2-Difluoromethylthio-4,6-bis(monoalkylamino)-1,3,5-triazines as Inhibitors of Hill Reaction: A QSAR Study with Orthogonalized Descriptors

Milan Šoškić*

Department of Chemistry, Faculty of Agriculture, University of Zagreb, HR-41000 Zagreb, The Republic of Croatia

Dejan Plavšić and Nenad Trinajstić

The Ruđer Bošković Institute, P.O.B. 1016, HR-41001 Zagreb, The Republic of Croatia

Received October 25, 1995[⊗]

The QSAR models for herbicidal activity of 2-difluoromethylthio-4,6-bis(monoalkylamino)-1,3,5-triazines have been developed, using orthogonalized molecular connectivity indices. The orthogonalization of the descriptor variables were performed by a recently proposed dominant component method and well established principal component analysis. The three parameter model based on dominant component descriptors and four parameter principal component regression model are similar in their statistical properties. However, direct correspondence between dominant component descriptors and original variables makes the former model easier to interpret. The obtained models are simpler than the earlier proposed six parameter model, based on *n*-octanol/water partition coefficients and Taft's steric substituent constants, but are as accurate as this model.

1. INTRODUCTION

The strong correlation between predictor variables (multicollinearity) may lead to misinterpretation of the resulting regression model.[1] Several techniques have been proposed to overcome this problem.[2] One of them is well-known principal component regression,[2−4] where orthogonal mathematical transformations of original descriptors (principal components) are used as predictor variables. Recently, a different orthogonalization method, dominant component analysis, has been proposed.[5] The essence of this approach is sequential removal of overlapping information from a set of correlated descriptors. An important result of the modeling with orthogonal descriptors is the stability of estimated regression coefficients, which permits one to assess the relative importance of an individual descriptor directly. Thus, even when correlation between the variables is not high, the orthogonalization of the descriptors may facilitate the interpretation of the model.

The objective of this study was to compare the modeling performances of the two aforementioned methods with each other and with the standard regression approach, based on the nonorthogonalized descriptors. The comparative analysis were performed on 2-difluoromethylthio-4,6-bis(monoalkylamino)-1,3,5-triazines,[6] which act as inhibitors of photosynthesis.[7] Their mechanism of action is based on displacing of secondary electron acceptor Q_B from its binding site on protein D1, a subunit of photosystem II complex. According to the present concept, triazines are classified as serine type inhibitors, which indicates that these molecules are supposed to orient themselves toward serine 264 in the receptor cavity.[7,8]

The inhibitory potencies of the triazine derivatives were modeled with molecular connectivity indices[9−11] as structural descriptors. These indices were successfully applied in numerous QSAR (quantitative structure−activity relationships) studies,[9−12] including the modeling of herbicidal activity of photosystem II inhibitors.[13,14]

2. METHOD OF CALCULATION AND EXPERIMENTAL DATA

2.1. Molecular Connectivity Indices.[9−11] The zero and first-order molecular connectivity indices are calculated from the hydrogen suppressed graph (structural formula) of an alkylamino substituent, in the usual way

$$^0\chi = \sum(\delta_i)^{-0.5} \quad (1)$$

$$^1\chi = \sum(\delta_i \cdot \delta_j)^{-0.5} \quad (2)$$

where δ stands for the number of non-hydrogen atoms bound to a particular skeletal atom and the sums are over all non-hydrogen atoms and over all bonds between adjacent non-hydrogen atoms, respectively.

2.2. Dominant Component Analysis.[5] A descriptor variable from the set of correlate variables can be made orthogonal, by removing the part of its information content that it shares with the other variables in the set. The order in which variables are orthogonalized is important, because it strongly affects the information content of so obtained orthogonal descriptors.[15−17] For the set of three correlated variables and the arbitrarily selected order of orthogonalization, the construction of orthogonal descriptors goes as follows: The first orthogonal descriptor is always an original variable. Suppose it is variable x. The residual of variable y, when it is regressed against variable x (res y/x) is the second orthogonal descriptor $\Omega(y)$. To make variable z orthogonal, variable z is first regressed against variable x. The residual obtained in this regression (res z/x) is then regressed against $\Omega(y)$. The residual of the last regression (res z/x/res y/x) is the third orthogonal descriptor $\Omega(z)$. A larger set of descriptor variables can be orthogonalized in the analogous way.

