Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 1297–1303
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Exploration of physicochemical properties and molecular modelling studies of 2-sulfonyl-phenyl-3-phenyl-indole analogs as cyclooxygenase-2 inhibitors

Arun Kumar Gupta, Revathi A. Gupta, Love Kumar Soni, S.G. Kaskhedikar*

Molecular Modelling Study Group, CADD Laboratory, Computational Chemistry Research, Department of Pharmacy, Shri G. S. Institute of Technology and Science, 23 Park Road, Indore 452003, Madhya Pradesh, India

Received 12 April 2007; received in revised form 24 June 2007; accepted 28 June 2007
Available online 12 July 2007

Abstract

In the present work, modelling study has been performed to explore the physicochemical requirements of 2-sulfonyl-phenyl-3-phenyl-indole analogs as COX-2 enzyme inhibitors. The multivariant regression expressions were developed using sequential multiple linear regression (SEQ-MLR) technique, considering adjustable correlation coefficient (r^2_{adj}). The statistical quality of SEQ-MLR equations was evaluated considering parameters like correlation coefficient (r), standard error of estimation (SEE), and variance ratio (F) at explicit degree of freedom (df). Orthogonality of the descriptors in SEQ-MLR was established through variance inflation factor (VIF). Developed equations have been internally validated using leave-one-out technique and further validated with test set, considering predictive squared correlation coefficient (r^2_{pred}). The orientation of the most potent and selective COX-2 inhibitor of training set, 2-(4-phenyl sulfonamide)-3-phenyl-5-methylindole, in the COX-2 active site was explored by docking. The phenyl sulfonamide moiety positioned in secondary pocket of enzyme which consists of amino acid residues Phe_{518}, Gln_{192}, Arg_{513}, Leu_{352}, Ser_{353} and Val_{523} is responsible for the selectivity. The unsubstituted phenyl ring positions in a hydrophobic cavity are lined by Tyr_{385}, Trp_{387}, Tyr_{348}, Leu_{384} and Met_{522}. Interestingly, the indole C-5 CH_3-substituent is located in a hydrophobic region formed by Ile_{345}, Val_{349}, Ala_{527}, Leu_{531} and Leu_{534}. The hydrophobic interactions of methyl group might be crucial for the potency of 2-sulfonyl-phenyl-3-phenyl-indole analogs. Study has revealed that atomic van der Waals volume and atomic masses explain COX-2 inhibitory activity of 2-sulfonyl-phenyl-3-phenyl-indole analogs significantly.

Keywords: COX-2 inhibitor; QSAR; SEQ-MLR; Docking; 2-Sulfonyl-phenyl-3-phenyl-indole analogs

1. Introduction

NSAIDs are among the most widely used prescription and over the counter medications primarily for the treatment of pain and inflammation particularly in arthritis. The puzzle of NSAIDs' action was resolved by the elegantly simple experiment of Sir John Vane [1], who in 1971 demonstrated that aspirin and other NSAIDs could prevent the synthesis of the prostaglandins. Needleman and Isakson [2] postulated two isoenzymes of cyclooxygenase; ''House Keeping'' enzyme (COX-1) which is responsible for a basic level of PGs and an ''Inducible enzyme'' (COX-2) which is activated by different stimuli mediating inflammatory reactions. Several works have been carried out on cyclooxygenase pathway inhibition, leading to a development of specific COX-2 inhibitors. A most promising approach seemed to be the designing of novel NSAIDs targeted at the inducible isoform COX-2 that appear to be devoid of gastrointestinal toxicity in that they spare mucosal prostaglandin synthesis [3].

In the commencement of our work in selective COX-2 inhibitor, we were particularly interested in investigating structure–activity relationship of diaryl-substituted bicyclic heterocycles [4–10] to establish whether these agents would

* Corresponding author. Tel./fax: +91 0731 2546031.
E-mail address: arunkg_73@hotmail.com (S.G. Kaskhedikar).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2007.06.022

demonstrate increased selectivity for the COX-2 isoform. However, these agents have more efficient occupation on the larger active site in the COX-2 isoform in compare to COX-1.

The major objective of this study is to explore the physicochemical properties which are helpful in the designing of selective COX-2 inhibitors with better efficacy and reduced toxicity. The objective can be fulfilled by structural requirement which is explored through QSAR and docking study and then exploited to optimize activity of compounds of selected series.

QSAR has been traditionally perceived as a means of establishing correlation between trend in the chemical structure modifications and respective changes of biological activity [11]. A part of our efforts to create QSAR models shows substantial predictive promise for the designing of new compounds with enhanced biological activity. In the present work, we correlated the COX-2 inhibitory activity of substituted 2-sulfonyl-phenyl-3-phenyl-indoles reported by Hu et al. [12] (Table 1).

