Ligand-based QSAR Studies on the Indolinones Derivatives as Inhibitors of the Protein Tyrosine Kinase of Fibroblast Growth Factor Receptor by CoMFA and CoMSIA

Kwan Hoon Hyun, In Young Kwack, Do Young Lee, Hyung Yeon Park, Bon-Su Lee, and Chan Kyung Kim[*]

Department of Chemistry, Inha University, Incheon 402-751, Korea
Received June 30, 2004

Ligand-based quantitative structure-activity relationship (QSAR) studies were performed on indolinones derivatives as a potential inhibitor of the protein tyrosine kinase of fibroblast growth factor receptor (FGFR) by comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) implemented in the SYBYL packages. The initial X-ray structure of docked ligand (Su5402) to FGFR was used to minimize the 27 training set molecules using TRIPOS force field. Seven models were generated using CoMFA and CoMSIA with grid spacing 2 Å. After the PLS analysis the best predicted CoMSIA model with hydrophobicity, hydrogen bond donor and acceptor property showed that a leave-one out (LOO) cross validated value (r^2_{cv}) and non-cross validated conventional value (r^2_{ncv}) are 0.543 and 0.938, respectively.

Key Words : Ligand-based QSAR, CoMFA, CoMSIA, Protein tyrosine kinase inhibitor, Fibroblast growth factor receptor

Introduction

Computer-aided drug design[1] is a new field to develop new biologically active compounds based on the knowledge of previously synthesized ones. The most widely used approach, quantitative structure-activity relationship (QSAR)[2] is to correlate observed biological activities with structural changes of ligands. 3D QSAR techniques, such as the comparative molecular field analysis (CoMFA)[3] and the comparative molecular similarity indices analysis (CoMSIA),[4] are based on the experimental structure-activity relationship on specific bio-macromolecule and ligand pair. This method is based only on the ligand structure and thus the spatial arrangement (or alignment) is crucial in determining the accuracy of these approaches.

Selective inhibition of the protein tyrosine kinases (PTKs) inhibitor of Fibroblast Growth Factor Receptor (FGFR) is considered as a major emerging strategy in cancer therapy because PTKs are critical components of signal pathways that control cell proliferation and differentiation and enhanced PTK activity due to activating mutations or over-expression has been implicated in many human cancers.[5]

Indolinone derivative inhibitors (IDIs) of protein tyrosine kinase of FGFR with high potency and selectivity have already been developed by Cho Tang etc. in Sugen Inc. and have been under clinical evaluation.[6] QSAR study on PTKs of FGFR was studied using the 2D descriptors by Corwin Hansch group.[7]

In this work, however, we plan to perform QSAR study on IDIs using the ligand-based CoMFA and CoMSIA to correlate their biological activities with three-dimensional structures and to provide useful information necessary for designing improved lead compounds.

Methods

Data Set for Analysis. To perform the ligand-based QSAR study, 27 IDIs, which have been synthesized and tested for biological inhibitory activity toward tyrosine phosphorylation were taken from the literature.[6,8] Experimentally, the indoline-2-one core is known to occupied the adenine binding site of ATP[8] and thus the information deduced from the substitution on the core is useful for further modification of indoline-2-one as inhibitors against FGFR. The molecular structures are shown in Figure 1. From the reference data, IC_{50} values were defined as the concentration of a compound required to achieve 50% inhibition of maximal tyrosine phosphorylation as measured by bromodeoxyuridine (BrdU) incorporation when compared to vehicle-treated controls (DMSO).

Computational Methods and Ligand-based Molecular Alignment. The crystal structure of Su5402 (3-[(3-(2-carboxethyl)-4-methylpyrrol-2-yl) methylene]-2-indolinone) was used to generate initial structures of 27 IDIs.[6,8] All ligands were sketched using the sketch module in the SYBYL package and conformational searches were performed by grid search which calculates energies by systematically changing the dihedral angles of each ligand using standard TRIPOS force field.[9] Among them, the lowest energy structures were selected as the conformers for the 3D-QSAR studies. Finally all ligands were fully optimized using the standard TRIPOS force field with Gasteiger-Hückel charges until the energy gradient converged to below 0.05 kcal/mol. Total 27 ligand structures were selected as training set and indolinone fragment of the molecules were used for the alignment. Result of the superimposed image of 27 ligand structures is shown in Figure 2. All calculations were done on a SGI octane 2 workstation using SYBYL 6.9 software packages.[10]