[⊗] Abstract published in *Advance ACS Abstracts,* January 1, 1996.

0095-2338/96/1636-0146$12.00/0 © 1996 American Chemical Society

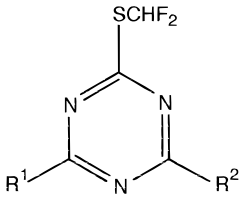

Figure 1. Structure of 2-difluoromethylthio-4,6-bis(monoalkylamino)-1,3,5-triazine derivatives.

2.3. Principal Component Analysis.[1,2] The aim of principal component analysis is to transform a set of correlated variables into a smaller set of uncorrelated variables (e.g., principal components) which accounts for most of the variance in the original data set. Principal components are derived from the correlation or covariance matrix and represent linear weighted sums of original (standardized) variables. Principal components are calculated in such a way that the first principal component explains the largest part of the total variance of the data, the second explains the largest part of the remaining variance, and so on until all the variance is explained. The amount of the variance explained by a principal component is directly proportional to the magnitude of its eigenvalue. Relationships between the original variables and the principal components are given by their loadings (correlation coefficient between an original variable and a certain principal component).

Since principal components are not always easy to interpret, they are often rotated to obtain a simplified loading structure. The orthogonal Varimax rotation[18] is most frequently used in that sense.

2.4. QSAR Calculation. The calculations of molecular connectivity indices, the orthogonalization of descriptor variables and multiple regression analysis were carried out by our own programs, written in Fortran 77. (The program for the orthogonalization of the variables will be reported in one of our future studies.) Principal component analysis was performed using the statistical package SAS (statistical analysis system).[19]

In all regression equations *n* is number of compounds used in the analysis, r^2 is the squared correlation coefficient, *s* is standard deviation of the estimates, and *F* is *F*-test. PRESS is the sum of squared prediction errors from the leave-one-out cross-validation analysis. The standard error of coefficients is given in parentheses. The significance of all the derived models and the regression coefficients is above 95% level or higher.

2.5. Experimental Data. The inhibitory activities of the triazine derivatives in Chlorella are taken from the study of Morita, Nagare, and Hayashi.[6] They are expressed as pI_{50}, the negative logarithm of concentration required for 50% inhibition of Hill reaction.

3. RESULTS AND DISCUSSION

3.1. Model Based on the Nonorthogonalized Descriptors. The inhibitory potencies of the triazine derivatives (Figure 1) examined in this study are displayed in Table 1, together with the substituent molecular connectivity indices used as structural descriptors. Preliminary analysis has shown that more accurate models can be obtained if the influence of substituents R^1 and R^2 on the activity of the triazine molecules are considered separately. Therefore, the molecular connectivity indices were calculated for the substituents R^1 and R^2. Initially, we searched for the best QSAR model using original, nonorthogonalized molecular connectivity indices as predictor variables. In the process of the model building the following structural descriptors were screened: the zero- and first-order molecular connectivity indices, their reciprocal and/or square terms and the third-order cluster molecular connectivity index. The zero- and first-order molecular connectivity indices are closely related to the size of substituents, and the third-order cluster molecular connectivity index is very sensitive to the branching pattern of substituents.[9,11] It was found by multiple regression analysis that the following three parameter model best describes the variation in the inhibitory potencies of the triazine derivatives