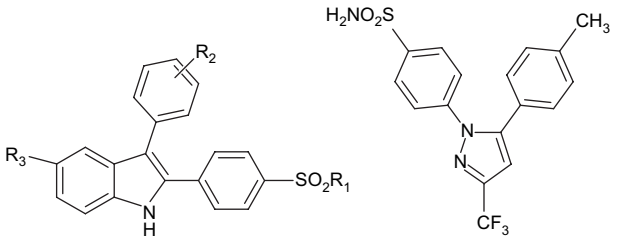
2. Experimental

In QSAR study, the logarithmic form of depending data set was considered which is having less skewness as compared to the non-logarithmic one. The inhibitory concentration (IC_{50} in nM) of COX-2 was converted into pIC_{50} (negative logarithm of IC_{50} in mole) used as a dependent variable. The series was divided into a training set of 22 compounds including TR-1 to TR-22 (Table 1), and a test set of 9 compounds including T-1 to T-9 (Table 1), on the basis of structural diversity and variation in inhibitory activity.

The molecular modelling study was performed using HyperChem [13], CS ChemOffice [14], and Dragon [15] program while the regression analysis was carried out on VALSTAT [16]. Structures of all compounds were sketched using builder module of the program. The sketched structures were subjected to energy minimization via steepest descent method using OPLS force field until the RMS gradient value become smaller than 0.1 kcal/mol Å. The energy minimized molecules have been subjected to re-optimization via Austin model 1 (AM1) [17–20] method until the RMS gradient attained a value lesser than 0.0001 kcal/mol Å using MOPAC. The geometry optimization of the lowest energy structure was carried out using EF routine. The descriptor values for all the molecules were calculated using ''compute properties'' module of the program. The minimized molecule was saved as MOL file format. These files were used for calculation of various physicochemical properties with the help of Dragon.

Correlation was established by sequential multiple linear regression (SEQ-MLR) technique. In SEQ-MLR, the program searches all the permutation/combination for the given data set sequentially. The statistical quality of the SEQ-MLR equations were assessed by parameters like explained variance (r^2_{adj}), correlation coefficient (r), standard error of estimate (SEE), and variance ratio (F) at specified degree of freedom (df). The generated QSAR equations validated by predicted residual sum of squares (PRESS), (leave-one-out or loo) [21,22] cross-validation squared correlation coefficient (Q^2), standard deviation based on PRESS (S_{PRESS}) and standard deviation of error of prediction (S_{DEP}). Finally, selected equations have been validated using test set considering predictive squared correlation coefficient (r^2_{pred}). The ± data within the parenthesis is the standard deviation associated with the coefficient of descriptor in regression equation.

Table 1
Structure and activity of 2-sulfonyl-phenyl-3-phenyl-indole analogs used in QSAR analysis



Comp. no.	R_1	R_2	R_3	IC_{50} in nM
TR-1	NH_2	H	H	0.09
TR-2	CH_3	H	H	0.60
TR-3	NH_2	4-F	H	5.15
TR-4	NH_2	4-Cl	H	33.5
TR-5	CH_3	4-Br	H	0.37
TR-6	NH_2	4-CH_3	H	0.07
TR-7	CH_3	4-CH_3	H	0.09
TR-8	NH_2	3,4-$(CH_3)_2$	H	1.46
TR-9	NHAc	H	H	0.18
TR-10	NH_2	4-Cl	Cl	0.54
TR-11	CH_3	4-Cl	Cl	85.1
TR-12	CH_3	3-Cl	Cl	0.8
TR-13	NH_2	2-Cl	Cl	100.0
TR-14	CH_3	2-Cl	Cl	10.0
TR-15	NH_2	H	5-F	2.0
TR-16	NH_2	H	5-Cl	0.14
TR-17	CH_3	H	5-Cl	0.36
TR-18	NH_2	H	5-Br	5.0
TR-19	CH_3	H	5-Br	1.41
TR-20	NH_2	H	5-CH_3	0.02
TR-21	CH_3	H	5-CH_3	0.28
TR-22(Cel)	–	–	–	0.52
T-1	CH_3	4-F	H	0.02
T-2	CH_3	4-Cl	H	146.0
T-3	NH_2	4-Br	H	8.36
T-4	NH_2	4-OCH_3	H	0.006
T-5	CH_3	4-OCH_3	H	0.02
T-6	CH_3	3,4-$(CH_3)_2$	H	0.17
T-7	NH_2	4-OH	H	6.44
T-8	NH_2	3-Cl	Cl	3.13
T-9	CH_3	H	5-F	5.0

The orientation and interactions of the 2-sulfonyl-phenyl-3-phenyl-indole analogs with COX-2 enzyme were examined by ArgusLab docking program [23]. For docking study 6cox protein database (PDB) file was considered. The coordinate of 6cox PDB file was taken from Brookhaven Protein Database (http://www.rcsb.org/pdb). Protein structure is prepared by considering protonation state while ligand structure is generated by adding hydrogen followed by energy minimization. ArgusDock engine default parameters were considered for the docking study.

