J Mol Model (2008) 14:1087–1099
DOI 10.1007/s00894-008-0350-8

ORIGINAL PAPER

Molecular modeling study on chemically diverse series of cyclooxygenase-2 selective inhibitors: generation of predictive pharmacophore model using Catalyst

Madhu Chopra · Ruby Gupta · Swati Gupta · Daman Saluja

Received: 14 February 2008 / Accepted: 4 July 2008 / Published online: 30 July 2008

Abstract Cyclooxygenase (COX) enzymes catalyse the biosynthesis of prostaglandins and thromboxane from arachidonic acid (AA). We summarize in this paper, the development of pharmacophores of a dataset of inhibitors for COX-2 by using the Catalyst/Hypogen module using six chemically diverse series of compounds. Training set consisting of 24 compounds was carefully selected. The activity spread of the training set molecules was from 0.1 to 10000 nM. The most predictive pharmacophore model (hypothesis 1), consisting of four features, namely, one hydrogen bond donor, one hydrogen bond acceptor, one hydrophobic aliphatic and one ring aromatic feature, had a correlation (r) of 0.954 and a root mean square deviation of 0.894. The entropy (configuration cost) value of the hypotheses was 16.79, within the allowed range. The difference between the null hypothesis and the fixed cost and between the null hypothesis and the total cost of the best hypothesis (hypothesis 1) was 88.37 and 78.51, respectively. The model was validated on a test set consisting of six different series of structurally diverse 22 compounds and performed well in classifying active and inactive molecules correctly. This validation approach provides confidence in the utility of the predictive pharmacophore model developed in this work as a 3D query tool in the virtual screening of drug like molecules to retrieve new chemical entities as potent COX-2 inhibitors. The model can also be used to predict the biological activities of compounds prior to their costly and time-consuming synthesis.

M. Chopra (✉) · R. Gupta · S. Gupta · D. Saluja
Laboratory of Molecular Modeling and Drug Design, Dr. B. R. Ambedkar Center for Biomedical Research,
University of Delhi, Delhi, India
e-mail: mchopradu@gmail.com

Keywords Catalyst · Cyclooxygenase · Hypogen · Pharmacophore · 3D-QSAR · NSAIDS · Cyclooxygenase-2 inhibitors

Introduction

Cyclooxygenase (COX) enzymes catalyse the biosynthesis of prostaglandins and thromboxane from arachidonic acid (AA). Cyclooxygenation of AA results in the formation of prostaglandinG_2 (PGG_2) [1]. PGG_2 is the precursor to numerous prostaglandins including those that possess analgesic, antipyretic and anti-inflammatory activity and those that provide protection for the gastric mucosa [2]. Until recently, a single COX enzyme was thought to be responsible for whole of the catalysis of AA to PGG_2, it was in 1990s that a second isoform of COX termed COX-2 was discovered [3]. This discovery made clear the existence of COX as its two isoforms, *i.e.*, COX-1 and COX-2 with individual modes of expression. COX-1 being constitutive in most cells perform housekeeping functions (gastrointestinal tolerability, keeping vascular and renal homeostasis) whereas COX-2 is inducible, induced in pathological conditions by pro-inflammatory cytokines and playing a major role in inflammation, pain and fever [4]. Therefore adverse gastrointestinal (GI) effects of traditional non-steroidal anti-inflammatory drugs (NSAIDS), e.g., aspirin, ibuprofen and naproxen would be largely due to COX-1 inhibition [5] and hence clarifies our objective to develop COX-2 selective inhibitors.

Selective COX-2 inhibitors are expected to play vital roles in ovulation and labor, as well as in the treatment of colon cancer and Alzheimer's disease [6–8]. In addition, there is a growing evidence that COX-2 contributes to carcinogenesis [9]. A broad range of laboratory investiga-

tions, animal models, and epidemiological studies provide convincing evidence of that inhibition of the COX-2 pathway might have significant implication for the treatment of colorectal cancer. As the inducible COX-2 isoform is over expressed in colorectal tissues and is associated with critical events of tumorigenesis [1]. COX-2 expression correlates with expression of angiogenic factors and new blood vessel formation. Inhibition of COX-2 favors apoptosis causes a dose dependent decline of tumor growth and metastasis in these models [1].

Currently, there is an ongoing research to design a new COX-2 inhibitors structurally different from the current ones. Indeed, recently Rofecoxib, from the diarylheterocycle family, was withdrawn from the market because of cardiotoxicity [10] which justifies that a constant effort is devoted to identify new scaffolds for the COX-2 inhibition. The present work applies pharmacophore development method on six chemically diverse series of compounds to give a hypothesis. **Pharmacophore** was first defined by Paul Ehlirch in 1909 as "a molecular framework that carries (*phoros*) the essential features responsible for a drug's (=*pharmacon's*) biological activity" [11]. In 1977, this definition was updated by Peter Gund to "a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity" [12]. The active molecule(s) are called active ingredients of a drug. Alternatively pharmacophore is described as an ensemble of interactive functional groups with a defined geometry.

A pharmacophore can help medicinal chemists visualise the potential interaction between ligands and receptor, it can be used as a query in a 3D database search to identify new structural classes of potential lead compounds and it can serve as a template for generating alignments for 3D-QSAR analysis. Thus the construction of an accurate pharmacophore is a key objective in many drug discovery efforts.

A drug discovery cycle, to identify, to optimize and eventually take a compound to the market is generally a long process (approx 12–15 years) and is very expensive (approx $500million R&D expense) [13]. The enormous pressure that pharmaceutical and biotech companies are facing, has created the need to apply all available techniques to decrease attrition rates, costs and the time to market [14]. Pharmacophore identification is now increasingly being handled by automated computational methods, e.g., Catalyst, GASP and DISCO are three commercially available programs [15–17]. We have used Catalyst [18] in this paper. The Catalyst software operates in 2 modes: HipHop and HypoGen [18]. These have been successfully used in drug discovery research [19–44] and toxicology.

Catalyst/HipHop generates pharmacophore hypothesis using actives only. Catalyst/HypoGen takes activity data into account and uses both active and inactive compounds in an attempt to identify hypothesis that are common among the active compounds but not among the inactives. Kaminski et al. reported the development of pharmacophore models from a series of farnesyl protein transferase (FPT) inhibitors [23]. The best derived pharmacophore was used to search a 3D database from Schering -Plough Research Institute and successfully identified several low micromolar FPT inhibitors with varied structures compared to the structures used in the training set to develop the pharmacophore.

Sprague [24] used this method in developing pharmacophores for inhibitors against angiotensin converting enzymes, protein farnesyl transferase, human immunodeficiency virus (HIV) protease, and HIV reverse transcriptase. Recently, Kurogi and Guner have used Catalyst®/HipHop generated pharmacophores in searching 3D databases to identify novel mesangial cell proliferation inhibitors [40]. These studies suggest that the Catalyst- generated pharmacophores can be effectively used for rational drug design.

We in our previous paper have developed pharmacophore hypotheses [45] for six structurally diverse series of cholecystokinin - B/Gastrin receptor antagonists. The most predictive model consisted of four features, namely, two hydrogen bond donors, one hydrophobic aliphatic and one hydrophobic aromatic feature. The pharmacophore could predict and discriminate the activities of all the compounds and even could explain the difference in activities of those highly active which were active in subnanomolar range over the others.

Palomer et al. identified three feature pharmacophore model from the 3D structure of four diarylheterocycles with knowledge of X-ray crystal structure of SC-558: two aromatic rings and a sulphonyl moiety [46, 47]. From the superposition of indomethacin, a non-selective inhibitor to this pharmacophore, it was observed that despite the absence of the sulphonyl group, this inhibitor reasonably maps the model. Using the mapping of the new compounds on to the three feature pharmacophoric model, analogues of indomethacin with sulphonyl group were developed.