Table 1. Substituent Molecular Connectivity Indices and Observed and Calculated pI_{50} Values of 2-Difluoromethylthio-4,6-bis(monoalkylamino)-1,3,5-triazines

compounds					pI_{50}		
no.	R^1	R^2	$^1\chi(R^1)$	$^0\chi(R^2)$	exp.	eq 3 eq 4	eq 5
1.	NH_2	NH_2	0.000	1.000	3.82	3.86	3.76
2.	NH_2	$NHCH_3$	0.000	2.000	5.20	5.14	5.16
3.	NH_2	NHC_2H_5	0.000	2.707	5.34	5.47	5.50
4.	NH_2	NH-i-C_3H_7	0.000	3.577	5.83	5.70	5.71
5.	$NHCH_3$	$NHCH_3$	1.000	2.000	6.01	6.16	6.21
6.	$NHCH_3$	NHC_2H_5	1.000	2.707	6.39	6.50	6.54
7.	$NHCH_3$	NHC_3H_7	1.000	3.414	6.75	6.70	6.73
8.	$NHCH_3$	NH-i-C_3H_7	1.000	3.577	6.76	6.73	6.76
9.	$NHCH_3$	NHC_4H_9	1.000	4.121	6.74	6.82	6.83
10.	$NHCH_3$	CH-s-C_4H_9	1.000	4.284	6.76	6.85	6.85
11.	$NHCH_3$	NH-t-C_4H_9	1.000	4.500	6.78	6.88	6.87
12.	$NHCH_3$	NHC_5H_{11}	1.000	4.828	7.12	6.91	6.89
13.	NHC_2H_5	NHC_2H_5	1.414	2.707	6.82	6.55	6.60
14.	NHC_2H_5	NHC_3H_7	1.414	3.414	6.74	6.75	6.78
15.	NHC_2H_5	NH-i-C_3H_7	1.414	3.577	6.89	6.78	6.81
16.	NHC_2H_5	NHC_4H_9	1.414	4.121	6.95	6.88	6.88
17.	NHC_2H_5	NH-i-C_4H_9	1.414	4.284	7.01	6.90	6.90
18.	NHC_2H_5	NH-s-C_4H_9	1.414	4.284	6.87	6.90	6.90
19.	NHC_2H_5	NH-t-C_4H_9	1.414	4.500	6.97	6.93	6.92
20.	NHC_2H_5	NHC_5H_{11}	1.414	4.828	6.94	6.97	6.94
21.	NHC_2H_5	NHC_6H_{13}	1.414	5.536	7.21	7.03	6.98
22.	NHC_2H_5	NHC_7H_{15}	1.414	6.243	7.01	7.09	7.00
23.	NHC_2H_5	NHC_8H_{19}	1.414	6.950	6.81	7.13	7.00
24.	NHC_3H_7	NHC_3H_7	1.914	3.414	6.45	6.52	6.55
25.	NH-i-C_3H_7	NHC_3H_7	1.732	3.414	6.75	6.64	6.67
26.	NH-i-C_3H_7	NH-i-C_3H_7	1.732	3.577	6.75	6.67	6.70
27.	NH-i-C_3H_7	NHC_4H_9	1.732	4.121	6.71	6.77	6.77
28.	NH-i-C_3H_7	NH-s-C_4H_9	1.732	4.284	6.88	6.90	6.79
29.	NH-i-C_3H_7	NH-t-C_4H_9	1.732	4.500	6.70	6.82	6.81
30.	NH-i-C_3H_7	NHC_5H_{11}	1.732	4.828	6.69	6.86	6.84

$$pI_{50} = 6.41(\pm 0.13) - 2.56(\pm 0.20)^0\chi(R^2) + 1.67(\pm 0.16)^1\chi(R^1) - 0.64(\pm 0.08)(^1\chi)^2(R^1) \quad (3)$$

$$n = 30 \quad r^2 = 0.967 \quad s = 0.134 \quad F^{3,26} = 254.5 \quad \text{PRESS} = 0.625$$

A plot of the observed inhibitory potencies of the triazine derivatives versus pI_{50}'s calculated by eq 3 is given in Figure 2.

