

Preliminary communication

QSAR analysis of 1,3-diaryl-4,5,6,7-tetrahydro-2H-isoindole derivatives as selective COX-2 inhibitors

Pratigya Silakari ^a, Savitri Devi Shrivastava ^{a,*}, Gyati Silakari ^b, Dharm Veer Kohli ^b, Gundla Rambabu ^c, Soumya Srivastava ^a, Santosh Kumar Shrivastava ^a, Om Silakari ^b^a Department of Chemistry, Dr. H.S. Gour University, Sagar 470 003, Madhya Pradesh, India^b Department of Pharmaceutical Sciences, Dr. H.S. Gour University, Sagar 470 003, Madhya Pradesh, India^c GVK BioSciences Pvt. Ltd., #210 'My Home Tycoon', 6-3-1192 Begumpet, Hyderabad 500 016, Andhra Pradesh, India

Received 7 March 2007; received in revised form 15 September 2007; accepted 27 September 2007

Available online 11 October 2007

Abstract

Quantitative structure–activity relationship (QSAR) analysis was performed on a series of 1,3-diaryl-4,5,6,7-tetrahydro-2H-isoindole for their cyclooxygenase-2 (COX-2) inhibition. QSAR investigations were based on Hansch's extra thermodynamic multi-parameter approach and receptor surface analysis (RSA). QSAR investigations reveal that steric and electrostatic interactions are primarily responsible for COX-2 enzyme–ligand interaction. QSAR model derived from Hansch analysis demonstrated that COX-2 inhibitory activity is correlated with sum of atomic polarizability (Apol), number of hydrogen-bond donor groups (HBD), energy of the highest occupied molecular orbital (HOMO), desolvation free energy for water (F_{H_2O}) and fraction of areas of molecular shadow in the XY and ZX planes over area of enclosing rectangle (Sxyf and Sxzf) with r ranges 0.870–0.904. The best model was obtained from RSA model having $r = 0.940$ with good predictive ability (predicted compounds in training set and test set within ± 1.0 unit of pIC_{50}) and can be used in designing better selective COX-2 inhibitors among the congeners in future.

© 2007 Elsevier Masson SAS. All rights reserved.

Keywords: QSAR; RSA; COX-2 inhibitors; HOMO; F_{H_2O}

1. Introduction

Non-steroidal anti-inflammatory drugs (NSAIDs) still remain among the most extensively used drugs world wide and have been used in the treatment of inflammatory conditions like rheumatoid arthritis, osteoarthritis, orthopedic injuries, post operative pain, etc. [1,2]. However, the use of conventional NSAIDs have been restricted due to their adverse effect especially gastrointestinal toxicity and renal insufficiency [3,4]. A major discovery in the search of novel anti-inflammatory agents without deleterious side effects exhibited by the conventional NSAIDs came from the identification of two different isoforms of the cyclooxygenase (COX) enzyme known as

COX-1 and COX-2. COX-1 is a constitutive isoform that is involved in normal cellular functions whereas COX-2 is an inducible isoform that is expressed only after inflammatory stimulus [5,6]. Greater deal of attention has been paid to the diaryl heterocyclic analogues as selective COX-2 inhibitors ever since the discovery of celecoxib [7] and rofecoxib [8]. Co-crystallization [9] and site directed mutagenesis studies [10] suggested that there are some non-conserved amino acid residues that are primarily responsible for selectivity of most selective COX-2 inhibitor of this class; Arg⁵¹³ and Val⁵²³ (His⁵¹³ and Ile⁵²³ in COX-1) have been demonstrated to be important. Arg⁵¹³ is a key residue for affinity through sulphone moiety. Val⁵²³ of COX-2 is smaller than Ile⁵²³ of COX-1 by a single methylene group and allows the access of side pocket, the additional binding site of most selective COX-2 inhibitors, whereas bulkier isoleucine in COX-1 blocks the access to that

* Corresponding author.

E-mail address: pratigya_silakari@rediffmail.com (S.D. Shrivastava).

side pocket. The investigation of selective COX-2 inhibitors in the treatment of colon cancer [11], Parkinson's [12] and Alzheimer's [13] diseases is the current highly interesting area of research in this therapeutic area.

Since the last decade QSAR analysis of different tricyclic class of selective COX-2 inhibitors with different heterocycles (Fig. 1) including thiazole, oxazole, furan, imidazole, iso-oxazole, pyrimidine, and thiophene as the central core has been studied [14–21]. Recently Portevin et al. [22] reported a novel series of 1,3-diaryl-4,5,6,7-tetrahydro-2*H*-isoindole derivatives as potent and selective COX-2 inhibitors in which a sulphonyl group is not a structural requisite. No QSAR work has been done to date on this class of compounds as it is the first time to introduce bicyclic ring system as central core of the diaryl heterocyclic class of selective COX-2 inhibitors. It is realized worthwhile to rationalize these compounds, in terms of physicochemical requirement.

In the present study, we have performed the quantitative structure–activity relationship analysis by conventional Hansch's extra thermodynamic multi-parameter approach [23] and receptor surface analysis (RSA) [24]. In Hansch's approach, structural features of drug molecules are quantified in terms of different parameters and these structural features are correlated to quantified biological activity through equation using regression analysis. RSA represents essential informations about hypothetical receptor site as a 3D surface with associated properties mapped onto the surface model. The location and shape of the surface represent information about the steric nature of the receptor site, while the associated properties represent other informations of interest such as hydrophobicity, partial charge, electrostatic potential and hydrogen-bonding propensity. The RSA model conveys important information in an intuitive manner and provides predictive capability for evaluating new compounds.

2. Material and methods

2.1. Data set

A data set of 30 molecules have been taken from the published results [22]. COX-2 inhibitory activity was expressed as IC_{50} values (IC_{50} values are concentration of compounds required to achieve 50% inhibition against COX-2 from mouse resident macrophages). The biological data have been converted to $-\log$ molar concentration (pIC_{50}) to reduce the skewness of the data set. It is essential to assess the predictive power of the models by using a test set of compounds. This was achieved by arbitrarily setting aside five compounds as a test set. The test set was created on the basis of suggestions by Oprea et al. [25] which are (i) the biological assay methods for both training set and test set should be compatible or same; (ii) for the test set, the biological activity values should span several times but should not exceed activity values in training set by more than 10%; (iii) the test set should represent a balanced number of both active and inactive compounds for uniform sampling of the data. The structures and COX-2 inhibition data of training set and test set are given in Table 1.