Michaux et al. used Catalyst/HipHop software in order to identify new potential lead compounds by generating pharmacophore model of 16 diverse and highly COX-2 selective inhibitors [4]. They developed a four point pharmacophore for COX-2 selective inhibitors using a training set of 16 compounds, using catalyast/HipHop module. The best hypothesis (h1) consisted of one Hydrogen bond acceptor, two hydrophobic groups and one aromatic ring. The pharmacophore was combined with exclusion volume spheres representing important residues of the COX-2 binding site and was used to virtually screen the Maybridge database. The resultant hit compounds were assayed for pharmacological activity against COX activity to provide initial lead compounds. The authors used four

classes of compounds viz., the diaryl hetrocycles, the sulphonamide derivatives, the analogues of flurbiprofen and the analogues of indomethacin. The pharmacophore was different from other generated three point [48–50] or have a second H-bond acceptor as in the hypothesis h5 [51–54].

We summarize in this paper, the development of pharmacophores of a dataset of inhibitors for COX-2 by using the Catalyst/ HypoGen module using six chemically diverse series of compounds. The training set includes compounds from viz., 1,2 diarylimadazole class, 1,5 diarylpyrazole class, acyclic 2-alkyl-1,2 diaryl (E) olefins, Aryl-substituted methyleneaminoxymethyl (MAOM), analogues of Diarylcyclopentenyl COX-2 inhibitors, 4''-(4-Cycloalkyl/aryl-oxazol-5yl)benzene sulphonamides, heteroaromatic analogues of (2-aryl-1-cyclopentenyl-1-alkylidene)-(arylmethyloxy) amine. There is, so far, no report on developing pharmacophore for COX-2 selective inhibitors using Catalyst/Hypogen module from all these different classes of compounds. This study is expected to provide useful knowledge for developing new drugs targeted to Cyclooxygenase-2.

Materials and methods

Molecular Modeling All molecular modeling works were performed on a Silicon Graphics Octane 2 computer running Irix 64 6.5, 600 MHz (SGI, 1600 Amphitheatre Parkway, Mountain View, CA 94043). Catalyst 4.11 software [30] was used to generate pharmacophore models.

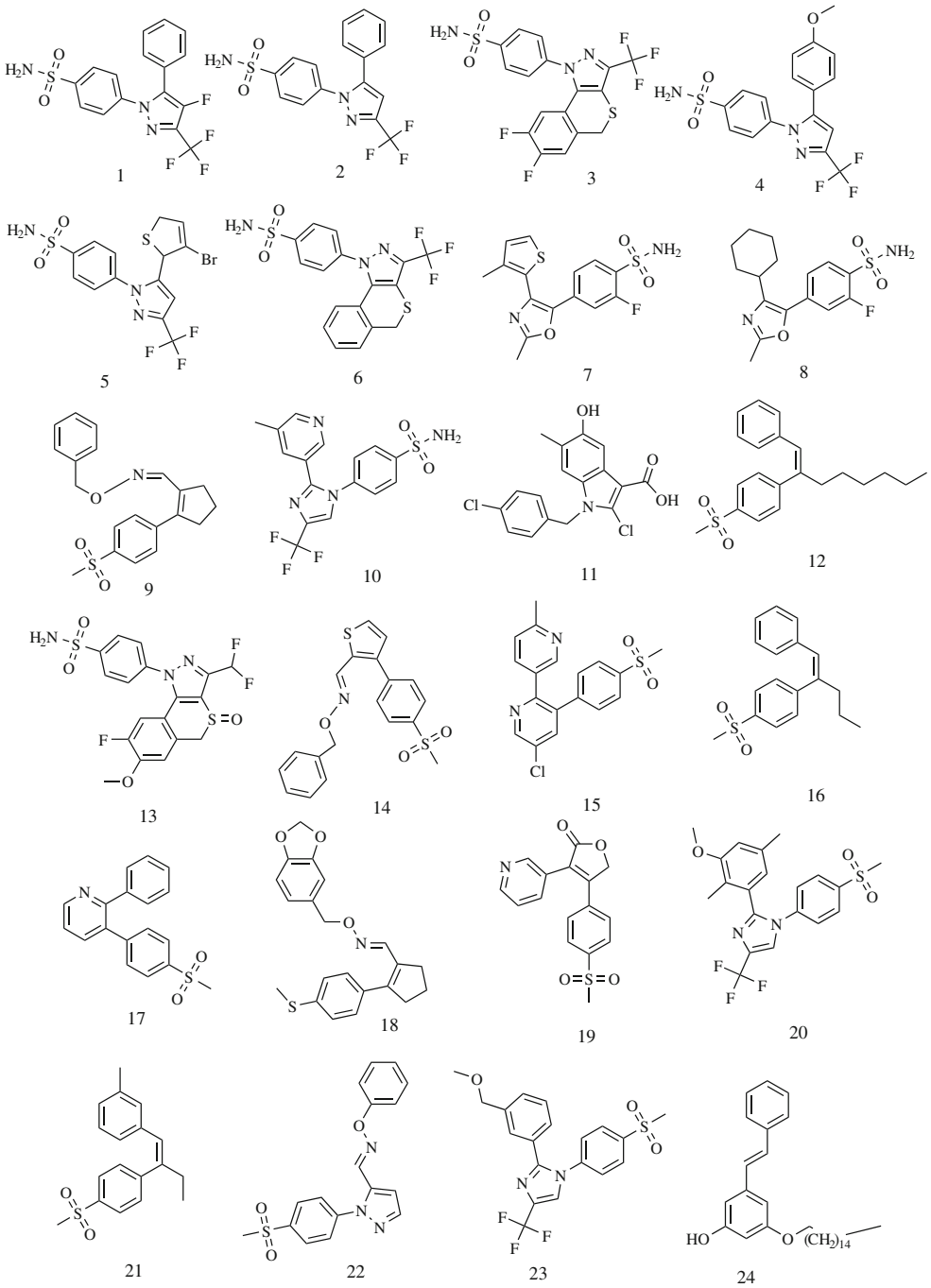
Selection of the training set The most important aspect of the hypothesis generation in HypoGen is the selection of the training set of molecules. The selection has to follow some basic requirements; such as a minimum of 16 structurally diverse compounds should be selected to avoid any chance correlation [40]. The activity data should have a range of 4–5 orders of magnitude. The selected compounds should provide clear and concise information. Any redundancy should be avoided in terms of structural features or activity range. A compound that is considered to be inactive because of steric factors should not be included because current Catalyst features in the Catalyst software cannot handle such cases.

Biological data The sources of the biological activity data, represented as IC_{50} in nM (compounds 1–46), were from the literature [55–68], and the chemical structures of the inhibitors are listed in Chart 1 and 2. The data sets were divided into a training set and a test set. The most active compounds were included so that they would provide critical information for pharmacophore requirements. Several moderately active and inactive compounds were also included to spread the activity ranges as wide as possible. The important aspect of this selection scheme is that each active compound should teach something new to the HypoGen module to help it uncover as much critical information as possible for predicting biological activity. In the case of COX-2 selective inhibitors, a training set of 24 compounds with the above criteria has been selected; the other 22 compounds were used as the test set. An uncertainty value of 3 (default) was used for compound activity, which is a ratio range of uncertainty in the activity value. The activities against COX-2 have been classified as follows: highly active (<50 nM), moderately active (50–2500 nM), and poorly active (>2500 nM). These activities are classified, somewhat arbitrarily, on the basis of the lowest and the highest activity ranges predicted (fit values) by the hypothesis for selected training set molecules.

Generation of Pharmacophores Details of the pharmacophore development procedures have been described in the literature [24, 69]. In brief, conformational models of all training set molecules for COX-2 were generated using the best quality conformational search option in Catalyst v4.11 (Accelrys) using a constraint of a 20 kcal mol^{-1} energy threshold above the global energy minimum and a modified version of the CHARMm force field parameters [46] as available in Catalyst. A maximum of 250 conformations were generated using the best-fit method to ensure maximum coverage in the conformational space. All other settings were kept as a default. Instead of using just the lowest energy conformation of each compound, all conformational models for molecules in each training set were used in Catalyst for pharmacophore hypothesis generation. An initial analysis revealed that four chemical feature types such as hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), and hydrophobic (aliphatic) (HY-ALI) and ring aromatic (RA) features could effectively map all critical chemical features of all molecules in the training and test sets. The minimum and maximum counts for HBA, HBD, HY-ALI and RA were set to 0 and 3, respectively. These four feature types were used to generate 10 pharmacophores from the training set. The uncertainty value was defaulted to 3, and MinPoints and MinSubsetPoints were 4 (default value). The MinPoints parameter controls the minimum number of location constraints required for any hypothesis. The MinSubsetPoint parameter defines the number of chemical features that a hypothesis must match in all the compounds set. The Catalyst software can generate pharmacophore hypotheses consisting of a maximum of five features.