It follows from eq 3 that inhibitory potency of the triazine derivatives is mainly controlled by the size of substituents R^1 and R^2. Exponential relationship between pI_{50}'s and the molecular connectivity indices suggests that the substituents interact with lipophilic areas of the receptor of the limited dimensions. It appears that substituents R^1 bind to a narrow

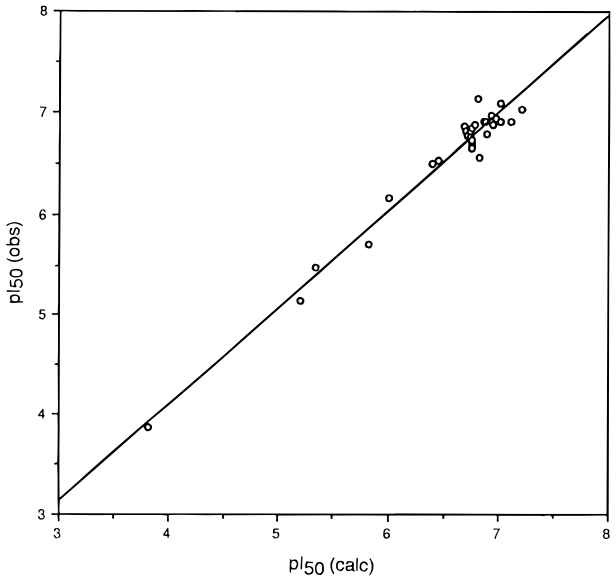

Figure 2. The observed versus calculated pI_{50} values for the studied triazine derivatives using eq 3 or 5.

zone of the receptor (which optimally accommodates ethylamino group), while the binding region for the substituents R^2 is more tolerant to the bulkier groups.

It should be emphasized that eq 3 is simpler than corresponding physicochemical model[6] based on *n*-octanol/water partition coefficient (log *P*) and Taft's steric substituent constants (E_s). This model (eq 4) has six predictor variables and very similar statistics to that of eq 3.

$$pI_{50} = 2.80 - 0.09(\log P)^2 + 0.96 \log P - 0.76(E_s(R^1))^2 - 1.48E_s(R^1) - 0.25(E_s(R^2))^2 - 1.04\,E_s(R^2) \quad (4)$$

$$n = 30 \quad r^2 = 0.973 \quad s = 0.130 \quad F^{6,23} = 134.7 \quad \text{PRESS} = 0.672$$

It is worth noting that the above model can be improved by the orthogonalization of the physicochemical variables. For instance, using the dominant component procedure the four parameter model of similar quality can be derived. In further analysis, however, we will concentrate on the models based on molecular connectivity indices.

3.2. Model Based on the Dominant Component Descriptors. In the next step, we formulated the QSAR models based on the orthogonal descriptors, which were previously derived by dominant and principal component analysis. For simplicity, we subjected to the orthogonalization only the predictor variables of regression eq 3 plus the linear term of the zero-order molecular connectivity index for substituent R^2. The latter was excluded from the regression eq 3 because its significance is below 95% (>90%).

The order in which predictor variables enter the orthogonalization process is of critical importance in dominant component analysis.[17] The different ordering of the variables results in the different orthogonal descriptors, or, put in a different way, the orthogonalization process dictates the information content of the resulting variables. Naturally, a whole set of the orthogonal descriptors, regardless of the sequence of orthogonalization, contains the same total information content as the set of original ones, since the orthogonalization is nothing more than removing of superfluous information from the correlated variables.