2.2. Molecular structure generation

All the molecular modeling and statistical analysis were performed using Cerius² [26] and QSAR easy software [27]. The structures of the compounds were built using molecular sketcher facilities provided in the modeling environment of Cerius². Geometric optimization was carried out using dreiding force field [28] while the partial atomic charges were calculated with Gasteiger method [29]. All the molecules were initially minimized with smart minimizer and further geometric optimization was carried out for each molecule with MOPAC 6 package using the semi-empirical AM1 Hamiltonian [30]. X-ray crystallographic structure of inhibitor SC-558 bound to the active site of COX-2 was obtained from Brookhaven Protein Data Bank (PDB entry 6COX).

2.3. Alignment

Initially all the molecules were manually aligned to most selective molecule ISOIN14 (Table 2) by considering the significant common structure (Fig. 2). Further refinement in the alignment was carried out by automated approach in Cerius². A stereo view of aligned molecules is observed in Fig. 3.

2.4. Hansch approach

The thermodynamic, spatial, electronic and topological parameters shown in Table 2 were calculated for QSAR analysis using Cerius². Thermodynamic parameters describe free energy change during drug receptor complex formation. Spatial parameters are the quantified steric features of drug molecules required for its complimentary fit with receptor. Electronic parameters describe weak non-covalent bonding between drug molecules and receptor.

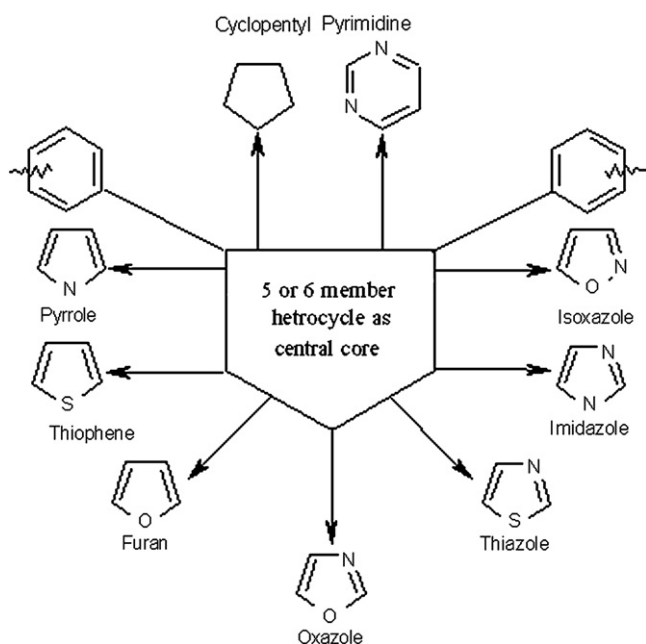
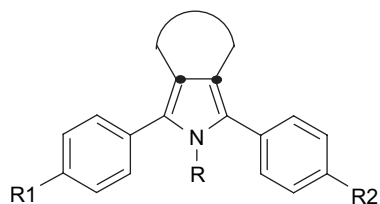


Fig. 1. Structural feature of tricyclic class of selective COX-2 inhibitors.

Table 1
Structure and COX-2 inhibition data of 1,3-diaryl-4,5,6,7-tetrahydro-2H-isoindole derivatives



Comp. No	Ring	R	R ¹	R ²	pIC ₅₀ ^a
ISOIN11		H	H	H	8.823
ISOIN12		H	H	H	8.481
ISOIN13 ^b		CH ₃	H	H	<6.000
ISOIN14		NH ₂	H	H	8.744
ISOIN15		NHSO ₂ CH ₃	H	H	6.301
ISOIN16		CH ₂ COOH	H	H	<7.000
ISOIN17		H	F	F	8.769
ISOIN18		H	CH ₃ S	CH ₃ S	6.301
ISOIN19		H	CH ₃	CH ₃	7.777
ISOIN20		H	OCH ₃	OCH ₃	7.671

(continued on next page)

Table 1 (continued)




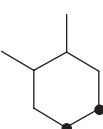
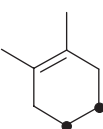
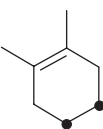


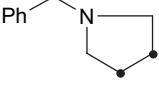
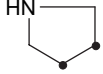








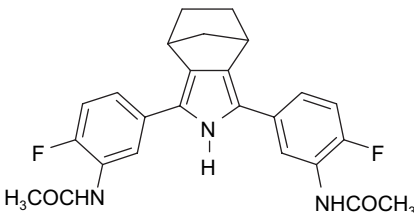
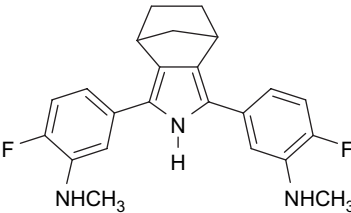
Comp. No	Ring	R	R ¹	R ²	pIC ₅₀ ^a
ISOIN21 ^b		H	Cl	Cl	8.301
ISOIN22		H	F	Imidazol-1-yl	7.376
ISOIN23 ^b		H	Imidazol-1-yl	Imidazol-1-yl	<7.00
ISOIN24		H	H	H	8.508
ISOIN25		H	H	H	7.838
ISOIN26		CH ₃	H	H	6.000
ISOIN27		H	H	H	9.154
ISOIN28		H	F	F	8.537
ISOIN29		H	H	H	<7.000
ISOIN30 ^b		H	H	H	<7.000
ISOIN31		H	H	H	8.585

Table 1 (continued)

Comp. No	Ring	R	R ¹	R ²	pIC ₅₀ ^a
ISOIN32		H	F	F	8.958
ISOIN33		H	F	F	8.346
ISOIN34		H	F	CH ₃ SO ₂	6.154
ISOIN35		H	CH ₃ SO ₂	CH ₃ SO ₂	5.000
ISOIN36		H	H	H	8.796
ISOIN37 ^b		H	F	F	9.221
ISOIN38		H	H	H	7.448
ISOIN39					<7.000
ISOIN40					<7.000

^a –log IC₅₀ in molar concentration (IC₅₀ values are concentration of compounds required to achieve 50% inhibition against COX-2 from mouse resident macrophages).

