSAR Environ. Res.* **2010**, *21*, 169-185.
[55]. Sharma, B. K.; Pilania, P.; Sarbhai, K.; Singh, P.; Prabhakar, Y. S. *Mol. Divers.* **2010**, *14*, 371-384.
[56]. Akaike, H. Second international symposium on information theory, Akademiai Kiado, Budapest, 1973.
[57]. Akaike, H. *IEEE Trans. Automat. Contr.* **1974**, *AC-19*, 716-723.
[58]. Kubinyi, H. *Quant. Struct.-Act. Relat.* **1994**, *13*, 285-294.
[59]. Kubinyi, H. *Quant. Struct.-Act. Relat.* **1994**, *13*, 393-401.
[60]. Friedman, J. Technical report no. 102. Laboratory for computational statistics. Stanford University, Stanford, 1990.
[61]. So, S. -S.; Karplus, M. *J. Med. Chem.* **1997**, *40*, 4347-4359.
[62]. Prabhakar, Y. S.; Solomon, V. R.; Rawal, R. K.; Gupta, M. K.; Katti, S. B. *QSAR Comb. Sci.* **2004**, *23*, 234-244.
[63]. Gramatica, P. *QSAR Comb. Sci.* **2007**, *26*, 694-701.
[64]. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Model.* **2002**, *20*, 269-276.