Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Bioorganic & Medicinal Chemistry 12 (2004) 1629–1641
Bioorganic & Medicinal Chemistry

Computational studies of COX-2 inhibitors: 3D-QSAR and docking☆

Hye-Jung Kim,[a,b] Chong Hak Chae,[a] Kyu Yang Yi,[a] Kyung-Lae Park[b] and Sung-eun Yoo[a,*]

[a]*Korea Research Institute of Chemical Technology, PO Box 107, Yusung-gu, Taejeon 305-343, South Korea*
[b]*Department of Pharmacy, Chungnam National University, 220 Gung-dong, Yusung-gu, Taejeon 305-764, South Korea*

Received 20 October 2003; revised 17 January 2004; accepted 19 January 2004

Abstract—The 3D-QSAR (three-dimensional quantitative structure–activity relationships) studies for 88 selective COX-2 (cyclooxygenase-2) inhibitors belonging to three chemical classes (triaryl rings, diaryl cycloalkanopyrazoles, and diphenyl hydrazides) were conducted using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). Partial least squares analysis produced statistically significant models with q^2 values of 0.84 and 0.79 for CoMFA and CoMSIA, respectively. The binding energies calculated from flexible docking were correlated with inhibitory activities by the least-squares fit method. The three chemical classes of inhibitors showed reasonable internal predictability ($r^2 = 0.51$, 0.49, and 0.54), but the sulfonyl-containing inhibitors demonstrated distinctively low binding energy compared to the others. The electrostatic interaction energy between the Arg513 of the COX-2 active site and sulfonyl group of the triaryl rings seemed to have the responsibility for difference in binding energy. Comparative binding energy (COMBINE) analyses gave q^2 values of 0.64, 0.63, and 0.50 for triaryl rings, diaryl cycloalkanopyrazoles, and diphenyl hydrazides, respectively. In this COMBINE model, some protein residues were highlighted as particularly important for inhibitory activity. The combination of ligand-based and structure-based models provided an improved understanding in the interaction between the three chemical classes and the COX-2.

1. Introduction

Nonsteroidal anti-inflammatory drugs (NSAIDs)[1] display their anti-inflammatory actions primarily through the inhibition of cyclooxygenase (COX), which catalyzes the conversion of arachidonic acid (AA) to prostaglandin (PG).[2] Cyclooxygenase exists in at least two isoforms, COX-1 and COX-2.[3] COX-1 is a constitutive enzyme,[4] and COX-2 is an inducible isoform that leads to inflammation.[5] All classical NSAIDs, such as aspirin, ibuprofen, and indomethacin, can inhibit both COX-1 and COX-2, but bind more tightly to COX-1.[6] Selective COX-2 inhibitors are proving to have the same anti-inflammatory, anti-pyretic, and analgesic activities as do nonselective NSAID inhibitors, but with few or none of their gastrointestinal side-effects.[7] At present, two orally administered selective COX-2 inhibitors, celecoxib[8] and rofecoxib,[9] have successfully reached the market, creating great interest in finding more isozyme-specific drugs. Valdecoxib[10] and etoricoxib[11] are currently being clinically evaluated; these drugs have shown efficacy in the treatment of acute pain, osteoarthritis, and rheumatoid arthritis in clinical trials.[12] A water-soluble prodrug of valdecoxib, parecoxib, has been marketed recently for the parenteral treatment of postoperative pain.[13] 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.[14–16] in addition, there is growing evidence that COX-2 contributes to carcinogenesis.[17]

There are at least four mechanisms of COX inhibition: competitive, tight binding/time dependent, weak binding/mixed, and covalent binding.[18] Some NSAIDs inhibit COX-1 and COX-2 by similar mechanisms, whereas other NSAIDs have distinct inhibition mechanisms for each isoform. For example, celecoxib has been reported as a reversible competitive inhibitor of COX-1, while demonstrating time-dependent irreversible inhibition of COX-2.[19]

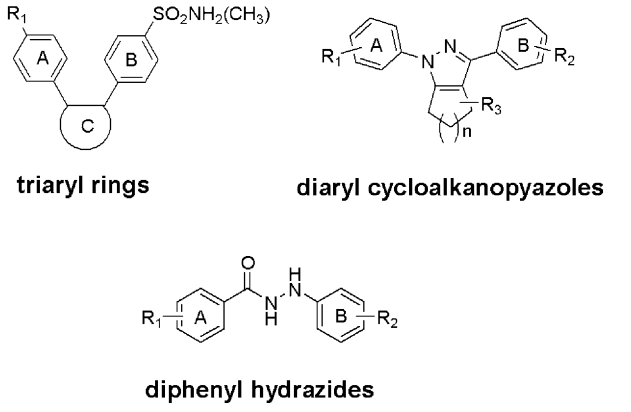
Keywords: COX-2; 3D-QSAR; CoMFA; CoMSIA; Docking; COMBINE.
☆Supplementary data associated with this article can be found, in the online version at, doi:10.1016/j.bmc.2004.01.027
* Corresponding author. Tel. +82-428-607-631; fax: +82-428-607-635; e-mail: seyoo@krict.re.kr

0968-0896/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.bmc.2004.01.027

Following the discovery of the original COX-2 inhibitor, Dup-697, many triaryl ring compounds with sulfonyl groups were reported as selective COX-2 inhibitors;[20–23] 3D-QSAR studies of these compounds were carried out.[24–29] Structure-based studies have been performed to identify binding modes and important interactions of triaryl rings at the COX-2 active site.[30–32] The triaryl ring moiety containing a *para*-sulfonyl group has been recognized as a pharmacophore for selective COX-2 inhibition.[33,34] Recently, a new series of diaryl heterocyclic compounds without sulfonyl groups has been reported to selectively inhibit COX-2;[35,36] 3D-QSAR analyses of these inhibitors were performed.[37] The present study applied 3D-QSAR[38,39] to the three chemical classes of selective COX-2 inhibitors, triaryl rings, diaryl cycloalkanopyrazoles, and diphenyl hydrazides. We also carried out flexible docking studies using the CHARMM program[40] to predict binding modes, and we derived the correlations between activities and binding energies using a least-squares fit method.[41] Furthermore, COMBINE analysis was performed to identify key residues for nonbonded (van der Waals, E_V and electrostatic, E_Q) interaction energy between ligand and enzyme.[42] These analyses defined energetic and structural differences in the interactions between the three chemical classes and the COX-2 enzyme.

2. Results and discussion

The data set consisted of triaryl ring, 1,3-diaryl cycloalkanopyrazole, and diphenyl hydrazide derivatives (Scheme 1), for which published biological activity data were available (Table S1, S2, and S3, Supporting Information). The triaryl rings belong to structurally different families, depending on the central C ring.[29,43–49] Activities were assayed using human COX-2 enzyme and were expressed as negative log values. Values greater than 10 μM were approximated as 20 μM for diaryl cycloalkanopyrazole and diphenyl hydrazide derivatives.[36] Inhibitors were grouped into the following subgroups for 3D-QSAR: the entire set (model A), triaryl rings (model B), diaryl cycloalkanopyrazoles (model C), diphenyl hydrazides (model D), triaryl rings and diaryl cycloalkanopyrazoles (model E), triaryl rings and diphenyl hydrazides (model F), and diaryl cycloalkanopyrazoles and diphenyl hydrazides (model G).



Scheme 1. Chemical classes of inhibitors studied.

