Quantitative Structure-Activity Relationship (QSAR) Study of New Fluorovinyloxyacetamides

Doo Ho Cho,[†‡] Sung Kwang Lee,[†] Bum Tae Kim,[‡*] and Kyoung Tai No[†*]

[†]*Department of Chemistry and CAMD Research Center, Soongsil University, Seoul 156-743, Korea*
[‡]*Korea Research Institute of Chemical Technology, Taejon 305-600, Korea*
Received November 13, 2000

Quantitative Structure-Activity Relationship (QSAR) have been established of 57 fluorovinyloxyacetamides compounds to correlate and predict EC_{50} values. Genetic algorithm (GA) and multiple linear regression analysis were used to select the descriptors and to generate the equations that relate the structural features to the biological activities. This equation consists of three descriptors calculated from the molecular structures with molecular mechanics and quantum-chemical methods. The results of MLR and GA show that dipole moment of z-axis, radius of gyration and logP play an important role in growth inhibition of barnyard grass.

Keywords : QSAR, Genetic algorithm (GA), Fluorovinyloxyacetamides

Introduction

Oxyacetamide has been developed as a herbicide, which shows good herbicidal activity for annual weeds through inhibition of cell division and amino acid biosynthesis.[1] Especially, the excellent selectivity between crop (rice) and weeds (barnyard grass), an important weed in paddy fields, is a merit of oxyacetamide herbicide.[2] To exploit a new and highly active herbicide, fluorovinyloxyacetamides, which was introduced fluorovinyl group into oxyacetamide, were synthesized and evaluated their herbicidal activities. In this research, we tried to describe the synthetic method and QSAR study of fluorovinyloxyacetamides.

Synthesis of fluorovinyloxyacetamides[3-5] is as follows (Figure 1). The chloroacetamides (**e**) were prepared from the substitution reaction of amines (**g**) with chloroacetyl chloride (**f**) using sodium hydroxide as a base in THF. The reaction of (**e)** with sodium acetate in DMF gave acetates (**d**) in high yields. The hydroxyacetamides (**b**) were obtained from base catalyzed hydrolysis of (**d)**. On the other hand, the vinyl fluorides (**c**) were provided by the Wittig reaction of trifluoromethyl ketones (**h**) with dibromodifluoromethane, triphenylphosphine in THF. Finally, addition-elimination reaction of (**b)** with (**c)** in the basic condition gave the fluorovinyloxyacetamides (**a**). The structures of new fluorovinyloxyacetamides (**a**) are summarized in Table 1.

Quantitative Structure-Activity Relationship (QSAR)[6-11] is a powerful method for the design of bioactive compounds and the prediction of corresponding activity with physical and chemical properties. Usually there are two major approaches to analyze QSAR data : i) the property (or activity) of a series of compounds is expressed as a multiple linear regression of descriptors, and ii) the non-linear regression method represents the property (or activity) with artificial neural network (ANN).[12,13] ANN is an information-processing paradigm inspired by the densely interconnected, parallel structure of the mammalian brain processes information.

This study was based on a Genetic Function Approximation (GFA) algorithm.[14,15] GFA can not only automatically select the optimum number of descriptors in regression analysis but also construct Multiple Linear Regression (MLR) models through the use of linear, higher order polynomials, splines and gaussians. The GFA algorithm method was used to select the optimum number of descriptors for use in regression analysis. The GFA algorithm could be a useful technique for searching the large probability space with a large number of descriptors for a small number of molecules.

The purpose of this research was to determine predictive QSAR models[16-19] by analysis of training set containing 57 molecules. If the models are reasonable, it is possible to predict biological activity of non-tested molecules. Finally, the successful models of QSAR certainly decrease the number of compounds to be synthesized, by making it possible to select the most promising compounds.

Methods

Experimental; Chemicals. 1-Piperidino-2-[(*E*)-1,3,3,3-tetrafluoro-2-(3-methylphenyl)-1-propenyloxy]-1-ethanone (compound 1): A solution of 286 mg (2 mmol) of *N*-2-hydroxyacetylpiperidine in 10 mL THF was added to 444 mg (2 mmol) of 1,1,3,3,3-pentafluoro-2-(3-methylphenyl)-1-propene at room temperature. The reaction mixture was treated with 0.2 mL of 10 M-NaOH solution and stirred for 30 min. The mixture was washed with 10 mL of water and extracted with ethyl acetate. The extract was dried over anhydrous $MgSO_4$ and concentrated under reduced pressure. The 310 mg (90%) of compound 1 (an *E* and *Z* isomeric mixtures on silica gel) was extracted with mixture of solvents (ethyl acetate/*n*-hexane = 1 : 2). The separation of geometrical isomers by silica gel chromatography (ethyl acetate/*n*-hexane = 1 : 4) was provided by 230 mg of *E*-isomer. oil; ^{1}H-NMR (300 MHz; $CDCl_3$, TMS) δ: 7.31-7.04 (m, 4H), 4.78 (E) 4.66 (Z) (s, 2H), 3.64-3.03 (m, 4H), 2.37 (s, 3H), 1.74-1.38 (m, 6H); ^{19}F-NMR (200 MHz; $CDCl_3$, $CFCl_3$) δ:

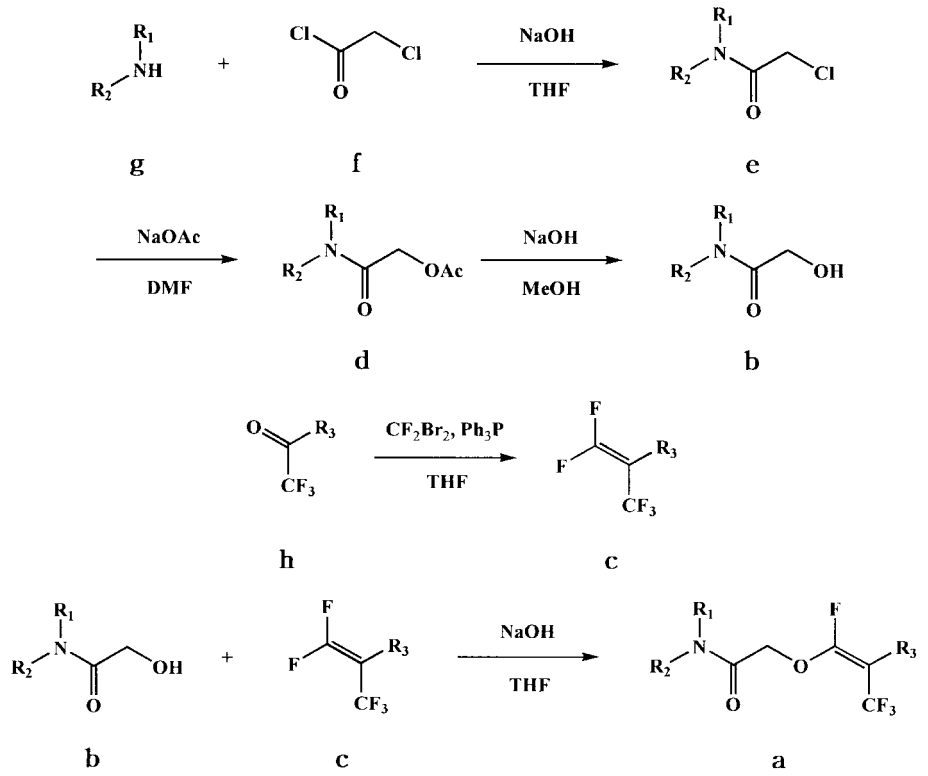


Figure 1. Synthesis of fluorovinyloxyacetamides.

-57.83 (d, 3F), -82.24 (q, 1F); MS m/z (rel. int.) 345 (26), 126 (71), 98 (70), 84 (100).

The other compounds (2 to 57) were obtained about 70 to 90% yields from the same reaction condition using corresponding hydroxyacetamides (**b**) and vinyl fluorides (**c**).

Computations; Data Construction. A series of 57 fluorovinyloxyacetamides derivatives was used in this work (Table 1). These molecules consisted of mixture of *E* and *Z* isomer, but the energy of all molecules was minimized at *E* isomer configuration. A grid search procedure was performed with core structure (R_1=CH_3, R_2=phenyl, R_3=phenyl) to identify their lowest energy conformation. The conformational energy of each conformer was generated by increasing the torsional angle ψ_1 and ψ_2 by 30 degrees. Geometry optimization was performed to obtain fixed core structure using the results of this conformational search (the minimum energy conformations have 30-130° angle range of ψ_1 and 150-200° angle range of ψ_2) by *ab initio* calculation with 4-31G basis set of Gaussian 94 program (see Figure 2).[20] After geometry optimization, functional groups were added to the core structure and the full geometry optimization was performed by Merck Molecular Force Field (MMFF)[21,22] method. After minimization, all molecules were aligned in the orientation that they were assumed to bind to the putative receptor. The method for performing the alignment was Maximum Common Sub Group (MCSG).[23,24] The MCSG was carried out rigid fit to superimpose each structure to overlays the shape reference compound (compound 45).