Important output parameters that determine the quality of the pharmacophore hypothesis **1. The cost function analysis**. The algorithm employed for the Catalyst auto-

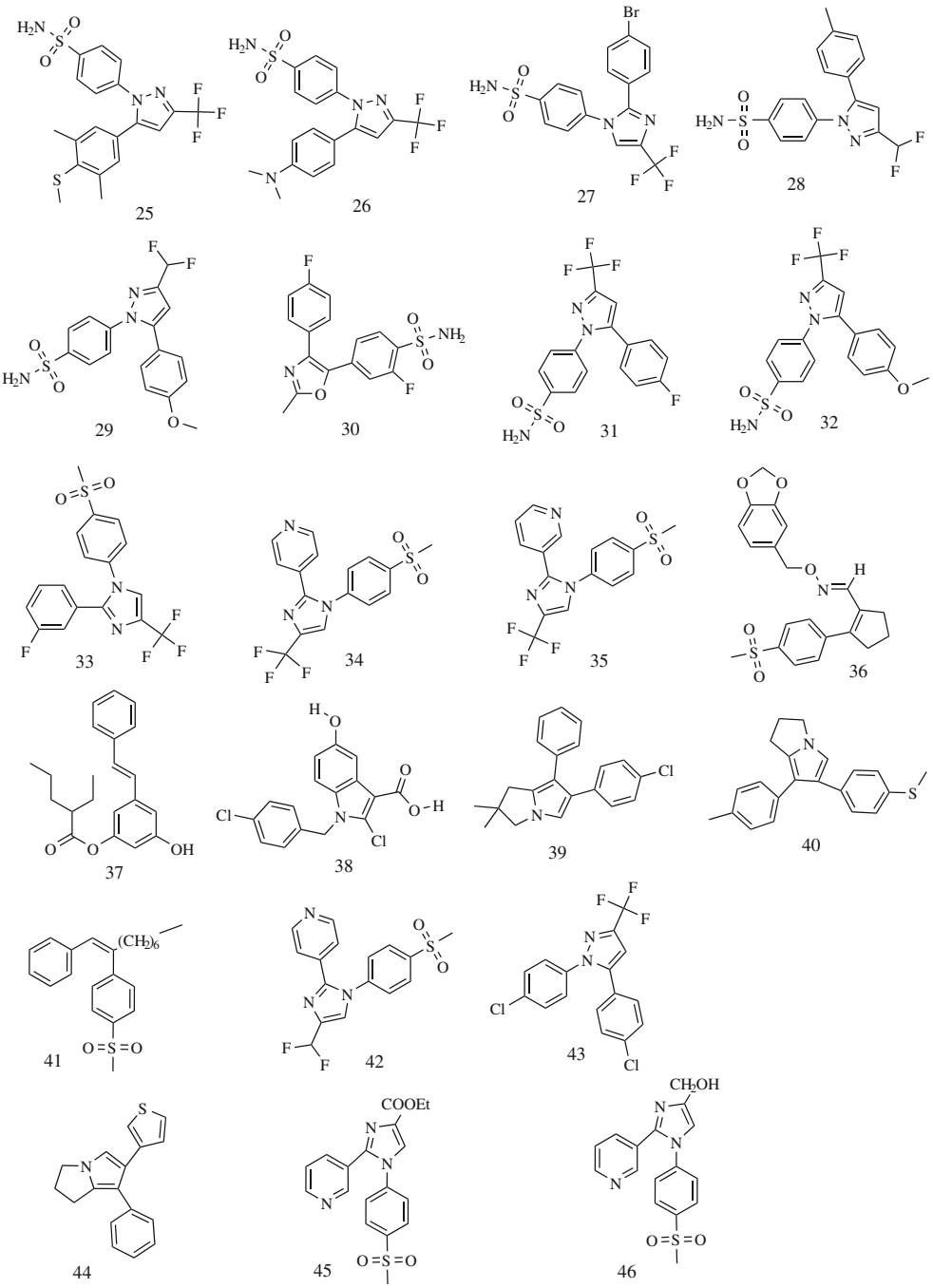
Chart 1 Chemical structures of 24 training set compounds[a]



matic hypothesis generation (HypoGen) optimizes hypotheses that are common to the active compounds in the training set but not shared by the inactive compounds. This is done in three phases. In the constructive phase, hypotheses common to all actives are defined. The subtractive phase removes hypotheses common to the inactive compounds. The third phase optimizes the resultant hypotheses from phase 2 that have survived the subtractive phase.

Catalyst uses bits for language and assigns costs to hypotheses in terms of number of bits required to describe them fully. The HypoGen module in Catalyst performs two important theoretical cost calculations (represented in bit units) that determine the success of any pharmacophore hypothesis. One is known as the fixed cost, which represents the simplest model that fits all data perfectly (all compounds fall along a line of slope 1), and is calculated by adding the minimum achievable error and weight cost and the constant configuration cost. The second one is known as null cost, which represents the highest cost of a pharmacophore with no features and which estimates activity to be the average of the activity data of the training

Chart 2 Chemical structures of 22 test set compounds[a]



set molecules. Its absolute value is equal to the maximum occurring error cost. A meaningful pharmacophore hypothesis may result when the difference between these two values (null cost - fixed cost) is large; a value of 70–100 bits suggests an excellent chance (>90%) for a true correlation; if the difference is between 40 and 70 bits for a pharmacophore hypothesis, it may indicate that it has a 75–90% probability of correlating the data (Catalyst 4.11 documentation). The total cost of any pharmacophore hypothesis should be close to the fixed cost to provide any useful models. The total cost value is the sum of three costs: a weight, an error, and a configuration cost value. These three cost components could be described as follows: Each feature of a hypothesis represents certain orders of magnitude of the compounds activity. With the default setting of 0.302, the represented orders of magnitude are kept as close to 2 as possible. The weight component is a value that increases in a Gaussian form as these function weights in a model deviate from the ideal value of 2. The error cost is dependent on the root mean- square (rms) differences between the estimated and actual activities of the training set molecules. The rms deviations represent the

quality of the correlation between the estimated and actual activity data. The configuration cost is also known as the entropy cost and depends on the complexity of the pharmacophore hypothesis space. Mathematically, the configuration cost is expressed as $\log_2 P$, where P is the number of initial hypothesis created in the constructive phase and that survived in the subtractive phase. It should not be greater than 17 (corresponds to a number of 217 pharmacophore models). Higher values would lead, more likely, to a chance correlation of the generated hypothesis, since Catalyst cannot consider more than 217 models in the optimization phase, and so, the rest is left out of the process. Any number greater than 17 suggests that some attention has to be given in selecting the training set molecules. Limiting the minimum and maximum features can reduce the entropy cost.

2. Fisher's cross-validation test. To evaluate the statistical relevance of the model, the Fischer's randomization test was applied. The Fisher's randomization test is used to validate the strong correlation between chemical structures and biological activity. The purpose of this technique is to randomize the activity data associated with the training set compounds, and the randomized training sets are used to generate pharmacophore hypotheses using the same features and parameters to develop the original pharmacophore hypothesis. If the randomized data set results in the generation of a pharmacophore with similar or better cost values, rms, and correlation, the original hypothesis is considered to have been generated by chance [70]. The statistical significance is given by the equation significance $= [1 - (1 + x)/y]100$, where x = total number of hypotheses having a total cost lower than a best significant hypothesis and y = the number of initial HypoGen runs+ random runs. With the aid of the Cat Scramble program available in Catalyst/HypoGen module, the experimental activities in the training set were scrambled randomly, and the resulting training set was used for a HypoGen run. Thereby, all parameters were adopted from the initial HypoGen calculation. The number of such random trials depends on what level of statistical significance is to be achieved. For a 95% confidence level, 19 spreadsheets were created. For 98% and 99% confidence levels, 49 and 99 spreadsheets, respectively, are created. This procedure was reiterated 19 times to achieve a 95% confidence level.