Superimposition of the molecules was carried out using the pose of crystallized SC-558 as the template. Figure 1 shows the aligned molecules within the grid box used to generate the CoMFA column.[38]

2.1. CoMFA analysis

PLS results for the different combinations of the three chemical classes are summarized in Table 1. Although cross-validation reflects the predictive power of all the models, the results show good predictability for models A–G, which yielded high cross-validated correlation coefficients, q^2, with reasonable standard errors of prediction (SDEP). Model A yielded a q^2 of 0.73 and a SDEP of 0.61. However, this model did not predict the $p\mathrm{IC}_{50}$ values accurately for the four compounds **68**, **77**, **84**, and **88**, which were estimated with error values of −1.46, −1.64, 1.91, and −1.61 log units, respectively. Omission of the four outliers increased q^2 to 0.84 for the remaining 84 compounds. In all models, these four compounds had high residual values in common. Several factors may contribute to the outlier status of compounds, including uncertainty in the biological activity, limited number of compounds, and the differences in inhibition mechanism. For compound **84**, our model cannot explain the sharp increase in activity observed when simply replacing the chlorine and methyl group in the A or B ring with other groups. Compound **68**, **77**, **88** appear to have a higher activity than expected when compared to the structurally similar compounds **69**, **78**, **87,** respectively. For model B, q^2 was 0.54, reflecting reasonable predictability. Although the numbers of compounds in models C and D were certainly too small to obtain meaningful CoMFA models, models C and D had q^2 values of greater than 0.5. Models E and F exhibited q^2 values of 0.86 and 0.71, respectively, suggesting the goodness of these models. To support this inference, we performed a CoMFA analysis of

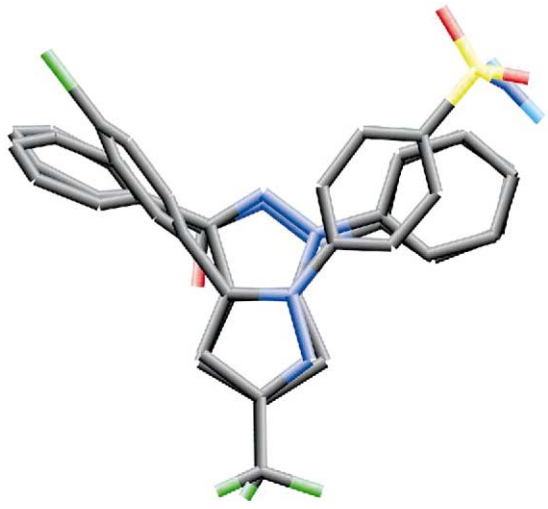


Figure 1. Superposition modes.

Table 1. Statistical results for CoMFA models

Model	A	B	C	D	E	F	G
NC[a]	88	58	18	12	76	70	30
q^{2}[b]	0.84	0.54	0.53	0.5	0.86	0.71	0.67
SDEP[c]	0.46	0.49	0.62	0.83	0.43	0.49	0.51
N[d]	6	4	5	2	6	6	4
r^{2}[e]	0.95	0.84	0.96	0.88	0.97	0.95	0.93
S[f]	0.26	0.29	0.17	0.41	0.2	0.2	0.23
F value[g]	237.96	67.04	62.82	32.6	352.95	192.64	54.84
Ste:Ele[h]	45.3:54.7	43.2:56.8	46:54:00	37.5:62.5	50.7:49.3	52.1:47.9	46.7:53.3

[a] The number of compounds.
[b] Squared correlation coefficient of a cross-validated analysis.
[c] Standard deviation of error of predictions.
[d] Number of PLS components in analysis.
[e] Squared correlation coefficient of a non-cross-validated analysis.
[f] Standard deviation of a non-cross-validated analysis.
[g] F-ratio.
[h] Molecular filed used in CoMFA (Ste: steric field, Ele: electrostatic field).

model G. Omitting the four outliers, as described above, resulted in a significant change in q^2 from 0.16 to 0.67 for model G. Figure 2 shows the cross-validated correlation curves obtained from CoMFA models.

For model A, more rigorous statistical tests, that is, prediction and scrambling test,[50] were performed. Noticeable difference of the statistical parameters of the prediction test ($r^2_{\text{pred}} = 0.74 \pm 0.07$, $s_{\text{pred}} = 0.56 \pm 0.04$) from those of the scrambling test ($q^2 = 0.10 \pm 0.09$, SDEP $= 1.21 \pm 0.05$) reflects the stability and robustness of our CoMFA model.

2.2. CoMSIA analysis

CoMSIA analyses were performed using the following descriptor fields: steric and electrostatic, hydrogen bond donor and acceptor, and lipophilicity. The same models used for the CoMFA study also served as models for the CoMSIA analyses. The PLS results of the CoMSIA models are summarized in Table 2. For the models that included triaryl rings (model A, B, E, and F), the q^2 values and the contributions of the fields demonstrated the importance of these fields to biological activity. However, other models revealed that the hydrogen bond donor and acceptor field did not contribute significantly to activity. PLS analysis of model A generated a q^2 of 0.72 as compared to a q^2 of 0.73 for CoMFA and a SDEP value of 0.63 as compared to a SDEP of 0.60 for CoMFA. The elimination of the four outliers, **68**, **77**, **84**, and **88**, increased q^2 to 0.79 as compared to a q^2 of 0.84 for CoMFA and gave a SDEP of 0.52 as compared to a SDEP of 0.46 for CoMFA. The residual values of these outliers were similar to the values of the CoMFA model as −1.46, −1.43, 1.81 and −1.61 log unit, respectively. PLS results for models B–D also showed reasonable q^2 values, which were similar to the CoMFA results. Models E and F produced good correlations between the chemical classes and activities. For model G, the exclusion of these four outliers improved the correlations from 0.21 to 0.51. Figure 3 illustrates the cross-validated correlation curves.

Table 2. Statistical results for seven CoMSIA models

Model	A	B	C	D	E	F	G
NC[a]	88	58	18	12	76	70	30
q^{2}[b]	0.79	0.51	0.54	0.53	0.78	0.67	0.51
SDEP[c]	0.52	0.5	0.62	1.06	0.53	0.53	0.61
N[d]	6	4	5	5	6	4	4
r^{2}[e]	0.91	0.76	0.95	0.95	0.92	0.89	0.82
S[f]	0.35	0.35	0.2	0.36	0.31	0.31	0.36
F value[g]	122.8	42.72	42.1	18.11	132.87	83.07	34.37
Ste:Ele[h]	8.4:31.4	6.9:31.9	12.8:62.6	22.8:37.6	9.0:29.3	6.8:28.5	16.4:54.0
HP[h]	20	25.6	24.6	39.6	21.6	25.1	29.7
HD:HA[h]	22.7:17.4	13.2:22.4			23.4:16.7	18.9:20.8	

[a] The number of compounds.
[b] Squared correlation coefficient of a cross-validated analysis.
[c] Standard deviation of error of predictions.
[d] Number of PLS components in analysis.
[e] Squared correlation coefficient of a non-cross-validated analysis.
[f] Standard deviation of a non-cross-validated analysis.
[g] F-ratio.
[h] Molecular filed used in CoMSIA (Ste: steric field, Ele: electrostatic field, HP: Hydrophobic field, HD: hydrogen-bond donor field, HA: hydrogen-bond acceptor field).

Prediction and scrambling tests for model A were performed as for CoMFA. As in CoMFA model, the result of prediction test ($r^2_{\text{pred}} = 0.70 \pm 0.05$, $s_{\text{pred}} = 0.59 \pm 0.04$) showed obvious difference from those of the scrambling test ($q^2 = 0.10 \pm 0.09$, SDEP $= 1.13 \pm 0.06$).

2.3. CoMFA and CoMSIA contour maps

To visualize the information content of the model, CoMFA contour maps were generated (Fig. 4). In model A, the green contours around triaryl ring display

