

## QM and Pharmacophore based 3D-QSAR of MK886 Analogues against mPGES-1

F. A. Pasha, M. Muddassar, Hwanwon Jung, Beom-Seok Yang, Cheolju Lee, Jung Soo Oh, Seung Joo Cho,\* and Hoon Cho†\*

Computational Science Center, Future Fusion Technology Division, Korea Institute of Science and Technology, P.O. Box 131, Seoul 130-650, Korea. \*E-mail: chosj@kist.re.kr

†College of Engineering, Chosun University, Gwangju 501-579, Korea. \*E-mail: hcho@chosun.ac.kr

Received October 22, 2007

Microsomal prostaglandin E<sub>2</sub> synthase (mPGES-1) is a potent target for pain and inflammation. Various QSAR (quantitative structure activity relationship) analyses used to understand the factors affecting inhibitory potency for a series of MK886 analogues. We derived four QSAR models utilizing various quantum mechanical (QM) descriptors. These QM models indicate that steric, electrostatic and hydrophobic interaction can be important factors. Common pharmacophore hypotheses (CPHs) also have studied. The QSAR model derived by best-fitted CPHs considering hydrophobic, negative group and ring effect gave a reasonable result ( $q^2 = 0.77$ ,  $r^2 = 0.97$  and  $R_{\text{testset}} = 0.90$ ). The pharmacophore-derived molecular alignment subsequently used for 3D-QSAR. The CoMFA (Comparative Molecular Field Analysis) and CoMSIA (Comparative Molecular Similarity Indices Analysis) techniques employed on same series of mPGES-1 inhibitors which gives a statistically reasonable result (CoMFA:  $q^2 = 0.90$ ,  $r^2 = 0.99$ . CoMSIA:  $q^2 = 0.93$ ,  $r^2 = 1.00$ ). All modeling results (QM-based QSAR, pharmacophore modeling and 3D-QSAR) imply steric, electrostatic and hydrophobic contribution to the inhibitory activity. CoMFA and CoMSIA models suggest the introduction of bulky group around ring B may enhance the inhibitory activity.

**Key Words :** 3D-QSAR, Drug design, CoMFA, CoMSIA, mPGES-1

### Introduction

Prostaglandin (PG),<sup>1</sup> are important mediators of various physiological processes such as regulation of gastrointestinal, renal and blood homeostasis. On the other hand, they also act as potent mediator of inflammation and fever. They acts as a biological mediator to produce signals in the human body that in turn induce pain and inflammation. There are three forms of prostaglandin E synthase (PGES), namely, microsomal prostaglandin E<sub>2</sub> synthase-1 (mPGES-1), microsomal prostaglandin E<sub>2</sub> synthase-2 (mPGES-2) and cytosolic PGES. The pathway linkage preference of mPGES-1, mPGES-2 and cPGES is, both COX-1 and COX-2 respectively.<sup>2</sup> The mPGES-1 is an important enzyme because it catalyzes the conversion of prostaglandin endoperoxide (PG) H<sub>2</sub> to PGE<sub>2</sub>. PGE<sub>2</sub> in turn controls biological activities such as relaxation and contraction of muscles. There are several reported compounds which act as inhibitors of mPGES-1.<sup>3,4</sup> Recently, a series of MK886 compounds also showed selectivity and higher activity against the inducible mPGES-1 with the lowest IC<sub>50</sub> value found being 3 nM.<sup>5</sup> The pharmacophore based QSAR,<sup>6-9</sup> QM based QSAR<sup>10,11</sup> and 3D QSAR<sup>12-17</sup> shown good predictivity for other datasets. The current study deals the molecular modeling of MK886 analogues with mPGES-1 to access further possibility of improved ligands. Specifically, QM-based QSAR, pharmacophore-based QSAR and 3D-QSAR (CoMFA and CoMSIA)<sup>18-20</sup> have been performed to study MK886 series.

### Material and Methods

**Data sets.** Thirty-two MK886 derivatives<sup>5</sup> were taken

from literature with their biological activities in terms of IC<sub>50</sub> values. The IC<sub>50</sub> values, *i.e.*, the concentration ( $\mu\text{M}$ ) of inhibitor that produces 50% inhibition of mPGES-1 were converted into pIC<sub>50</sub> ( $-\log\text{IC}_{50}$ ) as reported in Table 1.

**Quantum mechanical QSAR.** The quantum mechanical descriptors like chemical potential ( $\mu$ ),<sup>21</sup> electrophilicity index ( $\omega$ ),<sup>9,22</sup> electrophilic frontier densities,<sup>23</sup> molar refractivity (MR)<sup>24,25</sup> and solvent assessable surface area (SASA)<sup>26</sup> have been considered. Recently Parr *et al.* define the electronegativity<sup>21</sup> and chemical potential as equation (1).

$$\chi = -\mu = -(\partial E / \partial N)_{V(r)} \quad (1)$$

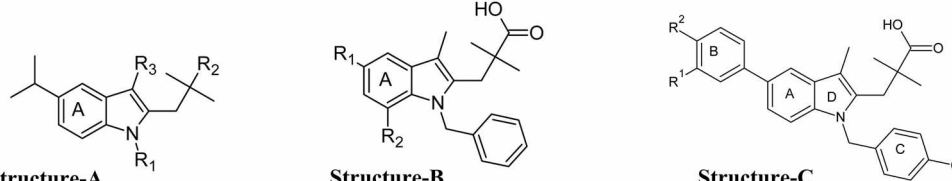
Where  $E$  is total energy,  $N$  is number of electrons of the chemical species and  $V(r)$  is external potential. According to the Koopman's theorem<sup>27</sup> the operational definitions of electro negativity or chemical potential may be given as equation (2)

$$\chi = -\mu = \frac{1}{2}(\epsilon_{\text{LUMO}} + \epsilon_{\text{HOMO}}) \quad (2)$$

Fukui *et al.* proposed the frontier electron density<sup>23</sup> for the electrophilic attack at the  $r^{\text{th}}$  atom in molecule  $f_r^{(E)}$  which may be written as equation (3)

$$f_r^{(E)} = 2 \frac{|C_r^{(1)}|^2 + |C_r^{(2)}|^2 e^{-D\Delta\lambda}}{1 + e^{-D\Delta\lambda}} \quad (3)$$

Where  $\Delta\lambda$  is energy difference between the two orbital in units of  $\beta$ .  $C_r^{(1)}$  and  $C_r^{(2)}$  are the coefficient of LCAO MO at  $r^{\text{th}}$  atom corresponding to the highest and next orbital respectively.  $D$  is constant which determines the degree of contribution of lower MO to the frontier electron density.