Result and discussion

Pharmacophore generation A pharmacophore model has been generated using a set of 24 training set compounds representing six series of structurally diverse compounds existing in the literature. Sets of 10 hypotheses were generated using the data from 24 training set compounds. Different cost values, correlation coefficients (r), rms deviations, and pharmacophore features are listed in Table 1.

The value of the total cost of each hypothesis was close to the fixed cost values, which is expected for good hypotheses. The entropy (configuration cost) value of the hypotheses was 16.79, within the allowed range. The difference between the null hypothesis and the fixed cost and between the null hypothesis and the total cost of the best hypothesis (hypothesis 1) was 88.37 and 78.51, respectively. All 10 hypotheses consist of four features: one HBD feature, one HBA, one HY-AR feature and one RA feature. The best pharmacophore (hypothesis 1), which has the highest cost difference, lowest error cost, lowest rms difference, and the best correlation coefficient, has one HBD, one HBA feature, one RA feature and one HY-AR feature. A regression of the predicted activities for the training set by the best hypothesis versus the actual activities results in the same relationship for those predicted by hypothesis 1. The graph of estimated activities against actual activities for the training set is shown in Fig. 1. The best hypothesis is different from the pharmacophore described by Michaux et al. [4]. The pharmacophore

Table 1 Results obtained from pharmacophore hypothesis generation using the training set molecules

Hypothesis number	Total cost	Error cost	RMS	Correlation (r)	Features[b]
1.	108.509	90.3154	0.894	0.954	HBA, HBD, HY-AR, RA
2	108.982	90.9657	0.924	0.950	HBA, HBD, HY
3	109.959	91.8718	0.964	0.946	HBA, HBD, HY
4	110.441	92.414	0.987	0.943	HBA, HBD, HY
5	110.709	92.497	0.990	0.943	HBA, HBD, HY
6	110.756	92.614	0.995	0.942	HBA, HBD, HY
7	110.994	92.974	1.010	0.940	HBA, HBD, HY
8	112.04	94.1003	1.055	0.935	HBA, HBD, HY
9	112.23	94.225	1.060	0.934	HBA, HBD, HY
10	112.612	80.7265	1.062	0.934	HBA, HBD, HY

[a] Null cost=187.021, Fixed cost=98.6513, Configuration cost=16.79. All costs are in units of bits. [b] HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; HY-AR, hydrophobic aromatic; RA, ring aromatic

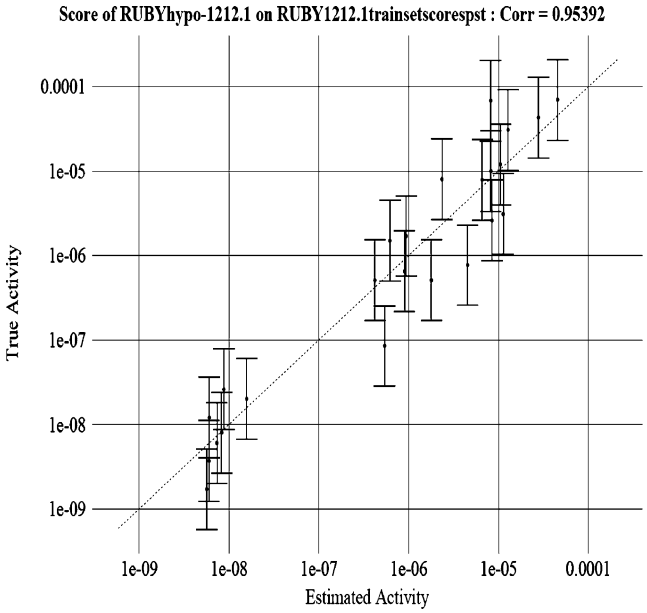


Fig. 1 The regression of actual versus predicted activities by hypothesis 1 for the training set inhibitors. The training set includes 24 structurally diverse COX-2 inhibitors

developed by Michaux et al. hypothesis h1 showed one HBA, One RA, and two hydrophobic (HY) features are important for the activity. The HBA feature was shown to be mapped by the S = O of the sulphonamide moiety. In another pharmacophore, h5 they showed two HBA features, one RA and one hydrophobic features, which could not explain the SAR of the diaryl heterocyclic class of compounds. The pharmacophore h1 could correlate the SAR for the different inhibitors and fitted the COX-2 selective site. We selected to generate hypothesis including more classes of compounds and searched for which of the features would be essential in all the classes of compounds. Our resultant hypothesis Hypo 1 (Fig. 2a) gave one HBD feature, one HBA, one HY-AR feature and one RA feature as essential features. The HBD feature in our hypothesis is also essential, to which NH of the sulphonamide maps in compounds containing the sulphonamide functionality. This feature is shown to interact with the Leu352 (C = O...H-N), Gln192 (C = O...H-N) and Ser353 (C = O...H-N) in the COX-2 binding site with SC558 (Fig. 3). In order to search for a pharmacophore comprising of one additional HBA feature in the form of S = O, we reduced the spacing parameter from 297 Å to 5 Å as advance parameter modification but the Hypogen output selected HBD feature

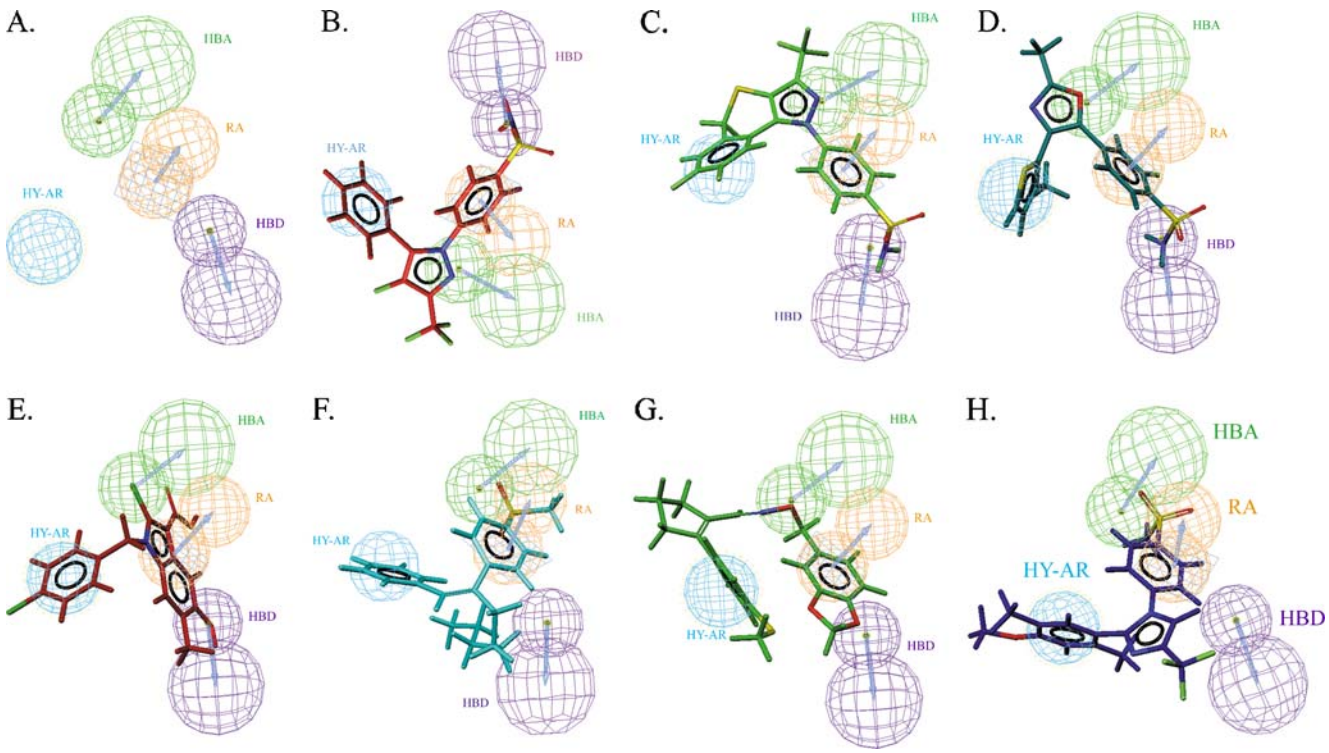


Fig. 2 Mapping of the five chemotypes of COX-2 inhibitors from training set onto the selected pharmacophore (hypothesis 1) (**a**) pharmacophore (hypothesis 1) (**b**) compound 1, celecoxib derivative (**c**) compound 3, diarylpyrazole derivative (**d**) compound 7, isooxazole derivative (**e**) compound 11, indole derivative (**f**) compound 12, diaryl olefin derivative (**g**) compound 18, cyclopentene derivative (**h**) compound 20, diarylimidazole derivative. The cyan contour represents hydrophobic aromatic (HY-AR), green contour represents hydrogen bond acceptor (HBA), purple contour represents the hydrogen bond donor (HBD) and orange contour represents the ring aromatic (RA) features, respectively

