

3D-QSAR study of microsomal prostaglandin E₂ synthase (mPGES-1) inhibitors

Amor A. San Juan · Seung Joo Cho

Received: 30 September 2006 / Accepted: 15 January 2007 / Published online: 28 March 2007
© Springer-Verlag 2007

Abstract Microsomal prostaglandin E₂ synthase (mPGES-1) has been identified recently as a novel target for treating pain and inflammation. The aim of this study is to understand the binding affinities of reported inhibitors for mPGES-1 and further to design potential new mPGES-1 inhibitors. 3D-QSAR-CoMFA (comparative molecular field analysis) and CoMSIA (comparative molecular similarity indices analysis) - techniques were employed on a series of indole derivatives that act as selective mPGES-1 inhibitors. The lowest energy conformer of the most active compound obtained from systematic conformational search was used as a template for the alignment of 32 compounds. The models obtained were used to predict the activities of the test set of eight compounds, and the predicted values were in good agreement with the experimental results. The 3D-QSAR models derived from the training set of 24 compounds were all statistically significant (CoMFA; $q^2=0.89$, $r^2=0.95$, $r_{bs}^2=0.98$, $r_{pred}^2=0.83$ and CoMSIA; $q^2=0.84$, $r^2=0.93$, $r_{bs}^2=0.93$, $r_{pred}^2=0.94$). Contour plots generated for the CoMFA and CoMSIA models reveal useful clues for improving the activity of mPGES-1

inhibitors. In particular, substitutions of an electronegative fluorine atom or a bulky hydrophilic phenoxy group at the *meta* or *para* positions of the biphenyl rings might improve inhibitory activity. A plausible binding mode between the ligands and mPGES-1 is also proposed.

Keywords 3D-QSAR · Drug design · Inflammation · mPGES-1 · Pain

Introduction

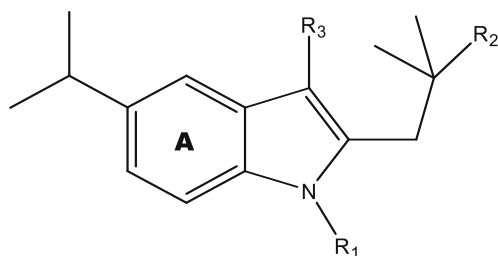
Pain is often associated with diseases of the bones, muscles, joints and skin. Pain and inflammation may occur as a reaction of living tissue to injury, which are mediated by various receptors [1–4] and signaling protein systems [5, 6]. One of these is prostaglandin (PG) [7] which acts as a biological mediator producing signals in the human body that in turn induce pain and inflammation.

The biosynthesis of prostaglandin proceeds through three sequential enzymatic steps [8]: phospholipase-A₂, cyclooxygenase (COX) and several lineage-specific terminal prostaglandin synthases. The phospholipase-A₂ catalyzes the formation of arachidonic acid, then passes into the COX pathway to form either prostaglandin D, E and F. There are three forms of prostaglandin E synthase (PGES) namely, microsomal prostaglandin E₂ synthase-1 (mPGES-1), microsomal prostaglandin E₂ synthase-2 (mPGES-2) and cytosolic PGES. The pathway linkage preference of mPGES-1, mPGES-2 and cPGES is COX-2, both COX-1 and COX-2, and COX-1, respectively [9]. This suggests that microsomal PGES (mPGES-1) acts downstream to cyclooxygenase-2 (COX-2) and affects PGE₂ production [10]. The mPGES-1 is an inducible enzyme that catalyzes the conversion of prostaglandin endoperoxide (PG)H₂ to PGE₂. PGE₂ in turn controls biological activities such as

A. A. San Juan · S. J. Cho
Biochemicals Research Center, Life Science Division,
Korea Institute of Science and Technology,
Cheongryang,
P.O. Box 131, Seoul 136-791, South Korea

A. A. San Juan
School of Science, University of Science and Technology,
52 Eoeun-dong, Yuseoung-gu,
Daejeon 305-333, South Korea
e-mail: asanjuan@kist.re.kr

S. J. Cho (✉)
Life Science Division, Korea Institute of Science and Technology,
Seoul 130-650, South Korea
e-mail: chosj@kist.re.kr

Table 1 Indole derivatives from MK-886 compound as inhibitors of mPGES-1

Compound	Substituent			IC ₅₀ ^a	pIC ₅₀ ^b
	R1	R2	R3		
1 ^c	CH ₂ (4-Cl-Ph)	CO ₂ H	<i>S-tert</i> Bu	1.60	-0.204
2	H	CO ₂ H	<i>S-tert</i> Bu	10	-1.041
3	Me	CO ₂ H	<i>S-tert</i> Bu	10	-1.041
4	CH ₂ (CH=CH ₂)	CO ₂ H	<i>S-tert</i> Bu	6.70	-0.826
5	(CH ₂) ₃ Ph	CO ₂ H	<i>S-tert</i> Bu	3.20	-0.505
6	CH ₂ (4-Cl-Ph)	CO ₂ Me	<i>S-tert</i> Bu	7.20	-0.857
7 ^c	CH ₂ (4-Cl-Ph)	CONH ₂	<i>S-tert</i> Bu	10	-1.041
8	CH ₂ (4-Cl-Ph)	CO ₂ H	Ph	6.40	-0.806
9 ^c	CH ₂ (4-Cl-Ph)	CO ₂ H	Oph	0.650	0.187
10	CH ₂ (4-Cl-Ph)	CO ₂ H	CH ₂ (4- <i>tert</i> Bu-Ph)	0.290	0.538
11	CH ₂ (4-Cl-Ph)	CO ₂ H	CO(2-Me-Ph)	0.900	0.046
12	CH ₂ (4-Cl-Ph)	CO ₂ H	COCH ₂ <i>S-tert</i> Bu	0.260	0.585
13	CH ₂ (4-Cl-Ph)	CO ₂ H	COCH ₂ - <i>tert</i> Bu	0.250	0.602
14	CH ₂ (4-Cl-Ph)	CO ₂ H	Me	1.10	-0.041

^a IC₅₀=50% inhibition (μM) to mPGES-1 enzyme

^b pIC₅₀=-log IC₅₀

^c = included in the test set of compounds

relaxation and contraction of muscles. Hence, an increased level of mPGES-1 is one of the cause of triggering pain and inflammation in the body. Studies [11, 12] have shown that mPGES-1 is responsible for the production of PGE₂ to mediate pain during an inflammatory response. Therefore, when the body experiences pain or inflammation, the mPGES-1 stimulation occurs and its level is increased. On the other hand, a decreased level of mPGES-1 may also reduce the inflammatory response as was reported by Uematsu et al. [13]. A recent gene targeting study [8] revealed that mPGES-1 enzyme represents a novel target for anti-inflammatory and anti-cancer drugs. Thus, the modulation of mPGES-1 is among the current goals in anti-inflammatory drug development.