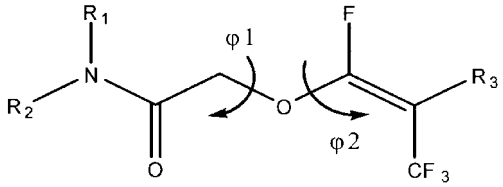
The activity was expressed in terms of 50%-growth-inhibition concentration (EC_{50}) of barnyard grass. The EC_{50} values were estimated by fitting data to sigmoidal type function (1) because the experiments were performed in a range of concentration (1.000, 0.2500, 0.0625, 0.0156, 0.0040 kg/ha) and activities were represented by 10 percent unit. In the equation (1), the initial and final value were fixed. Activities were demonstrated to 100 percent, and other variables were changed to fit. The concentration and EC_{50} value were converted by -log function to fit scale.

$$y = \frac{A_1 - A_2}{1+e^{(x-x_0)/dx}} + A_2 \quad (1)$$

(A_1: initial value = 0, A_2: final value = 100, χ_0: center, *dx*: time const.)

Selection of the Descriptors and the Activity Description. It is necessary to construct a numerical descriptors of a set of molecule in order to build QSAR models. A descriptor can be a quantitative property that depends on the structure of molecule. In this study, all 118 descriptors such as topological, spatial, electronic, quantum mechanical, and thermodynamic descriptors were calculated by Cerius2[25] program. Genetic Function Approximation (GFA) technique was utilized to select descriptors and to generate different QSAR models from various descriptors. GFA technique began with a population of 100 random models and 10000 iterations to evolution. The descriptors were selected by a few steps using GFA method: i) 118 descriptors were divided by four or five groups which were randomly selected, ii) GFA allowed the selection of some descriptor that is frequently used from each group of descriptors. Repeating this step,

Table 1. Structures of Fluorovinyloxyacetamides in Training Set



No.	R_1	R_2	R_3	No.	R_1	R_2	R_3
1	-$(CH_2)_5$-		3-CH_3-C_6H_4	31	4-CH_3O-C_6H_4	C_2H_5	C_6H_5
2	-$(CH_2)_5$-		4-C_2H_5-C_6H_4	32	3-CF_3-C_6H_4	C_2H_5	4-CH_3-C_6H_4
3	-$(CH_2)_5$-		3,5-Cl_2-C_6H_3	33	4-Cl-C_6H_4	C_2H_5	4-CH_3-C_6H_4
4	-$CH(CH_3)(CH_2)_4$-		4-CH_3-C_6H_4	34	4-CH_3O-C_6H_4	*n*-C_3H_7	4-CH_3-C_6H_4
5	-$CH(C_2H_5)(CH_2)_4$-		4-CH_3-C_6H_4	35	C_6H_5	*i*-C_3H_7	3,5-$(CH_3)_2$-C_6H_3
6	-$CH(CH_3)(CH_2)_3CH(CH_3)$-		3-Cl-C_6H_4	36	C_6H_5	*i*-C_3H_7	4-CH_3O-C_6H_4
7	-$(CH_2)_6$-		3-CH_3-C_6H_4	37	C_6H_5	*i*-C_3H_7	4-C_2H_5O-C_6H_4
8	-$(CH_2)_6$-		4-C_2H_5-C_6H_4	38	3-CH_3-C_6H_4	*i*-C_3H_7	C_6H_5
9	-$(CH_2)_6$-		3,4-$(CH_3)_2$-C_6H_3	39	3-CH_3-C_6H_4	*i*-C_3H_7	4-C_2H_5-C_6H_4
10	-$(CH_2)_6$-		3,5-$(CH_3)_2$-C_6H_3	40	3-CH_3-C_6H_4	*i*-C_3H_7	3-CH_3O-C_6H_4
11	-$(CH_2)_6$-		3-CH_3O-C_6H_4	41	3-CH_3-C_6H_4	*i*-C_3H_7	3-F-C_6H_4
12	-$(CH_2)_6$-		4-Cl-C_6H_4	42	3-CH_3O-C_6H_4	*i*-C_3H_7	C_6H_5
13	C_2H_5	n-C_4H_9	4-CH_3-C_6H_4	43	3-CH_3O-C_6H_4	*i*-C_3H_7	4-CH_3-C_6H_4
14	C_6H_5	CH_3	C_6H_5	44	4-CH_3O-C_6H_4	*i*-C_3H_7	4-CH_3-C_6H_4
15	C_6H_5	CH_3	3-CH_3-C_6H_4	45	4-CH_3O-C_6H_4	*i*-C_3H_7	4-CH_3O-C_6H_4
16	C_6H_5	CH_3	4-CH_3-C_6H_4	46	4-CH_3O-C_6H_4	*i*-C_3H_7	4-F-C_6H_4
17	C_6H_5	CH_3	4-C_2H_5-C_6H_4	47	4-F-C_6H_4	*i*-C_3H_7	3-CF_3-C_6H_4
18	C_6H_5	CH_3	4-C_2H_5O-C_6H_4	48	4-F-C_6H_4	*i*-C_3H_7	4-F-C_6H_4
19	C_6H_5	CH_3	3-CF_3-C_6H_4	49	3-Cl-C_6H_4	*i*-C_3H_7	4-F-C_6H_4
20	C_6H_5	CH_3	3-F-C_6H_4	50	4-CH_3-C_6H_4	CH_3	C_6H_5
21	C_6H_5	CH_3	4-Cl-C_6H_4	51	3,4-$(CH_3)_2$-C_6H_3	CH_3	4-CH_3-C_6H_4
22	4-CH_3O-C_6H_4	CH_3	3-CH_3-C_6H_4	52	3-Cl-C_6H_4	CH_3	3,4-$(CH_3)_2$-C_6H_3
23	4-CH_3O-C_6H_4	CH_3	3,4-$(CH_3)_2$-C_6H_3	53	3-Cl-C_6H_4	CH_3	4-C_2H_5-C_6H_4
24	4-CH_3O-C_6H_4	CH_3	3,4-OCH_2O-C_6H_3	54	3-Cl-C_6H_4	CH_3	4-CH_3O-C_6H_4
25	4-F-C_6H_4	CH_3	C_6H_5	55	4-Cl-C_6H_4	CH_3	C_6H_5
26	4-F-C_6H_4	CH_3	4-CH_3-C_6H_4	56	4-Cl-C_6H_4	CH_3	3-CH_3O-C_6H_4
27	4-F-C_6H_4	CH_3	4-CH_3O-C_6H_4	57	4-Cl-C_6H_4	CH_3	4-CH_3O-C_6H_4
28	2,4-F_2-C_6H_3	CH_3	3-Cl-C_6H_4				
29	2,4-Cl_2-C_6H_3	CH_3	4-F-C_6H_4				
30	C_6H_5	C_2H_5	4-F-C_6H_4				