^b Compound in test set.

Table 2
List of descriptor calculated for Hansch approach

Type	Symbol	Description
Electronic	APOL	Sum of atomic polarizabilities
	DIP	Dipole moment measured in debye units
	HOMO	Energy of highest occupied molecular orbital in electron volts
	LUMO	Energy of lowest unoccupied molecular orbital in electron volts
Structural	MW	Molecular weight
	RBOND	Number of rotatable bonds
	HBA	Number of hydrogen-bond acceptor groups
	HBD	Number of hydrogen-bond donors groups
Thermodynamic	$A \log P$	log of the partition coefficient
	F_{H_2O}	Desolvation free energy for water in kcal mol ⁻¹
	F_{oct}	Desolvation free energy for 1-octanol in kcal mol ⁻¹
	MR	Molar refractivity
Spatial	ROG	Radius of gyration
	AREA	Molecular surface area
	DEN	Density in g ml ⁻¹
	PMI	Principal moment of inertia
	Vm	Molecular volume
	Wiere	The sum of the chemical bonds existing between all pairs of heavy atoms in the molecule
	Zagreb	The sum of the squares of vertex valencies
	Sxy	Area of the molecular shadow in the XY plane
	Sxz	Area of the molecular shadow in the YZ plane
	Syz	Area of the molecular shadow in the XZ plane
	Sxyf	Fraction of area of molecular shadow in the XY plane over area of enclosing rectangle
	Syzf	Fraction of area of molecular shadow in the YZ plane over area of enclosing rectangle
	Sxzf	Fraction of area of molecular shadow in the XZ plane over area of enclosing rectangle
	η	Ratio of largest to smallest dimension
	L_x	Length of molecule in the X dimension
	L_y	Length of molecule in the Y dimension
	L_z	Length of molecule in the Z dimension

Stepwise multiple regression analysis was used to generate QSAR equations. Statistical measures used were: n – number of sample in the regression, r – correlation coefficient, r^2 – squared correlation coefficient (coefficient of determination), s – standard deviation, Fischer's value calculated (F_{cal}) and tabulated (F_{tab}) for statistical significance and correlation matrix to show mutual correlation among the parameters (Table 3) [13]. Mutual correlation among the parameters present in multi-parameter containing equations should be less than 0.5.

2.5. RSA

According to the technique described by Hahn [24], hypothetical receptor surface was generated. Compounds ISOIN11,

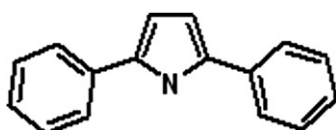


Fig. 2. Significant common structure for molecular alignment.

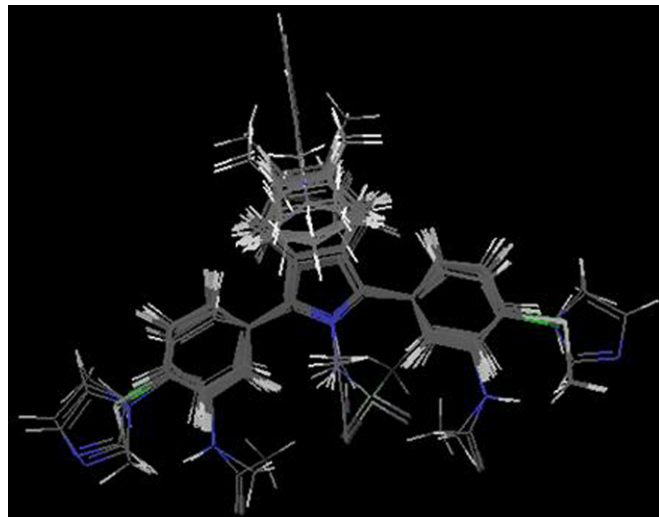


Fig. 3. A stereo view of aligned molecules.

ISOIN14, ISOIN17, ISOIN27 and ISOIN32 were selected as a template to build the receptor surface model. The interaction energy of all the molecules was evaluated within this receptor surface. The receptor surface descriptors, expressed as 3D field descriptors, were derived from Van der Waals and electrostatic interaction energies between the receptor surface and CH₃ and H⁺ groups as probes. Fig. 4 is a stereo view of receptor surface also showing the template molecules inside the receptor surface. These descriptors were used as independent variables for the generation of QSAR equations. Regression analysis was carried out using genetic function approximation (GFA) method where genetic operations have been performed over 50,000 generations. Statistical parameters used for selection of the model are number of compounds (n), adjusted correlation coefficient (r^2_{adj}), conventional correlation coefficient (r^2), predicted sum of squared (PRESS) residuals, Fischer's value calculated (F_{cal}) and tabulated (F_{tab}) for statistical significance, bootstrap correlation coefficient (r^2_{BS}), and least square error (LSE). Equations with highest r^2 , r^2_{adj} and F -test values and lowest LSE and PRESS values were considered for further discussion.