Few reported compounds act as inhibitors of mPGES-1. NS-398 [2-cyclohexyloxy-4-nitrophenyl)-methanesulfonamide] compound is a COX-2 inhibitor that inhibits mPGES-1 with an IC₅₀ of 20 μM in vitro [14]. Furthermore, 15-deoxy-Δ^{12,14}-prostaglandin J₂ inhibits mPGES-1 with an IC₅₀ of 0.3 μM [15]. Recently, a series of MK-886 compounds

showed selectivity and higher activity against the inducible mPGES-1 with the lowest IC₅₀ value found being 3 nM [16].

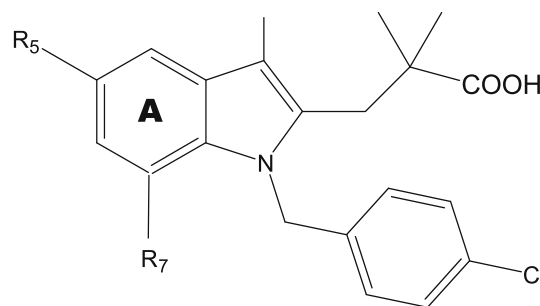
The aim of this study is to gain insights into the molecular interactions between MK-886 analogues that influence their binding to the mPGES-1 protein. We have applied comparative molecular field analysis (CoMFA) [17] and comparative molecular similarity indices analysis (CoMSIA) [18] 3D-QSAR techniques to a series of MK-886 compounds. Further, based on the 3D-QSAR models obtained we have proposed a potent new inhibitor to mPGES-1.

CoMFA has the ability to predict the biological activity of molecules and to represent the relationship between steric and electrostatic fields with the biological activity through 3D contour maps. CoMSIA is a complementary technique that includes the hydrophobic and hydrogen bonding interactions of the molecules as well. Since steric, hydrophobic, hydrogen bond and electrostatic interactions are crucial to whether a molecule will interact optimally at its active site, it is logical to model the distribution of these potential interactions in space around the molecule to identify regions that are active ligands.

Methodology

Data sets

Thirty-two MK-886 derivatives reported by Riendeau et al. [16] were used as input data for 3D-QSAR analyses. Their

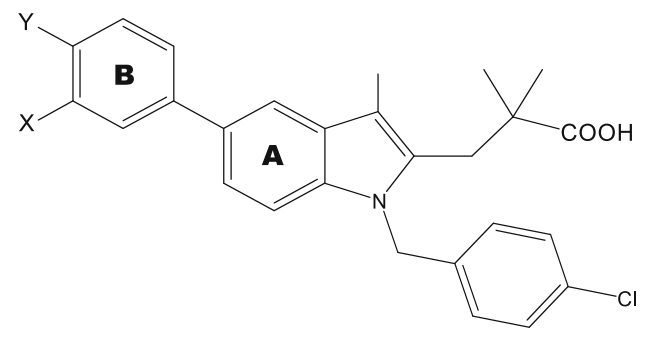
Table 2 Structure activity relationship at C₅ and C₇ positions of indole as inhibitors of mPGES-1

Compound	Substituent		IC ₅₀ ^a	pIC ₅₀ ^b
	R5	R7		
15	H	<i>iso</i> -propyl	4.3	-0.633
16 ^c	H	H	3.2	-0.505
17	F	H	2.6	-0.415
18	<i>tert</i> -butyl	H	0.330	-0.481
19	Ph	H	0.600	0.222

^a IC₅₀=50% inhibition (μM) to mPGES-1 enzyme

^b pIC₅₀=-log IC₅₀

^c = included in the test set of compounds

Table 3 Structure activity relationship of 5-phenyl indole as inhibitors of mPGES-1


Compound	Substituent		IC ₅₀ ^a	pIC ₅₀ ^b
	X	Y		
20 ^c	Ph	H	0.160	0.796
21	H	Ph	0.016	1.796
22	Cl	Ph	0.022	1.658
23	F	Ph	0.007	2.155
24 ^c	F	1,3-pyrazinyl	0.032	1.495
25	F	3-pyridinyl	0.012	1.921
26 ^c	F	2-MeO-Ph	0.005	2.301
27	F	2-Cl-Ph	0.004	2.398
28	F	2-F-Ph	0.008	2.097
29	F	2-MeCO-Ph	0.006	2.222
30	F	2-Me-Ph	0.003	2.523
31 ^c	F	3-Me-Ph	0.033	1.481
32	F	4-Me-Ph	0.031	1.509

^a IC₅₀=50% inhibition (μ M) to mPGES-1 enzyme^b pIC₅₀=-log IC₅₀^c = included in the test set of compounds

structures and potencies are given in Tables 1, 2 and 3. Careful selection of training and test sets is critical to obtain reliable validated QSAR model. Manual selection of training and test sets was based on structural diversity and wide range of activity. The inhibition constant (IC₅₀) values, i.e., the concentration (μ M) of inhibitor that produces 50% inhibition of mPGES-1 were converted into pIC₅₀ (-logIC₅₀) and subsequently used as dependent variables in the QSAR analyses.

Molecular modeling and alignment

Molecular modeling calculations were performed by SYBYL 7.2 [19] on a Silicon Graphics Octane (R1200) workstation with an IRIX 6.5 operating system. All structures were minimized using the Tripos force field [20] and Gasteiger-Huckel charges. The geometry optimization of the compounds was carried out by removing the constraints using conjugate gradient at convergence criterion of 0.005 kcal mol⁻¹.

In standard CoMFA procedure, bioactive conformations [21] are desired for superimposing ligands. In the absence of available crystallographic data information on mPGES-1 and inhibitor complexes, we assumed that the active conformer corresponds to the energetically minimized conformation. Systematic search was applied to compound **30** (the most active compound), with seven torsions by which conformational analysis was considered for all rotatable bonds in 10° increments from 0° to 360°. Conformational energies were computed with electrostatic terms included and the low energy conformer was used as template for superimposition. All 32 compounds were aligned to atoms of the template (Fig. 1a, mark atoms in blue color) by the atom-fit method.