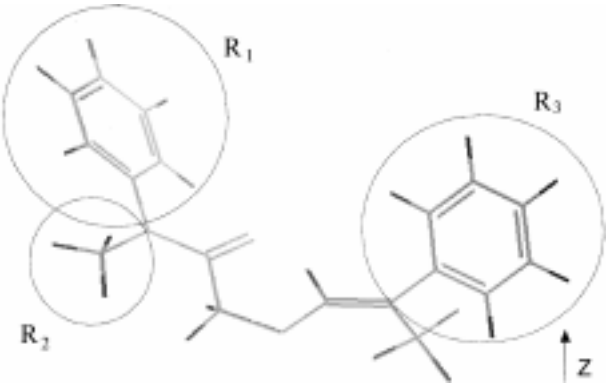


Figure 2. Geometry optimized core structure.

finally 16 descriptors (Table 2) were selected. These descriptors have high divergency and good representation of biological activity.

Results and Discussion

In this study, we screened 16 preselected descriptors for 57 fluorovinyloxyacetamides compounds using GFA method. Finally, we generated 100 QSAR equations that consist of four descriptors among QSAR random models. As a rule of thumb,[26] data set should be approximately 5 times more than the number of selected descriptors for good results. The results of the best QSAR model using 1-4 descriptors are given in Table 3. Regression models are all significant at *p*-

Table 2. List of Descriptors Used in This Study

Abbreviation	Definition
Dipole-X, Y, Z	X, Y, Z component of the dipole moment
LUMO-MOPAC	The energy of Lowest Unoccupied Molecular Orbital (LUMO) by MOPAC AM1 calculation
HOMO-MOPAC	The energy of Highest Occupied Molecular Orbital (HOMO) by MOPAC AM1 calculation
RadOfGyration	Radius of Gyration
PMI-X, Y, Z	X, Y, Z component of principle moment of inertia
logP	Calculated LogP (the partition coefficient) by Desolvation free energy of water (Fh2o) and octanol (Foct)
Vm	Molecular volume
Foct	Desolvation free energy of octanol
ShapeRMS	RMS value to shape reference
Dipole-MOPAC	Dipole moment by MOPAC AM1 calculation
Density	Density
MolRef	Molar refractivity by Ghose and Crippen calculation