3. Results and discussion

Before using the biological data for QSAR, they should be transformed to $-\log IC_{50} \times 10^{-9}$ or $-\log$ molar dose or pIC_{50} to get all the values positive, normal distribution of errors and

Table 3
Correlation matrix for Eqs. (2)–(6)

	pIC_{50}	Apol	HBD	HOMO	F_{H_2O}	Sxyf	Sxzf
pIC_{50}	1.000						
Apol	0.870	1.000					
HBD	0.011	0.260	1.000				
HOMO	0.248	0.208	0.016	1.000			
F_{H_2O}	0.237	0.490	0.786	0.210	1.000		
Sxyf	0.371	0.460	0.156	0.120	0.220	1.000	
Sxzf	0.271	0.398	0.352	0.429	0.506	0.365	1.000

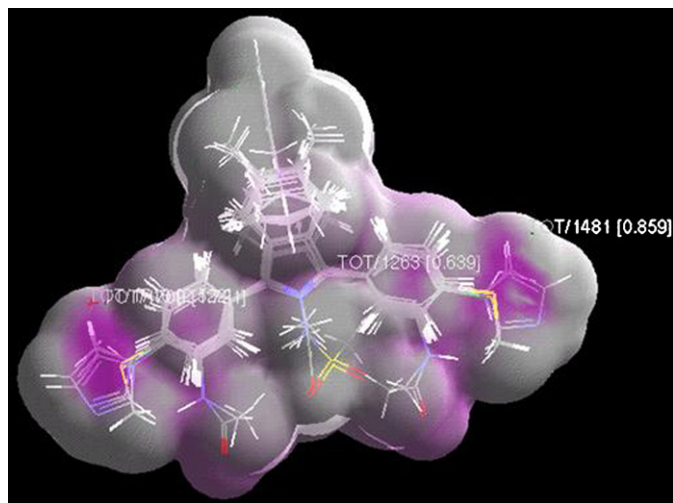


Fig. 4. Receptor surface model generated at 0.01 Å surface fit and 50% transparency level.

to get linear free energy relationship of these data with physicochemical properties.

Alignment rule, i.e., molecular conformation and orientation, is one of the most sensitive input areas for QSAR studies especially for 3D QSAR, since the relative interaction energies depend strongly on relative molecular position [31]. MOPAC minimized structures were read in Cerius² and aligned on each other using active analogue alignments by taking the most selective compound of training set as reference (ISOIN14) and on this all other molecules were aligned, i.e., they were adjusted so as to get maximum overlapping of each on the most selective one (minimum root mean square deviation).

For Hansch analysis regression was performed using pIC_{50} values as dependent variables and calculated parameters as independent variables. Out of hundreds of equations generated, some of the best QSAR equations are given below.

These equations were generated in stepwise manner by forward selection method starting with best single variable and adding further significant variable according to their contribution to the model. Initially, equations of first generation, which showed individual correlation of all calculated parameters with pIC_{50} of COX-2, were obtained. Out of these equations, only Eq. (1) that shows negative correlation with the sum of atomic polarizabilities (Apol) showed the highest correlation coefficients at the 95% level of significance ($F_{1,23 \text{ cal}} = 32.673 > F_{\alpha \text{ } 5\% \text{ } 1,23 \text{ tab}} = 4.284$). For the remaining parameters, r^2 is less as well as $F_{1,23 \text{ cal}} < F_{\alpha \text{ } 5\% \text{ } 1,23 \text{ tab}}$ and therefore they were not significant.

Eqs. (2)–(6) are the equations of second generation. Apol having the highest correlation in first generation is fixed and only those parameters, i.e., HBD, HOMO, F_{H_2O} , Sxyf and Sxzf, which have intercorrelation with Apol less than 0.5, i.e., assigned value of relaxation factor (see in Table 3) are added to Eq. (1) one by one. Addition of hydrogen-bond donor (HBD) group parameter to Eq. (1) generates Eq. (2), which shows favorable effect of HBD along with unfavorable effect of Apol on pIC_{50} of COX-2 and also has the highest overall statistics. The addition of all five parameters to Eq. (1) is 95% significant as $F_{2,22 \text{ cal}}$ for all five equations is greater than $F_{\alpha \text{ } 5\% \text{ } 2,22 \text{ tab}}$ (3.45).

In any thorough investigation of the effects of molecular properties, it is essential to prove that the results are both statistically valid and make chemical sense. Neither approach is sufficient alone. It would be appropriate to obtain insight into the physical meaning of the correlation obtained as an output of the regression analysis.

The Apol descriptor computes the sum of the atomic polarizabilities [32]. The polarizabilities are calculated from A coefficients used for molecular mechanics calculations.

$$P\alpha = \sum A_i$$

$$pIC_{50} = -0.000617 \times \text{Apol} + 16.269758$$

$$n = 25, r = 0.870, r^2 = 0.757, F_{1,23 \text{ cal}} = 32.673, F_{\alpha \text{ } 5\% \text{ } 1,23 \text{ tab}} = 4.284, \text{std} = 0.581 \quad (1)$$

$$pIC_{50} = -0.000664 \times \text{Apol} + 0.449673 \times \text{HBD} + 16.405326$$

$$n = 25, r = 0.904, r^2 = 0.817, F_{2,22 \text{ cal}} = 29.819, F_{\alpha \text{ } 5\% \text{ } 2,22 \text{ tab}} = 3.45, \text{std} = 0.516 \quad (2)$$

$$pIC_{50} = -0.000606 \times \text{Apol} - 0.327658 \times \text{HOMO} + 13.056474$$

$$n = 25, r = 0.873, r^2 = 0.761, F_{2,22 \text{ cal}} = 21.276, F_{\alpha \text{ } 5\% \text{ } 2,22 \text{ tab}} = 3.45, \text{std} = 0.590 \quad (3)$$

$$pIC_{50} = -0.000703 \times \text{Apol} - 0.063167 \times F_{H_2O} + 16.543090$$

$$n = 25, r = 0.897, r^2 = 0.804, F_{2,22 \text{ cal}} = 27.334, F_{\alpha \text{ } 5\% \text{ } 2,22 \text{ tab}} = 3.45, \text{std} = 0.535 \quad (4)$$

$$pIC_{50} = -0.000629 \times \text{Apol} - 1.443421 \times \text{Sxyf} + 17.278286$$

$$n = 25, r = 0.871, r^2 = 0.758, F_{2,22 \text{ cal}} = 20.872, F_{\alpha \text{ } 5\% \text{ } 2,22 \text{ tab}} = 3.45, \text{std} = 0.594 \quad (5)$$

$$pIC_{50} = -0.000642 \times \text{Apol} - 1.990139 \times \text{Sxzf} + 17.990429$$

$$n = 25, r = 0.874, r^2 = 0.764, F_{2,22 \text{ cal}} = 21.530, F_{\alpha \text{ } 5\% \text{ } 2,22 \text{ tab}} = 3.45, \text{std} = 0.587 \quad (6)$$

HBD is a structural descriptor that counts the number of hydrogen-bonding donor groups in the current molecule [33]. In Eq. (2), HBD shows positive correlation along with Apol. As we know that there are large number of hydrogen-bond acceptor groups like OH, NH₂, CO, etc. are present in the side chain of amino acid residues of enzyme, therefore, in some cases hydrogen bonding is the most prominent interaction for the immobilization of ligand with hydrogen-bond donor groups in the active site of enzyme. Positive contribution of HBD descriptor of these derivatives to COX-2 enzyme inhibition indicates the effectiveness of the hydrogen bonding on their interaction.