3. Result and discussion

In the present study, QSAR analysis has been performed to explore structural requirement for inhibitory activity of 2-sulfonylphenyl-3-phenyl-indole analogs against COX-2 enzyme using Hansch approach. The multivariant expressions were developed on the basis of adjustable correlation coefficient (r^2_{adj}). This parameter explains statistical significance of incorporated physicochemical descriptors in SEQ-MLR. r^2_{adj} takes into account of adjustment of conventional correlation coefficient (r^2). If r^2_{adj} value decline by the addition of a physicochemical descriptor to the equation it is indicated that descriptor was not contributed fairly. Adjustable correlation coefficient is a measure of % explained variation of regression expression. Whereas r^2 always increases when an independent variable is added. Study has furnished uni and bi-variant expression with moderate correlation coefficient (Eqs. (1) and (2)) but the r^2_{adj} value is increasing significantly from uni to bi-variant expression.

$$\begin{aligned} \mathrm{pIC}_{50} &= 1.839(\pm 0.526)Mor07v + 1.546 \\ n &= 22,\ r = 0.616,\ r^2_{adj} = 0.349,\ \mathrm{SEE} = 0.803, \\ F &= 12.235 \end{aligned} \quad (1)$$

$$\begin{aligned} \mathrm{pIC}_{50} &= 0.031(\pm 0.012)\mathrm{ISIZ} - 0.301(\pm 0.078)RDF080m \\ &\quad + 5.467 \\ n &= 22,\ r = 0.722,\ r^2_{adj} = 0.471,\ \mathrm{SEE} = 0.724, \\ F &= 10.339 \end{aligned} \quad (2)$$

Significant improvement in r^2_{adj} value emphasizes to explore the higher variant expressions. Proposed models should have to satisfy both statistical quality and predictive power. Therefore, all the expressions were tested for internal and external corroboration. Both the validations put forward decision-making input for selection of QSAR models. Internal corroboration was carried out using leave-one-out cross-validation method, bootstrapping technique and randomized biological activity test while external corroboration confirmed with the help of test set data. Tri-variant expressions (Eqs. (3) and (4)) which fulfill all the corroboration criteria up to significant echelon were considered as QSAR model 1 and 2, respectively.

$$\begin{aligned} \mathrm{pIC}_{50} &= 0.184(\pm 0.032)RDF085u \\ &\quad - 0.392(\pm 0.056)RDF080m \\ &\quad + 0.368(\pm 0.131)RDF090v + 10.224 \\ n &= 22,\ r = 0.893,\ r^2_{adj} = 0.763,\ \mathrm{SEE} = 0.484, \\ F &= 23.594 \end{aligned} \quad (3)$$

$$\begin{aligned} \mathrm{pIC}_{50} &= 0.227(\pm 0.038)RDF085u \\ &\quad - 0.391(\pm 0.059)RDF080m \\ &\quad - 3.402(\pm 1.382)Mor08v + 5.302 \\ n &= 22,\ r = 0.884,\ r^2_{adj} = 0.746,\ \mathrm{SEE} = 0.502, \\ F &= 21.481 \end{aligned} \quad (4)$$

Model 1 has a correlation coefficient ($r = 0.893$), which accounts for more than 76.3% of the explained variance in the activity, calculated as $r^2_{adj} = r^2(1 - 1/F)$ that accounts in percentage when multiplied by 100 (Table 2). Model revealed that the dependent variable can be predicted from a linear combination of the independent variables. The P value is less than 0.05 for each physiochemical parameters involved in model generation. The data show an overall internal statistical significance level better than 99.9% as calculated variance ratio i.e. Fischer value (F) exceeded the tabulated $F_{(3,18\alpha 0.001)} = 9.69$. Fischer value suggested that the equations are applicable for more than 999 out of 1000 times. The orthogonality of the descriptors in the model was established through variance inflation factor (VIF) [24,25] values (Table 3) and inter-correlation among the descriptors (ICAP) (Table 4). The VIF is defined as $1/(1 - r_i^2)$, where r_i is the multiple correlation coefficient for the ith variable regressed on the $p - 1$ others, p is being the number of variables contributed to the model. VIF value larger than 10 indicates that the information of the descriptors may be hidden by the correlation of the other descriptors [26]. VIF is less than 1.2 for all the contributing descriptors revealed that the descriptors are fairly independent to each other. The low value of ICAP ($< \sim 0.350$) also supported comparatively independent contribution. We have also made efforts to investigate predictive power of the proposed model by using quality factor (QF) using Pogliani's method [27,28]. QF is defined as the ratio of correlation coefficient to standard error of estimation (SEE). The larger value of QF (1.845) signifies better predictive power of the model. For reliability of the model, we have calculated regression associated statistical parameter called probable error of correlation (PE) [29]. Goodness of fit is calculated as $PE = 2(1 - r^2)/3\sqrt{n}$, if the value of correlation coefficient (r) is more than six times of PE than the