**Table 1.** Indole derivatives from MK886 compound as inhibitors of mPGES-1


No.	Structure	R1	R2	R3	IC50	pIC50
1	A	CH2(4-Cl-Ph)	COOH	S-tertBu	1.6	-0.204
2*	A	H	COOH	S-tertBu	10	-1.041
3*	A	Me	COOH	S-tertBu	10	-1.041
4*	A	CH2(CH=CH2)	COOH	S-tertBu	6.7	-0.826
5	A	(CH2)3Ph	COOH	S-tertBu	3.2	-0.50
6	A	CH2(4-Cl-Ph)	COOMe	S-tertBu	7.2	-0.857
7	A	CH2(4-Cl-Ph)	CONH2	S-tertBu	10	-1.041
8	A	CH2(4-Cl-Ph)	COOH	Ph	6.4	-0.806
9*	A	CH2(4-Cl-Ph)	COOH	OPh	0.65	0.187
10*	A	CH2(4-Cl-Ph)	COOH	CH2(4-tertBu-Ph)	0.29	0.538
11	A	CH2(4-Cl-Ph)	COOH	CO(2-Me-Ph)	0.9	0.046
12	A	CH2(4-Cl-Ph)	COOH	COCH2S-tertBu	0.26	0.585
13	A	CH2(4-Cl-Ph)	COOH	COCH2-tertBu	0.25	0.602
14	A	CH2(4-Cl-Ph)	COOH	Me	1.1	-0.041
15	B	H	iso-propyl	-	4.3	-0.633
16	B	H	H	-	3.2	-0.505
17	B	F	H	-	2.6	-0.415
18	B	tert-butyl	H	-	0.33	-0.481
19*	B	Ph	H	-	0.6	0.222
20	C	Ph	H	-	0.16	0.796
21*	C	H	Ph	-	0.016	1.796
22	C	Cl	Ph	-	0.022	1.658
23*	C	F	Ph	-	0.007	2.155
24	C	F	1,3-pyrazinyl	-	0.032	1.495
25*	C	F	3-pyridinyl	-	0.012	1.921
26	C	F	2-MeO-Ph	-	0.005	2.301
27	C	F	2-Cl-Ph	-	0.004	2.398
28	C	F	2-F-Ph	-	0.008	2.097
29*	C	F	2-MeCO-Ph	-	0.006	2.222
30	C	F	2-Me-Ph	-	0.003	2.523
31	C	F	3-Me-Ph	-	0.033	1.481
32	C	F	4-Me-Ph	-	0.031	1.509

<sup>a</sup>IC50 = 50% inhibition ( $\mu\text{M}$ ) to mPGES-1 enzyme. pIC50 =  $-\log \text{IC}_{50}$ . \* = included in the test set of compounds

The molar refractivity is a constitutive-additive property that is calculated by the Lorenz-Lorentz formula<sup>24,25</sup> as given in equation (4)

$$MR = \frac{(n^2 - 1) \times MW}{(n^2 + 2) \times d} \quad (4)$$

Where  $MW$  is the molecular weight,  $n$  is the refraction index and  $r$  the density, and its value depends only of the wave longitude of the light used to measure the refraction index.

**Solvent assessable surface area (SASA).** The molecular surface is defined in COSMO<sup>26</sup> as the sum of overlapping van der Waals radii,  $R_{\alpha}^{vdw}$ , about each atom,  $\alpha$ . The solvent approximated as a sphere of radius  $R^{sol}$ . The surface

available to the solvent's centers is therefore given as the surface defined by the sum of overlapping radii,  $R_{\alpha}$ , where

$$R_{\alpha} = R_{\alpha}^{vdw} + R^{sol} \quad (5)$$

The effective charges, which are responsible for the dielectric screening, will not be located at the centers of solvent molecules but instead located at distance,  $\delta^{SC}$  from the molecular center. The solvent accessible surface is then defined by the sum of overlapping radii,  $R_{\alpha}^*$ , where

$$R_{\alpha}^* = R_{\alpha} - \delta^{SC} \quad (6)$$

The solvent assessable surface area calculated by using AM1 'COnductor-like Screening MOdel' (COSMO). Molecular geometry optimization carried out by AM1 semi

empirical method in conjunction with molecular mechanics using CAChe pro software.

**Multiple Linear Regression Analysis (MLR).** MLR analyses performed using SPSS software. The quantum mechanical descriptors used as independent variables and the pIC<sub>50</sub> values as the dependent variable. In the statistical analyses, the systematic search performed to determine the significant descriptors. In order to minimize the effect of collinearity and to avoid redundancy correlation matrix developed with a cutoff value of 0.6 and the variables physically removed from the analysis which shows exact linear dependencies between subsets of the variables and multi-collinearity (high multiple correlations between subsets of the variables). In order to explore the reliability of the proposed model we used the crossvalidation method. Prediction error sum of squares (*PRESS*) is a standard index to measure the accuracy of a modeling method based on the crossvalidation technique. The  $r_{cv}^2$  calculated by using equation-7 based on the *PRESS* and *SSY* (Sum of squares of deviations of the experimental values from their mean).

$$r_{cv}^2 = 1 - \frac{PRESS}{SSY} = 1 - \frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{\sum_{i=1}^n (y_{exp} - \bar{y})^2} \quad (7)$$

### Pharmacophore Based QSAR

**Generation of the Common Pharmacophore Hypothesis (CPH).** The common pharmacophore hypotheses were generated using PHASE.<sup>28</sup> Conformers were generated by MCOMM/LOD with OPLS-2005 force field. A set of conformers for each molecule with maximum energy difference of 10 kcal/mol relative to global energy minima were retained. Pharmacophore features; hydrogen bond acceptor(A), hydrogen bond donor (D), hydrophobic group (H), negatively charged group (N), positively charged group(P), and aromatic ring (R) were defined by a set of chemical structure patterns as SMARTS queries and assigned one of three possible geometries, which define physical characteristics of the site:

Point – the site is located on a single atom in the SMARTS query.

Vector – the site is located on a single atom in the SMARTS query, and assigned directionality according to one or more vectors originating from the atom.

Group – the site is located at the centroid of a group of atoms in the SMARTS query. For aromatic rings, the site includes directionality, defined by a vector that is normal to the plane of the ring.