The correlation of the frontier orbital energies suggests the involvement of some electronic interactions in the process under study. HOMO (highest occupied molecular orbital) is the highest energy level in the molecule that contains electrons and LUMO (lowest unoccupied molecular orbital) is the lowest energy level in the molecule that contains no electrons. Both these energies are important in governing molecular reactivity. When a molecule acts as a Lewis base (an electron-pair donor) in bond formation, the electrons are supplied from the molecule's HOMO. How readily this occurs is reflected in the energy of the HOMO. Molecules with high HOMOs are more able to donate their electrons and hence are relatively reactive compared to molecules with low-lying HOMOs, thus the HOMO descriptor should measure the nucleophilicity of a molecule [34]. This correlation of frontier orbital energies suggests that the molecules, which are less susceptible to attack by electrophile, i.e., less reactive or more stable, would be better COX-2 inhibitors.

$F_{\text{H}_2\text{O}}$ is the aqueous desolvation free energy in kcal mol⁻¹ derived from a hydration shell model developed by Hopfinger [35]. This is a physicochemical property associated with linear free energy (LFE) models that measure absolute hydrophilicity of the molecule. LFE computations are based solely on the connectivity of the atoms in a molecule and are not conformationally dependent. This property is useful as molecular descriptor in QSAR. QSAR calculates $F_{\text{H}_2\text{O}}$ for each molecule by searching the molecule for recognizable substituent groups and their bonding patterns and summing the substituent constants' contributions for each group that is present in the molecule. Negative contribution of $F_{\text{H}_2\text{O}}$ in Eq. (4) suggests that increase in aqueous solubility would be unfavorable for COX-2 inhibition.

Sxyf and Sxzf are geometric descriptors that help to characterize the shape of the molecules. Sxyf and Sxzf are fraction of areas of molecular shadow in the XY and ZX planes, respectively, over area of enclosing rectangle. The descriptors are calculated by projecting the molecular surface on mutually perpendicular planes, XY and XZ [36]. These descriptors depend not only on the conformation but also on the orientation of the molecule. To calculate them, the molecules are first rotated to align the principle moments of inertia with the X, Y, and Z-axes. Negative effect of both these indices indicates that after substitution, orientation of the molecule should be such that the extension in the shape of molecules in XY and XZ planes as compared to reference molecule ISOIN14 is minimum for better COX-2 inhibition.

Activities of compounds in training set and test set were predicted using the best two models, i.e., Eqs. (2) and (4). Results are given in Table 4. Graphs between the predicted activities calculated from Eqs. (2) and (4) vs observed activities and the residuals of prediction vs compounds in training and test sets are shown in Figs. 5 and 6, respectively. From these graphs it can be observed that all the compounds in training set were predicted within ± 1.0 unit of the $p\text{IC}_{50}$ of COX-2. Compounds in test set were predicted with less accuracy, ± 2.0 unit of $p\text{IC}_{50}$, indicating that external predictability is lower. Although conventional statistics of these equations are significant with less external predictability, these equations indicate that steric and electronic effects of substituents might have a role to play in determining the COX-2 inhibition. The more advanced and accurate RSA approach was further considered for this data set by considering steric and electrostatic 3D field descriptors.

The receptor surface model is normally generated from the most active compounds in the data set. The rationale is that the most active molecules tend to explore the best spatial and electronic interactions with the receptor, while the least active do not tend to have unfavorable steric and electrostatic interactions. We initiated our study with receptor models' generation using the five potent and selective compounds (ISOIN11,

Table 4

Predicted activity from best equation obtained in Hansch analysis and RSA

S. No.	Comp. code	Observed $p\text{IC}_{50}$	Predicted $p\text{IC}_{50}$			Residual		
			Eq. (2)	Eq. (4)	RSA	Eq. (2)	Eq. (4)	RSA
1	ISOIN11	8.823	8.680	8.650	8.926	0.143	0.173	-0.103
2	ISOIN12	8.481	8.471	8.468	8.385	0.010	0.013	0.096
3	ISOIN13 ^a	<6.000	7.890	8.060	6.415	-1.890	-2.060	-0.415
4	ISOIN14	8.744	8.842	8.706	8.937	-0.098	0.038	-0.193
5	ISOIN15	6.301	6.445	6.926	6.120	-0.144	-0.625	0.181
6	ISOIN16	<7.000	7.468	7.630	6.671	-0.468	-0.630	0.329
7	ISOIN17	8.769	8.531	8.482	8.871	0.238	0.287	-0.102
8	ISOIN18	6.301	5.697	5.516	6.733	0.604	0.785	-0.432
9	ISOIN19	7.777	7.934	7.849	7.534	-0.157	-0.072	0.243
10	ISOIN20	7.671	7.477	7.838	7.834	0.194	-0.167	-0.163
11	ISOIN21 ^a	8.301	6.635	6.483	8.289	1.666	1.818	0.012
12	ISOIN22	7.376	6.802	7.065	7.365	0.574	0.311	0.011
13	ISOIN23 ^a	<7.000	5.074	5.648	6.800	1.926	1.352	0.200
14	ISOIN24	8.508	7.998	7.906	8.504	0.510	0.602	0.004
15	ISOIN25	7.838	7.789	7.695	7.903	0.049	0.143	-0.065
16	ISOIN26	6.000	6.999	7.104	6.343	-0.999	-1.104	-0.343
17	ISOIN27	9.154	9.021	9.024	8.487	0.133	0.130	0.667
18	ISOIN28	8.537	8.872	8.856	9.100	-0.335	-0.319	-0.563
19	ISOIN29	<7.000	6.315	6.450	6.737	0.685	0.550	0.263
20	ISOIN30 ^a	<7.000	9.524	9.469	7.854	-2.524	-2.469	-0.854
21	ISOIN31	8.585	8.433	8.448	7.882	0.152	0.137	0.703
22	ISOIN32	8.958	8.283	8.279	8.257	0.675	0.679	0.701
23	ISOIN33	8.346	8.075	8.097	8.203	0.271	0.249	0.143
24	ISOIN34	6.154	6.866	6.796	6.533	-0.712	-0.642	-0.379
25	ISOIN35	5.000	5.449	5.313	4.784	-0.449	-0.313	0.216
26	ISOIN36	8.796	8.092	8.074	8.032	0.129	0.147	0.189
27	ISOIN37 ^a	9.221	7.943	7.905	8.954	1.278	1.316	0.267
28	ISOIN38	7.448	7.883	7.892	7.957	-0.435	-0.444	-0.509
29	ISOIN39	<7.000	6.577	6.361	6.725	0.423	0.639	0.275
30	ISOIN40	<7.000	7.862	7.608	7.378	-0.862	-0.608	-0.378