Table 2. All Possible Regression Equations for a Given Set of Predictor Variables[a]

	r^2	s
$^1\chi(R^1)$	0.579	0.460
$(^1\chi)^2(R^1)$	0.344	0.575
$^0\chi(R^2)$	0.514	0.495
$1/^0\chi(R^2)$*	0.803	0.316
$^1\chi(R^1),(^1\chi)^2(R^1)$	0.769	0.348
$^1\chi(R^1),1/^0\chi(R^2)$*	0.889	0.241
$^1\chi(R^1),^0\chi(R^2)$	0.726	0.378
$(^1\chi)^2(R^1),1/^0\chi(R^2)$	0.836	0.293
$(^1\chi)^2(R^1),^0\chi(R^2)$	0.615	0.448
$1/^0\chi(R^2),^0\chi(R^2)$	0.803	0.321
$^1\chi(R^1),(^1\chi)^2(R^1),^0\chi(R^2)$	0.847	0.288
$^1\chi(R^1),(^1\chi)^2(R^1),1/^0\chi(R^2)$*	0.967	0.134
$^1\chi(R^1),^0\chi(R^2),1/^0\chi(R^2)$	0.891	0.243
$(^1\chi)^2(R^1),^0\chi(R^2),1/^0\chi(R^2)$	0.837	0.298
$^1\chi(R^1),(^1\chi)^2(R^1),^0\chi(R^2),1/^0\chi(R^2)$	0.971	0.130

[a] The best combinations of variables are indicated with an asterisk.

The order of orthogonalization can be defined in various ways.[15,16] We adopted here an interactive approach, where the orthogonalization of the variables parallels the process of the model building. The order of orthogonalization, for the analyzed triazine derivatives, has been deduced from the list of all possible regression subsets (displayed in Table 2), using r^2 and s as criteria. It corresponds to the sequence in which the predictor variables from the final (best) regression equation were introduced in the process of the model building. The following order of orthogonalization was established: $1/^0\chi(R^2)$, $^1\chi(R^1)$, $(^1\chi)^2(R^1)$, and $^0\chi(R^2)$. Although the variable $^0\chi(R^2)$ does not contribute significantly to the explained variance in pI_{50}'s, it was orthogonalized for the subsequent comparison with the nonorthogonal descriptors and principal components. The model based on the orthogonal descriptors obtained by the method of dominant component analysis is as follows

$$pI_{50} = 7.73(\pm 0.05) - 3.98(\pm 0.16)/\Omega(^0\chi(R^2)) + 0.46(\pm 0.06)\Omega(^1\chi(R^1)) - 0.64(\pm 0.08)\Omega((^1\chi)^2(R^1)) \quad (5)$$

$$n = 30 \quad r^2 = 0.967 \quad s = 0.134 \quad F^{3,26} = 254.5 \quad \text{PRESS} = 0.625$$

Ω stands for a dominant component (DC) descriptor and the original variable from which the descriptor is derived is given in parentheses. The DC descriptors in the model are arranged according to the sequence of their orthogonalization. It enables one to determine the importance and meaning of the orthogonal predictor variables by a simple inspection of the model. Namely, in the interactive approach, the sequence of orthogonalization of variables coincides with the decreasing order of their relative importance. On the other hand, the orthogonalized descriptors are not only mutually uncorrelated among themselves but also mutually uncorrelated with the original variables that precede them in the orthogonalization process (Table 3). Thus, direct interpretation of DC predictors in terms of the starting variables is possible. Generally, a DC descriptor can be defined as the part of information of an original variable that is unique concerning the preceding variable in the orthogonalization procedure.

Comparison of the regression eqs 3 and 5 shows that statistical parameters of these models (i.e., r^2, s, and F) are the same. This is not surprising,[16,20] since, as we pointed

Table 3. Correlation Matrix between Dominant Component Descriptors (Ω) and Original Variables

	$1/^0\chi(R^2)$	$^1\chi(R^1)$	$(^1\chi)^2(R^1)$	$^0\chi(R^2)$
$\Omega(1/^0\chi(R^2))$	1.00	−0.58	−0.48	−0.82
$\Omega(^1\chi(R^1))$	0.00	0.81	0.83	0.04
$\Omega((^1\chi)^2(R^1))$	0.00	0.00	0.30	−0.03
$\Omega(^0\chi(R^2))$	0.00	0.00	0.00	0.58

out, the original (eq 3) and corresponding DC descriptors (eq 5) have the same total information content. Moreover, the coefficients of the orthogonal predictors in eq 5 can be extracted[16,20] from the gradual inclusion of the starting variables (eq 3) in the model. The same is true for the contribution (r^2) of a DC descriptor to the regression. It is equal to the increase in r^2, when the corresponding original variable is introduced in the nonorthogonal model. Consequently, eq 5 can be interpreted in the same terms as the eq 3.