Table 2
Statistical parameters of models 1 and 2

Statistical parameters	Model 1	Model 2
r	0.893	0.884
r^2_{adj}	0.763	0.746
SEE	0.484	0.502
F	23.594	21.481
PE	0.029	0.031
QF	1.845	1.761
ICAP	<0.352	<0.432
r^2_{bs}	0.821	0.794
S_{bs}	0.110	0.126
Chance	<0.001	<0.001
Q^2	0.693	0.670
S_{PRESS}	0.595	0.617
S_{DEP}	0.538	0.558
r^2_{pred}	0.442	0.379

Table 3
t-Values and VIF values of the descriptors used in QSAR models

Intercept/descriptors	Model 1		Model 2	
	t-Value	VIF	*t*-Value	VIF
RDF080m	6.983	1.170	6.642	1.194
RDF090v	2.813	1.142	–	–
RDF085u	5.714	1.026	5.909	1.360
MoRSE08v	–	–	2.462	1.433

expression is good and reliable. In model 1 the value of correlation coefficient is significantly higher then 6PE supporting reliability and goodness.

The model was further analyzed for the outlier by the Z-score method (*Z*-value), the outlier test helps in the identification of unexplainable structurally diverse analogs. The persuasive QSAR model should not have any outlier. The *Z*-value for individual compounds lies within the specific range ($<|2.5|$), which indicated the absence of outliers. Test revealed that the model is able to explain the structurally diverse analogs and is helpful in the designing of more potent compounds using physiochemical descriptors.

The chance of fortuitous correlation is checked with the help of randomized biological activity test (Chance), which is evaluated as ratio of the equivalent regression equations to the total number of randomized sets. Chance value of 0.001 corresponds to 0.1% chance of fortuitous correlation. Model 1 having Chance value less than 0.001 suggested that the result was not based on prospective correlation.

Bootstrapping analysis was performed for further access of the robustness and statistical confidence. The bootstrapping analysis gives an overview about contribution of individual molecules to the QSAR model. The r^2_{bs} is the average squared correlation coefficient calculated during the validation procedure which is computed from a subset of compounds used one at a time for the validation procedure. S_{bs} is the standard deviation in multiple run of a given data set. Bootstrapping correlation coefficient is at par to conventional r^2 and S_{bs} is low, then the model is robust and promising. In this study both the values ($r^2_{bs} = 0.821$ and $S_{bs} = 0.110$) fall within the agreement.

Internal predictivity of the model 1 was assured with the help of cross-validated constraints like Q^2, S_{PRESS} and S_{DEP} obtained by 'leave-one-out (loo)' method. This model was built by $N-1$ compounds and the *N*th compound considered as a predicted (Table 5 and Fig. 1). The value of $Q^2 > 0.5$ is the basic requirement for declaring a model to be a valid one [30]. The internal consistency of the model supported by $Q^2 = 0.693$, $S_{PRESS} = 0.595$ and $S_{DEP} = 0.538$ values (Table 2).

Although equation shows good internal consistency, they may not be applicable for the analogs which were never used in the generation of the correlation. Therefore, the external extrapolation power of the equation was further authenticated by a test set of nine compounds. The value of predictive squared correlation coefficient ($r^2_{pred} = 0.442$) of test set supported robustness, predictiveness and applicability of the model (Table 6 and Fig. 2). In general the model fulfills the statistical validation criteria to a significant extent.

The result and discussion made above indicate that the model is positively contributed by *RDF085u* and *RDF090v* while negatively contributed by *RDF080m*. RDF descriptors belonging to the class of radial distribution function descriptors [31–33] are based on the distance distribution in the geometrical representation of the molecule. In addition to interatomic distances in the entire molecule, the RDF also provides valuable information about bond distances, ring types, planar and non-planar systems, atom types and other important structural motifs.

The RDF code has been proven to be a good representation for the 3D structure which has several merits like independence from the number of atoms; unambiguity regarding the three-dimensional arrangement of the atoms and invariance against translation and rotation of the entire molecule.

The RDF of an ensemble of *N* atoms can be interpreted as the probability distribution to find an atom in a spherical volume of radius *r*. The RDF used in this work is as follows:

$$g(r) = f\sum_{1}^{N-1}\sum_{j>i}^{N} A_i A_j e^{-B\left(r-r_{ij}\right)^2}$$

$$f = 1\Big/\sqrt{\sum_{r}[g(r)]^2}$$

where *f* is a scaling factor, *N* is the number of atoms, *A* is the atomic propertie of atoms *i* and *j*, *B* is smoothing parameter that defines the probability distribution of the individual distances, r_{ij} is the distance between the atoms *i* and *j*, $g(r)$ was calculated at a number of discrete points with defined intervals.

Each molecule was represented by a vector of length 32. The parameter *B* was set to 25 $Å^{-2}$ corresponding to a total resolution of 0.2 Å in the defined distance *r*. The RDF for the structure derivations was calculated with the atomic properties. *RDF090v* is the radial distribution function at 9.0 Å interatomic distance weighted by atomic van der Waals volumes contributed positively. The contribution of *RDF090v* revealed that van der Waals volume is crucial for the interaction and it might be responsible for the interaction with hydrophobic pocket of the COX-2 enzyme. *RDF085u* is the radial distribution function at 8.5 Å interatomic distance unweighted and contributed positively. *RDF080m* is the radial distribution function at 8.0 Å interatomic distance weighted by atomic mass and contributed negatively.