^a Compound in test set.

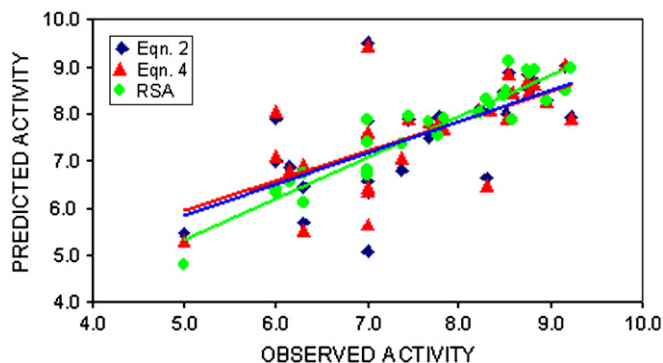


Fig. 5. Graph of observed vs predicted activity from Eqs. (2) and (4) and RSA.

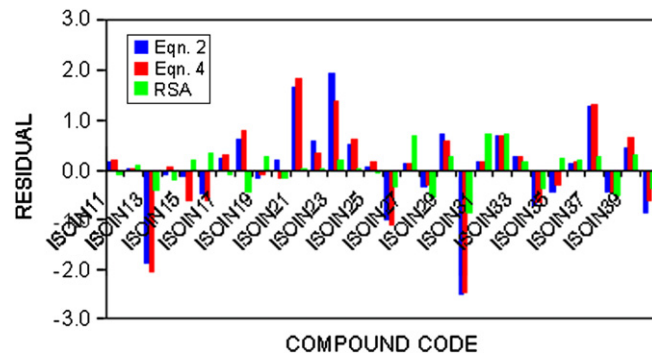


Fig. 6. Graph of compound vs residual from Eqs. (2) and (4) and RSA.

ISOIN14, ISOIN17, ISOIN27 and ISOIN32) as template molecules. The receptor surface was generated at 0.01 Å surface fit and 50% transparency level. After the receptor surface model has been generated, all the structures in the training and test sets can be evaluated against the model. The model can be used to calculate the energy associated with the binding of a molecule in the model. It can also be used to minimize a molecule by adjusting the geometry of the structure into a “best-fit configuration” based on the constraints imposed by the receptor model. All the interaction energies and pIC_{50} of the derivatives were subjected to regression by GFA. The GFA algorithm approach has a number of important advantages over the other techniques: (i) it builds multiple models rather than a single model; (ii) it automatically selects which features are to be used in the models; (iii) it is better at discovering combinations of features that take advantage of correlations between multiple features; (iv) it incorporates Friedman’s lack-of-fit (LOF) error measure, which estimates the most appropriate number of features, resists over fitting, and allows control over the smoothness of fit; (v) it can use a larger variety of equation term types in construction of its models (for example, step functions or high-order polynomials); (vi) it provides thorough study of the evolving models, additional information not available from standard regression analysis (such as the preferred model length and useful partitions of the data set). The following equation was obtained as the best equation in the population of 100 equations:

$$\begin{aligned}
 pIC_{50} = & 6.687 \times (TOT/1481) + 0.790 \\
 & \times (InterVDWEnergy) - 11.0154 \times (TOT/1710) + 4.321 \\
 & \times (TOT/1263) + 10.411 \times (TOT/1709) + 12.254 \\
 n = & 25, LOF = 0.123, r = 0.940, r^2 = 0.883, \\
 r^2_{adj} = & 0.859, F\text{-test} = 36.311, \\
 PRESS = & 6.464, LSE = 0.139
 \end{aligned} \quad (7)$$

The descriptors TOT/1481, TOT/1710, etc. are added energy of both electrostatic and van der Waals interaction energies at points 1481, 1710, etc. Descriptor InterVDWEnergy is non-bond van der Waals energy between molecule and receptor. The above equation explains 88.3% variance in the activity with respect to steric and electrostatic interaction

energies between receptor points and molecules in data set. The statistical measures, r^2 , LSE and F , determine the estimation power of model for the same data from which it has been determined and evaluate it only internally. On the other hand, cross-validated parameters, LOF, r^2_{adj} and PRESS, determine the prediction power for the data not included in deriving the model and evaluate the model externally to avoid chance of correlation completely. It can be observed that the overall statistics of the equation generated in RSA are excellent and their prediction ability is also significant (Fig. 6). From this graph it can be seen that compounds in training set and test set were predicted within ± 1.0 unit of pIC_{50} .