3.3. Principal Component Regression Model. Principal components (PCs) were extracted from the correlation matrix of standardized variables (Table 4), and their loadings are presented in the upper part of Table 5. It is seen that all four variables have approximately equal high loadings on the first PC and moderate or small loadings on the other components. To make interpretation easier, the PCs were rotated, using Varimax procedure. Although the structure of the loadings was considerably simplified by the rotation, the complete separations of the variables were not attained (lower part of Table 5). The first rotated PC is highly loaded by the variables $^1\chi(R^1)$ and $(^1\chi)^2(R^1)$. The second and the third PC are dominated by $^0\chi(R^2)$ and $1/^0\chi(R^2)$, respectively, and the fourth PC is characterized by small loadings for $^1\chi(R^1)$ and $(^1\chi)^2(R^1)$ and near zero loadings for $^0\chi(R^2)$ and $1/^0\chi(R^2)$.

The question which PCs to include in regression analysis is still without definite answer.[21] There is a tendency to use only the most informative PCs (PCs whose eigenvalues exceed a prescribed cut-off level), but the problem associated with this strategy is that the large variance-components are not necessarily the best predictors of the dependent variable. Here, all Varimax rotated PCs were subjected to regression, and the contribution of an individual PC to the explained variance was used as a criterion for its inclusion in the model. To obtain the model of similar predictive ability as the previous ones (eqs 3 and 5), all four PCs had to be introduced in eq 6

$$pI_{50} = 6.56(\pm0.02) - 0.48(\pm0.02)PC3 + 0.33(\pm0.02)PC1 - 0.28(\pm0.02)PC2 - 0.24(\pm0.02)PC4 \quad (6)$$

$$n = 30 \quad r^2 = 0.971 \quad s = 0.130 \quad F^{4,25} = 206.6 \quad PRESS = 0.980$$

The three-parameter model (PC3, PC1, and PC2) is of significantly lower statistical quality ($r^2 = 0.852$, $s = 0.285$). A somewhat better three-parameter model can be obtained if the unrotated PCs (PC1, PC4, and PC3) are used as predictor variables ($r^2 = 0.933$, $s = 0.190$).

To examine why the additional (the fourth) predictor variable is needed in the model based on the PCs, we correlated the Varimax rotated PCs with the dominant component descriptors (Table 6). Parallel to this, we compared the individual contributions of the PCs and corresponding DC descriptors to the explained variance (Table 7). It is noteworthy that there is significant overlapping in the information contents of these descriptors, but at the same time they differ considerably in the amount of the information that is relevant for describing variability in pI_{50}'s. The latter is the consequence of the fact that DC descriptors were derived by taking into account their correlation with the dependent variable, while the PCs were generated independently of the response variable. The result of this is that almost all relevant information is concentrated on the first three DC descriptors, while in PCA approach this information is distributed over four PCs. Except for PC2, which corresponds to $\Omega(^0\chi(R^2))$, the order of importance of the other PCs is in accord with that of DC descriptors. Since PC2 practically does not correlate with $\Omega(^1\chi(R^1))$ and $\Omega((^1\chi)^2(R^1))$ and contribution of $\Omega(^0\chi(R^2))$ to explained variability in pI_{50}'s can be ignored, it follows that the relevant information in PC2 mostly comes from $\Omega(1/^0\chi(R^2))$. Hence, basically the same conclusion can be drawn from eq 6 as from the previous models.