The final size of pharmacophore box was 1 Å, which governs the tolerance on matching; the smaller the box size, the more closely pharmacophores must match. Any single pharmacophore in the group could ultimately become a CPH. The analyses indicate that maximum three sites can match only up to 30 molecules out of 32. These CPHs examined using a scoring function to yield the best alignment

of the active ligands and quality of alignment measured by a survival score, which defined as:

$$S = W_{site}S_{site} + W_{vec}S_{vec} + W_{vol}S_{vol} + W_{sel}S_{sel} + W_{rew}^m \quad (8)$$

Where  $W_{Ds}$  are weights and  $S_{Ds}$  are scores,  $S_{site}$  represents an alignment score, the root mean square deviation at the site point position.  $S_{vec}$  represents vector score, and averages cosine of the angles formed by corresponding pairs of vector features in aligned structures.  $S_{vol}$  represents volume score based on overlap of van der Waals models of non-hydrogen atoms in each pair of structures.  $S_{sel}$  represents the selectivity score, and accounts for what fractions of molecules are likely to match the hypothesis regardless of their activity toward a receptor. Weights are user adjustable.  $W_{site}$ ,  $W_{vec}$ ,  $W_{vol}$ , and  $W_{rew}$  have a default value of 1.0 while  $W_{sel}$  has a default value of 0.0, so that a useful hypothesis are not missed out.  $W_{rew}^m$  represents the reward weights, where  $m$  is the number of actives that match the hypothesis minus one. In the hypothesis generation, all default values used.

**Assessment of significant CPH using Partial Least Square Analysis (PLS).** The evaluation of generated CPHs performed by correlating the observed and estimated activities of training and test sets of 20 and 10 molecules respectively. The PLS analyses carried out using PHASE with maximum of N/3 PLS factors. N/3 number of ligands in training set, and either atom, or pharmacophore-based model using grid spacing of 1 Å. CPHs of best predictivity and significant statistics were selected for molecular alignments and QSAR model. The same alignment used for further 3D-QSAR (CoMFA and CoMSIA).

**3D-QSAR (CoMFA and CoMSIA).** In standard CoMFA and CoMSIA procedure, a suitable conformation<sup>29</sup> is desired for superimposing the ligands which is assumed to be bioactive. The alignment based on CPH with significant statistical data was imported in to SYBYL 7.3<sup>30</sup> running on linux cluster and directly used for 3D-QSAR. Lennard-Jones and Coulomb potentials based CoMFA has been performed and the steric as well as electrostatic energies were calculated by using sp<sup>3</sup> carbon probe atom with Van der Waals radius of 1.52 Å and (+1) charge. The energies truncated to ±30 kcal mol<sup>-1</sup> and the electrostatic contributions ignored at lattice interactions with maximum steric interactions. The CoMFA generated by standard method in SYBYL. The CoMSIA models derived with the same lattice box as in CoMFA. All five CoMSIA similarity index (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) 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 distribution function.<sup>18</sup> The default value (0.3) of attenuation factor was used, which is the standard distance dependence of molecular similarity. The effect of using the standard attenuation factor displayed in contour maps with prominent molecular features.

**Partial Least Square (PLS) analysis and validation of QSAR models:** To derive 3D-QSAR models, the CoMFA and CoMSIA descriptors used as independent variables and

the  $\text{pIC}_{50}$  as the dependent variable. PLS analysis<sup>31,32</sup> used to correlate these CoMFA and CoMSIA descriptors as a function of variation of inhibitory activity values. The CoMFA cutoff values were set to 30 kcal mol<sup>-1</sup> for both steric and electrostatic fields, and all fields scaled by the default options in SYBYL. The crossvalidation analysis performed by using the leave one out (LOO) method in which one compound removed from the data set and its activity predicted using the model derived from the rest of the data points. The cross-validated correlation coefficient ( $q^2$ ) that resulted in optimum number of components and lowest standard error of prediction were considered for further analysis and calculated using following equations (9)-(10)

$$q^2 = 1 - \frac{\sum_y (y_{\text{pred}} - y_{\text{observed}})^2}{\sum_y (y_{\text{observed}} - y_{\text{mean}})^2} \quad (9)$$

$$\text{PRESS} = \sum_y (y_{\text{predicted}} - y_{\text{observed}})^2 \quad (10)$$

Where,  $y_{\text{pred}}$ ,  $y_{\text{actual}}$  and  $y_{\text{mean}}$  are predicted, actual, mean values of the target property ( $\text{pIC}_{50}$ ), respectively, and PRESS is the sum of predictive sum of squares. The non-crossvalidated PLS analyses were performed with 2.0 column filter, to reduce computation time with small effect on the  $q^2$  values. To assess the robustness and statistical confidence of the derived models, bootstrapping analysis for 10 runs performed. To assess the predictive power of the 3D-QSAR models derived using the training set, biological activities of an external test set of twelve molecules predicted. The predictive ability of the models is expressed by the  $r^2_{\text{predictive}}$  value, which is analogous to cross-validated  $r^2$  ( $q^2$ ) and is calculated using the formula-11

$$r^2_{\text{pred}} = \frac{SD - \text{PRESS}}{SD} \quad (11)$$

Where  $SD$  is the sum of the squared deviations between the biological activities of the test set and mean activities of the training molecules and PRESS is the sum of squared deviation between predicted and actual activities of the test set molecules.

## Results

**QM-based QSAR.** The necessary input values extracted from MOPAC calculation result. MLRA (multiple linear regression analysis) employed to correlate the variation of activity with the values of chemical potential, electrophilic frontier density, molar refractivity and solvent accessible surface area.

In initial step of regression, no significant model obtained but after careful data mining, based on number of rings one-indicator parameter "I" introduced. All molecules having four rings they have been allotted  $I = 1$  while for rest all molecules  $I = 0$ . This indicator parameter significantly contributes to every models, clearly indicates the pharmacophore feature of fourth ring. In general, the aromatic rings are responsible for hydrophobicity so the presence of

indicator parameter in every model indicates the probability of hydrophobic interaction. A significant model PA1 was reported with better statistics ( $r^2_{\text{CV}} = 0.73$ ,  $r^2 = 0.79$ ) which involve molar refractivity with a coefficient 0.004 and chemical potential with a coefficient 0.82. The major contribution of MR and small contribution of chemical potential indicates that, there is a contribution of steric and electrostatic field effect to activity. Based on this model the predicted activities of training and test sets are reported in Table 2.

$$\begin{aligned} \text{PA1} &= 0.004 \times \text{MR} + 0.82 \times \mu + 1.83 \times \text{I} - 4.22 \\ \text{N} &= 20, r^2_{\text{CV}} = 0.73, r^2 = 0.79, \text{SEE} = 0.52, \text{F} = 23.34, \\ \text{Pearson } R_{\text{testset}} &= 0.89 \end{aligned} \quad (12)$$