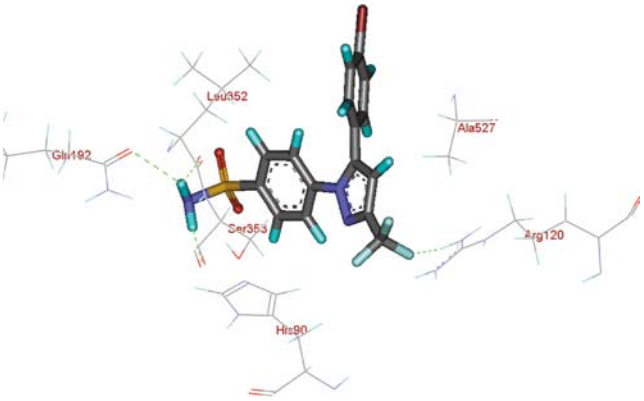


Fig. 3 SC558 (in bold stick) in the COX-2 binding site is shown to interact with Glu192, Ser353, Arg120 (prepared from PDB ID 6COX)

of NH to be important over HBA feature to explain the activities of all the classes of compounds. This could be explained as probably the Catalyst does not seem to support two closely placed features such as those present in SO_2NH_2 group. Either the generated hypothesis has one HBA mapping to S = O as in the case of h1 and h5 generated by Michaux et al. or one HBD mapping to NH of the sulphonamide (HYPO 1 in our case). The software could not detect simultaneous presence of closely placed features one HBA and one HBD on same functional group (SO_2 NH_2).

Table 2 shows the actual and estimated inhibitory activities for training set molecules for COX-2 inhibitors. In the training set compounds, all highly active compounds (<50 nM) were predicted correctly as highly active. Among moderately active compounds all were predicted rightly except one compound 12 (50–2500 nM), which was predicted to be poorly active. Another compound 18 a poorly active one (>2500 nM) was predicted to be moderately active.

Validation of pharmacophore model **1. Fisher's cross validation test**. A pharmacophore generated by the Cat Scramble software in Catalyst was assessed for quality by Fisher's randomization test method. The purpose of this technique is to reshuffle the activity data associated with the training set compounds and the randomized training sets are then used to generate pharmacophore hypotheses using the same features and parameters as had been used to develop the original pharmacophore hypothesis. If the randomized data result in the generation of pharmacophore with similar or better cost values, RMS's, or correlations,

Table 2 Actual and predicted activities of training set molecules calculated on the basis of hypothesis 1

Number	Comp. number	Fit	IC_{50}(nM)[a]		Activity scale[b]	
			Actual	Estimated	Actual	Estimated
1	1	8.41	1.7	5.6	+ + +	+ + +
2	2	8.39	3.7	5.9	+ + +	+ + +
3	3	8.30	6	7.2	+ + +	+ + +
4	4	8.25	8	8.2	+ + +	+ + +
5	5	8.39	12	5.9	+ + +	+ + +
6	6	7.97	20	16	+ + +	+ + +
7	7	8.22	26	8.6	+ + +	+ + +
8	8	6.43	85	530	+ +	+ +
9	9	5.91	1900	1800	+ +	+ +
10	10	6.54	510	420	+ +	+ +
11	11	6.21	650	890	+ +	+ +
12	12	5.51	770	4500	+ +	+
13	13	6.37	1500	610	+ +	+ +
14	14	6.19	1700	930	+ +	+ +
15	15	5.24	2600	8400	+	+
16	16	5.11	3100	11000	+	+
17	17	5.34	7900	6500	+	+
18	18	5.79	8000	2300	+	+ +
19	19	5.25	10000	8100	+	+
20	20	5.14	12000	10000	+	+
21	21	5.06	31000	13000	+	+
22	22	4.72	43000	28000	+	+
23	23	5.25	68000	8100	+	+
24	24	4.51	70000	45000	+	+

[a] Data for activities of Cyclooxygenase inhibition are from references listed in Materials and methods section.
[b] Activity scale: + + + (0–50 nM, highly active), + + (50–2500 nM, moderately active), + (>2500 nM, inactives)

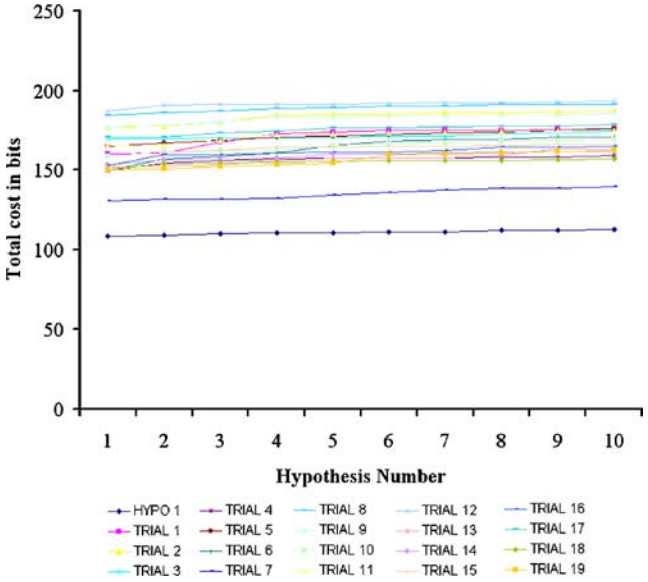


Fig. 4 Fisher's cross validation run using CatScramble[a]

the original hypothesis is considered as having been generated by chance. The results of such a run are shown in Fig. 4, and the resultant data clearly shows that none of the generated hypothesis after reshuffling had a cost value lower than that of hypothesis 1. However, it is clearly depicted from the graph shown in Fig. 4 that among the 19 trials conducted for the randomization of activity data only one randomization trial 7 gave the results having the total cost 130.44 for hypothesis 1 which is close to the total cost 108.50 of hypothesis 1 for the unscrambled data. According to this software documentation and the literature available, this result indicates that there is a 95% chance for hypo1 to represent the correlation in the training set.