CoMFA and CoMSIA

Models of steric and electrostatic fields for CoMFA were based on both Lennard-Jones and Coulomb potentials [17]. Steric and electrostatic energies were calculated using an sp³ carbon probe atom with van der Waals radius of 1.52 Å and a +1 charge. Energies were truncated to ± 30 kcal mol⁻¹ and the electrostatic contributions were ignored at lattice

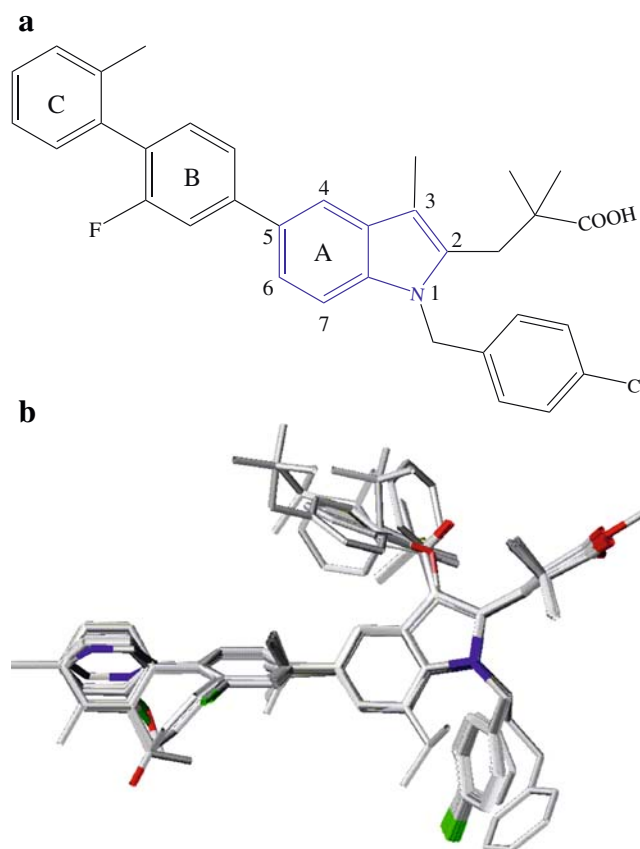


Fig. 1 (a) The two-dimensional structure of compound **30** with designated codes of phenyl moieties (b) The alignment by atom-fit of 32 indole derivatives. The lowest energy conformer of the most active compound (compound **30**) derived from systematic search was used as the template for alignment

Table 4 Summary of CoMFA and CoMSIA results by PLS analysis

	CoMFA		CoMSIA		
		S E	S E H	S E D A	S E D A H
q^2	0.89	0.88	0.89	0.84	0.86
C	3	2	2	2	2
SEP	0.42	0.43	0.42	0.49	0.46
F	180.07	183.31	205.21	191.77	206.76
r^2	0.95	0.93	0.93	0.93	0.93
SEE	0.28	0.25	0.32	0.33	0.32
r_{pred}^2	0.83	0.93	0.87	0.94	0.91
r_{bs}^2	0.98	0.93	0.93	0.93	0.95
SD_{bs}	0.01	0.02	0.02	0.02	0.01
Field Contribution					
S	57.90	53.80	35.60	41.50	29.00
E	42.10	46.20	32.30	35.30	25.90
H			32.00		26.30
D				13.70	10.80
A				9.50	8.00

q^2 = Cross-validated r^2 by leave-one-out method

r_{pred}^2 = Predictive r^2

C = Number of components

r_{bs}^2 = Bootstrap run (10)

SEP = Standard error of prediction

SD_{bs} = Standard deviation from bootstrap

F = F-test value

S = Steric

r^2 = non-cross-validated r^2

E = Electrostatic

SEE = Standard error of estimate

H = Hydrophobic

D = Hydrogen bond donor

A = Hydrogen bond acceptor

interactions with maximum steric interactions. The CoMFA fields generated automatically were block scaled by the CoMFA standard method in SYBYL.

CoMSIA similarity indices descriptors were derived with the same lattice box used in the CoMFA calculations. All five CoMSIA similarity index fields available within SYBYL (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) were evaluated using the probe atom. The CoMSIA models from hydrophobic and hydrogen bonds were calculated between the grid point and each atom of the molecule by a Gaussian distance function [18]. The attenuation factor's default value of 0.30 was used, which is the standard distance dependence of molecular similarity. The effect of using the standard attenuation factor is shown in contour maps with prominent molecular features. Higher value of attenuation factor resulted in fewer molecular features.

Partial least square calculations

Partial least squares (PLS) [22, 23] methodology was employed for all 3D-QSAR modeling analysis. The

CoMFA and CoMSIA fields were used as independent variables, and the pIC_{50} values were utilized as dependent variables. The result of these analyses corresponds to a regression equation with thousands of coefficients. To select and verify the best model, leave-one-out (LOO) [24] cross-validation was employed. Cross-validation tests a model by omitting one or more compounds (rows) from the analysis, re-deriving the model, and then predicting the omitted observations. The cross-validated coefficient, q^2 , was calculated using Eq. 1 below:

$$q^2 = 1 - \frac{\sum(Y_{\text{predicted}} - Y_{\text{observed}})^2}{\sum(Y_{\text{observed}} - Y_{\text{mean}})^2} \quad (1)$$

where $Y_{\text{predicted}}$, Y_{observed} , and Y_{mean} are predicted, actual, and mean values of the target property (pIC_{50}), respectively. $\sum(Y_{\text{predicted}} - Y_{\text{observed}})^2$ is the predictive sum of squares (PRESS).

To determine the optimum number of PLS components (C) and avoid overfitting the data, the number of components corresponding to the lowest PRESS value was employed to identify the final PLS regression models. In addition to q^2 and C, the conventional correlation coefficient r^2 and the standard error of prediction (SEP) were also computed. The visualization of results from the best CoMFA and CoMSIA models in 3D was performed by plotting the standard deviation coefficient with default values of 80 % favored and 20 % disfavored contributions.

Three methods of model validation

Bootstrapping

To obtain statistical confidence limits for the analyses' statistics, bootstrapping [24, 25] (10 runs) was performed at optimum C value in which the r^2 mean is given as q^2 bootstrap. Bootstrapping employs random sampling to

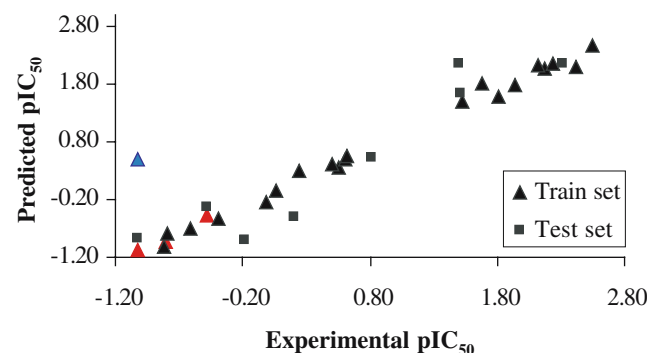


Fig. 2 The correlation ($r^2=0.92$) between experimental pIC_{50} and CoMFA predicted pIC_{50} for training and test sets