The model was validated against test set of 10 molecules (Pearson  $R_{\text{testset}} = 0.89$ ). Similarly another significant model ( $r^2_{\text{CV}} = 0.75$ ,  $r^2 = 0.80$ ) obtained by solvent assessable surface area (SASA) with a coefficient 0.005 and chemical potential with a coefficient 0.767. The SASA is also steric and hydrophobic parameter while  $\mu$  is electrostatic parameter, which is in consonance with model PA1 and gives emphasis to contribution of steric bulk to activity.

$$\begin{aligned} \text{PA2} &= 0.005 \times \text{SASA} + 0.767 \times \mu + 1.759 \times \text{I} - 4.360 \\ \text{N} &= 20, r^2_{\text{CV}} = 0.75, r^2 = 0.80, \text{SEE} = 0.50, \text{F} = 2501, \\ \text{Pearson } R_{\text{testset}} &= 0.91 \end{aligned} \quad (13)$$

Based on this model the regression equation (13) developed and the predicted activities of training and test sets are reported in Table 2. This model was also validated against test set of 10 molecules (Pearson  $R_{\text{testset}} = 0.91$ ). In order to have further insight for electrostatic interaction a quantum mechanical atomic level calculation was performed and the electrophilic frontier density (EFD) at every atom of each molecules were calculated, the highest EFD of every molecule has been used as electrostatic descriptor. The model PA3 derived by MR and EFD with indicator parameter "I" which gives better statistics ( $r^2_{\text{CV}} = 0.79$ ,  $r^2 = 0.85$ ) than corresponding model PA1. The regression equation (14) has been derived and the predicted activities of training and test sets are reported in Table 2.

$$\begin{aligned} \text{PA3} &= 0.013 \times \text{MR} - 1.66 \times \text{EFD} + 1.395 \times \text{I} - 1.034 \\ \text{N} &= 20, r^2_{\text{CV}} = 0.79, r^2 = 0.85, \text{SEE} = 0.50, \text{F} = 25.63, \\ \text{Pearson } R_{\text{testset}} &= 0.94 \end{aligned} \quad (14)$$

The routine model validation carried out using test set of 10 molecules (Pearson  $R_{\text{testset}} = 0.94$ ). Similarly, the contribution of EFD tested in conjunction with SASA and the regression model PA-4 was derived. This model gives better statistics ( $r^2_{\text{CV}} = 0.79$ ,  $r^2 = 0.85$ ) than corresponding model PA-2 and the regression equation (15) has been developed. The predicted activities of training and test sets by model PA-4 are reported in Table 2.

$$\begin{aligned} \text{PA4} &= 0.013 \times \text{SAS} - 1.758 \times \text{EFD} + 1.187 \times \text{I} - 1.795 \\ \text{N} &= 20, r^2_{\text{CV}} = 0.79, r^2 = 0.85, \text{SEE} = 0.38, \text{F} = 28.35, \\ \text{Pearson } R_{\text{testset}} &= 0.95 \end{aligned} \quad (15)$$

Like before this model was also validated against test set of 10 molecules (Pearson  $R_{\text{testset}} = 0.95$ ).

The linear dependency of Observed  $\text{pIC}_{50}$  jointly on steric,

**Table 2.** The Observed and predicted pIC<sub>50</sub> values of mPEGS-1 inhibitors by quantum mechanical descriptors

No.	pIC <sub>50</sub>	PA1	PA2	PA3	PA4
1	-0.2	-0.334	-0.315	-0.706	-0.717
5	-0.51	-0.273	-0.258	-0.659	-0.681
6 <sup>a</sup>	-0.86	-0.252	-0.225	-0.543	-0.522
7 <sup>a</sup>	-1.04	-0.143	-0.111	-0.744	-0.701
8	-0.81	-0.246	-0.232	0.027	0.066
11	0.05	-0.018	-0.007	0.261	0.315
12	0.59	-0.064	-0.054	0.437	0.486
13	0.6	-0.085	-0.079	0.152	0.193
14	-0.04	-0.387	-0.401	-0.261	-0.283
15	-0.63	-0.372	-0.426	-0.211	-0.337
16	-0.51	-0.437	-0.475	-0.472	-0.538
17	-0.42	-0.295	-0.328	-0.411	-0.433
18	0.48	-0.393	-0.388	-0.17	-0.149
20	0.8	1.652	1.634	1.765	1.695
22	1.66	1.783	1.785	1.815	1.813
24	1.5	2.016	1.993	1.677	1.662
26	2.3	1.709	1.733	1.875	1.916
27	2.4	1.823	1.846	1.807	1.869
28	2.1	1.842	1.841	1.764	1.774
30	2.52	1.801	1.798	1.847	1.833
31	1.48	1.828	1.823	1.85	1.837
32	1.51	1.815	1.816	1.87	1.872
<b>Test Set</b>					
2	-1.04	-0.41	-0.44	-1.24	-1.35
3	-1.04	-0.40	-0.45	-1.16	-1.30
4	-0.83	-0.45	-0.49	-0.80	-0.96
9	0.19	-0.19	-0.16	0.10	0.20
10	0.54	-0.11	-0.04	0.37	0.53
19	0.22	1.51	1.44	1.41	1.24
21	1.8	1.70	1.68	1.80	1.73
23	2.16	1.81	1.79	1.78	1.74
25	1.92	1.90	1.88	1.72	1.70
29	2.22	1.85	1.87	1.84	1.88

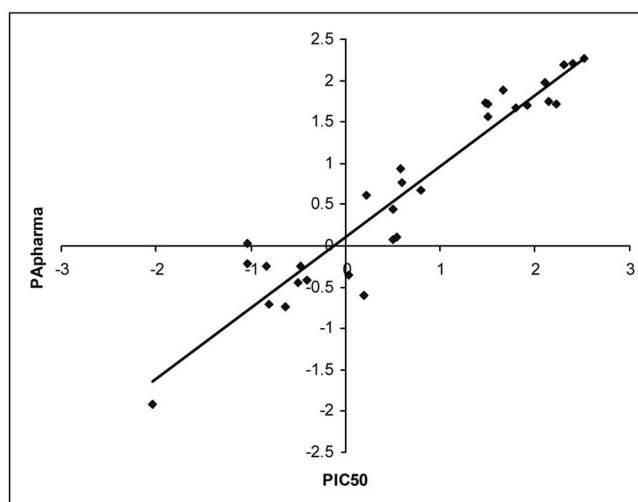
<sup>a</sup>datapoints not used inequations

electrostatic and hydrophobic parameters indicates that ligands have binding affinity with receptor due to steric, hydrophobic and electrostatic interaction. To gain further insight we have developed pharmacophore based QSAR models.