Table 4
Inter-correlation matrix of physicochemical properties used in QSAR analysis of COX-2 inhibitory activity

	RDF085u	*RDF080m*	*RDF090v*	*Mor07v*	*Mor08v*	*ISIZ*
RDF085u	1.000					
RDF080m	0.153	1.000				
RDF090v	0.016	0.351	1.000			
Mor07v	0.254	0.730	0.240	1.000		
Mor08v	0.432	0.270	0.550	0.414	1.000	
ISIZ	0.554	0.048	0.254	0.490	0.301	1.000

Table 5
Observed, calculated and calculated (loo) pIC_{50} values with Z-score and residual of 2-sulfonyl-phenyl-3-phenyl-indole analogs used in QSAR analysis for COX-2 inhibitory activity for model 1 and model 2

Comp. no.	Obs[a] (pIC_{50})	Model 1 (pIC_{50})					Model 2 (pIC_{50})				
		Cal[b]	Cal_{res}[c]	Z-value	Cal(loo)[d]	$Cal(loo)_{res}$[e]	Cal[b]	Cal_{res}[c]	Z-value	Cal(loo)[d]	$Cal(loo)_{res}$[e]
TR-1	10.046	9.244	0.802	1.790	9.109	0.937	9.002	1.044	2.247	8.819	1.227
TR-2	9.222	9.711	−0.490	−1.093	9.806	−0.584	9.568	−0.346	−0.744	9.679	−0.457
TR-3	8.288	8.849	−0.561	−1.252	8.940	−0.652	8.874	−0.586	−1.260	8.981	−0.693
TR-4	7.475	8.299	−0.824	−1.839	8.603	−1.128	8.011	−0.536	−1.154	8.121	−0.646
TR-5	9.432	9.081	0.351	0.783	9.062	0.369	8.669	0.762	1.641	8.547	0.885
TR-6	10.155	10.145	0.010	0.022	10.141	0.014	9.988	0.167	0.360	9.933	0.222
TR-7	10.046	9.944	0.102	0.228	9.896	0.150	10.100	−0.054	−0.116	10.126	−0.080
TR-8	8.836	8.919	−0.083	−0.186	8.943	−0.107	8.614	0.221	0.476	8.564	0.272
TR-9	9.745	9.741	0.004	0.008	9.741	0.004	9.465	0.280	0.602	9.439	0.306
TR-10	9.268	8.264	1.004	2.242	7.915	1.352	8.244	1.024	2.203	7.866	1.401
TR-11	7.070	7.246	−0.176	−0.393	7.309	−0.239	7.333	−0.263	−0.566	7.422	−0.352
TR-12	9.097	8.738	0.359	0.801	8.716	0.381	9.316	−0.219	−0.471	9.366	−0.269
TR-13	7.000	7.160	−0.160	−0.356	7.220	−0.220	7.130	−0.130	−0.281	7.188	−0.188
TR-14	8.000	7.824	0.176	0.392	7.770	0.230	8.016	−0.016	−0.035	8.019	−0.019
TR-15	8.699	8.869	−0.170	−0.380	8.891	−0.192	9.063	−0.364	−0.783	9.121	−0.422
TR-16	9.854	9.661	0.193	0.431	9.622	0.232	9.882	−0.028	−0.060	9.890	−0.036
TR-17	9.444	9.276	0.167	0.374	9.249	0.195	9.626	−0.182	−0.391	9.646	−0.203
TR-18	8.301	9.127	−0.826	−1.845	9.186	−0.885	8.899	−0.598	−1.288	8.970	−0.669
TR-19	8.851	8.494	0.357	0.797	8.447	0.404	8.754	0.096	0.207	8.748	0.102
TR-20	10.699	10.647	0.052	0.116	10.630	0.069	10.497	0.202	0.435	10.445	0.254
TR-21	9.553	9.806	−0.253	−0.565	9.831	−0.278	9.583	−0.031	−0.066	9.588	−0.035
TR-22	9.284	9.317	−0.033	−0.075	9.330	−0.046	9.728	−0.444	−0.956	9.821	−0.537

[a] Observed data of the compounds used in generation of model.
[b] Calculated data of the compounds using model.
[c] Residual value of calculated data.
[d] Calculated (loo) data of the compounds using leave-one-out method.
[e] Residual value of calculated (loo) data.