Table 3. The best QSAR equation using 1-4 descriptors and their regression statistics

No. of descriptor	Equation	r^2	r^2_{CV}	SE	F	p-value
1	5.415-0.0109 (Vm)	0.627	0.598	0.190	92.357	2.27E-13
2	5.477-0.355 (logP) -0.545 (RadOfGyration)	0.810	0.789	0.137	115.021	3.41E-20
3	5.365-0.350 (logP) -0.524(RadOfGyration) +0.0577 (Dipole-Z)	0.853	0.832	0.121	102.684	4.52E-22
4	5.600-0.341 (logP) -0.560 (RadOfGyration) +0.0581 (Dipole-Z) -0.122 (ShapeRMS)	0.859	0.829	0.120	79.317	1.71E-21

r^2: correlation coefficient, r^2_{CV}: cross-validated r^2, E: standard error, F: Fisher test value, p-valule: significance level

value < 0.001 using the F statistics. The p-value is the observed significance probability of obtaining a greater F value by chance alone if a model fits no better than the overall response mean. The lower the p-value, the more significant the QSAR equation. As a result of the lowest p-value and the highest cross-validated r^2, an optimum fit was found to require three descriptors. This equation produces the best description for the activity of the fluorovinyloxyacetamides.

$$\begin{aligned}\text{Activity} = 5.365 \\ &-0.350\ (\text{Log P}) \\ &-0.524\ (\text{RadOfGyration}) \\ &+0.0577\ (\text{Dipole-Z})\end{aligned} \quad (2)$$

(r^2=0.853, Cross-validated r^2=0.832, F=102.684, p-value=4.52E-22)

The good relationship between the observed and the GFA predicted -logEC_{50} for training set is shown in Table 4 and Figure 3. The residual of those values was 0.003-0.284 and mean residual is 0.09095. The test set (Table 5) which is composed of 14 compounds were used to prove prediction ability of this equation. The prediction results of test set using equation (2) are depicted in Figure 4 and Table 6. These have also small error value that represents the prediction ability of final QSAR equation.

The Figure 5 represents the descriptors selected in the elite population of this models in the GFA calculation. Each curve of descriptor usage versus crossover operation number reaches a plateau after about 1000 crossovers in the GFA procedure which indicate a convergent optimization of QSAR equations. The most often used descriptor is clearly logP, which is found in about 30% of all QSAR models. It is considered that logP played an important role on the biological activity of molecule that might be concerned with herbicides distribution. That is, herbicides should be solved in water environment to penetrate into plant. Generally, LogP value is inversely proportional to solubility of water environment. Other descriptors, radius of gyration and dipole-Z in equation (2), were rarely selected in the elite population. However, they were effective descriptors combined with logP in equation (2). Radius of gyration confirms the significance of steric hinderance caused by the size of functional groups (R_1, R_2, R_3). Dipole-Z accounts for dipole-dipole interaction of functional group R_1. They may be related to bind between drug and receptor because drug size and charge distributions are essential factors to bind active site of receptor molecule.

The multicolinearity of three descriptors in equation (2) is represented by variance inflation factor (VIF). The effect of multicolinearity is to inflate the variance of the least squares estimator and possibly any predictions made, and also to restrict the generality and applicability of the QSAR model. VIF value is calculated from $1/1\text{-}r^2$, where r^2 is the multiple