[*]To whom correspondence should be addressed. Tel: +82-32-860-7684; Fax: +82-32-867-5604; e-mail: kckyung@inha.ac.kr
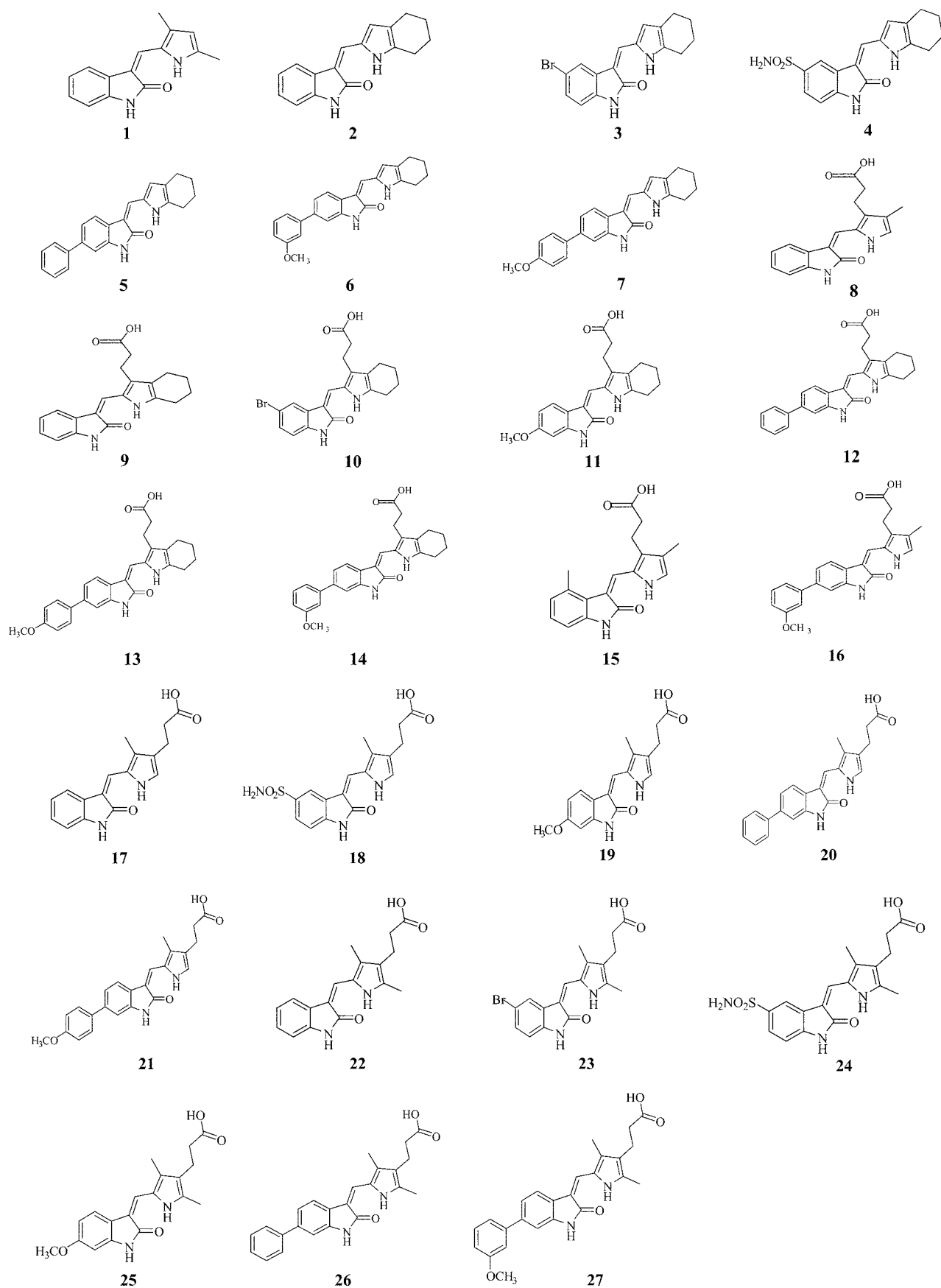
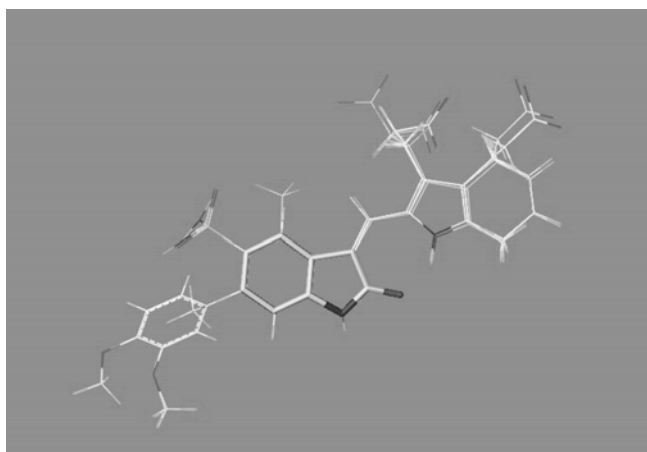**Figure 1**. Structures of indolinone-based protein tyrosine kinase inhibitors used as a training set.