Table 5 Experimental and predicted activities with residuals by CoMFA and CoMSIA

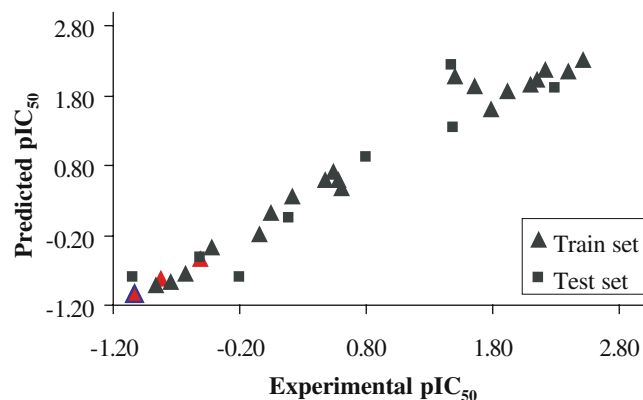
Compound	pIC ₅₀ ^a	CoMFA		CoMSIA	
		PA ^b	R ^c	PA ^b	R ^c
1 ^d	-0.20	-0.84	0.63	-0.81	0.60
2	-1.04	0.56	-1.60	-1.05	0.01
3	-1.04	-1.01	-0.03	-1.04	0.00
4	-0.83	-0.86	0.03	-0.83	0.00
5	-0.51	-0.40	-0.11	-0.53	0.03
6	-0.86	-0.92	0.06	-0.91	0.05
7 ^d	-1.04	-0.80	-0.24	-0.81	-0.23
8	-0.81	-0.72	-0.09	-0.73	-0.08
9 ^d	0.19	-0.42	0.61	0.06	0.13
10	0.54	0.44	0.10	0.70	-0.16
11	0.05	0.03	0.02	0.13	-0.08
12	0.59	0.56	0.02	0.60	-0.02
13	0.60	0.64	-0.04	0.47	0.13
14	-0.04	-0.18	0.14	-0.18	0.14
15	-0.63	-0.62	-0.01	-0.74	0.11
16 ^d	-0.51	-0.27	-0.24	-0.51	0.01
17	-0.42	-0.45	0.04	-0.37	-0.05
18	0.48	0.48	0.00	0.59	-0.11
19	0.22	0.37	-0.15	0.35	-0.13
20 ^d	0.80	0.60	0.20	0.92	-0.13
21	1.80	1.66	0.14	1.60	0.20
22	1.66	1.90	-0.24	1.93	-0.27
23	2.16	2.15	0.00	2.03	0.13
24 ^d	1.50	1.71	-0.22	1.34	0.16
25	1.92	1.87	0.05	1.86	0.06
26 ^d	2.30	2.24	0.06	1.91	0.39
27	2.40	2.17	0.23	2.16	0.24
28	2.10	2.20	-0.10	1.97	0.13
29	2.22	2.23	-0.01	2.18	0.04
30	2.52	2.53	-0.01	2.32	0.20
31 ^d	1.48	2.24	-0.76	2.25	-0.77
32	1.51	1.56	-0.05	2.09	-0.58

^apIC₅₀ = -log IC₅₀ as experimental activity^bPA = predicted activity^cR = residuals^d= included in the test set of compounds

pseudo-replicates of compounds from the original dataset. A confidence measure by bootstrap validation is determined by the standard deviation of the statistics obtained from bootstrapping.

Predictive R squared r_{pred}^2

Another way to validate a model derived from CoMFA or CoMSIA is to evaluate the correlation between actual and predicted activities. The biological activities of the eight test set of compounds were predicted using the model derived from the training set. The predictive ability of the model is expressed by predictive r^2 values, which is similar

**Fig. 3** The correlation ($r^2=0.98$) between experimental pIC₅₀ and CoMSIA predicted pIC₅₀ for training and test sets

to cross-validated $r^2(q^2)$ and is calculated by using the formula below:

$$r_{pred}^2 = \frac{SS - PRESS}{SS}$$

wherein, SS is the sum of squared deviation between biological activities of the test set compounds and mean activity of the training set compounds; and PRESS is the sum of squared deviation between observed and predicted activities of the test set compounds.

Progressive scrambling

Progressive scrambling [26] evaluates the sensitivity of a QSAR model to small changes in the dataset. Thirty scramblings were carried out with a maximum of 10 bins and a minimum of two bins with a critical point of 0.85. Progressive scrambling of the biological data produces three statistics namely q^2 , cSDEP and slope of q^2 . The susceptibility of the model to chance correlation can be gauged by the slope of q^2 (as originally determined using SAMPLS) with respect to the correlation of the original biological activity versus the scrambled biological activity. The slope is denoted as dq^2/dx_{yy} , wherein, an ideal model has the slope equal to 1. In a model where the correlation of activity is not by chance, the slope should be from the range of 0.8 to 1.2.

Results and discussion

CoMFA model

A CoMFA model was derived for a set of indole-based mPGES-1 inhibitors. The analysis was carried out by aligning the compounds to the low energy conformer obtained from systematic conformational search of compound **30**. Figure 1b shows the alignment of the 32 compounds noted by molecular superposition of the biphenyl, indole and carboxylic moieties. To derive a model from CoMFA, a PLS

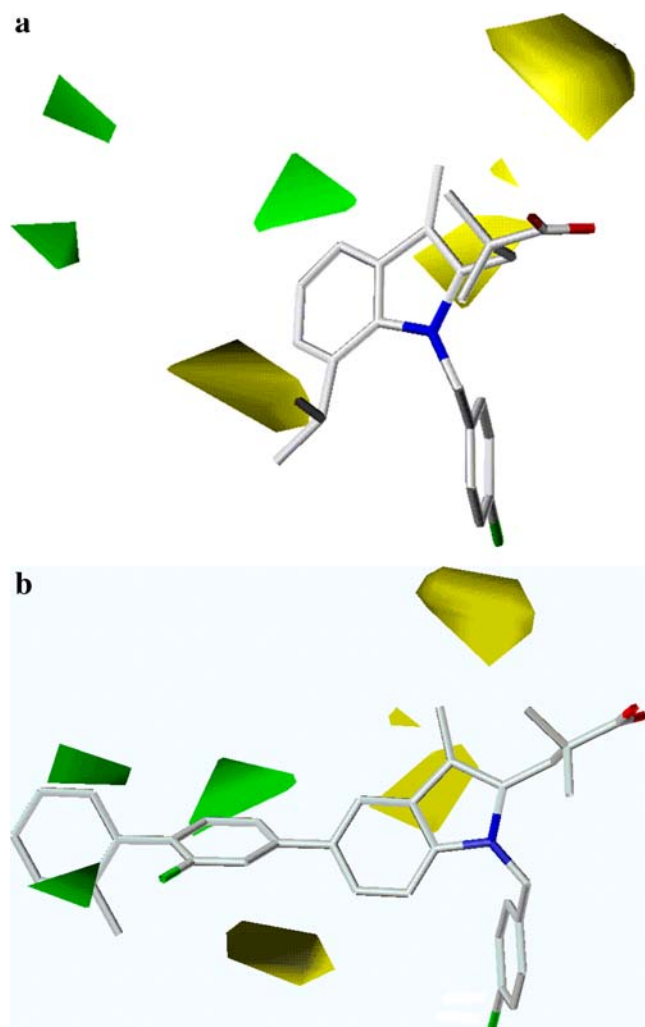


Fig. 4 CoMFA steric contour maps for compound **15** (mark **a**, lower activity) as compared to compound **30** (mark **b**, highest activity). The sterically favored green regions were found on rings B and C, whereas sterically disfavored regions were found near the substituents of rings A and B