The aforementioned discussion indicated that the regression and statistical parameters are good enough to establish an equation as model 1. However, the single model is not enough for the factual extrapolation of analogs therefore another model was explored from the remaining QSAR expression. Eq. (4) was considered as model 2. Model 2 shows correlation coefficient at par with model 1 ($r = 0.884$). Model 2 accounts for more than 74.6% of the explained variance in the activity. The linear contribution of each physicochemical parameter to the model was significant by more than 95.0% ($P < 0.05$). The orthogonality of the descriptors (VIF) in model 2 is in agreement with the limit. Reliability of the model was supported by the lower value of probable error of correlation ($r > 6PE$). The *Z*-value for individual compounds lies within the specific range ($<|2.5|$), which indicated the absence of outliers. The results of chance statistics and bootstrapping analysis are in conformity. Model 2 has somewhat

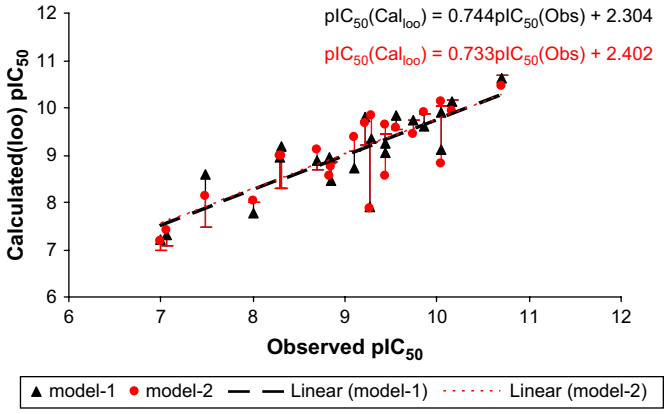


Fig. 1. Plot of calculated (loo) pIC_{50} against observed pIC_{50} of training set for models 1 and 2.

Table 6
Observed and predicted pIC_{50} values with residual of test set of 2-sulfonyl-phenyl-3-phenyl-indole analogs obtained by extrapolation through model 1 and model 2

Comp. no.	Obs[a] (pIC_{50})	Model 1 (pIC_{50})		Model 2 (pIC_{50})	
		Pred[b]	$Pred_{res}$[c]	Pred[b]	$Pred_{res}$[c]
T-1	10.699	9.375	1.324	9.408	1.291
T-2	6.836	8.728	−1.892	8.575	−1.739
T-3	8.078	8.997	−0.920	8.064	0.013
T-4	11.222	10.831	0.391	9.916	1.306
T-5	10.699	11.011	−0.312	10.113	0.586
T-6	9.770	9.156	0.614	9.330	0.440
T-7	8.191	9.708	−1.517	9.971	−1.780
T-8	8.504	8.167	0.338	8.278	0.226
T-9	8.301	9.219	−0.918	9.324	−1.023

[a] Observed data of the compounds used in test set.
[b] Predicted data of test set compounds.
[c] Residual value of predicted data.

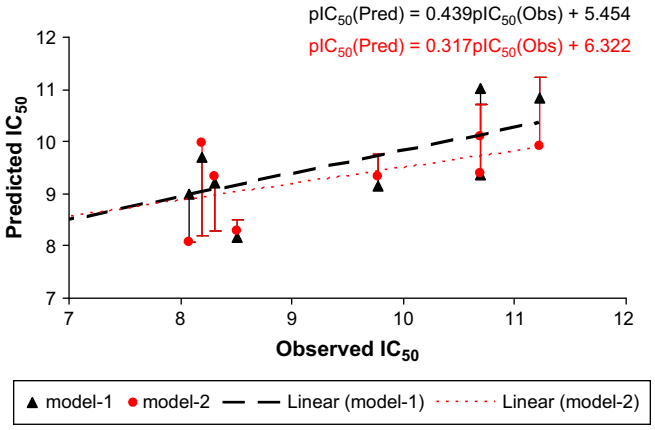


Fig. 2. Plot of predicted pIC_{50} against observed pIC_{50} of test set for models 1 and 2.

lower internal consistency (Fig. 1 and Table 5) and predictive ability of test set r^2_{pred} (0.379) as compared to those of model 1 (Fig. 2 and Table 6). Model 2 also fulfills the statistical criteria but the level is somewhat low as compare to model 1. However, amalgamation of both models put forwards useful theoretical base for proposing more active compounds.

From abovementioned result and discussion model 2 is positively contributed by *RDF085u* while negatively contributed by *RDF080m* and *Mor08v*.

Mor08v, [34–37] is 3D molecular representation of structure based on electron diffraction code (MoRSE code). MoRSE code was calculated by summing atom weights viewed by a different angular scattering function. The values of these code functions were calculated at 32 evenly distributed values of scattering angle(s) in the range of 0–31 $Å^{-1}$ from the three-dimensional atomic coordinates of a molecule. The 3D-MoRSE code was calculated using following expression:

$$I(s) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} A_i A_j \frac{\sin sr_{ij}}{sr_{ij}}$$

where s is scattering angle, r_{ij} is the interatomic distance of ith and jth atom.

A_i and A_j are atomic properties of ith and jth atom, respectively, including van der Waals volume, atomic mass, Sanderson atomic electronegativity and atomic polarizability. The negative contribution of *Mor08v* revealed that sum of the properties calculated for the atoms (weighted by van der Waals volume) from the three-dimensional atomic coordinates of a molecule is decisive for explaining the enzyme–ligand interaction. *Mor08v* might be responsible for change in orientation of the ligand at binding site therefore extent of the macromolecule–ligand interaction reduced.