Figure 2. Superimposed complexes model using indolinone based molecular alignment. Atoms with capped sticks type selected as the fitting fragment.

PLS Analysis Using CoMFA and CoMSIA. CoMFA is one of the well known 3D-QSAR descriptors which has been used to produce the 3D models to indicate the regions that affect biological activity with a change in the chemical substitution.[11-13] CoMFA quantifies the statistical relationship between the 3D properties of a set of small molecules and a global property, such as their potency in a particular biological assay. CoMFA was performed with the QSAR option of SYBYL. For all steps of conventional CoMFA, the default SYBYL settings were used except otherwise noted. For each CoMFA analysis, the minimum s was set to 2.0 to expedite the calculation. The steric and electrostatic field energies were calculated using sp^3 carbon probe atoms with +1 charge. CoMFA grid spacing used in this work was 2.0 Å in all X, Y and Z directions. CoMFA QSAR equations were calculated with the partial least square (PLS) algorithm. The optimal number of components (ONC) in the final PLS model was determined by the r^2_{cv} and standard error of estimate values, obtained from the leave-one-out cross-validation technique. The van der Waals potential and columbic terms, which represent the steric and electrostatic terms, respectively, were calculated using the standard TRIPOS force field. A distance dependent dielectric constant of 1.00 was used. Values of the steric and electrostatic energy were truncated at 30 kcal/mol.

CoMSIA calculates similarity indices at the intersections of a surrounding lattice. The five CoMSIA fields of steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor were calculated at the grid lattice point using a probe atom of 2.0 Å radius as well as the charge, hydrophobic and hydrogen bond properties of H, and an attenuation factor of 0.3. The CoMSIA approach can avoid some inherent deficiencies arising from the functional form of Lennard-Jones and Coulomb potentials used in the conventional CoMFA. In CoMSIA, a distance-dependent Gaussian-type functional form has been introduced, which can avoid singularities at the atomic positions and the dramatic changes of potential energy for these grids in the proximity

of the surface. Compared with usual CoMFA, CoMSIA has better ability to visualize and interpret correlations obtained in terms of field contributions. The unique difference between conventional CoMFA and CoMSIA is the field type and the field calculation function.

The partial least-squares (PLS) analysis algorithm was used in conjugation with the cross-validation (leave-one-out) option to obtain an optimum number of components, which were used to generate the final CoMFA and CoMSIA models without cross validation. The result from a cross validation analysis was expressed as r^2_{cv} defined in eq. (1):

$$r^2_{cv} = 1 - PRESS / \Sigma(Y_{obs} - Y_{mean})^2 \qquad (1)$$

The PRESS statistic is calculated by the following equation, eq. (2).

$$PRESS = \Sigma (Y_{obs} - Y_{pred})^2 \qquad (2)$$

where, Y_{obs}, Y_{mean} and Y_{pred} are observed, the mean of the predicted and predicted values of the activity, respectively.

Results and Discussion

In this work, CoMFA and CoMSIA approaches were used to elucidate the QSAR as descriptors for protein tyrosine kinase inhibitors of FGFR biological activity.