The receptor surface model represents essential information about the hypothetical receptor site as a 3D surface with associated properties mapped onto the surface model. The location and magnitude of a descriptor can be used as a guideline to improve the COX-2 inhibitory activity of molecules. The surface represents information about the steric nature of the receptor site and the associated properties of interest, such as hydrophobicity, partial charge, and hydrogen-bonding propensity. Interaction energies’ correlations were mapped on the surface of model. Fig. 4 is a stereo view of the receptor surface also showing the molecule in training set. The magenta color represents negative energy values as favorable interaction sites, while the green-colored regions represent positive energy values that are unfavorable sites for binding of the molecule on the receptor surface. (For interpretation of the references to color in this text, the reader is referred to the web version of this article.) This color-coding anticipates the characteristic structural features of the active site of COX-2 where the derivatives are being accommodated. This information can offer a qualitative way of examining compounds, by introducing them into the virtual receptor and visually inspecting the favorable/unfavorable interactions; substituents that increase or decrease the binding affinity can be easily recognized, and one can make easily simple but accurate structure–activity estimations.

To understand the interactions between amino acid residues and molecules under study, the most potent and selective molecule ISOIN14 (grey in color) was superimposed on the highly selective COX-2 inhibitor SC-558 (yellow in color) in the active site of tricyclic inhibitor (Fig. 7A and 7B). The active site

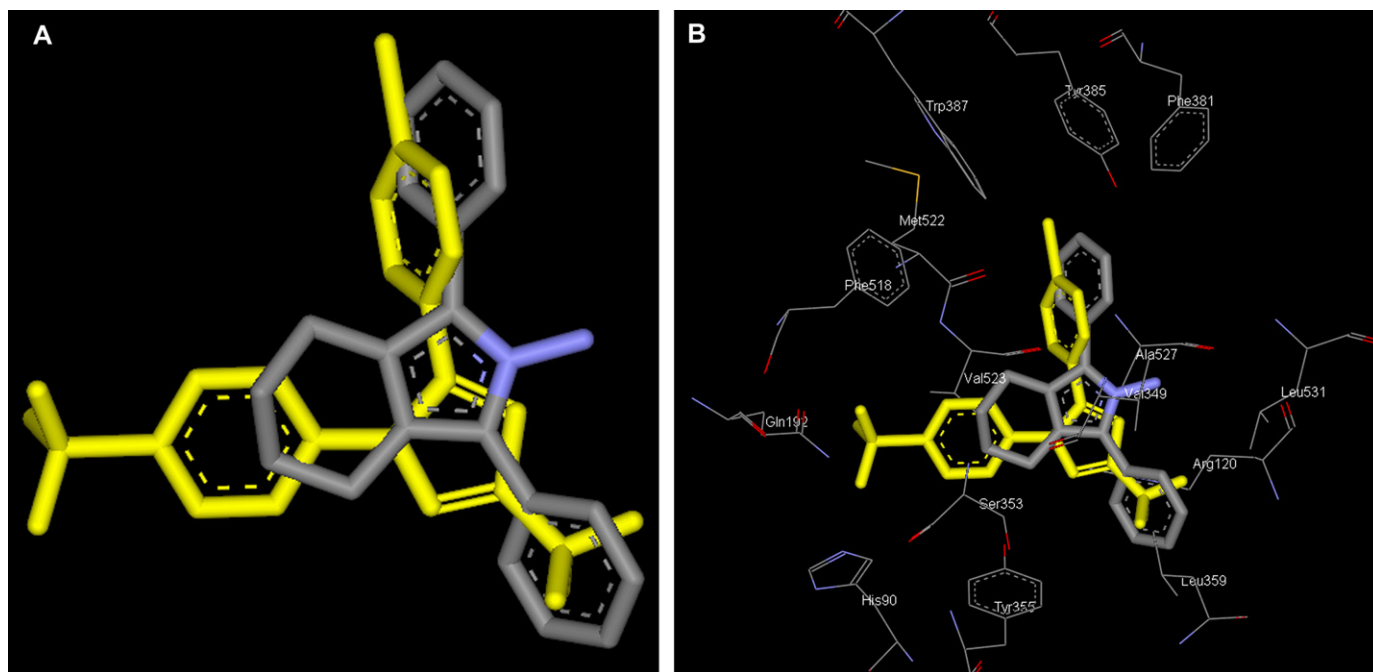


Fig. 7. Superimposition of ISOIN14 (in grey color) on highly selective inhibitor SC-558 (in yellow color): (A) without active site and (B) with active site of COX-2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was viewed within 5 Å areas. The comparison of the orientation of molecule under study with the orientation of SC-558 in the active site shows that substituent NH_2 at R of ISOIN14 is being occupied in the region, which is surrounded by hydrogen-bond acceptor groups (red 'O' and blue 'N' colored atom of Ala⁵²⁷, Leu⁵³¹, and Arg¹²⁰). This might be the region of correlation of HBD parameter with $p\text{IC}_{50}$ of COX-2 in Hansch approach. Two aryl rings of ISOIN14 overlap with 5-aryl ring and 3-trifluoromethyl group of pyrazole in SC-558, respectively. Alicyclic ring fused with centre core of pyrrole is so oriented that it is making perfect hydrophobic interaction with Val⁵²³ which is known to be important for selectivity COX-2 [9]. This cavity marked by Val⁵²³ and Ser³⁵³ residues is completely filled by alicyclic ring of ISOIN14 and aromatic ring containing sulfonamide of SC-558. These docking results further substantiate our earlier results correlating Sxyf and Sxzf to COX-2 inhibition. The information obtained in this manner can be used in designing new ligand molecules for better complementarity fit with COX-2.

4. Conclusions

In summary, from the derived QSAR model it can be concluded that selective COX-2 inhibition by the 1,3-diaryl-4,5,6,7-tetrahydro-2H-isoindole derivatives is strongly influenced by the steric and electrostatic nature of substituents. Pattern of substitution can be extracted from the developed model. Consequently this study may prove to be helpful in development and optimization of existing selective COX-2 inhibitors of this class of compounds.

Acknowledgement

Authors wish to thank The Director, GVK Biocampus Balanagar, Hyderabad, for providing molecular modeling facilities. Authors are also thankful to all faculty members of GVK BioSciences Balanagar, Hyderabad for their help and support to perform various computational and modeling steps.