2. Using the test set. Apart from predicting the activity of the training set compounds accurately, one of the major objectives of pharmacophore hypothesis generation is to verify whether pharmacophore models are capable of predicting the activities of test series and classifying them correctly as actives or inactives. We have constructed a test set which consists of 22 compounds and conformational studies for each compound were done as described earlier. The estimated activities were scored using hypothesis 1 as the pharmacophore (Table 3). All highly active compounds (<50 nM) were predicted to be highly active except compound 27, which was inaccurately predicted as moderately active. Among the moderately active compounds (50–2500 nM) all were correctly predicted as moderately active except three compounds. One moderately active compound 38 was predicted to be highly active whereas

Table 3 Actual and estimated activities of test set molecules calculated on basis of hypothesis 1

Number	Comp. number	Fit	IC_{50} (nM)[a]		Activity scale[b]	
			Actual	Estimated	Actual	Estimated
1	25	8.399	3.7	5.8	+ + +	+ + +
2	26	8.432	4.7	5.3	+ + +	+ + +
3	27	6.634	7.0	340	+ + +	+ +
4	28	8.454	13	5.1	+ + +	+ + +
5	29	8.487	15	4.7	+ + +	+ + +
6	30	8.421	40	5.5	+ + +	+ + +
7	31	8.563	41	4.0	+ + +	+ + +
8	32	8.522	15	4.3	+ + +	+ + +
9	33	5.995	120	1500	+ +	+ +
10	34	5.93	130	1700	+ +	+ +
11	35	5.792	1690	2300	+ +	+ +
12	36	6.635	400	330	+ +	+ +
13	37	6.265	1700	780	+ +	+ +
14	38	7.599	1700	36	+ +	+ + +
15	39	5.623	370	5600	+ +	+
16	40	3.979	900	9400	+ +	+
17	41	5.987	10000	1500	+	+ +
18	42	5.890	20700	1900	+	+ +
19	43	6.3639	4790	720	+	+ +
20	44	4.897	20% at 10 μM	4400	+	+
21	45	4.398	>100 μM	8700	+	+
22	46	4.855	934000	4000	+	+

[a] Data for activities of cyclooxygenase inhibition are from references listed in Materials and methods section.
[b] Activity scale: + + + (0–50 nM, highly active), + + (50–2500 nM, moderately active), + (>2500 nM, inactives)

compound 39 and 40 were predicted to be inactive. All inactive compounds were predicted as inactive ones except the compound 41–43, which were predicted to be moderately active.

Finally the compounds were mapped onto the best hypothesis using the best fit and a conformational energy constraint of 10 kcalmol^{-1} One of the most active as well as most selective COX-2 inhibitors among all compounds tested in all the series (compound 1) was selected from the training set to show mapping of this compound on the selected pharmacophore (hypothesis 1, Fig. 2a and b). The pharmacophore predicted the inhibitory activity of this compound against COX-2 excellently. Compound 1 mapped all four features of the hypothesis quite well with a fit value of 8.41 (actual activity 1.7 nM and estimated 5.6 nM). Similarly, compounds 2–7 also mapped all four features well with a fit value range from 8.39–7.97 (figure not shown). Only highly active compounds mapped all the four features in the training set. Out of 24 compounds moderately active compounds, e.g., 11 and 12 missed either HBD (compound 11, Fig. 2e) or HBA (compound 12, Fig. 2f) one feature or partially overlapped either of them. The poorly active compounds missed two of the features. Among poorly active compounds, 18 misses HBD completely and partially overlaps with HBA; compound 20 misses HBA and HBD completely. All the 24 compounds mapped RA and HY-AR feature. As all the highly active compounds with fit value 8.41 –7.97 mapped four features, this suggested that presence of four features one HBD, one HBA, one RA and one HY-AR are minimum essential for the highly active compounds and three feature hypothesis missing either HBA or HBD is sufficient to account for the moderately active compounds. Inclusion of fourth feature as HBA /HBD results in the improvement of potency from moderately active (50–2500 nM) to highly active (1–50 nM). In other words, three feature hypothesis explains the moderately active compounds in the activity range of 50 to 2500 nM, but the four feature hypothesis explains the improvement of activity of compound 1 (1.7 nM) over that of compound 9 (1900 nM). To show that the other chemotypes also map well to hypothesis 1, one compound from each type - diarylpyrazole derivative compound 3, Fig. 2c; Isooxazole derivative compound 7, Fig. 2d; Indole derivative compound 11, Fig. 2e; diaryl olefin derivative compound 12, Fig. 2f, Cyclopentene derivative compound 18, Fig. 2g and diarylimidazole derivative compound 20, Fig. 2h were mapped to hypothesis 1. To

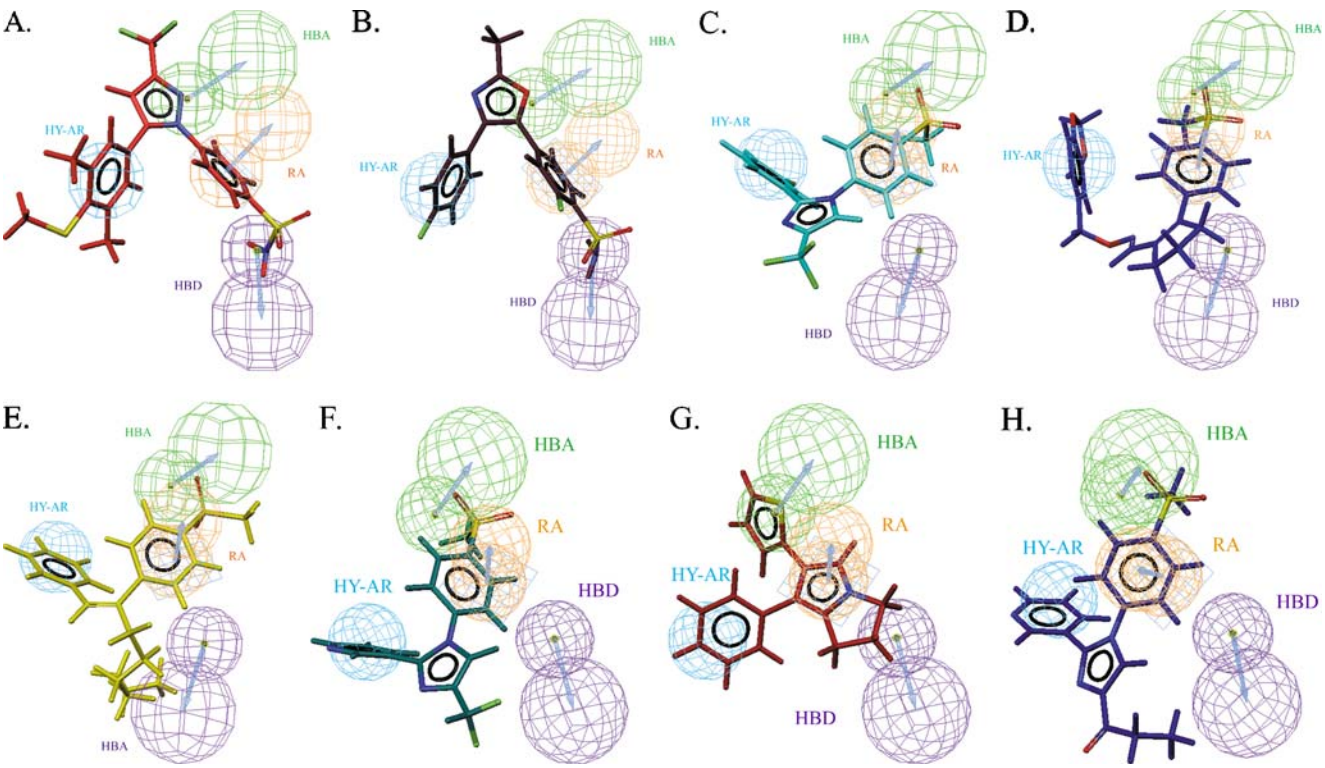


Fig. 5 Mapping of the five chemotypes of COX-2 inhibitors from test set onto the selected pharmacophore (hypothesis 1) (**a**) compound 25, celecoxib derivative (**b**) compound 30, isooxazole derivative (**c**) compound 33, diarylimidazole derivative (**d**) compound 36 cyclopentene derivative (**e**) compound 41, diarylimidazole derivative, (**f**) compound 42, diarylimidazole derivative (**g**) compound 44, dihydropyrolizine derivative (**h**) compound 45. The cyan contour represents hydrophobic aromatic (HY-AR), green contour represents hydrogen bond acceptor (HBA), purple contour represents the hydrogen bond donor (HBD) and orange contour represents the ring aromatic (RA) features, respectively