CoMFA offers two different descriptors and CoMSIA offers five different descriptors; therefore, diverse combinations of descriptors are possible for CoMSIA.[14] Seven different models were generated using 2 Å grid spacing with diverse descriptors - Model 1: CoMFA with steric and electrostatic descriptors, Model 2: CoMSIA with steric, electrostatic, hydrogen bond donor, acceptor and hydrophobic descriptors, Model 3: CoMSIA with steric and electrostatic descriptors, Model 4: CoMSIA with hydrogen bond donor and acceptor, Model 5: CoMSIA with steric, electrostatic and hydrophobic descriptors, Model 6: CoMSIA with steric, electrostatic, hydrogen bond donor and acceptor descriptors, and Model 7: CoMSIA with hydrogen bond donor, acceptor and hydrophobic descriptors.

The results obtained from the PLS analysis are summarized in Table 1. Inspection of Table 1 shows that leave-one out cross-validated value (r^2_{cv}) is 0.494 and non cross-validated conventional value is 0.840 for CoMFA Model 1 and the results of PLS analysis for CoMSIA Model 2 show that r^2_{cv} is 0.527 and r^2_{ncv} is 0.900. This means that the CoMSIA model is better than CoMFA model to elucidate the QSAR as descriptors for PTK inhibitors of FGFR biological activity. In order to obtain better predictive values using CoMSIA, five more models were considered.

In these approaches, Model 7 with hydrogen bond donor, acceptor and hydrophobic descriptors shows best correlation - leave-one out cross-validated value (r^2_{cv}) and non cross-validated conventional value (r^2_{ncv}) are 0.543 and 0.938, respectively. Smaller standard error of estimate values (SEE) and higher F values also suggests that the selected CoMSIA Model 7 is better than the others. From this analysis, we can conclude that Model 7 is the best model.

Table 1. The results of PLS analysis in the Training Set using CoMFA and CoMSIA

		CoMFA		CoMSIA				
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Selected Fields[a]	SE	SEDAH	SE	DA	SEH	SEDA	DAH
PLS Analysis	r^2_{cv}	0.494	0.527	0.397	0.567	0.481	0.504	0.543
	Components	3	3	5	4	5	3	4
	r^2_{ncv}	0.840	0.900	0.859	0.684	0.911	0.891	0.938
	SEE	0.323	0.255	0.318	0.464	0.253	0.267	0.206
	F Value	40.25	69.05	25.48	11.91	47.73	62.60	82.81
Contribution	Steric	0.470	0.060	0.296		0.131	0.072	
	Electrostatic	0.530	0.307	0.704		0.481	0.390	
	H Bond Donor		0.449		0.911		0.488	0.068
	H Bond Acceptor		0.043		0.089		0.050	0.498
	Hydrophobic		0.144			0.388		0.434

[a]S=steric, E=electrostatic, D=hydrogen bond donor A=hydrogen bond acceptor, H=hydrophobic

Table 2. Actual and predicted activities (pIC_{50}) of the Training Set Molecules using Model 7

	pIC_{50}				pIC_{50}		
NO.	Actual	Pred.	Residual	NO.	Actual	Pred.	Residual
1	5.15	5.13	0.02	15	7.52	7.23	0.29
2	4.98	4.80	0.18	16	5.92	6.25	−0.33
3	4.88	4.67	0.21	17	5.43	5.59	−0.16
4	6.66	6.45	0.21	18	6.34	6.59	−0.25
5	4.70	4.93	−0.23	19	5.41	5.57	−0.16
6	4.76	4.90	−0.14	20	5.98	5.72	0.26
7	4.70	4.89	−0.19	21	5.81	5.68	0.13
8	7.52	7.48	0.04	22	5.52	5.51	0.01
9	6.57	6.79	−0.22	23	5.69	5.39	0.30
10	6.57	6.66	−0.09	24	6.55	6.49	0.06
11	5.97	5.83	0.14	25	5.27	5.49	−0.22
12	5.90	5.99	−0.09	26	5.64	5.65	−0.01
13	5.87	5.96	−0.09	27	5.85	5.61	0.24
14	6.06	5.95	0.11		Ave.		0.16

Figure 3. Comparison of actual *vs* predicted pIC_{50} (based on the data of Table 2).

Activities of 27 compounds in the training set are predicted from the PLS analysis using the best model and the results along with their actual pIC_{50} values and residuals are summarized in Table 2. A plot of actual pIC_{50} vs. predicted pIC_{50} is shown in Figure 3. Close examination of Table 2 shows that average residual of the actual and predicted values is 0.16. Therefore we can confirm that Model 7 is the best, which is also apparent from the linear plot depicted in Figure 3.