Table 4. Correlation Matrix between Four Original Variables

	$1/^0\chi(R^2)$	$^1\chi(R^1)$	$(^1\chi)^2(R^1)$	$^0\chi(R^2)$
$1/^0\chi(R^2)$	1.00			
$^1\chi(R^1)$	−0.58	1.00		
$(^1\chi)^2(R^1)$	−0.48	0.95	1.00	
$^0\chi(R^2)$	−0.82	0.51	0.41	1.00

Table 5. Principal Component Loadings Before (Upper Part) and After Varimax Rotation (Lower Part) Plus Corresponding Eigenvalues and Percents of Explained Total Variance

	PC1	PC2	PC3	PC4
Unrotated Principal Components				
$^1\chi(R^1)$	0.905	0.398	−0.003	−0.150
$^0\chi(R^2)$	0.797	−0.530	0.290	0.003
$1/^0\chi(R^1)$	−0.842	0.443	0.308	−0.021
$(^1\chi)^2(R^2)$	0.846	0.513	0.037	0.137
eigenvalue	2.879	0.899	0.180	0.042
% of var	71.98	22.47	4.50	1.04
Varimax Rotated Principal Components				
$^1\chi(R^1)$	0.924	−0.235	−0.249	−0.168
$^0\chi(R^2)$	0.221	−0.906	−0.361	−0.010
$1/^0\chi(R^2)$	−0.280	0.509	0.814	0.010
$(^1\chi)^2(R^1)$	0.967	−0.161	−0.154	0.122
eigenvalue	1.917	1.161	0.878	0.043
% of var	47.93	29.03	21.96	1.08

Table 6. Correlation Matrix between Principal Components (PC) and Dominant Component Descriptors (Ω)

	PC1	PC2	PC3	PC4
$\Omega(^1\chi(R^1))$	0.94	0.08	0.28	−0.20
$\Omega(^0\chi(R^2))$	−0.07	−0.86	0.51	0.06
$\Omega(1/^0\chi(R^2))$	−0.28	0.51	0.81	0.01
$\Omega((^1\chi)^2(R^1))$	0.20	0.06	0.02	0.98

Table 7. Comparison of Contributions of Principal Components (PC) and Corresponding Dominant Component Descriptors (Ω) to the Explained Variability in pI_{50} Values

	r^2		r^2
PC3	0.466	$\Omega(1/^0\chi(R^2))$	0.803
PC1	0.226	$\Omega(^1\chi(R^1))$	0.086
PC2	0.160	$\Omega(^0\chi(R^2))$	0.004
PC4	0.119	$\Omega((^1\chi)^2(R^1))$	0.078

4. CONCLUSION

Inhibitory potencies of 2-difluoromethylthio-4,6-bis-(monoalkylamino)-1,3,5-triazines have been modeled, using nonorthogonalized and orthogonalized molecular connectivity indices, as structural descriptors. The orthogonalization of descriptor variables is of interest in QSAR studies, even when interrelatedness between predictor variables is not a problem, because of stability of the orthogonal regression models. The two methods, dominant component and principal component analysis, were used here for the orthogonalization of descriptors and QSAR models obtained by these methods were compared. The model based on dominant component descriptors and principal component regression model are similar in their statistical properties, but the former is easier to interpret. The reasons for this are (a) direct correspondence between dominant component descriptors and original variables and (b) ability of dominant component method to eliminate unimportant descriptors. In principal component approach such variables may,[22] as in our case (i.e., ${}^0\chi(R^2)$), aggravate the interpretation of the resulting model.

Applying molecular connectivity indices we were able to formulate simpler QSAR models, when compared to the existing physicochemical model, without losing accuracy.

ACKNOWLEDGMENT

This work was supported by the Ministry of Science and Technology of the Republic of Croatia through Grants 1-07-159 and 1-07-165.

REFERENCES AND NOTES

(1) Dillon, W. R.; Goldstein, M. Multivariate Analysis; Wiley: New York, 1984.