References

- [1] W.L. Smith, *Annu. Rev. Physiol.* 48 (1986) 251–262.
- [2] J.R. Vane, Y.S. Bakhle, R.M. Botting, *Annu. Rev. Pharmacol. Toxicol.* 38 (1998) 97–120.
- [3] H. Herschman, *Biochim. Biophys. Acta* 1299 (1996) 125–140.
- [4] D. Munroe, C.Y. Lau, *Chem. Biol.* 2 (1995) 343–350.
- [5] W. Xie, J.G. Chipmann, D.L. Robertson, R.L. Erikson, D.L. Simmons, *Proc. Natl. Acad. Sci. U.S.A.* 88 (1991) 2692–2696.
- [6] W.L. Smith, R.M. Garavito, D.L. DeWitt, *J. Biol. Chem.* 271 (1996) 33157–33160.
- [7] T.D. Penning, J.J. Talley, S.R. Bertenshaw, J.S. Carter, P.W. Collins, S. Docter, M.J. Graneto, L.F. Lee, J.W. Malecha, J.M. Miyashiro, R.S. Rogers, D.J. Rogier, S.S. Yu, G.D. Anderson, E.G. Burton, J.N. Cogburn, S.A. Gregory, C.M. Koboldt, W.E. Perkins, K. Seibert, A.W. Veenhuizen, Y.Y. Zhang, P.C. Isakson, *J. Med. Chem.* 40 (1997) 1347–1365.
- [8] L.A. Sorbera, P.A. Leeson, J. Castañer, *Drugs Future* 23 (1998) 1287–1296.
- [9] R.G. Kurumball, A.M. Stevens, J.K. Gierse, J.J. McDonald, R.A. Stegeman, J.Y. Pak, D. Gildehaus, J.M. Miyashiro, T.D. Penning, K. Seibert, P.C. Isakson, W.C. Stallings, *Nature* 384 (1996) 644–648.
- [10] E. Wong, C. Bayly, H.L. Watermann, D. Riendeau, J.A. Mancini, *J. Biol. Chem.* 272 (1997) 9280–9286.
- [11] M. Oshima, J.E. Dinchuk, S.L. Kargman, H. Oshima, B. Hancock, E. Kwong, J.M. Trzaskos, J.F. Evans, M.M. Taketo, *Cell* 87 (1996) 803–809.

- [12] P. Teismann, K. Tieu, D.K. Choi, D.C. Wu, A. Naini, S. Hunot, M. Vila, V. Jackson-Lewis, S. Przedborski, *Proc. Natl. Acad. Sci. U.S.A.* 100 (2003) 5473–5478.
- [13] G.M. Pasinetti, *Neurosignals* 11 (2002) 293–307.
- [14] P. Chavatte, A. Farce, *Anti-cancer Agents Med. Chem.* 6 (2006) 239–249.
- [15] S. Shahapurkar, N. Kawathekar, S.C. Chaturvedi, *Pharmazie* 60 (2005) 254–258.
- [16] S. Prasanna, E. Manivannan, S.C. Chaturvedi, *Bioorg. Med. Chem. Lett.* 15 (2005) 2097–2102.
- [17] S. Shahapurkar, T. Pandya, N. Kawathekar, S.C. Chaturvedi, *Eur. J. Med. Chem.* 39 (2004) 899–904.
- [18] S. Prasanna, E. Manivannan, S.C. Chaturvedi, *Bioorg. Med. Chem. Lett.* 14 (2004) 4005–4011.
- [19] H.X. Liu, R.S. Zhang, X.J. Yao, M.C. Liu, Z.D. Hu, B.T. Fan, *J. Comput. Aided Mol. Des.* 18 (2004) 389–399.
- [20] P. Chavatte, S. Yous, C. Marot, N. Baurin, D. Lesieur, *J. Med. Chem.* 44 (2001) 3223–3230.
- [21] C. Almansa, A.F. de Arriba, F.L. Cavalcanti, L.A. Gomez, A. Miralles, M. Merlos, J. Garcia-Rafanell, J. Forn, *J. Med. Chem.* 44 (2001) 350–361.
- [22] B. Portevin, C. Tordjman, P. Pastoureaux, J. Bonnet, G.D. Nanteuil, *J. Med. Chem.* 43 (2000) 4582–4593.
- [23] C. Hansch, T. Fujita, *J. Am. Chem. Soc.* 86 (1964) 1616–1626.
- [24] M. Hahn, *J. Med. Chem.* 38 (1995) 2080–2090.
- [25] T.J. Oprea, G.L. Waller, G.R. Marshall, *J. Med. Chem.* 37 (1994) 2206–2215.
- [26] Cerius² Molecular Modelling Program Package, Molecular Simulations (Accelrys) Inc., San Diego, CA, 92121–93752.
- [27] QSAR Easy software was developed in S.G.S.I.T.S. Indore (M.P.), India.
- [28] S.L. Mayo, B.D. Olafson, I. Goddard, *J. Phys. Chem.* 94 (1990) 8897–8909.
- [29] J. Gasteiger, M. Marsili, *Tetrahedron* 36 (1980) 3219–3288.
- [30] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, *J. Am. Chem. Soc.* 107 (1985) 3902–3909.
- [31] R.D. Cramer III, D.E. Patterson, J.D. Bunce, *J. Am. Chem. Soc.* 110 (1988) 5959–5967.
- [32] M. Marsili, J. Gasteiger, *Croat. Chem. Acta* 53 (1980) 601–605.
- [33] C. Hansch, A. Vittoria, C. Silipo, P.Y.C. Jow, *J. Med. Chem.* 18 (1975) 546–548.
- [34] N. Bodor, Z. Gabanyi, C.K. Wong, *J. Am. Chem. Soc.* 111 (1989) 3783–3786.
- [35] A.J. Hopfinger, *Conformational Properties of Macromolecules*, Academic Press, New York, 1973.
- [36] R.H. Rohrbaugh, P.C. Jurs, *Anal. Chim. Acta* 9 (1987) 99–109.