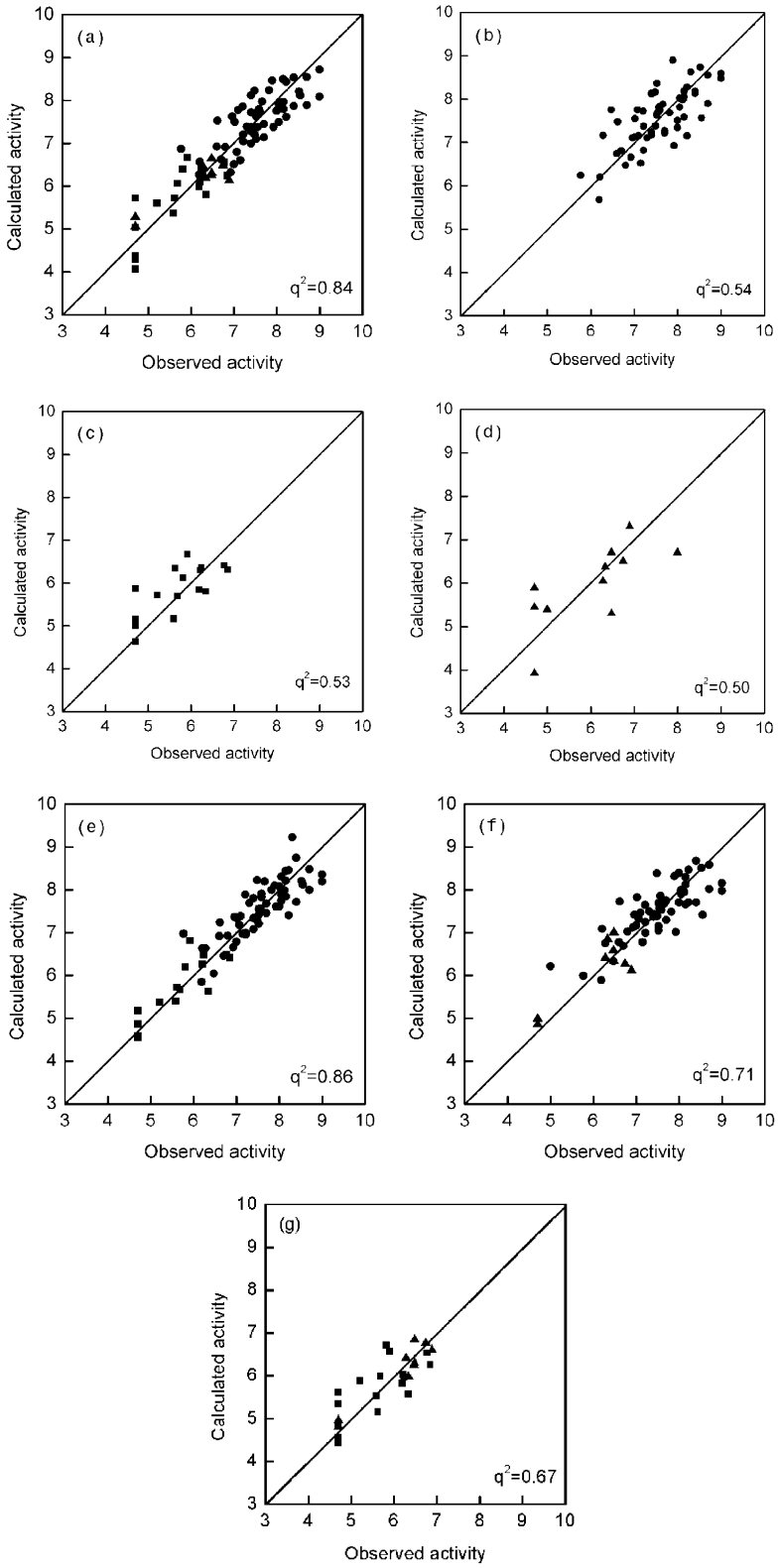


Figure 2. Observed versus calculated activity ($p\text{IC}_{50}$) for the CoMFA models. (a) Model A. (b) Model B. (c) Model C. (d) Model D. (e) Model E. (f) Model F. (g) Model G. ● Triaryl rings. ■ Diaryl cycloalkanopyrazoles. ▲ Diphenyl hydrazides.

that more bulky groups are required to increase activity. The yellow contours near the phenyl rings of diaryl cylcoalkanopyrazole and diphenyl hydrazide suggest that bulky groups in these regions are not beneficial to activity. The blue contours show that groups with positive charge in this area increase activity. The red contour near the trifluoromethyl displays that groups with negative charge in this region are favorable to activity. In model B, green contours near the *para* position of the A ring and C ring indicate favorable steric regions. The yellow contours near the trifluoromethyl and sulfonamide groups indicate intolerance for bulky constituents.

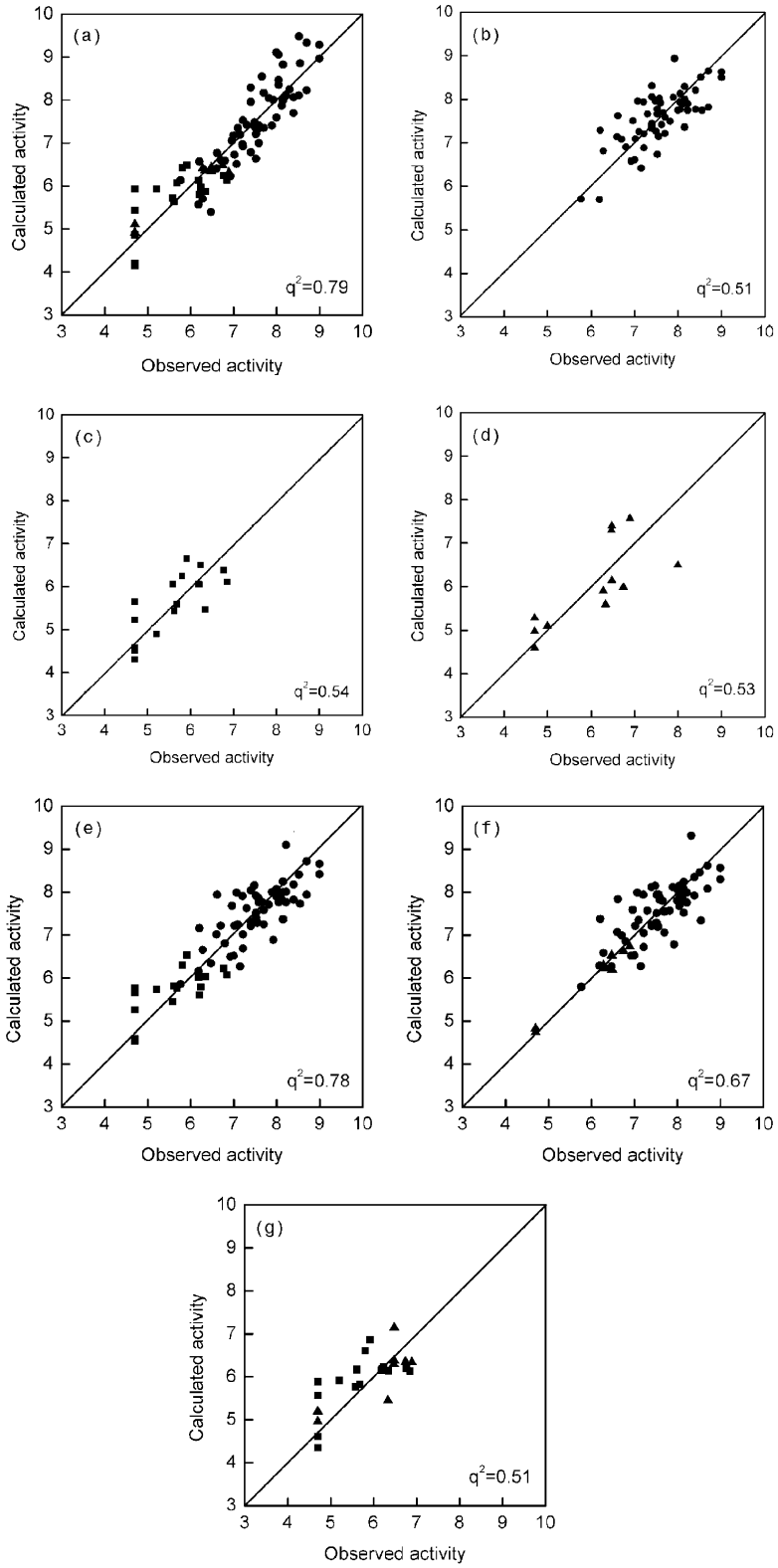


Figure 3. Observed versus calculated activity (pIC_{50}) of CoMSIA models. (a) Model A. (b) Model B. (c) Model C. (d) Model D. (e) Model E. (f) Model F. (g) Model G. ● Triaryl rings. ■ Diaryl cycloalkanopyrazoles. ▲ Diphenyl hydrazides.