Graphical representations of the selected CoMSIA maps for the best model are displayed in Figures 4-6. The contour maps were superimposed on Su 5402 (entry 8) shown in atom type color. The field type "stdev*coeff" was used to obtain contour maps in CoMSIA.

Hydrogen bond donor contour map of CoMSIA is depicted in Figure 4. Cyan colored region where hydrogen bond donor is associated with enhanced affinity is not found in this map. But orange colored regions where hydrogen

bond donor is associated with diminished affinity are found in the near 5-position of indolinone ring, which is found for the compounds with 5-SO_2NH_2 group (entry 4, 18, and 24). Hydrogen bond acceptor contour map of CoMSIA is described in Figure 5. Blue colored regions where hydrogen bond acceptor is associated with enhanced affinity is found largely in near carboxyethyl group of methyl pyrrol ring. It indicates that carboxyl group found for entry 8 or 15 is essential for the interaction with amino acid residues of the active site. But red colored region where hydrogen bond acceptor is associated with diminished affinity is not found in this map. In Figure 6, hydrophobic contour map of CoMSIA is shown. Yellow color contour region where hydrophobicity is associated with enhanced affinity is near the 6 position of indolinone ring (compare entry 19 vs. entry 20 or entry 25 vs. entry 26), aliphatic hydrophobic side chain of carboxyethyl and near the methyl pyrrol ring, which is found for entry 8 and 15 with carboxyethyl group at 4 position and CH_3 group at 3 position. On the other hand,
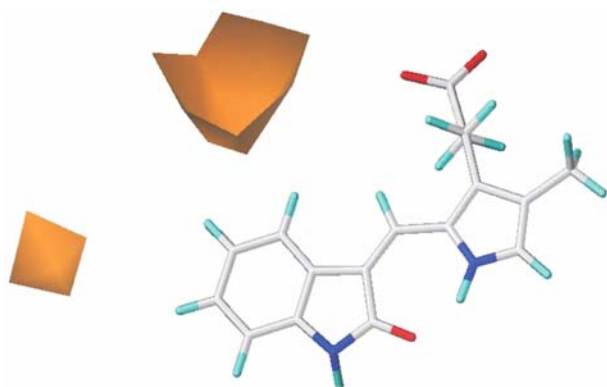
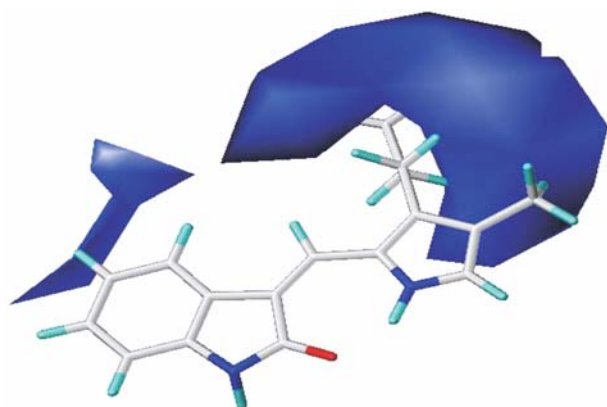**Figure 4**. Superposition of the CoMSIA Hydrogen bond donor stdev*coeff contour plots

Figure 5. Superposition of the CoMSIA Hydrogen bond acceptor stdev*coeff contour plots
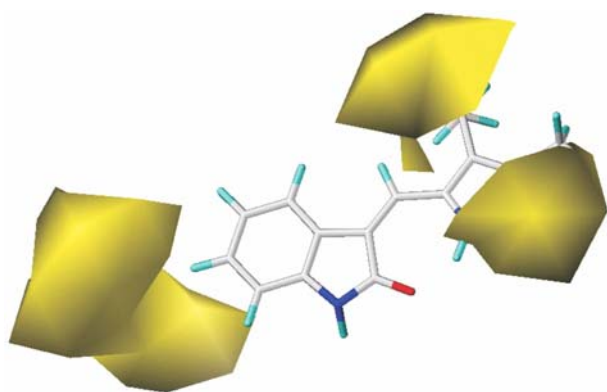

Figure 6. Superposition of the CoMSIA Hydrophobic stdev*coeff contour plots.

white color contour region where hydrophobicity is associated with diminished affinity is not found.

To further validate our results, five compounds which were not included in the training set molecules with pIC_{50} range between 4.70 and 7.10 were assigned as a test set and their biological activities are predicted from the PLS equation derived from Model 7. Predicted and actual activities of test set molecules are summarized in Table 3.