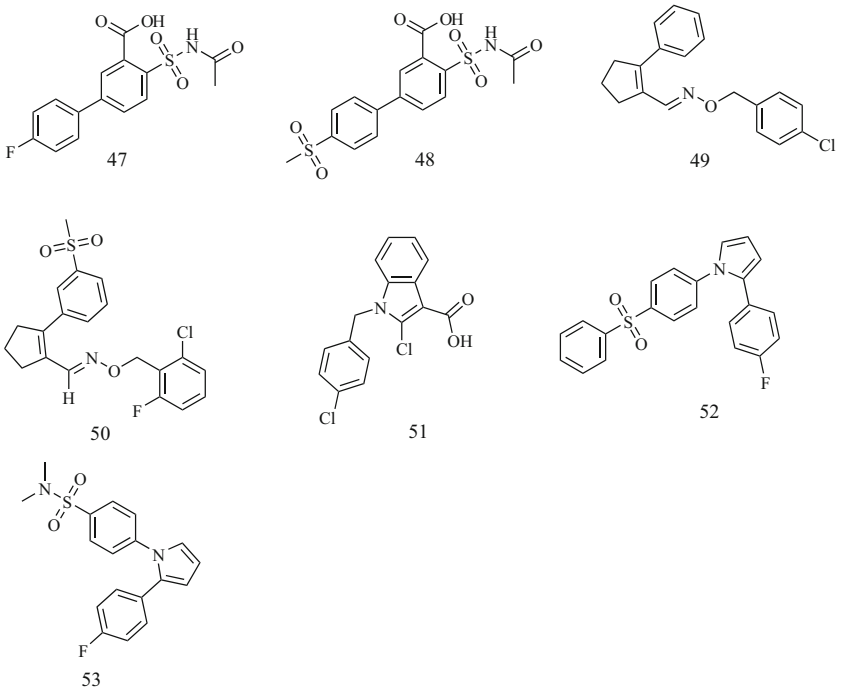


Chart 3 Chemical structures of 7 test set compounds with COX-1 selectivity[a]

give some indications of how well the structurally diverse compounds from the test set mapped this pharmacophore, one derivative from each subtype was selected and mapped on this pharmacophore and is shown in Fig. 5. Among the test set compounds, compound 25, (Fig. 5a) mapped the pharmacophore (fit value 8.399) very well and the estimated activity calculated on the basis of this pharmacophore was close to the actual activity (actual activity of 3.7 nM vs estimated value of 5.8 nM) and is shown as + + + in Table 3 (The classification of high 0–50 nM, moderate 50–2500 nM and weak activity >2500 nM was done arbitrarily). Moderately active compounds were mapping to three features with partial overlaps e.g compound 11 (fit value 6.21, Fig. 2e) from the training set and compound 33 (fit value 5.99, Fig. 5c) from the test set. Inactive compound mapped only two features, e.g., compound 41, Fig. 5f. The inactive compound either missed both HBD and HBA features, or missed one of them with partial overlap of other feature.

The validation study with five different classes of COX-2 inhibitors suggested that pharmacophore was capable of mapping structurally diverse group of compounds quite effectively and provided confidence that this pharmacophore could be used as a search query to identify new COX-2 inhibitors from drug like chemical libraries.

Validation of the pharmacophore model using cyclooxygenase-1 selective inhibitors as a test set (test for selectivity)

A set of seven compounds [64–68] (Chart 3) that target COX-1 selectively were taken as a negative controls to check the selectivity of the hypothesis for COX-2 targeted compounds versus COX-1 compounds. In fact the most selective COX-1 ligands were predicted as poorly active

Table 4 Actual and estimated activities of test set molecules (COX-1 selective inhibitors) calculated on basis of hypothesis 1

Number	Comp number	Fit	Actual activity (COX-1) nM	Actual activity (COX-2) nM	Estimated activity (COX-2) nM	predicted activity scale
1	47	5.37	680	>100 μM	6100	+
2	48	5.14	1170	>100 μM	10000	+
3	49	6.63	43% at10 μM	1800	3400	+
4	50	6.631	33% at10 μM	Not active	3400	+
5	51	6.23	95±15	230±120	830	+ +
6	52	6.11	100000	>100 μM	1100	+ +
7	53	5.66	100000	>100 μM	3100	+

using pharmacophore mapping, thereby showing some degree of selectivity of generated hypothesis. Hypothesis 1 predicted the activities of five out of seven tested activities to be inactive against COX-2, thereby showing the selectivity of the hypothesis to some extent. Compound 51 and 52 were predicted to be moderately active against the COX-2. This suggests that there are similarities for feature requirements of the two kinds of isoforms, i.e., COX-1 and COX-2 binding sites, and hypothesis1generated in this work, though predictive for the COX enzyme may not be fully selective for the COX-2. The results of the activity prediction of these compounds are shown in Table 4.

Conclusions

The work presented in this study shows how chemical features of a set of compounds along with their activities ranging over several orders of magnitudes can be used to generate pharmacophore hypothesis that can successfully predict activity. The models were not only predictive within the same series of compounds but six different classes of diverse compounds also effectively mapped onto most of the features important for activity. The pharmacophores generated from COX-2 inhibitors can be used (1) as a three dimensional query in database searches to identify compounds with diverse structures that can potentially inhibit COX-2 enzyme selectively and (2) to evaluate how well any newly designed compounds maps on the pharmacophore before undertaking any further study including synthesis. Both these applications may help in identifying or designing compounds for further biological evaluation and optimization. The pharmacophore developed in this study using inhibitors for COX-2 showed distinct chemical features that may be responsible for the activity of the inhibitors. We intend to utilize the information to undertake 3D searches on large databases of drug like molecules to identify a new generation of COX-2 inhibitors.

Acknowledgements This work was supported by grants from Department of Biotechnology (DBT), Ministry of Science and Technology, India.

References

1. Stoehlmacher J, Lenz H-J (2003) Sem Oncol 30(3), Suppl. 6:10–16
2. Kim HJ, Chae CH, Yi KY, Park KL, Sung-eun Y (2004) Bioorg Med Chem 12:1629–1641
3. Holtzman MJ, Turk J, Shornick LP (1992) J Biol Chem 267:21438–21445
4. Michaux C, de Leval X, Julemont F, Dogne JM, Pirotte B, Durant F (2006) Eur J Med Chem 41:1446–1455
5. Lanza FL, Scand J (1989) Gastroenterol 163(Suppl):24–31
6. Sawdy R, Slater D, Fisk N, Edmonds DK, Benett P (1997) Lancet 350:265
7. Kutchra W, Jones DA, Matsunami N, Groden J, McIntyre TM, Zimmerman GA, White RL, Presscott SM (1996) Proc Natl Acad Sci USA 93:4816
8. Stewart WF, Kawas C, Corrada M, Metter EJ (1997) Neurology 48:626
9. Subbaramaiah K, Dannerberg AJ (2003) Trends Pharmacol Sci 24:96
10. Thompson CA (2004) Am J Health Syst Pharm 61:2234–2236, 2238
11. Ehrlich P (1909) Dtsch Chem Ges 42:17
12. Gund P (1977) Prog Mol Subcell Biol 5:117–143
13. PhRMA: New Medicines in Development Series (2001) (http://www.phrma.org/publications/documents/factsheets/2001-03-01.210.phtml)
14. Barril X, Robert S (2006) Mol Biosyst 2:660–681
15. Barnum D, Greene J, Smellie A, Sprague P (1996) J Chem Inf Comput Sci 36:563–571
16. Jones G, Willett P, Glen RC (1995) J Comp Aided Mol Des 9:532–549
17. Martin YC, Bures MG, Danaher EA, DeLazzer J, Lico I, Pavlik PA (1993) J Comp Aided Mol Des 7:83–102
18. CatalystÒ, Accelrys, 9685 North Scranton Road, San Diego,CA 92121, USA. http://www.accelrys.com/
19. Greenidge PA, Weiser J (2001) Mini-Rev Med Chem 1:79–87
20. Palomer A, Pascual J, Cabre F, Garcia ML, Mauleon D (2000) J Med Chem 43:392–400
21. Duffy JC, Deardon JC, Green DSV Sanz F, Giraldo J, Manaut F (eds) (1995) Prous Science, Barcelona, pp 289–291
22. Hoffman RD, Bourguignon JJ (1995) Building a hypothesis for CCK-B antagonists using Catalyst program. In: Sanz F, Giraldo J, Manaut F (eds) QSAR and molecular modelling: concepts, computational tools and biological applications. Prous Science, Barcelona, pp 298–300
23. Kaminski JJ, Rane DF, Snow ME, Weber L, Rothofsky ML, Anderson SD, Lin SL (1997) J Med Chem 40:4103–4112
24. Sprague PW (1995) Automated chemical hypothesis generation and database searching with Catalyst. In: Muller K (ed) Perspectives in drug discovery and design. ESCOM Science, Leiden, The Netherlands, pp 1–21
25. Quintana J, Contijoch M, Cuberes R, Frigola J (1995) Structure reactivity relationships and molecular modeling studies of a series of H1 antihistamines. In: Sanz F, Giraldo J, Manaut F (eds) QSAR and molecular modelling: concepts, computational tools and biological applications. Prous Science, Barcelona, pp 282–288
26. Barbaro R, Betti L, Botta M, Corelli F, Giannaccini G, Maccari L, Manetti F, Strappaghetti G, Corsano S (2001) J Med Chem 44:2118–2132
27. Baringhaus KH, Matter H, Stengelin S, Kramer W (1999) J Lipid Res 40:2158–2168
28. Ekins S, de Groot MJ, Jones JP (2001) Drug Metab Dispos 29:936–944
29. Hirashima A, Rafaeli A, Gileadi C, Kuwano E (1999) J Mol Graphics Model 17:43–44
30. Karki RG, Kulkarni VM (2001) Eur J Med Chem 36:147–163
31. Manetti F, Corelli F, Biava M, Fioravanti R, Porretta GC, Botta M (2000) Farmaco 55:484–491
32. Ekins S, Bravi G, Ring BJ, Gillespie TA, Gillespie JS, Vandenbranden M, Wrighton SA, Wikel JH (1999) J Pharmacol Exp Ther 288:21–29
33. Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH, Wrighton SA (1999) J Pharmacol Exp Ther 290:429–438
34. Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH, Wrighton SA (1999) Pharmacogenetics 9:477–489
35. Grigorov M, Weber J, Tronchet JM, Jefford CW, Milhous WK, Maric D (1997) J Chem Inf Comput Sci 37:124–130