The blue region around the A ring is favorable positive potential area. This moiety binds to the hydrophobic pocket of the active site. In model C, the favorable steric region is close to the cycloalkane ring. The yellow region near the *para* position of the B ring represents that less bulky groups are favorable to activity. The blue contour indicates that groups with negative charge are unfavorable to activity. In model D, the yellow regions near the *para* position of the A and B ring indicate that any bulky groups at these positions would decrease

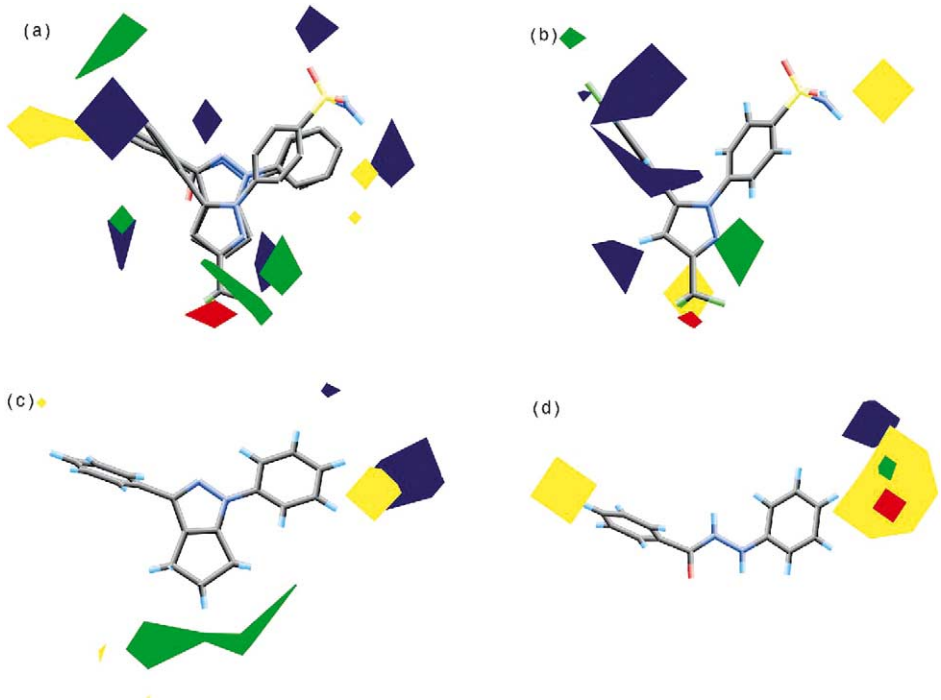


Figure 4. CoMFA STDEV*COEFF contour plots. Green contour (contribution level of 80%) indicates region where bulky group increases activity, whereas yellow contour (contribution level of 20%) indicates region where bulky group decreases activity. Blue contour (contribution level of 80%) indicates region where positive group increases activity, whereas red contour (contribution level of 20%) indicates region where negative charge increases activity. (a) Model A. (b) Model B. (c) Model C. (d) Model D.

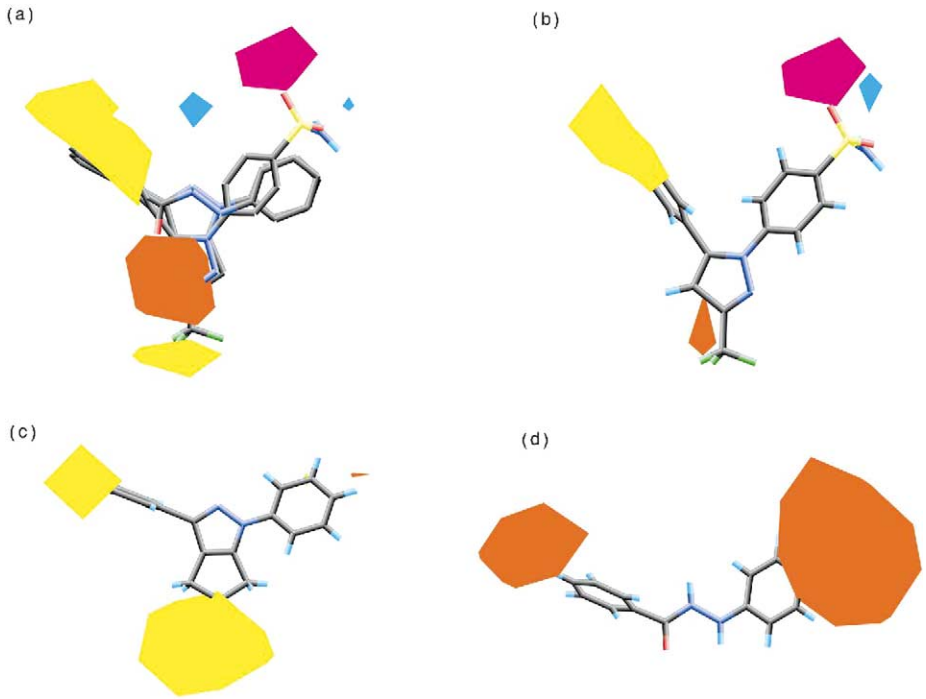


Figure 5. CoMSIA STDEV*COEFF contour plots. Yellow contour (contribution level of 80%) represents area where hydrophobic group enhances activity, and orange contour (contribution level of 20%) indicates regions where hydrophobic group decreases activity. Cyan contour (contribution level of 80%) indicates region where H-bond donor group increases activity, and magenta contour (contribution level of 80%) indicates region where H-bond acceptor group increases activity. (a) Model A. (b) Model B. (c) Model C. (d) Model D.

activity. Electronegative groups in the red region (*para* position of the B ring) enhance activity. The blue contours near the *para* position of the B rings display favorable positive potential region. The steric and electrostatic fields of CoMSIA were generally in accordance with the field distribution of CoMFA maps. The CoMSIA hydrophobic and hydrogen bond donor and acceptor contour plots are represented in Figure 5. In model A, the yellow contours around the A and C rings indicate that the lipophilic substituents at these positions enhance activity. The orange contour near the C ring suggests that hydrophilic groups are favorable to activity. The cyan contour near the sulfonamide group indicates that a hydrogen bond donor at this position enhances activity by forming hydrogen bonds with the appropriate enzyme residues. The magenta contour observed near the sulfonyl group indicates that the substitution of hydrogen bond donor at this position may decrease activity. The CoMSIA map of model B is similar to that of model A. In model C, hydrophobic contours appear near the cycloalkane and the A ring. In model D, the orange contours near the *para* position of the phenyl rings indicate that hydrophilic groups in these regions are beneficial to activity.

2.4. Flexible docking

Flexible docking of all data sets used for the QSAR study was carried out on the active site of a monomeric unit of the catalytic core of the COX-2 enzyme. For the

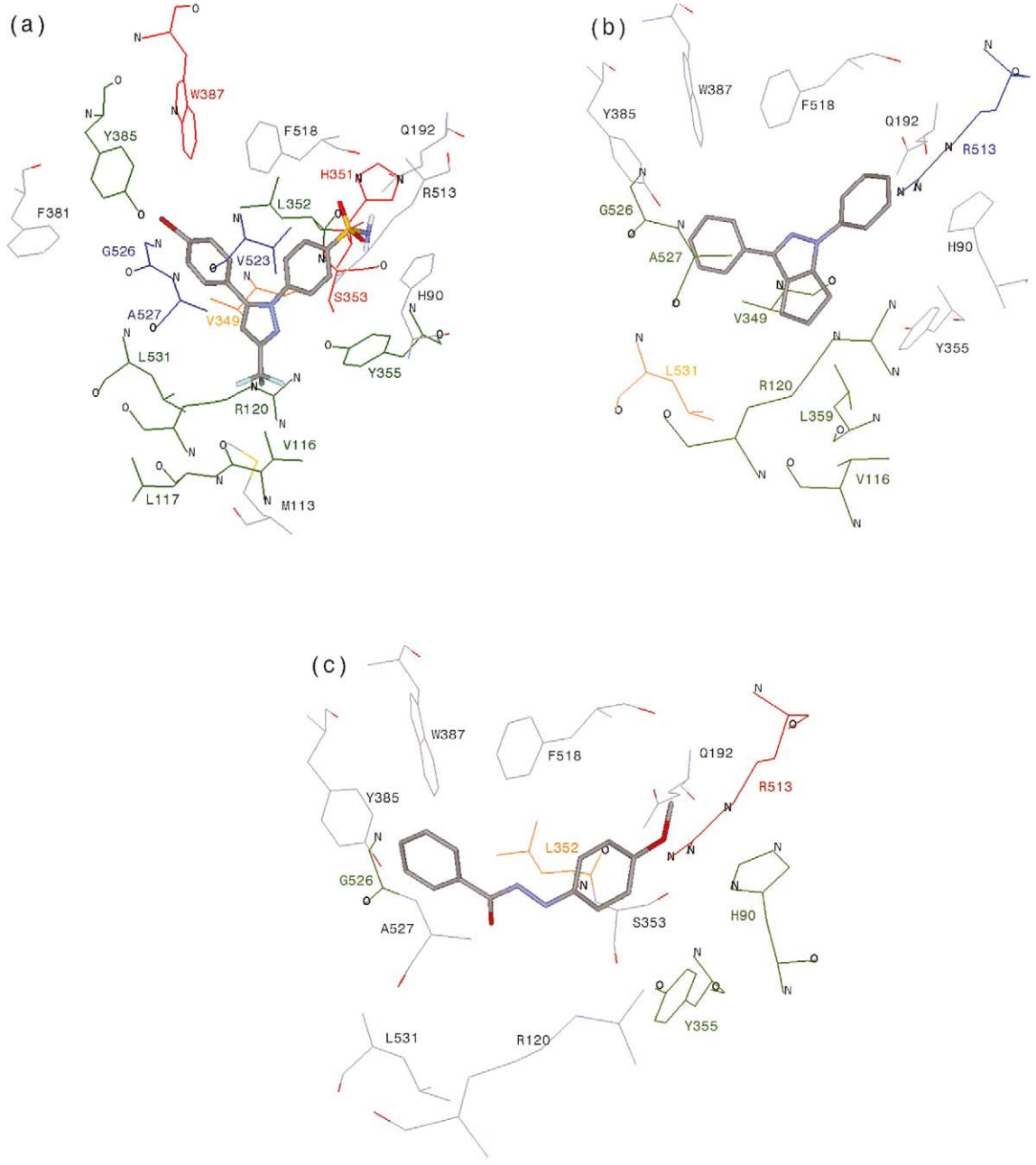


Figure 6. Docking pose and COMBINE Model. Green displays favorable E_V residue, whereas yellow reveals unfavorable E_V residue. Red indicates favorable E_Q residue, whereas blue displays unfavorable E_Q residue. (a) Triaryl rings. (b) Diaryl cycloalkanopyrazoles. (c) Diphenyl hydrazides.