The role of atomic van der Waals volume and atomic masses was further examined with the help of docking. The orientation of the most potent and selective COX-2 inhibitor of training set, 2-(4-phenyl sulfonamide)-3-phenyl-5-methylindole (TR-20), was anticipated by docking (Fig. 3) in the

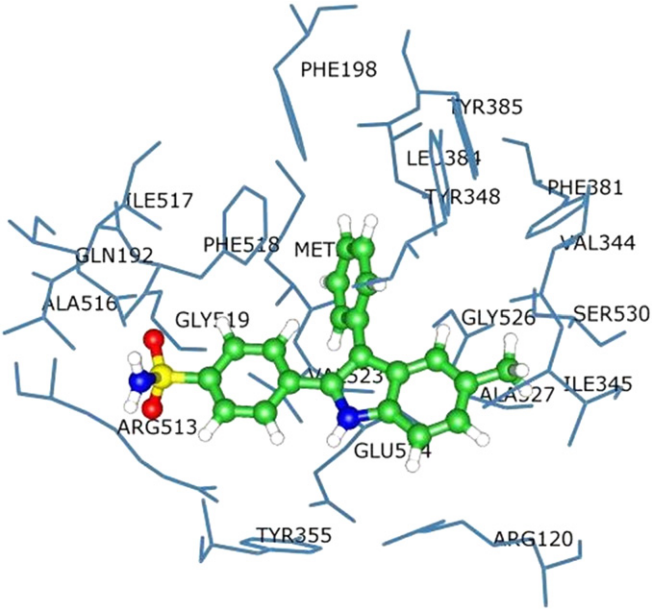


Fig. 3. Docking of 2-(4-phenyl-sulfonamide)-3-phenyl-5-methylindole on the active site of COX-2 enzyme (stick). Hydrogen atoms of amino acid residues are removed to increase the clarity.

COX-2 active site. This study showed that TR-20 binds in the center of the primary binding site of COX-2 with the SO_2NH_2 moiety interacting with the secondary pocket amino acid residues Phe_{518}, Gln_{192}, Arg_{513}, Leu_{352}, Ser_{353} and Val_{523}. One of the O-atoms of the SO_2NH_2 substituent forms a hydrogen bond with the amide hydrogen of Phe_{518} (1.978 Å). One of the H-atom of SO_2NH_2 substituent forms a hydrogen bond with the side chain of OE1 of Gln_{192}. The unsubstituted phenyl ring positions in a hydrophobic cavity are lined by Tyr_{385}, Trp_{387}, Tyr_{348}, Leu_{384} and Met_{522}. Incredibly, *p*-methyl or *p*-methoxy substituted phenyl ring analogs are more potent; probably these substitution enhanced hydrophobic interactions. Interestingly, the indole C-5 CH_3-substituent is located in a hydrophobic region formed by Ile_{345}, Val_{349}, Ala_{527}, Leu_{531} and Leu_{534}. The hydrophobic interactions of methyl group might be crucial for the potency of 2-sulfonyl-phenyl-3-phenyl-indole analogs although C-5 chloro-atom forms weak hydrogen bond with the OH of Ser_{530}. This study shows that appropriately substituted 2-sulfonyl-phenyl-3-phenyl-indole analogs interact favorably within the COX-2 binding site.

SEQ-MLR analysis reveals that RDF code and MoRSE code depend on atomic van der Waals volume, atomic masses and atomic unweighted properties that explains COX-2 inhibitory activity significantly. This study suggested that these descriptors could be explored efficiently for the development of new analogs for COX-2 inhibition from the virtual library.

Acknowledgment

Authors are thankful to the Director, Shri G. S. Institute of Technology and Science, Indore for providing facilities to

complete this work. The author AKG is thankful to CSIR, New Delhi, for providing SRF.