analysis was conducted to correlate the variations of biological activities of ligands to the chemical structures as a function of variations in CoMFA fields. Table 4 shows the results obtained: cross-validated q^2 of 0.89 with 3 components, non-cross-validated r^2 of 0.95, bootstrapped r^2 of 0.98 and predictive r^2 of 0.83. These values indicate a good statistical correlation and reasonable predictability of the CoMFA model. Furthermore, the high correlation ($r^2=0.92$) between the actual and predicted values of activities suggests that the CoMFA model is reliable (see Fig. 2 and Table 5). The plot of the predicted versus experimental activity values show that compound **2** is an outlier in CoMFA (blue color, Fig. 2), but not in CoMSIA (Fig. 3). The outlier that is poorly predicted by CoMFA has a residual of -1.60 (Table 5). This is probably due to the unique shape of compound **2** (hydrogen at N_1 position of indole moiety) as

compared with bulky substituents (red color, compounds **3–5**).

The steric and electrostatic contributions to activity were 57.90 % and 42.10 %, respectively. The CoMFA steric map (Fig. 4) encompasses green contours (80 % contribution) corresponding to the regions in space where the steric bulk confers an increase in activity. The green polyhedra along the rings B and C of compound **30** (refer to Fig. 1a for benzene codes) suggest that steric substitutions are favorable. In particular, the presence of a phenyl group at ring C (compound **30**) results to higher activity, as compared to compound **15** with hydrogen substitution that results to poor activity. This is consistent with the fact that the relocation of isopropyl group (compound **15**) from R_5 to R_7 or its removal (compound **16**) resulted to a 3 to 4 fold decrease in potency as compared to compound **14**, suggesting that a bulky group is favored.

The yellow regions (20 % contribution) near the vicinity of ring B and surrounding the ring A of compound **30**

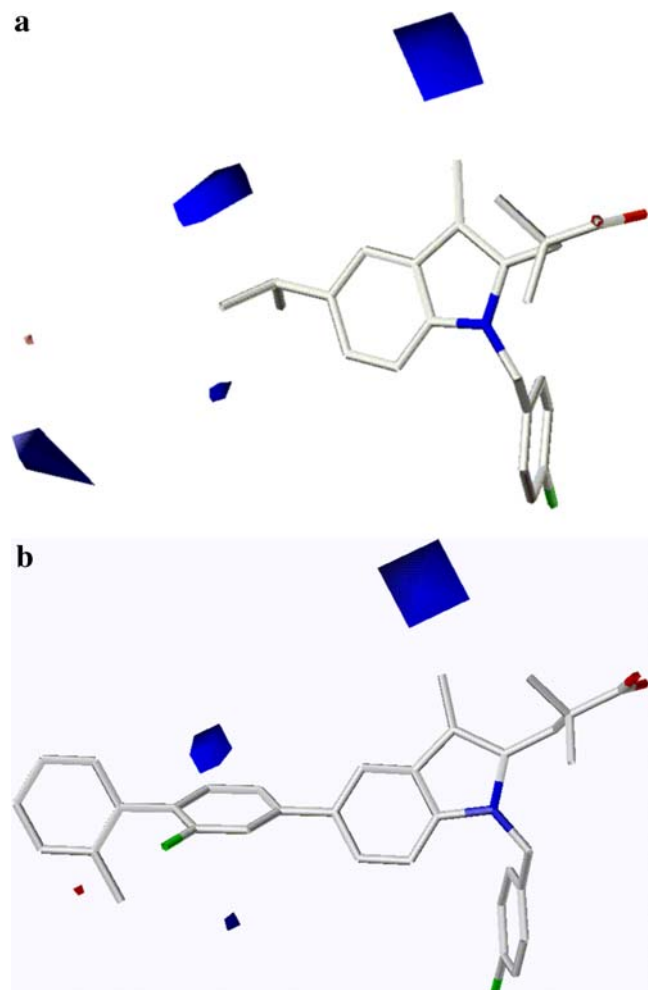


Fig. 5 CoMFA electrostatic contour maps for compound **14** (mark **a**, lower activity) as compared to compound **30** (mark **b**, highest activity). The electropositive favored blue regions were found around the ring B and pyrrole moiety near ring A

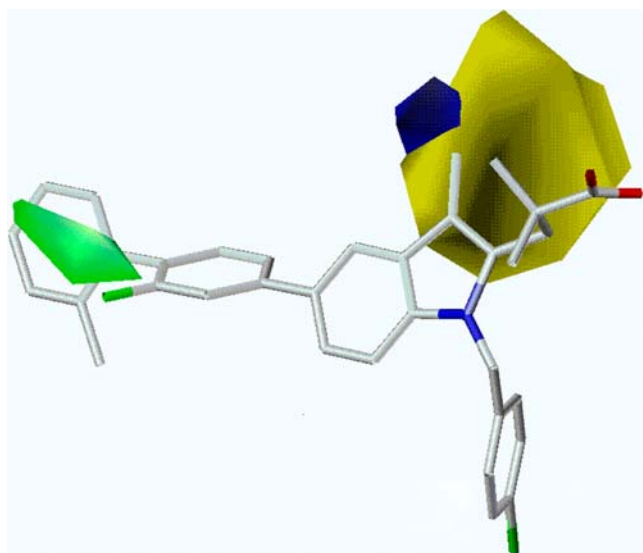


Fig. 6 CoMSIA steric and electrostatic maps. Green (sterically favored), yellow (sterically disfavored), and blue (electropositive favored) regions. Compound **30**, the most active inhibitor, is overlaid in the contour map

indicate that activity decreases due to increased steric bulk. This agrees with the compound **20** in which the incorporation of phenyl group in the X position (Table 3) of ring B results to lower activity. Furthermore, in compound **8** the presence of bulky phenyl at R₃ position (Table 1) of indole results to poor activity. In contrast, when a methyl group (compound **14**) is inserted into the same position, the activity is enhanced.