docking study, we applied an initial structure similar to that of SC-558 to all data sets, based on the results of the 3D-QSAR studies. The analysis showed that in all cases the inhibitor-enzyme complex remained stable throughout the simulation without suffering remarkable structural changes. The best possible binding mode of the three representative inhibitors at the COX-2 active site are displayed in Figure 6. For triaryl rings, the sulfonamide group forms hydrogen bonds with His90, Gln192, and Arg513, and with the backbones of Ser353 and Phe518. The hydrophobic residues Leu352, Phe518, and Val523 surround the B ring. The A ring binds in a hydrophobic cavity formed by Phe381, Tyr385, Trp387, Gly526, and Ala527. The trifluoromethyl group binds through hydrophobic interactions in a pocket formed by Met113, Val116, Val349, Tyr355, Leu359, and Leu531. These binding regions have similar feature to the binding of diaryl cycloalkanopyrazoles and diphenyl hydrazides; the A ring binds to the hydrophobic pocket, and the B ring binds near the side pocket. The cycloalkane ring of the diaryl cycloalkanopyrazoles is placed in the pocket in which the trifluoromethyl group bound.

The predicted binding energies of these inhibitors into the active site are listed in Table 3. Least-squares fit analyses were performed to explore whether the binding energies could be correlated with activities. The equations were obtained for the inhibitory activities represented as $p\mathrm{IC}_{50}$ values, using the binding energies as the sole descriptor variable. For triaryl rings, a model with good predictability ($r^2 = 0.51$) was obtained for 58 compounds:

$$p\mathrm{IC}_{50} = -8.75 - 0.22\Delta E$$

For diaryl cycloalkanopyrazoles, compounds **61**, **64**–**65**, **71**, and **76** were not included in the least-squares fit, due to uncertain IC_{50} values. As predicted by the CoMFA model, the size of the active site restricted the binding of compounds **64**–**65** and **71**. The correlation derived using the binding energies of 13 compounds gave a poor model. Compound **75**, in which the R is substituted with *N*-acetylhydroxylamino group, showed an activity much greater than predicted. Omitting compound **75** resulted in a correlation of $r^2 = 0.49$ with the equation:

$$p\mathrm{IC}_{50} = -8.77 - 0.31\Delta E$$

Table 3. Observed activity ($p\mathrm{IC}_{50}$) and predicted binding energy (kcal/mol)

Compd	$p\mathrm{IC}_{50}$	Binding energy	Compd	$p\mathrm{IC}_{50}$	Binding energy
1	7.22	−70.09	**41**	7.48	−74.98
2	7.22	−73.19	**42**	8.7	−74.79
3	7.1	−69.12	**43**	8.22	−72.54
4	7.4	−70.61	**44**	7.59	−74.82
5	6.6	−70.72	**45**	7.57	−74.47
6	7.7	−74.41	**46**	6.21	−71.82
7	7.3	−73.26	**47**	6.19	−68.57
8	7	−71.98	**48**	7.4	−75.86
9	6.8	−69.73	**49**	7.07	−68.96
10	7.4	−72.42	**50**	7.15	−69.52
11	6.7	−69.46	**51**	7.62	−70.97
12	6.92	−72.64	**52**	7.52	−75.25
13	7.52	−75.96	**53**	7.4	−74.15
14	5.77	−68.01	**54**	8.05	−75.34
15	6.28	−71	**55**	8.55	−75.38
16	7.59	−72.27	**56**	7.7	−71.83
17	8.15	−72.92	**57**	7.92	−74.47
18	7.52	−74.54	**58**	8.05	−76.91
19	8.15	−73.79	**59**	5.21	−44.84
20	8.52	−73.33	**60**	5.59	−47.48
21	7.21	−75.1	**62**	6.24	−48.83
22	8.22	−75.4	**63**	6.21	−48.05
23	8.15	−74.85	**66**	6.77	−48.03
24	8.4	−77.33	**67**	6.85	−49.68
25	7.89	−74.44	**68**	5	−47.29
26	8.4	−74.82	**69**	6.19	−48.03
27	6.62	−68.42	**70**	5.81	−47.22
28	8.7	−77.38	**72**	5.91	−47
29	7.49	−72.75	**73**	5.68	−48.63
30	6.96	−71.19	**74**	6.34	−49.42
31	6.92	−74.32	**75**	5.62	−66.62
32	7.55	−72.03	**77**	5	−36.59
33	8	−73.82	**78**	6.48	−43.16
34	7.4	−73.01	**79**	6.47	−39.33
35	8.05	−74.27	**80**	6.48	−38.2
36	7.02	−72.58	**81**	6.28	−37.8
37	8.12	−72.56	**83**	6.74	−44.74
38	9	−73	**84**	8	−46.94
39	9	−75.04	**85**	6.89	−46.71
40	7.66	−72.26	**87**	6.33	−44.99

For diphenyl hydrazide inhibitors, a good correlation of $r^2 = 0.54$ was found between the binding energy alone and biological activity:

$$p\mathrm{IC}_{50} = -0.64 - 0.14\Delta E$$

Compounds **82**, **86**, and **88** were not included in this correlation, due to uncertain IC_{50} values. Figure 7 shows the correlation between binding energies and observed activities for each chemical class, and for the total set. In these binding models, the diaryl cycloalkanopyrazoles and diphenyl hydrazides showed correlations between activities and binding energies that were as good as the correlations for the triaryl rings. Triaryl rings bound to the active site with low energy, whereas the diaryl cycloalkanopyrazoles and diphenyl hydrazides exhibited relatively high binding energies. An inspection of the separate contribution to the binding energy showed that difference of binding energy corresponds almost in difference of electrostatic energy. The mean electrostatic interaction energies between inhibitors and COX-2 were -20.00 ± 3.96, -1.30 ± 1.77, and -2.36 ± 3.36 kcal/mol for triaryl rings, diaryl cycloalkanopyrazoles, and diphenyl hydrazides, respectively. And the electrostatic interaction of -15.43 ± 6.02 kcal/mol was obtained between sulfonamide or methylsulfonyl group of triaryl rings and Arg513.

The experimental evidence that triaryl rings with sulfonyl group appear to be irreversible has been reported previously.[19] Diaryl cycloalkanopyrazole and diphenyl hydrazide inhibitors may belong to the weak binding categories of COX-2 inhibition.