References

[1] J.R. Vane, Nature 231 (1971) 232–235.
[2] P. Needleman, P.C. Isakson, J. Rheumatol. Suppl. 49 (1997) 6–8.
[3] S. Revathi, A.K. Gupta, L.K. Soni, S. Kavitha, R. Wagh, S.G. Kaskhedikar, Indian J. Chem. 45B (2006) 2763–2769.
[4] H.C. Huang, T.S. Chamberlain, K. Seibert, C.M. Koboldt, P.C. Isakson, D.B. Reitz, Bioorg. Med. Chem. Lett. 5 (1995) 2377–2380.
[5] M. Therien, C. Brideau, C.C. Chan, W.A. Cromlish, J.Y. Gauthier, R. Gordon, G. Greig, S. Kargman, C.K. Lau, Y. Leblanc, C.S. Li, G.P. O'Neill, D. Riendeau, P. Roy, Z. Wang, L. Xu, P. Prasit, Bioorg. Med. Chem. Lett. 7 (1997) 47–52.
[6] P. Roy, Y. Leblanc, R.G. Ball, C. Brideau, C.C. Chan, N. Chauret, W. Cromlish, D. Ethier, J.Y. Gauthier, R. Gordon, G. Greig, J. Guay, S. Kargman, C.K. Lau, G. O'Neill, J. Silva, M. Therien, C.V. Staden, E. Wong, L. Xu, P. Prasit, Bioorg. Med. Chem. Lett. 7 (1997) 57–62.
[7] D.J.P. Pinto, D.G. Batt, W.J. Pitts, J.J. Petraitis, M.J. Orwat, S. Wang, J.W. Jetter, S.R. Sherk, G.C. Houghton, R.A. Copeland, M.B. Covington, J.M. Trzaskos, R.L. Magolda, Bioorg. Med. Chem. Lett. 9 (1999) 919–924.
[8] B. Portvein, C. Tordjman, P. Postoureau, J. Bonnet, G. De Nanteuil, J. Med. Chem. 43 (2000) 4582–4593.
[9] 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.
[10] A. Palomer, F. Cabre, J. Pascual, J. Campos, M.A. Trujillo, A. Entrena, M.A. Gallo, L. Garcia, D. Mauleon, A. Espinosa, J. Med. Chem. 45 (2002) 1402–1411.
[11] X.J. Yao, J.P. Panaye, J.P. Doucet, R.S. Zhang, H.F. Chen, M.C. Liu, Z.D. Hu, B.T. Fan, J. Chem. Inf. Comput. Sci. 44 (2004) 1257–1266.
[12] W. Hu, Z. Guo, F. Chu, A. Bai, X. Yi, G. Cheng, J. Li, Bioorg. Med. Chem. 11 (2003) 1153–1160.
[13] HyperChem 7.52, Demo version, Hypercube, Inc. 1115 NW 4th Street, Gainesville, FL 32601 USA, http://www.hyper.com/.
[14] CS ChemOffice, Version 8.0, Cambridge Soft Corporation, Software Publishers Association, 1730 M Street, Suite 700, Washington D.C. 20036 (202) 452-1600, U.S.A.
[15] R. Todeschini, V. Consonni, DRAGON-Software for the Calculation of Molecular Descriptors, rel. 1.12 for Windows (2001).
[16] A.K. Gupta, B.M. Arockia, S.G. Kaskhedikar, Indian J. Pharm. Sci. 66 (2004) 396–402.
[17] M.J.S. Dewar, E.G. Zoebisch, E.F. Healey, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902–3909.
[18] M.J.S. Dewar, C.H. Hwang, D.R. Kuhn, J. Am. Chem. Soc. 113 (1991) 735–741.
[19] P.U. Civcir, J. Mol. Struct: THEOCHEM 535 (2001) 121–129.
[20] L.B. Kier, Molecular Orbital Theory in Drug Research, Academic Press, New York, 1971, pp. 1–62.
[21] K.J. Schaper, Quant. Struct.–Act. Relat. 18 (1999) 354–360.
[22] S. Wold, L. Eriksson, Chemometric Methods in Molecular Design, in: H. van de Waterbeemd (Ed.), VCH, Weinheim, 1995, p. 321.
[23] ArgusLab 4.0.1, Mark A. Thompson, Planaria Software LLC, Seattle, WA, http://www.arguslab.com.
[24] S. Chatterjee, A. Hadi, B. Price, Regression Analysis by Examples, third ed. Wiley-VCH, New York, 2000.
[25] S. Shapiro, B. Guggenheim, Quant. Struct. Act. Relat. 17 (1998) 327–337.
[26] D.H. Cho, S.K. Lee, B.T. Kim, K.T. No, Bull. Korean Chem. Soc. 22 (2001) 388–394.
[27] L. Pogliani, Amino Acids 6 (1994) 141–153.
[28] L. Pogliani, J. Phys. Chem. 100 (1996) 18065–18077.
[29] D. Mandloi, S. Joshi, P.V. Khadikar, N. Khosla, Bioorg. Med. Chem. Lett. 15 (2005) 405–411.
[30] A. Golbraikh, A. Tropsha, J. Mol. Graphics 20 (2002) 269–276.
[31] M.C. Hemmer, V. Steinhauer, J. Gasteiger, Vib. Spectrosc. 19 (1999) 151–164.
[32] A. Fedorowicz, L. Zheng, H. Singh, E. Demchuk, Int. J. Mol. Sci. 5 (2004) 56–66.
[33] M.C. Hemmer, J. Gasteiger, Anal. Chim. Acta 420 (2000) 145–154.
[34] J. Gasteiger, J. Sadowski, P. Schuur, L. Selzer, V. Steinhauer, J. Steinhauer, J. Chem. Inf. Comput. Sci. 36 (1996) 1030–1037.
[35] J.H. Schuur, P. Selzer, J. Gasteiger, J. Chem. Inf. Comput. Sci. 36 (1996) 334–344.
[36] J. Schuur, J. Gasteiger, J. Anal. Chem. 69 (1997) 2398–2405.
[37] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.