Table 4. Observed and GFA predicted -log EC_{50} for Training Set

No.	Observed Activity (-logEC_{50})	GFA predic-tion	GFA residual	No.	Observed Activity (-logEC_{50})	GFA predic-tion	GFA residual
1	2.114	2.132	-0.019	31	1.937	2.009	-0.072
2	1.738	1.798	-0.061	32	1.259	1.311	-0.052
3	1.674	1.901	-0.227	33	1.517	1.552	-0.035
4	1.574	1.888	-0.315	34	1.360	1.345	0.015
5	1.807	1.706	0.101	35	1.309	1.553	-0.244
6	1.738	1.762	-0.024	36	1.860	1.796	0.065
7	2.102	2.119	-0.017	37	1.608	1.457	0.151
8	1.867	1.746	0.115	38	2.036	1.830	0.205
9	1.992	1.954	0.038	39	1.070	1.164	-0.093
10	1.910	1.936	0.026	40	1.470	1.537	-0.067
11	2.036	2.111	-0.075	41	1.673	1.735	-0.062
12	2.066	2.101	-0.035	42	1.806	1.925	-0.019
13	1.614	1.862	-0.247	43	1.581	1.524	0.057
14	2.408	2.470	-0.062	44	1.574	1.449	0.125
15	2.366	2.153	0.213	45	1.259	1.336	-0.078
16	2.398	2.188	0.210	46	1.618	1.643	-0.035
17	1.860	1.880	-0.020	47	1.360	1.304	0.056
18	1.937	1.914	0.023	48	1.896	1.749	0.147
19	2.046	1.949	0.098	49	1.672	1.577	0.095
20	2.387	2.400	-0.013	50	2.143	2.207	-0.063
21	2.169	2.183	-0.013	51	1.672	1.813	-0.141
22	2.030	1.866	0.165	52	1.579	1.569	0.010
23	1.737	1.746	-0.009	53	1.548	1.406	0.142
24	1.896	2.014	-0.118	54	1.670	1.708	-0.038
25	2.387	2.238	0.150	55	1.988	2.006	-0.018
26	1.941	1.924	0.017	56	1.691	1.774	-0.083
27	1.763	1.938	-0.175	57	1.670	1.653	0.017
28	2.034	1.942	0.092				
29	1.896	1.719	0.176				
30	2.209	2.167	0.042				

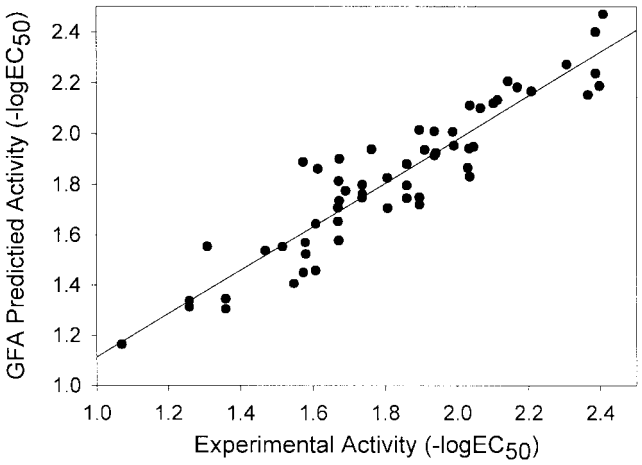


Figure 3. Activity values predicted by GFA Model of Training Set.

correlation coefficient of one descriptor's effect regressed on the remaining molecular descriptors. If VIF value is larger than 5, information of descriptors can be hidden by correlation of descriptors. In this model, the VIF value of these descriptors are 1.0155 (dipole-Z), 1.0292 (radius of gyration), 1.0229 (logP). Therefore, these descriptors showed no intercorrelation.

Relations between descriptors and activity can be seen from Figure 6. Solid and Dot lines in Figure 6 depict respectively regression line and 95% confidence level of regression line. The molecules 12, 27, 45, 54 show great deviation in

Table 5. Structures of Fluorovinyloxyacetamides in Test Set

No.	R_1	R_2	R_3
1	4-CH_3O-C_6H_4	CH_3	3,5-$(CH_3)_2$-C_6H_3
2	4-CH_3O-C_6H_4	CH_3	4-CH_3O-C_6H_4
3	4-CH_3O-C_6H_4	CH_3	4-Cl-C_6H_4
4	4-F-C_6H_4	CH_3	3-CH_3-C_6H_4
5	4-F-C_6H_4	CH_3	3,5-$(CH_3)_2$-C_6H_3
6	4-F-C_6H_4	CH_3	3-CH_3O-C_6H_4
7	C_6H_5	C_2H_5	4-CF_3-C_6H_4
8	4-CH_3-C_6H_4	C_2H_5	4-F-C_6H_4
9	4-CH_3O-C_6H_4	C_2H_5	4-F-C_6H_4
10	2-Cl-C_6H_4	C_2H_5	4-CH_3-C_6H_4
11	3-Cl-C_6H_4	C_2H_5	C_6H_5
12	4-Cl-C_6H_4	C_2H_5	C_6H_5
13	C_6H_5	*i*-C_3H_7	4-CH_3-C_6H_4
14	C_6H_5	*i*-C_3H_7	3,4-OCH_2O-C_6H_4

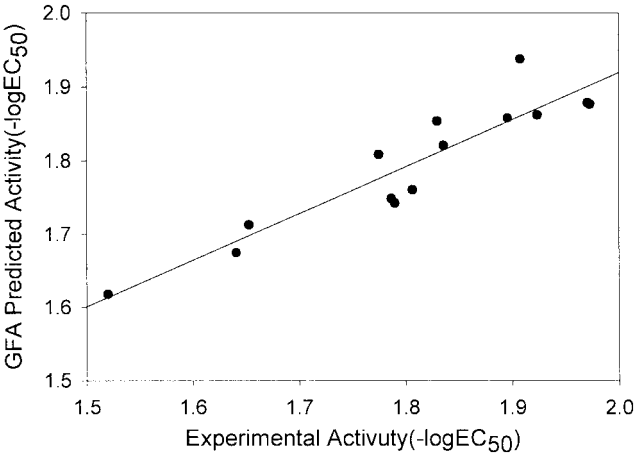


Figure 4. Activity values predicted by GFA Model of Test Set.