The CoMFA electrostatic map (Fig. 5), displays blue contours (80 % contribution) at regions in space where electropositive group is associated with increased activity. The insertion of a methyl group at the C₃ position of the indole moiety (compound **14**) indicates that an electropositive group is associated with increased activity. As a consequence, when the electronegative fluorine is connected to the R₅ position in ring A (Table 2) of compound **17**, the activity is reduced. However, when an electron donating *tert*-butyl group (compound **18**) is attached to the same position, the activity is enhanced. The favorable substitution of *tert*-butyl indicates that an electron donating bulky group is important at the R₅ position of ring A (Table 2). As a result, when this position is unsubstituted (compound **16**), activity is reduced.

CoMSIA model

As well as steric and electrostatic fields analogous to those generated in CoMFA, CoMSIA analyses define the hydrophobic and hydrogen bond (donor and acceptor) fields. The same procedure used for the CoMFA study was also employed in CoMSIA study. The CoMSIA analyses were performed based on combination of steric, electrostatic,

hydrophobic, and hydrogen bond donor and acceptor fields. The best model was selected based on the criteria of high q^2 value, low number of components (C) and high predictive r^2 value. The CoMSIA results are summarized in Table 5 wherein only the best combination of descriptors is shown.

The CoMSIA model combining all five descriptors (S E D A H) yield a cross-validated q^2 of 0.86 with 2 components, a non-cross-validated r^2 of 0.93, an F value of 206.76, and a predictive r^2 of 0.91. The field contributions of this CoMSIA model were 29.0%, 25.9%, 26.30%, 10.8% and 8.0% for steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor field contributions, respectively.

A CoMSIA model based on steric, electrostatic and hydrophobic fields yield the highest cross-validated q^2 of 0.89 with 2 components, a non-cross-validated r^2 of 0.93, an F value of 205.21, and a predictive r^2 of 0.87. The field contributions of this CoMSIA model were 35.6 %, 32.3 %, and 32.0 % for steric, electrostatic and hydrophobic fields, respectively.

The combination of steric, electrostatic, donor and acceptor fields yield a cross-validated q^2 of 0.84 with 2 components, a non-cross-validated r^2 of 0.93, an F value of 191.77 and the highest predictive r^2 of 0.94. The field contributions for this CoMSIA model were 41.5 % (steric), 35.3 % (electrostatic), 13.7 % (donor) and 9.5% (acceptor). The models generated by other combinations of CoMSIA fields (Table 4) showed statistically significant internal and external predictions, in which the steric, electrostatic and hydrogen bond donor fields dominate the CoMSIA models. Figure 3 shows the correlation ($r^2=0.98$) of actual versus predicted activities for the CoMSIA model involving steric, electrostatic and hydrogen bond donor fields.

Figure 6 shows the steric and electrostatic maps derived from CoMSIA analysis. The contours found were similar to those for the CoMFA model. Green contours found along the rings B and C (compound **30**) suggest favorable steric regions, which was also seen in the CoMFA steric map (Fig. 4). Interestingly, the yellow sterically unfavored region (20 % contribution) in CoMSIA is centered at the

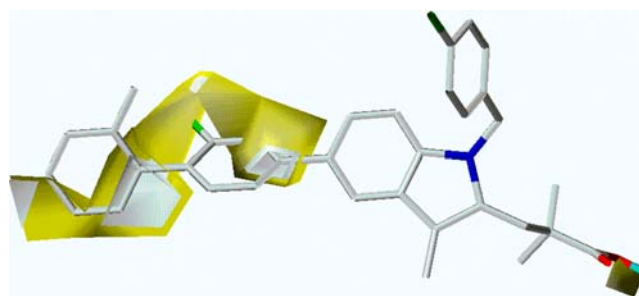


Fig. 7 CoMSIA hydrophobic favored yellow region was found in between rings B and C of compound **30**

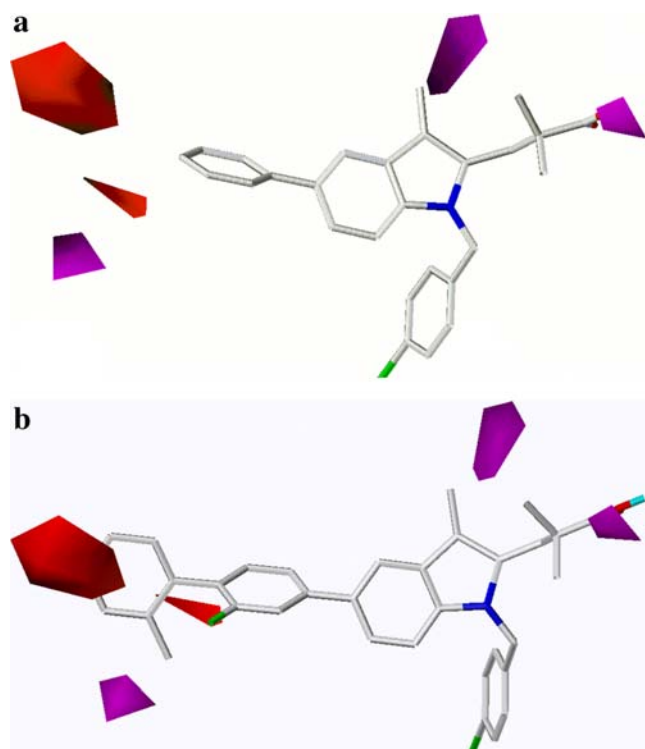


Fig. 8 CoMSIA H-bond acceptor contour maps for compound 19 (mark **a**, lower activity) as compared to compound 30 (mark **b**, highest activity). Magenta color contours indicate regions favorable with hydrogen bond acceptor substituents. Purple contours connote regions disfavored with hydrogen bond acceptor groups

C₂ and C₃ positions of indole moiety (compound 30). As a result, the replacement of a methyl group at R₃ position of compound 14 is favored, in contrast to the effect of placing a phenyl group at the R₃ position of compound 8.

Figure 7 shows the hydrophobic contour map of CoMSIA. The large yellow isopleths indicate that any hydrophobic group in the rings B and C is favorable. This reflects the fact that structural modification of compound 1 (MK-886) favored hydrophobic benzene ring at R₅ position of indole ring (compound 19) and at Y position of ring B (compound 21). A remarkable 37.5-fold increase in potency is observed after the double insertions of phenyl groups to the indole ring. The result agrees with the contour map indicated by prominent hydrophobic regions encompassing rings B and C of compound 30.