2.5. COMBINE analysis

The COMBINE approach was applied to derive a quantitative model for predicting activity from structural parameters. In a COMBINE model, the values of the weighted PLS pseudo-coefficients can be analyzed to identify the most relevant ligand-residue interactions. Figure 6 displays important interaction residues determined by COMBINE analysis for each three chemical class. For 58 triaryl ring inhibitors, the PLS model exhibited $r^2 = 0.72$, $q^2 = 0.64$, and SDEP = 0.44. Important residues in E_V or E_Q are listed with coefficients in Table 4. Five regions of the binding site were highlighted in the COMBINE analysis. One important region located near the hydrophobic pocket suggests that a more favorable E_V with the Tyr385 and a more favorable E_Q with Trp387 are required for greater activity. A second region is located near the entrance. Residues Val116, Leu117, Arg120, Tyr355, and Leu531 have negative coefficients in E_V and an unfavorable steric region related to the accessibility of inhibitors exists near Val349. Third, a negative correlation of the E_V is observed between the B ring and the main chain atoms of Leu352 surrounding it. The fourth highlighted aspect is the negative coefficients for Ser353 and His351; the backbone of these residues shows a more favorable E_Q with the sulfonamide group than with the methylsulfone group. The final region contains Val523, Gly526, and Ala527 residues, which show positive coefficients in E_Q term. In the COMBINE model for diaryl cycloalkanopyrazoles, the values of $r^2 = 0.89$, $q^2 = 0.63$, and SDEP = 0.35 were obtained. A more favorable E_V with the residues Val116, Arg120, Val349, Leu359,

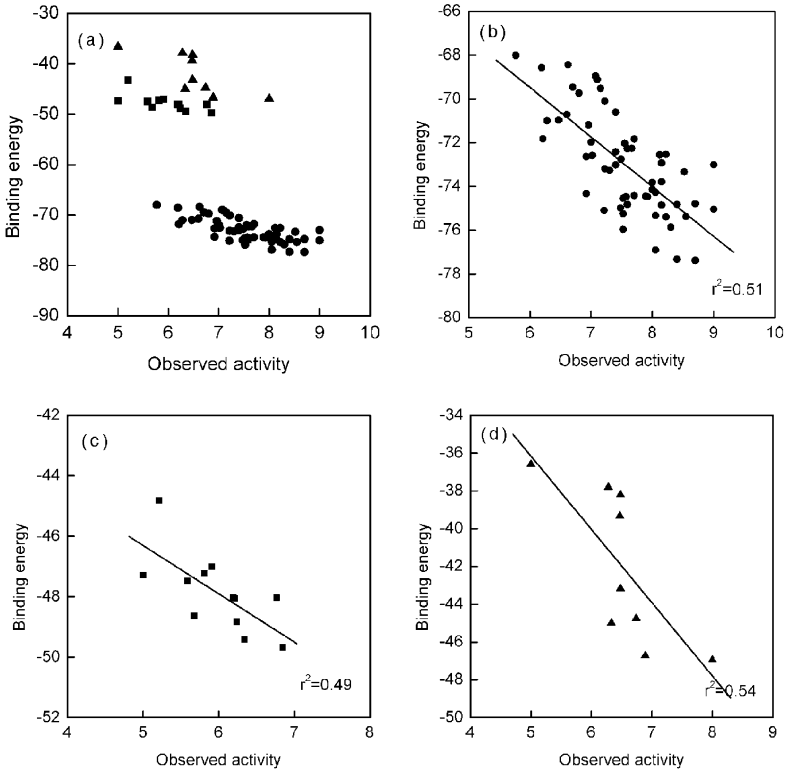


Figure 7. Observed activity ($p\mathrm{IC}_{50}$) versus predicte binding energy (kcal/mol) by flexible docking. (a) Entire inhibitors. (b) Triaryl rings. (c) Diaryl cycloalkanopyrazoles. (d) Diphenyl hydrazides. ● Triaryl rings. ■ Diaryl cycloalkanopyrazoles. ▲ Diphenyl hydrazides.

Table 4. Result of COMBINE analysis

Model	Equation
Triaryl rings	$pIC_{50} = 4.42 - 0.23\ V116^{V} - 0.31\ L117^{V} - 0.14\ R120^{V+} 0.23\ V349^{V} - 0.45\ L352^{V} - 0.18\ Y355^{V} - 0.33\ Y385^{V} - 0.10\ L531^{V} - 0.59 H351^{Q} - 0.15\ S353^{Q} - 0.40\ W387^{Q+} 0.30\ V523^{Q} + 0.27\ L525^{Q} + 0.11\ G526^{Q}$
Diaryl cycloalkanopyrazoles	$pIC_{50} = 2.94 - 0.09\ V116^{V} - 0.12\ R120^{V} - 0.17\ V349^{V} - 0.05\ L359^{V} - 0.23\ G526^{V} - 0.58\ A527^{V+} 0.11\ L531^{V+} 0.10\ R513^{Q}$
Diphenyl hydrazides	$pIC_{50} = 3.85 - 0.70 H90^{V+} 0.18\ L352^{V} - 0.60\ Y355^{V} - 0.32\ G526^{V} - 0.07\ R513^{Q}$

^{V}Van der Waals interaction; ^{Q}Electrostatic interaction.

Gly526, and Ala527 enhance activity. Leu531 exhibits a positive coefficient in E_V term due to the presence of the R group in compounds **73**–**75**. The positive coefficient for Arg513 indicates that unfavorable E_Q with it increases activity.

Good correlations of $r^2 = 0.77$, $q^2 = 0.50$, and SDEP = 0.58 were obtained for the diphenyl hydrazides with a relatively small sample size ($n = 9$). The residues His90, Tyr355, and Gly526 display negative coefficient in E_V term and the positive coefficient for Leu352 indicates that favorable E_V with it decreases activity. The negative coefficient for Arg513 indicates that favorable E_V in this region are beneficial to activity. These COMBINE models show good agreement with 3D-QSAR maps. Figure 8 shows the relationships between the calculated and observed activities.

3. Conclusion

A 3D-QSAR approach using CoMFA and CoMSIA was applied to 88 structurally heterogeneous derivatives known as selective COX-2 inhibitors. The alignment of the compounds is one of the critical elements for 3D-QSAR studies. In the present study, we aligned the ligands onto the template structure of the compound SC-558. This alignment was validated by CoMFA and CoMSIA studies. The model for the entire inhibitor set was validated by prediction and scrambling tests in both 3D-QSAR methods. Flexible docking study showed good correlations between activities and binding energies for each three chemical class. But no correlations were observed between activities and energies for the entire inhibitor set. An inspection of the separate contribution to the binding energy revealed a marked

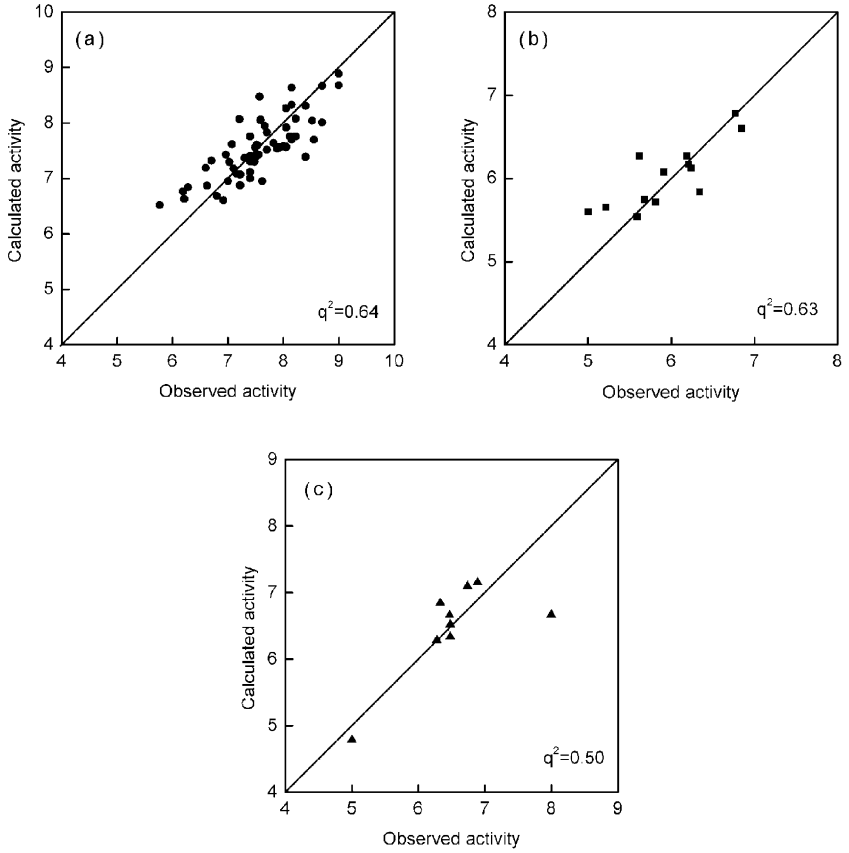


Figure 8. Observed activity versus calculated activity (pIC_{50}) by COMBINE analysis. (a) Triaryl rings. (b) Diaryl cycloalkanopyrazoles. (c) Diphenyl hydrazides.

dependence on the electrostatic interaction energy term. COMBINE analysis showed that electrostatic interaction of inhibitors with Arg513 has responsibility to difference of binding energies. The COMBINE analysis produced reasonably good QSAR models with high cross-validated and conventional r^2 values for all three chemical classes. COMBINE model coincided well to information of QSAR map. In this study, combination of ligand-based and structure-based models may provide a helpful guideline for novel compounds design with enhanced inhibitory activities.