36. Lopez-Rodriguez ML, Porras E, Benhamu B, Ramos JA, Morcillo MJ, Lavandera JL (2000) Bioorg Med Chem Lett 10:1097–1100
37. Daveu C, Bureau R, Baglin I, Prunier H, Lancelot JC, Rault S (1999) J Chem Inf Comput Sci 39:362–369
38. Norinder U (2000) J Comput Aided Mol Des 14:545–557
39. Bureau R, Daveu C, Baglin I, Sopkova-De Oliveira SJ, Lancelot JC, Rault S (2001) J Chem Inf Comput Sci 41:815–823
40. Kurogi Y, Guner OF (2001) Curr Med Chem 8:1035–1055
41. Guner OF, Waldman M, Hoffman R, Kim J-H (2000) Strategies for database mining and pharmacophore development. In: Guner OF (ed) Pharmacophore perception, development, and use in drug design. International University Line, La Jolla, CA, pp 213–236
42. Langer T, Hoffman RD, Bachmair F, Begle S (2000) J Mol Struct: THEOCHEM 500:59–72
43. Kurogi Y, Miyata K, Okamura T, Hashimoto K, Tsutsumi K, Nasu M, Moriyasu M (2001) J Med Chem 44:2304–2307
44. Briens F, Bureau R, Rault S (1999) Ecotoxicol Environ Saf 43:241–251
45. Chopra M, Mishra AK (2005) J Chem Inf Model 45:1934–1942
46. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187–217
47. Llorens O, Perez JJ, Palomer A, Mauleon D (1999) Bioorg Med Chem Lett 9:2779–2784
48. Palomer A, Cabre F, Pascual J, Campos J, Trujillo MA, Entrena A, Gallo MA, Garcia L, Mauleon D, Espinosa A (2002) J Med Chem 45:1402–1411
49. Palomer A, Pascual J, Cabre M, Borra's L, Gonzalez G, Aparici M, Carabaza A, Cabre F, Garcija ML, Mauleon D (2002) Bioorg Med Chem Lett 12:533–537
50. Sutherland JJ, O'Brien LA, Weaver DF (2004) J Med Chem 47:3777–3787
51. Rodrigues CR, Veloso MP, Verli H, Fraga CA, Miranda AL, Barreiro EJ (2002) Curr Med Chem 9:849–867
52. Prasanna S, Manivannan E, Chaturvedi SC (2004) Biorg Med Chem. Lett 14:4005–4011
53. Rollinger JM, Haupt S, Stuppner H, Langer TJ (2004) J Chem Inf Comput Sci 44:480–488
54. Renner S, Schneider G (2004) J Med Chem 47:4653–4664
55. Hashimoto H, Imamura K, Haruta J, Wakitani K (2002) J Med Chem 45:1511–1517
56. Balsamo A, Coletta I, Guglielmotti A, Landolfi C, Mancini F, Martinelli A, Milanese C, Minutolo F, Nencetti S, Orlandini E, Pinza M, Rapposelli S, Rosello A (2003) Eur J Med Chem 38:157–168
57. Khanna IK, Yu Y, Huff RM, Weier RM, Xu X, Koszyk Francis J, Collins PW, Cogburn JN, Isakson PC, Koboldt CM, Masferrer JL, Perkins WE, Seibert K, Veenhuizen AW, Yuan J, Yang D-Chang, Zhang YY (2000) J Med Chem 43:3168–3185
58. Balsamo A, Coletta I, Guglielmotti A, Landolfi C, Mancini F, Martinelli A, Milanese C, Minutolo F, Orlandini E, Ortore G, Pinza M, Rapposelli S (2002) Eur J Med Chem 37:585–594
59. Khanna IK, Weier R, Yu Y, Xu XD, Koszyk FJ, Collins PW, Koboldt CM, Veenhuizen AW, Perkins WE, Casler JJ, Masferrer JL, Zhang YY, Gregory SA, Seibert K, Isakson PC (1997) J Med Chem 40:1634–1647
60. Penning TD, Talley JJ, Bertenshaw SR, Carter JS, Collins PW, Docter S, Graneto MJ, Lee LF, Malecha JW, Miyashiro JM, Rogers RS, Rogier DJ, Yu SS, Anderson GD, Burton EG, Cogburn JN, Gregory SA, Koboldt CM, Perkins WE, Seibert K, Veenhuizen AW, Zhang YY, Isakson PC (1997) J Med Chem 40:1347–1363
61. Bertanshaw SR, Talley JJ, Rogier DJ, Graneto MJ, Koboldt CM, Zhang Y et al. (1996) Bioorg Med Chem Lett 6:23, 2827–2830
62. Uddin MDJ, Praveen Rao PN, Rahim AMD, McDonal R, Knaus EE (2004) Bioorg Med Chem Lett 14:4911–4914
63. Andreani A, Granaiola M, Leoni A, Lacatelli A, Morigi R, Rambaldi M, Roda A, Guardigli M, Traniello S, Spisani S (2004) Eur J Med Chem 39:785–791
64. Balsamo A, Coletta I, Domiano P, Guglielmotti A, Landolfi C, Mancini F, Milanese C, Orlandini E, Rapposelli S, Pinza M, Macchia B (2002) Eur J Med Chem 37:391–398
65. Andreani A, Granaiola M, Leoni A, Locatelli A, Morigi R, Rambaldi M, Roda A, Guardigli M, Traniello S, Spisani S (2004) Eur J Med Chem 39:785–791
66. Khanna IK, Weier RM, Yu Y, Collins PW, Miyashiro JM, Koboldt CM, Veenhiuzen AW, Currie JL, Siebert K, Isakson PC (1997) J Med Chem 40:1619–1633
67. Chen QH, Praveen Rao PN, Knaus EE (2005) Bioorg Med Chem Lett 13:2459–2468
68. Ulbrich H, Fiebich B, Dannhardt G (2002) Eur J Med Chem 37:953–959
69. Gu¨ner OF (2000) Pharmacophore perception, development, and use in drug design Ed, International University Line, La Jolla, CA
70. Tromelin A, Guichard E (2003) J Agric Food Chem 51:1977–1983