The CoMSIA analysis provide additional structural insights into the probable binding sites of the ligand-receptor complex. The hydrogen bond acceptor contour map (Fig. 8) connotes magenta contours (80% contribution) located in the proximity of rings B and C of compound 30. This result indicates that electronegative substituents are favored for the biphenyl rings, and thus a cationic subsite of the receptor can be deduced. In compound 21, after electronegative fluorine and steric phenyl ring were substituted into *meta* and *para* positions of ring B

(compound 23), there is 85.7-fold increase of activity with respect to compound 19. A 200-fold increase in potency of compound 30 is observed after the replacement (compound 19) of hydrogen by fluorine at the *meta*-position, and by conversion of hydrogen to a 2-methylphenyl group at the *para*-position. A great improvement in potency is shown when combined electronegative and steric groups are inserted into the *meta* and *para* positions of ring B. In Figs. 4 and 6, the *para*-position of the sterically favored phenyl group agrees with the importance of this group.

The CoMSIA hydrogen bond donor contour map (Fig. 9) indicates cyan contours (80% contribution) near the carboxylic moiety at C₂ position of indole ring. The cyan isopleths besides the carboxylic moiety denote the regions where proton donors are expected in the receptor and thus, a hydrogen-bond-acceptor substituent should be favored when protonation occurs in carboxylic acid. This can be explained by poor activity when the hydrogen (at R₂ position of compound 1) is replaced by methyl (compound 6), or is replaced by amino group (compound 7), that results to decrease in potency of 4.5-fold and 6.25-fold, respectively. The results show that the inactivity of ester and amide analogue favor the ligand carboxylate to interact via ionic and hydrogen bonds with the anionic binding site.

Based on the present 3D-QSAR studies, a hypothetical binding model of indole derivatives to mPGES-1 can be proposed (Fig. 10). The large hydrophobic region/binding

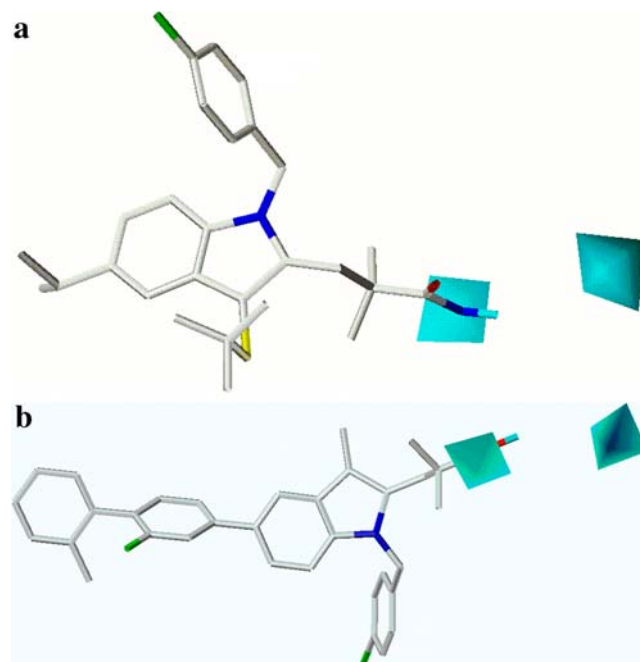


Fig. 9 CoMSIA H-bond donor contour maps for compound 7 (mark **a**, lower activity) as compared to compound 30 (mark **b**, highest activity). Cyan color contour indicates regions favorable with hydrogen bond donor substituents

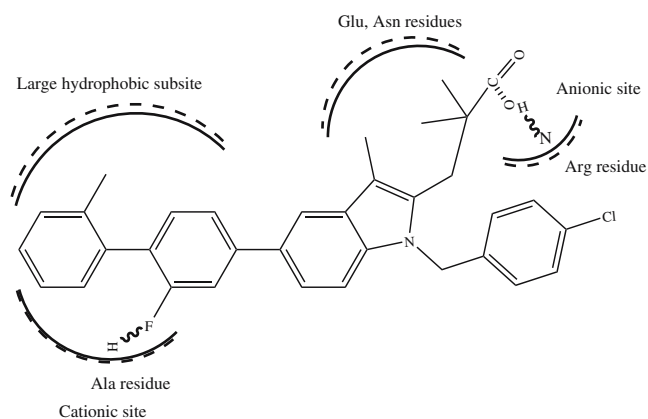


Fig. 10 The proposed model of the binding site of mPGES-1 protein based on 3D-QSAR analysis

site may play a significant role in the selectivity of ligands to mPGES-1 because a substitution at rings B and C (i.e., compounds **19–21**) drastically changed activity. The biphenyl may have π – π interactions with the receptor, wherein the planarity in the ligand structure is significant. In the absence of crystal structure of mPGES-1, the important amino acid residues in the active site were reported [27] based on alignment technique. The important residues involved in the active site can be classified into an aliphatic alanine (*Ala*), an acidic glutamic acid (*Glu*) and an asparagine (*Asn*), and a basic group such as arginine (*Arg*). Based on the study of Mancini et al. [27], the negative charged residue *Glu* and *Asn* are pivotal in the binding of ligands to mPGES-1. Furthermore, the residue *Arg* was reported [28] to have catalytic function to the receptor. In the proposed cationic site of the receptor (Fig. 10), it indicates hydrogen bond formation with the electronegative substituent of the ligand. In particular, the amino acid residue *Ala* can donate a proton to form a hydrogen bond to the electronegative substituent fluorine

located at the *meta*-position in ring B of compound **30**. For the plausible anionic site of the receptor, the hydrogen bond formation may occur between the hydrogen proton from carboxylic acid of compound **30** with the basic residue *Arg*. The secondary donor site shown by a cyan block located farther away from ligand carboxylate supports this interpretation.

Model validations

Bootstrapping

The bootstrap results for both CoMFA and CoMSIA models were obtained from 10 runs (Table 4). A bootstrapped value, r_{bs}^2 of 0.98 and 0.93, and bootstrapped standard deviation, SD_{bs} of 0.01 and 0.02, were obtained for CoMFA (steric and electrostatic) and CoMSIA (steric, electrostatic, donor and acceptor) models. The results indicate that reliable models were successfully constructed.