Table 6. Observed and GFA Predicted -log EC_{50} for Test Set

No.	Observed Activity (-logEC_{50})	GFA predic-tion	GFA residual	No.	Observed Activity (-logEC_{50})	GFA predic-tion	GFA residual
1	1.641	1.675	-0.034	8	1.836	1.821	0.015
2	1.924	1.862	0.062	9	1.830	1.854	-0.024
3	1.807	1.761	0.046	10	1.520	1.678	-0.158
4	1.971	1.879	0.092	11	1.973	1.877	0.096
5	1.653	1.713	-0.060	12	1.775	1.809	-0.034
6	1.908	1.938	-0.030	13	1.790	1.743	0.047
7	1.787	1.749	0.038	14	1.896	1.858	0.038

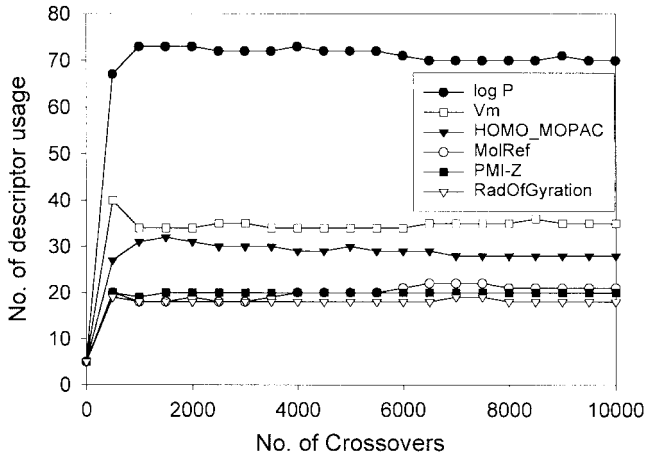


Figure 5. Population of GFA Selected Descriptors.

correlation between logP and -logEC_{50} (Figure 6(a)). On the other hands, these molecules show good correlations between RadOfGyration and -logEC_{50} or dipole-Z and -logEC_{50} (Figure 6(b) and (c)). In the case of the molecules 14, 22, 32, 39, they express great deviation with RadOfGyration (Figure 6(b)) while they express good correlations with logP (Figure 6(a)). This correlation patterns represent that these descriptors can describe biological activities effectively.

According to these descriptors, this fluorovinyloxyacetamide herbicides have important characteristics such as great plant penetration ability and molecular size. The highly active molecules might have smaller logP, smaller radius of gyration and larger dipole moment than less active molecules.

Further analysis of this data indicated that simple calculation of logP, radius of gyration, dipole moment might predict the biological activity of this herbicides.

Acknowledgment. This work was supported by the Korea Research Foundation (1997-012-D00035).

References

1. Matsumoto, H. *Proceedings I(A), Weed and Environmental Impact; The 17th Asian-Pacific Weed Science Society Conference*, **1999**, 39.
2. Ito, S.; Kamochi, A.; Sawada, K.; Goto, T.; Yasui, K. *Proceedings, The 12th Asian-Pacific Weed Science Society Conference*, **1989**, 255.
3. Kim, B. T.; Park, N. K.; Hong, K. S.; Park, J. E.; Kwon, Y. W. *PCT/KR 99/00116*.
4. Kim, B. T.; Park, N. K.; Hong, K. S.; Park, J. E. *KR Pat. 98-8750*.
5. Kim, B. T.; Park, N. K.; Hong, K. S.; Kwon, Y. W. *KR Pat. 98-8751*.
6. Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616.
7. Fujita, T. *QSAR and Drug Design; New Developments and Applications*, Elsevier: Amsterdam, 1995.
8. Lee, D. L.; Kollman, P. A.; Marsh, F. J.; Wolff, M. E. *J. Med. Chem.* **1977**, *20*, 1139.
9. Hugo Kubinyi *3D QSAR in Drug Design: Theory, Methods and Applications*; ESCOM: Leiden, 1993.
10. Zhou, Y.-x.; Xu, L.; Wu, Y.-p.; Liu, B.-l. *J. Chem. Intell, Lab. Sys.* **1999**, *45*, 95.
11. Hou, T. J.; Wang, J. M.; Liao, N.; Xu, X. J. *J. Chem. Inf. Comput Sci.* **1999**, *39*, 775.
12. Gasteiger, J.; Zupan, J. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 503.
13. David, J.; David, W. S. *Bioorg. Med. Chem. Lett.* **1992**, *2*, 213.
14. Leardi, R.; Boggia, R.; Terrile, M. *J. Chemomertrics*