**Pharmacophore.** The same training and test sets used to develop the pharmacophore based QSAR models by using tree based partition algorithms. No CPHs obtained common in all 32 molecules but after elimination of compound 6 and 7 a number of CPHs were reported using sites: hydrophobic (H), negative (N) and ring (R). Maximum three features were allowed to develop hypothesis and there were three hypotheses based on NRR, 17 hypotheses based on HRR, 97 hypotheses based on HNR and 23 hypotheses based on HHR. On applying the scoring function for 3 features CPHs the 28 alignment rules were identified. The training set molecules were aligned by these different rules and the partial least square (PLS) analysis employed to correlates the

**Table 3.** The Statistical summary of Pharmacophore based Models

No.	Model	Factors	q <sup>2</sup>	SD	r <sup>2</sup>	F	RMSE	Pearson R <sub>testset</sub>
A1	HNR	1	0.8	0.6	0.78	64.2	0.56	0.9
A2	HNR	2	0.59	0.4	0.91	83.1	0.81	0.79
A3	HNR	3	0.77	0.23	0.97	173.2	0.6	0.9
B1	HNR	1	0.74	0.6	0.78	63.7	0.64	0.87
B2	HNR	2	0.51	0.27	0.96	188.3	0.88	0.72
B3	HNR	3	0.62	0.21	0.97	202.5	0.78	0.79
C1	HNR	1	0.8	0.58	0.79	68.9	0.56	0.9
C2	HNR	2	0.72	0.4	0.91	83	0.66	0.85
C3	HNR	3	0.78	0.25	0.97	157.3	0.59	0.89
D1	HNR	1	0.68	0.78	0.63	30.2	0.71	0.83
D2	HNR	2	0.82	0.38	0.92	92.3	0.53	0.91
D3	HNR	3	0.79	0.27	0.96	132.3	0.58	0.89

**Figure 1.** Trend of observed and predicted activity by pharmacophore based model.

biological activities with pharmacophore scores. In PLS analysis, 3 factors used with a grid spacing 1 Å. The three different regression models for each alignment were derived and the regression summary of top four hypotheses (A1-A3, B1-B3, C1-C3 and D1-D3) is reported in Table 3. The top models were selected by values of survival score of hypotheses A (survival score = 7.3), B (survival score = 7.27), C (survival score = 7.23) and D (survival score = 7.19). All four rules of alignment were based on pharmacophore combination (HNR). Model A3 is statistically (q<sup>2</sup> = 0.77, r<sup>2</sup> = 0.97, F = 173.2, R<sub>testset</sub> = 0.9) best fitted and consequently used for prediction of activities of training and test sets of molecules as reported in Table 4. Figure 1 shows the trend of observed and predicted activities.

Figure 2 demonstrates the selected pharmacophore on template molecule-30. The green ball showing the H (hydrophobic) pharmacophore while red ball showing N (negative) site. The brown ring demonstrates the R (ring) pharmacophore. The H and R are located in same zone, which is quite reasonable. The hydrophobicity not related to parent skeleton but with fourth ring, which is in accordance

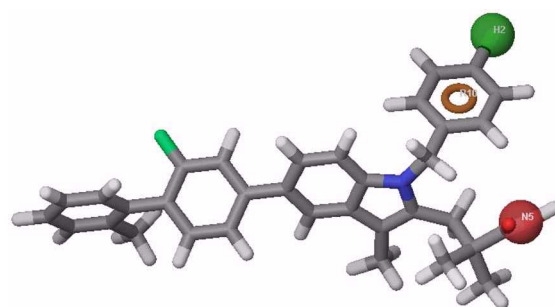
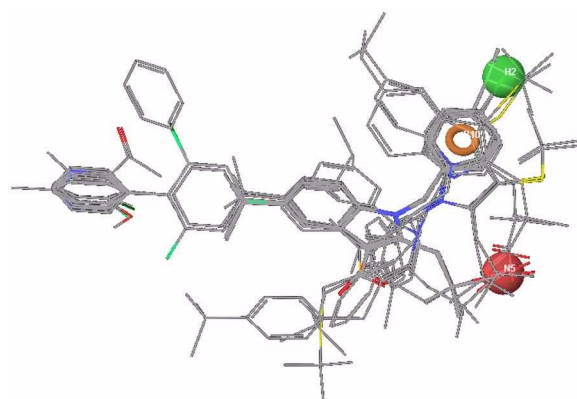
**Table 4.** The Observed and predicted activities by Pharmacophore based Model

No.	pIC <sub>50</sub>	PA <sub>Pharma</sub>	Resid.
1	-2.04	-1.92	-0.12
5	0.5	0.44	0.06
8	-0.81	-0.7	-0.11
11	0.5	0.07	0.43
12	0.58	0.94	-0.36
13	0.6	0.76	-0.16
14	0.04	-0.36	0.4
15	-0.633	-0.74	0.107
16	-0.5	-0.44	-0.06
17	-0.41	-0.42	0.01
18	-0.48	-0.25	-0.23
20	0.8	0.67	0.13
22	1.67	1.89	-0.22
24	1.5	1.56	-0.06
26	2.3	2.2	0.1
27	2.4	2.21	0.19
28	2.1	1.98	0.12
30	2.52	2.27	0.25
31	1.48	1.73	-0.25
32	1.51	1.72	-0.21
<b>Test Set</b>			
2	-1.04	0.03	-1.07
3	-1.04	-0.21	-0.83
4	-0.83	-0.24	-0.59
9	0.19	-0.6	0.79
10	0.54	0.1	0.44
19	0.22	0.61	-0.39
21	1.8	1.67	0.13
23	2.15	1.75	0.4
25	1.92	1.71	0.21
29	2.22	1.72	0.5

compound 6 and 7 omitted from analysis due to lack of CPHs

to the assumption made for indicator parameters in QM based QSAR.

The molecules were aligned HNR based CPHs and shown in Figure 3. Same alignment subsequently used for predic-

**Figure 2.** The perception of pharmacophore on template molecule-30.**Figure 3.** Pharmacophore based molecular alignment.

tion of activities by pharmacophore based QSAR and 3D-QSAR.