Table 3. Actual and predicted activities (pIC_{50}) of test set molecules using CoMSIA Model 7

No.	Structure	Actual pIC_{50}	CoMSIA Model 7	
			Predicted pIC_{50}	Residual
T1[a]		4.70	4.78	-0.08
T2[b]		5.24	5.60	-0.36
T3[b]		6.44	5.69	0.75
T4[a]		6.70	6.79	-0.09
T5[b]		7.10	7.00	0.10
	AVE.			0.28

[a]From reference 6(b). [b]From reference 6(a).

Predicted pIC_{50} values agree well with the experimental ones with average deviation of 0.28, which suggests that our best model is good in predicting the pIC_{50} values.

Conclusion

In order to perform ligand-based QSAR studies, 27 training set molecules as an inhibitor of the protein tyrosine kinase of FGFR were selected as a training set and indolinone fragment of the molecules were used for the alignment. With combination of diverse descriptors, seven models were generated from CoMFA and CoMSIA. The best model, Model 7, was CoMSIA with hydrogen bond donor, acceptor and hydrophobic descriptors, and this model showed good

1806   *Bull. Korean Chem. Soc.* **2004**, Vol. 25, No. 12

   Kwan Hoon Hyun et al.

statistical results. From above PLS correlations, activities of five test set molecules were predicted satisfactorily. Based on above 3D-QSAR results, we are currently performing virtual screen work to find novel inhibitors.

References

1. (a) Perun, T. J.; Propst, C. L. In *Computer-Aided Drug Design*; Marcel Dekker Inc.: New York, 1989. (b) Charifson, P. S. In *Practical Application of Computer-Aided Drug Design*; Marcel Dekker Inc.: New York, 1997.

2. (a) Hansch, C.; Leo, A. In *Exploring QSAR*; American Chemical Society: Washington DC, 1995. (b) Sung, N.-D.; Cheun, Y. G.; Kwon, B.-M.; Park, H. Y.; Kim, C. K. *Bull. Korean Chem. Soc.* **2003**, *10*, 1509.

3. (a) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959. (b) Dean, P. M. In *Molecular Simularity in Drug Design*; Blackie Academic & Professional: Glasgow, 1995; p 291. (c) Suh, M.-E.; Park, S.-Y.; Lee, H.-J. *Bull. Korean Chem. Soc.* **2002**, *23*, 417.

4. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.

5. Levitzki, A.; Gazit, A. *Science* **1995**, *267*, 1782.

6. (a) Sun, L.; Tran, N.; Liang, C.; Tang, F.; Rice, A.; Schreck, R.; Waltz, K.; Shawver, L. K.; McMahon, G.; Tang, C. *J. Med. Chem.* **1999**, *42*, 5120. (b) Sun, L.; Tran, N.; Liang, C.; Hubbard, S.; Tang, F.; Lipson, K.; Schreck, R.; Zhou, Y.; McMahon, G.; Tang, C. *J. Med. Chem.* **2000**, *43*, 2655.

7. Kurup, A.; Garg, R.; Hansch, C. *Chem. Rev.* **2001**, *101*, 2573.

8. Mohammadi, M.; McMahon, G.; Sun, L.; Tang, C.; Hirth, P.; Yeh, B. K.; Hubbard, S. R.; Schlessinger, J. *Science* **1997**, *276*, 955.

9. (a) Clark, M.; Cramer, R. D.; Opdenbosch, N. V. *J. Comp. Chem.* **1989**, *10*, 982. (b) Vinter, J. G.; Davis, A.; Saunders, M. R. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 31.

10. *SYBYL Molecular Modeling Software*; Tripos Inc.: 1699 South Hanley Rd, Suite 303, St. Louis, MO 63144, 2002.

11. Huang, M.; Yang, D.-Y.; Shang, Z.; Zou, J.; Yu, Q. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2271.

12. Zhu, L.; Hou, T.; Xu, X. *J. Mol. Model.* **2001**, *7*, 223.

13. Cho, S. J.; Garsia, M. L. S.; Bier, J.; Tropsha, A. *J. Med. Chem.* **1996**, *39*, 5064.

14. Raichurkar, A. V.; Kulkarni, V. M. *J. Med. Chem.* **2003**, *46*, 4419.