(2) Jolliffe, I. T. Principal Component Analysis; Springer-Verlag: New York, 1986.

(3) Lukovits, I. Quantitative Structure−Activity Relationships Employing Independent Quantum Chemical Indices. *J. Med. Chem.* **1983**, *26*, 1104−1109.

(4) Kubota, T.; Hanamura, J.; Kano, K.; Uno, B. Application of Principal Component Analysis to the Study of Quantitative Structure−Activity Relationships by Means of Multiple Regression Analysis. *Chem. Pharm. Bull.* **1985**, *33*, 1488−1495.

(5) Randić, M. Orthogonal Molecular Descriptors. *New J. Chem.* **1991**, *15*, 517−525.

(6) Morita, K.; Nagare, T.; Hayashi, Y. Quantitative Structure−Activity Relationships for Herbicidal 2-Difluoromethylthio-4-6-bis(monoalkylamino)-1,3,5-triazines. *Agric. Biol. Chem.* **1987**, *51*, 1955−1957.

(7) Draber, W.; Tietjen, K.; Kluth, J. F.; Trebst, A. Herbicides in Photosynthesis Research. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 1621−1633.

(8) Trebst, A. The Three-Dimensional Structure of the Herbicide Binding Niche on the Reaction Center Polypeptides of Photosystem II. *Z. Naturforsch.* **1987**, *42c*, 742−750.

(9) Kier, L. B.; Hall, L. H. Molecular Connectivity in Structure−Activity Analysis; Research Studies Press Ltd: Letchworth, 1986.

(10) Trinajstić, N. Chemical Graph Theory, 2nd ed.; CRC Press: Boca Raton, FL, 1992.

(11) Sabljić, A.; Horvatić, D. Graph III: A Computer Program for Calculating Molecular Connectivity Indices on Microcomputers. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 292−295.

(12) Šoškić, M.; Sabljić, A. Herbicidal Selectivity of (*E*)-3-(2,4-Dichlorophenoxy)acrylates: QSAR Study with Molecular Connectivity Indices. *Pestic. Sci.* **1993**, *39*, 245−250.

(13) Šoškić, M.; Sabljić, A. Inhibition of Hill Reaction by 3-Alcoxyuracil Derivatives: QSAR Study with Topological Indices. *Croat. Chem. Acta.* **1987**, *60*, 755−763.

(14) Šoškić, M.; Sabljić, A. Inhibition of Hill Reaction by 2-Azido-s-triazine Derivatives: QSAR Study with Molecular Connectivity Indices. *Z. Naturforsch.* **1989**, *44c*, 255−261.

(15) Randić, M. Resolution of Ambiguities in Structure−Property Studies by Use of Orthogonal Descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311−320.

(16) Randić, M.; Trinajstić, N. Isomeric Variations in Alkanes: Boiling Points of Nonanes. *New. J. Chem.* **1994**, *18*, 179−189.

(17) Lučić, B.; Nikolić, S.; Trinajstić, N.; Juretić, D. The Structure−Property Models Can Be Improved Using the Orthogonalized Descriptors. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 532−538.

(18) Kaiser, H. F. The Varimax Criterion for Analytic Rotation in Factor Analysis. Psychometrika **1958**, *23*, 187−200.

(19) SAS Institute Inc., SAS/STAT User's Guide, Version 6, 4th ed.; SAS Institute Inc.: Cary, NC, 1989; Vol. 2.

(20) Amić, D.; Davidović-Amić, D.; Trinajstić, N. Calculation of Retention Times of Anthocyanins with Orthogonalized Topological Indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 136−139.

(21) Sutter, J. M.; Kalivas, J. H.; Lang, P. M. Which Principal Components to Utilize for Principal Component Regression. *J. Chemometrics* **1992**, *6*, 217−225.

(22) Randić, M. Correlation of Enthalphy of Octanes with Orthogonal Connectivity Indices. *J. Mol. Struct. (Theochem)* **1991**, *233*, 45−59.

CI950270E