**CoMFA model.** CoMFA developed by using statistically significant CPH-based alignment. The aligned molecules imported in Sybyl and charges were assigned with the Gasteiger-Hückel method. Previously defined training and test sets were used to correlate the biological activities against CoMFA fields. Three different CoMFA models were derived using steric "S", electrostatic "E" and jointly both fields. The regression summary reported in Table 5. The model based on both steric and electrostatic field is most successful as clear from statistics, the crossvalidated leave one out  $q^2 = 0.90$  with 6 components, non-crossvalidated  $r^2$

**Table 5.** Regression summary of CoMFA and CoMSIA

Field	n	$q^2$	$r^2$	SE	F	$r^2_{hs}$	SD	$r^2_{predictive}$
<b>CoMFA</b>								
S	4	0.89	0.99	0.13	387.4	-	-	-
E	3	0.46	0.79	0.57	20.35	-	-	-
0.71S+0.28E	6	0.902	0.99	0.06	1083.59	0.99	0.001	0.8
<b>CoMSIA</b>								
S	3	0.87	0.93	0.33	70.88	-	-	-
E	3	0.64	0.83	0.51	25.71	-	-	-
H	7	0.82	0.99	0.12	222.83	-	-	-
D	1	-0.59	-	-	-	-	-	-
A	2	-0.1	-	-	-	-	-	-
0.6S+0.4E	6	0.9	0.99	0.11	311.51	-	-	-
0.33S+0.30E+0.37H	10	0.93	1	0.01	6565.86	1	0	0.8

**Table 6.** The Observed and predicted activities by CoMFA and CoMSIA based Models

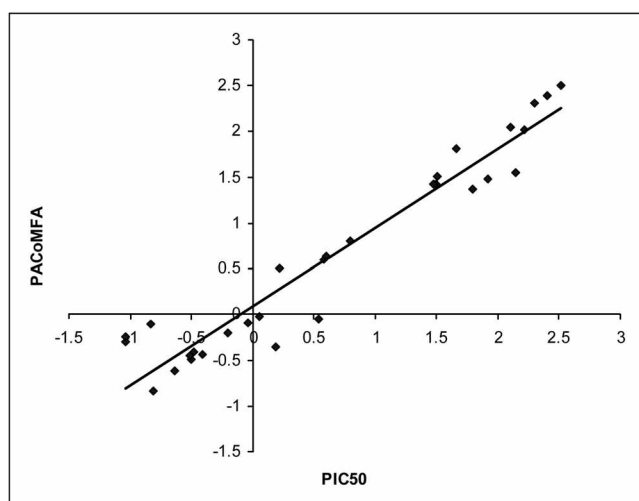
No.	pIC50	PA <sub>CoMFA</sub>	Resid.	PA <sub>CoMSIA</sub>	Resid.
1	-0.2	-0.208	0.008	-0.183	-0.017
5	-0.5	-0.498	-0.002	-0.502	0.002
8	-0.81	-0.835	0.025	-0.799	-0.011
11	0.05	-0.025	0.075	0.038	0.012
12	0.58	0.604	-0.024	0.585	-0.005
13	0.6	0.639	-0.039	0.601	-0.001
14	-0.04	-0.089	0.049	-0.048	0.008
15	-0.63	-0.614	-0.016	-0.617	-0.013
16	-0.51	-0.451	-0.059	-0.504	-0.006
17	-0.41	-0.437	0.027	-0.431	0.021
18	-0.48	-0.416	-0.064	-0.493	0.013
20	0.8	0.806	-0.006	0.806	-0.006
22	1.66	1.806	-0.146	1.666	-0.006
24	1.5	1.425	0.075	1.49	0.01
26	2.3	2.312	-0.012	2.329	-0.029
27	2.4	2.392	0.008	2.414	-0.014
28	2.1	2.048	0.052	2.074	0.026
30	2.52	2.509	0.011	2.499	0.021
31	1.48	1.431	0.049	1.476	0.004
32	1.51	1.507	0.003	1.504	0.006
<b>Test Set</b>					
2	-1.04	-0.294	-0.746	-0.378	-0.662
3	-1.04	-0.247	-0.793	-0.524	-0.516
4	-0.83	-0.108	-0.722	-0.352	-0.478
9	0.19	-0.357	0.547	-0.478	0.668
10	0.54	-0.048	0.588	-0.507	1.047
19	0.22	0.498	-0.278	0.402	-0.182
21	1.8	1.365	0.435	1.227	0.573
23	2.15	1.544	0.606	1.599	0.551
25	1.92	1.483	0.437	1.707	0.213
29	2.22	2.024	0.196	2.116	0.104

compound 6 and 7 omitted from analysis due to lack of CPHs

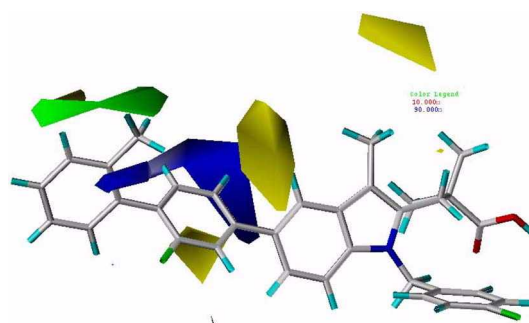
= 0.99 with standard error = 0.06 and F value = 1083.59. The success of model was tested for internal predictivity ( $r^2_{\text{boot strapping}} = 0.99$ , SD = 0.001) and external predictivity ( $r^2_{\text{predictive}} = 0.80$ ) for test set of 10 molecules. Based on this model the predicted activities are presented in Table 6. The trend of observed and predicted activities of training and test sets are shown in Figure 4 with a value of  $R_{\text{testset}} = 0.92$ .

The Figure 5 shows CoMFA contour map based on CoMFA model-3 with ligand 30. The model holds green and blue contour around ring B indicates that a bulky and positive group around this ring might have good effect over activity. The ring A and pentagonal ring holds a yellow contour. indicates the favorable sites for small groups. The data also supports this observation and in case of compound 20-32 as bulk increases at ring B the activity also increases while in case of compound 1-11 as bulk increases at pentagonal ring the activity decreases.

**CoMSIA model.** The CoMSIA models also have been made by using five field descriptors namely (steric, electrostatic, hydrophobic, Hydrogen bond donor and acceptor)



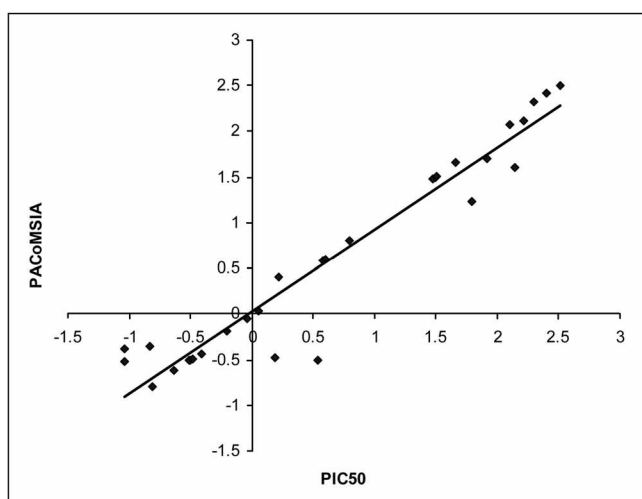
**Figure 4.** Trend of Observed and predicted pIC50 by CoMFA based model.