4. Methods

4.1. Molecular modeling

All molecular modeling and 3D-QSAR studies were performed using the SYBYL 6.7[51] molecular modeling software package on a Silicon Graphics workstation (Origin R1000, 256 Mbytes memory, 2 CPU, 180 MHz). All compounds were fully geometry optimized using a Tripos molecular mechanics force field with a distance-dependent dielectric constant and an energy gradient convergence criterion of 0.05 kcal/mol. The partial atomic charges required for the calculation of the electrostatic potential were assigned using the PEOE method.[52]

The crystal structure of murine COX-2 complexed with SC-558 was obtained from the Brookhaven Protein Data Bank (1cx2).[53] The crystal structure was stripped of inhibitor and water molecules, and hydrogen atoms were generated in standard geometry. Protonation states were assumed to be those most common at pH 7, i.e., lysines, arginines, aspartates, and glutamates were ionized.

4.2. CoMFA Analysis

Superimposition of the molecules was carried out using the pose of crystallized SC-558 as the template. The overlapped molecules were surrounded by a 3D grid of points extending in three dimensions to at least 4Å beyond the union volume occupied by the superimposed molecules. The default sp^3 carbon atom with +1 charge was selected as the probe atom for the calculations of steric and electrostatic fields around the aligned molecules. The probe-ligand interaction energies were calculated using a Lennard-Jones 6–12 potential and Coulombic potential with a distance-dependent dielectric, respectively. Values of steric and electrostatic energies were truncated at 30 kcal/mol. The entire set of inhibitors was divided into two groups in the approximate ratio of 2:1, and about 30 inhibitors were used as a test set to assess the predictive power. This experiment was repeated ten times. Furthermore, to detect possible chance correlations, the scrambling test was repeated ten times.[50]

4.3. CoMSIA Analysis

The same grid that was constructed for the CoMFA field calculation was used for the CoMSIA field calculation.[39] Five physicochemical properties (steric, electrostatic, hydrophobic, and hydrogen bond donor and acceptor) were evaluated using a common probe atom with a radius of 1 Å, a +1 charge, hydrophobicity of +1, and hydrogen bond donor and acceptor properties of +1. The contributions from these descriptors were truncated at 0.3 kcal/mol. Validation of the CoMSIA analysis was performed as described for CoMFA.

4.4. PLS Analysis

Partial least squares (PLS) analysis, the statistical method used in deriving the 3D QSAR models, is a variation of principal component regression in which the original variables are replaced by a small subset of linear combinations.[54] PLS analysis was carried out using the leave-one-out option to obtain the optimal number of components to be used subsequently in the final analysis. The cross-validated coefficient q^2 was calculated using formula

$$q^2 = 1 - \frac{\sum(Y_{\text{actual}} - Y_{\text{pred}})^2}{\sum(Y_{\text{actual}} - Y_{\text{mean}})^2}$$

where Y_{pred}, Y_{actual}, and Y_{mean} are predicted, actual, and mean values of the target property ($p\text{IC}_{50}$), respectively. The optimal number of components was designated as that which yielded the highest q^2 values and the smallest rms error values. CoMFA coefficient contour maps were generated by interpolation of the pairwise products between the 3D-QSAR coefficients and the standard deviation of the associated energy variables.

4.5. Flexible docking

Flexible docking was carried out for all 88 compounds in continuum solvation environment using the CHARMM program[40] with the force field version 27. Ligand molecules were manually docked into the active site of protein based on the orientation of SC-558 in the crystal structure. For the nonbonded interaction energy calculation, a distance dependent dielectric constant ($\varepsilon = r$) was used with a 12-Å cutoff. All residues within 15-Å range from the center of SC-558 were included in the simulation, whereas others were fixed to their initial positions.

Each complex was simulated using the following protocol:

1. The system was energy-minimized by 1000 steps of steepest descendent method with the convergence criteria of 0.01 kcal/mol/Å to release the bad contacts.
2. After minimization, the system was gradually heated up to 300 K for 20 ps with a 100 steps interval per 5 K, and equilibrated for 100 ps.
3. The resulting structure was energy-minimized with the same protocol as (i) to get the refined coordinate.

We have checked the adequate convergence of the measured quantities, such as potential energy as a function

of time, and equilibration for each simulation. All of the simulations are well-converged in the equilibration runs (data not shown).

The protein conformation of the crystal and simulated structure did not show any significant difference. Thus we assumed that the complex structure undergo limited conformational changes with respect to the various ligand structure except for several flexible sidechains. The approximated ligand-enzyme binding energy is defined as

$$\Delta E_{\text{bind}} = E_{\text{complex}} - (E_{\text{enzyme}} + E_{\text{ligand}})$$

where E_{bind}, E_{complex}, E_{enzyme}, and E_{ligand} are the binding energy, complex energy, enzyme energy, and ligand energy in the complex state, respectively. The binding energies were correlated with the inhibitory activities for each ligand using the least-squares fit method.

4.6. COMBINE Analysis

COMBINE analysis[42] was performed to assess most important interacting residues in protein related to the biological activities of the ligands. Interaction energies (E_{V} and E_{Q}) between the ligand and the protein for each refined complex obtained from flexible docking were calculated and partitioned on a per residue basis by using the CHARMM program. Only the residues in the range of 6-Å from the ligand molecule were included in this COMBINE analysis. These interaction energies were input to the PLS spreadsheet module of SYBYL program. No weight pretreatment for E_{V} and E_{Q} was performed because the difference in the variances of E_{V} and E_{Q} of residues was considered to be not so significant.[42] Variable selection did not improve the statistics significantly in our experiment (data not shown), and the PLS model was built using all variables as input descriptors.[42]

The final COMBINE model was built with the number of latent variables that produced the most predictive model as in the CoMFA method.

5. Supporting Information

Tables of structures and COX-2 inhibitory activities of triaryl rings (Table S1), diaryl cycloalkanopyrazoles (Table S2), and diphenyl hydrazides (Table S3).

Acknowledgements

We thank Tripos Associates for providing us with the SYBYL program. We thank Nam Sook Kang for many suggestions and advices in preparation of this paper.