Predictive r^2

Model validation is the most critical and important part of the QSAR model building, where the internal predictive power of the model and its ability to reproduce biological activities of untested compounds is established. The eight molecules selected in the test set (compounds **1**, **7**, **9**, **16**, **20**, **24**, **26** and **31**) were used to verify the validity of the CoMFA and CoMSIA models. The predicted activities for both the training and test sets (shown in Table 5) showed high correlation with experimental activities, $r^2=0.92$ (Fig. 2) and $r^2=0.98$ (Fig. 3), for CoMFA and CoMSIA models, respectively. Moreover, good predictive r^2 values of 0.83 for the CoMFA model and 0.94 for the best CoMSIA model, suggest that the models were predictive. Based on the predictive r^2 validation results, the models

Table 6 Model stability test for CoMFA and CoMSIA by progressive scrambling

Component	CoMFA			CoMSIA		
	q^2	cSDEP	$dq^{2'}/dr^{2yy'}$	q^2	cSDEP	$dq^{2'}/dr^{2yy'}$
2	0.660	0.700	1.060	0.660	0.760	1.090
3	0.640	0.730	1.660	0.640	0.800	1.340
4	0.590	0.820	1.930	0.610	0.850	1.480
5	0.580	0.830	1.610	0.610	0.880	1.580

$q^2 = 1 - (sSDEP)^2$ Predictivity of the model

SDEP = standard cross-validated error taken from SAMPLS

cSDEP = Calculated cross-validated standard error as function of correlation coefficient between the true values (y) of the dependent variables and the perturbed values (y') of the dependent variables

$dq^{2'}/dr^{2yy'}$ = Slope of q^2 (cross-validated correlation coefficient from SAMPLS) with respect correlation of the original dependent variables versus the perturbed dependent variables

could be reliably used in a new potent inhibitor design for mPGES-1 drug leads.

Progressive scrambling

Progressive scrambling of the CoMFA model gave the best evaluation at two components, with $q^2=0.66$ and $dq^2/dr_{yy}^2=1.06$ (Table 6). For CoMSIA, the best evaluation is with two components, producing a q^2 of 0.66 and $dq^2/dr_{yy}^2=1.09$. The progressive scrambling evaluation demonstrates that the generated 3D-QSAR models are stable and does not depend on chance correlation.

Conclusion

As the crystallographic data for the ligand bound to mPGES-1 protein is not available, ligand-based 3D-QSAR analyses have been carried out. Using the atom-fit technique for alignment, highly predictive models were built from indole derivatives that inhibit mPGES-1. Overall, the CoMFA and CoMSIA models exhibited good statistical properties, which implies that the steric, electrostatic, hydrophobic and hydrogen bond donor fields all contribute significantly to the activity of the ligands. Bootstrapping, predictive r^2 and progressive scrambling analysis were all carried out for model validations. Moreover, the consistency of the CoMSIA and CoMFA results show the robustness of the constructed 3D-QSAR models. The structural requirements for the ligand, as well as a plausible binding mode of the mPGES-1 receptor, that were identified in the present study can be utilized in the design of novel compounds for treating pain and inflammation.

Acknowledgement The study was supported by the Korea Research Foundation.

References

- Schmidt RF, Schaible HG, Messlinger K, Heppelmann B, Hanesch U, Pawlak M (1994) Progress in pain and management. In: Gebhart GF, Hammond DL, Jensen TS (eds) Proc 7th World Congress on Pain (2nd edn.) Seattle, IASP, Chap 16
- Hoogerwerf WA, Zou L, Shenoy M, Sun D, Micci MA, Lee-Hellmich H, Xiao SY, Winston JH, Pasricha PJ (2001) *J Neurosci* 21(22):9036–9042
- Yang HYT, Iadarola MJ (2006) *Peptides* 27(5):943–952
- Wan WZ, Zhang N (2006) *Zhongguo Shengwu Huaxue Yu Fenzi Shengwu Xuebao* 22(2):101–105
- Shukla AK, Haase W, Reinhart C, Michel H (2006) *Biol Chem* 387(5):569–576
- Qin X, Zhang H (2005) *Zhongguo Linchuang Kangfu* 9(17):145–147
- Gilmour RS, Mitchell MD (2001) *Exp Biol Med* 226:1–4
- Kudo I, Murakami M (2005) *J Biochem Mol Biol* 38(6):633–638
- Murakami K, Kudo M (2004) *Prog Lipid Res* 43:3–35
- Sampey AV, Monrad S, Crofford LJ (2005) *Arthritis Res Therapy* 7(3):114–117
- Trebino CE, Stock JL, Gibbons CP, Naiman BN, Wachtmann TS, Umland JP, Pandher K, Lapointe JM, Saha S, Roach ML, Carter D, Thomas NA, Durtschi BA, McNeish JD, Hambor JE, Jakobsson P, Carty TJ, Perez JR, Audoly LP (2003) *Proc Natl Acad Sci USA* 100:9044–9049
- Duffy D, Seachord C, Dozier B (2005) *Hum Reprod* 20(6):1485–1492
- Uematsu S, Matsumoto M, Takeda K, Akira S (2002) *J Immunol* 168:5811–5816
- Thoren S, Jakobsson PJ (2000) *Eur J Biochem* 267:6428–6434
- Quraishi O, Mancini J, Riendeau D (2002) *Biochem Pharmacol* 63:1183–1189
- Riendeau D, Aspiotis R, Ethier D, Gareau Y, Grimm E, Guay J, Guiral S, Juteau H, Mancini J, Methot N, Rubin J, Friesen R (2005) *Bioorg Med Chem Lett* 15:3352–3355
- Cramer III RD, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959–5967
- Klebe G, Abraham U, Mietzner T (1994) *J of Med Chem* 37:4130–4146
- SYBYL 7.2 (2006) Tripos Inc., St. Louis, MO 63144
- Clark M, Cramer III R, Van Opdenbosch N (1989) *J Comput Chem* 10:982–1012
- Kim K, Greco G, Novellino E (1998) *Perspect Drug Discov Des* 12–14:257–315
- Wold S, Albano C, Dunn W, Edlund U, Esbensen K, Geladi P, Hellberg S, Johansson E, Lindberg W, Sjostrom M (1987) *Chemometrics: mathematics and statistics in chemistry*. In: Kowalski B (ed) Reidel, Dordrecht, The Netherlands pp 17–95
- Bush BL, Nachbar RB (1993) *J Comput Aided Mol Des* 7:587–619
- Stahle L, Wold S (1988) *Prog Med Chem* 25:292–338
- Cramer III RD, Bunce JD, Patterson DE, Frank IE (1988) *Quant Struc-Act Relat* 7:18–25
- Clark RD, Sprous DG, Leonard JM (2001) Validating models based on large data sets. In: Höltje HD, Sippl W (eds) 13th european symposium on quantitative structure-activity relationships: rational approaches to drug design. Prous Science, Barcelona-Philadelphia pp 475–485
- Mancini JA, Blood K, Guay J, Gordon R, Claveau D, Chan C, Riendeau D (2001) *J Biol Chem* 276(6):4469–4475
- Jakobsson P, Morgenstern R, Mancini J, Hutchinson AF, Persson B (1999) *Prot Sci* 8:689–692