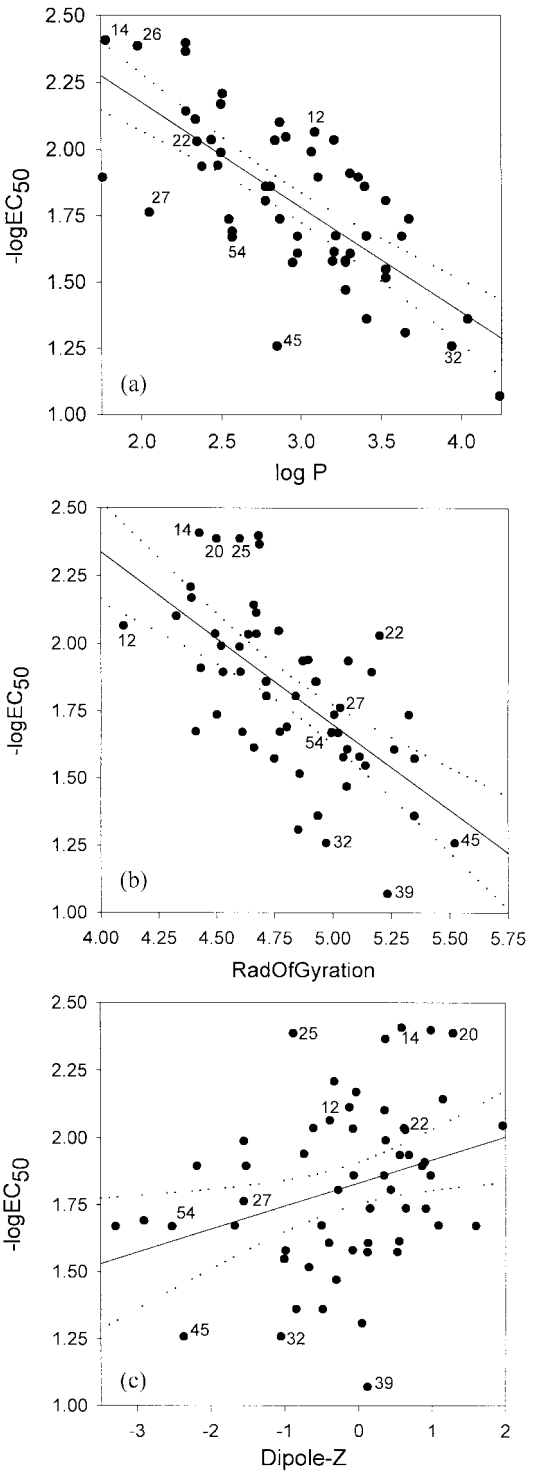


Figure 6. Plots of -logEC_{50} vs. (a) logP, (b) radius of gyration and (c) dipole-Z.

1992, *6*, 267.
15. David, R.; Hofinger, A. J. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 854.
16. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
17. Clare, B. W. *J. Med. Chem.* **1998**, *41*, 3845.
18. Baumann, K. *J. Anal. Chem*. **1999**, *18*, 36.
19. Sweet, R. M.; Eisenberg, D. *J. Mol. Biol.* **1983**, *171*, 479.
20. *Gaussian 94, Revision D.2*; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 1995.
21. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
22. Gundertofte, K.; Liljefors, T.; Norrby, P. O.; Pettersson, I. *J. Comput. Chem*. **1996**, *17*, 429.
23. Crandell, C. W.; Smith, D. H. *J. Chem. Inf. Comput. Sci.* **1983,** *23*, 186.
24. Brint, A. T.; Willett, P. *J. Mol. Graphics* **1987**, *5*, 200.
25. *Cerius2 QSAR+ (version 3.8)*; Molecular Simulations Inc.: 1998.
26. Topliss, J. G.; Edwards, R. P. *J. Med. Chem.* **1979**, *22*, 1238.