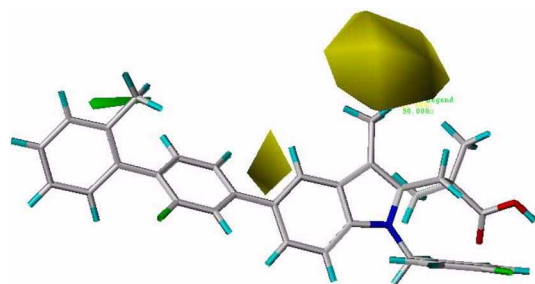
**Figure 5.** The CoMFA steric and electrostatic contour maps.

with same molecular alignment as for CoMFA. The steric field alone shows good relationship with the value of  $q^2 = 0.87$  while in conjunction with electrostatic field the result become more prominent ( $q^2 = 0.90$  and  $r^2 = 0.99$ ). The most fitted model ( $q^2 = 0.93$ ,  $r^2 = 1.00$ ) was obtained by combination of steric, electrostatic and hydrophobic fields. This model involves SE = 0.01 and F values = 6565.86 and the model was tested for internal predictivity ( $r^2_{\text{boot strap}} = 1.00$ ) and external predictivity ( $r^2_{\text{predictive}} = 0.80$ ) to test set of 10 molecules. The statistical summary and the predicted activities of training and test sets are reported in Table 5 and Table 6 respectively. The trend of observed and predicted activities of training and test sets is shown in Figure 6 with a value of Pearson  $R_{\text{testset}} = 0.91$ .

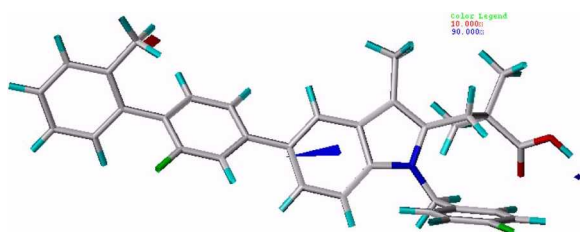
CoMSIA steric contour map was developed and shown in Figure 7 with template (ligand-30). The map is quite similar to CoMFA map and holds a green contour around ring B while yellow contour around ring A and pentagonal ring. It is clear from map the ring B is favorable for bulk while ring A and pentagonal ring is favorable for small groups. Similarly, the Figure 8 shows CoMSIA electrostatic contour map and a small blue contour appears around ring "A". indicates that a positive group around this site might improve the activity. The CoMSIA hydrophobic contour map shown in Figure 9 and hydrophobic favorable magenta



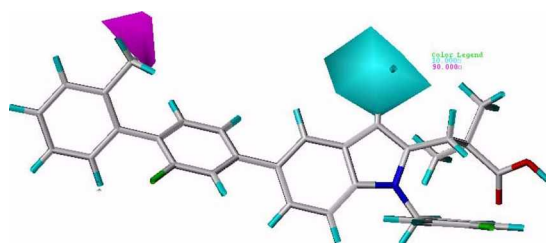
**Figure 6.** Trend of Observed and predicted pIC<sub>50</sub> by CoMSIA based model.



**Figure 7.** The CoMSIA steric map.



**Figure 8.** The CoMSIA electrostatic map.



**Figure 9.** The CoMSIA hydrophobic map.

contour appear around ring B. The same site holds a green contour in Figure 7 and red contour in Figure 8, which is quite reasonable, and a bulk with negative charge and hydrophobic nature might improve the activity. In quantum mechanical QSAR we have taken an Indicator parameter responsible for existence of ring B in molecule.

Since rings often related with hydrophobic nature and we

assumes that indicator parameter related with hydrophobicity, the CoMSIA gives a support to assumption and there is demand of hydrophobic group at fourth ring for which the "T" was considered. In same figure, a cyan contour appears near to pentagonal ring, which also holds a yellow contour in Figure 8 and indicates that a hydrophilic and small group is desirable around this site.

## Discussion

PGE synthase is a member of the MAPEG family, which includes FLAP and LTC<sub>4</sub> synthase. MK-886, which is a known inhibitor of PGE synthase, leukotriene biosynthesis,<sup>33</sup> and LTC<sub>4</sub> synthase.<sup>34</sup> Since there is no crystal structure of mPEGS-1, the ligand-based QSAR techniques are tools to understand inhibitor potency. Interestingly it is found that the region of FLAP essential for binding of MK-886<sup>35,36</sup> is highly conserved in PGE synthase. MK-886 appears to inhibit leukotriene biosynthesis by binding to an arachidonate binding site on FLAP.<sup>37</sup> The presence of a consensus amino acid sequence and sensitivity to indole inhibitors of the MK-886 series for FLAP, LTC<sub>4</sub> synthase, and PGE synthase suggest that this region might have similar interaction with MK-886. In fact, the negative charge of the aspartate or a glutamate at position 62 of FLAP is essential for binding MK-886 analogues. In present study it is clear from CoMFA map, there is a demand of positive groups to improve the activity. The positive group may facilitate the binding with negative zone of receptor. The position R1 of structure "C" is favorable for electropositive groups, which is also clear from compound 22 and 23. A fluorine atom instead of chlorine at R1 (structure-C) significantly increase the activity. Similarly, the site R2 of structure "C" is favorable for bulkiness. It is clear from data, the compounds 21 to 32 holds phenyl ring at site R2 that is favorable for bulkiness and the activity of all the compounds is comparatively higher than other molecules. As bulk increases to the phenyl ring of R2 (structure "C") the activity increases. The ring at R2 is very sensitive to activity, the compounds 30, 31 and 32 has similar structure but the position of methyl group varies (site 2, 3 and 4) which brought significant change in activity.

Recently Amor *et al.*<sup>38</sup> presented a systematic search based CoMFA and CoMSIA study of these compounds. The earlier work reports the contribution of hydrogen bond donor and acceptor field effect with steric electrostatic and hydrophobic interaction to activity but the model derived by steric, electrostatic and hydrophobic effect was statistically better than former model. The statistically refined results were presented but all available descriptors give high statistical values. In earlier CoMFA and CoMSIA study, there was no clear indication to key interaction. In this study, we have generated CPHs (common pharmacophore hypothesis) and there is no CPHs seen with hydrogen bond donor (D) or acceptor (A) but mostly CPHs based on HNR. We used the CPHs based alignment in CoMFA and CoMSIA, the model individually based on hydrogen bond donor field effect gives



leave one out  $q^2 = -0.60$  and hydrogen bond acceptor field effect gives leave one out  $q^2 = -0.10$  as reported in Table 5. The negative correlation coefficient supports to pharmacophore based assumption as there is no CPHs based on (D) or (A). In this way the results indicates that steric, electrostatic and hydrophobic effects are rather important and prominent factors for inhibition.