References and notes

1. Vane, J. R.; Botting, R. M. *Scand. J. Rheumatol. Suppl.* **1996**, *102*, 9.
2. Dannhardt, G.; Kiefer, W. *Eur. J. Med. Chem.* **2001**, *36*, 109.
3. Fu, J. Y.; Masferrer, J. L.; Seibert, K.; Raz, A.; Needleman, P. *J. Biol. Chem.* **1990**, *265*, 16737.
4. Crofford, L. J. *J. Rheumatol.* **1997**, *24* (Suppl 49), 15.
5. Marnett, L. J. *Curr. Opin. Chem. Biol.* **2000**, *4*, 545.
6. Laneuville, O.; Breuer, D. K.; Dewitt, D. L.; Hla, T.; Funk, C. D.; Smith, W. L. *J. Pharmacol. Exp. Ther.* **1994**, *271*, 927.
7. Marnett, L. J.; Kalgutkar, A. S. *Curr. Opin. Chem. Biol.* **1998**, *2*, 482.
8. Penning, T. D.; Talley, J. J.; Bertenshaw, S. R.; Carter, J. S.; Collins, P. W.; Docter, S.; Graneto, M. J.; Lee, L. F.; Malecha, J. W.; Miyashiro, J. M.; Rogers, R. S.; Rogier, D. J.; Yu, S. S.; Anderson, G. D.; Burton, E. G.; Cogburn, J. N.; Gregory, S. A.; Koboldt, C. M.; Perkins, W. E.; Seibert, K.; Veenhuizen, A. W.; Zhang, Y. Y.; Isakson, P. C. *J. Med. Chem.* **1997**, *40*, 1347.
9. Prasit, P.; Wang, Z.; Brideau, C.; Chan, C. C.; Charleson, S.; Cromlish, W.; Ethier, D.; Evans, J. F.; Ford-Hutchinson, A. W.; Gauthier, J. Y.; Gordon, R.; Guay, J.; Gresser, M.; Kargman, S.; Kennedy, B.; Leblanc, Y.; Léger, S.; Mancini, J.; O'Neill, G. P.; Ouellet, M.; Percival, M. D.; Perrier, H.; Riendeau, D.; Rodger, I.; Tagari, P.; Thérien, M.; Vickers, P.; Wong, E.; Xu, L.-J.; Young, R. N.; Zamboni, R. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1773.
10. Talley, J. J.; Brown, D. L.; Carter, J. S.; Graneto, M. J.; Koboldt, C. M.; Masferrer, J. L.; Perkins, W. E.; Rogers, R. S.; Shaffer, A. F.; Zhang, Y. Y.; Zweifel, B. S.; Seibert, K. *J. Med. Chem.* **2000**, *43*, 775.
11. Riendeau, D.; Percival, M. D.; Brideau, C.; Charleson, S.; Dube, D.; Ethier, D.; Falgueyret, J. P.; Friesen, R. W.; Gordon, R.; Greig, G.; Guay, J.; Mancini, J.; Ouellet, M.; Wong, E.; Xu, L.; Boyce, S.; Visco, D.; Girard, Y.; Prasit, P.; Zamboni, R.; Rodger, I. W.; Gresser, M.; Ford-Hutchinson, A. W.; Young, R. N.; Chan, C. C. *J. Pharmacol. Exp. Ther.* **2001**, *296*, 558.
12. Vane, J. R.; Botting, R. M. *Inflamm. Res.* **1998**, *47* (Suppl 2), S78.
13. Talley, J. J.; Bertenshaw, S. R.; Brown, D. L.; Carter, J. S.; Graneto, M. J.; Kellogg, M. S.; Koboldt, C. M.; Yuan, J.; Zhang, Y. Y.; Seibert, K. *J. Med. Chem.* **2000**, *43*, 1661.
14. Sawdy, R.; Slater, D.; Fisk, N.; Edmonds, D. K.; Bennett, P. *Lancet* **1997**, *350*, 265.
15. Kutchera, W.; Jones, D. A.; Matsunami, N.; Groden, J.; McIntyre, T. M.; Zimmerman, G. A.; White, R. L.; Prescott, S. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4816.
16. Stewart, W. F.; Kawas, C.; Corrada, M.; Metter, E. J. *Neurology* **1997**, *48*, 626.
17. Subbaramaiah, K.; Dannenberg, A. J. *Trends Pharmacol. Sci.* **2003**, *24*, 96.
18. Gierse, J. K.; Koboldt, C. M.; Walker, M. C.; Seibert, K.; Isakson, P. C. *Biochem. J.* **1999**, *339*, 607.
19. Walker, M. C.; Kurumbail, R. G.; Kiefer, J. R.; Moreland, K. T.; Koboldt, C. M.; Isakson, P. C.; Seibert, K.; Gierse, J. K. *Biochem. J.* **2001**, *357*, 709.
20. Bosch, J.; Roca, T.; Catena, J. L.; Llorens, O.; Perez, J.-J.; Lagunas, C.; Fernández, A. G.; Miquel, I.; Fernández-Serrat, A.; Farrerons, C. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1745.
21. Habeeb, A. G.; Praveen Rao, P. N.; Knaus, E. E. *Drug. Des. Res.* **2000**, *51*, 273.
22. Habeeb, A. G.; Praveen Rao, P. N.; Knaus, E. E. *J. Med. Chem.* **2001**, *44*, 2921.
23. Almansa, C.; Alfon, J.; De Arriba, A. F.; Cavalcanti, F. L.; Escamilla, I.; Gomez, L. A.; Miralles, A.; Soliva, R.; Bartroli, J.; Carceller, E.; Merlos, M.; Garcia-Rafanell, J. *J. Med. Chem.* **2003**, *46*, 3463.

24. Marot, C.; Chavatte, P.; Lesieur, D. *Quant. Struct.-Act. Relat.* **2000**, *19*, 127.
25. Desiraju, G. R.; Gopalakrishnan, B.; Jetti, R. K. R.; Raveendra, D.; Sarma, J. A. R. P., et al. *Molecules* **2000**, *5*, 945.
26. Desiraju, G. R.; Sarma, J. A. R. P.; Raveendra, D.; Gopalakrishnan, B.; Thilagavathi, R., et al. *J. Phys. Org. Chem.* **2001**, *14*, 481.
27. Desiraju, G. R.; Gopalakrishnan, B.; Jetti, R. K.; Nagaraju, A.; Raveendra, D.; Sarma, J. A.; Sobhia, M. E.; Thilagavathi, R. *J. Med. Chem.* **2002**, *45*, 4847.
28. Liu, H.; Huang, X.; Shen, J.; Luo, X.; Li, M.; Xiong, B.; Chen, G.; Shen, J.; Yang, Y.; Jiang, H.; Chen, K. *J. Med. Chem.* **2002**, *45*, 4816.
29. Chavatte, P.; Yous, S.; Marot, C.; Baurin, N.; Lesieur, D. *J. Med. Chem.* **2001**, *44*, 3223.
30. Plount Price, M. L.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 9455.
31. Soliva, R.; Almansa, C.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Med. Chem.* **2003**, *46*, 1372.
32. Pouplana, R.; Lozano, J. J.; Perez, C.; Ruiz, J. *J. Comput. Aided. Mol. Des.* **2002**, *16*, 683.
33. Wong, E.; Bayly, C.; Waterman, H. L.; Riendeau, D.; Mancini, J. A. *J. Biol. Chem.* **1997**, *272*, 9280.
34. Palomer, A.; Cabre, F.; Pascual, J.; Campos, J.; Trujillo, M. A.; Entrena, A.; Gallo, M. A.; Garcia, L.; Mauleon, D.; Espinosa, A. *J. Med. Chem.* **2002**, *45*, 1402.
35. Portevin, B.; Tordjman, C.; Pastoureau, P.; Bonnet, J.; De Nanteuil, G. *J. Med. Chem.* **2000**, *43*, 4582.
36. Sui, Z.; Guan, J.; Ferro, M. P.; McCoy, K.; Wachter, M. P.; Murray, W. V.; Singer, M.; Steber, M.; Ritchie, D. M.; Argentieri, D. C. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 601.
37. Chakraborti, A. K.; Thilagavathi, R. *Bioorg. Med. Chem.* **2003**, *11*, 3989.
38. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
39. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
40. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S., et al. *J. Comput. Chem.* **1983**, *4*, 187.
41. York, D. *Canada. J. Phys* **1966**, *44*, 1079.
42. Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. *J. Med. Chem.* **1995**, *38*, 2681.
43. Hashimoto, H.; Imamura, K.; Haruta, J.; Wakitani, K. *J. Med. Chem.* **2002**, *45*, 1511.
44. Hashimoto, H.; Maeda, K.; Ozawa, K.; Haruta, J.; Wakitani, K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 65.
45. Lau, C. K.; Brideau, C.; Chan, C. C.; Charleson, S.; Cromlish, W. A.; Ethier, D.; Gauthier, J. Y.; Gordon, R.; Guay, J.; Kargman, S.; Li, C.-S.; Prasit, P.; Riendeau, D.; Thérien, M.; Visco, D. M.; Xu, L. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 3187.
46. Friesen, R. W.; Dube, D.; Fortin, R.; Frenette, R.; Prescott, S.; Cromlish, W.; Greig, G. M.; Kargman, S.; Wong, E.; Chan, C. C.; Gordon, R.; Xu, L.; Riendeau, D. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2677.
47. Huang, H. C.; Chamberlain, T. S.; Selbert, K.; Koboldt, C. M.; Isakson, P. C.; Reitz, D. B. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2377.
48. Bertenshaw, S. R.; Talley, J. J.; Rogier, D. J.; Graneto, M. J.; Koboldt, C. M.; Yan, Z. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2827.
49. Therien, M.; Brideau, C.; Chan, C. C.; Cromlish, W. A.; Gauthier, J. Y.; Gordon, R.; Greig, G.; Kargman, S.; Lau, C. K.; Leblanc, Y.; Li, C.-S.; O'Neill, G. P.; Reindeau, D.; Roy, P.; Wang, Z.; Xu, L.; Prasit, P. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 47.
50. van der Voet, H. *Chemometrics and Intelligent Laboratory Systems* **1994**, *25*, 313.
51. Tripos Associates, I. **1699**, MO 63144.
52. Gasteiger, H.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
53. Kurumbail, R. G.; Stevens, A. M.; Gierse, J. K.; McDonald, J. J.; Stegeman, R. A., et al. *Nature* **1996**, *384*, 644.
54. Norinder, U.; Rivera, C.; Unden, A. *J. Pept. Res.* **1997**, *49*, 155.