### Conclusion

To identify the key factors more systematically, we have considered the three different QSAR methods. The quantum mechanical QSAR based on molecular and atomic descriptors indicate, steric bulk, electrostatic fields effect and hydrophobicity jointly contributes to activity of a series of mPGES-1 inhibitors. In the current study, we have generated CPHs which are based on HNR. To identify the factors more systematically, we have considered the three different methods of QSAR. The QM-based QSAR indicates, steric bulk electrostatic and hydrophobic effect jointly contributes to activity of a series of mPGES-1 inhibitors. The best CPHs identified which indicates that HNR are mainly responsible. In ligand-based 3D-QSAR analyses (CoMFA and CoMSIA) imply that the steric, electrostatic and hydrophobic effects jointly contribute to the inhibitory activity. The contour maps indicate the bulky group around ring B and small group near pentagonal ring may be desirable for better activity. These findings might be helpful to further design novel compounds with enhanced activity against mPGES-1.

**Acknowledgments.** The study was supported by Computational science center of KIST.

### References

- Gilmour, R. S.; Mitchell, M. D. *Exp. Biol. Med. (Maywood)* **2001**, *226*, 1-4.
- Murakami, M.; Kudo, I. *Progress in Lipid Research* **2004**, *43*, 3-35.
- Thoren, S.; Jakobsson, P. J. *Eur. J. Biochem.* **2000**, *267*, 6428-6434.
- Qurashi, O.; Mancini, J. A.; Riendeau, D. *Biochem Pharmacol* **2002**, *63*, 1183-9.
- Riendeau, D.; Aspiotis, R.; Ethier, D.; Gareau, Y.; Grimm, E. L.; Guay, J.; Guiral, S.; Juteau, H.; Mancini, J. A.; Methot, N.; Rubin, J.; Friesen, R. W. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3352-3355.
- Bharatham, N.; Bharatham, K.; Lee, K. W. *Bull. Korean Chem. Soc.* **2007**, *28*, 200-206.
- Lee, Y.; Bharatham, N.; Bharatham, K.; Lee, K. W. *Bull. Korean Chem. Soc.* **2007**, *28*, 561-566.
- Padmanabhan, J.; Parthasarathi, R.; Subramanian, V.; Chattaraj, P. K. *Journal of Physical Chemistry A* **2007**, *111*, 1358-1361.
- Chattaraj, P. K.; Roy, D. R.; Elango, M.; Subramanian, V. *J. of Molecular Structure-Theochem* **2006**, *759*, 109-110.
- Pasha, F. A.; Chung, H. W.; Cho, S. J.; Kang, S. B. *International Journal of Quantum Chemistry* **2008**, *108*, 391-400.
- Pasha, F. A.; Nam, K. D.; Cho, S. J. *Molecular and Cellular Toxicology* **2007**, *3*, 145-149.
- Kim, J.; Han, J. H.; Chong, Y. *Bull. Korean Chem. Soc.* **2006**, *27*, 1919-1922.
- Kim, M. K.; Choo, I. H.; Lee, H. S.; Woo, J. I.; Chong, Y. *Bull. Korean Chem. Soc.* **2007**, *28*, 1231-1234.
- Lee, D. Y.; Hyun, K. H.; Park, H. Y.; Lee, K. A.; Lee, B. S.; Kim, C. K. *Bull. Korean Chem. Soc.* **2006**, *27*, 273-276.
- Myung, P. K.; Park, K. Y.; Sung, N. D. *Bull. Korean Chem. Soc.* **2005**, *26*, 1941-1945.
- San Juan, A. A.; Cho, S. J. *Bull. Korean Chem. Soc.* **2005**, *26*, 952-958.
- Sung, N. D.; Jang, S. C.; Choi, K. S. *Bull. Korean Chem. Soc.* **2006**, *27*, 1741-1746.
- Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
- Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130-4146.
- Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130-4146.
- Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. *Journal of Chemical Physics* **1978**, *68*, 3801-3807.
- Parr, R. G.; Von Szentpaly, L.; Liu, S. B. *J. Am. Chem. Soc.* **1999**, *121*, 1922-1924.
- Fukui, K.; Yonezawa, T.; Nagata, C.; Shingu, H. *Journal of Chemical Physics* **1954**, *22*, 1433-1442.
- Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. *Journal of Chemical Information and Computer Sciences* **1989**, *29*, 163-172.
- Ghose, A. K.; Pritchett, A.; Crippen, G. M. *Journal of Computational Chemistry* **1988**, *9*, 80-90.
- Klamt, A.; Schuurmann, G. *Journal of the Chemical Society-Perkin Transactions 2* **1993**, 799-805.
- Koopmans, T. C. *Physica* **1934**, *1*, 104-113.
- Phase, version 8.0; Schroedinger Inc.: Portland, USA, 2007.
- Kim, K. H.; Greco, G.; Novellino, E. *Perspectives in Drug Discovery and Design* **1998**, *12*, 257-315.
- SYBYL 7.5; Tripos Inc.: St. Louis, USA, 2007.
- Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. *Siam Journal on Scientific and Statistical Computing* **1984**, *5*, 735-743.
- Geladi, P.; Xie, Y. L.; Polissar, A.; Hopke, P. *Journal of Chemometrics* **1998**, *2*, 231.
- Rouzer, C. A.; Ford-Hutchinson, A. W.; Morton, H. E.; Gillard, J. W. *J. Biol. Chem.* **1990**, *265*, 1436-1442.
- Gupta, N.; Nicholson, D. W.; Ford-Hutchinson, A. W. *Can. J. Physiol. Pharmacol.* **1997**, *75*, 1212-1219.
- Vickers, P. J.; Adam, M.; Charleson, S.; Coppolino, M. G.; Evans, J. F.; Mancini, J. A. *Mol. Pharmacol.* **1992**, *42*, 94-102.
- Mancini, J. A.; Coppolino, M. G.; Klassen, J. H.; Charleson, S.; Vickers, P. J. *Life Sci.* **1994**, *54*, PL137-142.
- Mancini, J. A.; Waterman, H.; Riendeau, D. *J. Biol. Chem.* **1998**, *273*, 32842-32847.
- San Juan, A. A.; Cho, S. J. *J. Mol. Model* **2007**, *13*, 601-610.